Strong Gaussian approximation of metastable density-dependent Markov chains on large time scales

Adrien Prodhomme

Strong Gaussian approximation of metastable density-dependent Markov chains on large time scales

Adrien Prodhomme*

October 13, 2020

Abstract

Density-dependent Markov chains form an important class of continuous-time Markov chains in population dynamics. On any fixed time window $[0, T]$, when the scale parameter $K > 0$ is large such chains are well approximated by the solution of an ODE (the fluid limit), with Gaussian fluctuations superimposed upon it. In this paper we quantify the period of time during which this Gaussian approximation remains precise, uniformly on the trajectory, in the case where the fluid limit converges to an exponentially stable equilibrium point. We provide a new coupling between the density-dependent chain and the approximating Gaussian process, based on a construction of Kurtz using the celebrated Komlós-Major-Tusnády theorem for random walks. We show that under mild hypotheses the time $T(K)$ necessary for the strong approximation error to reach a threshold $\varepsilon(K) \ll 1$ is at least of order $\exp(VK\varepsilon(K))$, for some constant $V > 0$. This notably entails that the Gaussian approximation yields the correct asymptotics regarding the time scales of moderate deviations. We also present applications to the Gaussian approximation of the logistic birth-and-death process conditioned to survive, and to the estimation of a quantity modeling the cost of an epidemic.

1 Introduction and main result

Density-dependent Markov chains are widely used, in ecology, biology, chemistry and epidemiology, to model the evolution of populations. Let us cite [1, 6, 22] for numerous examples, including stochastic Lotka-Volterra models, chemical reaction networks and epidemic models. Such chains record the abundances of a finite set of populations, in interaction with one another. They involve a scale parameter $K > 0$, which can have different interpretations depending on the context (quantity of resources, volume of reaction, or total size of the population). As shown by Kurtz [20], density-dependent families $\left(N^K\,;\,K > 0\right)$ satisfy a functional law of large numbers and a central limit theorem. On a fixed time window $[0, T]$, when K is large the trajectory of the process $X^K = N^K/K$, called the density, is well approximated by the solution of an ODE (the fluid limit), with Gaussian fluctuations of order $1/\sqrt{K}$ superimposed upon it. Since in a number of applications, notably in ecology and evolution, the relevant periods of time are very long, we are led to the following question : on which time scales $T(K)$ does the Gaussian approximation of the trajectories remain valid ?

To answer this question, we first construct a coupling between the density X^K and its Gaussian approximation, based on a construction of Kurtz [21]. It relies on the possibility to represent density-dependent Markov chains using time-changed Poisson processes, combined with the powerful strong approximation theorem of Komlós, Major and Tusnády (KMT)[18, 19] for one-dimensional random walks. This theorem entails the existence of a coupling between a Poisson process P and a Brownian motion B, and constants $a, b, c > 0$ such that

$$\mathbf{P}\left(\sup_{0\le t\le T} |P(t) - t - B(t)| > c\log(T) + x\right) \le ae^{-bx} \tag{1}$$

*Institut Denis Poisson, Université de Tours, France ; CMAP, Ecole Polytechnique, France ; adrien.prodhomme@univ-tours.fr

for all $T \geq 1$ and $x \geq 0$ [13, Chapter 7, Corollary 5.3]. Generalizations of the KMT result to independent, nonidentically distributed, and multidimensional increments were obtained by Sakhanenko, Einhmahl and Zaitsev, see [15] for a review on the subject. More recently, KMT type results were obtained in various weakly-dependent cases, with applications to mixing dynamical systems and ergodic Markov chains, see [5, 16, 24] and the references therein.

In our context, the chain X^K is subject to a drift, given by the vector field of the limiting ODE. We focus on the case where the limiting ODE admits an exponentially stable equilibrium point. This is common in applications : let us mention coexistence equilibriums in competitive population models, endemic equilibriums in epidemic models, and chemical equilibriums. In this situation, near the equilibrium the drift tends to reduce the gap between X^K and its strong (path-by-path) Gaussian approximation. We show in our main result, Theorem 1.1, that it allows the Gaussian approximation to remain precise for very large periods of time.

Let us set the framework precisely. Fix $d \in \mathbf{N}^*$, and for all $e \in \mathbf{Z}^d \setminus \{0\}$, let β_e be a non-negative function defined on $\mathbf{R}^d$. For all $K > 0$, let N^K be a $\mathbf{Z}^d$-valued continous-time Markov chain, with transition rate from n to $m \neq n$ given by

$$q^K_{n,m} = K\beta_{m-n}(n/K). \tag{2}$$

The family $\left(N^K\,;\,K>0\right)$ is called a density-dependent family of Markov chains, associated to the functions β_e, $e \in \mathbf{Z}^d \setminus \{0\}$. For the sake of concision, a given N^K is called a density-dependent Markov chain. We make the following assumptions :

- There exists a finite, non empty subset E of $\mathbf{Z}^d \setminus \{0\}$ such that $\beta_e \equiv 0$ for all $e \notin E$.
- There exists an open subset $\mathscr{U}$ of $\mathbf{R}^d$ such that, for all $e \in E$:
  - β_e is differentiable on $\mathscr{U}$ and its gradient is locally Lispchitz;
  - $\sqrt{\beta_e}$ is locally Lipschitz on $\mathscr{U}$.
- The lifetime of N^K, i.e. the limit of the time of the i-th jump of N^K as i goes to infinity, is almost surely infinite.

These assumptions are satisfied in all the applications we have in mind. Note that if β_e is indeed C^1, a sufficient condition for its square root to be locally Lipschitz is that β_e does not vanish on $\mathscr{U}$. The last assumption is only made for mathematical comfort.

Let $F\colon \mathscr{U} \to \mathbf{R}^d$ be the vector field defined by

$$F(x) = \sum_{e\in E} \beta_e(x)e, \tag{3}$$

and for all $x \in \mathscr{U}$, let φ_x be the maximal solution of the Cauchy problem

$$\begin{cases} \dot{\varphi_x} = F(\varphi_x) \\ \varphi_x(0) = x \end{cases}, \tag{4}$$

where $\dot{\varphi_x}$ denotes the time derivative of φ_x. The flow $\varphi : (x,t) \mapsto \varphi_x(t)$ is of class $\mathcal{C}^1$ on its domain of definition, which is an open subset of $\mathscr{U} \times \mathbf{R}$. Let us fix $x = (x_1, \ldots, x_d) \in \mathscr{U}$, assume $N^K(0) = \lfloor Kx \rfloor := (\lfloor Kx_1 \rfloor, \ldots, \lfloor Kx_d \rfloor)$ and set $X^K_x = N^K/K$. We have the following functional central limit theorem [20]. For all $T > 0$ such that φ_x is defined on $[0,T]$, we have

$$\sqrt{K}\left(X^K_x - \varphi_x\right) \underset{K\to\infty}{\Longrightarrow} U_x$$

in the Skorokhod space $\mathcal{D}([0,T], \mathbf{R}^d)$, and U_x satisfies, almost surely for all $t \in [0,T]$,

$$U_x(t) = \int_0^t F'(\varphi_x(s))U_x(s)\mathrm{d}s + \sum_{e\in E}\left(\int_0^t \sqrt{\beta_e(\varphi_x(s))}\mathrm{d}W_e(s)\right)e$$

where $W = (W_e(t)\,;\, e \in E, t \geq 0)$ is a $\mathbf{R}^E$-valued standard Brownian motion, and $F'(y)$ denotes the Jacobian matrix of F at y. Moreover, Kurtz showed that we can construct X_x^K and U_x on the same probability space such that

$$\mathbf{P}\left(\sup_{0\leq t\leq T} \left\| X_x^K(t) - \varphi_x(t) - U_x(t)/\sqrt{K} \right\| \geq C_T \log(K)/K \right) \underset{K\to+\infty}{\longrightarrow} 0, \tag{5}$$

where C_T is a constant which grows exponentially fast as T increases, due to the use of Grönwall lemma (see [21], or [13, Chapter 11, Section 3]). This suggests that with high probability the gap between X_x^K and its Gaussian approximation is negligible with respect to $1/\sqrt{K}$ during a period of time of order $\log(K)$. In the present paper, we show that with additional stability assumptions on the limiting ODE, we can obtain much longer time scales. Starting from now, we make the following assumption :

Assumption A. *There exists $x_* \in \mathscr{U}$ such that $F(x_*) = 0$ and all the eigenvalues of the Jacobian $F'(x_*)$ have a negative real part.*

This entails that x_* is an exponentially stable equilibrium point of F, see e.g. [31, Corollary 3.27]. Let $\mathscr{U}_*$ denote its basin of attraction, i.e. the set of all $x \in \mathscr{U}$ such that $\varphi_x(t) \to x_*$ as $t \to +\infty$. It is an open subset of $\mathscr{U}$.

Let us take $x \in \mathscr{U}_*$. It is known that X_x^K shows a metastable behaviour : the theory of large deviations for dynamical systems perturbed with Poissonian noise [29, 6] predicts that X_x^K stays in a small neighbourhood of the equilibrium for a time which is exponentially large in K. Similarly, the vector field F tends to bring the trajectories of the density X_x^K and the Gaussian approximation $\varphi_x + U_x/\sqrt{K}$ closer together, keeping the gap small between them. In this paper we construct a coupling between these two processes by essentially concatenating couplings like the one constructed by Kurtz, on intervals of length one. Then, roughly speaking, reaching an error threshold $\varepsilon(K)$ can be compared to a succession of independent trials of low probability of success.

Let us introduce some notation before stating our main result. We denote by μ_x^K the probability distribution of X_x^K on the Skorokhod space $\mathcal{D}(\mathbf{R}_+, \mathbf{R}^d)$, and by ν_x the probability distribution of U_x on $\mathcal{C}(\mathbf{R}_+, \mathbf{R}^d)$. Given two probability spaces $(E_1, \mathscr{E}_1, \mu_1)$ and $(E_2, \mathscr{E}_2, \mu_2)$, a *coupling* of (μ_1, μ_2) is a random element (X_1, X_2) of $(E_1 \times E_2, \mathscr{E}_1 \otimes \mathscr{E}_2)$ such that X_1 is distributed as μ_1 and X_2 as μ_2. For two functions $f, g : \mathbf{R}_+^* \to \mathbf{R}_+^*$, we write $f(K) \ll g(K)$, or $f(K) = o(g(K))$, if $f(K)/g(K) \to 0$ as $K \to +\infty$. We fix a compact subset $\mathscr{D}$ of $\mathscr{U}_*$, containing x_* in its interior.

Theorem 1.1. *There exist constants $C, V, \alpha > 0$ such that for every $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ satisfying $\alpha \log(K)/K \leq \varepsilon(K) \ll 1$, the following holds. For all K large enough and for all $x \in \mathscr{D}$, there exists a coupling (X_x^K, U_x) of (μ_x^K, ν_x) such that for all $T \geq 0$,*

$$\mathbf{P}\left(\sup_{0\leq t\leq T} \left\| X_x^K(t) - \varphi_x(t) - U_x(t)/\sqrt{K} \right\| > \varepsilon(K) \right) \leq C(T+1)\exp\left(-VK\varepsilon(K)\right).$$

With this coupling, the gap between the density and its Gaussian approximation remains smaller than $\varepsilon(K)$ during a period of time of order $\exp(VK\varepsilon(K))$. The main choices for $\varepsilon(K)$ and the corresponding time scales are regrouped in the following table.

Precision $\varepsilon(K)$	$C_T \log(K)/K$	$\alpha \log(K)/K$	$o(1/\sqrt{K})$	$K^{-p},\ 0 < p < 1/2$
Time scale	T	$K^{V\alpha}$	$\exp(o(\sqrt{K}))$	$\exp(VK^{1-p})$

The first column recalls the result (5) obtained by Kurtz on a fixed time window in the general case. We see in the second column that with Assumption A, a precision of order $\log(K)/K$ can be achieved uniformly during polynomial time scales. The third colum shows that, roughly speaking, the functional central limit theorem can be extended to any time scale of the form $\exp(o(\sqrt{K}))$. Choosing $K^{-1/2} \ll \varepsilon(K) \ll 1$ yields interesting results too and enables to explore the whole range of subexponential time scales. We cannot expect more, since exponential time scales are associated to large deviations [6, 29], and the rate functions associated to the large deviations of $X_x^K - \varphi_x$ and $U_x/\sqrt{K}$ are different [14]. Note that the time scale $\exp(VK\varepsilon(K))$ coincides with the time needed for $\|X_x^K - \varphi_x\|$ to reach a level of order $\sqrt{\varepsilon(K)}$ (see Lemma 3.8), which corresponds to moderate deviations of $X_x^K - \varphi_x$ since $\sqrt{K} \ll \sqrt{\varepsilon(K)} \ll 1$. We refer to [26] for a detailed study of moderate deviations of density-dependent Markov chains.

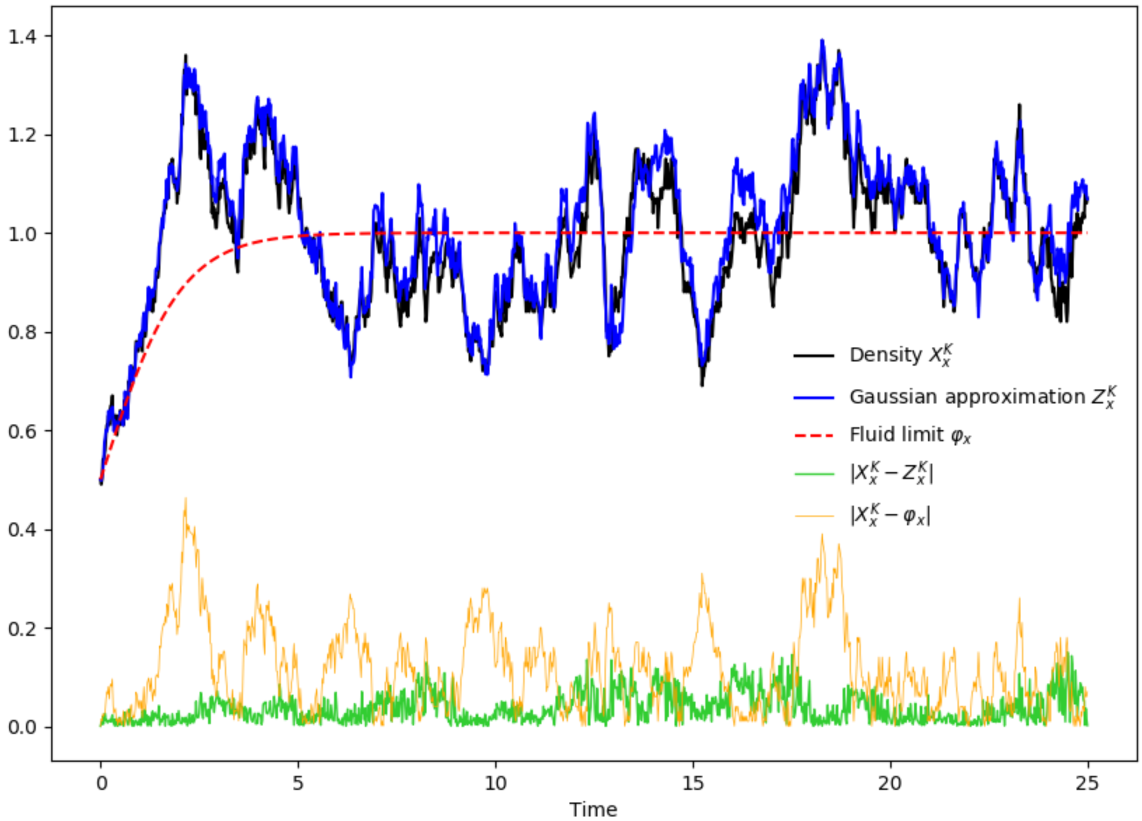

Figure 1: Simulation of the coupling (X_x^K, Z_x^K) given by Theorem 1.1, where $Z_x^K := X_x^K + U_x/\sqrt{K}$, for the logistic birth-and-death process (see Section 2.2). Here $d = 1$, $E = \{-1, 1\}$, $\beta_1(x) = 2x\mathbf{1}_{x\geq 0}$, $\beta_{-1}(x) = x(1+x)\mathbf{1}_{x\geq 0}$, $K = 100$ and $x = 0.5$. The coupling is described in Section 3.2. We use the algorithm presented in [25] to generate 'KMT couplings' of Poisson processes and Brownian motions, satisfying (1).

Of course, it is important to understand the large time behaviour of the process $\varphi_x + U_x/\sqrt{K}$. Actually, after a transitory period, it can be well approximated by a stationary process. Set $S_* = \sum_{e\in E} \beta_e(x_*)\, e\, e^T$, and

$$\Sigma_* = \int_0^\infty e^{sF'(x_*)} S_* e^{sF'(x_*)^T} \mathrm{d}s,$$

which is well defined since $s \mapsto \|e^{sF'(x_*)}\|$ is exponentially decreasing. We can show that for all $x \in \mathscr{U}_*$,

$$U_x(t) \underset{t\to+\infty}{\Longrightarrow} \mathscr{N}\left(0, \Sigma_*\right). \tag{6}$$

Moreover, if we let $W = (W_e(t)\,;\, e \in E, t \geq 0)$ be a $\mathbf{R}^E$-valued standard Brownian motion and $U_*(0)$ be distributed as $\mathscr{N}(0, \Sigma_*)$ and independent of W, then the unique strong solution U_* of the SDE

$$U_*(t) = U_*(0) + \int_0^t F'(x_*)U_*(s)\mathrm{d}s + \sum_{e\in E} \sqrt{\beta_e(x_*)} W_e(t) e$$

is a stationary process. Let us denote by ν_* its probability distribution on $\mathcal{C}\left(\mathbf{R}_+, \mathbf{R}^d\right)$, and set $\rho_* = \min\{-\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(F'(x_*))\}$, which is positive due to Assumption A. From Theorem 1.1 we can deduce the following corollary, which gives a simpler approximation for X_x^K, valid after a transitory period of order $\log(K)$.

Corollary 1.2. *There exists $C', V, \alpha > 0$ such that for every $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ satisfying $\alpha \log(K)/K \leq \varepsilon(K) \ll 1$, the following holds. For all K large enough and for all $x \in \mathscr{D}$, there exists a coupling $\left(X_x^K, U_*\right)$ of $\left(\mu_x^K, \nu_*\right)$ such that for all $T \geq (6/\rho_*)\log(K)$,*

$$\mathbf{P}\left(\sup_{(6/\rho_*)\log(K) \leq t \leq T} \left\| X_x^K(s) - x_* - U_*(s)/\sqrt{K} \right\| > \varepsilon(K)\right) \leq C'(T+1)\exp\left(-VK\varepsilon(K)\right).$$

Regarding marginal distributions of the density process, the Gaussian approximation remains valid for much a much longer period of time. Indeed, X_x^K stays in $\mathscr{D}$ with high probability for a period of time that is exponentially large in K (time scale of large deviations). Combining this fact with Corollary 1.2 yields the following result. We denote by $\mathcal{W}_c$ the Wasserstein distance on $\mathcal{P}(\mathbf{R}^d)$ associated to the truncated distance $c(x, y) = \|x - y\| \wedge 1$, i.e.

$$\begin{array}{rccl} \mathcal{W}_c\colon & \mathcal{P}(\mathbf{R}^d) \times \mathcal{P}(\mathbf{R}^d) & \to & \mathbf{R}_+ \\ & (\mu_1, \mu_2) & \mapsto & \inf_{\pi \in \Pi(\mu_1,\mu_2)} \int_{\mathbf{R}^d \times \mathbf{R}^d} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y), \end{array}$$

where $\Pi(\mu_1, \mu_2)$ is the set of probability measures on $\mathbf{R}^d \times \mathbf{R}^d$ with first marginal μ_1 and second marginal μ_2.

Corollary 1.3. *There exists $V' > 0$ such that for all $x \in \mathscr{D}$,*

$$\sup_{(12/\rho_*)\log(K) \leq t \leq \exp(V'K)} \mathcal{W}_c\Big[\mathbf{P}\left(\sqrt{K}\left(X_x^K(t) - x_*\right) \in \cdot\right), \mathscr{N}\left(0, \Sigma_*\right)\Big] \underset{K \to +\infty}{\longrightarrow} 0.$$

This paper contains a Section 2 which is devoted to various applications of our main result, and a Section 3 which contains all the proofs.

Notation. Given a topological space Y, $\mathscr{B}(Y)$ stands for its Borel sigma-algebra, and $\mathcal{P}(Y)$ stands for the set of probability measures on $(Y, \mathscr{B}(Y))$. If Z, Z' are random elements of Y and $\lambda \in \mathcal{P}(Y)$, the notation $Z \sim Z'$ (resp. $Z \sim \lambda$) means that Z and Z' share the same distribution (resp. that Z has distribution λ). We use the notation $\|\cdot\|$ both for the Euclidean norm on $\mathbf{R}^d$ and for the associated operator norm on the set $M_d(\mathbf{R})$ of $d \times d$ real matrices. We denote by A^T the transpose of a matrix A. The notation $\dot{u}$ refers to the derivative of a function of time $t \mapsto u(t)$. For all $x \in \mathbf{R}^d$ and $r \geq 0$, $B(x, r)$ (resp. $\bar{B}(x, r)$) stands for the open (resp. closed) euclidean ball of center x and radius r. Given a function $f : E_1 \to E_2$, where E_2 is some normed vector space, we denote by $\|f\|_\infty$ the supremum norm of f. If E_1 is a subset of $\mathbf{R}^d$, we let $\|f\|_{\mathrm{Lip}}$ denote the quantity $\sup\left\{\|f(y) - f(x)\|/\|x - y\| \,;\, x, y \in \mathbf{R}^d, x \neq y\right\}$. Finally, if $S \subset E_1$, then we write $\|f\|_{\infty,S} := \|f_{|S}\|_\infty$ and $\|f\|_{\mathrm{Lip},S} := \|f_{|S}\|_{\mathrm{Lip}}$.

