Breadth first search coding of multitype forests with application to Lamperti representation

Loïc Chaumont

Abstract We obtain a bijection between some set of multidimensional sequences and the set of d-type plane forests which is based on the breadth first search algorithm. This coding sequence is related to the sequence of population sizes indexed by the generations, through a Lamperti type transformation. The same transformation in then obtained in continuous time for multitype branching processes with discrete values. We show that any such process can be obtained from a d^2-dimensional compound Poisson process time changed by some integral functional. Our proof bears on the discretisation of branching forests with edge lengths.

1 Introduction

A famous result from Lamperti [8] asserts that any continuous state branching process can be represented as a spectrally positive Lévy process, time changed by the inverse of some integral functional. This transformation is invertible and defines a bijection between the set of spectrally positive Lévy processes and this of continuous state branching processes. The same type of transformation holds between continuous time, integer valued branching processes and downward skip free compound Poisson processes. Lamperti's result is the source of an extensive mathematical literature in which it is mainly used as a tool in branching theory. However, recently Lamperti representation itself has been the focus of some research papers. In [2] several proofs of this result are proposed and in [4] an extension of the transformation to the case of branching processes with immigration is proved. The case of affine processes, which includes continuous state multitype branching processes with immigration, was investigated in [10]. In the latter paper a Lamperti type rep-

Loïc Chaumont – LAREMA – UMR CNRS 6093, Université d'Angers, 2 bd Lavoisier, 49045 Angers cedex 01 e-mail: loic.chaumont@univ-angers.fr

resentation in law of affine processes was obtained. Then right after our paper was submitted, the two prepublications [6] and [3], appeared where pathwise construction of affine processes is obtained through Lamperti representation.

In the present work we show through a combinatorial method, an extension of Lamperti representation to continuous time, integer valued, multitype branching processes. More specifically, let $Z=(Z^{(1)},\dots,Z^{(d)})$ be such a process issued from $x\in\mathbb{Z}_+^d$, then we shall prove that Z can be represented as

$$(Z_t^{(1)},\dots,Z_t^{(d)})=x+\left(\sum_{i=1}^d X^{i,1}_{\int_0^t Z_s^{(i)}\,ds},\dots,\sum_{i=1}^d X^{i,d}_{\int_0^t Z_s^{(i)}\,ds}\right),\quad t\ge 0\,,$$

where $X^{(i)}=(X^{i,1},\dots,X^{i,d})$, $i=1,\dots,d$, are d independent $\mathbb{Z}_+^d$-valued compound Poisson processes. Conversely, given $X^{(i)}$, $i=1,\dots,d$, the above equation admits a unique solution Z, and Z is a multitype branching process. Since 0 is an absorbing state for Z, it is plain that the whole path of $X=(X^{(1)},\dots,X^{(d)})$ is not always needed in the above transformation. For instance, when $d=1$, we see that the process X can be stopped at its first passage time T_x at level $-x$. In the multitype case, we shall prove that the processes $X^{(1)},\dots,X^{(d)}$ can be stopped at times $T_x^{(1)},\dots,T_x^{(d)}$, respectively, where $T_x=(T_x^{(1)},\dots,T_x^{(d)})$ is defined as the 'first passage time' at level $-x$ by the multidimensional random field,

$$\mathrm{X}=(t_1,\dots,t_d)\mapsto\left(\sum_{i=1}^d X^{i,j}_{t_i},\, j\in[d]\right)=X^{(1)}_{t_1}+\ldots+X^{(d)}_{t_d}\,.$$

The law of the multivariate random time T_x will be given in Theorem 1.

