

## CONTINUOUS-TIME MULTITYPE BRANCHING PROCESSES CONDITIONED ON VERY LATE EXTINCTION<sup>\*,\*\*</sup>

SOPHIE PÉNISSON<sup>1</sup>

**Abstract.** Multitype branching processes and Feller diffusion processes are conditioned on very late extinction. The conditioned laws are expressed as Doob  $h$ -transforms of the unconditioned laws, and an interpretation of the conditioned paths for the branching process is given, *via* the immortal particle. We study different limits for the conditioned process (increasing delay of extinction, long-time behavior, scaling limit) and provide an exhaustive list of exchangeability results.

**Mathematics Subject Classification.** 60J60, 60J80.

Received October 19, 2009.

### INTRODUCTION

In this paper we analyse conditional limit theorems for multitype branching processes. As described in the early papers of Kolmogorov ([15]) and Yaglom ([22]), interesting non-degenerate limits for branching processes are obtained by conditioning on non-extinction. The study of the multitype case leads to the same statement, and Yaglom's theorem asserting the existence of a limiting probability measure for a process  $X_t$  conditioned on  $X_t$  to be alive can be generalized to the class of multitype irreducible subcritical processes ([21]). A natural alternative is to condition the population to be still extant at some fixed time  $T$ , but this yields time-inhomogeneous kernels. However, considering solely the particles at each time  $t \in [0, T]$  having descendants at time  $T$ , one can obtain asymptotical results as  $T$  tends to infinity. For this topic we refer to the literature on reduced branching processes (see for example the seminal work of Fleischmann and Prehn in [9], or [10]). It is also meaningful to generalize Yaglom's results by conditioning the process  $X_t$  on the event that it is not extinct at time  $t + \theta$ , but does eventually die out. The extinction is thus delayed of at least  $\theta$ . The first published result on this question is due to Lamperti and Ney ([17]) and applies to one-dimensional Bienaymé-Galton-Watson processes: letting the delay of extinction  $\theta$  tend to infinity yields a Markov chain known as the  $Q$ -process. This result has been later extended to multitype irreducible Bienaymé-Galton-Watson processes ([3]).

We are here concerned with multitype continuous-time irreducible branching processes conditioned on very late extinction, *i.e.* the continuous-time analog of the  $Q$ -process, as well as with their continuous-state counterpart. As proved by Feller in [6], one can indeed apply a continuous approximation for large branching

---

*Keywords and phrases.* Multitype branching process, Feller diffusion process, conditioned branching process, diffusion limit, extinction, immortal particle, long-time behavior.

\* *Research supported by the Université franco-allemande and the IRTG Stochastic Models of Complex Processes.*

\*\* *The author especially wishes to thank her Ph.D. thesis supervisor, Prof. Dr. Sylvie Røelly.*

<sup>1</sup> Laboratoire d'Analyse et de Mathématiques Appliquées UMR 8050, Université Paris-Est Créteil, 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France. [sophie.penisson@u-pec.fr](mailto:sophie.penisson@u-pec.fr)

populations, leading to processes of the diffusion type (Feller diffusion processes). Such a diffusion approximation has also been investigated in [17] for branching processes conditioned on very late extinction. However, the result only concerns single-type processes and the convergence of the finite-dimensional distributions. In this paper we seek to analyse the connection between conditioned multitype branching processes and conditioned Feller diffusion processes. Multitype Feller diffusion processes conditioned on very late extinction first appeared in [2] (see [19] for the monotype case).

It is known that the  $Q$ -process introduced in [17] is positive recurrent under additional assumptions, and admits as stationary measure a size-biased Yaglom distribution ([1]). In [2], the authors prove that the multitype conditioned Feller diffusion process has a stationary distribution as well, which raises many natural questions. Does the multitype conditioned branching process also admit a stationary distribution? Can this stationary distribution be related to that of the conditioned Feller process? Considering the process  $X_t$  conditioned on  $X_{t+\theta} \neq 0$ , the asymptotic behavior of the process conditioned on very late extinction is obtained by letting first  $\theta$  tend to infinity and then  $t$  tend to infinity. Would we obtain the same limit by taking first the limit  $t \rightarrow \infty$  and next the limit  $\theta \rightarrow \infty$ ? We know from [2] that the answer is positive for a Feller diffusion process, and will discuss the same question for branching processes.

The first section of this paper is dedicated to the multitype branching process conditioned on very late extinction. We precisely define its law and in Theorem 1.1 express it as a Doob  $h$ -transform of the law of the unconditioned process. In a second instance we provide an interpretation of the conditioned paths *via* an immortal particle, and finally discuss the commutativity of the long-time limits in  $t$  and  $\theta$ . The second section is devoted to the conditioned Feller diffusion process. After recalling its definition and known results, we present in Theorem 2.2 the conditioned Feller process as a solution to a limit martingale problem, which identifies it as a limit of rescaled conditioned branching processes. Since this convergence also holds for unconditioned processes, we deduce from this result that “rescaling and conditioning commute”. We finally investigate the many possible commutativity results between the long-time and scaling limits of a rescaled conditioned branching process.

**Notation**

Let  $d$  be the number of types. In this paper we use the following notation.

$$\mathbb{N} := \{0, 1, 2, \dots\}, \quad \mathbb{N}^* := \{1, 2, \dots\},$$

$$\mathbb{R}_+ := [0, \infty[, \quad \mathbb{R}_+^d := [0, \infty[^d.$$

Any  $d$ -dimensional vector  $\mathbf{x} \in \mathbb{R}^d$  is denoted  $(x_1, \dots, x_d)$ , and its transpose  $\mathbf{x}^t$ .  $\mathbf{1}$  and  $\mathbf{0}$  denote the vectors  $(1, \dots, 1)$  and  $(0, \dots, 0) \in \mathbb{R}^d$ , and for all  $i = 1 \dots d$ ,  $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$  the basis vector of  $\mathbb{R}^d$ .  $\mathbf{x} \cdot \mathbf{y}$  denotes the scalar product between  $\mathbf{x}$  and  $\mathbf{y}$  in  $\mathbb{R}^d$ , and  $\|\mathbf{x}\|$  the euclidian norm. We define moreover

$$\mathbf{x}\mathbf{y} := (x_1y_1, \dots, x_dy_d) \quad \text{and} \quad \mathbf{x}^{\mathbf{y}} := x_1^{y_1} \dots x_d^{y_d}.$$

We introduce the following partial order on  $\mathbb{R}^d$ :  $\mathbf{x} \leq \mathbf{y}$  (resp.  $\mathbf{x} < \mathbf{y}$ ) means that for all  $i = 1 \dots d$ ,  $x_i \leq y_i$  (resp.  $x_i < y_i$ ). We call a matrix *positive* (resp. *non-negative*) if all its coefficients are  $> 0$  (resp.  $\geq 0$ ).

Throughout this paper we work on the probability space  $(\Omega, (\mathbf{X}_t)_{t \geq 0}, (\mathcal{F}_t)_{t \geq 0})$ , where  $\Omega := D(\mathbb{R}_+, \mathbb{R}_+^d)$  is the canonical space of càdlàg functions from  $\mathbb{R}_+$  to  $\mathbb{R}_+^d$ . For every  $t \geq 0$ ,  $\mathbf{X}_t$  denotes the canonical  $t$ -th projection, and  $(\mathcal{F}_t)_{t \geq 0}$  the right-continuous filtration generated by the canonical process  $(\mathbf{X}_t)_{t \geq 0}$ .

1. MULTITYPE BRANCHING PROCESSES CONDITIONED ON VERY LATE EXTINCTION

1.1. The model

In this section we consider a continuous-time  $\mathbb{N}^d$ -valued branching process with law  $\mathbb{P}$  on the probability space  $(\Omega, (\mathbf{X}_t)_{t \geq 0}, (\mathcal{F}_t)_{t \geq 0})$ . Let  $(\mathbf{p}(\mathbf{j}))_{\mathbf{j} \in \mathbb{N}^d}$  be the offspring distribution of the branching process, where for all  $i = 1 \dots d$  and all  $\mathbf{j} \in \mathbb{N}^d$ ,  $p_i(\mathbf{j}) \in [0, 1]$  denotes the probability that a particle of type  $i$  produces  $j_1$  particles

of type 1,  $j_2$  particles of type 2 etc. We denote by  $\mathbf{f}(\mathbf{r})$  the generating function of the offspring distribution, defined for all  $i = 1 \dots d$  and all  $\mathbf{r} \in [0, 1]^d$  by  $f_i(\mathbf{r}) := \sum_{\mathbf{j} \in \mathbb{N}^d} p_i(\mathbf{j}) \mathbf{r}^{\mathbf{j}}$ . Let  $\alpha_i > 0, i = 1 \dots d$ , be the branching rates: each particle of type  $i$  lives an exponentially distributed lifetime of parameter  $\alpha_i$  before branching. The infinitesimal generator of the process is then, for all smooth function  $f : \mathbb{R}_+ \times \mathbb{N}^d \rightarrow \mathbb{R}$  and all  $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{N}^d$  (see *e.g.* Sect. 9.2 in [4]),

$$(Lf)(t, \mathbf{x}) := \frac{\partial f}{\partial t}(t, \mathbf{x}) + \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} p_i(\mathbf{k}) [f(t, \mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(t, \mathbf{x})]. \tag{1.1}$$

From now on we assume that the first order moments of the offspring distribution are finite, and denote by  $\mathbf{M}$  the *mean matrix* with entries defined for all  $i, j = 1 \dots d$  by  $m_{ij} := \sum_{\mathbf{k} \in \mathbb{N}^d} k_j p_i(\mathbf{k})$ . We assume moreover that  $\mathbf{M}$  is *irreducible*, *i.e.* that there exists no permutation matrix  $\mathbf{S}$  such that  $\mathbf{S}^{-1} \mathbf{M} \mathbf{S}$  is block triangular. We will thus work under the following assumption.

**(A1)** The mean matrix  $\mathbf{M}$  is *finite* and *irreducible*.

Under assumption **(A1)**, the matrix defined by  $\mathbf{C} := \mathbf{A}(\mathbf{M} - \mathbf{I})$ , where  $\mathbf{A}$  is the diagonal matrix  $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_d)$ , is irreducible too and all its non-diagonal elements are non-negative. An extension of Perron-Frobenius theorem (see Satz 13.2.2 in [7] or Thm. 2.5 in [20]) implies that  $\mathbf{C}$  admits a real eigenvalue  $\rho$  larger than the real part of any other eigenvalue. The so-called Perron’s root  $\rho$  is simple, with a one-dimensional eigenspace, and there correspond right and left eigenvectors with positive coordinates. In the following we denote by  $\boldsymbol{\xi}$  (resp.  $\boldsymbol{\eta}$ ) the associated right (resp. left) eigenvector with normalization  $\boldsymbol{\xi} \cdot \mathbf{1} = 1, \boldsymbol{\eta} \cdot \boldsymbol{\xi} = 1$ . The branching process with law  $\mathbb{P}$  is called *critical, subcritical* or *supercritical* according as  $\rho = 0, \rho < 0$  or  $\rho > 0$ .

### 1.2. The conditioned branching process as $h$ -process

We are interested in conditioning the discrete-state branching process on *very late extinction*. In the (sub)critical non-simple case (see definition below) the extinction of the process occurs almost surely, hence the conditioning simply consists in delaying the extinction: we consider the law of the process at time  $t$  conditionally on non-extinction at time  $t + \theta$ , and let  $\theta$  tend to infinity. In the supercritical case however we face the “extinction *versus* explosion” dichotomy: the process can escape extinction with positive probability, and if so it explodes to infinity. In order to condition a supercritical process on very late extinction we thus need firstly to condition the process on extinction, and secondly to delay the extinction similarly as in the (sub)critical case.

Let us now define more precisely the law  $\mathbb{P}^*$  of the process conditioned on very late extinction, for any class of criticality. We first introduce the law  $\tilde{\mathbb{P}}$  of the process conditioned on *extinction*,

$$\tilde{\mathbb{P}}(\cdot) := \mathbb{P}\left(\cdot \mid \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0}\right). \tag{1.2}$$

Denoting by a subscript the initial condition, this definition makes sense if for any  $\mathbf{x} \in \mathbb{N}^d, \mathbb{P}_{\mathbf{x}}(\lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0}) > 0$ , or equivalently if for any  $\mathbf{x} \in \mathbb{N}^d, \lim_{s \rightarrow \infty} \mathbb{P}_{\mathbf{x}}(\mathbf{X}_s = \mathbf{0}) > 0$ . Thanks to the branching property, this is satisfied if for all  $i = 1 \dots d, \lim_{s \rightarrow \infty} \mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_s = \mathbf{0}) > 0$ . We define  $\mathbf{q}(t)$  the extinction probability vector at time  $t$  (resp.  $\mathbf{q}$  the extinction probability vector) as follows, for all  $i = 1 \dots d$  and all  $t \geq 0$ ,

$$q_i(t) := \mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_t = \mathbf{0}),$$

$$q_i := \lim_{t \rightarrow \infty} q_i(t),$$

and the assumption

**(A2)** The branching process has a *positive risk of extinction*  $\mathbf{q} > \mathbf{0}$ .

From what precedes it thus appears that under **(A2)**, which will be assumed in all this paper,  $\tilde{\mathbb{P}}$  is a well-defined probability measure on  $(\Omega, (\mathbf{X}_t)_{t \geq 0}, (\mathcal{F}_t)_{t \geq 0})$ . If the branching process is supercritical, assumption **(A2)** excludes the degenerate case for which the process explodes almost surely. If the process is critical, assumption **(A2)** avoids the trivial case of a *simple* process, *i.e.* a process with a generating function  $f_i(\mathbf{r})$  linear in  $r_1, \dots, r_d$  with no constant term. In this case each particle has exactly one offspring and the process has a constant number of particles.

Note that under assumption **(A2)** a (sub)critical process almost surely dies out ([21], Satz 5.1.7), hence conditioning on extinction in (1.2) does not change the measure and we have  $\tilde{\mathbb{P}} = \mathbb{P}$ .

We can now condition on a *delayed extinction*. More precisely we define the law  $\mathbb{P}^*$  by first delaying the extinction of at least  $\theta$  and then by letting  $\theta$  tend to infinity: for all  $t \geq 0$  and all  $B \in \mathcal{F}_t$ ,

$$\mathbb{P}^*(B) := \lim_{\theta \rightarrow \infty} \tilde{\mathbb{P}}(B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}), \tag{1.3}$$

if this limit exists. To draw a parallel with the  $Q$ -process defined in [3] for a multitype Bienaymé-Galton-Watson (BGW) process, we can write the conditioned law  $\mathbb{P}^*$  as follows,

$$\mathbb{P}^*(B) = \lim_{\theta \rightarrow \infty} \mathbb{P}\left(B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0}\right). \tag{1.4}$$

It appears that the law  $\mathbb{P}^*$  can be roughly thought as the law of the process “conditioned on not being extinct in the distant future and on being extinct in the even more distant future” ([1] Sect. 1.14).

In the following Theorem 1.1 we prove that the limit  $\mathbb{P}^*|_{\mathcal{F}_t}$  given by (1.3) is a well-defined probability measure on  $\mathcal{F}_t$  which is absolutely continuous with respect to  $\mathbb{P}|_{\mathcal{F}_t}$ .

In order to handle with the supercritical case, for which  $\tilde{\mathbb{P}} \neq \mathbb{P}$ , we need the intermediate result given by Lemma 1.2, which states that  $\tilde{\mathbb{P}}$  is the law of a specific subcritical branching process. In [11], the authors prove that general branching processes conditioned on extinction remain branching processes, and more specifically that supercritical general branching processes conditioned on extinction are subcritical. In Lemma 1.2 we provide the explicit parameters of the subcritical process with law  $\tilde{\mathbb{P}}$  which will be needed in Theorem 1.1.