2 Applications

We discuss some consequences of Theorem 1.1. In Section 2.1 we show that we can use the Gaussian approximation to estimate the time scales of moderate deviations of $X_x^K - \varphi_x$. The next two subsections are devoted to concrete examples of density-dependent Markov chains. In Section 2.2 we consider the logistic birth-and-death process, and we obtain Gaussian approximation estimates for the process conditioned to survive, and for the associated quasi-stationary distribution. Then, in Section 2.3 we consider the stochastic SIRS epidemic model and we apply Theorem 1.1 to give a Gaussian estimation of a quantity modeling the cost of the epidemic.

2.1 Moderate deviations

We know that the density $X_{x_*}^K$ stays close to the equilibrium point x_* for a long time. In population models it is useful to estimate precisely the time needed for deviations to occur (in particular, the extinction of a population). In this section we consider deviations of order $\eta(K)$, where $K^{-1/2} \ll \eta(K) \ll 1$. They are called *moderate* deviations : $K^{-1/2}$ is the natural scale of the fluctuations given by the central limit theorem, while taking η constant would correspond to large deviations.

In what follows, we assume that $S_* \in S_d^{++}$, where S_d^{++} denotes the set of symmetric, positive-definite $d \times d$ real matrices. It entails that $\Sigma_* \in S_d^{++}$. We denote by $\|\cdot\|_A$ the norm associated to a matrix $A \in S_d^{++}$, defined by $\|y\|_A = \sqrt{y^T A y}$. For all $\delta > 0$, the process $U^\delta := \delta U_{x_*}$ satisfies the SDE

$$U^\delta(t) = \int_0^t F'(x_*)U^\delta(s)\mathrm{d}s + \delta\, S_*^{1/2} B(t),$$

where B is a d-dimensional Brownian motion and $S_*^{1/2}$ denotes the symmetric square root of S_*. Set $\delta(K) = 1/(\sqrt{K}\eta(K))$. Then the process $U^{\delta(K)}$ is an approximation of $\eta^{-1}(K)\left(X_{x_*}^K - x_*\right)$, and its large deviations are well described by the Freidlin-Wentzell theory. For all $T > 0$, let $I_T : \mathcal{C}([0,T], \mathbf{R}^d) \to \mathbf{R}_+$ be the Freidlin-Wentzell action associated to the family $(U^\delta\,;\,\delta > 0)$, defined by

$$I_T(u) = \begin{cases} \frac{1}{2}\int_0^T \|(\dot{u}(s) - F'(x_*)u(s))\|_{S_*^{-1}}^2 \mathrm{d}s & \text{if } u(0) = 0,\ u \text{ is absolutely continuous} \\ & \text{and } \dot{u}(s) - F'(x_*)u(s) \in \mathrm{Im}(\sigma) \text{ a.e.;} \\ +\infty & \text{otherwise,} \end{cases},$$

The associated quasipotential is explicit [10, Proposition 2.3.6], for all $y \in \mathbf{R}^d$ we have

$$\inf\left\{I_T(u)\,;\,T > 0, u \in \mathcal{C}([0,T], \mathbf{R}^d), u(T) = y\right\} = \frac{1}{2}\|y\|_{\Sigma_*^{-1}}^2.$$

For all y on the unit sphere of $\|\cdot\|_{\Sigma_*^{-1}}$, the vector $F'(x_*)y$ points towards the interior of the ball. Indeed, an integration by parts shows that Σ_* solves the Lyapunov equation $F'(x_*)\Sigma_* + \Sigma_* F'(x_*)^T = -S_*$, hence $(F'(x_*)y)^T(2\Sigma_*^{-1}y) = -(\Sigma_*^{-1}y)^T S_* \Sigma_*^{-1} y < 0$. Thus, Theorem 4.2 in Chapter 4 of [14] entails that for all $h > 0$,

$$\mathbf{P}\left[\exp\left(\delta^{-2}\left(\frac{1}{2} - h\right)\right) < \inf\left\{t \geq 0 : \left\|U^\delta(t)\right\|_{\Sigma_*^{-1}} \geq 1\right\} < \exp\left(\delta^{-2}\left(\frac{1}{2} + h\right)\right)\right] \underset{\delta \to 0}{\longrightarrow} 1. \quad (7)$$

Using Theorem 1.1, we can show that the Gaussian approximation yields the good asymptotics for the moderate deviations of $X_{x_*}^K$. This was already noticed by Pardoux in [26] (and before that by Barbour [4] under the assumption $\eta(K) \ll K^{-3/8}$) : the rate function associated to the moderate deviations of $X_{x_*}^K$ coincides with I_T. Let us stress that it is not the case for large deviations [6, 29].

Proposition 2.1. *Let* $\tau_\eta^K = \inf\left\{t \geq 0 : \left\|X_{x_*}^K(t) - x_*\right\|_{\Sigma_*^{-1}} \geq \eta(K)\right\}$. *For all* $h > 0$,

$$\mathbf{P}\left[\exp\left(\left(\frac{1}{2} - h\right)K\eta^2(K)\right) < \tau_\eta^K < \exp\left(\left(\frac{1}{2} + h\right)K\eta^2(K)\right)\right] \underset{K \to +\infty}{\longrightarrow} 1. \quad (8)$$

2.2 Logistic birth-and-death process conditioned to survival

Consider the logistic birth-and-death process. It is a density-dependent Markov chain with $d = 1$, $E = \{-1, 1\}$, $\beta_1(x) = px\mathbf{1}_{x \geq 0}$ for some $p > 0$, and $\beta_{-1}(x) = \beta_{-1}(x) = x(q + x)\mathbf{1}_{x \geq 0}$ for some $0 < q < p$. We suppose $N^K(0) \in \mathbf{N}$. Then N^K is a $\mathbf{N}$-valued continuous-time Markov chain with transition rates from $n \in \mathbf{N}$ to $m \in \mathbf{N} \setminus \{n\}$ given by

$$q_{n,m}^K = \begin{cases} pn & \text{if } m = n + 1 \quad \text{(birth)} \\ n(q + n/K) & \text{if } m = n - 1 \quad \text{(death)}\,. \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

This is a very classical model in population dynamics. The quadratic form of the death rate models the competition between individuals. The scale paramater $K > 0$ is the inverse of the intensity of the competition : it can be interpreted as an amount of resources available to the population. We refer to [3] for the proof that the lifetime of N^K is almost surely infinite. Note that the functions β_{-1} and β_1 are of class $\mathcal{C}^\infty$ and do not vanish on $\mathscr{U} = \mathbf{R}_+^*$, and the vector field $F\colon \mathbf{R}_+^* \ni x \mapsto \beta_1(x) - \beta_{-1}(x) = x(p-q-x)$ admits $x_* = p - q$ as an equilibrium point. Since $\rho_* := -F'(x_*) = x_* > 0$, Assumption A is met, and the basin of attraction of x_* is $\mathscr{U}_* = \mathbf{R}_+^*$. Moreover $\Sigma_* = p$ and the process U_* satisfies

$$U_*(t) = U_*(0) - x_* \int_0^t U_*(s)\mathrm{d}s + \sqrt{2px_*}W(t)$$

where W is a real Brownian motion and $U_*(0) \sim \mathscr{N}(0, p)$ is independent of W.

Almost surely, the process N^K hits the absorbing point 0 in finite time (see e.g. [3]). Suppose that we know that a population, whose size is well modeled by N^K, has survived for a long time. What can we say about the present size of this population, and how it evolved in the past ? Such questions are classical, and a vast literature exists about the large time behaviour of Markov processes conditioned to non extinction, and the related notion of quasi-stationary distribution (see for instance the survey [23] and the book [11]).

Using Theorem 1.1, we show that one can strongly approximate the past trajectory with good precision on large time scales by the trajectory of a stationary Gaussian process. For all $t, T \geq 0$, we denote by $\tilde{\mu}_x^{K;t,T}$ the probability distribution of the process $\left(X_x^K(t+s)\right)_{0 \leq s \leq T}$ conditional on the event $\left\{X_x^K(t+T) > 0\right\}$, where $X_x^K \sim \mu_x^K$, and we denote by $\tilde{\mu}_x^{K;t}$ the probability distribution of $X_x^K(t)$ conditional on $\left\{X_x^K(t) > 0\right\}$.

Proposition 2.2. *There exists constants $C''', V'', \alpha, \beta > 0$ such that the following holds. For all $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ satisfying $\alpha \log(K)/K \leq \varepsilon(K) \ll 1$, for all K large enough, for all $t \geq (6/x_*)\log(K)$, all $T \geq 0$ and all $x \in K^{-1}\mathbf{N}^*$, there exists a coupling $(\tilde{X}, U_*)$ of $\left(\tilde{\mu}_x^{K;t,T}, \nu_*\right)$, such that*

$$\mathbf{P}\left(\sup_{0 \leq s \leq T} \left|\tilde{X}(s) - x_* - U_*(s)/\sqrt{K}\right| > \varepsilon(K)\right) \leq C''' \left(\exp\left(-\frac{t}{\beta \log(K)}\right) + (T+1)\exp\left(-V'' K \varepsilon(K)\right)\right).$$

Note that it is important that the estimate be uniform in x when the initial size of the population is not known. The scale $\beta \log(K)$ corresponds to the time needed to 'forget' the initial starting point, conditional on survival.

Let us mention an interesting consequence of this result. It is shown in [32] that for all $K > 0$, there exists a unique probability distribution γ^K on $K^{-1}\mathbf{N}^*$ such that, if $X^K(0) \sim \gamma^K$, then $\mathbf{P}\left(X^K(t) \in \cdot \,|\, X^K(t) > 0\right) = \gamma^K$ for all $t \geq 0$. Such a probability distribution is called a *quasi-stationary distribution* (QSD). Moreover, the QSD satisfies, for all $x \in K^{-1}\mathbf{N}^*$ and $A \subset K^{-1}\mathbf{N}^*$,

$$\tilde{\mu}_x^{K;t}(A) \underset{K \to +\infty}{\longrightarrow} \gamma^K(A).$$

In [7], Chazottes, Collet and Méléard use spectral methods to obtain sharp bounds on the total variation distance between the distributions $\tilde{\mu}_x^{K;t}$ and γ^K, and between γ^K and the discrete Gaussian distribution

$$\lambda^K = \frac{1}{Z(K)} \sum_{n \in \mathbf{N}} \exp\left(-\frac{(n - \lfloor K x_* \rfloor)^2}{Kp}\right) \delta_{n/K},$$

where $Z(K)$ is a normalization constant. More precisely, they show that

$$\mathrm{d}_{TV}\left(\gamma^K, \lambda^K\right) = \mathcal{O}(1/\sqrt{K}), \tag{10}$$

where d_{TV} stands for the total variation distance.

We can obtain an analogous result, with respect to the distance W_c on $\mathcal{P}(\mathbf{R})$. We recall that $\mathcal{W}_c$ was defined in the introduction as the Wasserstein distance on $\mathcal{P}(\mathbf{R})$ associated to the truncated distance $c(x, y) = |x - y| \wedge 1$. Using Proposition 2.2, we obtain the following result.

Corollary 2.3. *Let $\tilde{\gamma}^K$ be the rescaled quasistationary distribution defined by*

$$\tilde{\gamma}^K(A) = \gamma^K\left(\left\{x > 0 : \sqrt{K}(x - x_*) \in A\right\}\right),$$

for all $A \in \mathscr{B}(\mathbf{R})$. Then, we have

$$\mathcal{W}_c\left(\tilde{\gamma}^K, \mathscr{N}(0, p)\right) = \mathcal{O}\left(K^{-1/2}\log(K)\right).$$

Note that this result is not optimal, as one could remove the logarithmic factor using (10). Its interest lies in the trajectorial approach we used. We expect to generalize Corollary 2.3 to a multidimensional context, such as the one studied in [8] and [9].

2.3 SIRS model and cost of an epidemic

Consider the following epidemic model (SIRS). An infectious disease spreads in a population of individuals which can be either susceptible, infected, of recovered. We assume that the total size of the population is constant equal to a large $K \in \mathbf{N}_*$, hence it is enough to record the amounts of susceptibles and infected, denoted by N_s^K and N_i^K respectively. We model the evolution of the epidemic by assuming that the process $N^K = (N_s^K(t), N_i^K(t))_{t\geq 0}$ is a $\mathbf{N}^2$-valued continuous-time Markov chain, such that the transition rate from $n = (n_s, n_i)$ to $m \neq n$ is given by

$$q_{n,m}^K = \begin{cases} \lambda n_s(n_i/K) & \text{if } m = (n_s - 1, n_i + 1) \quad \text{(infection)} \\ \gamma n_i & \text{if } m = (n_s, n_i - 1) \quad \text{(recovery)} \\ \theta(K - n_i - n_s) & \text{if } m = (n_s + 1, n_i) \quad \text{(loss of immunity)} \\ 0 & \text{otherwise} \end{cases},$$

for some $\lambda, \gamma, \theta > 0$. The loss of immunity may arise if the pathogen mutates (one may have in mind, for instance, the influenza virus). In other words, N^K is a density-dependent Markov chain with $d = 2$, $E = \{e_{inf} = (-1, 1), e_{rec} = (0, -1), e_{loi} = (1, 0)\}$, and, for all $(s, i) \in \mathbf{R}^2$, setting $\Delta = \{s, i \in \mathbf{R}_+ : i + s \leq 1\}$,

$$\beta_{inf}(s, i) = \lambda\, s\, i\, \mathbf{1}_\Delta(s, i), \quad \beta_{rec}(s, i) = \gamma\, i\, \mathbf{1}_\Delta(s, i), \quad \beta_{loi}(s, i) = \theta\,(1 - i - s)\, \mathbf{1}_\Delta(s, i),$$

with obvious notations. Their restrictions to the open set $\mathscr{U} = \mathring{\Delta}$ are $\mathcal{C}^\infty$ and positive. If we let $F = \beta_{inf}e_{inf} + \beta_{rec}e_{rec} + \beta_{loi}e_{loi}$, then for $(s, i) \in \mathscr{U}$, we have

$$F(s, i) = (-\lambda\, s\, i + \theta(1 - i - s),\ \lambda\, s\, i - \gamma\, i).$$

From now on, we assume that $\lambda > \gamma$. Then, F has a unique zero on $\mathscr{U}$, given by

$$x_* = (s_*, i_*) = \left(\frac{\gamma}{\lambda}, \frac{\theta(\lambda - \gamma)}{\lambda(\gamma + \theta)}\right).$$

The equilibrium point x_* satisfies Assumption A. Indeed,

$$F'(x_*) = \begin{pmatrix} -\theta(\lambda + \theta)/(\gamma + \theta) & -(\gamma + \theta) \\ \theta(\lambda - \gamma)/(\gamma + \theta) & 0 \end{pmatrix}$$

has a positive determinant and a negative trace. We call x_* the endemic equilibrium : it corresponds to a persistence of the disease in the population. Here, it is made possible by the supply of new susceptibles due to the loss of immunity.

Say we want to estimate the cost of the epidemic for the population, on a time interval $[0, t]$. A simple model is to consider that the total cost is proportional to the sum, on all individuals, of the total time during which they were infected. The cost per person of the epidemic on $[0, t]$ is then proportional to

$$\int_0^t I^K(s)\mathrm{d}s,$$

where $I^K = N_i^K/K$. To estimate this quantity, we can make use of the trajectorial Gaussian approximation given by Theorem 1.1. To simplify, we suppose that $N^K(0) = \lfloor Kx_* \rfloor$. Set

$$S_* = \beta_{inf}(x_*)e_{inf}\, e_{inf}^T + \beta_{rec}(x_*)e_{rec}\, e_{rec}^T + \beta_{loi}(x_*)e_{loi}\, e_{loi}^T = \frac{\gamma\theta(\lambda-\gamma)}{\lambda(\gamma+\theta)} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix},$$

and let $C, V, \alpha > 0$ be given by Theorem 1.1. We can assume that, on the same probability space as N^K, there is a continuous 2-dimensional process U_{x_*} satisfying the SDE

$$U_{x_*}(t) = \int_0^t F'(x_*)U_{x_*}(s)\mathrm{d}s + S_*^{1/2}B(t) \tag{11}$$

where B is a 2-dimensional Brownian motion, such that the following holds. For every $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ satisfying $\alpha \log(K)/K \leq \varepsilon(K) \ll 1$, we have, for all K large enough and for all $t \geq 0$:

$$\mathbf{P}\left(\sup_{0\leq s\leq t} \left\| N^K(s)/K - x_* - U_{x_*}(s)/\sqrt{K} \right\| > \varepsilon(K) \right) \leq C(t+1)\exp\left(-VK\varepsilon(K)\right).$$

From this we can deduce a Gaussian approximation for the cost per person of the epidemic. Given a family of (real-valued) random variables $\left(A^K;\, K > 0\right)$ and a function $f\colon \mathbf{R}_+^* \to \mathbf{R}_+^*$, we write $A^K = \mathcal{O}_{\mathbf{P}}(f(K))$ if for all $\delta > 0$, there exists $M > 0$ and $K_0 > 0$ such that $\mathbf{P}\left(|A^K| > Mf(K)\right) \leq \delta$ for all $K \geq K_0$. We denote by $y^{(1)}$ and $y^{(2)}$ the first and second coordinates of a vector $y \in \mathbf{R}^2$.

Proposition 2.4. *Set*

$$\sigma = \left[\frac{2\gamma\theta}{\lambda(\lambda-\theta)(\gamma+\theta)^3} \left((\lambda-\gamma)^2 + (\gamma+\theta)(\lambda+\theta)\right) \right]^{1/2}.$$

For all $t \geq 0$, we have

$$\int_0^t I^K(s)\mathrm{d}s = i_*\, t + \sigma W(t)/\sqrt{K} + \left(F'(x_*)^{-1}U_{x_*}\right)^{(2)}(t)/\sqrt{K} + \int_0^t \left(I^K(s) - i_* - U_{x_*}^{(2)}(s)/\sqrt{K}\right)\mathrm{d}s, \tag{12}$$

where W is a real Brownian motion. For all $T\colon \mathbf{R}_+^* \to \mathbf{R}_+^*$ such that $1 \ll T(K) \ll K^p$ for some $p > 1$, we have

$$\int_0^{T(K)} I^K(s)\mathrm{d}s = i_*\, T(K) + \sigma\sqrt{T(K)/K}\,\mathscr{N}(0,1) + \mathcal{O}_{\mathbf{P}}\left(1/\sqrt{K} + T(K)\log(K)/K\right). \tag{13}$$

The leading term $i_*\, T(K)$ is given by the deterministic approximation of I^K. Note that in the regime $K/(\log(K))^2 \ll T(K) \ll K^p$, the Gaussian term becomes negligible with respect to $T(K)\log(K)/K$ and the approximation reduces to

$$\int_0^{T(K)} I^K(s)\mathrm{d}s = T(K)\left(i_* + \mathcal{O}_{\mathbf{P}}\left(\log(K)/K\right)\right).$$

3 Proofs

3.1 Preliminaries

3.1.1 Limiting ODE

We first prove some basic consequences of Assumption A on the fluid limit φ_x and on solutions of the ODE

$$\dot{y} = F'\left(\varphi_x\right)y, \tag{14}$$

for $x \in \mathscr{U}_*$. The ODE (14) is the linearization of the limiting ODE, driven by the vector field F, along the trajectory φ_x. It is important for our purposes because X_x^K is a random perturbation of φ_x. For all $s \geq 0$, let $\Psi_x(\cdot, s) \colon \mathbf{R}_+ \to M_d(\mathbf{R})$ be the unique matrix solution of the Cauchy problem

$$\begin{cases} \dfrac{\partial \Psi_x(t,s)}{\partial t} = F'\big(\varphi_x(t)\big)\Psi_x(t,s), \quad t \in \mathbf{R}_+ \\ \Psi_x(s,s) = I_d. \end{cases}.$$

This is a classical object known as the principal matrix solution of the ODE (14) at time s. Given $r, s \geq 0$, the functions $t \mapsto \Psi_x(t,r)$ and $t \mapsto \Psi_x(t,s)\Psi_x(s,r)$ satisfy the same Cauchy problem at time s, hence

$$\Psi_x(t,r) = \Psi_x(t,s)\Psi_x(s,r) \quad \text{and} \quad \Psi_x(t,s) = \Psi_x(s,t)^{-1}$$

for all $r, s, t \geq 0$. Thus Ψ_x is also differentiable with respect to its second variable and we have

$$\frac{\partial \Psi_x(t,s)}{\partial s} = \Psi_x(t,s) F'\big(\varphi_x(s)\big). \tag{15}$$

The following lemma gives classical bounds related to the exponential stability of x_* (resp. 0) for the limiting ODE (resp. its linearization). Recall that $\rho_* = \min\{-\mathrm{Re}(\lambda)\,;\, \lambda \in \mathrm{Sp}(F'(x_*))\}$.