The multitype Lamperti representation is not invertible as in the one dimensional case. However, by considering the whole branching structure behind the multitype branching process, we actually obtain in Theorems 2 and 3 a one-to-one pathwise transformation between $(X^{(1)},\dots,X^{(d)})$ and a particular family of processes $(Z^{i,j}:i,j=1,\dots d)$ satisfying the decomposition

$$Z_t^{(i)}=\sum_{i=1}^d Z_t^{i,j}\quad\text{and}\quad Z_t^{i,j}=x_i\mathbf{1}_{i=j}+X^{i,j}_{\int_0^t Z_s^{(i)}\,ds},\ \ t\ge 0,$$

where $Z_t^{i,j}$ is the total number of individuals of type j whose parent has type i and who were born before time t.

The proofs of our results are achieved by using a special coding of multitype plane forests based on the breadth first search algorithm. This deterministic one-to-one correspondence between some set of multivariate sequences and multitype forests is stated in Theorem 4 and leads to a Lamperti type representation of discrete time, multitype branching processes in Theorem 5. Results in discrete time are displayed and proved in Section 3, whereas the next section is devoted to the statements of our results in continuous time. The latter will be proved in Section 4.

2 Main results in continuous time

In all this work, we use the notation $\mathbb{R}_+ = [0,\infty)$, $\mathbb{Z}_+ = \{0,1,2,\dots\}$ and for any positive integer d, we set $[d] = \{1,\dots,d\}$. We also define the following partial order on $\mathbb{R}^d$ by setting $x = (x_1,\dots,x_d) \geq y = (y_1,\dots,y_d)$, if $x_i \geq y_i$, for all $i \in [d]$. Moreover, we write $x > y$ if $x \geq y$ and if there exists $i \in [d]$ such that $x_i > y_i$. We will denote by e_i the i-th unit vector of $\mathbb{Z}_+^d$.

Fix $d \geq 2$ and let $\nu = (\nu_1,\dots,\nu_d)$, where ν_i is any probability measure on $\mathbb{Z}_+^d$ such that $\nu_i(e_i) < 1$. Let $Z = (Z^{(1)},\dots,Z^{(d)})$ be a d-type continuous time and $\mathbb{Z}_+^d$-valued branching process with progeny distribution $\nu = (\nu_1,\dots,\nu_d)$ and such that type $i \in [d]$ has reproduction rate $\lambda_i > 0$. For $i, j \in [d]$, the quantity

$$m_{ij} = \sum_{x \in \mathbb{Z}_+^d} x_j \nu_i(x)\,,$$

corresponds to the mean number of children of type j, given by an individual of type i. Let $M := (m_{ij})_{i,j\in[d]}$ be the mean matrix of Z. We say that the progeny distribution ν is irreducible if the matrix $M = (m_{ij})$ satisfies $m_{ij}^{(n)} > 0$, for some n and for all $i, j \in [d]$, where $m_{ij}^{(n)}$ is the (i, j)-th element of the matrix M^n. Recall that if ν is irreducible, then according to Perron-Frobenius Theorem, it admits a unique eigenvalue ρ which is simple, positive and with maximal modulus. If moreover, ν is non degenerate, that is if individuals have exactly one offspring with probability strictly less than 1, then extinction holds if and only if $\rho \leq 1$, see [7], [12] and Chapter V of [1]. If $\rho = 1$, we say that Z is critical and if $\rho < 1$, we say that Z is subcritical.

We now define the underlying compound Poisson process in the Lamperti representation of Z that will be presented in Theorems 2 and 3. Let $X = (X^{(1)},\dots,X^{(d)})$, where $X^{(i)}$, $i \in [d]$ are d independent $\mathbb{Z}^d$-valued compound Poisson processes. We assume that $X_0^{(i)} = 0$ and that $X^{(i)}$ has rate λ_i and jump distribution

$$\mu_i(\mathrm{k}) = \frac{\tilde{\nu}_i(\mathrm{k})}{1-\tilde{\nu}_i(0)}\,, \text{ if } \mathrm{k} \neq 0 \text{ and } \mu_i(0) = 0\,, \tag{1}$$

where

$$\tilde{\nu}_i(k_1,\dots,k_d) = \nu_i(k_1,\dots,k_{i-1},k_i+1,k_{i+1},\dots,k_d)\,. \tag{2}$$