**Theorem 1.1.** *Let  $\mathbb{P}$  be the law of an irreducible branching process with positive risk of extinction. We assume moreover that*

- (i) if  $\rho = 0$ , all the second order moments of the offspring distribution  $\frac{\partial^2 f_i}{\partial r_j \partial r_k}(\mathbf{1})$  are finite,
- (ii) if  $\rho < 0$ , for all  $t > 0$  and all  $i, j = 1 \dots d$ ,  $\mathbb{E}_{e_i} [X_{t,j} \ln X_{t,j}] < \infty$ ,
- (iii) if  $\rho > 0$ , for all  $t > 0$  and all  $i, j = 1 \dots d$ ,  $\mathbb{E}_{e_i} [q^{X_t} X_{t,j} \ln X_{t,j}] < \infty$ .

Then  $\mathbb{P}^*$  is a Doob  $h$ -transform of  $\mathbb{P}$  satisfying for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$d\mathbb{P}^*_{\mathbf{x}}|_{\mathcal{F}_t} = e^{-\tilde{\rho}t} \frac{q^{\mathbf{X}_t}}{q^{\mathbf{x}}} \frac{\mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}}}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} d\mathbb{P}_{\mathbf{x}}|_{\mathcal{F}_t}, \tag{1.5}$$

where  $\tilde{\rho}$  is the Perron’s root of the irreducible matrix  $\tilde{\mathbf{C}} := \mathbf{A}(\tilde{\mathbf{M}} - \mathbf{I})$ , with  $\tilde{\mathbf{M}}$  defined by

$$\tilde{m}_{ij} = \frac{q_j}{q_i} \frac{\partial f_i}{\partial r_j}(q), \tag{1.6}$$

and  $\tilde{\boldsymbol{\xi}}$  is the right eigenvector of  $\tilde{\mathbf{C}}$  for  $\tilde{\rho}$  with norm  $\tilde{\boldsymbol{\xi}} \cdot \mathbf{1} = 1$ .

In particular, if the process is (sub)critical, then  $\mathbb{P}^*$  satisfies for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$d\mathbb{P}_{\mathbf{x}}^*|_{\mathcal{F}_t} = e^{-\rho t} \frac{\mathbf{X}_t \cdot \boldsymbol{\xi}}{\mathbf{x} \cdot \boldsymbol{\xi}} d\mathbb{P}_{\mathbf{x}}|_{\mathcal{F}_t}. \tag{1.7}$$

Let us first introduce the following preliminary result.

**Lemma 1.2.** *Let  $\mathbb{P}$  be the law of an irreducible supercritical branching process with positive risk of extinction. Then  $\tilde{\mathbb{P}}$  is a Doob  $h$ -transform of  $\mathbb{P}$  satisfying for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,*

$$d\tilde{\mathbb{P}}_{\mathbf{x}}|_{\mathcal{F}_t} = \frac{\mathbf{q}^{\mathbf{X}_t}}{\mathbf{q}^{\mathbf{x}}} d\mathbb{P}_{\mathbf{x}}|_{\mathcal{F}_t}. \tag{1.8}$$

Moreover,  $\tilde{\mathbb{P}}$  is the law of an irreducible subcritical branching process with branching rates  $(\alpha_1, \dots, \alpha_d)$  and offspring generating function  $\tilde{\mathbf{f}}$

$$\tilde{f}_i(\mathbf{r}) := \frac{1}{q_i} f_i(\mathbf{q}\mathbf{r}). \tag{1.9}$$

**Remark 1.3.** For any class of criticality we will in the following use the definition (1.9) for  $\tilde{\mathbf{f}}$ , and the associated probability distributions

$$\tilde{p}_i(\mathbf{k}) := \frac{1}{q_i} \mathbf{q}^{\mathbf{k}} p_i(\mathbf{k}), \quad i = 1 \dots d, \quad \mathbf{k} \in \mathbb{N}^d. \tag{1.10}$$

In the (sub)critical case we clearly have  $\tilde{\mathbf{f}} = \mathbf{f}$  and  $\tilde{\mathbf{p}}(\mathbf{k}) = \mathbf{p}(\mathbf{k})$ .

*Proof of Lemma 1.2.* Let  $t \geq 0$ ,  $B \in \mathcal{F}_t$  and  $\mathbf{x} \in \mathbb{N}^d$ . By definition,

$$\tilde{\mathbb{E}}_{\mathbf{x}}(\mathbf{1}_B) = \frac{\mathbb{E}_{\mathbf{x}}[\mathbf{1}_B \mathbb{P}_{\mathbf{x}}(\lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} | \mathcal{F}_t)]}{\mathbb{P}_{\mathbf{x}}(\lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0})}.$$

The branching property implies that for all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbb{P}_{\mathbf{x}}(\lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0}) = \mathbf{q}^{\mathbf{x}}$ , which together with the Markov property leads to

$$\tilde{\mathbb{E}}_{\mathbf{x}}(\mathbf{1}_B) = \mathbb{E}_{\mathbf{x}} \left[ \mathbf{1}_B \frac{\mathbf{q}^{\mathbf{X}_t}}{\mathbf{q}^{\mathbf{x}}} \right].$$

It ensues (1.8) and that  $(\mathbf{q}^{\mathbf{X}_t})_{t \geq 0}$  is a  $(\mathbb{P}, \mathcal{F}_t)$ -martingale. Defining for all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\tilde{h}(\mathbf{x}) := \mathbf{q}^{\mathbf{x}}$ , the infinitesimal generator  $\tilde{L}$  of the conditioned process with law  $\tilde{\mathbb{P}}$  is then given for all smooth function  $f : \mathbb{N}^d \rightarrow \mathbb{R}$  by

$$\tilde{L}f := \frac{1}{\tilde{h}} L(\tilde{h}f).$$

Hence, for all  $\mathbf{x} \in \mathbb{N}^d$ ,

$$(\tilde{L}f)(\mathbf{x}) = \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} p_i(\mathbf{k}) \left[ \frac{1}{q_i} \mathbf{q}^{\mathbf{k}} f(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(\mathbf{x}) \right].$$

As a fixed point of the generating function  $\mathbf{f}$ , the extinction probability vector  $\mathbf{q}$  satisfies for all  $i = 1 \dots d$ ,  $\sum_{\mathbf{k} \in \mathbb{N}^d} p_i(\mathbf{k}) \mathbf{q}^{\mathbf{k}} = q_i$ , and it follows

$$(\tilde{L}f)(\mathbf{x}) = \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \tilde{p}_i(\mathbf{k}) [f(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(\mathbf{x})]. \tag{1.11}$$

Knowing from [11] that  $\tilde{\mathbb{P}}$  is the law of a subcritical branching process, we deduce from (1.11) its offspring generating function  $\tilde{\mathbf{f}}$  and branching rates  $\alpha_i$ . We easily check that the associated irreducible mean matrix is the matrix  $\tilde{\mathbf{M}}$  introduced in Theorem 1.1.  $\square$

**Remark 1.4.** Due to the equivalence between extinction of the continuous-time process and extinction of its embedded generation counting process, we obtain unsurprisingly the same offspring generating function  $\tilde{\mathbf{f}}$  as the one computed in [3] for conditioned BGW processes.

*Proof of Theorem 1.1.* The proof relies mostly on the asymptotical properties of the extinction probability vector  $\mathbf{q}(t)$  as  $t \rightarrow \infty$ , in both critical and subcritical cases. Thanks to Lemma 1.2, the supercritical case is then simply reduced to the subcritical case.

Let  $t \geq 0$  and  $B \in \mathcal{F}_t$ . By definition, for all  $\theta \geq 0$  and all  $\mathbf{x} \in \mathbb{N}^d, \mathbf{x} \neq \mathbf{0}$ ,

$$\tilde{\mathbb{E}}_{\mathbf{x}}[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \frac{\tilde{\mathbb{E}}_{\mathbf{x}} \left[ \mathbf{1}_B \tilde{\mathbb{P}}_{\mathbf{x}}(\mathbf{X}_{t+\theta} \neq \mathbf{0} \mid \mathcal{F}_t) \right]}{\tilde{\mathbb{P}}_{\mathbf{x}}(\mathbf{X}_{t+\theta} \neq \mathbf{0})}.$$

By virtue of the Markov and branching properties we obtain

$$\tilde{\mathbb{E}}_{\mathbf{x}}[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \frac{\tilde{\mathbb{E}}_{\mathbf{x}} \left[ \mathbf{1}_B (1 - \tilde{\mathbf{q}}(\theta)^{\mathbf{X}_t}) \right]}{1 - \tilde{\mathbf{q}}(t + \theta)^{\mathbf{x}}}. \tag{1.12}$$

In the critical case  $\rho = 0$  we have  $\tilde{\mathbb{P}} = \mathbb{P}$ , and (1.12) becomes

$$\mathbb{E}_{\mathbf{x}}[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \mathbb{E}_{\mathbf{x}} \left[ \mathbf{1}_B \frac{1 - \mathbf{q}(\theta)^{\mathbf{X}_t}}{1 - \mathbf{q}(t + \theta)^{\mathbf{x}}} \right]. \tag{1.13}$$

Let us define

$$\zeta := \sum_{i,j,k=1}^d \alpha_i \frac{\partial^2 f_i}{\partial r_j \partial r_k}(\mathbf{1}) \eta_i \xi_j \xi_k < \infty.$$

Then, according to Satz 6.3.4 in [21],  $\mathbf{q}(t)$  has the following asymptotic behavior: for all  $i = 1 \dots d$ ,

$$q_i(t) \sim 1 - \frac{2\xi_i}{\zeta t} \text{ as } t \rightarrow \infty.$$

It ensues that for all  $t \geq 0$  and all  $\mathbf{x}, \mathbf{y} \in \mathbb{N}^d, \mathbf{x} \neq \mathbf{0}$ ,

$$\lim_{\theta \rightarrow \infty} \frac{1 - \mathbf{q}(\theta)^{\mathbf{y}}}{1 - \mathbf{q}(t + \theta)^{\mathbf{x}}} = \lim_{\theta \rightarrow \infty} \frac{1 - \prod_{i=1}^d \left[ 1 - \frac{2\xi_i}{\zeta \theta} \right]^{y_i}}{1 - \prod_{i=1}^d \left[ 1 - \frac{2\xi_i}{\zeta(t+\theta)} \right]^{x_i}} = \frac{\sum_{i=1}^d y_i \xi_i}{\sum_{i=1}^d x_i \xi_i}.$$

Using dominated convergence in (1.13) we thus obtain that

$$\lim_{\theta \rightarrow \infty} \mathbb{E}_{\mathbf{x}}[\mathbf{1}_B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \mathbb{E}_{\mathbf{x}} \left[ \mathbf{1}_B \frac{\mathbf{X}_t \cdot \boldsymbol{\xi}}{\mathbf{x} \cdot \boldsymbol{\xi}} \right],$$

which leads to (1.7).

In the subcritical case  $\rho < 0$  we similarly use the known asymptotic behavior of  $\mathbf{q}(t)$  given in [21] Satz 6.2.7, which holds under assumption (ii): there exists  $K > 0$  such that for all  $i = 1 \dots d$ ,

$$q_i(t) \sim 1 - K \xi_i e^{\rho t} \text{ as } t \rightarrow \infty,$$

hence for all  $t \geq 0$  and for all  $\mathbf{x}, \mathbf{y} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$\lim_{\theta \rightarrow \infty} \frac{1 - \mathbf{q}(\theta)^{\mathbf{y}}}{1 - \mathbf{q}(t + \theta)^{\mathbf{x}}} = \lim_{\theta \rightarrow \infty} \frac{1 - \prod_{i=1}^d [1 - K\xi_i e^{\rho\theta}]^{y_i}}{1 - \prod_{i=1}^d [1 - K\xi_i e^{\rho(t+\theta)}]^{x_i}} = e^{-\rho t} \frac{\sum_{i=1}^d y_i \xi_i}{\sum_{i=1}^d x_i \xi_i},$$

which by dominated convergence in (1.13) leads to (1.7).

In the supercritical case  $\rho > 0$  we apply the previous result to the subcritical process with law  $\tilde{\mathbb{P}}$  (which is possible under assumption (iii) since  $\tilde{\mathbb{E}}_{\mathbf{e}_i} [X_{t,j} \ln X_{t,j}] = q_i^{-1} \mathbb{E}_{\mathbf{e}_i} [\mathbf{q}^{\mathbf{X}_t} X_{t,j} \ln X_{t,j}]$ ): for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$d\mathbb{P}_{\mathbf{x}}^* |_{\mathcal{F}_t} = e^{-\tilde{\rho}t} \frac{\mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}}}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} d\tilde{\mathbb{P}}_{\mathbf{x}} |_{\mathcal{F}_t}, \tag{1.14}$$

which combined with Lemma 1.2 leads to (1.5).

We finally check that for any class of criticality  $(e^{-\tilde{\rho}t} \mathbf{q}^{\mathbf{X}_t} \mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}})_{t \geq 0}$  is a  $(\mathbb{P}, \mathcal{F}_t)$ -martingale. Denoting for all  $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{N}^d$ ,  $h(t, \mathbf{x}) := e^{-\tilde{\rho}t} \mathbf{x} \cdot \tilde{\boldsymbol{\xi}}$  and using the fact that by definition  $\tilde{\mathbf{C}} \tilde{\boldsymbol{\xi}}^t = \rho \tilde{\boldsymbol{\xi}}^t$ , we see that for all  $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{N}^d$ ,

$$(\tilde{L}h)(t, \mathbf{x}) = e^{-\tilde{\rho}t} \left[ -\tilde{\rho} \mathbf{x} \cdot \tilde{\boldsymbol{\xi}} + \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \tilde{p}_i(\mathbf{k}) (\mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}} \right] = e^{-\tilde{\rho}t} \left[ -\tilde{\rho} \mathbf{x} \cdot \tilde{\boldsymbol{\xi}} + \sum_{i=1}^d x_i \sum_{j=1}^d \tilde{c}_{ij} \tilde{\xi}_j \right] = 0,$$

which proves that  $(e^{-\tilde{\rho}t} \mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}})_{t \geq 0}$  is a  $(\tilde{\mathbb{P}}, \mathcal{F}_t)$ -martingale. This together with the fact that  $(\mathbf{q}^{\mathbf{X}_t})_{t \geq 0}$  is a  $(\mathbb{P}, \mathcal{F}_t)$ -martingale then leads to the result, and implies that  $\mathbb{P}^*$  is a Doob  $h$ -transform of  $\mathbb{P}$ .  $\square$

### 1.3. The conditioned paths and the immortal particle

As already mentioned, it has been proved in [11] that a branching process conditioned to die out remains a branching process, and that in the supercritical case the conditioned process becomes subcritical. Conditioning on extinction in the distant future thus influences the life careers of the particles (and more precisely modifies the offspring distribution) but it preserves the branching property. We will see that the same does not occur when conditioning on very late extinction, and the purpose of this section is to describe the structure of the conditioned process obtained in Theorem 1.1.

One can easily verify that the branching property is not preserved for the process with law  $\mathbb{P}^*$ . For  $d = 1$  and  $\rho = 0$  we have for instance  $\mathbb{E}_x^*(X_t) = x^{-1} \mathbb{E}_x(X_t^2) = x + \alpha \sigma^2 t$ , where  $\sigma^2$  is the variance of the offspring distribution. We thus obviously have  $\mathbb{E}_x^*(X_t) \neq x \mathbb{E}_1^*(X_t)$  for  $x \neq 1$ . Nevertheless we will see that the branching structure is somehow preserved: the conditioned process with law  $\mathbb{P}^*$  behaves like an unconditioned (sub)critical branching process to which an external structure is added, forcing the process to die out very late.