Lemma 3.1. i) *There exists $\Gamma_1 \geq 1$ such that, for all $t \geq 0$ and $x \in \mathscr{D}$:*

$$\|\varphi_x(t) - x_*\| \leq \Gamma_1 e^{-\rho_* t/2}. \tag{16}$$

ii) *The set $\overline{\varphi(\mathbf{R}_+ \times \mathscr{D})}$ is compact.*

iii) *There exists $\Gamma_2 \geq 1$ such that, for all $t \geq s \geq 0$ and $x \in \mathscr{D}$:*

$$\|\Psi_x(t,s)\| \leq \Gamma_2 e^{-\rho_*(t-s)/2}. \tag{17}$$

Proof. Let us prove i). Assumption A entails that there exist $\rho > 0$, $\delta > 0$ and $\Gamma_0 \geq 1$ such that, for all $x \in \bar{B}(x_*, \delta)$ and all $t \geq 0$,

$$\|\varphi_x(t) - x_*\| \leq \Gamma_0 \|x - x_*\| e^{-\rho t}.$$

Moreover, we can choose ρ to be any value in the open interval $(0, \rho_*)$ (see e.g. [31, Corollary 3.27]). We take $\rho = \rho_*/2$. Consider the function $T : \mathscr{U}_* \to \mathbf{R}_+$, defined by $T(x) = \inf\{t \geq 0 : \|\varphi_x(t) - x_*\| < \delta\}$. By continuity of the flow φ with respect to the space variable, the function T is upper semi-continuous, and consequently it is bounded from above on the compact $\mathscr{D}$. Let $\bar{T} \geq \|T\|_{\infty,\mathscr{D}}$, define $M = \sup\big\{\|\varphi_x(t) - x_*\|\,;\, x \in \mathscr{D},\ 0 \leq t \leq \bar{T}\big\}$, which is finite due to the continuity of φ, and set

$$\Gamma_1 = (M \vee \Gamma_0 \delta)\, e^{\rho_* \bar{T}/2},$$

where $p \vee q$ (resp. $p \wedge q$) stands for the maximum (resp. the minimum) of p and q. Let $x \in \mathscr{D}$. There exist $0 \leq s \leq \bar{T}$ such that $\varphi_x(s) \in \bar{B}(x_*, \delta)$. For all $0 \leq t \leq s$, (16) holds by definition of Γ_1, and for all $t \geq s$,

$$\|\varphi_x(t) - x_*\| = \|\varphi_{\varphi_x(s)}(t-s) - x_*\| \leq \Gamma_0 \delta e^{-\rho_*(t-s)/2} \leq \Gamma_1 e^{-\rho_* t/2}.$$

Let us prove ii). Let $(x_n, t_n)_{n \in \mathbf{N}}$ be a sequence of elements of $\mathscr{D} \times \mathbf{R}_+$. If (t_n) is not bounded, there exists a subsequence $t_{n_k} \underset{k \to +\infty}{\longrightarrow} +\infty$, and item i) entails that $\varphi(x_{n_k}, t_{n_k}) \underset{k \to +\infty}{\longrightarrow} x_*$. Otherwise, if (t_n) is bounded, we can a extract subsequence (x_{n_k}, t_{n_k}) which converges to a limit $(\tilde{x}, \tilde{t}) \in \mathscr{D} \times \mathbf{R}_+$, so that $\varphi(x_{n_k}, t_{n_k}) \underset{k \to +\infty}{\longrightarrow} \varphi(\tilde{x}, \tilde{t})$. Hence, $\varphi(\mathscr{D} \times \mathbf{R}_+)$ is relatively compact.

Now, we turn to iii). The proof is very similar to [31, Theorem 3.20]. Due to iii), we may suppose without loss of generality that $\mathscr{D}$ is positively invariant by the flow φ. Let $x \in \mathscr{D}$, $s \geq 0$, and set $Y : \mathbf{R}_+ \ni t \mapsto \Psi_x(t,s)$. We have

$$\dot{Y} = F'(\varphi_x)Y = F'(x_*)Y + \big[F'(\varphi_x) - F'(x_*)\big]Y,$$

thus, by variation of constants,

$$Y(t)=e^{(t-s)F'(x_*)}+\int_s^t e^{(t-r)F'(x_*)}\big[F'(\varphi_x(r))-F'(x_*)\big]Y(r)\mathrm{d}r. \tag{18}$$

Let $C\geq 1$ be such that for all $t\geq 0$, $\|e^{tF'(x_*)}\|\leq Ce^{-3\rho_* t/4}$. Recall that the gradient of the β_e are locally Lipschitz, hence $\|F'\|_{\mathrm{Lip},\mathscr{D}}<\infty$. Set $z(t)=e^{3\rho_*(t-s)/4}\|Y(t)\|$. Equation (18) yields, for all $t\geq s$,

$$z(t)\leq C+C\|F'\|_{\mathrm{Lip},\mathscr{D}}\Gamma_1 e^{-\rho_* s/2}\int_s^t z(r)\mathrm{d}r.$$

Obviously there exists $t_0\geq 0$ such that $C\|F'\|_{\mathrm{Lip},\mathscr{D}}\Gamma_1 e^{-\rho_* s/2}\leq \rho_*/4$. For all $t\geq s\geq t_0$, Grönwall's lemma yields $z(t)\leq Ce^{\rho_*(t-s)/4}$, hence

$$\|\Psi_x(t,s)\|\leq Ce^{-\rho_*(t-s)/2}. \tag{19}$$

We end the proof by showing that the condition $s\geq t_0$ can be removed, if we change the constant C. For $t_0\geq t\geq s\geq 0$, Grönwall's lemma entails $\|\Psi_x(t,s)\|\leq e^{t_0\|F'\|_{\infty,\mathscr{D}}}$. Finally, the case $t\geq t_0\geq s\geq 0$ can be reduced to the previous ones thanks to the relation $\Psi_x(t,s)=\Psi_x(t,t_0)\Psi_x(t_0,s)$. Hence, setting $\Gamma_2=C\exp\big(t_0(\|F'\|_{\infty,\mathscr{D}}+\rho_*/2)\big)$, the inequality

$$\|\Psi_x(t,s)\|\leq \Gamma_2 e^{-\rho_*(t-s)/2}$$

holds for all $t\geq s\geq 0$. □

From now on, we make the following assumption. It entails no loss of generality due to item ii).

Assumption. *The compact $\mathscr{D}$ is positively invariant by the flow φ, i.e. $\varphi(\mathscr{D}\times\mathbf{R}_+)=\mathscr{D}$.*

3.1.2 Perturbed linear ODE

We know from Lemma 3.1 that the solutions of the linear ODE (14) are killed exponentially fast. In the following lemma, we consider a solution y of a perturbed version of the integral equation associated to this ODE. We show that if one wants the norm of y to reach high values, then the pertubation term must oscillate strongly enough, to be able to compensate the killing effect of the (non-perturbed) ODE. This lemma is crucial for the proof of Theorem 1.1.

Lemma 3.2. *There exists $\Gamma\geq 1$ such that for every $x\in\mathscr{D}$ and every Borel measurable locally bounded functions $y,h:\mathbf{R}_+\to\mathbf{R}^d$ satisfying*

$$y(t)=y(0)+\int_0^t F'\big(\varphi_x(s)\big)y(s)\mathrm{d}s+h(t),\quad t\geq 0, \tag{20}$$

we have, for all $t\geq 0$:

$$\sup_{0\leq s\leq t}\|y(s)\|\leq \Gamma\left(\|y(0)\|\vee \sup_{\substack{0\leq r,s\leq t\\ |s-r|\leq 1}}\big\|h(s)-h(r)\big\|\right). \tag{21}$$

Proof. Let $\mathscr{D}$ be a compact subset of $\mathscr{U}_*$, which we may suppose positively invariant by φ due to Lemma 3.1, ii). Let $x\in\mathscr{D}$, and let $y,h:\mathbf{R}_+\to\mathbf{R}^d$ be Borel measurable locally bounded functions satisfying (20). For all $t\geq 0$, we have

$$(y-h)(t)=y(0)+\int_0^t F'(\varphi_x(s))y(s)\mathrm{d}s.$$

By variation of constants, we can deduce from the above equation that

$$(y-h)(t)=\Psi_x(t,0)y(0)+\int_0^t \Psi_x(t,s)F'\big(\varphi_x(s)\big)h(s)\mathrm{d}s. \tag{22}$$

Let us precise the argument. We have, for all $t\geq 0$, using (15),

$$\Psi_x(0,t)=I_d-\int_0^t \Psi_x(0,s)F'(\varphi_x(s))\mathrm{d}s. \tag{23}$$

For all functions $a_1,A_1:\mathbf{R}_+\to M_d(\mathbf{R})$ and $a_2,A_2:\mathbf{R}_+\to\mathbf{R}^d$ such that a_1,a_2 are locally integrable and $A_i(t)=A_i(0)+\int_0^t a_i(s)\mathrm{d}s$, Fubini's theorem yields

$$A_1(t)A_2(t)=A_1(0)A_2(0)+\int_0^t \left(a_1(s)A_2(s)+A_1(s)a_2(s)\right)\mathrm{d}s.$$

Equation (22) is obtained by applying this formula to $A_1=\Psi_x(0,\cdot)$ and $A_2=y-h$, before left-multiplying by $\Psi_x(t,0)$. More generally, if we set, for each $j\in\mathbf{N}$,

$$h_j(t)=h(t)-h(j)\quad\text{and}\quad \tilde{h}_j(t)=h_j(t)+\int_j^t \Psi_x(t,s)F'(\varphi_x(s))h_j(s)\mathrm{d}s,$$

then the same argument yields, for all $t\geq j$,

$$y(t)=\Psi_x(t,j)y(j)+\tilde{h}_j(t).$$

By induction on $\lfloor t\rfloor$, we obtain

$$y(t)=\Psi_x(t,0)y(0)+\sum_{j=1}^{\lfloor t\rfloor}\Psi_x(t,j)\tilde{h}_{j-1}(j)+\tilde{h}_{\lfloor t\rfloor}(t).$$

The term $\tilde{h}_{\lfloor t\rfloor}(t)$ represents the contribution of the most recent increments of h to the value of $y(t)$. Lemma **??** entails that the other terms, which represent contribution of increments of h that more distant in the past, are killed exponentially fast. Letting $\Gamma_2\geq 1$ be given by **??**, we get

$$\begin{aligned}\|y(t)\| &\leq \Gamma_2 e^{-\frac{\rho_*}{2}t}\|y(0)\|+\sum_{i=1}^{\lfloor t\rfloor}\Gamma_2 e^{-\frac{\rho_*}{2}(t-j)}\tilde{h}_{j-1}(j)+\sup_{\lfloor t\rfloor\leq s\leq t}\left\|\tilde{h}_{\lfloor t\rfloor}(s)\right\|\\ &\leq \left(\|y(0)\|\vee\max_{0\leq j\leq\lfloor t\rfloor}\sup_{j\leq s\leq t\wedge(j+1)}\left\|\tilde{h}_j(s)\right\|\right)\left(1+\frac{\Gamma_2}{1-e^{-\frac{\rho_*}{2}}}\right).\end{aligned}$$

Now, the definition of $\tilde{h}_j$ implies that for all $j\in\mathbf{N}$ and $j\leq r\leq j+1$,

$$\sup_{i\leq s\leq r}\left\|\tilde{h}_i(s)\right\|\leq\sup_{i\leq s\leq r}\left\|h_i(s)\right\|\left(1+\Gamma_2\|F'\|_{\infty,\mathscr{D}}\right).$$

Finally, we obtain

$$\sup_{0\leq s\leq t}\|y(s)\|\leq\Gamma\left(\|y(0)\|\vee\max_{0\leq i\leq\lfloor t\rfloor}\sup_{i\leq s\leq t\wedge(i+1)}\left\|h(s)-h(i)\right\|\right).$$

with

$$\Gamma=\left(1+\frac{\Gamma_2}{1-e^{-\frac{\rho_*}{2}}}\right)\left(1+\Gamma_2\|F'\|_{\infty,\mathscr{D}}\right).$$

3.1.3 Gaussian process

We show that the Gaussian processes U_x and U_* are well defined and satisfy the properties given in the introduction.

Proposition 3.3. *Let $W = (W_e(t)\,;\, e \in E, t \geq 0)$ be a $\mathbf{R}^E$-valued Brownian motion, and let $U_*(0) \sim \mathscr{N}(0, \Sigma_*)$ be independent of W. For all $x \in \mathscr{D}$, there exist unique strong solutions U_x and U_* to the SDEs*

$$U_x(t) = \int_0^t F'(\varphi_x(s))U_x(s)\mathrm{d}s + \sum_{e\in E}\left(\int_0^t \sqrt{\beta_e(\varphi_x(s))}\mathrm{d}W_e(s)\right)e, \tag{24}$$

$$U_*(t) = U_*(0) + \int_0^t F'(x_*)U_*(s)\mathrm{d}s + \sum_{e\in E}\sqrt{\beta_e(x_*)}W_e(t)e. \tag{25}$$

Moreover, $U_x(t) \Rightarrow \mathscr{N}(0, \Sigma_*)$ as $t \to +\infty$, and the process U_* is stationary,i.e. $U_*(t+\cdot) \sim U_*$ for all $t \geq 0$.

Proof. Let $x \in \mathscr{D}$. Using Itô's lemma and the relation $\mathrm{d}\Psi_x(t,0)/\mathrm{d}t = F'\big(\varphi_x(t)\big)\Psi_x(t,0)$, we see that U_x solves the SDE (24) if and only if

$$\Psi(0,t)U_x(t) = \sum_{e\in E}\int_0^t \sqrt{\beta_e(\varphi_x(s))}\Psi_x(0,s)e\,\mathrm{d}W_e(s)$$

almost surely for all $t \geq 0$. Thus (24) has a unique strong solution given by the formula

$$U_x(t) = \sum_{e\in E}\int_0^t \sqrt{\beta_e(\varphi_x(s))}\Psi_x(t,s)e\,\mathrm{d}W_e(s).$$

A similar argument shows that (25) has a unique strong solution given by

$$U_*(t) = e^{tF'(x_*)}U_*(0) + \sum_{e\in E}\int_0^t \sqrt{\beta_e(\varphi_x(s))}e^{(t-s)F'(x_*)}e\,\mathrm{d}W_e(s).$$

For all $t \geq 0$,

$$\begin{aligned}\mathbf{E}\Big(U_x(t)U_x(t)^T\Big) &= \int_0^t \Psi_x(t,s)\left(\sum_{e\in E}\beta_e\big(\varphi_x(s)\big)e\,e^T\right)\Psi_x(t,s)^T\mathrm{d}s \\ &= \int_0^\infty \mathbf{1}_{\{u\leq t\}}\Psi_x(t,t-u)\left(\sum_{e\in E}\beta_e\big(\varphi_x(t-u)\big)e\,e^T\right)\Psi_x(t,t-u)^T\mathrm{d}u.\end{aligned}$$

Using Lemma 3.1 and the boundedness of the functions β_e on $\mathscr{D}$, we see that the norm of the above integrand is dominated by $u \mapsto Ce^{-\rho_* u}$, for some $C > 0$. Let Γ_1 and Γ_2 be given by Lemma 3.1. For all $t \geq u$, (18) yields

$$\left\|\Psi_x(t,t-u) - e^{uF'(x_*)}\right\| \leq \int_{t-u}^t \Gamma_2\|F'\|_{\mathrm{Lip},\mathscr{D}}\Gamma_1 e^{-\rho_* r/2}\Gamma_2\mathrm{d}r \leq 2\rho_*^{-1}\Gamma_1\Gamma_2^2\|F'\|_{\mathrm{Lip},\mathscr{D}}e^{-\rho_*(t-u)/2}.$$

Hence, as $t \to +\infty$, we have $\Psi_x(t,t-u) \to e^{uF'(x_*)}$. Moreover $\varphi_x(t-u) \to x_*$, thus by dominated convergence we obtain

$$\mathbf{E}\Big(U_x(t)U_x(t)^T\Big) \underset{t\to+\infty}{\longrightarrow} \Sigma_*.$$

Since U_x is a centered Gaussian process, this implies $U_x(t) \Rightarrow \mathscr{N}(0,\Sigma_*)$ as $t \to +\infty$.

Let us show that the process U_* is stationary. Since it satisfies the SDE with constant coefficients (25), it is Markovian, thus it is enough to show that all its marginals are distributed as $\mathscr{N}(0,\Sigma_*)$. For all $t \geq 0$, the independence of $U_*(0)$ and U_{x_*} entails that $U_*(t)$ is a centered Gaussian random vector with covariance matrix

$$\begin{aligned}\mathbf{E}\left(U_*(t)U_*(t)^T\right) &= e^{tF'(x_*)}\Sigma^* e^{tF'(x_*)^T} + \int_0^t e^{(t-s)F'(x_*)} S_* e^{(t-s)F'(x_*)^T} \mathrm{d}s \\ &= e^{tF'(x_*)}\Sigma^* e^{tF'(x_*)^T} + \Sigma_* - \int_t^\infty e^{sF'(x_*)} S_* e^{sF'(x_*)^T} \mathrm{d}s \\ &= \Sigma_*,\end{aligned}$$

which ends the proof.

3.1.4 Chernoff bounds for Poisson process and Brownian motion

We give exponential bounds on the tail probabilities of the supremum norm of a compensated Poisson Process (Lemma 3.4), and a Brownian stochastic integral (Lemma 3.5) on a given time interval. These results are not new and we provide the proofs here for the sake of completeness. We follow the standard method which consists in optimizing over a one-parameter family of Chernoff bounds obtained via Doob's maximal inequality.

Lemma 3.4. *Let P be a standard Poisson process. For all $S, A > 0$ such that $A \leq 2\log(2)S$, we have*

$$\mathbf{P}\left(\sup_{0\leq s\leq S} |P(s) - s| \geq A\right) \leq 2\exp\left(-\frac{A^2}{4S}\right). \tag{26}$$

Proof. Let $\tilde{P}$ be the compensated process defined by $\tilde{P}(t) = P(t) - t$. Let $S, A > 0$ such that $A \leq 2\log(2)S$, and let $\xi > 0$. We have

$$\mathbf{P}\left(\sup_{0\leq s\leq S} \left|\tilde{P}(s)\right| \geq A\right) \leq \mathbf{P}\left(\sup_{0\leq s\leq S} e^{\xi\tilde{P}(s)} \geq e^{\xi A}\right) + \mathbf{P}\left(\sup_{0\leq s\leq S} e^{-\xi\tilde{P}(s)} \geq e^{\xi A}\right). \tag{27}$$

Since $\tilde{P}$ is a càdlàg martingale (with respect to its canonical filtration), $e^{\xi\tilde{P}}$ and $e^{-\xi\tilde{P}}$ are càdlàg submartingales and Doob's maximal inequality yields

$$\begin{aligned}\mathbf{P}\left(\sup_{0\leq s\leq S} \left|\tilde{P}(s)\right| \geq A\right) &\leq e^{-\xi A}\left(\mathbf{E}\left(e^{\xi\tilde{P}(S)}\right) + \mathbf{E}\left(e^{-\xi\tilde{P}(S)}\right)\right) \\ &\leq e^{-\xi A}\left(e^{S(e^\xi - \xi - 1)} + e^{S(e^{-\xi}+\xi-1)}\right).\end{aligned}$$

Let us choose $\xi = A/2S$. By hypothesis, $\xi \leq \log(2)$, hence $e^{\pm\xi} \mp \xi - 1 \leq \xi^2$. Consequently,

$$\mathbf{P}\left(\sup_{0\leq s\leq S} \left|\tilde{P}(s)\right| \geq A\right) \leq 2e^{-\xi A + S\xi^2} = 2\exp\left(-\frac{A^2}{4S}\right).$$

□

We say that a filtration $(\mathscr{F}_t)_{0\leq t\leq\infty}$ defined on probability space $(\Omega, \mathscr{F}, \mathbf{P})$ satisfies the *usual conditions* if it is complete, i.e. $\mathscr{F}_0$ contains the $\mathbf{P}$-null sets of $\mathscr{F}_\infty$, and right-continuous.