In particular, with the notation $X^{(i)} = (X^{i,1},\dots,X^{i,d})$, for all $i = 1,\dots,d$, the process $X^{i,i}$ is a $\mathbb{Z}$-valued, downward skip free, compound Poisson process, i.e. $\Delta X_t^{i,i} = X_t^{i,i} - X_{t-}^{i,i} \geq -1$, $t \geq 0$, with $X_{0-} = 0$ and for all $i \neq j$, the process $X^{i,j}$ is an increasing compound Poisson process. We emphasize that in this definition, some of the processes $X^{i,j}$, $i, j \in [d]$ can be identically equal to 0.

We first present a result on passage times of the multidimensional random field

$$\mathrm{X} : (t_1,\dots,t_d) \mapsto \left(\sum_{i=1}^d X^{i,j}_{t_i},\, j\in[d]\right) = X^{(1)}_{t_1}+\dots+X^{(d)}_{t_d}\,,$$

which is a particular case of additive Lévy process, see [11] and the references therein. Henceforth, a process such as X will be called an *additive* (*downward skip free*) *compound Poisson process*.

Theorem 1. *Let $x=(x_1,\dots,x_d)\in\mathbb{Z}_+^d$. Then there exists a (unique) random time $T_x=(T_x^{(1)}\dots,T_x^{(d)})\in\overline{\mathbb{R}}_+^d$ such that almost surely,*

$$x_j+\sum_{i,T_x^{(i)}<\infty} X^{i,j}(T_x^{(i)})=0\,,\quad \textit{for all } j \textit{ such that } T_x^{(j)}<\infty\,, \tag{3}$$

and if T_x' is any random time satisfying (3), then $T_x'\ge T_x$. The time T_x will be called the first passage time of the additive compound Poisson process X at level $-x$.

The process $(T_x, x\in\mathbb{Z}_+^d)$ is increasing and additive, that is, for all $x,y\in\mathbb{Z}_+^d$,

$$T_{x+y}\stackrel{(d)}{=}T_x+\tilde{T}_y\,, \tag{4}$$

where $\tilde{T}_y$ is an independent copy of T_y. The law of T_x on $\mathbb{R}_+^d$ is given by

$$\mathbb{P}(T_x\in dt,\, X^{i,j}_{t_i}=x_{i,j},\, 1\le i,j\le d)=$$

$$\frac{det(-x_{i,j})}{t_1t_2\dots t_d}\prod_{i=1}^d \mathbb{P}(X^{i,j}_{t_i}=x_{i,j},\,1\le j\le d)dt_1dt_2\dots dt_d\,,$$

where the support of this measure is included in $\left\{x_{ij}\in\mathbb{Z}: x_{ij}\ge 0,\, x_{ii}\le 0,\, \sum_{i=1}^d x_{i,j}=-x_i\right\}$.

Note that from the additivity property (4) of $(T_x,\, x\in\mathbb{Z}_+^d)$, we derive that the law of this process is characterised by the law of the variables T_{e_i} for $i\in[d]$.

As the above statement suggests, some coordinates of the time T_x may be infinite. More specifically, we have:

Proposition 1. *Assume that ν is irreducible and non degenerate.*

1. *If ν is (sub)critical, then almost surely, for all $x\in\mathbb{Z}_+^d$ and for all $i\in[d]$, $T_x^{(i)}<\infty$.*
2. *If ν is super critical, then for all $x\in\mathbb{Z}_+^d$, with some probability $p>0$, $T_x^{(i)}=\infty$, for all $i\in[d]$ and with probability $1-p$, $T_x^{(i)}<\infty$, for all $i\in[d]$.*

There are instances of reducible distributions ν such that for some $x\in\mathbb{Z}_+^d$, with positive probability, $T_x^{(i)}<\infty$, for all $i\in A$ and $T_x^{(i)}=\infty$, for all $i\in B$, (A,B) being a non trivial partition of $[d]$. It is the case for instance when $d=2$, for $x=(1,1)$, $X^{1,2}=X^{2,1}\equiv 0$, $0<m_{11}<1$ and $m_{22}>1$.