In the monotype case  $d = 1$ , this external input is a standard immigration. It has indeed been shown in [14] that a critical BGW process conditioned on very late extinction and from which one removes one particle has the same law as a branching process with immigration. This result could be generalized to one-dimensional continuous-time branching processes of any class of criticality. One obtains that under the assumptions of Theorem 1.1, the law of the process with law  $\mathbb{P}^*$  shifted downwards by 1 is the law of a branching process with immigration: the branching process has branching rate  $\alpha$  and offspring generating function  $\tilde{f}$  defined in (1.9), while the immigration is given by  $\alpha(\tilde{f}' - \tilde{m})$ .

In the multitype case, however, such an interpretation with a classical branching process with immigration is not possible, and we will prove in Proposition 1.5 that the external input comes from an *immortal particle* (so called in reference to [5]). In the monotype case this is clear thanks to the previous interpretation: the conditioned process can be seen as the independent sum of an unconditioned branching process and of an immortal particle, which produces offspring according to the distribution of the immigration, and corresponds to the removed particle mentioned in the previous paragraph. The generalization to the multitype case then

differs by the fact that the immortal particle can mutate from one type to another. More precisely, the behavior of the immortal particle is the one of the trunk of a size-biased multitype Galton-Watson tree in its continuous-time version, introduced in [8].

Let us introduce the following *size-biased* offspring distribution  $(s_i(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$  with respect to the offspring distribution  $(\tilde{p}_i(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$  defined in (1.10). For all  $i = 1 \dots d$ ,

$$s_i(\mathbf{k}) := \frac{\alpha_i}{(\alpha_i + \tilde{\rho})\tilde{\xi}_i} \mathbf{k} \cdot \tilde{\boldsymbol{\xi}} \tilde{p}_i(\mathbf{k}), \quad \mathbf{k} \in \mathbb{N}^d. \tag{1.15}$$

We easily check that for all  $i = 1 \dots d$ ,  $(s_i(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$  is a probability distribution:

$$\sum_{\mathbf{k} \in \mathbb{N}^d} s_i(\mathbf{k}) = \frac{1}{(\alpha_i + \tilde{\rho})\tilde{\xi}_i} \sum_{j=1}^d \alpha_i \tilde{m}_{ij} \tilde{\xi}_j = \frac{1}{(\alpha_i + \tilde{\rho})\tilde{\xi}_i} \left[ \sum_{j=1}^d \tilde{c}_{ij} \tilde{\xi}_j + \alpha_i \tilde{\xi}_i \right] = 1,$$

and that it is concentrated on  $\mathbb{N}^d \setminus \{\mathbf{0}\}$ .

Let us now describe in detail the structure of the conditioned process with law  $\mathbb{P}^*$ .

**Proposition 1.5.** *Under the assumptions of Theorem 1.1,  $\mathbb{P}^*$  is the law of the independent sum of a (sub)critical branching process and of an “immortal particle”. The branching process has branching rates  $\alpha_1, \dots, \alpha_d$  and offspring generating function  $\tilde{\mathbf{f}}$ . Given that the immortal particle is of type  $i$ , it has an exponential life-time of parameter  $\alpha_i + \tilde{\rho}$  and an offspring distribution  $(s_i(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$ . Its initial type is  $i$  with probability  $x_i \tilde{\xi}_i / \mathbf{x} \cdot \tilde{\boldsymbol{\xi}}$ , where  $\mathbf{x} \in \mathbb{N}^d$  is the initial number of particles, and if it produces  $\mathbf{k}$  offspring it mutates to type  $j$  with probability  $k_j \tilde{\xi}_j / \mathbf{k} \cdot \tilde{\boldsymbol{\xi}}$ .*

*Proof.* Let us compute the infinitesimal generator  $L^*$  of the process with law  $\mathbb{P}^*$ . According to Theorem 1.1, denoting  $\tilde{h}(t, \mathbf{x}) = e^{-\tilde{\rho}t} \mathbf{x} \cdot \tilde{\boldsymbol{\xi}}$ , we have for all smooth function  $f : \mathbb{R}^+ \times \mathbb{N}^d \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$

$$L^* f := \frac{1}{\tilde{h}} \tilde{L}(\tilde{h}f),$$

and thus

$$(L^* f)(t, \mathbf{x}) = \frac{\partial f}{\partial t}(t, \mathbf{x}) - \tilde{\rho} f(t, \mathbf{x}) + \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \tilde{p}_i(\mathbf{k}) \left[ (\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}} f(t, \mathbf{x} + \mathbf{k} - \mathbf{e}_i) - \mathbf{x} \cdot \tilde{\boldsymbol{\xi}} f(t, \mathbf{x}) \right]. \tag{1.16}$$

Using the definition of  $\tilde{\rho}$  and  $\tilde{\boldsymbol{\xi}}$  we obtain that

$$\tilde{\rho} = \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \tilde{p}_i(\mathbf{k}) (\mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}},$$

hence (1.16) becomes

$$(L^* f)(t, \mathbf{x}) = \frac{\partial f}{\partial t}(t, \mathbf{x}) + \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \tilde{p}_i(\mathbf{k}) \frac{(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}}}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} [f(t, \mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(t, \mathbf{x})]. \tag{1.17}$$

On the other hand, denoting by  $\overline{\mathbb{P}}$  the law of the time-homogeneous Markov process described in Proposition 1.5, its infinitesimal generator  $\overline{L}$  is by definition

$$(\overline{L}f)(t, \mathbf{x}) := \lim_{h \rightarrow 0} \frac{1}{h} \overline{\mathbb{E}}_{\mathbf{x}} [f(t+h, \mathbf{X}_h) - f(t, \mathbf{x})].$$



Let  $h > 0$ . It appears that during the time-interval  $[0, h]$ , the only non-trivial events whose probabilities are not dominated by  $h$  as  $h \rightarrow 0$  are the ones consisting of exactly one branching event, either of the unconditioned branching process or of the immortal particle. The first possibility is that during  $[0, h]$ , the immortal particle of type  $i$  splits into  $\mathbf{k}$  offspring, with no other event occurring. The probability of this event is then

$$\frac{x_i \tilde{\xi}_i}{\mathbf{x} \cdot \tilde{\xi}} (\alpha_i + \tilde{\rho}) s_i(\mathbf{k}) h = \alpha_i x_i \frac{\mathbf{k} \cdot \tilde{\xi}}{\mathbf{x} \cdot \tilde{\xi}} \tilde{p}_i(\mathbf{k}) h.$$

The second possible type of event appearing in the computation of  $\bar{L}$  is that one particle of type  $i$  in the unconditioned process splits into  $\mathbf{k}$  offspring, while the immortal particle is of type  $j$  and no other event occurs. The probability of this event is then

$$\frac{x_j \tilde{\xi}_j}{\mathbf{x} \cdot \tilde{\xi}} \alpha_i (x_i - \delta_{ij}) \tilde{p}_i(\mathbf{k}) h.$$

The infinitesimal generator  $\bar{L}$  is thus given by

$$\begin{aligned} (\bar{L}f)(t, \mathbf{x}) &= \frac{\partial f}{\partial t}(t, \mathbf{x}) + \sum_{i=1}^d \sum_{\mathbf{k} \in \mathbb{N}^d} \alpha_i x_i \frac{\mathbf{k} \cdot \tilde{\xi}}{\mathbf{x} \cdot \tilde{\xi}} \tilde{p}_i(\mathbf{k}) [f(t, \mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(t, \mathbf{x})] \\ &\quad + \sum_{i=1}^d \sum_{j=1}^d \sum_{\mathbf{k} \in \mathbb{N}^d} \frac{x_j \tilde{\xi}_j}{\mathbf{x} \cdot \tilde{\xi}} \alpha_i (x_i - \delta_{ij}) \tilde{p}_i(\mathbf{k}) [f(t, \mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(t, \mathbf{x})], \end{aligned}$$

and we observe that  $\bar{L} = L^*$ . □

**Remark 1.6.** In the critical case we can draw an analogy between the (state-dependent) offspring distribution of the conditioned process with law  $\mathbb{P}^*$ , and the transition probabilities of its discrete-time analog studied in [3], the so-called  $Q$ -process. In the critical case the infinitesimal generator  $L^*$  becomes indeed

$$(L^*f)(\mathbf{x}) = \sum_{i=1}^d \alpha_i x_i \sum_{\mathbf{k} \in \mathbb{N}^d} p_i(\mathbf{k}) \frac{(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \xi}{\mathbf{x} \cdot \xi} [f(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) - f(\mathbf{x})],$$

and we see that in this case, for every  $i = 1 \dots d$  and  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,  $(p_i^*(\mathbf{x}, \mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$  is a probability distribution, where

$$p_i^*(\mathbf{x}, \mathbf{k}) := \frac{(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \xi}{\mathbf{x} \cdot \xi} p_i(\mathbf{k}). \tag{1.18}$$

The analogy between (1.18) and the formula (1.19) given in [3] for the  $Q$ -process, providing the relation between the transition probabilities  $P^*(\mathbf{i}, \mathbf{j})$  (resp.  $P(\mathbf{i}, \mathbf{j})$ ) of the conditioned process (resp. unconditioned process)

$$P^*(\mathbf{i}, \mathbf{j}) = \frac{\mathbf{j} \cdot \mathbf{u}}{\mathbf{i} \cdot \mathbf{u}} P(\mathbf{i}, \mathbf{j}), \quad \mathbf{i}, \mathbf{j} \in \mathbb{N}^d, \mathbf{i} \neq \mathbf{0}, \tag{1.19}$$

is then obvious (here  $\mathbf{u}$  denotes the normalized right eigenvector of the mean matrix  $\mathbf{M}$  for its Perron's root 1).

### 1.4. Long-time behavior of the conditioned branching process

We are now interested in the long-time behavior of the discrete-state branching process conditioned on very late extinction introduced in Section 1.2. We show in Proposition 1.7 that the long-time limit is non-degenerate in the subcritical and supercritical cases, and that it is a probability distribution independent of the initial condition. By definition of  $\mathbb{P}^*$ , this limit is obtained by letting first  $\theta$  and then  $t$  tend to infinity in the law of  $\mathbf{X}_t$  conditioned on extinction in the distant future and on  $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ . It is thus a natural question to ask whether

the order of those two limits in  $t$  and  $\theta$  can be exchanged. Proposition 1.7 provides an affirmative answer: the non-degenerate limit mentioned above can also be obtained by letting first  $t$  and then  $\theta$  tend to infinity, *i.e.* as the limit as  $\theta$  tends to infinity of the asymptotic law of the process  $\mathbf{X}_t$  conditioned on extinction in the distant future and on  $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ .

Furthermore, we will relate in (1.26) the asymptotic law when  $t \rightarrow \infty$  of the process  $\mathbf{X}_t$  conditioned on extinction in the distant future and on  $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ ,  $\theta$  fixed, to the well-known Yaglom distribution, which is the exact equivalent asymptotic law obtained for  $\theta = 0$ .

**Proposition 1.7.** *Under the assumptions of Theorem 1.1, the following holds for all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,*

$$\lim_{t \rightarrow \infty} \lim_{\theta \rightarrow \infty} \mathbb{P}_{\mathbf{x}} \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right) = \lim_{\theta \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{P}_{\mathbf{x}} \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right). \tag{1.20}$$

Furthermore, if  $\rho \neq 0$  then this limit is non-degenerate and is a probability distribution which does not depend on  $\mathbf{x}$ .

If  $\rho = 0$  however, this limit is degenerate and for all  $\mathbf{u} \geq \mathbf{0}$  we have

$$\lim_{t \rightarrow \infty} \lim_{\theta \rightarrow \infty} \mathbb{P}_{\mathbf{x}} (\mathbf{X}_t > \mathbf{u} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}) = \lim_{\theta \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{P}_{\mathbf{x}} (\mathbf{X}_t > \mathbf{u} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}) = 1. \tag{1.21}$$

*Proof.* We first assume that  $\rho \neq 0$  and focus on the right term of (1.20). We denote by  $F^0$  the generating function of the Yaglom limit of the subcritical process with law  $\tilde{\mathbb{P}}$ . We thus have, for all  $\mathbf{r} \in [0, 1]^d$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$F^0(\mathbf{r}) = \lim_{t \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} \mid \mathbf{X}_t \neq \mathbf{0}]. \tag{1.22}$$

By means of Theorem 2 in [13] and its extension to the continuous-time case *via* the embedded process in Theorem 6.1 in [18], we obtain that there exists a non-negative real function  $\gamma$  on  $[0, 1]^d$  such that

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}t} (\mathbf{1} - \tilde{\mathbf{F}}_t(\mathbf{r})) = \gamma(\mathbf{r})\tilde{\boldsymbol{\xi}}, \tag{1.23}$$

where  $\tilde{\mathbf{F}}_t$  is the generating function at time  $t$  of the subcritical process with law  $\tilde{\mathbb{P}}$ ,  $\tilde{F}_{t,i}(\mathbf{r}) := \tilde{\mathbb{E}}_{\mathbf{e}_i}(\mathbf{r}^{\mathbf{X}_t})$ . It follows from the Markov and branching properties of  $\tilde{\mathbb{P}}$  that for all  $\theta \geq 0$ ,  $\mathbf{r} \in [0, 1]^d$  and  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$\tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \frac{\tilde{\mathbf{F}}_t(\mathbf{r})^{\mathbf{x}} - \tilde{\mathbf{F}}_t(\mathbf{r}\tilde{\mathbf{q}}(\theta))^{\mathbf{x}}}{1 - \tilde{\mathbf{F}}_{t+\theta}(\mathbf{0})^{\mathbf{x}}},$$

from which it ensues together with (1.23) that

$$\lim_{t \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \frac{\gamma(\mathbf{r}\tilde{\mathbf{q}}(\theta)) - \gamma(\mathbf{r})}{e^{\tilde{\rho}\theta}\gamma(\mathbf{0})}. \tag{1.24}$$

Now for  $\theta = 0$  this relation becomes

$$F^0(\mathbf{r}) = 1 - \frac{\gamma(\mathbf{r})}{\gamma(\mathbf{0})}, \tag{1.25}$$

and from the properties of the generating function  $F^0$  it follows that the function defined by (1.24) is a probability generating function as well. Moreover, denoting by  $F^\theta$  this generating function, we obtain thanks to (1.24) the announced relation between the Yaglom distribution and its generalization to  $\theta > 0$ :

$$F^\theta(\mathbf{r}) = e^{-\tilde{\rho}\theta} [F^0(\mathbf{r}) - F^0(\mathbf{r}\tilde{\mathbf{q}}(\theta))]. \tag{1.26}$$

We now have all the tools in hand to study the right term of (1.20). By (1.23) and (1.24) we have indeed, for all  $\mathbf{r} \in [0, 1]^d$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$\lim_{\theta \rightarrow \infty} \lim_{t \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \lim_{\theta \rightarrow \infty} \frac{\gamma(\mathbf{r} - e^{\tilde{\rho}\theta} \gamma(\mathbf{0}) \tilde{\boldsymbol{\xi}}) - \gamma(\mathbf{r})}{e^{\tilde{\rho}\theta} \gamma(\mathbf{0})} = - \sum_{i=1}^d r_i \tilde{\xi}_i \frac{\partial \gamma(\mathbf{r})}{\partial r_i}, \tag{1.27}$$

the differentiability of  $\gamma$  stemming from (1.25).