Lemma 3.5. *Let B be a real Brownian motion with respect to a filtration $(\mathscr{F}_t)_{0\leq t\leq\infty}$ satisfying the usual conditions. Let $A, S, \rho > 0$, and let R be a $(\mathscr{F}_t)$-progressive process such that, almost surely,*

$$|R| \leq \rho \quad \textit{almost everywhere on } [0, S].$$

Then

$$\mathbf{P}\left(\sup_{0\leq s\leq S} \left|\int_0^s R(r)\mathrm{d}B(r)\right| \geq A\right) \leq 2\exp\left(-\frac{A^2}{2S\rho^2}\right). \tag{28}$$

Proof. We can proceed in a similar way as in the proof of Lemma 3.4. Let M be the $(\mathscr{F}_t)$-martingale defined by

$$M(t) = \int_0^t R(s) \mathrm{d}B(s).$$

For all $\xi > 0$, we have

$$\mathbf{P}\left(\sup_{0\leq s\leq S} |M(s)| \geq A\right) \leq e^{-\xi A}\left(\mathbf{E}\left(e^{\xi M(S)}\right) + \mathbf{E}\left(e^{-\xi M(S)}\right)\right). \tag{29}$$

The Doléans-Dade exponentials

$$\mathcal{E}_{\pm\xi M}(t) = \exp\left(\pm\xi M(t) - \frac{\xi^2}{2}\int_0^t R^2(s)\mathrm{d}s\right)$$

are positive local martingales as shown by Ito's lemma, thus supermartingales. Consequently,

$$\mathbf{E}\left(e^{\pm\xi M(S)}\right) \leq e^{\xi^2 S\rho^2/2}\mathbf{E}\big(\mathcal{E}_{\pm\xi M}(S)\big) \leq e^{\xi^2 S\rho^2/2}.$$

We conclude by plugging this inequality into (29) and choosing $\xi = A/(S\rho^2)$. $\square$

From Lemma 3.5 we can deduce an exponential bound on the probability of large oscillations of Brownian motion.

Lemma 3.6. *Let B be a real Brownian motion. For all $S, T, A > 0$, we have*

$$\mathbf{P}\left(\sup_{\substack{0\leq s,t\leq T\\ |t-s|\leq S}} |B(t) - B(s)| \geq A\right) \leq 2\left\lceil\frac{T}{S}\right\rceil \exp\left(-\frac{A^2}{18S}\right). \tag{30}$$

Proof. Let $S, T, A > 0$. It follows easily from the triangular inequality that

$$\sup_{\substack{0\leq s,t\leq T\\ |t-s|\leq S}} |B(t) - B(s)| \leq 3 \sup_{k\in\{0,\ldots,\lceil T/S\rceil-1\}} \sup_{0\leq r\leq S} |B(kS+r) - B(kS)|.$$

Since for all $k \in \mathbf{N}$ the process $\big(B(kS+r) - B(kS)\,;\, r\geq 0\big)$ is a Brownian motion, we get

$$\mathbf{P}\left(\sup_{\substack{0\leq s,t\leq T\\ |t-s|\leq S}} |B(t) - B(s)| \geq A\right) \leq \left\lceil\frac{T}{S}\right\rceil \mathbf{P}\left(\sup_{0\leq r\leq S} |B(r)| \geq A/3\right).$$

We conclude by applying Lemma 3.5 with $R \equiv 1$. $\square$

3.2 Proof of Theorem 1.1

Let $r_0 > 0$ be such that the compact $\mathscr{D}' := \mathscr{D} + \bar{B}(0, r_0)$ is a subset of $\mathscr{U}$. Let $\Gamma \geq 1$ be given by Lemma 3.2, and set

$$M_0 = \sum_{e\in E}\|e\|,\ M_1 = \max_{e\in E}\|\beta_e\|_{\infty,\mathscr{D}'},\ M_2 = \max_{e\in E}\|\beta_e\|_{\mathrm{Lip},\mathscr{D}'},\ M_3 = \|F'\|_{\mathrm{Lip},\mathscr{D}'},\ M_4 = \max_{e\in E}\left\|\sqrt{\beta_e}\right\|_{\mathrm{Lip},\mathscr{D}'}.$$

Note that M_1 and M_2 are finite due to the fact the β_e are $\mathcal{C}^1$ on $\mathscr{U}$, while M_3 and M_4 are finite because for each $e \in E$ the gradient of β_e and $\sqrt{\beta_e}$ are locally Lipschitz on $\mathscr{U}$. We say that (B, P) is a *KMT coupling* when B is a Brownian motion and P is a Poisson Process such that

$$\mathbf{P}\left[\sup_{0\leq t\leq T} |P(t) - t - B(t)| > c\log(T) + u\right] \leq ae^{-bu}, \tag{31}$$

for all $T \geq 1$ and $u \geq 0$, where a, b, c are the constants appearing in (1).

Let us fix $K > 0$ and $x \in \mathscr{D}$ for the rest of the proof. The first step is to construct the coupling $\big(X_x^K, U_x\big)$ of (μ_x^K, ν_x).

Proposition 3.7. *We can construct a probability space $(\Omega, \mathscr{F}, \mathbf{P})$, equipped with*

a) *a filtration $(\mathscr{F}_t)_{0\le t\le\infty}$ satisfying the usual conditions, a $\mathbf{R}^E$-valued $(\mathscr{F}_t)$-Brownian motion $W = (W_e(t)\,;\, e\in E, t\ge 0)$, and a $(\mathscr{F}_t)$-adapted, d-dimensional continuous process Y_x^K such that, $\mathbf{P}$-almost surely for all $t \le \inf\{s\ge 0 : Y(s)\notin \mathscr{D}'\}$,*

$$Y_x^K(t) = \frac{\lfloor Kx\rfloor}{K} + \int_0^t F(Y_x^K(s))\mathrm{d}s + \frac{1}{\sqrt{K}}\sum_{e\in E}\left(\int_0^t \sqrt{\beta_e\left(Y_x^K(s)\right)}\,\mathrm{d}W_e(s)\right)e\,; \tag{32}$$

b) *a family $(B_{e,j}\,;\, e\in E, j\in\mathbf{N})$ of mutually independent real Brownian motions such that, for all $e\in E$, $j\in\mathbf{N}$, and $\mathbf{P}$-almost surely for all $j\le t\le (j+1)\wedge\inf\{s\ge 0 : Y(s)\notin\mathscr{D}'\}$,*

$$Y_x^K(t) = Y_x^K(j) + \int_j^t F(Y_x^K(s))\mathrm{d}s + \frac{1}{K}B_{e,j}\left(K\int_j^t \beta_e(Y_x^K(s))\mathrm{d}s\right); \tag{33}$$

c) *a $(\mathscr{F}_t)$-adapted, d-dimensional continuous process U_x, of probability distribution ν_x, solution of*

$$U_x(t) = \int_0^t F'(\varphi_x(s))U_x(s)\mathrm{d}s + \sum_{e\in E}\left(\int_0^t \sqrt{\beta_e(\varphi_x(s))}\,\mathrm{d}W_e(s)\right)e\,; \tag{34}$$

d) *a family $(P_{e,j}\,;\, e\in E, j\in\mathbf{N})$ of mutually independent Poisson processes, such that for all $e\in E$, $j\in\mathbf{N}$, $(B_{e,j}, P_{e,j})$ is a KMT coupling;*

e) *a d-dimensional càdlàg process X_x^K of probability distribution μ_x^K, such that for all $j\in\mathbf{N}$, $\mathbf{P}$-almost surely for all $j\le t\le j+1$,*

$$X_x^K(t) = X_x^K(j) + \frac{1}{K}\sum_{e\in E} P_{e,j}\left(K\int_j^t \beta_e(X_x^K(s))\,\mathrm{d}s\right)e. \tag{35}$$

Let us make some comments. This construction involves a diffusion Y_x^K, which admits the representation (33), involving time-changed Brownian motions. The analogous representation (35) of X_x^K, involving time-changed Poisson processes, enables to use KMT couplings. The coupling between Y_x^K and the Gaussian process $\varphi_x + U_x/\sqrt{K}$ is more straightforward : we use the same family of Brownian motions W_e to drive them both.

The coupling is based on the construction of Kurtz in [21], but here we use different KMT couplings on each time interval $[j, j+1]$. That way, gaps between the time changes of Poisson processes and Brownian motion are suitably controlled, even for large t. This is crucial since we are interested in large time scales.

Proof. Let $\chi\colon \mathbf{R}^d\to\mathbf{R}_+$ be a continuous function with compact support such that $\chi_{|\mathscr{D}'} = 1$. The functions χF and $\chi\beta_e$ are continuous and bounded, thus Theorem 2.2 in [17, Chapter IV] yields the existence of a probability space $(\Omega,\mathscr{F},\mathbf{P})$ equipped with a filtration $(\mathscr{F}_t)_{0\le t\le\infty}$ satisfying the usual conditions, a $\mathbf{R}^E$-valued $(\mathscr{F}_t)$-Brownian motion $W = (W_e(t)\,;\, e\in E, t\ge 0)$, and a $(\mathscr{F}_t)$-adapted, d-dimensional càdlàg process Y_x^K such that, $\mathbf{P}$-almost surely for all $t\ge 0$,

$$Y_x^K(t) = \frac{\lfloor Kx\rfloor}{K} + \int_0^t (\chi F)\big(Y_x^K(s)\big)\mathrm{d}s + \frac{1}{\sqrt{K}}\sum_{e\in E}\left(\int_0^t \sqrt{(\chi\beta_e)\big(Y_x^K(s)\big)}\,\mathrm{d}W_e(s)\right)e.$$

This equation implies (32) almost surely for all $t\le\inf\left\{s\ge 0 : Y_x^K(s)\notin\mathscr{D}'\right\}$, using that $\chi_{|\mathscr{D}'} = 1$.

Enlarging the filtered probability space, we may suppose that there exists mutually independent real Brownian motions $\left(\tilde{B}_{e,j}\,;\, e\in E, j\in\mathbf{N}\right)$ and mutually independent random variables $(V_{e,j}\,;\, e\in E, j\in\mathbf{N})$ uniformly distributed on $[0,1]$, such that the sigma-fields $\sigma\left(V_{e,j}\,;\, e\in E, j\in\mathbf{N}\right)$, $\sigma\left(\tilde{B}_{e,j}\,;\, e\in E, j\in\mathbf{N}\right)$ and $\mathscr{F}_\infty$ are mutually independent.

Let us deal with b). Given Equation (32), what we need is to construct a family $(B_{e,j}\,;\,e\in E, j\in\mathbf{N})$ of mutually independent real Brownian motions such that, for all $e\in E$ and $j\in\mathbf{N}$, we have, almost surely for all $t\geq 0$,

$$B_{e,j}\left(K\int_j^t \beta_e(Y_x^K(s))\mathrm{d}s\right)=\sqrt{K}\int_j^t\sqrt{\beta_e(Y_x^K(s))}\mathrm{d}W_e(s). \tag{36}$$

For all $e\in E$ and $j\in\mathbf{N}$, define the process $M_{e,j}$ by

$$M_{e,j}(t)=\int_0^t \mathbf{1}_{\{j\leq s\leq j+1\}}\sqrt{K(\chi\beta_e)\big(Y_x^K(s)\big)}\mathrm{d}W_e(s).$$

It is a continuous $(\mathscr{F}_t)$-local martingale starting from 0 with quadratic variation given by

$$\langle M_{e,j}\rangle(t)=\int_0^t \mathbf{1}_{\{j\leq s\leq j+1\}}K(\chi\beta_e)\big(Y_x^K(s)\big)\mathrm{d}s.$$

Moreover, $(M_{e,j}\,;\,e\in E, j\in\mathbf{N})$ is an orthogonal family, in the sense that $(e,j)\neq(e',j')$ implies $\langle M_{e,j},M_{e',j'}\rangle\equiv 0$, where $\langle\cdot,\cdot\rangle$ denotes the quadratic covariation. For all $u\geq 0$, define the $(\mathscr{F}_t)$-stopping time

$$\tau_{e,j}(u)=\inf\left\{t\geq 0:\langle M_{e,j}\rangle(t)>u\right\},$$

and define the process $B_{e,j}$ by

$$B_{e,j}(u)=\begin{cases}M_{e,j}\left(\tau_{e,j}(u)\right) & \text{if } u<\langle M_{e,j}\rangle(j+1)\\ M_{e,j}\left(j+1\right)+\tilde{B}_{e,j}\left(u-\langle M_{e,j}\rangle(j+1)\right) & \text{if } u\geq\langle M_{e,j}\rangle(j+1)\end{cases}.$$

Then, $(B_{e,j}, e\in E, j\in\mathbf{N})$ is a family of mutually independent real Brownian motions, see Theorem 1.10 in [27, Chapter V]. The fact that $B_{e,j}$ is a Brownian motion is essentially Dambis-Dubins-Schwarz's theorem, but we need $\tilde{B}_{e,j}$ to extend $B_{e,j}$ after time $\langle M_{e,j}\rangle(j+1)$, which is finite. As for the independence of the $B_{e,j}$, it comes from the orthogonality of the $M_{e,j}$.

To conclude the proof of b), we still need to verify (36). Set $\tau_{e,j}^-(0)=0$ and $\tau_{e,j}^-(u)=\lim_{v\to u,v<u}\tau_{e,j}(v)$ for all $u>0$. The process $\langle M_{e,j}\rangle$ is constant on $[\tau_{e,j}^-(u),\tau_{e,j}(u)]$ for all $u\geq 0$, almost surely, hence this is also the case for $M_{e,j}$. Moreover, almost surely for all $j\leq t\leq j+1$, we have $t\in\left[\tau_{e,j}^-\big(\langle M_{e,j}\rangle(t)\big),\tau_{e,j}\big(\langle M_{e,j}\rangle(t)\big)\right]$, thus

$$M_{e,j}(t)=M_{e,j}\Big[\tau_{e,j}\big(\langle M_{e,j}\rangle(t)\big)\Big]=B_{e,j}\big(\langle M_{e,j}\rangle(t)\big).$$

This entails (36) almost surely for all $t\leq\inf\left\{s\geq 0:Y_x^K(s)\notin\mathscr{D}'\right\}$.

Now, let us turn to d). The definition of the $P_{e,j}$ should satisfy two constraints : $(B_{e,j},P_{e,j})$ should be a KMT coupling, and the $P_{e,j}$ should form an independent family. In order to do that, we use the $V_{e,j}$ and Lemma 3.12, which guarantees the existence of a measurable function $G:\mathcal{C}(\mathbf{R}_+,\mathbf{R}^d)\times[0,1]\to\mathcal{D}(\mathbf{R}_+,\mathbf{R}^d)$ such that, if B is a real Brownian motion, and V is uniformly distributed on $(0,1)$ and independent of B, then $(B,G(B,V))$ is a KMT coupling. Thus, we set $P_{e,j}=G(B_{e,j},V_{e,j})$, and d) is satisfied.

Finally, let us define the process X_x^K. For each $y\in\mathscr{D}\cap K^{-1}\mathbf{Z}^d$, it follows from Theorem 4.1 in [13, Chapter 6] that there exists a unique d-dimensional càdlàg process $\big(X'_{y,j}(t)\,;\,t\geq 0\big)$ satisfying the equation

$$X'_{y,j}(t)=y+\frac{1}{K}\sum_{e\in E}P_{e,j}\left(\int_0^t\beta_e(X'_{y,j}(s))\mathrm{d}s\right)e,$$

and we have $X'_{y,j}\sim\mu_y^K$. What's more, $\sigma(X'_{y,j})\subset\sigma(P_{e,j}\,;\,e\in E)$. Now, define X_x^K by $X_x^K(0)=\lfloor Kx\rfloor/K$ and, for all $j\in\mathbf{N}$ and $j<t\leq j+1$,

$$X_x^K(t)=X'_{X_x^K(j),j}(t-j).$$

It is not hard to prove by induction that $\sigma\left(X_x^K(t)\,;\,0\le t\le j\right)\subset\sigma(P_{e,i}\,;\,e\in E, 0\le i\le j-1)$ and that $(X_x^K(t)\,;\,0\le t\le j)$ is a $K^{-1}\mathbf{Z}^d$-valued continuous-time Markov chain, with transition rate from y to $z\neq y$ equal to $\tilde{q}^K_{y,z}:=q^K_{Ky,Kz}=K\beta_{K(z-y)}(y)$. The key point is that conditional on $X_x^K(j)$, the process $(X_x^K(t)\,;\,j\le t\le j+1)$ is a continuous-time Markov chain with transition rates $\left(\tilde{q}^K_{y,z}\right)$ starting from $X_x^K(j)$, and independent of $\sigma\left(X_x^K(t)\,;\,0\le t\le j\right)$. Hence, $X_x^K\sim\mu_x^K$. □

In the rest of the proof, we generally write $(X,Y,U)=\left(X_x^K,Y_x^K,U_x\right)$. The quantities we call constants may only depend on the M_i, r_0, the constants a,b,c involved in (1), and the cardinal of E, which we denote by $|E|$. When we say that an assertion holds 'for K large enough', we mean that there exists $K_0>0$, *independent of* x, such that the assertion is true if $K\ge K_0$. For all $(e,j)\in E\times\mathbf{N}$, we denote by $\tilde{P}_{e,j}$ the compensated Poisson process associated to $P_{e,j}$, defined by $\tilde{P}_{e,j}(t)=P_{e,j}(t)-t$.

The next step is to study the deviations of X_x^K from the fluid limit φ_x. Taking $K^{-1/2}\ll\eta(K)\ll 1$ in the next proposition corresponds to moderate deviations of $X_x^K-\varphi_x$, while taking η constant corresponds to large deviations.

Proposition 3.8. *There exist constants $V_0,\eta_0>0$ such that for all $\eta:\mathbf{R}_+^*\to\mathbf{R}_+^*$ satisfying $K^{-1/2}\ll\eta(K)\le\eta_0$, we have, for K large enough and for all $t\ge 0$:*

$$\mathbf{P}\left(\sup_{0\le s\le t}\left\|X_x^K(s)-\varphi_x(s)\right\|>\eta(K)\right)\le 2|E|(t+1)\exp\left(-V_0K\eta^2(K)\right). \tag{37}$$

Proof. Set

$$\eta_0=(6\Gamma M_3)^{-1}\wedge(12\log(2)\Gamma M_0M_1)\wedge r_0.$$

Let $\eta:\mathbf{R}_+^*\to\mathbf{R}_+^*$ be such that $K^{-1/2}\ll\eta(K)\le\eta_0$, and set

$$\tau_\eta=\inf\Big\{t\ge 0:\left\|X(t)-\varphi_x(t)\right\|>\eta(K)\Big\}.$$

Note that since $\eta_0\le r_0$, we have $X(t)\in\mathscr{D}'$ almost surely for all $t<\tau_\eta$. Using (35), we get, almost surely for all $t\ge 0$,

$$X(t)=\frac{\lfloor Kx\rfloor}{K}+\int_0^t F\big(X(s)\big)\mathrm{d}s+\sum_{0\le j\le\lfloor t\rfloor}A_j(t)$$

where

$$A_j(t)=\mathbf{1}_{\{t\ge j\}}\frac{1}{K}\sum_{e\in E}\tilde{P}_{e,j}\left(K\int_j^{t\wedge(j+1)}\beta_e\big(X(s)\big)\mathrm{d}s\right)e.$$

Recalling that

$$\varphi_x(t)=x+\int_0^t F\big(\varphi_x(s)\big)\mathrm{d}s,$$

we can write

$$X(t)-\varphi_x(t)=\left(\frac{\lfloor Kx\rfloor}{K}-x\right)+\int_0^t F'\big(\varphi_x(s)\big)\big(X(s)-\varphi_x(s)\big)\mathrm{d}s+\sum_{0\le j\le\lfloor t\rfloor}\left(A_j+D_j\right)(t) \tag{38}$$

where

$$D_j(t)=\mathbf{1}_{\{t\ge j\}}\int_j^{t\wedge(j+1)}\Big[F\big(X(s)\big)-F\big(\varphi_x(s)\big)-F'\big(\varphi_x(s)\big)\big(X(s)-\varphi_x(s)\big)\Big]\mathrm{d}s.$$

Let $t\ge 0$. On the event $\{\tau_\eta\le t\}$, we have $\|X(t\wedge\tau_\eta)\|=\|X(\tau_\eta)\|\ge\eta(K)$ by right-continuity of X, thus

$$\{\tau_\eta\le t\}\subset\left\{\sup_{0\le s\le t\wedge\tau_\eta}\left\|X(s)-\varphi_x(s)\right\|\ge\eta(K)\right\}.$$

Now, consider Equation (38): it shows that $X - \varphi_x$ can be seen as a perturbation of a solution of the linear ODE $\dot{y} = F'(\varphi_x)y$. Thus, we can use the key Lemma 3.2, which allows us to control $X - \varphi_x$ in terms of the A_j and the D_j. Since for all $f : \mathbf{R}_+ \to \mathbf{R}^d$ and $T > 0$,

$$\sup_{\substack{0\le r,s\le T\\ |s-r|\le 1}} \|f(s) - f(r)\| \le 3 \max_{0\le j\le \lfloor T\rfloor} \sup_{j\le s\le (j+1)\wedge T} \|f(s) - f(j)\|,$$

and since $\|\lfloor Kx\rfloor/K - x\| \le \sqrt{d}/K \ll \eta(K)$, we obtain, for K large enough,

$$\left\{ \sup_{0\le s\le t\wedge\tau_\eta} \big\|X(s) - \varphi_x(s)\big\| \ge \eta(K) \right\} \subset \left\{ \max_{0\le j\le\lfloor t\rfloor} \sup_{j\le s\le t\wedge\tau_\eta} \big\|(A_j + D_j)(s)\big\| \ge \eta(K)/(3\Gamma) \right\}.$$

We recall that $\Gamma \ge 1$ is given by Lemma 3.2. Consequently, setting $\eta'(K) = \eta(K)/(3\Gamma)$,

$$\mathbf{P}\left(\tau_\eta \le t\right) \le (t+1) \max_{0\le j\le\lfloor t\rfloor} \mathbf{P}\left(\sup_{j\le s\le t\wedge\tau_\eta} \big\|(A_j + D_j)(s)\big\| \ge \eta'(K) \right). \tag{39}$$

Let us bound the right handside of this inequality. For K large enough, for all $0 \le j \le \lfloor t\rfloor$ and $j \le s \le t\wedge\tau_\eta$, we have

$$\begin{aligned}
\|D_j(s)\| &\le \int_j^{s\wedge(j+1)} \Big\| F\big(X(r)\big) - F\big(\varphi_x(r)\big) - F'\big(\varphi_x(r)\big)\big(X(r) - \varphi_x(r)\big) \Big\| \mathrm{d}r \\
&\le \int_j^{s\wedge(j+1)} M_3 \big\|X(r) - \varphi_x(r)\big\|^2 \mathrm{d}r \\
&\le M_3 \eta(K)^2 \\
&\le \eta'(K)/2,
\end{aligned}$$

where we used that $6\Gamma M_3 \eta_0 \le 1$ for the last inequality. Hence (39) yields, for K large enough,

$$\mathbf{P}\left(\tau_\eta \le t\right) \le (t+1) \max_{0\le j\le\lfloor t\rfloor} \mathbf{P}\left(\sup_{j\le s\le t\wedge\tau_\eta} \big\|A_j(s)\big\| \ge \eta'(K)/2 \right).$$

Now, for all $0 \le j \le \lfloor t\rfloor$ and all $j \le s \le t\wedge\tau_\eta$, we have

$$\max_{e\in E} \left(\int_j^{s\wedge(j+1)} \beta_e\big(X(r)\big)\mathrm{d}r \right) \le \max_{e\in E} \|\beta_e\|_{\infty,\mathscr{D}'} \le M_1,$$

thus

$$\sup_{j\le s\le t\wedge\tau_\eta} \big\|A_j(s)\big\| \le K^{-1} M_0 \max_{e\in E} \sup_{0\le s\le KM_1} |\tilde{P}_{e,j}\left(s\right)|.$$

Letting $\tilde{P}$ denote a compensated Poisson process, we get

$$\mathbf{P}\left(\tau_\eta \le t\right) \le |E|(t+1)\mathbf{P}\left(\sup_{0\le s\le KM_1} |\tilde{P}(s)| \ge K\eta''(K) \right), \tag{40}$$

where $\eta''(K) = \eta'(K)/(2M_0) = \eta(K)/(6\Gamma M_0)$. The inequality $\eta(K) \le \eta_0$ entails $\eta''(K) \le 2\log(2)M_1$, thus we can use Lemma 3.4, which yields

$$\mathbf{P}\left(\tau_\eta \le t\right) \le 2|E|(t+1)\exp\left(-\frac{K\eta''(K)^2}{4M_1} \right).$$

This entails (37) with $V_0 = \big(144\Gamma^2 M_0^2 M_1\big)^{-1}$, hence the proposition is proved. $\square$

Next, in Proposition 3.9 (resp. Proposition 3.10), we obtain upper bounds on the probability that $\|X - Y\|$ (resp. $\|Y - \varphi_x - U/\sqrt{K}\|$) exceeds a level $\varepsilon(K)$. Once again, the idea of the proof is to see the process as solution of a perturbed linear ODE and use Lemma 3.2.