Then we define d-type branching forests with edge lengths as finite sets of independent branching trees with edge lengths. We say that such a forest is issued from $x=(x_1,\dots,x_d)\in\mathbb{Z}_+^d$ (at time $t=0$), if it contains x_i trees whose root is of type i. The discrete skeleton of a branching forest with edge lengths is a discrete branching forest with progeny distribution ν. The edges issued from vertices of type i are exponential random variables with parameter λ_i. These random variables are mutually independent and are independent of the discrete skeleton. A realisation of such a forest is represented in Figure 1. Then to each d-type forest with edge lengths, F, is associated the branching process $Z=(Z^{(1)},\dots,Z^{(d)})$, where $Z^{(i)}$ is the number of individuals in F, alive at time t.

Definition 1. For $i\neq j$, we denote by $Z_t^{i,j}$ the total number of individuals of type j whose parent has type i and who were born before time t. The definition of $Z_t^{i,i}$ for $i\in[d]$ is the same, except that we add the number of roots of type i and we subtract the number of individuals of type i who died before time t.

More formal definitions of branching forests with edge lengths and processes $Z^{i,j}$ will be given in sections 4.2 and 4.3, see in particular (24).

Then we readily check that the branching process $Z=(Z_1,\dots,Z_d)$ which is associated to this forest can be expressed in terms of the processes $Z^{i,j}$, as follows:

$$Z_t^{(j)}=\sum_{i=1}^{d}Z_t^{i,j}\,,\quad j\in[d]\,.$$

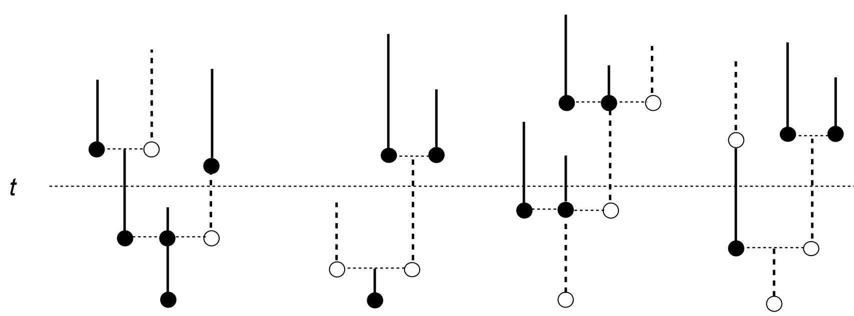

Fig. 1 A two type forest with edge lengths issued from $x=(2,2)$. Vertices of type 1 (resp. 2) are represented in black (resp. white). At time t, $Z_t^{1,1}=1$, $Z_t^{1,2}=2$, $Z_t^{2,1}=3$ and $Z_t^{2,2}=2$.

The next theorem asserts that from a given d-type forest with edge lengths F, issued from $x=(x_1,\dots,x_d)$, we can construct a d-dimensional additive compound Poisson process $\mathrm{X}=(\sum_{i=1}^d X_{t_i}^{i,j},\, j\in[d],\,(t_1,\dots,t_d)\in\mathbb{R}_+^d)$ stopped at its first passage time of $-x$, such that the branching process Z associated to F can be represented as a time change of X. This extends the Lamperti representation to multitype branching processes.