Let us now focus on the left term of (1.20). Using (1.14) we obtain

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}] &= e^{-\tilde{\rho}t} \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}} \mathbf{r}^{\mathbf{X}_t}] = e^{-\tilde{\rho}t} \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \sum_{i=1}^d r_i \tilde{\xi}_i \frac{\partial}{\partial r_i} [\tilde{\mathbf{F}}_t(\mathbf{r})^{\mathbf{x}}] \\ &= e^{-\tilde{\rho}t} \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \sum_{i=1}^d r_i \tilde{\xi}_i \sum_{j=1}^d x_j \frac{\partial}{\partial r_i} [\ln \tilde{F}_{t,j}(\mathbf{r})] \tilde{\mathbf{F}}_t(\mathbf{r})^{\mathbf{x}}. \end{aligned} \tag{1.28}$$

But for all  $i, j = 1 \dots d$  and all  $\mathbf{r} \in [0, 1]^d$  such that  $r_i > 0$  we have (see Lemma 1.8 below)

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} [\ln \tilde{F}_{t,j}(\mathbf{r})] = - \frac{\partial \gamma(\mathbf{r})}{\partial r_i} \tilde{\xi}_j,$$

which together with (1.27) and (1.28) leads to the equality

$$\lim_{\theta \rightarrow \infty} \lim_{t \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \lim_{t \rightarrow \infty} \lim_{\theta \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}],$$

for all  $\mathbf{r} \in [0, 1]^d$ . Now from [21] Satz 6.2.8 we know that under the assumptions (ii), (iii) of Theorem 1.1, the generating function  $F^0$  is differentiable in  $\mathbf{r} = \mathbf{1}$ , and its derivative satisfies, for all  $i = 1 \dots d$ ,

$$\frac{\partial F^0(\mathbf{1})}{\partial r_i} = \frac{\tilde{\eta}_i}{\gamma(\mathbf{0})}, \tag{1.29}$$

where  $\tilde{\boldsymbol{\eta}}$  is the left eigenvector of  $\tilde{\mathbf{M}}$  for  $\tilde{\rho}$ , with  $\tilde{\boldsymbol{\eta}} \cdot \tilde{\boldsymbol{\xi}} = 1$ . Using Lebesgue's dominated convergence theorem together with the fact that (1.29) is finite, it comes that  $\frac{\partial F^0(\mathbf{r})}{\partial r_i}$  is continuous in  $\mathbf{1}$ , which thanks to (1.25) implies the continuity in  $\mathbf{r} = \mathbf{1}$  of the right term of (1.27). The second assertion of Proposition 1.7 then ensues from Lévy's continuity theorem.

We now consider the critical case  $\rho = 0$ . It is known (Satz 6.3.5 in [21]) that in the critical case the random vector  $(\frac{2X_{t,1}}{\zeta \eta_1 t}, \dots, \frac{2X_{t,d}}{\zeta \eta_d t})$  converges conditionally on  $\mathbf{X}_t \neq \mathbf{0}$  to a random vector  $\mathbf{Y}$  independent of  $\mathbf{x}$ , with coordinates  $Y_1 = \dots = Y_d$  almost surely, and such that each  $Y_i$  is exponentially distributed with parameter 1. We can easily show that the same holds when conditioning on  $\mathbf{X}_{t+\theta} \neq \mathbf{0}$ , and that the limiting vector  $\mathbf{Y}$  depends neither on  $\mathbf{x}$  nor on  $\theta$  (this comes intuitively from the fact that rescaling the process by  $t$  or by  $t + \theta$  does not make a difference any more once  $t$  tends to  $\infty$ ). An immediate consequence is that for any  $\theta \geq 0$ ,  $\mathbf{X}_t$  explodes conditionally on  $\mathbf{X}_{t+\theta} \neq \mathbf{0}$  when  $t \rightarrow \infty$ . Hence for all  $\theta \geq 0$ ,  $\mathbf{r} \in C_d$ ,  $\mathbf{r} \neq \mathbf{1}$ ,  $\lim_{t \rightarrow \infty} \mathbb{E}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}] = 0$ . On the other hand,  $\lim_{\theta \rightarrow \infty} \mathbb{E}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}] = \frac{1}{\mathbf{x} \cdot \boldsymbol{\xi}} \mathbb{E}_{\mathbf{x}} [\mathbf{X}_t \cdot \boldsymbol{\xi} \mathbf{r}^{\mathbf{X}_t}] \leq \frac{1}{\mathbf{x} \cdot \boldsymbol{\xi}} \mathbb{E}_{\mathbf{x}} [\mathbf{X}_t \cdot \boldsymbol{\xi}]$ , and it ensues from the asymptotic behavior of the mean matrix in the critical case that  $\lim_{t \rightarrow \infty} \lim_{\theta \rightarrow \infty} \mathbb{E}_{\mathbf{x}} [\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_{t+\theta} \neq \mathbf{0}] = 0$ .  $\square$

We finally provide the following technical lemma, needed in the proof of Proposition 1.7.

**Lemma 1.8.** *For all  $i, j = 1 \dots d$  and all  $\mathbf{r} \in [0, 1]^d$ ,  $\mathbf{r} \neq \mathbf{1}$ , such that  $r_i > 0$ ,*

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} [\ln \tilde{F}_{t,j}(\mathbf{r})] = - \frac{\partial \gamma(\mathbf{r})}{\partial r_i} \tilde{\xi}_j. \tag{1.30}$$

*Proof.* Deducing from (1.23) that  $\lim_{t \rightarrow \infty} e^{-\tilde{\rho}t} \ln \tilde{F}_{t,j}(\mathbf{r}) = -\gamma(\mathbf{r})\tilde{\xi}_j$ , we obtain that for all  $h \geq 0$ ,

$$\gamma(\mathbf{r})\tilde{\xi}_j - \gamma(\mathbf{r} + h\mathbf{e}_i)\tilde{\xi}_j = \lim_{t \rightarrow \infty} \int_0^h e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}(\mathbf{r} + u\mathbf{e}_i) \right] du. \tag{1.31}$$

Moreover,

$$0 \leq e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}(\mathbf{r}) \right] \leq e^{-\tilde{\rho}t} \frac{1}{\tilde{F}_{t,j}(\mathbf{r})} \frac{1}{r_i} \tilde{\mathbb{E}}_{\mathbf{e}_j} [X_{t,i}],$$

and as a consequence of the Perron-Frobenius theorem we know that  $\lim_{t \rightarrow \infty} e^{-\tilde{\rho}t} \tilde{\mathbb{E}}_{\mathbf{e}_j} [X_{t,i}] = \tilde{\xi}_j \tilde{\eta}_i$  (see [21]). We then have by the continuity of  $\tilde{F}_{t,j}$  in  $\mathbf{r}$  the existence of a constant  $C > 0$  such that for all  $t \geq 0$  and all  $u \in [0, h]$ ,

$$\left| e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}(\mathbf{r} + u\mathbf{e}_i) \right] \right| \leq \frac{C}{r_i + u}.$$

Using this upper bound integrable on  $u \in [0, h]$  together with Lebesgue’s dominated convergence theorem, (1.31) leads to

$$\gamma(\mathbf{r})\tilde{\xi}_j - \gamma(\mathbf{r} + h\mathbf{e}_i)\tilde{\xi}_j = \int_0^h \limsup_{t \rightarrow \infty} \left[ e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} \left( \ln \tilde{F}_{t,j}(\mathbf{r} + u\mathbf{e}_i) \right) \right] du,$$

and thus  $\limsup_{t \rightarrow \infty} e^{-\tilde{\rho}t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}(\mathbf{r}) \right] = -\frac{\partial \gamma(\mathbf{r})}{\partial r_i} \tilde{\xi}_j$ . Proving the same way the result for the limit inferior we finally obtain (1.30).  $\square$

## 2. THE CONDITIONED MULTITYPE FELLER DIFFUSION PROCESS AS A LIMIT OF CONDITIONED BRANCHING PROCESSES

We are now interested in the continuous-state analog of the conditioned process studied in Section 1, as well as the possible relation between continuous-state and discrete-state branching processes conditioned on very late extinction. Monotype Feller diffusion processes conditioned to be never extinct have been studied in [19] in the broader context of conditioned Dawson-Watanabe processes, and the multitype case has been introduced in [2]. The structure of multitype conditioned Feller diffusion processes is thus well-known, and it is shown in [2] that an irreducible (sub)critical multitype Feller diffusion process conditioned on non-extinction is a Doob  $h$ -transform of the unconditioned process, *via* the harmonic function  $e^{-\rho t} \mathbf{x} \cdot \boldsymbol{\xi}$ , where  $\rho$  (resp.  $\boldsymbol{\xi}$ ) is the Perron’s root (resp. the right normalized eigenvector for  $\rho$ ) of the irreducible mutation matrix of the diffusion process. We generalize in Proposition 2.1 this result to multitype Feller diffusion processes of any class of criticality, in order to obtain the continuous-state equivalent of Theorem 1.1.

In this section we present the conditioned Feller diffusion process as the solution to a limit martingale problem. It is well known that a Feller diffusion process can be obtained as the scaling limit of discrete-state branching processes (see *e.g.* Thm. 4.4.2 in [12] for the multitype case), and we show in Theorem 2.2 that an appropriate approximation also holds for a *conditioned* Feller diffusion process. According to the intuition, the approximating discrete-state processes are branching processes conditioned on very late extinction. As a corollary we conclude that “rescaling and conditioning commute”, or in other words that “the diffusion limit of conditioned branching processes is the conditioned diffusion limit”.

Our fundamental tools are martingale problems. We use the following notation. For a given infinitesimal generator  $G$  with domain  $D(G)$  and a given subset  $D_0(G) \subseteq D(G)$ , we say that  $\mathbf{P}$  is a solution to the martingale problem  $\mathcal{MP}(G, D_0(G))$  (or  $\mathcal{MP}(G)$  to avoid heavy notation) if for all  $f \in D_0(G)$ ,

$$f(t, \mathbf{X}_t) - f(0, \mathbf{X}_0) - \int_0^t (Gf)(s, \mathbf{X}_{s-}) ds \text{ is a } (\mathbf{P}, \mathcal{F}_t)\text{-martingale.}$$

When imposing the initial condition  $\mathbf{x}_0$  we write  $\mathcal{MP}(G, D_0(G), \mathbf{x}_0)$ .

Denoting by  $L$  the infinitesimal generator of the unconditioned Feller diffusion process, we show in Proposition 2.1 that the conditioned Feller diffusion process is the unique solution to the martingale problem  $\mathcal{MP}(\frac{1}{h}L(h))$ , where  $h$  is a space-time harmonic function for  $L$ . Denoting by  $L^n$  the infinitesimal generator of the rescaled discrete-state branching process (in a sense that will be detailed later), we know from Theorem 1.1 that the corresponding conditioned process is a solution to the martingale problem  $\mathcal{MP}(\frac{1}{h^n}L^n(h^n))$ , where  $h^n$  is an appropriate space-time harmonic function for  $L^n$ . On the other side it is known that any limit of solutions to the martingale problems  $\mathcal{MP}(L^n)$  is a solution to the martingale problem  $\mathcal{MP}(L)$ . Our aim is now to prove that any limit of solutions to  $\mathcal{MP}(\frac{1}{h^n}L^n(h^n))$  is a solution to  $\mathcal{MP}(\frac{1}{h}L(h))$ . The result is illustrated in the following commutative diagram, where  $\dashrightarrow$  stands for the transform by conditioning on very late extinction, and  $\Rightarrow$  for the weak convergence of probability measures when the scaling parameter  $n$  tends to infinity:

$$\begin{array}{ccc} \mathcal{MP}(L^n) & \dashrightarrow & \mathcal{MP}(\frac{1}{h^n}L^n(h^n)) \\ \Downarrow & & \Downarrow \\ \mathcal{MP}(L) & \dashrightarrow & \mathcal{MP}(\frac{1}{h}L(h)). \end{array} \tag{2.1}$$

### 2.1. The Feller diffusion process conditioned on very late extinction

Let us present a first statement on Feller diffusion processes conditioned on very late extinction. Our result is a generalization of Theorem 2.2 in [2] to processes of any class of criticality. It is straightforward by using the fact that by conditioning a supercritical processes on extinction, one recovers a subcritical process. For this reason we omit the proof of Proposition 2.1.

We consider a  $d$ -type Feller diffusion process with sample paths in  $D(\mathbb{R}^+, \mathbb{R}_+^d)$ , and we denote by  $\mathbb{P}$  its law on  $(\Omega, (\mathbf{X}_t)_{t \geq 0}, (\mathcal{F}_t)_{t \geq 0})$ . Let  $\mathbf{C} \in \mathcal{M}_d(\mathbb{R})$  be the *mutation matrix* of the process, with non-negative non-diagonal elements. We denote by  $\sigma_1^2, \dots, \sigma_d^2$  the variance parameters of the process, and assume that  $\inf_i \sigma_i^2 > 0$ . The infinitesimal generator is then given on  $D(L) := C^2(\mathbb{R}_+^d, \mathbb{R})$  by (see e.g. [4], Sect. 8.1),

$$(Lf)(\mathbf{x}) := \frac{1}{2} \sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\mathbf{x}) + \sum_{i=1}^d \sum_{j=1}^d c_{ji} x_j \frac{\partial f}{\partial x_i}(\mathbf{x}), \tag{2.2}$$

(to avoid heavy notation we only give the generator for time-independent functions). We denote  $D_0(L) := C_b^2(\mathbb{R}_+^d, \mathbb{R})$  the set of bounded  $C^2$ -functions on  $\mathbb{R}_+^d$ . Then  $\mathbb{P}$  is the unique solution to the martingale problem  $\mathcal{MP}(L, D_0(L))$  (see for example [4], Sect. 8.1, Thm. 1.7).

Assuming that the mutation matrix  $\mathbf{C}$  is irreducible, and denoting by  $\rho$  its maximal eigenvalue, we call the process *subcritical*, *critical* or *supercritical* according as  $\rho < 0$ ,  $\rho = 0$  or  $\rho > 0$ . In the following, we denote by  $\boldsymbol{\xi}$  (resp.  $\boldsymbol{\eta}$ ) the right (resp. left) eigenvector for  $\rho$  with normalization  $\boldsymbol{\xi} \cdot \mathbf{1} = 1$ ,  $\boldsymbol{\eta} \cdot \boldsymbol{\xi} = 1$ .

We denote by  $\mathbf{u}_t(\boldsymbol{\lambda})$  the cumulant of the process, such that for all  $t \geq 0$ ,  $\boldsymbol{\lambda} \in \mathbb{R}_+^d$  and  $\mathbf{x} \in \mathbb{R}_+^d$ ,

$$\mathbb{E}_{\mathbf{x}}(e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_t}) = e^{-\mathbf{x} \cdot \mathbf{u}_t(\boldsymbol{\lambda})}.$$

Then the probability of extinction at time  $t$  (resp. probability of extinction) is given by the vector  $\mathbf{u}_t := \lim_{\boldsymbol{\lambda} \rightarrow \infty} \mathbf{u}_t(\boldsymbol{\lambda})$  (resp.  $\mathbf{u} := \lim_{t \rightarrow \infty} \mathbf{u}_t$ ). We have indeed, for all  $i = 1 \dots d$ ,

$$\mathbb{P}_{\mathbf{e}_i}(\mathbf{X}_t = \mathbf{0}) = e^{-\mathbf{x} \cdot \mathbf{u}_t},$$

and

$$\mathbb{P}_{\mathbf{e}_i}(\lim_{t \rightarrow \infty} \mathbf{X}_t = \mathbf{0}) = e^{-\mathbf{x} \cdot \mathbf{u}}.$$

Let us now condition the process on very late extinction. Assuming from now on that the Feller diffusion process has a *positive risk of extinction* (i.e.  $\mathbf{u} < \infty$ ), we introduce as in Section 1.2 the conditioned law

$$\tilde{\mathbb{P}}(\cdot) := \mathbb{P}\left(\cdot \mid \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0}\right), \tag{2.3}$$

which is well-defined under the previous assumption. Similar to Section 1.2, we can prove that in the supercritical case, the law  $\tilde{\mathbb{P}}$  is a Doob  $h$ -transform of the unconditioned law, satisfying for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{R}_+^d$ ,

$$d\tilde{\mathbb{P}}_{\mathbf{x}}|_{\mathcal{F}_t} = \frac{e^{-\mathbf{X}_t \cdot \mathbf{u}}}{e^{-\mathbf{x} \cdot \mathbf{u}}} d\mathbb{P}_{\mathbf{x}}|_{\mathcal{F}_t}. \tag{2.4}$$

Computing the infinitesimal generator of the process with law  $\tilde{\mathbb{P}}$ , we obtain that it is a Feller diffusion process with rates  $\alpha_1, \dots, \alpha_d$ , variance parameters  $\sigma_1^2, \dots, \sigma_d^2$ , and irreducible mutation matrix  $\tilde{\mathbf{C}}$  with entries  $\tilde{c}_{ij} := c_{ij} - \sigma_i^2 u_i \delta_{ij}$ . In the following we denote by  $\tilde{\rho}, \tilde{\boldsymbol{\xi}}$  and  $\tilde{\boldsymbol{\eta}}$  its Perron's root and associated right and left eigenvectors with the usual normalization convention. We then show by dominated convergence that  $\lim_{t \rightarrow \infty} \tilde{\mathbb{E}}_{\mathbf{e}_i}(X_{t,j}) = 0$ , which proves that the process is subcritical.