Proposition 3.9. *There exist constants $C_1, V_1, \alpha_1 > 0$, such that for every $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ satisfying $\alpha_1 \log(K)/K \le \varepsilon(K) \ll 1$, we have, for K large enough and for all $t \ge 0$:*

$$\mathbf{P}\left(\sup_{0\le s\le t} \left\| X_x^K(s) - Y_x^K(s) \right\| > \varepsilon(K) \right) \le C_1(t+1)\exp\left(-V_1 K\varepsilon(K)\right). \tag{41}$$

Proof. Let $\varepsilon, \eta : \mathbf{R}_+^* \to \mathbf{R}_+^*$ be such that $K^{-1} \ll \varepsilon(K) \ll 1$ and $K^{-1/2} \ll \eta(K) \ll 1$. Set

$$\sigma_\varepsilon = \inf\Big\{ t \ge 0 : \big\|X(t) - Y(t)\big\| > \varepsilon(K) \Big\} \quad \text{and} \quad \tau_\eta = \inf\Big\{ t \ge 0 : \big\|X(t) - \varphi_x(t)\big\| > \eta(K) \Big\}.$$

The scale η will be specified later, as a function of ε. Since $\varepsilon(K) + \eta(K) \ll 1$, for K large enough both $Y(t)$ and $X(t)$ belong to $\mathscr{D}'$ almost surely for all $t < \tau_\eta \wedge \sigma_\varepsilon$. Thus, using equations (38) and (**??**), we obtain that almost surely for all $t \le \tau_\eta \wedge \sigma_\varepsilon$,

$$X(t) - Y(t) = \int_0^t F'\big(\varphi_x(s)\big)\big(X(s) - Y(s)\big)\mathrm{d}s + \sum_{0\le j\le \lfloor t\rfloor} \left(H_j + J_j + L_j\right)(t), \tag{42}$$

where

$$\begin{aligned}
H_j(t) &= \mathbf{1}_{\{t\ge j\}} \frac{1}{K} \sum_{e\in E} \left(\tilde{P}_{e,j} - B_{e,j}\right)\left(K\int_j^{t\wedge(j+1)} \beta_e\big(X(s)\big)\mathrm{d}s \right) e\\
I_j(t) &= \mathbf{1}_{\{t\ge j\}} \frac{1}{K} \sum_{e\in E} \left[B_{e,j}\left(K\int_j^{t\wedge(j+1)} \beta_e\big(X(s)\big)\mathrm{d}s \right) - B_{e,j}\left(K\int_j^{t\wedge(j+1)} \beta_e\big(Y(s)\big)\mathrm{d}s \right)\right] e\\
L_j(t) &= \mathbf{1}_{\{t\ge j\}} \int_j^{t\wedge(j+1)} \Big[F\big(X(s)\big) - F\big(Y(s)\big) - F'\big(\varphi_x(s)\big)\big(X(s) - Y(s)\big)\Big]\mathrm{d}s.
\end{aligned}$$

Let $t \ge 0$. We have

$$\begin{aligned}
\big\{\sigma_\varepsilon \le t\big\} &\subset \big\{\tau_\eta < t\big\} \cup \big\{\sigma_\varepsilon \le t \le \tau_\eta\big\}\\
&\subset \big\{\tau_\eta < t\big\} \cup \left\{ \sup_{0\le s\le t\wedge\tau_\eta\wedge\sigma_\varepsilon} \big\|X(s) - Y(s)\big\| \ge \varepsilon(K) \right\}.
\end{aligned}$$

Using Equation (42), we apply Lemma 3.2 with $X - Y$ playing the role of y and we get

$$\left\{ \sup_{0\le s\le t\wedge\tau_\eta\wedge\sigma_\varepsilon} \big\|X(s) - Y(s)\big\| \ge \varepsilon(K) \right\} \subset \left\{ \max_{0\le j\le \lfloor t\wedge\tau_\eta\wedge\sigma_\varepsilon\rfloor} \sup_{j\le s\le t\wedge\tau_\eta\wedge\sigma_\varepsilon} \big\|(H_j + I_j + L_j)(s)\big\| \ge \varepsilon'(K) \right\},$$

where $\varepsilon'(K) = \varepsilon(K)/(3\Gamma)$. Hence, for K large enough we have

$$\begin{aligned}
\mathbf{P}\left(\sigma_\varepsilon < t\right) &\le \mathbf{P}\left(\tau_\eta < t\right) + \mathbf{P}\left(\max_{0\le j\le\lfloor t\rfloor} \sup_{j\le s\le t\wedge\tau_\eta\wedge\sigma_\varepsilon} \big\|(H_j + I_j + L_j)(s)\big\| \ge \varepsilon'(K) \right)\\
&\le (t+1)\left[2|E|\exp\big(-V_0 K\eta^2(K)\big) + \max_{0\le j\le\lfloor t\rfloor} \mathbf{P}\left(\sup_{j\le s\le t\wedge\tau_\eta\wedge\sigma_\varepsilon} \big\|(H_j + I_j + L_j)(s)\big\| \ge \varepsilon'(K) \right)\right],
\end{aligned}$$

where $V_0 > 0$ is given by Lemma 3.8. Moreover, for K large enough, for all $0 \le j \le \lfloor t\rfloor$ and for all $j \le s \le t \wedge \tau_\eta \wedge \sigma_\varepsilon$,

$$\begin{aligned}
\|L_j(s)\| &\le \int_j^{s\wedge(j+1)} \sup_{0\le\theta\le 1} \Big\|F'\big(\theta X(r) + (1-\theta)Y(r)\big) - F'\big(\varphi_x(r)\big)\Big\| \big\|Y(r) - X(r)\big\|\mathrm{d}r\\
&\le M_3\left(\eta(K) + \varepsilon(K)\right)\varepsilon(K)\\
&\le \varepsilon'(K)/3,
\end{aligned}$$

hence

$$\mathbf{P}\left(\sigma_\varepsilon \leq t\right) \leq 2|E|(t+1)\exp\left(-V_0 K\eta^2(K)\right) + (t+1)\max_{0\leq j\leq \lfloor t\rfloor}\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\left\|H_j(s)\right\| \geq \varepsilon'(K)/3\right)$$
$$+ (t+1)\max_{0\leq j\leq \lfloor t\rfloor}\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\left\|I_j(s)\right\| \geq \varepsilon'(K)/3\right). \tag{43}$$

Let $0 \leq j \leq \lfloor t\rfloor$. The term H_j is the error term due to the difference between the $\tilde{P}_{e,j}$ and the $B_{e,j}$, and we bound it thanks to the KMT estimate (31). We have

$$\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\|H_j(s)\| \geq \varepsilon'(K)/3\right) \leq \mathbf{P}\left(\max_{e\in E}\sup_{0\leq u\leq KM_1}\left|\tilde{P}_{e,j}(u) - B_{e,j}(u)\right| \geq K\varepsilon''(K)\right),$$

where $\varepsilon''(K) = \varepsilon'(K)/(3M_0) = \varepsilon(K)/(9\Gamma M_0)$. Let $a, b, c > 0$ be the constants involved in (31). If we suppose $\varepsilon''(K) \geq 2c\log(K)/K$, then

$$K\varepsilon''(K) \geq c\log(KM_1) + \left(K\varepsilon''(K)/2 - c\log(M_1)\right),$$

and the use of the KMT estimate yields, for K large enough :

$$\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\|H_j(s)\| \geq \varepsilon'(K)/3\right) \leq |E|\, a\exp\left(bcM_1\right)\exp\left(-\frac{bK\varepsilon''(K)}{2}\right). \tag{44}$$

Finally, let us bound the last term in (43), which comes from the difference between the time changes of $\tilde{P}_{e,j}$ and $B_{e,j}$. Recalling that $M_2 = \max_{e\in E}\|\beta_e\|_{\mathrm{Lip},\mathscr{D}'}$, we have, for all $j \leq s \leq t\wedge\tau_\eta\wedge\sigma_\varepsilon$,

$$\max_{e\in E}\left(\int_j^{s\wedge(j+1)}\left|\beta_e\big(X(r)\big) - \beta_e\big(Y(r)\big)\right|\mathrm{d}r\right) \leq M_2\varepsilon(K),$$

hence

$$\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\|I_j(s)\| \geq \varepsilon'(K)/3\right) \leq \mathbf{P}\left(\max_{e\in E}\sup_{\substack{0\leq r,s\leq KM_1\\ |s-r|\leq M_2K\varepsilon(K)}}|B_{e,j}(s) - B_{e,j}(r)| \geq K\varepsilon''(K)\right).$$

We control the oscillations of Brownian motion thanks to Lemma 3.6 and we get, setting $\varepsilon'''(K) = \varepsilon''(K)/(162\Gamma M_0 M_2)$:

$$\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\|I_j(s)\| \geq \varepsilon'(K)/3\right) \leq 2|E|\left\lceil\frac{M_1}{M_2\varepsilon(K)}\right\rceil\exp\left(-K\varepsilon'''(K)\right).$$

If we suppose $\varepsilon'''(K) \geq 2\log(K)/K$, then for K large enough $\lceil M_1/(M_2\varepsilon(K))\rceil \leq \exp\left(K\varepsilon'''(K)\right)$, thus

$$\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\tau_\eta\wedge\sigma_\varepsilon}\|I_j(s)\| \geq \varepsilon'(K)/3\right) \leq 2|E|\exp\left(-\frac{K\varepsilon'''(K)}{2}\right). \tag{45}$$

Now, it is time to fix η. We choose $\eta = \sqrt{\varepsilon}$, which satisfies the condition $K^{-1/2} \ll \eta(K) \ll 1$, and set

$$\alpha = 9\Gamma M_0\left(2c\wedge(324\Gamma M_0M_2)\right),\ C_1 = 2|E|\left(2 + a\exp\left(bcM_1\right)\right),\ V_1 = \left(V_0\wedge\frac{b}{18\Gamma M_0}\wedge\frac{1}{2916\Gamma^2M_0^2M_2}\right).$$

We conclude by combining (43), (44) and (45). We obtain that if $\varepsilon(K) \geq \alpha\log(K)/K$, then for K large enough, for all $t \geq 0$,

$$\mathbf{P}\left(\sigma_\varepsilon \leq t\right) \leq C_1(t+1)\exp\left(-V_1K\varepsilon(K)\right).$$

□

Next proposition provides the last piece of the puzzle..

Proposition 3.10. *There exists a constant $V_2 > 0$ such that for every $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ satisfying $K^{-1} \ll \varepsilon(K) \ll 1$, we have, for K large enough and for all $t \geq 0$:*

$$\mathbf{P}\left(\sup_{0\leq s\leq t}\left\|Y_x^K(s) - \varphi_x(s) - U_x(s)/\sqrt{K}\right\| > \varepsilon(K)\right) \leq (4|E|+1)(t+1)\exp\left(-V_2 K\varepsilon(K)\right). \tag{46}$$

Proof. Let $Z = \varphi_x + K^{-1/2}U$. It follows from (34) and the definition of φ_x that

$$Z(t) = x + \int_0^t F\big(\varphi_x(s)\big)\mathrm{d}s + \int_0^t F'\big(\varphi_x(s)\big)\big(Z(s) - \varphi_x(s)\big)\mathrm{d}s + \frac{1}{\sqrt{K}}\sum_{e\in E}\left(\int_0^t \sqrt{\beta_e\big(\varphi_x(s)\big)}\mathrm{d}W_e(s)\right)e,$$

almost surely for all $t \geq 0$. Using (32), we obtain that almost surely for all $t \leq \inf\{s \geq 0 : Y(s) \notin \mathscr{D}'\}$,

$$Y(t) - Z(t) = \left(\frac{\lfloor Kx\rfloor}{K} - x\right) + \int_0^t F'\big(\varphi_x(s)\big)\big(Y(s) - Z(s)\big)\mathrm{d}s + \sum_{0\leq j\leq \lfloor t\rfloor}\left(S_j + T_j\right)(t), \tag{47}$$

where

$$S_j(t) = \mathbf{1}_{\{t\geq j\}}\frac{1}{\sqrt{K}}\sum_{e\in E}\left(\int_j^{t\wedge(j+1)}\left(\sqrt{\beta_e\big(Y(s)\big)} - \sqrt{\beta_e\big(\varphi_x(s)\big)}\right)\mathrm{d}W_e(s)\right)e,$$
$$T_j(t) = \mathbf{1}_{\{t\geq j\}}\int_j^{t\wedge(j+1)}\Big[F\big(Y(s)\big) - F\big(\varphi_x(s)\big) - F'\big(\varphi_x(s)\big)\big(Y(s) - \varphi_x(s)\big)\Big]\mathrm{d}s.$$

The term S_j comes from the fact that the dispersion matrix appearing in the equation (34) defining U is $\sqrt{\beta_e(\varphi_x)}$, which follows the deterministic trajectory φ_x, whereas the equation (32) defining Y involves $\sqrt{\beta_e(Y)}$. As for the term T_j, it comes from the linearization of F along φ_x.

Let $\varepsilon, \eta : \mathbf{R}_+^* \to \mathbf{R}_+^*$ be such that $K^{-1} \ll \varepsilon(K)$ and $K^{-1/2} \ll \eta(K) \ll 1$. Set

$$\zeta_\varepsilon = \inf\{t \geq 0 : \|Y(t) - Z(t)\| > \varepsilon(K)\} \quad \text{and} \quad \theta_\eta = \inf\{t \geq 0 : \|Y(t) - \varphi_x(t)\| > \eta(K)\}.$$

Let $t \geq 0$, we have

$$\begin{aligned}\{\zeta_\varepsilon \leq t\} &\subset \{\theta_\eta < t\} \cup \{\zeta_\varepsilon \leq t \leq \theta_\eta\}\\ &\subset \{\theta_\eta < t\} \cup \left\{\sup_{0\leq s\leq t\wedge\theta_\eta}\|Y(s) - Z(s)\| \geq \varepsilon(K)\right\}.\end{aligned}$$

Since $\eta(K) \ll 1$, for K large enough we have $\theta_\eta \leq \inf\{s \geq 0 : Y(s) \notin \mathscr{D}'\}$ and thus (47) holds almost surely for all $t \leq \theta_\eta$. Applying Lemma 3.2 with $Y - Z$ playing the role of y, we get, for K large enough,

$$\left\{\sup_{0\leq s\leq t\wedge\theta_\eta}\big\|Y(s) - Z(s)\big\| \geq \varepsilon(K)\right\} \subset \left\{\max_{0\leq j\leq\lfloor t\rfloor}\sup_{j\leq s\leq t\wedge\theta_\eta}\big\|(S_j + T_j)(s)\big\| \geq \varepsilon'(K)\right\}$$

where $\varepsilon'(K) = \varepsilon(K)/(3\Gamma)$. Hence,

$$\begin{aligned}\mathbf{P}\left(\zeta_\varepsilon \leq t\right) \leq \mathbf{P}\left(\theta_\eta < t\right) &+ (t+1)\max_{0\leq j\leq\lfloor t\rfloor}\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\theta_\eta}\|S_j(s)\| \geq \varepsilon'(K)/2\right)\\ &+ (t+1)\max_{0\leq j\leq\lfloor t\rfloor}\mathbf{P}\left(\sup_{j\leq s\leq t\wedge\theta_\eta}\|T_j(s)\| > \varepsilon'(K)/2\right).\end{aligned} \tag{48}$$

We bound successively each term, before choosing an adequate η. First, if we let V_0 and V_1 be given by Proposition 3.8 and Proposition 3.9 respectively, then for K large enough, we have

$$\begin{aligned}\mathbf{P}(\theta_\eta < t) &\leq \mathbf{P}\left(\sup_{0\leq s<t}\|X(s) - \varphi_x(s)\| > \eta(K)/2\right) + \mathbf{P}\left(\sup_{0\leq s<t}\|X(s) - Y(s)\| > \eta(K)/2\right)\\ &\leq (t+1)\left[2|E|\exp\left(-V_0K\eta^2(K)/4\right) + C_1\exp\left(-V_1K\eta(K)/2\right)\right]\\ &\leq (2|E|+1)(t+1)\exp\left(-V_0K\eta^2(K)/4\right).\end{aligned} \tag{49}$$

Next, let us bound the second term of (48). Let $0 \leq j \leq \lfloor t \rfloor$. We have

$$\sup_{j \leq s \leq t \wedge \theta_\eta} \|S_j(s)\| \leq K^{-1/2} M_0 \max_{e \in E} \sup_{j \leq s \leq j+1} \left| \int_j^s R_e(r) \mathrm{d}W_e(r) \right|$$

where

$$R_e(r) = \left(\sqrt{\beta_e\big(Y(r)\big)} - \sqrt{\beta_e\big(\varphi_x(r)\big)} \right) \mathbf{1}_{\{r \leq \theta_\eta\}}.$$

Since the $\sqrt{\beta_e}$ are M_4-Lipschitz on $\mathscr{D}'$, we have $|R_e(r)| \leq M_4 \eta(K)$ for all $r \neq \theta_\eta$, hence

$$\begin{aligned} \max_{0 \leq j \leq \lfloor t \rfloor} \mathbf{P} \left(\sup_{j \leq s \leq t \wedge \theta_\eta} \|S_j(s)\| \geq \varepsilon'(K)/2 \right) &\leq \max_{0 \leq j \leq \lfloor t \rfloor} \mathbf{P} \left(\max_{e \in E} \sup_{j \leq s \leq j+1} \left| \int_j^s R_e(r) \mathrm{d}W_e(r) \right| \geq \frac{\sqrt{K} \varepsilon'(K)}{2M_0} \right) \\ &\leq 2|E| \exp \left(- \frac{K \varepsilon'^2(K)}{8 M_0^2 M_4^2 \eta^2(K)} \right), \end{aligned} \tag{50}$$

using Lemma 3.5 for the last inequality. In addition, we have

$$\sup_{j \leq s \leq t \wedge \theta_\eta} \|T_j(s)\| \leq M_3 \eta^2(K).$$

Let us choose $\eta = \sqrt{\varepsilon'/(2M_3)}$, which satisfies the condition $K^{-1/2} \ll \eta(K) \ll 1$. Due to the above inequality, the last term of the right handside of (48) vanishes. If we set

$$V_2 = \frac{V_0}{24 \Gamma M_3} \wedge \frac{M_3}{12 \Gamma M_0^2 M_4^2},$$

then the bounds (49) and (50) yield, for K large enough and for all $t \geq 0$,

$$\mathbf{P}\left(\zeta_\varepsilon < t\right) \leq (4|E| + 1)(t+1) \exp\left(-V_2 K \varepsilon(K)\right),$$

which ends the proof of the proposition. □

We conclude the proof of Theorem 1.1 by combining Proposition 3.9 and Proposition 3.10, using the triangular inequality : take $\alpha = 2\alpha_1$, $C = C_1 + 4|E| + 1$, $V = (V_1 \wedge V_2)/2$. ■

Let us mention that if we do not assume that the functions $\sqrt{\beta_e}$ are locally Lipschitz, we can prove a theorem similar to Theorem 1.1, with the following modifications : consider $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ such that $K^{-3/4} \ll \varepsilon(K) \ll 1$, and replace $\exp(-VK\varepsilon(K))$ by $\exp(-\tilde{V} K \varepsilon^{4/3}(K))$. Thus, in that context the gap between X_x^K and its Gaussian approximation remains smaller than $\delta/\sqrt{K}$ for a period of time of order $\exp\left(\tilde{V}\delta^{4/3}K^{1/3}\right)$. The proof is the same except that in Proposition 3.10 we can only dominate R_e by the square root of $\eta(K)$. This result can be useful for instance if the trajectory of φ_x spends time in a region where one of the functions β_e vanishes. However this is not the case in the models we consider in Section 2, at least in the neighbourhood of the equilibrium point x_*.

3.3 Proof of Corollary 1.2

We start by the following lemma, which shows that U_x can be well approximated by U_* after a period of time of order $\log(K)$.

Lemma 3.11. *Let $W = (W_e(t)\,;\, e \in E, t \geq 0)$ be a $\mathbf{R}^E$-valued Brownian motion, let $U_*(0) \sim \mathscr{N}(0, \Sigma_*)$ be independent of W, and let $(U_x\,;\, x \in \mathscr{U}_*)$ and U_* be defined as in Proposition 3.3. Then, for all K large enough and all $x \in \mathscr{D}$, we have :*

$$\mathbf{P}\left(\sup_{t \geq (6/\rho_*)\log(K)} \big\|U_x(t) - U_*(t)\big\| \geq 1/\sqrt{K} \right) \leq \exp\left(-K\right). \tag{51}$$

Proof. Let $x \in \mathscr{D}$, and let $t_1 \geq 0$. Set $x_1 = \varphi_x(t_1)$. In what follows, when we say that an assertion holds 'for K large enough', we mean that there exist $K_0 > 0$, independent of x, such that the assertion is true if $K \geq K_0$. Setting

$$\Delta U(t) = U_*(t) - U(t), \tag{52}$$

$$A(t) = \int_0^t \left(F'(x_*) - F'\big(\varphi_x(s)\big)\right) U_*(s)\mathrm{d}s, \tag{53}$$

$$D(t) = \sum_{e \in E} \left(\int_0^t \left[\sqrt{\beta_e(x_*)} - \sqrt{\beta_e\big(\varphi_x(s)\big)} \right] \mathrm{d}W_e(s) \right) e, \tag{54}$$

we have, almost surely for all $t \geq 0$,

$$\Delta U(t_1 + t) = \Delta U(t_1) + \int_0^t F'(\varphi_{x_1}(s))\Delta U(t_1 + s)\mathrm{d}s + A(t_1 + t) - A(t_1) + D(t_1 + t) - D(t_1).$$

The application of Lemma 3.2 with $\Delta U(t_1 + \cdot)$ playing the role of y yields $\Gamma \geq 1$ such that

$$\sup_{t \geq t_1} \|\Delta U(t)\| \leq \Gamma \left(\big\|\Delta U(t_1)\big\| \vee 3 \sup_{j \in \mathbf{N}} \sup_{j \leq t \leq j+1} \big\|(A + D)(t_1 + t) - (A + D)(t_1 + j)\big\| \right).$$

Let $K \geq 1$, and choose $t_1 = t_1(K) = (6/\rho_*) \log(K)$. We get

$$\begin{aligned} \mathbf{P}\left(\sup_{t \geq t_1(K)} \|\Delta U(t)\| \geq 1/\sqrt{K} \right) &\leq \mathbf{P}\left(\big\|\Delta U(t_1)\big\| \geq 1/(2\Gamma\sqrt{K}) \right) \\ &\quad + \sum_{j \in \mathbf{N}} \mathbf{P}\left(\sup_{j \leq t \leq j+1} \|A(t_1 + t) - A(t_1 + j)\| \geq 1/(12\Gamma\sqrt{K}) \right) \\ &\quad + \sum_{j \in \mathbf{N}} \mathbf{P}\left(\sup_{j \leq t \leq j+1} \|D(t_1 + t) - D(t_1 + j)\| \geq 1/(12\Gamma\sqrt{K}) \right) \end{aligned} \tag{55}$$

We bound each term of the right handside of this inequality. We start by the second term. Let Γ_1, Γ_2 be given by Lemma 3.1, $M_0 = \sum_{e \in E}\|e\|$, $M_1 = \max_{e \in E}\|\beta_e\|_{\infty,\mathscr{D}}$, $M_2 = \max_{e \in E}\|\beta_e\|_{\mathrm{Lip},\mathscr{D}}$, and $M_3 = \|F'\|_{\mathrm{Lip},\mathscr{D}}$. Let $j \in \mathbf{N}$. We have

$$\begin{aligned} \sup_{j \leq t \leq j+1} \|A(t_1 + t) - A(t_1 + j)\| &\leq \int_{t_1+j}^{t_1+j+1} \big\|F'(x_*) - F'\big(\varphi_x(t)\big)\big\|\big\|U_*(t)\big\|\mathrm{d}t \\ &\leq M_3\Gamma_1 e^{-\frac{\rho_*}{2}(t_1+j)} \sup_{t_1+j \leq t \leq t_1+j+1} \big\|U_*(t)\big\|, \end{aligned}$$

hence, recalling that $t_1 = (6/\rho_*) \log(K)$,

$$\mathbf{P}\left(\sup_{j \leq t \leq j+1} \|A(t_1 + t) - A(t_1 + j)\| \geq 1/(12\Gamma\sqrt{K}) \right) \leq \mathbf{P}\left(\sup_{0 \leq t \leq 1} \|U_*(t)\| \geq \frac{e^{\frac{\rho_*}{2}j} K^{5/2}}{12\Gamma\Gamma_1 M_3} \right).$$

Now, Lemma 3.2 entails that for all $\delta > 0$,

$$\begin{aligned} \mathbf{P}\left(\sup_{0 \leq t \leq 1} \big\|U_*(t)\big\| \geq \delta \right) &\leq \mathbf{P}\left[\Gamma \left(\big\|U_*(0)\big\| \vee 2 \sup_{0 \leq t \leq 1} \left\| \sum_{e \in E} \int_0^t \sqrt{\beta_e(x_*)}\mathrm{d}W_e(s)e \right\| \right) \geq \delta \right] \\ &\leq \mathbf{P}\left(\big\|U_*(0)\big\| \geq \frac{\delta}{\Gamma} \right) + |E|\mathbf{P}\left(\sup_{0 \leq t \leq 1} |B(t)| \geq \frac{\delta}{2\Gamma M_0 \sqrt{M_1}} \right), \end{aligned}$$

where B denotes a real Brownian motion. If the rank r of Σ_* is not zero, then thre exists $(G_1, \ldots, G_r) \sim \mathscr{N}(0, I_r)$ and $(\sigma_1, \ldots, \sigma_r) \in \mathbf{R}_+^r$ such that $\|U_*(0)\|^2 = \sum_{1 \leq i \leq r} \sigma_i^2 G_i^2 \leq \mathrm{Tr}(\Sigma_*) \max_{1 \leq i \leq r} G_i^2$, and thus for all $\delta > 0$,

$$\mathbf{P}\left(\|U_*(0)\| \geq \delta\right) \leq r\mathbf{P}\left(G_1^2 \geq \delta^2/\mathrm{Tr}(\Sigma_*)\right) \leq 2r \exp\left(-\frac{\delta^2}{2\mathrm{Tr}(\Sigma_*)} \right). \tag{56}$$

If $r = 0$, then $U_* \equiv 0$ and this bound also holds, with the convention $\exp(-\infty) = 0$. Using Lemma 3.5 to bound the Brownian term we obtain that, for all $\delta > 0$,

$$\mathbf{P}\left(\sup_{0\le t\le 1}\big\|U_*(t)\big\| \ge \delta\right) \le (2r+2|E|)\exp\left(-\frac{\delta^2}{2\Gamma^2\left(\mathrm{Tr}(\Sigma_*)\vee 4M_0^2M_1\right)}\right), \tag{57}$$

which entails

$$\mathbf{P}\left(\sup_{j\le t\le j+1}\|A(t_1+t)-A(t_1+j)\| \ge 1/(12\Gamma\sqrt{K})\right) \le (2r+2|E|)\exp\left(-\frac{e^{\rho_* j}K^5}{C_1}\right) \tag{58}$$

where $C_1 = 288\Gamma^4\Gamma_1^2M_3^2\left(\mathrm{Tr}(\Sigma_*)\vee 4M_0^2M_1\right)$. Using that $e^{\rho_* j} \ge (1+\rho_* j)$, we get, for K large enough,

$$\begin{aligned}\sum_{j\in\mathbf{N}}\mathbf{P}\left(\sup_{j\le t\le j+1}\|A(t_1+t)-A(t_1+j)\| \ge 1/(12\Gamma\sqrt{K})\right) &\le (2r+2|E|)\exp\left(-\frac{K^5}{C_1}\right)\sum_{j\in\mathbf{N}}\exp\left(-\frac{K^5\rho_* j}{C_1}\right)\\ &\le \exp(-K)/3. \end{aligned}\tag{59}$$

Now, let us bound the second term of the right handside of (55). For all $j \in \mathbf{N}$, we have

$$\begin{aligned}&\mathbf{P}\left(\sup_{j\le t\le j+1}\|D(t_1+t)-D(t_1+j)\| \ge 1/(12\Gamma\sqrt{K})\right)\\ &\le \sum_{e\in E}\mathbf{P}\left(\sup_{j\le t\le j+1}\left|\int_{t_1+j}^{t_1+t}\left(\sqrt{\beta_e(x_*)}-\sqrt{\beta_e(\varphi_x(t))}\right)\mathrm{d}W_e(t)\right| \ge 1/(12\Gamma M_0\sqrt{K})\right).\end{aligned}\tag{60}$$

Using that $(\sqrt{u}-\sqrt{v})^2 \le |u-v|$, we obtain that for all $e \in E$,

$$\int_{t_1+j}^{t_1+j+1}\left(\sqrt{\beta_e(x_*)}-\sqrt{\beta_e(\varphi_x(t))}\right)^2\mathrm{d}t \le M_2\Gamma_1K^{-3}e^{-\frac{\rho_*}{2}j}.$$

Thus, Lemma 3.6 yields

$$\mathbf{P}\left(\sup_{j\le t\le j+1}\|D(t_1+t)-D(t_1+j)\| \ge 1/(12\Gamma\sqrt{K})\right) \le 2|E|\exp\left(-\frac{e^{\frac{\rho_*}{2}j}K^2}{C_2}\right), \tag{61}$$

where $C_2 = 288\Gamma^2\Gamma_1M_0^2M_2$, and therefore, for K large enough

$$\sum_{j\in\mathbf{N}}\mathbf{P}\left(\sup_{j\le t\le j+1}\|D(t_1+t)-D(t_1+j)\| \ge 1/(12\Gamma\sqrt{K})\right) \le \exp(-K)/3.$$

Finally, let us bound the first term of the right handside of (55). We have

$$\Delta U(t) = \Delta U(0) + \int_0^t F'\big(\varphi_x(s)\big)\Delta U(s)\mathrm{d}s + A(t) + D(t),$$

thus, applying Itô's lemma to $\Psi_x(0,t)\Delta U(t)$ and left multiplying by $\Psi_x(t,0)$ after that yields

$$\begin{aligned}\Delta U(t) = \Psi_x(t,0)\Delta U(0) &+ \int_0^t \Psi_x(t,s)\left(F'(x_*)-F'\big(\varphi_x(s)\big)\right)U_*(s)\mathrm{d}s\\ &+\sum_{e\in E}\int_0^t\left[\sqrt{\beta_e(x_*)}-\sqrt{\beta_e\big(\varphi_x(s)\big)}\right]\Psi_x(t,s)e\ \mathrm{d}W_e(s).\end{aligned}$$

It follows from Lemma 3.1 that

$$\big\|\Psi_x(t_1,0)\Delta U(0)\big\| \le \Gamma_2K^{-3}\big\|U_*(0)\big\|$$

and

$$\left\|\int_0^{t_1}\Psi_x(t,s)\left(F'(x_*)-F'\big(\varphi_x(s)\big)\right)U_*(s)\mathrm{d}s\right\| \leq \int_0^{t_1}\Gamma_2 e^{-\frac{\rho_*}{2}(t-s)}M_3\Gamma_1 R e^{-\frac{\rho_*}{2}s}\big\|U_*(s)\big\|\mathrm{d}s$$
$$\leq 6\rho_*^{-1}\Gamma_1\Gamma_2 M_3\log(K)K^{-3}\sup_{0\leq t\leq t_1}\big\|U_*(t)\big\|.$$

Using inequalities (56) and (57), we obtain, for K large enough

$$\mathbf{P}\left(\big\|\Psi_x(t_1,0)\Delta U(0)\big\|\geq 1/(6\Gamma\sqrt{K})\right)\leq \mathbf{P}\left(\big\|U_*(0)\big\|\geq \frac{K^{5/2}}{6\Gamma\Gamma_2}\right)\leq \exp(-K)/9 \tag{62}$$

and

$$\begin{aligned}
&\mathbf{P}\left(\left\|\int_0^{t_1}\Psi_x(t,s)\left(F'(x_*)-F'\big(\varphi_x(s)\big)\right)U_*(s)\mathrm{d}s\right\|\geq \frac{1}{6\Gamma\sqrt{K}}\right)\\
&\leq \mathbf{P}\left(\sup_{0\leq t\leq t_1}\big\|U_*(t)\big\|\geq \frac{\rho_* K^{5/2}}{36\Gamma\Gamma_1\Gamma_2 M_3\log(K)}\right)\\
&\leq \left\lceil\frac{6\log(K)}{\rho_*}\right\rceil \mathbf{P}\left(\sup_{0\leq t\leq 1}\big\|U_*(t)\big\|\geq \frac{\rho_* K^{5/2}}{36\Gamma\Gamma_1\Gamma_2 M_3\log(K)}\right)\\
&\leq \exp(-K)/9.
\end{aligned}\tag{63}$$

Moreover,

$$\begin{aligned}
&\mathbf{P}\left(\left\|\sum_{e\in E}\int_0^{t_1}\left[\sqrt{\beta_e(x_*)}-\sqrt{\beta_e\big(\varphi_x(s)\big)}\right]\Psi_x(t_1,s)e\;\mathrm{d}W_e(s)\right\|\geq\frac{1}{6\Gamma\sqrt{K}}\right)\\
&\leq \mathbf{P}\left(\max_{1\leq i\leq d, e\in E}\left|\int_0^{t_1}\left[\sqrt{\beta_e(x_*)}-\sqrt{\beta_e\big(\varphi_x(s)\big)}\right](\Psi_x(t_1,s)e)_i\mathrm{d}W_e(s)\right|\geq \frac{1}{6\Gamma\sqrt{d}|E|\sqrt{K}}\right),
\end{aligned}$$

and for all $1\leq i\leq d$ and $e\in E$,

$$\int_0^{t_1}\left[\sqrt{\beta_e(x_*)}-\sqrt{\beta_e\big(\varphi_x(s)\big)}\right]^2(\Psi_x(t_1,s)e)_i^2\mathrm{d}s \leq \int_0^{t_1}M_2\Gamma_1 e^{-\frac{\rho_*}{2}s}(\Gamma_2 e^{-\frac{\rho_*}{2}(t_1-s)}\|e\|)^2\mathrm{d}s$$
$$\leq \Gamma_1\Gamma_2^2 M_0^2 M_2 t_1 e^{-\frac{\rho_*}{2}t_1}.$$

Thus, Lemma 3.5 entails that we have, for K large enough,

$$\mathbf{P}\left(\left\|\sum_{e\in E}\int_0^{t_1}\left[\sqrt{\beta_e(x_*)}-\sqrt{\beta_e\big(\varphi_x(s)\big)}\right]\Psi_x(t_1,s)e\;\mathrm{d}W_e(s)\right\|\geq\frac{1}{6\Gamma\sqrt{K}}\right)\leq\exp(-K)/9, \tag{64}$$

The bounds (62), (63) and (64) yield

$$\mathbf{P}\left(\big\|\Delta U(t_1)\big\|\geq 1/(2\Gamma\sqrt{K})\right)\leq\exp(-K)/3.$$

Plugging this into (55) together with (58) and (61), we obtain that, for K large enough, for all $x\in\mathscr{D}$,

$$\mathbf{P}\left(\sup_{t\geq(6/\rho_*)\log(K)}\big\|U_*(t)-U_x(t)\big\|\geq 1/\sqrt{K}\right)\leq\exp(-K).$$

□

Now, let us prove Corollary 1.2. Let C,V,α be given by Theorem 1.1 and let $\varepsilon:\mathbf{R}_+^*\to\mathbf{R}_+^*$ be such that $\alpha\log(K)/K\leq\varepsilon(K)\ll 1$. Let $K>0$ be large enough, let $x\in\mathscr{D}$, and let $\big(X_x^K,U_x\big)$ be the coupling given by Theorem 1.1. Using Lemma 3.11, we may suppose that there exists, on the same probability space as X_x^K and U_x, a process $U_* \sim \nu_*$ satisfying (51) if K is large enough (independently of x). Set $t(K) = (6/\rho_*)\log(K)$. Letting $\Gamma_1 \geq 1$ be given by Lemma 3.1, we have

$$\sup_{s \geq t(K)} \|\varphi_x(s) - x_*\| \leq \Gamma_1 e^{-\frac{\rho_*}{2} t(K)} \leq \Gamma_1 K^{-3}. \tag{65}$$

Hence, for K large enough (indepentl) we have, for all $T \geq t(K)$,

$$\begin{aligned}
&\mathbf{P}\left(\sup_{t(K) \leq t \leq T} \left\|X_x^K(t) - x_* - U_*(t)/\sqrt{K}\right\| > \varepsilon(K)\right) \\
&\leq \mathbf{P}\left(\sup_{t(K) \leq t \leq T} \left\|X_x^K(t) - \varphi_x(t) - U_x(t)/\sqrt{K}\right\| > \varepsilon(K) - 2/K\right) \\
&\quad + \mathbf{P}\left(\sup_{t(K) \leq t \leq T} \|\varphi_x(t) - x_*\| > 1/K\right) + \mathbf{P}\left(\sup_{t(K) \leq t \leq T} \left\|U_x^K(t) - U_*(t)\right\| > 1/\sqrt{K}\right) \\
&\leq C(T+1)\exp\left(-VK\varepsilon(K)\right) + \exp(-K).
\end{aligned}$$

using (65) and Lemma 3.11 for the last inequality. Using that $\varepsilon(K) \ll 1$, the corollary is proved with $C' = C + 1$.

3.4 Proof of Corollary 1.3

For all $K > 0$, $x \in \mathbf{R}^d$ and $t \geq 0$, let $\mu_x^{K;t}$ denote the probability distribution of $X_x^K(t)$, where $X_x^K \sim \mu_x^K$. Let $C', V, \alpha > 0$ be given by Proposition 1.2, and let $t(K) = (6/\rho_*)\log(K)$. It follows from Corollary 1.2 that there exists $K_0 \geq 1$ such that for all $K \geq K_0$ and for all $x \in \mathscr{D}$ there exists a coupling $\left(X_x^K, U_*\right)$ of (μ_x^K, ν_*) such that,

$$\mathbf{P}\Big(\left\|X_x^K\big(t(K)\big) - x_* - U_*\big(t(K)\big)\right\| > \alpha\log(K)/K\Big) \leq C'(t(K)+1)K^{-V\alpha}. \tag{66}$$

For all $K \geq K_0$, we denote by π_x^K the probability distribution of the coupling $\left(X_x^K\big(t(K)\big), U_*\big(t(K)\big)\right)$ for $x \in \mathscr{D}$, while we set $\pi_x^K = \mu_x^{K;t(K)} \otimes \mathscr{N}(0, \Sigma_*)$ for $x \notin \mathscr{D}$.