We finally introduce the law of the process conditioned on very late extinction. For all  $t \geq 0$  and all  $B \in \mathcal{F}_t$ ,

$$\mathbb{P}^*(B) := \lim_{\theta \rightarrow \infty} \tilde{\mathbb{P}}(B \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}), \tag{2.5}$$

provided this limit exists.

Combining (2.4) with Theorem 2.2 in [2], we come to the following statement.

**Proposition 2.1.** *Let  $\mathbb{P}$  be the law of an irreducible Feller diffusion process with positive risk of extinction. Then  $\mathbb{P}^*$  is a Doob  $h$ -transform of  $\mathbb{P}$  satisfying for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{R}_+^d, \mathbf{x} \neq \mathbf{0}$ ,*

$$d\mathbb{P}_{\mathbf{x}}^*|_{\mathcal{F}_t} = e^{-\tilde{\rho}t} \frac{e^{-\mathbf{X}_t \cdot \mathbf{u}}}{e^{-\mathbf{x} \cdot \mathbf{u}}} \frac{\mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}}}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} d\mathbb{P}_{\mathbf{x}}|_{\mathcal{F}_t}, \tag{2.6}$$

where  $\tilde{\rho}$  is the Perron's root of the irreducible matrix  $\tilde{\mathbf{C}}$  with entries

$$\tilde{c}_{ij} := c_{ij} - \sigma_i^2 u_i \delta_{ij}, \tag{2.7}$$

and  $\tilde{\boldsymbol{\xi}}$  is the associated right eigenvector with norm  $\tilde{\boldsymbol{\xi}} \cdot \mathbf{1} = 1$ .

In particular, if the process is (sub)critical, then  $\mathbb{P}^*$  satisfies for all  $t \geq 0$  and all  $\mathbf{x} \in \mathbb{N}^d, \mathbf{x} \neq \mathbf{0}$ ,

$$d\mathbb{P}_{\mathbf{x}}^*|_{\mathcal{F}_t} = e^{-\rho t} \frac{\mathbf{X}_t \cdot \boldsymbol{\xi}}{\mathbf{x} \cdot \boldsymbol{\xi}} d\mathbb{P}_{\mathbf{x}}|_{\mathcal{F}_t}. \tag{2.8}$$

It is simple to deduce from (2.6) that the infinitesimal generator  $L^*$  of the conditioned process with law  $\mathbb{P}^*$  is given on  $D(L^*) := C^2(\mathbb{R}_+^d \setminus \{\mathbf{0}\}, \mathbb{R})$  by

$$(L^* f)(\mathbf{x}) := \frac{1}{2} \sum_{i=1}^d \sigma_i^2 x_i \frac{\partial^2 f}{\partial x_i^2}(\mathbf{x}) + \sum_{i=1}^d \left( \sum_{j=1}^d \tilde{c}_{ji} x_j + \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \sigma_i^2 x_i \tilde{\xi}_i \right) \frac{\partial f}{\partial x_i}(\mathbf{x}). \tag{2.9}$$

The conditioned Feller diffusion process can thus be considered as a Feller diffusion with variance parameters  $\sigma_1^2, \dots, \sigma_d^2$  and state-dependent mutation matrix  $\mathbf{C}(\mathbf{x})$  defined by

$$c_{ij}(\mathbf{x}) := \tilde{c}_{ij} + \frac{\sigma_i^2 \tilde{\xi}_i}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \delta_{ij}. \tag{2.10}$$

2.2. Limit of conditioned branching processes

We now approximate the conditioned Feller diffusion process with law  $\mathbb{P}^*$  by discrete-state processes. Those processes are rescaled branching processes conditioned on very late extinction. For every  $n \in \mathbb{N}^*$  we consider  $\mathbb{P}^n$  the law of a continuous-time branching process with offspring distribution  $(\mathbf{p}^n(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$ , branching rates  $\alpha_i = n$ ,  $i = 1 \dots d$ , and rescaled by  $\frac{1}{n}$ . The process with law  $\mathbb{P}^n$  consequently takes its values in  $\frac{1}{n}\mathbb{N}^d$ . We denote by  $\mathbf{M}^n$  its mean matrix, and whenever  $\mathbf{M}^n$  is finite and irreducible, we denote by  $\rho^n$  the Perron's root of the matrix  $\mathbf{C}^n := n(\mathbf{M}^n - \mathbf{I})$ . Finally, we denote by  $\boldsymbol{\xi}^n$  and  $\boldsymbol{\eta}^n$  the right and left eigenvectors of  $\mathbf{C}^n$  for the eigenvalue  $\rho^n$ , with the usual normalization convention  $\boldsymbol{\xi}^n \cdot \mathbf{1} = 1$  and  $\boldsymbol{\eta}^n \cdot \boldsymbol{\xi}^n = 1$ . Note that here and it was follows the superscript  $n$  stands for the rescaling parameter and not for an exponentiation.

It is known that under appropriate assumptions on the initial distributions and on the first and second-order moments, the sequence of branching processes with law  $\mathbb{P}^n$  is a nice approximation of the Feller diffusion process with law  $\mathbb{P}$  ([12], Thm. 4.4.2). For each  $n \in \mathbb{N}^*$ , we denote by  $\mathbb{P}^{n,*}$  the law of the process with law  $\mathbb{P}^n$  conditioned on very late extinction, as defined in (1.3). The following theorem then states that under a technical additional assumption on the third-order moments (assumption (2.16)), the sequence of conditioned laws  $\mathbb{P}^{n,*}$  converges weakly to the conditioned law  $\mathbb{P}^*$ .

We denote by  $m_{ij}^n$  the first-order moments of the offspring distribution  $\mathbf{p}^n(\mathbf{k})$ ,

$$m_{ij}^n := \sum_{\mathbf{k} \in \mathbb{N}^d} k_j p_i^n(\mathbf{k}), \tag{2.11}$$

and introduce the second-order moments: for all  $i, j, k = 1 \dots d$ ,

$$\sigma_{ij}^n(l) := \sum_{\mathbf{k} \in \mathbb{N}^d} (k_i - \delta_{li})(k_j - \delta_{lj}) p_l^n(\mathbf{k}). \tag{2.12}$$

**Theorem 2.2.** *Let  $\mathbb{P}$  be the law of a Feller diffusion process with positive risk of extinction, irreducible mutation matrix  $\mathbf{C}$  and variance parameters  $\sigma_1^2, \dots, \sigma_d^2$ . Let  $\{(\mathbf{p}^n(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}, n \in \mathbb{N}^*\}$  be a sequence of offspring distributions satisfying for all  $i, j, l = 1 \dots d$ , as  $n \rightarrow \infty$ ,*

$$m_{ij}^n = \delta_{ij} + \frac{1}{n} c_{ij} + o\left(\frac{1}{n}\right), \tag{2.13}$$

$$\sigma_{ii}^n(i) = \sigma_i^2 + o(1), \tag{2.14}$$

$$\lim_{N \rightarrow \infty} \sup_{n \in \mathbb{N}^*} \sum_{\mathbf{k} / \|\mathbf{k}\| > N} \|\mathbf{k}\|^2 p_i^n(\mathbf{k}) = 0, \tag{2.15}$$

$$\sup_{n \in \mathbb{N}^*} \sum_{\mathbf{k} \in \mathbb{N}^d} k_j^2 k_i p_i^n(\mathbf{k}) < \infty. \tag{2.16}$$

Let  $(\mathbb{P}^n)_{n \in \mathbb{N}^*}$  be the sequence of the associated rescaled branching processes, and  $(\mathbb{P}^{n,*})_{n \in \mathbb{N}^*}$  the sequence of associated conditioned laws.

Then, for any  $\mathbf{y} \in \mathbb{R}_+^d$ ,  $\mathbf{y} \neq \mathbf{0}$ , and any  $\mathbb{N}^d$ -valued sequence  $\mathbf{y}^n$  such that  $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{y}^n = \mathbf{y}$ , the following diagram is commutative,

$$\begin{array}{ccc} \mathbb{P}^n_{\frac{1}{n} \mathbf{y}^n} & \dashrightarrow & \mathbb{P}^{n,*}_{\frac{1}{n} \mathbf{y}^n} \\ \downarrow & & \downarrow \\ \mathbb{P}_{\mathbf{y}} & \dashrightarrow & \mathbb{P}^*_{\mathbf{y}} \end{array}, \tag{2.17}$$

where  $\dashrightarrow$  stands for the transform by conditioning on very late extinction, and  $\Rightarrow$  for the weak convergence of probability measures as  $n$  tends to infinity.

*Proof.* It is already known (see *e.g.* [12]) that under assumptions (2.13)–(2.15), the following weak convergence holds

$$\mathbb{P}_{\frac{1}{n}\mathbf{y}^n}^n \implies \mathbb{P}_{\mathbf{y}}, \quad n \rightarrow \infty. \tag{2.18}$$

The infinitesimal generator of the rescaled branching process with law  $\mathbb{P}^n$  is, for all smooth function  $f : \frac{1}{n}\mathbb{N}^d \rightarrow \mathbb{R}$  and all  $\mathbf{x} \in \frac{1}{n}\mathbb{N}^d$ ,

$$(L^n f)(\mathbf{x}) := n^2 \sum_{i=1}^d x_i \sum_{\mathbf{k} \in \mathbb{N}^d} p_i^n(\mathbf{k}) \left[ f\left(\mathbf{x} + \frac{\mathbf{k} - \mathbf{e}_i}{n}\right) - f(\mathbf{x}) \right]. \tag{2.19}$$

Similar to (1.10) we denote by  $(\tilde{\mathbf{p}}^n(\mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$  the probability distribution

$$\tilde{p}_i^n(\mathbf{k}) := q_{n,i}^{-\frac{1}{n}} (\mathbf{q}_n)^{\frac{\mathbf{k}}{n}} p_i^n(\mathbf{k}), \tag{2.20}$$

where  $\mathbf{q}_n$  stands for the extinction probability vector

$$q_{n,i} := \lim_{t \rightarrow \infty} \mathbb{P}_{\mathbf{e}_i}^n(\mathbf{X}_t = \mathbf{0}). \tag{2.21}$$

We introduce the associated first and second-order moments

$$\tilde{m}_{ij}^n := \sum_{\mathbf{k} \in \mathbb{N}^d} k_j \tilde{p}_i^n(\mathbf{k}), \tag{2.22}$$

$$\tilde{\sigma}_{ij}^n(l) := \sum_{\mathbf{k} \in \mathbb{N}^d} (k_i - \delta_{li})(k_j - \delta_{lj}) \tilde{p}_i^n(\mathbf{k}), \tag{2.23}$$

and define the matrices  $\tilde{\mathbf{M}}^n := (\tilde{m}_{ij}^n)_{i,j=1\dots d}$  and  $\tilde{\mathbf{C}}^n := n(\tilde{\mathbf{M}}^n - \mathbf{I})$ . Assumption (2.13) means that  $\lim_{n \rightarrow \infty} \mathbf{C}^n = \mathbf{C}$ , which implies that for  $n$  large enough, every matrix  $\mathbf{C}^n$  has at least as many positive non-diagonal entries as  $\mathbf{C}$  and is thus irreducible. As shown in the proof of Lemma 1.2, this implies that the matrices  $\tilde{\mathbf{C}}^n$  are irreducible as well, and we denote by  $\tilde{\rho}^n$ ,  $\tilde{\boldsymbol{\xi}}^n$  and  $\tilde{\boldsymbol{\eta}}^n$  their Perron’s roots and right and left eigenvectors. Moreover, from (2.18) it follows that for all  $i = 1 \dots d$ ,

$$\lim_{n \rightarrow \infty} q_{n,i} = e^{-u_i}, \tag{2.24}$$

which implies that for  $n$  large enough the rescaled branching process with law  $\mathbb{P}^n$  has a positive risk of extinction  $\mathbf{q}^n > \mathbf{0}$ . Besides, assumptions (2.13)–(2.15) imply

$$\lim_{n \rightarrow \infty} \sigma_{ij}^n(l) = \begin{cases} \sigma_i^2 & \text{if } (i, j) = (l, l), \\ 0 & \text{otherwise.} \end{cases} \tag{2.25}$$

Indeed if  $(i, j) \neq (l, l)$ , say  $i \neq l$ , we have for any  $N > 0$ ,

$$\begin{aligned} \sigma_{ij}^n(l) &= \sum_{\mathbf{k} \in \mathbb{N}^d} k_i k_j p_l^n(\mathbf{k}) - \delta_{li} m_{li}^n \leq \sum_{\|\mathbf{k}\| \leq N} k_i k_j p_l^n(\mathbf{k}) + \sum_{\|\mathbf{k}\| > N} k_i k_j p_l^n(\mathbf{k}) \\ &\leq N \sum_{\|\mathbf{k}\| \leq N} k_i p_l^n(\mathbf{k}) + \frac{1}{2} \left[ \sum_{\|\mathbf{k}\| > N} k_i^2 p_l^n(\mathbf{k}) + \sum_{\|\mathbf{k}\| > N} k_j^2 p_l^n(\mathbf{k}) - \sum_{\|\mathbf{k}\| > N} (k_i - k_j)^2 p_l^n(\mathbf{k}) \right] \\ &\leq N m_{li}^n + \frac{1}{2} \sup_{n \in \mathbb{N}^*} \sum_{\|\mathbf{k}\| > N} \|\mathbf{k}\|^2 p_l^n(\mathbf{k}). \end{aligned}$$



Let  $\varepsilon > 0$ . (2.15) implies that there exists  $N$  such that  $\sup_{n \in \mathbb{N}^*} \sum_{\|\mathbf{k}\| > N} \|\mathbf{k}\|^2 p_l^n(\mathbf{k}) < \varepsilon$ , and from (2.13) there exists  $N_0 \in \mathbb{N}^*$  such that for all  $n \geq N_0$ ,  $m_{ii}^n < \frac{\varepsilon}{2N}$ . Then for all  $n \geq N_0$ ,  $\sigma_{ij}^n(l) < \varepsilon$ , which completed with (2.14) leads to (2.25).