Now, for $t \geq 2t(K)$ we get an upper bound on the probability that $X_x^K(t - t(K)) \notin \mathscr{D}$ and combine it with (66) using the Markov property. Let $r > 0$ be such that $\mathscr{D} \supset \bar{B}(x_*, r)$. Proposition 3.8 yields constants $\eta_0, V_0 > 0$ such that, setting $\tilde{\eta}_0 = \eta_0 \wedge (r/2)$, we have, for all K large enough, for all $x \in \mathscr{D}$ and for all $t \geq 0$,

$$\mathbf{P}\left(\sup_{0 \leq s \leq t} \left\|X_x^K(s) - \varphi_x(s)\right\| > \tilde{\eta}_0\right) \leq 2|E|(t+1)\exp\left(-V_0\tilde{\eta}_0^2 K\right).$$

In addition, it follows from Lemma 3.1 that

$$\sup\left\{\|\varphi_x(t) - x_*\|\,;\, x \in \mathscr{D}, t \geq t(K)\right\} = \mathcal{O}(K^{-3}).$$

Consequently, for K larger than some $K_1 \geq K_0$, for all $x \in \mathscr{D}$ and $t \geq 2t(K)$, we have

$$\mathbf{P}\left(\left\|X_x^K(t - t(K)) - x_*\right\| > r\right) \leq \mathbf{P}\left(\left\|X_x^K(t - t(K)) - \varphi_x(t - t(K))\right\| > \tilde{\eta}_0\right),$$

hence

$$\mathbf{P}\left(X_x^K(t - t(K)) \notin \mathscr{D}\right) \leq 2|E|(t - t(K) + 1)\exp\left(-V_0\tilde{\eta}_0^2 K\right). \tag{67}$$

Set $V' = V_0\tilde{\eta}_0^2/2$. Let $K \geq K_1$, $x \in \mathscr{D}$, $t \in \left[2t(K),\, e^{V'K}\right]$ and define $\tilde{\pi} \in \mathcal{P}\left(\mathbf{R}^d \times \mathbf{R}^d\right)$ by

$$\tilde{\pi} = \sum_{y \in K^{-1}\mathbf{Z}^d} \mathbf{P}\left(X_x^K(t - t(K)) = y\right)\pi_y^K.$$

The first marginal of $\tilde{\pi}$ is $\mu_x^{K;t}$ as a consequence of the Markov property of X_x^K at time $t - t(K)$, while the second marginal of $\tilde{\pi}$ is $\mathscr{N}(0, \Sigma_*)$. Letting $X, G : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d$ denote the first and second canonical projections, and $\mathbf{E}_{\tilde{\pi}}$ the expectation with respect to $\tilde{\pi}$, we have

$$\begin{aligned}\mathbf{E}_{\tilde{\pi}}\left[c\left(\sqrt{K}(X - x_*), G\right)\right] &\leq \alpha \log(K)/\sqrt{K} + \tilde{\pi}\left(\left|\sqrt{K}(X - x_*) - G\right| > \alpha \log(K)/\sqrt{K}\right) \\ &\leq \alpha \log(K)/\sqrt{K} + \mathbf{P}\left(X_x^K(t - t(K)) \notin \mathscr{D}\right) \\ &\quad + \sup_{y \in \mathscr{D}} \pi_y^K\left(\left|\sqrt{K}(X - x_*) - G\right| > \alpha \log(K)/\sqrt{K}\right) \\ &\leq \alpha \log(K)/\sqrt{K} + 2|E|(e^{V'K} + 1)e^{-2V'K} + C'(t(K) + 1)K^{-V\alpha},\end{aligned}$$

using (66) and (67) for the last inequality. We conclude the proof using the definition of $\mathcal{W}_c$: for all $x \in \mathscr{D}$,

$$\sup_{2t(K) \leq t \leq \exp(V'K)} \mathcal{W}_c\left[\mathbf{P}\left(\sqrt{K}\left(X_x^K(t) - x_*\right) \in \cdot\right), \mathscr{N}\left(0, \Sigma_*\right)\right] \underset{K \to +\infty}{\longrightarrow} 0.$$

3.5 Proof of Proposition 2.1

Let $0 < h < 1$. Set

$$t_1(K) = \exp\left(\left(\frac{1}{2} - h\right)(1 - h)^2 K\eta^2(K)\right) \quad \text{and} \quad t_2(K) = \exp\left(\left(\frac{1}{2} + h\right)(1 + h)^2 K\eta^2(K)\right).$$

We have $1 \ll K\eta^2(K) \ll K\eta(K)$, hence the coupling $(X_{x_*}^K, U_{x_*})$ given by Theorem 1.1 satisfies

$$\mathbf{P}\left(\sup_{0 \leq s \leq t_2(K)} \left\|X_{x_*}^K(s) - x_* - U_{x_*}(s)/\sqrt{K}\right\|_{\Sigma_*^{-1}} \geq h\,\eta(K)\right) \underset{K \to +\infty}{\longrightarrow} 0, \tag{68}$$

using the equivalence of norms on $\mathbf{R}^d$. Now, (7) entails

$$\mathbf{P}\left(\sup_{0 \leq s < t_2(K)} \left\|U_{x_*}(s)/\sqrt{K}\right\|_{\Sigma_*^{-1}} \geq (1 + h)\eta(K)\right) \underset{K \to +\infty}{\longrightarrow} 1, \tag{69}$$

and

$$\mathbf{P}\left(\sup_{0 \leq s \leq t_1(K)} \left\|U_{x_*}(s)/\sqrt{K}\right\|_{\Sigma_*^{-1}} \geq (1 - h)\eta(K)\right) \underset{K \to +\infty}{\longrightarrow} 0. \tag{70}$$

Combining (68), (69), (70) with the triangular inequality yields

$$\mathbf{P}\left[\exp\left(\left(\frac{1}{2} - h\right)(1 - h)^2 K\eta^2(K)\right) \leq \tau_\eta^K < \exp\left(\left(\frac{1}{2} + h\right)(1 + h)^2 K\eta^2(K)\right)\right] \underset{K \to +\infty}{\longrightarrow} 1.$$

This holds for all $h > 0$, thus the proposition is proved.

3.6 Proof of Proposition 2.2

We start by showing that when we condition a process $X_x^K \sim \mu_x^K$ to survive for a large time, then for t much larger than $\log(K)$, $X_x^K(t)$ belongs to the compact $[x_*/4, 3x_*]$ with high probability, uniformly in x.

First, we compare the logistic birth-and-death process with a supercritical branching process at the neighbourhood of 0. Let $\mathscr{M}$ be a Poisson point measure on $\mathbf{R}_+^2$, of intensity the Lebesgue measure. Let $K > 0$. For all $n \in \mathbf{N}$, we can construct a logistic birth-and-death process N_n^K starting from n

(with the transition rates defined in (9)), as the unique real-valued process, up to indinstiguishability, satisfying

$$N_n^K(t) = n + \int_{]0,t]\times \mathbf{R}_+} \left(\mathbf{1}_{\{u\leq pN_n^K(s_-)\}} - \mathbf{1}_{\{pN_n^K(s_-)<u\leq N_n^K(s_-)(p+q+N_n^K(s_-)/K)\}}\right) \mathscr{M}(\mathrm{d}s, \mathrm{d}u)$$

almost surely for all $t \geq 0$. The unique process L solution of

$$L(t) = 1 + \int_{]0,t]\times \mathbf{R}_+} \left(\mathbf{1}_{\{u\leq pL(s_-)\}} - \mathbf{1}_{\{pL(s_-)<u\leq (p+q+x_*/2)L(s_-)\}}\right) \mathscr{M}(\mathrm{d}s, \mathrm{d}u)$$

is a branching process starting from $L(0) = 1$, with transition rate from m to $m+1$ given by pm and transition rate from m to $m-1$ given by $(q + x_*/2)m$. It is supercritical because $p - q - x_*/2 = x_*/2 > 0$. This coupling between N_1^K and L has the useful property that $\tau_N \leq \tau_L$, where $\tau_N = \inf\left\{t \geq 0 : N_1^K(t) \geq Kx_*/2\right\}$ and $\tau_L = \inf\left\{t \geq 0 : L(t) \geq Kx_*/2\right\}$. Indeed, setting $\sigma = \inf\left\{t \geq 0 : L(t) > N_1^K(t)\right\}$, we see that on the event $\{\sigma < \infty\}$, it is necessary that $L(\sigma_-) = N_1^K(\sigma_-)$ and that a death happens at time σ for N_1^K but not for L. Hence $N_1^K(\sigma_-) > x_*/2$, which entails $\tau_N < \sigma$, and $\tau_L \geq \tau_N$.

Now, it is a classical result that $\left(e^{-tx_*/2}L(t)\,;\, t \geq 0\right)$ is a martingale which converges almost-surely, when $t \to +\infty$, to a nonnegative random variable W such that $\mathbf{E}(W) = 1$, see e.g.[2, Chapter III]. Hence, there exists $\eta < 1$ such that, for t large enough,

$$\mathbf{P}\left(e^{-tx_*/2}L(t) \leq 1/2\right) \leq \eta.$$

From this we deduce that, for K larger than some $K_0 > 0$,

$$\begin{aligned}\mathbf{P}\left(\tau_N > 3\log(K)/x_*\right) &\leq \mathbf{P}\left(\tau_L > 3\log(K)/x_*\right) \\ &\leq \mathbf{P}\left(K^{-3/2}L\left(3\log(K)/x_*\right) \leq 1/2\right) \\ &\leq \eta.\end{aligned}$$

Moreover, we can see that for all $n \geq 1$, we have $N_1^K(t) \leq N_n^K(t)$ almost surely for all $t \geq 0$. Hence, the same inequality holds when we replace N_1^K by N_n^K in the definition of τ_N.

Let us introduce the canonical real càdlàg process $X = (X(t)\,;\, t \geq 0)$, defined by $X(t)(\omega) = \omega(t)$ for all $\omega \in \mathcal{D}(\mathbf{R}_+, \mathbf{R})$, and set, for all $h \in \mathbf{R}_+$, $\tau_h^+ = \inf\{t \geq 0 : X(t) \geq h\}$ and $\tau_h^- = \inf\{t \geq 0 : X(t) \leq h\}$. The result we just obtained can be restated as follows : for $K \geq K_0$, for all $x \in K^{-1}\mathbf{N}^*$,

$$\mu_x^K\left(\tau_{x_*/2}^+ > 3\log(K)/x_*\right) \leq \eta.$$

When we condition a birth-and-death process to survive, we favour trajectories that go away from zero. Setting $t_1(K) = 3\log(K)/x_*$, we have, for $K \geq K_0$, for all $t \geq t_1(K)$ and all $x \in K^{-1}\mathbf{N}^*$,

$$\begin{aligned}\mu_x^K\left(\tau_{x_*/2}^+ \leq t_1(K), X(t) > 0\right) &\geq \mu_x^K\left(\tau_{x_*/2}^+ \leq t_1(K), X(\tau_{x_*/2}^+ + t) > 0\right) \\ &= \mu_x^K\left(\tau_{x_*/2}^+ \leq t_1(K)\right)\mu_{\lceil Kx_*/2\rceil/K}^K\left(X(t) > 0\right) \\ &\geq \mu_x^K\left(\tau_{x_*/2}^+ \leq t_1(K)\right)\mu_x^K\left(X(t) > 0\right),\end{aligned}$$

where the equality comes from the strong Markov property at time $\tau_{x_*/2}^+$. Thus,

$$\mu_x^K\left(\tau_{x_*/2}^+ > t_1(K) \,\middle|\, X(t) > 0\right) \leq \eta.$$

Moreover, for all $x \in K^{-1}\mathbf{N}^*$, for all $t, T \geq 0$, for all $A \in \mathscr{B}(\mathcal{D}([0,t], \mathbf{R}))$ and for all $B \in \mathscr{B}(\mathcal{D}([0,T], \mathbf{R}))$, the Markov property entails that

$$\begin{aligned}&\mu_x^K\left[\,(X(s))_{0\leq s\leq t} \in A, (X(t+s))_{0\leq s\leq t} \in B \,\middle|\, X(t+T) > 0\right] \\ &= \sum_{y\in K^{-1}\mathbf{N}^*} \mu_x^K\left[\,(X(s))_{0\leq s\leq t} \in A, X(t) = y \,\middle|\, X(t+T) > 0\right]\mu_y^K\left[X(s)_{0\leq s\leq T} \in B \,\middle|\, X(T) > 0\right]. \quad (71)\end{aligned}$$

Let $t_i = t_i(K) = it_1(K)$ for all $i \in \mathbf{N}$. For $K \geq K_0$, for all $i \in \mathbf{N}^*$, for all $t \geq t_i$ and for all $x \in K^{-1}\mathbf{N}^*$, we get

$$\begin{aligned}
&\mu_x^K\Big[\tau^+_{x_*/2} > t_i \,\Big|\, X(t) > 0\Big] \\
&= \sum_{y \in K^{-1}\mathbf{N}^*} \mu_x^K\Big[\tau^+_{x_*/2} > t_{i-1}, X(t_{i-1}) = y \,\Big|\, X(t) > 0\Big] \mu_y^K\Big[\tau^+_{x_*/2} > t_1, \,\Big|\, X(t - t_{i-1}) > 0\Big] \\
&\leq \eta \sum_{y \in K^{-1}\mathbf{N}^*} \mu_x^K\Big[\tau^+_{x_*/2} > t_{i-1}, X(t_{i-1}) = y \,\Big|\, X(T) > 0\Big] \\
&= \eta\, \mu_x^K\Big[\tau^+_{x_*/2} > t_{i-1} \,\Big|\, X(t) > 0\Big],
\end{aligned}$$

hence, by induction,

$$\mu_x^K\Big[\tau^+_{x_*/2} > t_i(K) \,\Big|\, X(t) > 0\Big] \leq \eta^i. \tag{72}$$

Let $t, T \geq 0$, let $K \geq K_0$ and let $x \in K^{-1}\mathbf{N}^*$. We have

$$\begin{aligned}
&\mu_x^K\left(X(t+T) > 0 \,\middle|\, 0 < X(t) < x_*/4\right) \\
&= \sum_{y \in (0, x_*/4) \cap K^{-1}\mathbf{N}^*} \mu_x^K\left(X(t) = y \,\middle|\, 0 < X(t) < x_*/4\right) \mu_y^K\left(X(T) > 0\right) \\
&\leq \mu^K_{\lfloor Kx_*/4 \rfloor / K}\left(X(T) > 0\right) \\
&\leq \mu_x^K\left(X(t+T) > 0 \,\middle|\, X(t) \geq x_*/4\right),
\end{aligned}$$

hence

$$\mu_x^K\left(X(t+T) > 0 \,\middle|\, 0 < X(t) < x_*/4\right) \leq \mu_x^K\left(X(t+T) > 0 \,\middle|\, X(t) > 0\right).$$

Given that $\{X(t+T) > 0\} \subset \{X(t) > 0\}$, this inequality is equivalent to

$$\mu_x^K\left(X(t) < x_*/4 \,\middle|\, X(t+T) > 0\right) \leq \mu_x^K\left(X(t) < x_*/4 \,\middle|\, X(t) > 0\right). \tag{73}$$

Now, the right handside satisfies

$$\begin{aligned}
\mu_x^K\left(X(t) < x_*/4 \,\middle|\, X(t) > 0\right) &\leq \mu_x^K\left(\tau^+_{x_*/2} > t \,\middle|\, X(t) > 0\right) \\
&\quad + \mu_x^K\left(\tau^+_{x_*/2} \leq t, X(t) < x_*/4 \,\middle|\, X(t) > 0\right) \\
&\leq \eta^{\lfloor t/t_1(K) \rfloor} + \frac{\mu_x^K\left(\tau^+_{x_*/2} \leq t, X(t) < x_*/4\right)}{\mu_x^K\left(\tau^+_{x_*/2} \leq t, X(\tau^+_{x_*/2} + t) > 0\right)} \\
&\leq \eta^{\lfloor t/t_1(K) \rfloor} + \frac{\mu^K_{\lceil Kx_*/2 \rceil / K}\left(\tau^-_{x_*/4} \leq t\right)}{\mu^K_{\lceil Kx_*/2 \rceil / K}\left(X(t) > 0\right)},
\end{aligned}$$

using (72) for the first inequality and the strong Markov property at time $\tau^+_{x_*/2}$ for the last one. It follows from Proposition 3.8 that there exists $V_0' > 0$ such that for all K large enough, for all $t \geq 0$,

$$\sup_{x_*/4 \leq x \leq 3x_*} \mu_x^K\left(\sup_{0 \leq s \leq t} \left|X(s) - \varphi_x(s)\right| > x_*/4\right) \leq 4(t+1)\exp\left(-V_0' K\right), \tag{74}$$

Using that $\varphi_{\lceil Kx_*/2 \rceil / K}(s) \geq x_*/2$ for all $s \geq 0$, this yields, for all K large enough and for all $t \geq 0$,

$$\mu^K_{\lceil Kx_*/2 \rceil / K}\left(\tau^-_{x_*/4} \leq t\right) \leq 4(t+1)\exp\left(-V_0' K\right),$$

hence

$$\sup_{x\in K^{-1}\mathbf{N}^*} \mu_x^K \left(X(t) < x_*/4 \,\middle|\, X(t) > 0 \right) \leq \eta^{\lfloor t/t_1(K) \rfloor} + \frac{4(t+1)\exp\left(-V_0'K\right)}{1-4(t+1)\exp\left(-V_0'K\right)}.$$

Set $\beta = 3/(x_*|\log(\eta)|)$. The above inequality yields, for all K large enough and for all $t \leq \beta V_0' K \log(K)/2$,

$$\sup_{x\in K^{-1}\mathbf{N}^*} \mu_x^K \left(X(t) < x_*/4 \,\middle|\, X(t) > 0 \right) \leq \eta^{-1} \exp\left(-\frac{t}{\beta \log(K)}\right) + \exp\left(-\frac{V_0'K}{2}\right).$$

Moreover, as a consequence of (71), the left handside of this inequality is a non-increasing function of t. Thus, for $t \geq \beta V_0' K \log(K)/2$, the left handside is less than the right handside evaluated at time $\beta V_0' K \log(K)/2$. Using (73), we obtain that for all K large enough, and for all $t, T \geq 0$,

$$\sup_{x\in K^{-1}\mathbf{N}^*} \mu_x^K \left(X(t) < x_*/4 \,\middle|\, X(t+T) > 0 \right) \leq \eta^{-1} \exp\left(-\frac{t}{\beta \log(K)}\right) + (\eta^{-1}+1)\exp\left(-\frac{V_0'K}{2}\right). \quad (75)$$

Now, we bound $\mu_x^K \left(X(t) > 3x_* \,\middle|\, X(t+T) > 0 \right)$. Let us set, for all $n \in \mathbf{N}^*$,

$$\pi_n^K = \prod_{j=1}^{n} \frac{q_{j,j-1}^K}{q_{j,j+1}^K} = \prod_{j=1}^{n} \frac{q+j/K}{p},$$

and denote by $\mathbf{E}_x^K$ the expectation under μ_x^K. For all $x \in (2x_*, +\infty) \cap K^{-1}\mathbf{N}^*$, we have the following explicit formulas for expectations of first passage times (see e.g. [28]) :

$$\mathbf{E}_x^K \left(\tau_{2x_*}^- \right) = \sum_{i=\lfloor 2Kx_* \rfloor +1}^{Kx} \sum_{n=i}^{\infty} \frac{\pi_{i-1}}{n(q+n/K)\pi_{n-1}}.$$

Hence,

$$\begin{aligned}
\sup_{x\in K^{-1}\mathbf{N}^*} \mathbf{E}_x^K \left(\tau_{2x_*}^- \right) &= \sum_{i=\lfloor 2Kx_* \rfloor +1}^{\infty} \sum_{n=i}^{\infty} \frac{\pi_{i-1}}{n(q+n/K)\pi_{n-1}} \\
&= \sum_{n=\lfloor 2Kx_* \rfloor +1}^{\infty} \frac{1}{n(q+n/K)} \sum_{i=\lfloor 2Kx_* \rfloor +1}^{n} \prod_{j=i}^{n-1} \frac{p}{q+j/K} \\
&\leq \sum_{n=\lfloor 2Kx_* \rfloor +1}^{\infty} \frac{1}{n(q+n/K)} \sum_{i=\lfloor 2Kx_* \rfloor +1}^{n} \left(\frac{p}{p+x_*} \right)^{n-i} \\
&\leq (1+p/x_*) \int_{\lfloor 2Kx_* \rfloor /K}^{\infty} \frac{1}{u(q+u)} \mathrm{d}u.
\end{aligned}$$

Let $\theta = 2(1+p/x_*) \int_{x_*}^{\infty} \frac{1}{u(q+u)} \mathrm{d}u$, it is finite and for K large enough, Markov inequality yields

$$\sup_{x\in K^{-1}\mathbf{N}^*} \mu_x^K \left(\tau_{2x_*}^- > \theta \right) \leq 1/2,$$

and then Markov property entails that for all $t \geq 0$,

$$\sup_{x\in K^{-1}\mathbf{N}^*} \mu_x^K \left(\tau_{2x_*}^- > t \right) \leq 2^{-\lfloor t/\theta \rfloor}.$$

Let K be large enough so that the above inequality holds, let $x \in K^{-1}\mathbf{N}^*$ and let $t \geq 0$. We have

$$\begin{aligned}
\mu_x^K \left(X(t) > 3x_* \right) &\leq \mu_x^K \left(\tau_{2x_*}^- > t \right) + \mu_x^K \left(\tau_{2x_*}^- \leq t, X(t) > 3x_* \right) \\
&\leq 2^{-\lfloor t/\theta \rfloor} + \mu_{\lfloor 2Kx_* \rfloor /K}^K \left(\tau_{3x_*}^+ \leq t \right),
\end{aligned}$$

and thus, using (74),

$$\mu_x^K\left(X(t) > 3x_*\right) \le 2^{1-t/\theta} + 4(t+1)\exp(-V_0'K)$$

Now, let $T \ge 0$. Since

$$\begin{aligned}\mu_x^K\left(X(t+T) > 0\right) &\ge \mu_x^K\left(\tau_{x_*/2}^+ < \infty, X(\tau_{x_*/2}^+ + t + T) > 0\right)\\ &= \mu_x^K\left(\tau_{x_*/2}^+ < \infty\right)\mu_{\lceil Kx_*/2\rceil/K}^K\left(X(t+T) > 0\right)\\ &\ge (1-\eta)\left(1 - 4(t+T+1)\exp\left(-V_0'K\right)\right),\end{aligned}$$

we obtain

$$\sup_{x\in K^{-1}\mathbf{N}^*}\mu_x^K\left(X(t) > 3x_* \,\middle|\, X(t+T) > 0\right) \le \frac{2^{1-t/\theta} + 4(t+1)\exp(-V_0'K)}{(1-\eta)\left(1 - 4(t+T+1)\exp\left(-V_0'K\right)\right)}.$$

Moreover, as consequence of (71), the left handside of the above inequality is a non-increasing function of t. Setting $t_2(K) = V_0'\theta K/(2\log(2))$, we get, for all K large enough, $t \ge 0$ and $T \le \exp\left(V_0'K/2\right)$,

$$\begin{aligned}\sup_{x\in K^{-1}\mathbf{N}^*}\mu_x^K\left(X(t) > 3x_* \,\middle|\, X(t+T) > 0\right) &\le \sup_{x\in K^{-1}\mathbf{N}^*}\mu_x^K\left(X\left(t\wedge t_2(K)\right) > 3x_* \,\middle|\, X\left(t\wedge t_2(K) + T\right) > 0\right)\\ &\le C_1\left(2^{-t/\theta} + \exp(-V_0'K/2)\right), \qquad (76)\end{aligned}$$

where $C_1 = 5/(1-\eta)$. Now we can combine (75) and (76), and we obtain that for all K large enough, $t \ge 0$ and $T \le \exp\left(V_0'K/2\right)$,

$$\sup_{x\in K^{-1}\mathbf{N}^*}\mu_x^K\left(X(t) \notin [x_*/4, 3x_*] \,\middle|\, X(t+T) > 0\right) \le C_2\left(\exp\left(-\frac{t}{\beta\log(K)}\right) + \exp\left(-\frac{V_0'}{2}K\right)\right), \quad (77)$$

where $C_2 = C_1 + \eta^{-1} + 1$.

Now that we control, uniformly in $X(0)$, the probability that $X(t)$ belongs to some fixed compact, conditional on later survival, we use Corollary 1.2 to build the desired couplings. Let $C', V, \alpha > 0$ be given by the application of Corollary 1.2 to $\mathscr{D} = [x_*/4, 3x_*]$. Note that $\rho_* = -F'(x_*) = x_*$ and $(6/\rho_*)\log(K) = 2t_1(K)$. Let $\varepsilon : \mathbf{R}_+^* \to \mathbf{R}_+^*$ such that $\alpha\log(K)/K \le \varepsilon(K) \ll 1$. For K larger than some $K_0 \ge 1$, for all $T \ge 0$ and all $y \in [x_*/4, 3x_*]$, Corollary 1.2 yields a coupling $\left(X_y^K, U_*\right)$ of $\left(\mu_y^K, \nu_*\right)$ such that

$$\begin{aligned}\mathbf{P}\left(\sup_{2t_1(K)\le s\le 2t_1(K)+T}\left\|X_y^K(s) - x_* - U_*(s)/\sqrt{K}\right\| > \varepsilon(K)\right) &\le C'(2t_1(K) + T + 1)\exp\left(-VK\varepsilon(K)\right)\\ &\le (T+1)\exp\left(-(V/2)K\varepsilon(K)\right).\end{aligned}$$

Enlarging the probability space, we may suppose that there exists a process Y independent of X_y^K and distributed as $\tilde{\mu}_y^{K;0,2t_1(K)+T}$. The process $(Z(s)\,;\, 0 \le s \le 2t_1(K) + T)$ defined by

$$Z = X_y^K\mathbf{1}_{\left\{X_y^K(2t_1(K)+T)>0\right\}} + Y_y^K\mathbf{1}_{\left\{X_y^K(2t_1(K)+T)=0\right\}}$$

is then also distributed as $\tilde{\mu}_y^{K;0,2t_1(K)+T}$. It satisfies, for K large enough and any $y \in [x_*/4, 3x_*]$,

$$\begin{aligned}&\mathbf{P}\left(\sup_{2t_1(K)\le s\le 2t_1(K)+T}\left\|Z(s) - x_* - U_*(s)/\sqrt{K}\right\| > \varepsilon(K)\right)\\ &\le (T+1)\exp\left(-(V/2)K\varepsilon(K)\right) + \mathbf{P}\left(X_y^K(2t_1(K)+T) = 0\right)\\ &\le (T+1)\exp\left(-(V/2)K\varepsilon(K)\right) + 4(2t_1(K) + T + 1)\exp\left(-V_0'K\right)\\ &\le 2(T+1)\exp\left(-(V/2)K\varepsilon(K)\right). \qquad (78)\end{aligned}$$

Let us denote by $\Gamma_y^{K;T} \in \mathcal{P}\left(\mathcal{D}\left([0,T],\mathbf{R}\right)\times\mathcal{C}\left(\mathbf{R}_+,\mathbf{R}\right)\right)$ the probability distribution of

$$\big(Z(2t_1(K)+s)_{0\le s\le T},\ U_*(2t_1(K)+s)_{s\ge 0}\big).$$

For all $K \ge K_0$, $t \ge 2t_1(K)$, $T \ge 0$, and $x \in K^{-1}\mathbf{N}^*$, let us set

$$\tilde{\Gamma}_x^{K;t,T} = \sum_{y\in K^{-1}\mathbf{N}^*} \mu_x^K\left(X\left(t-2t_1(K)\right)=y \,\middle|\, X(t+T)>0\right)\Gamma_y^{K;T}.$$

As a consequence of (71), its first marginal is $\tilde{\mu}_x^{K;t,T}$, while its second marginal is ν_*. Let the processes $\left(\tilde{X}(s)\,;\,0\le s\le T\right)$ and $\left(\tilde{U}(s)\,;\,s\ge 0\right)$ be defined on the space $\mathcal{D}\left([0,T],\mathbf{R}\right)\times\mathcal{C}\left(\mathbf{R}_+,\mathbf{R}\right)$ by $\tilde{X}(s)(\omega_1,\omega_2)=\omega_1(s)$ and $\tilde{U}(s)(\omega_1,\omega_2)=\omega_2(s)$. Combining (77) and (78) yields, for K large enough, for all $t\ge 2t_1(K)$, $x\in K^{-1}\mathbf{N}^*$ and $T\le \exp\left(V_0'K/2\right)/2$,

$$\begin{aligned}
&\tilde{\Gamma}_x^{K;t,T}\left(\sup_{0\le s\le T}\left|\tilde{X}(s)-x_*-\tilde{U}(s)/\sqrt{K}\right|>\varepsilon(K)\right)\\
&\le \mu_x^K\left(X(t-2t_1(K))\notin[x_*/4,3x_*]\,\middle|\,X(t+T)>0\right)+2(T+1)\exp\left(-(V/2)K\varepsilon(K)\right)\\
&\le C_2\left(\exp\left(-\frac{t-2t_1(K)}{\beta\log(K)}\right)+\exp\left(-\frac{V_0'}{2}K\right)\right)+2(T+1)\exp\left(-(V/2)K\varepsilon(K)\right).
\end{aligned}$$

Setting $C'''=C_2e^{6/(\beta x_*)}\vee 3$, and using that $\exp\left(-V_0'K/2\right)\ll\exp\left(-(V/2)K\varepsilon(K)\right)$ and that a probability is less than 1, we obtain that, for K large enough, for all $t\ge 2t_1(K)$, $x\in K^{-1}\mathbf{N}^*$ and $T\ge 0$,

$$\begin{aligned}
&\tilde{\Gamma}_x^{K;t,T}\left(\sup_{0\le s\le T}\left|\tilde{X}(s)-x_*-\tilde{U}(s)/\sqrt{K}\right|>\varepsilon(K)\right)\\
&\le C'''\left(\exp\left(-\frac{t}{\beta\log(K)}\right)+(T+1)\exp\left(-(V/2)K\varepsilon(K)\right)\right).
\end{aligned}$$

Since $\Gamma_x^{K;t,T}\circ\tilde{X}^{-1}=\tilde{\mu}_x^{K;t,T}$ and $\Gamma_x^{K;t,T}\circ\tilde{U}^{-1}=\nu_*$, the proposition is proved.

3.7 Proof of Corollary 2.3

Let $C''',V'',\alpha,\beta>0$ be given by Proposition 2.2. Let $\varepsilon:\mathbf{R}_+^*\to\mathbf{R}_+^*$ be such that $\alpha\log(K)/K\le\varepsilon(K)\ll 1$. Proposition 2.2 entails that for K large enough, and for all $t\ge(6/x_*)\log(K)$, we can construct a coupling $\left(\tilde{X},G\right)$ of $\left(\tilde{\mu}_{x_*}^{K;t},\mathscr{N}\left(0,\Sigma_*\right)\right)$ such that

$$\mathbf{P}\left(\left|\tilde{X}-x_*-G/\sqrt{K}\right|>\varepsilon(K)\right)\le C'''\left(\exp\left(-\frac{t}{\beta\log(K)}\right)+\exp\left(-V''K\varepsilon(K)\right)\right). \tag{79}$$

Let $\Gamma^{K;t}$ be the probability distribution of $(\tilde{X},G)$. We may suppose that the underlying probability space of $\left(\tilde{X},G\right)$ is $\left(\mathbf{R}^2,\mathscr{B}(\mathbf{R}^2),\Gamma^{K;t}\right)$ and that $\tilde{X}$ and G are respectively the first and second canonical projections from $\mathbf{R}^2$ to $\mathbf{R}$. We know that the first marginal of $\Gamma^{K;t}$ converges weakly to γ^K as $t\to+\infty$, while its second marginal is constant, hence $\left(\Gamma^{K;n}\,;\,n\in\mathbf{N}\right)$ is tight in $\mathcal{P}(\mathbf{R}^2)$. Therefore there exists an increasing integer sequence $(n_j\,;\,j\in\mathbf{N})$ and $\Gamma^K\in\mathcal{P}\left(\mathbf{R}^2\right)$ such that

$$\Gamma^{K;n_j}\underset{j\to+\infty}{\Longrightarrow}\Gamma^K.$$

The first marginal of Γ^K is γ^K, and its second marginal is $\mathscr{N}\left(0,\Sigma_*\right)$. For every open subset $\mathscr{O}$ of $\mathbf{R}^2$, we have $\Gamma^K(\mathscr{O})\le\liminf_{j\to\infty}\Gamma^{K;n_j}(\mathscr{O})$, thus (79) entails

$$\Gamma^K\left(\left|\tilde{X}-x_*-G/\sqrt{K}\right|>\varepsilon(K)\right)\le C'''\exp\left(-V''K\varepsilon(K)\right).$$

Let $\mathbf{E}_{\Gamma^K}$ denote the expectation under Γ^K. For K large enough, we have

$$\mathbf{E}_{\Gamma^K}\left[c\left(\sqrt{K}(\tilde{X}-x_*),G\right)\right] \leq \Gamma^K\left(\left|\sqrt{K}\left(\tilde{X}-x_*\right)-G\right| > \sqrt{K}\varepsilon(K)\right)+\sqrt{K}\varepsilon(K),$$

and this entails, by definition of $\mathcal{W}_c$ and $\tilde{\gamma}^K$,

$$\mathcal{W}_c\left(\tilde{\gamma}^K, \mathscr{N}(0,\Sigma_*)\right) \leq C''' \exp\left(-V'' K\varepsilon(K)\right)+\sqrt{K}\varepsilon(K).$$

Given that $\varepsilon(K) \geq \alpha \log(K)/K$, the second term of the above right handside is at best of order $\mathcal{O}\left(\sqrt{K}/\log(K)\right)$. Choosing $\varepsilon(K) = (\alpha \vee 1/(2V''))\log(K)/K$ we obtain, for K large enough,

$$\mathcal{W}_c\left(\tilde{\gamma}^K, \mathscr{N}(0,\Sigma_*)\right) \leq \frac{C''' + (\alpha \vee 1/(2V''))\log(K)}{\sqrt{K}}.$$

3.8 Proof of Proposition 2.4

The SDE (11) satisfied by U_{x_*} yields, for all $t \geq 0$,

$$\int_0^t U_{x_*}^{(2)}(s)\mathrm{d}s = \left(F'(x_*)^{-1}U_{x_*}\right)^{(2)}(t) - \left(F'(x_*)^{-1}S_*^{1/2}B\right)^{(2)}(t).$$

Then (12) follows from the equality

$$\begin{pmatrix}0 & 1\end{pmatrix} F'(x_*)^{-1}S_*\left(F'(x_*)^{-1}\right)^T \begin{pmatrix}0\\1\end{pmatrix} = \sigma^2.$$

Let $T\colon \mathbf{R}_+^* \to \mathbf{R}_+^*$ be such that $1 \ll T(K) \ll K^p$ for some $p > 1$. Theorem 1.1 yields

$$\mathbf{P}\left(\sup_{0\leq s\leq T(K)}\left|I^K(s) - i_* - U_{x_*}^{(2)}(s)/\sqrt{K}\right| > (\alpha \vee p/V)\log(K)/K\right) \underset{K\to\infty}{\longrightarrow} 0.$$

Moreover, since $U_{x_*}(T(K))$ converges in distribution as $K \to +\infty$, we have $U_{x_*}(T(K)) = \mathcal{O}_{\mathbf{P}}(1)$. Hence, we obtain

$$\int_0^{T(K)} I^K(s)\mathrm{d}s = i_*\, T(K) + \sigma\sqrt{T(K)/K}\,\mathscr{N}(0,1) + \mathcal{O}_{\mathbf{P}}\left(1/\sqrt{K} + T(K)\log(K)/K\right),$$

which concludes the proof.

3.9 A coupling lemma

In what follows, $\mathscr{U}(0,1)$ stands for the uniform distribution on the interval $(0,1)$.

Lemma 3.12. *Let E_1 and E_2 be two complete separable metric spaces, let μ be a probability distribution on $(E_1\times E_2, \mathscr{B}(E_1)\otimes\mathscr{B}(E_2))$. Let μ_1 denote the first marginal of μ. There exists a measurable function $G : E_1 \times (0,1) \to E_2$ such that if $(X_1, V) \sim \mu_1 \otimes \mathscr{U}(0,1)$, then $(X_1, G(X_1,V)) \sim \mu$.*

Proof. We may suppose without loss of generality that E_1 and E_2 are Borel subsets of $\mathbf{R}$, thanks to the Borel isomorphism theorem (see e.g. Theorem 13.1.1 in [12]). There exist a probability kernel $R : E_1 \times \mathscr{B}(E_2) \to [0,1]$ such that $\mu = \mu_1 \otimes R$, i.e. $\mu(A\times B) = \int_A \mu_1(\mathrm{d}x_1)R(x_1,B)$ for all $A \in \mathscr{B}(E_1)$ and $B \in \mathscr{B}(E_2)$ (see e.g. Theorem 9.2.2 in [30]). Define $G_0 : E_1 \times (0,1) \to \mathbf{R}$ by

$$G_0(x,v) = \inf\left\{y \in \mathbf{R} : R(x,(-\infty,y]) \geq v\right\}.$$

For all $x \in E_1$, $v \in (0,1)$ and $a \in \mathbf{R}$, we have

$$G_0(x,v) \leq a \Leftrightarrow R(x,(-\infty,a]) \geq v.$$

This entails that G_0 is measurable and that $G_0(x,V) \sim R(x,\cdot)$ for all $x \in E_1$ and $V \sim \mathscr{U}(0,1)$. Let $(X_1, V) \sim \mu_1 \otimes \mathscr{U}(0,1)$, we have, for all $A \in \mathscr{B}(E_1)$ and $B \in \mathscr{B}(E_2)$:

$$\begin{aligned}\mathbf{P}(X_1 \in A, G_0(X_1,V) \in B) &= \int_A \mu_1(\mathrm{d}x) \int_0^1 \mathrm{d}v \mathbf{1}_B(G_0(x,v)) \\ &= \int_A \mu_1(\mathrm{d}x) R(x,B) \\ &= \mu(A \times B),\end{aligned}$$

hence $(X_1, G_0(X_1,V)) \sim \mu$. We have almost finished, except that we still need to modify the function G_0 to get a function G taking values in E_2. But we know that $(X_1, G_0(X_1,V)) \in E_1 \times E_2$ a.s., thus if we fix $y \in E_2$ and define $G : E_1 \times (0,1) \to E_2$ by $G(x,v) = G_0(x,v)$ if $G_0(x,v) \in E_2$ and $G(x,v) = y$ otherwise, we have $(X_1, G(X_1,V)) = (X_1, G_0(X_1,V))$ a.s., which ends the proof.

Acknowledgements

I am grateful to Vincent Bansaye and Florent Malrieu for introducing this topic to me and guiding my research. This work has been supported by the Chair "Modélisation Mathématique et Biodiversité" of VEOLIA Environnement-Ecole Polytechnique-MNHN-F.X and by the ANR project ABIM (ANR-16-CE40-0001).

References

[1] L. J. S. Allen. *An introduction to stochastic processes with applications to biology.* CRC Press, Boca Raton, FL, second edition, 2011. 1

[2] K. B. Athreya and P. E. Ney. *Branching processes.* Springer-Verlag, New York-Heidelberg, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196. 29

[3] V. Bansaye and S. Méléard. *Stochastic models for structured populations*, volume 1 of *Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems.* Springer, Cham; MBI Mathematical Biosciences Institute, Ohio State University, Columbus, OH, 2015. Scaling limits and long time behavior. 7

[4] A. D. Barbour. Quasi-stationary distributions in Markov population processes. *Advances in Appl. Probability*, 8(2):296–314, 1976. 6

[5] I. Berkes, W. Liu, and W. B. Wu. Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.*, 42(2):794–817, 2014. 2

[6] T. Britton and E. Pardoux. Stochastic epidemics in a homogeneous community. *arXiv e-prints*, page arXiv:1808.05350, Aug. 2018. 1, 3, 4, 6

[7] J.-R. Chazottes, P. Collet, and S. Méléard. Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes. *Probab. Theory Related Fields*, 164(1-2):285–332, 2016. 7

[8] J.-R. Chazottes, P. Collet, and S. Méléard. On time scales and quasi-stationary distributions for multitype birth-and-death processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(4):2249–2294, 2019. 8

[9] J.-R. Chazottes, P. Collet, S. Méléard, and S. Martínez. Quasi-Stationary Distributions and Resilience: What to get from a sample? *Journal de l'École polytechnique - Mathématiques*, June 2020. 8

[10] Z. Chen. *Asymptotic problems related to Smoluchowski-Kramers approximation.* ProQuest LLC, Ann Arbor, MI, 2006. Thesis (Ph.D.)–University of Maryland, College Park. 6

[11] P. Collet, S. Martínez, and J. San Martín. *Quasi-stationary distributions*. Probability and its Applications (New York). Springer, Heidelberg, 2013. Markov chains, diffusions and dynamical systems. 7

[12] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original. 34

[13] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence. 2, 3, 17

[14] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs. 4, 6

[15] F. Götze and A. Y. Zaitsev. Bounds for the rate of strong approximation in the multidimensional invariance principle. *Teor. Veroyatn. Primen.*, 53(1):100–123, 2008. 2

[16] S. Gouëzel. Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.*, 38(4):1639–1671, 2010. 2

[17] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989. 16

[18] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975. 1

[19] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1):33–58, 1976. 1

[20] T. G. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probability*, 8:344–356, 1971. 1, 2

[21] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.*, 6(3):223–240, 1977/78. 1, 3, 16

[22] T. G. Kurtz. *Approximation of population processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981. 1

[23] S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. *Probab. Surv.*, 9:340–410, 2012. 7

[24] F. Merlevède and E. Rio. Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electron. J. Probab.*, 20:no. 14, 27, 2015. 2

[25] P. Mozgunov, M. Beccuti, A. Horvath, T. Jaki, R. Sirovich, and E. Bibbona. A review of the deterministic and diffusion approximations for stochastic chemical reaction networks. *Reaction Kinetics, Mechanisms and Catalysis*, 01 2018. 4

[26] E. Pardoux. Moderate deviations and extinction of an epidemic. *Electron. J. Probab.*, 25:27 pp., 2020. 4, 6

[27] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. 17

[28] S. Sagitov and A. Shaimerdenova. Extinction times for a birth-death process with weak competition. *Lith. Math. J.*, 53(2):220–234, 2013. 31

[29] A. Shwartz and A. Weiss. *Large deviations for performance analysis.* Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei. 3, 4, 6

[30] D. W. Stroock. *Probability theory.* Cambridge University Press, Cambridge, second edition, 2011. An analytic view. 34

[31] G. Teschl. *Ordinary differential equations and dynamical systems*, volume 140 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2012. 3, 10

[32] E. A. van Doorn. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. in Appl. Probab.*, 23(4):683–700, 1991. 7