This ensures finiteness of the second-order moments for the distribution  $\mathbf{p}^n(\mathbf{k})$  from a certain rank.  $\mathbb{P}^n$  thus satisfies assumptions (i)–(iii) of Theorem 1.1. From now on we work with  $n$  large enough such that Theorem 1.1 holds for  $\mathbb{P}^n$ . From (1.17) the infinitesimal generator  $L^{n,*}$  of the rescaled branching process conditioned on very late extinction with law  $\mathbb{P}^{n,*}$  is, for all smooth function  $f : \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$  and all  $\mathbf{x} \in \frac{1}{n}\mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$(L^{n,*} f)(\mathbf{x}) := n^2 \sum_{i=1}^d x_i \sum_{\mathbf{k} \in \mathbb{N}^d} \tilde{p}_i^n(\mathbf{k}) \frac{(n\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \tilde{\xi}^n}{n\mathbf{x} \cdot \tilde{\xi}^n} \left[ f\left(\mathbf{x} + \frac{\mathbf{k} - \mathbf{e}_i}{n}\right) - f(\mathbf{x}) \right]. \tag{2.26}$$

Introducing the state-dependent probability distribution  $(s^n(\mathbf{x}, \mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$  and branching rates  $\alpha_i^n(\mathbf{x})$  defined for all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ , by

$$s_i^n(\mathbf{x}, \mathbf{k}) := \frac{(\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \tilde{\xi}^n}{\mathbf{x} \cdot \tilde{\xi}^n + \tilde{\rho}^n \tilde{\xi}_i^n} \tilde{p}_i^n(\mathbf{k}),$$

$$\alpha_i^n(\mathbf{x}) := \frac{\mathbf{x} \cdot \tilde{\xi}^n + \tilde{\rho}^n \tilde{\xi}_i^n}{\mathbf{x} \cdot \tilde{\xi}^n},$$

the infinitesimal generator  $L^{n,*}$  can be written

$$(L^{n,*} f)(\mathbf{x}) = n^2 \sum_{i=1}^d \alpha_i^n(n\mathbf{x}) x_i \sum_{\mathbf{k} \in \mathbb{N}^d} s_i^n(n\mathbf{x}, \mathbf{k}) \left[ f\left(\mathbf{x} + \frac{\mathbf{k} - \mathbf{e}_i}{n}\right) - f(\mathbf{x}) \right]. \tag{2.27}$$

The law  $\mathbb{P}^{n,*}$  is thus the law of a branching process with state-dependent offspring distribution and branching rates.

We can now apply the convergence criterion provided in [12]. For this purpose, we introduce the moments associated with the probability distribution  $(s^n(\mathbf{x}, \mathbf{k}))_{\mathbf{k} \in \mathbb{N}^d}$ . For every  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$m_{ij}^n(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbb{N}^d} s_i^n(\mathbf{x}, \mathbf{k}) k_j,$$

$$\sigma_{ij}^n(l)(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbb{N}^d} s_l^n(\mathbf{x}, \mathbf{k}) (k_i - \delta_{li})(k_j - \delta_{lj}), \tag{2.28}$$

and we define

$$c_{ij}^n(\mathbf{x}) := n\alpha_i^n(\mathbf{x}) (m_{ij}^n(\mathbf{x}) - \delta_{ij}). \tag{2.29}$$

Then, according to Theorem 4.4.2 in [12], the weak convergence of  $\mathbb{P}_{n^{-1}\mathbf{y}^n}^{n,*}$  to  $\mathbb{P}_{\mathbf{y}}^*$  holds if for all  $i, j = 1 \dots d$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$\sup_{n \in \mathbb{N}^*} \sup_{\substack{\mathbf{x} \in \mathbb{N}^d \\ \mathbf{x} \neq \mathbf{0}}} \alpha_i^n(n\mathbf{x}) < \infty, \quad \lim_{n \rightarrow \infty} \alpha_i^n(n\mathbf{x}) = 1, \tag{2.30}$$

$$\sup_{n \in \mathbb{N}^*} \sup_{\substack{\mathbf{x} \in \mathbb{N}^d \\ \mathbf{x} \neq \mathbf{0}}} c_{ij}^n(n\mathbf{x}) < \infty, \quad \lim_{n \rightarrow \infty} c_{ij}^n(n\mathbf{x}) = c_{ij}(\mathbf{x}), \tag{2.31}$$

$$\lim_{N \rightarrow \infty} \sup_{n \in \mathbb{N}^*} \sup_{\substack{\mathbf{x} \in \mathbb{N}^d \\ \mathbf{x} \neq \mathbf{0}}} \sum_{\|\mathbf{k}\| > N} \|\mathbf{k}\|^2 s_i^n(\mathbf{x}, \mathbf{k}) = 0, \tag{2.32}$$

$$\sup_{n \in \mathbb{N}^*} \sup_{\substack{\mathbf{x} \in \mathbb{N}^d \\ \mathbf{x} \neq \mathbf{0}}} \sigma_{ii}^n(i)(\mathbf{x}) < \infty, \quad \lim_{n \rightarrow \infty} \alpha_i^n(n\mathbf{x}) \sigma_{ii}^n(i)(n\mathbf{x}) = \sigma_i^2, \tag{2.33}$$

and if the martingale problem  $\mathcal{MP}(L^*, C_b^2(\mathbb{R}_+^d \setminus \{\mathbf{0}\}, \mathbb{R}), \mathbf{y})$  has a unique solution.

First we can easily prove that  $\lim_{n \rightarrow \infty} \mathbf{C}^n = \mathbf{C}$  implies

$$\lim_{n \rightarrow \infty} \rho^n = \rho, \quad \lim_{n \rightarrow \infty} \boldsymbol{\xi}^n = \boldsymbol{\xi}. \quad (2.34)$$

Moreover, we have on the one hand  $\tilde{c}_{ij} = c_{ij} - \sigma_i^2 u_i \delta_{ij}$ , and on the other hand

$$\tilde{c}_{ij}^n = n \left( \frac{1}{q_{n,i}} \sum_{\mathbf{k} \in \mathbb{N}^d} (\mathbf{q}_n)^{\mathbf{k}} k_j p_i^n(\mathbf{k}) - \delta_{ij} \right) = c_{ij}^n + \sum_{\mathbf{k} \in \mathbb{N}^d} n \left( q_{n,i}^{-1} (\mathbf{q}_n)^{\mathbf{k}} - 1 \right) k_j p_i^n(\mathbf{k}).$$

We deduce from (2.24) that for all  $i = 1 \dots d$ ,

$$\lim_{n \rightarrow \infty} q_{n,i}^{\frac{1}{n}} = 1, \quad (2.35)$$

and that for all  $\mathbf{k} \in \mathbb{N}^d$ ,

$$\lim_{n \rightarrow \infty} n \left( q_{n,i}^{-\frac{1}{n}} (\mathbf{q}_n)^{\frac{\mathbf{k}}{n}} - 1 \right) = -(k_i - 1)u_i.$$

It ensues that

$$\lim_{n \rightarrow \infty} \tilde{\mathbf{C}}^n = \tilde{\mathbf{C}}, \quad (2.36)$$

from which it follows that

$$\lim_{n \rightarrow \infty} \tilde{\rho}^n = \tilde{\rho}, \quad \lim_{n \rightarrow \infty} \tilde{\boldsymbol{\xi}}^n = \tilde{\boldsymbol{\xi}}. \quad (2.37)$$

Let us now prove that (2.30)–(2.33) are satisfied. (2.30) is immediate. Let us show (2.31). For all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ , we have

$$\begin{aligned} c_{ij}^n(n\mathbf{x}) &= n \left[ \sum_{\mathbf{k} \in \mathbb{N}^d} k_j \frac{(n\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}}^n}{n\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \tilde{p}_i^n(\mathbf{k}) - \frac{n\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n + \tilde{\rho}^n \tilde{\xi}_i^n}{n\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \delta_{ij} \right] \\ &= n (\tilde{m}_{ij}^n - \delta_{ij}) + \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \left[ \sum_{\mathbf{k} \in \mathbb{N}^d} k_j (\mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}}^n \tilde{p}_i^n(\mathbf{k}) - \tilde{\rho}^n \tilde{\xi}_i^n \delta_{ij} \right] = \tilde{c}_{ij}^n + \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \sum_{l=1}^d \tilde{\xi}_l^n \tilde{\sigma}_{jl}^n(i), \end{aligned} \quad (2.38)$$

From the definition of  $\tilde{\sigma}_{ij}^n(l)$  we easily obtain thanks to (2.25) and (2.35) that

$$\lim_{n \rightarrow \infty} \tilde{\sigma}_{ij}^n(l) = \begin{cases} \sigma_i^2 & \text{if } (i, j) = (l, l), \\ 0 & \text{otherwise,} \end{cases} \quad (2.39)$$

which together with (2.36), (2.37) and (2.38) implies that  $\sup_{n \in \mathbb{N}^*} \sup_{\substack{\mathbf{x} \in \mathbb{N}^d \\ \mathbf{x} \neq \mathbf{0}}} c_{ij}^n(n\mathbf{x}) < \infty$ , and that

$$\lim_{n \rightarrow \infty} c_{ij}^n(n\mathbf{x}) = \tilde{c}_{ij} + \frac{\sigma_i^2 \tilde{\xi}_i}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}} \delta_{ij}.$$

Let us next prove (2.32). For all  $N \in \mathbb{N}^*$  and all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$\sum_{\|\mathbf{k}\| > N} \|\mathbf{k}\|^2 s_i^n(\mathbf{x}, \mathbf{k}) = \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n + \tilde{\rho}^n \tilde{\xi}_i^n} \left[ \mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n \sum_{\|\mathbf{k}\| > N} \|\mathbf{k}\|^2 \tilde{p}_i^n(\mathbf{k}) + \sum_{\|\mathbf{k}\| > N} \|\mathbf{k}\|^2 (\mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}}^n \tilde{p}_i^n(\mathbf{k}) \right].$$

We have on the one hand

$$\sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 \tilde{p}_i^n(\mathbf{k}) \leq \frac{1}{q_{n,i}} \sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 p_i^n(\mathbf{k}),$$

and on the other hand

$$\sum_{\|\mathbf{k}\|>N} \|\mathbf{k}\|^2 (\mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}} \tilde{p}_i^n(\mathbf{k}) \leq \frac{1}{q_{n,i}} \sum_{l=1}^d \tilde{\xi}_l^n \sum_{j=1}^d \sum_{\mathbf{k} \in \mathbb{N}^d} k_j^2 k_l p_i^n(\mathbf{k}).$$

This together with assumptions (2.15)–(2.16) and the convergence results given by (2.35) and (2.37) leads to (2.32).

It remains to prove (2.33). For all  $\mathbf{x} \in \mathbb{N}^d$ ,  $\mathbf{x} \neq \mathbf{0}$ ,

$$\begin{aligned} \alpha_i^n(n\mathbf{x}) \sigma_{ii}^n(i)(n\mathbf{x}) &= \sum_{\mathbf{k} \in \mathbb{N}^d} \frac{(n\mathbf{x} + \mathbf{k} - \mathbf{e}_i) \cdot \tilde{\boldsymbol{\xi}}^n}{n\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} (k_i - 1)^2 \tilde{p}_i^n(\mathbf{k}) \\ &= \tilde{\sigma}_{ii}^n(i) + \frac{1}{n} \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \sum_{l=1}^d \tilde{\xi}_l^n \sum_{\mathbf{k} \in \mathbb{N}^d} (k_l - \delta_{il}) (k_i^2 - 2k_i + 1) \tilde{p}_i^n(\mathbf{k}) \\ &= \tilde{\sigma}_{ii}^n(i) + \frac{1}{n} \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \sum_{l=1}^d \tilde{\xi}_l^n \left( \sum_{\mathbf{k} \in \mathbb{N}^d} k_i^2 k_l \tilde{p}_i^n(\mathbf{k}) - 2\tilde{\sigma}_{il}^n(i) + \tilde{m}_{il}^n - \delta_{il} \right), \end{aligned}$$

which thanks to (2.36), (2.37), (2.39) and assumption (2.16) converges to  $\sigma_i^2$  as  $n$  tends to infinity. Writing

$$\sigma_{ii}^n(i)(\mathbf{x}) = \frac{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n + \tilde{\rho}^n \tilde{\xi}_i^n} \left[ \tilde{\sigma}_{ii}^n(i) + \frac{1}{\mathbf{x} \cdot \tilde{\boldsymbol{\xi}}^n} \sum_{l=1}^d \tilde{\xi}_l^n \left( \sum_{\mathbf{k} \in \mathbb{N}^d} k_i^2 k_l \tilde{p}_i^n(\mathbf{k}) - 2\tilde{\sigma}_{il}^n(i) + \tilde{m}_{il}^n - \delta_{il} \right) \right],$$

we obtain (2.33) thanks to the same convergence results.

Let us finally prove that the martingale problem  $\mathcal{MP}(L^*, C_b^2(\mathbb{R}_+^d \setminus \{\mathbf{0}\}, \mathbb{R}), \mathbf{y})$  admits as unique solution the conditioned law  $\mathbb{P}_{\mathbf{y}}^*$ . For this purpose we define the following subset of bounded  $C^2$ -functions on  $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$ ,

$$D_0(L^*) := \{ \mathbf{x} \mapsto e^{-\boldsymbol{\lambda} \cdot \mathbf{x}}, \boldsymbol{\lambda} \in \mathbb{R}_+^d \}.$$

Since  $L^*$  maps  $D_0(L^*)$  into bounded continuous functions on  $\mathbb{R}_+^d \setminus \{\mathbf{0}\}$ ,  $D_0(L^*)$  is a core for  $L^*$ . It is thus enough to prove that  $\mathcal{MP}(L^*, D_0(L^*), \mathbf{y})$  admits a unique solution, *i.e.* that there exists at most one  $\mathbf{P}$  such that for all  $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ ,

$$e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_t} - e^{-\boldsymbol{\lambda} \cdot \mathbf{y}} - \int_0^t \left( \mathbf{X}_s \cdot \boldsymbol{\psi}(\boldsymbol{\lambda}) + \frac{\mathbf{X}_s \cdot \boldsymbol{\varphi}(\boldsymbol{\lambda})}{\mathbf{X}_s \cdot \tilde{\boldsymbol{\xi}}} \right) e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_s} ds \tag{2.40}$$

is a  $(\mathbf{P}, \mathcal{F}_t)$ -martingale, where for all  $i = 1 \dots d$  and all  $i, j = 1 \dots d$ ,

$$\psi_i(\boldsymbol{\lambda}) := - \sum_{j=1}^d \tilde{c}_{ij} \lambda_j + \frac{1}{2} \sigma_i^2 \lambda_i^2, \quad \varphi_i(\boldsymbol{\lambda}) := -\sigma_i^2 \tilde{\xi}_i \lambda_i. \tag{2.41}$$

Applying the martingale problem  $\mathcal{MP}(L, D_0(L), \mathbf{y})$  to the function  $f(t, \mathbf{x}) := e^{-\tilde{\rho}t} \mathbf{x} \cdot \tilde{\boldsymbol{\xi}} e^{-\boldsymbol{\lambda} \cdot \mathbf{x}}$ ,  $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ , we know that

$$e^{-\tilde{\rho}t} \mathbf{X}_t \cdot \tilde{\boldsymbol{\xi}} e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_t} - \mathbf{y} \cdot \tilde{\boldsymbol{\xi}} e^{-\boldsymbol{\lambda} \cdot \mathbf{y}} - \int_0^t e^{-\tilde{\rho}s} \mathbf{X}_s \cdot \tilde{\boldsymbol{\xi}} \left( \mathbf{X}_s \cdot \boldsymbol{\psi}(\boldsymbol{\lambda}) + \frac{\mathbf{X}_s \cdot \boldsymbol{\varphi}(\boldsymbol{\lambda})}{\mathbf{X}_s \cdot \tilde{\boldsymbol{\xi}}} \right) e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_s} ds \tag{2.42}$$

is a  $(\mathbb{P}_y, \mathcal{F}_t)$ -martingale. Furthermore, according to Theorem 1.1,  $\mathbb{P}_y^*$  is a Doob  $h$ -transform of  $\mathbb{P}_y$  via the harmonic function  $h(t, \mathbf{x}) := e^{-\tilde{\rho}t} \mathbf{x} \cdot \tilde{\xi}$ . Together with (2.42) this implies that (2.40) is a  $(\mathbb{P}_y^*, \mathcal{F}_t)$ -martingale for all  $\lambda \in \mathbb{R}_+^d$ . The law  $\mathbb{P}_y^*$  is thus a solution to  $\mathcal{MP}(L^*, D_0(L^*), y)$ . Any other solution  $\widehat{\mathbb{P}}_y^*$  would have the same representation  $d\widehat{\mathbb{P}}_y^*|_{\mathcal{F}_t} := \frac{h(0, y)}{h(t, \mathbf{X}_t)} d\mathbb{P}_y^*|_{\mathcal{F}_t}$ . By uniqueness of the solution to the martingale problem  $\mathcal{MP}(L)$  this implies  $\widehat{\mathbb{P}}_y = \mathbb{P}_y$ , and thus  $\widehat{\mathbb{P}}_y^* = \mathbb{P}_y^*$ .  $\square$

**Remark 2.3.** The proof of Theorem 2.2 ensures as well the commutativity between *rescaling* and *conditioning on extinction*. We obtain indeed thanks to (2.36) and (2.39) that under assumptions (2.13)–(2.15), the laws  $\widetilde{\mathbb{P}}_{n-1}^n$  of the rescaled branching processes conditioned on extinction converge weakly to the law  $\widetilde{\mathbb{P}}_y$  of the Feller diffusion process conditioned on extinction. The following diagram is thus commutative,

$$\begin{array}{ccc}
 \mathbb{P}_{\frac{1}{n}y^n}^n & \rightsquigarrow & \widetilde{\mathbb{P}}_{\frac{1}{n}y^n}^n \\
 \downarrow & & \downarrow \\
 \mathbb{P}_y & \rightsquigarrow & \widetilde{\mathbb{P}}_y
 \end{array}, \tag{2.43}$$

where  $\rightsquigarrow$  stands for the transform by conditioning on extinction, and  $\Rightarrow$  for the weak convergence of probability measures as  $n$  tends to infinity.

**2.3. Commutativity of the long-time limits**

The purpose of this section is to show the commutativity between the three possible limits in  $n$ ,  $t$  and  $\theta$  of

$$\mathbb{P}_{\frac{1}{n}x^n}^n \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right).$$

From the study of the long-time behavior of the conditioned branching processes in Proposition 1.7, we first obtain that under the assumptions of Theorem 2.2, for  $\rho \neq 0$ ,

$$\lim_n \lim_\theta \lim_t = \lim_n \lim_t \lim_\theta. \tag{2.44}$$

On the other hand, we immediately deduce from the commutativity result for the long-time limits of the conditioned Feller diffusion process (Thm. 3.7 in [2]) that

$$\lim_\theta \lim_t \lim_n = \lim_t \lim_\theta \lim_n, \tag{2.45}$$

and from Theorem 2.2 that

$$\lim_t \lim_\theta \lim_n = \lim_t \lim_n \lim_\theta. \tag{2.46}$$

In order to obtain the equality of the six possible combinations of limits in  $n$ ,  $t$  and  $\theta$ , we have to prove that for any  $\theta \geq 0$  fixed the limits in  $t$  and  $n$  commute. This would lead to the equality

$$\lim_\theta \lim_n \lim_t = \lim_\theta \lim_t \lim_n. \tag{2.47}$$

Let  $\theta \geq 0$  fixed. We want to show that the following holds

$$\lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{P}_{\frac{1}{n}x^n}^n \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right) = \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}_{\frac{1}{n}x^n}^n \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right). \tag{2.48}$$

From the proof of Proposition 1.7, we know that for  $n$  large enough the long-time limit

$$\lim_{t \rightarrow \infty} \mathbb{P}_{\frac{1}{n}x^n}^n \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right)$$

defines a probability distribution on  $\frac{1}{n}\mathbb{N}^d$  with generating function  $F^{\theta,n}$  given by

$$F^{\theta,n}(\mathbf{r}) = e^{-\tilde{\rho}^n \theta} [F^{0,n}(\mathbf{r}) - F^{0,n}(\mathbf{r}\tilde{\mathbf{q}}_n(\theta)^n)]. \tag{2.49}$$

Here  $\tilde{\mathbf{q}}_n(\theta)$  denotes the extinction probability vector at time  $\theta$  for the subcritical process with law  $\tilde{\mathbb{P}}^n$ , and  $F^{0,n}$  denotes the generating function of the Yaglom distribution associated with the subcritical process with law  $\tilde{\mathbb{P}}^n$ ,  $F^{0,n}(\mathbf{r}) := \lim_{t \rightarrow \infty} \tilde{\mathbb{E}}^n[\mathbf{r}^{\mathbf{X}_t} | \mathbf{X}_t \neq \mathbf{0}]$ .

On the other hand, by Remark 2.3, the right side of (2.48) is equal to  $\lim_{t \rightarrow \infty} \tilde{\mathbb{P}}_{\mathbf{x}_0}(\mathbf{X}_t \in \cdot | \mathbf{X}_{t+\theta} \neq \mathbf{0})$ . We easily show that this limit defines a probability distribution on  $\mathbb{R}_+^d$  with Laplace transform  $\Phi^\theta$  given by

$$\Phi^\theta(\boldsymbol{\lambda}) = e^{-\tilde{\rho} \theta} [\Phi^0(\boldsymbol{\lambda}) - \Phi^0(\boldsymbol{\lambda} + \tilde{\mathbf{u}}(\theta))]. \tag{2.50}$$

Here  $\tilde{\mathbf{u}}(\theta)$  denotes the cumulant at time  $\theta$  of the process with law  $\tilde{\mathbb{P}}$ , and  $\Phi^0$  denotes the Laplace transform of the Yaglom distribution associated with the subcritical process with law  $\tilde{\mathbb{P}}$ ,  $\Phi^0(\boldsymbol{\lambda}) := \lim_{t \rightarrow \infty} \tilde{\mathbb{E}}[e^{-\boldsymbol{\lambda} \cdot \mathbf{X}_t} | \mathbf{X}_t \neq \mathbf{0}]$ .

Hence we need to prove that for any  $\theta \geq 0$  and for all  $\boldsymbol{\lambda} \geq \mathbf{0}$ , denoting  $\mathbf{e}^{-\boldsymbol{\lambda}} := (e^{-\lambda_1}, \dots, e^{-\lambda_d})$ ,

$$\lim_{n \rightarrow \infty} F^{\theta,n}(\mathbf{e}^{-\boldsymbol{\lambda}}) = \Phi^\theta(\boldsymbol{\lambda}). \tag{2.51}$$

From Remark 2.3 we have  $\lim_{n \rightarrow \infty} \tilde{q}_{n,i}(\theta) = e^{-\tilde{u}_i(\theta)}$  and, as shown in (2.37),  $\lim_{n \rightarrow \infty} \tilde{\rho}^n = \tilde{\rho}$ . Hence we see by (2.49) and (2.50) that the convergence (2.51) holds as soon as it is true for  $\theta = 0$ , *i.e.*

$$\lim_{n \rightarrow \infty} F^{0,n}(\mathbf{e}^{-\boldsymbol{\lambda}}) = \Phi^0(\boldsymbol{\lambda}). \tag{2.52}$$

On the one hand we have, as seen in (1.25),

$$F^{0,n}(\mathbf{r}) = 1 - \frac{\gamma^n(\mathbf{r})}{\gamma^n(\mathbf{0})},$$

where  $\gamma^n(\mathbf{r})$  satisfies (see (1.23))

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}^n t} \tilde{\boldsymbol{\eta}}^n \cdot (\mathbf{1} - \tilde{\mathbf{F}}_t^n(\mathbf{r})) = \gamma^n(\mathbf{r}). \tag{2.53}$$

Here  $\tilde{\mathbf{F}}_t^n$  denotes the generating function at time  $t$  of the rescaled process with law  $\tilde{\mathbb{P}}^n$ , defined by  $\tilde{F}_{t,i}^n(\mathbf{r}) := \tilde{\mathbb{E}}_{\mathbf{e}_i}^n(\mathbf{r}^{\mathbf{X}_t})$ . On the other hand, a small computation shows that for all  $\boldsymbol{\lambda} \in \mathbb{R}_+^d$ ,

$$\Phi^0(\boldsymbol{\lambda}) = 1 - \frac{\kappa(\boldsymbol{\lambda})}{\kappa},$$

where  $\kappa(\boldsymbol{\lambda})$  and  $\kappa > 0$  satisfy (see proof of Theorem 3.7 in [2])

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho} t} \tilde{\boldsymbol{\eta}} \cdot \tilde{\mathbf{u}}_t(\boldsymbol{\lambda}) = \kappa(\boldsymbol{\lambda}),$$

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho} t} \tilde{\boldsymbol{\eta}} \cdot \left( \lim_{\lambda \rightarrow \infty} \tilde{\mathbf{u}}_t(\boldsymbol{\lambda}) \right) = \kappa.$$

In order to obtain (2.52) we first prove that the convergence (2.53) is uniform in  $n$ . As already mentioned, (1.23) (and thus (2.53)) is obtained as an extension *via* the embedded process of a convergence result for BGW processes (Thm. 2 in [13]), *via* a method detailed in the proof of Theorem 6.1 in [18]. Analysing this proof, it appears that the uniform convergence in  $n$  for (2.53) stems from the uniform convergence in  $n$  of

$$\lim_{\substack{k \rightarrow \infty \\ k \in \mathbb{N}}} e^{-\tilde{\rho}^n k} \tilde{\boldsymbol{\eta}}^n \cdot (\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r})) = \gamma^n(\mathbf{r}). \tag{2.54}$$

We thus consider the embedded subcritical BGW process with offspring generating function  $\tilde{\mathbf{F}}_1^n$ , mean matrix  $\exp(\tilde{\mathbf{C}}^n)$ , maximal eigenvalue  $e^{\tilde{\rho}^n}$ , associated eigenvectors  $\tilde{\boldsymbol{\xi}}^n$  and  $\tilde{\boldsymbol{\eta}}^n$ . By the integral form of the remainder term in the Taylor expansion of  $\tilde{\mathbf{F}}_1^n$ , there exists a non-negative matrix-valued rest  $\mathbf{A}^n(\mathbf{r})$  such that for all  $\mathbf{r} \in [0, 1]^d$ ,

$$\mathbf{1} - \tilde{\mathbf{F}}_1^n(\mathbf{r}) = \left( \exp(\tilde{\mathbf{C}}^n) - \mathbf{A}^n(\mathbf{r}) \right) (\mathbf{1} - \mathbf{r}), \tag{2.55}$$

satisfying  $\mathbf{A}^n(\mathbf{r}) = O(\|\mathbf{1} - \mathbf{r}\|)$  as  $\mathbf{r} \rightarrow \mathbf{1}$ . Moreover, the second-order derivatives of  $\tilde{\mathbf{F}}_1^n$  being finite and bounded in  $n$  thanks to (2.39), we have in the neighborhood of  $\mathbf{1}$

$$\mathbf{A}^n(\mathbf{r}) = O(\|\mathbf{1} - \mathbf{r}\|) \quad \text{uniformly in } n. \tag{2.56}$$

Let us denote

$$\Delta_k^n(\mathbf{r}) := e^{-\tilde{\rho}^n k} \tilde{\boldsymbol{\eta}}^n \cdot \left( \mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r}) \right).$$

By (2.55) we have

$$\Delta_{k+1}^n(\mathbf{r}) - \Delta_k^n(\mathbf{r}) = -e^{-(k+1)\tilde{\rho}^n} \tilde{\boldsymbol{\eta}}^n \cdot \mathbf{A}^n \left( \tilde{\mathbf{F}}_k^n(\mathbf{r}) \right) \left( \mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r}) \right),$$

hence for every  $n$  and  $\mathbf{r}$ ,  $\Delta_k^n(\mathbf{r})$  is decreasing in  $k$ . It follows that for every  $n, k \in \mathbb{N}$  and  $\mathbf{r} \in [0, 1]^d$ ,

$$\Delta_k^n(\mathbf{r}) \leq \Delta_0^n(\mathbf{r}) = \tilde{\boldsymbol{\eta}}^n \cdot (\mathbf{1} - \mathbf{r}) \leq \sup_{n \in \mathbb{N}} \tilde{\boldsymbol{\eta}}^n \cdot \mathbf{1},$$

the right term being finite by means of (2.36). This ensures that  $\sup_{n \in \mathbb{N}} \gamma^n(\mathbf{r}) < \infty$ . Let  $N \in \mathbb{N}$  such that  $\sup_{n \geq N} \tilde{\rho}^n < 0$ . From (1.23) we obtain that for every  $n$ , as  $k \rightarrow \infty$ ,

$$\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r}) \sim e^{\tilde{\rho}^n k} \gamma^n(\mathbf{r}) \tilde{\boldsymbol{\xi}}^n,$$

from which we deduce thanks to (2.37) that

$$\lim_{k \rightarrow \infty} \tilde{\mathbf{F}}_k^n(\mathbf{r}) = \mathbf{1} \quad \text{uniformly in } n \text{ up to } N.$$

(Note that the convergence is also uniform in  $\mathbf{r} \in [0, 1]^d$ ). Together with (2.56) this implies the existence of  $C_1 > 0$  and  $K > 0$  such that for all  $k \geq K$  and all  $n \geq N$ ,

$$\mathbf{A}^n \left( \tilde{\mathbf{F}}_k^n(\mathbf{r}) \right) \leq C_1 \|\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r})\| \mathbf{I}. \tag{2.57}$$

Since

$$e^{-\tilde{\rho}^n k} \left( \mathbf{1} - \tilde{\mathbf{F}}_{k,i}^n(\mathbf{r}) \right) \leq \frac{1}{\inf_{n,i} \tilde{\eta}_i^n} e^{-\tilde{\rho}^n k} \tilde{\boldsymbol{\eta}}^n \cdot \left( \mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r}) \right),$$

there exists  $C_2 > 0$  such that for all  $k$  and all  $n \geq N$ ,

$$e^{-\tilde{\rho}^n k} \|\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r})\| \leq C_2. \tag{2.58}$$

Now, for every  $k \geq K$ ,  $p \geq 0$  and every  $n \geq N$ ,

$$\begin{aligned} \Delta_k^n(\mathbf{r}) - \Delta_{k+p}^n(\mathbf{r}) &= \sum_{i=0}^{p-1} e^{-\tilde{\rho}^n(k+1+i)} \tilde{\boldsymbol{\eta}}^n \cdot \mathbf{A}^n \left( \tilde{\mathbf{F}}_{k+i}^n(\mathbf{r}) \right) \left( \mathbf{1} - \tilde{\mathbf{F}}_{k+i}^n(\mathbf{r}) \right) \\ &\leq e^{(k-1) \sup_n \tilde{\rho}^n} C_1 (C_2)^2 \sup_n (\tilde{\boldsymbol{\eta}}^n \cdot \mathbf{1}) \frac{1}{1 - e^{\sup_n \tilde{\rho}^n}}. \end{aligned}$$

We thus obtain by virtue of Cauchy criterion that the convergence (2.54) is uniform in  $n$  up to  $N$ . As a consequence, the convergence (2.53) is uniform as well, and we obtain that for all  $\lambda \in \mathbb{R}_+^d$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \gamma^n(\mathbf{e}^{-\lambda}) &= \lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} e^{-\tilde{\rho}^n t \tilde{\eta}^n} \cdot \left( \mathbf{1} - \tilde{\mathbf{F}}_t^n(\mathbf{e}^{-\lambda}) \right) = \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} e^{-\tilde{\rho}^n t \tilde{\eta}^n} \cdot \left( \mathbf{1} - \tilde{\mathbf{F}}_t^n(\mathbf{e}^{-\lambda}) \right) \\ &= \lim_{t \rightarrow \infty} e^{-\tilde{\rho} t \tilde{\eta}} \cdot \left( \mathbf{1} - \mathbf{e}^{-\tilde{\mathbf{u}}_t(\lambda)} \right) = \kappa(\lambda). \end{aligned}$$

Similarly

$$\lim_{n \rightarrow \infty} \gamma^n(\mathbf{0}) = \kappa.$$

This finally proves (2.52), which ensures (2.48).

It finally remains to prove that for  $\rho \neq 0$ ,

$$\lim_n \lim_t \lim_\theta = \lim_t \lim_n \lim_\theta, \tag{2.59}$$

for which we show that the convergence proved for every fixed  $n$  in Proposition 1.7,

$$\lim_{t \rightarrow \infty} \mathbb{E}_{\frac{1}{n} \mathbf{x}^n}^{n,*} [\mathbf{r}^{\mathbf{X}_t}] = - \sum_{i=1}^d r_i \tilde{\xi}_i^n \frac{\partial \gamma^n(\mathbf{r})}{\partial r_i},$$

is indeed uniform in  $n$ . As seen in the proof of Proposition 1.7, it is enough to prove that

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}^n t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}^n(\mathbf{r}) \right] = - \frac{\partial \gamma^n(\mathbf{r})}{\partial r_i} \tilde{\xi}_j^n \text{ uniformly in } n. \tag{2.60}$$

For this purpose we follow the steps of Lemma 1.8. First we prove that

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}^n t} (\mathbf{1} - \mathbf{F}_t^n(\mathbf{r})) = \gamma^n(\mathbf{r}) \tilde{\xi}^n \text{ uniformly in } n, \tag{2.61}$$

which similarly as for (2.53) can be deduced from the convergence of the embedded BGW process

$$\lim_{\substack{k \rightarrow \infty \\ k \in \mathbb{N}}} e^{-\tilde{\rho}^n k} (\mathbf{1} - \mathbf{F}_k^n(\mathbf{r})) = \gamma^n(\mathbf{r}) \tilde{\xi}^n \text{ uniformly in } n. \tag{2.62}$$

(2.62) is itself a consequence from (2.54) together with

$$\lim_{k \rightarrow \infty} \frac{\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r})}{\tilde{\eta}^n \cdot (\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r}))} = \tilde{\xi}^n \text{ uniformly in } n. \tag{2.63}$$

Let us prove (2.63). We denote by  $\{\lambda_l^n, l\}$  the eigenvalues of  $\tilde{\mathbf{C}}^n$  different from  $\tilde{\rho}^n$ , and define the spectral gap of  $\tilde{\mathbf{C}}^n$ ,  $R^n := \min_l (|\tilde{\rho}^n| - |\lambda_l^n|)$ . Let  $R$  be the spectral gap of the matrix  $\tilde{\mathbf{C}}$ . By (2.36) we have  $\lim_{n \rightarrow \infty} R^n = R$ . Moreover, for all  $i, j = 1 \dots d$  and  $k \in \mathbb{N}^*$ ,

$$\left[ \exp(k \tilde{\mathbf{C}}^n) \right]_{ij} = \tilde{\xi}_i^n \tilde{\eta}_j^n e^{\tilde{\rho}^n k} + \sum_l \varphi_{ij,l}^n(k) e^{\lambda_l^n k},$$

where  $\varphi_{ij,l}^n$  is a complex-valued polynomial with degree smaller than the algebraic multiplicity of  $\lambda_l^n$ . Since the  $\varphi_{ij,l}^n$  converge as  $n$  tends to infinity, we have  $\sup_n |\varphi_{ij,l}^n(k)| < \infty$  and we can write

$$\left| e^{-\tilde{\rho}^n k} \left[ \exp(k \tilde{\mathbf{C}}^n) \right]_{ij} - \tilde{\xi}_i^n \tilde{\eta}_j^n \right| \leq \sum_l \sup_n |\varphi_{ij,l}^n(k)| e^{-\inf_n R^n k}.$$

Consequently, denoting by  $\mathbf{P}^n$  the matrix with entries  $\tilde{\xi}_i^n \tilde{\eta}_j^n$ ,

$$\lim_{k \rightarrow \infty} e^{-\tilde{\rho}^n k} \exp(k \tilde{\mathbf{C}}^n) = \mathbf{P}^n \quad \text{uniformly in } n, \tag{2.64}$$

and we can find a null sequence  $u_k$  such that for all  $k, n \geq 1$ ,

$$(1 - u_k) \mathbf{P}^n \leq e^{-\tilde{\rho}^n k} \exp(k \tilde{\mathbf{C}}^n) \leq (1 + u_k) \mathbf{P}^n.$$

On the other hand, we know by (2.57) and (2.58) that  $\lim_{k \rightarrow \infty} e^{-\tilde{\rho}^n k} \mathbf{A}^n(\tilde{\mathbf{F}}_{k-1}^n(\mathbf{r})) = \mathbf{0}$  uniformly in  $n$  up to  $N$ , hence we can choose a null sequence  $v_k$  such that for all  $k \geq 1$  and all  $n \geq N$ ,

$$0 \leq e^{-\tilde{\rho}^n k} \mathbf{A}^n(\tilde{\mathbf{F}}_{k-1}^n(\mathbf{r})) \leq v_k \mathbf{P}^n.$$

Then, as detailed in the proof of Theorem 1 in [13], we have for any  $k \geq l \geq 1$  and any  $n \geq N$ ,

$$\left\| \frac{\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r})}{\tilde{\boldsymbol{\eta}}^n \cdot (\mathbf{1} - \tilde{\mathbf{F}}_k^n(\mathbf{r}))} - \tilde{\boldsymbol{\xi}}^n \right\| \leq \frac{2u_l + \sum_{m=k-l+1}^k v_m}{1 - \sum_{m=k-l+1}^k v_m - u_l}.$$

By letting first  $k$  tend to infinity, and then  $l$ , we obtain (2.63) and thus (2.61). Now from (2.61) we deduce that the convergence  $\lim_{t \rightarrow \infty} e^{-\tilde{\rho}^n t} \ln \tilde{F}_{t,j}^n(\mathbf{r}) = -\gamma^n(\mathbf{r}) \tilde{\xi}_j^n$  is uniform in  $n$  too, and for all  $h \geq 0$ ,

$$\gamma(\mathbf{r})^n \tilde{\xi}_j^n - \gamma^n(\mathbf{r} + h\mathbf{e}_i) \tilde{\xi}_j^n = \lim_{t \rightarrow \infty} \int_0^h e^{-\tilde{\rho}^n t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}^n(\mathbf{r} + u\mathbf{e}_i) \right] du \quad \text{uniformly in } n.$$

Moreover,

$$0 \leq e^{-\tilde{\rho}^n t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}^n(\mathbf{r}) \right] \leq e^{-\tilde{\rho}^n t} \frac{1}{\tilde{F}_{t,j}^n(\mathbf{r})} \frac{1}{r_i} \tilde{\mathbb{E}}_{\mathbf{e}_j}^n [X_{t,i}],$$

and by (2.64)

$$\lim_{t \rightarrow \infty} e^{-\tilde{\rho}^n t} \tilde{\mathbb{E}}_{\mathbf{e}_j}^n [X_{t,i}] = \tilde{\xi}_j^n \tilde{\eta}_i^n \quad \text{uniformly in } n.$$

Since  $\lim_{t \rightarrow \infty} \tilde{F}_{t,j}^n(\mathbf{r}) = 1$  uniformly in  $n$  and  $\mathbf{r}$ , there exists a constant  $C > 0$  such that for all  $n \in \mathbb{N}$ ,  $t \geq 0$  and all  $u \in [0, h]$ ,

$$\left| e^{-\tilde{\rho}^n t} \frac{\partial}{\partial r_i} \left[ \ln \tilde{F}_{t,j}^n(\mathbf{r} + u\mathbf{e}_i) \right] \right| \leq \frac{C}{r_i + u},$$

which is integrable in  $u$ . By Lebesgue's dominated convergence theorem we thus have

$$\gamma^n(\mathbf{r}) \tilde{\xi}_j^n - \gamma^n(\mathbf{r} + h\mathbf{e}_i) \tilde{\xi}_j^n = \int_0^h \lim_{t \rightarrow \infty} \left[ e^{-\tilde{\rho}^n t} \frac{\partial}{\partial r_i} \left( \ln \tilde{F}_{t,j}^n(\mathbf{r} + u\mathbf{e}_i) \right) \right] du \quad \text{uniformly in } n,$$



which leads to (2.60). Hence we can write

$$\lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{E}_{\frac{1}{n} \mathbf{x}^n}^{n,*} [\mathbf{r}^{\mathbf{X}_t}] = \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E}_{\frac{1}{n} \mathbf{x}^n}^{n,*} [\mathbf{r}^{\mathbf{X}_t}],$$

which ensures (2.59).

Summing up the results of this section we consequently obtain the following statement.

**Proposition 2.4.** *Under the assumptions of Theorem 2.2 and assuming that  $\rho \neq 0$ , the three limits interchange*

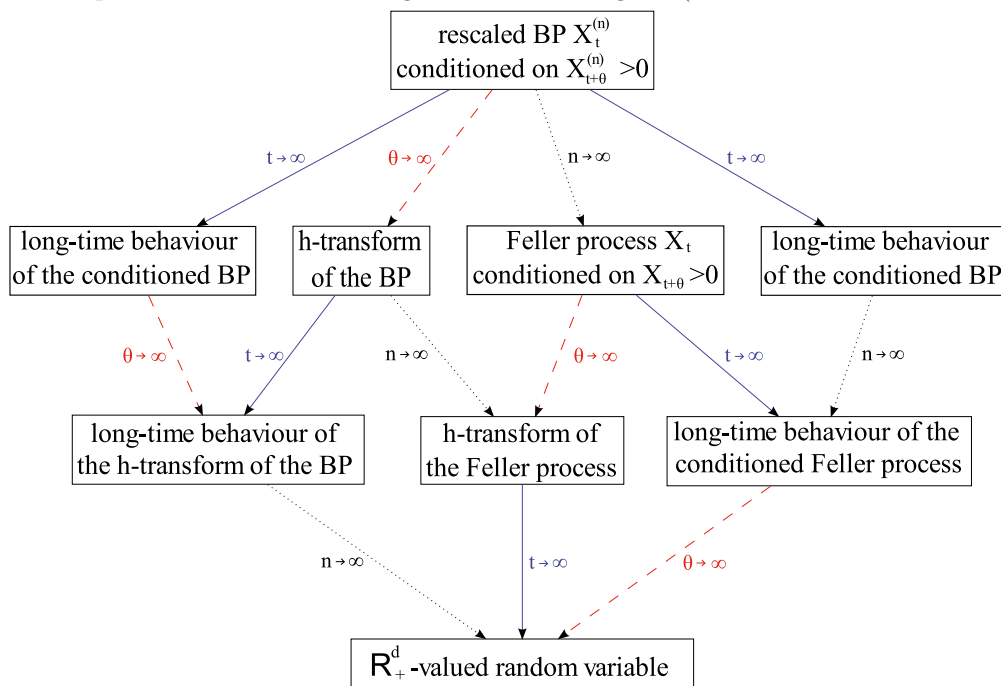
$$\lim_{n, t, \theta} \mathbb{P}_{\frac{1}{n} \mathbf{y}^n}^n \left( \mathbf{X}_t \in \cdot \mid \mathbf{X}_{t+\theta} \neq \mathbf{0}, \lim_{s \rightarrow \infty} \mathbf{X}_s = \mathbf{0} \right). \tag{2.65}$$

Furthermore, the obtained limit is non-degenerate and defines a probability distribution on  $\mathbb{R}_+^d$  which does not depend on  $\mathbf{y} := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{y}^n$ .

**Remark 2.5.** In the monotype case  $d = 1$ , this limit is known to be a Gamma distribution. Denoting by  $\rho$  the mutation matrix (reduced to its eigenvalue) of the unconditioned Feller diffusion process, by  $\sigma^2$  its variance parameter, and defining  $\tilde{\rho} := \rho - \sigma^2 u$  as in (2.7), we have indeed (see e.g. [2], Prop. 3.1),

$$\begin{aligned} \lim_{t \rightarrow \infty} \lim_{\theta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}_{\frac{1}{n} \mathbf{y}^n}^n \left( X_t \in \cdot \mid X_{t+\theta} \neq 0, \lim_{s \rightarrow \infty} X_s = 0 \right) &= \lim_{t \rightarrow \infty} \mathbb{P}^* (X_t \in \cdot) \\ &= \text{Exp}(-2\tilde{\rho}/\sigma^2) \otimes \text{Exp}(-2\tilde{\rho}/\sigma^2) = \Gamma(2, -2\tilde{\rho}/\sigma^2). \end{aligned}$$

We illustrate Proposition 2.4 in the following commutative diagram (where BP stands for *branching process*).



REFERENCES

[1] K.B. Athreya and P.E. Ney, *Branching Processes*. Springer-Verlag (1972).  
 [2] N. Champagnat and S. Røelly, Limit theorems for conditioned multitype Dawson-Watanabe processes and Feller diffusions. *Electronic Journal of Probability* **13** (2008) 777–810.

- [3] S. Dallaporta and A. Joffe, The  $Q$ -process in a multitype branching process. *Int. J. Pure Appl. Math.* **42** (2008) 235–240.
- [4] S.N. Ethier and T.G. Kurtz, *Markov processes: characterization and convergence*. Wiley (1986).
- [5] S.N. Evans, Two representations of a conditioned superprocess, in *Proc. R. Soc. Edinb. Sect. A* **123** (1993) 959–971.
- [6] W. Feller, Diffusion processes in genetics, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, University of California Press, Berkeley and Los Angeles (1951) 227–246.
- [7] F.R. Gantmacher, *Matrizentheorie*. Springer-Verlag (1986).
- [8] H.O. Georgii and E. Baake, Supercritical multitype branching processes: the ancestral types of typical individuals. *Adv. Appl. Probab.* **35** (2003) 1090–1110.
- [9] K. Fleischmann and U. Prehn, Ein Grenzwertsatz für subkritische Verzweigungsprozesse mit endlich vielen Typen von Teilchen. *Math. Nachr.* **64** (1974) 357–362.
- [10] K. Fleischmann and R. Siegmund-Schultze, The structure of reduced critical Galton-Watson processes. *Math. Nachr.* **79** (1977) 233–241.
- [11] P. Jagers and A.N. Lagerås, General branching processes conditioned on extinction are still branching processes. *Electronic Communications in Probability* **13** (2008) 540–547.
- [12] A. Joffe and M. Métivier, Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. Appl. Probab.* **18** (1986) 20–65.
- [13] A. Joffe and F. Spitzer, On multitype branching processes with  $\rho \leq 1$ . *J. Math. Anal. Appl.* **19** (1967) 409–430.
- [14] K. Kawazu and S. Watanabe, Branching processes with immigration and related limit theorems. *Theory Probab. Appl.* **16** (1971) 34–51.
- [15] A.N. Kolmogorov, Zur Lösung einer biologischen Aufgabe. *Comm. Math. Mech. Chebyshev Univ. Tomsk* **2** (1938).
- [16] A. Lambert, Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Electronic Journal of Probability* **12** (2007) 420–446.
- [17] J. Lamperti and P. Ney, Conditioned branching process and their limiting diffusions. *Theory Probab. Appl.* **13** (1968) 126–137.
- [18] Y. Ogura, Asymptotic behavior of multitype Galton-Watson processes. *J. Math. Kyoto Univ.* **15** (1975) 251–302.
- [19] S. Roelly and A. Rouault, Processus de Dawson-Watanabe conditionné par le futur lointain. *C. R. Acad. Sci. Sér. I Math.* **309** (1989) 867–872.
- [20] E. Seneta, *Non-negative matrices – An introduction to theory and applications*. Halsted Press (1973).
- [21] B.A. Sewastjanow, *Verzweigungsprozesse*. R. Oldenbourg Verlag (1975).
- [22] A.M. Yaglom, Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)* **56** (1947) 795–798.