Theorem 2. *Let $x=(x_1,\dots,x_d)\in\mathbb{Z}_+^d$ and let F be a d-type branching forest with edge lengths, issued from x, with progeny distribution ν and reproduction rates λ_i. Then the processes $Z^{i,j}$, $i,j\in[d]$ introduced in Definition 1 admit the following representation:*

$$Z_t^{i,j}=x_i\mathbf{1}_{i=j}+X^{i,j}_{\int_0^t Z_s^{(i)}\,ds}\,,\quad t\ge 0,\tag{5}$$

where

$$X^{(i)}=(X^{i,1},X^{i,2},\dots,X^{i,d})\,,\quad i=1,\dots,d\,,$$

are independent $\mathbb{Z}_+^d$-valued compound Poisson processes with respective rates λ_i and jump distributions μ_i defined in (1), stopped at the first hitting time T_x of $-x$ by the additive compound Poisson process, $\mathrm{X}=\left(\sum_{i=1}^d X^{i,j}_{t_i},\,j\in[d],\,(t_1,\dots,t_d)\in\mathbb{R}_+^d\right)$, that is,

$$X_t^{(i)}\mathbf{1}_{\{t<T_x^{(i)}\}}+(X^{i,1}_{T_x^{(i)}},\dots,X^{i,d}_{T_x^{(i)}})\mathbf{1}_{\{t\ge T_x^{(i)}\}}\,,\quad t\ge 0.$$

In particular the multitype branching process Z, issued from $x=(x_1,\dots,x_d)\in\mathbb{R}_+^d$ admits the following representation,

$$(Z_t^{(1)},\dots,Z_t^{(d)})=x+\left(\sum_{i=1}^d X^{i,1}_{\int_0^t Z_s^{(i)}\,ds},\dots,\sum_{i=1}^d X^{i,d}_{\int_0^t Z_s^{(i)}\,ds}\right),\quad t\ge 0.\tag{6}$$

Moreover, the transformation (5) is invertible, so that the processes $Z^{i,j}$, $i,j\in[d]$ can be recovered from the processes $X^{(i)}$, $i\in[d]$.

Note that in Theorem 2, $T_x^{(i)}$ actually represents the total length of the branches issued from vertices of type i in the forest F.

Conversely, the following theorem asserts that an additive compound Poisson process X being given, we can construct a multitype branching forest whose branching process Z is the unique solution to equation (6).

Theorem 3. *Let $x=(x_1,\dots,x_d)\in\mathbb{Z}_+^d$ and*

$$X^{(i)}=(X^{i,1},X^{i,2},\dots,X^{i,d})\,,\quad i=1,\dots,d\,,$$

be independent $\mathbb{Z}_+^d$ valued compound Poisson processes with respective rates $\lambda_i>0$ and jump distributions μ_i, stopped at the first hitting time T_x of $-x$ by the additive compound Poisson process $(t_1,\dots,t_d)\mapsto\left(\sum_{i=1}^d X^{i,j}_{t_i},\,j\in[d]\right)$. Then there is a branching forest with edge lengths, with progeny distribution ν and reproduction rates $\lambda_i>0$ such that the processes $Z^{i,j}$ of Definition 1 satisfy relation (5). Moreover, the branching process $Z=(Z^{(1)},\dots,Z^{(d)})$ associated to this forest is the unique solution of the equation,

$$(Z_t^{(1)},\dots,Z_t^{(d)})=x+\left(\sum_{i=1}^d X^{i,1}_{\int_0^t Z_s^{(i)}\,ds},\dots,\sum_{i=1}^d X^{i,d}_{\int_0^t Z_s^{(i)}\,ds}\right),\quad t\ge 0\,.$$

We emphasize that Theorems 2 and 3 do not define a bijection between the set of branching forests with edge lengths and this of additive compound Poisson processes, as in the discrete time case, see Section 3. Indeed, in the continuous time case, when constructing the processes $Z^{i,j}$ as in Definition 1, at each birth time, we lose the information of the specific individual who gives birth. In particular, the forest which is constructed in Theorem 3 is not unique. This lost information is preserved in discrete time and the breadth first search coding that is defined in Subsection 3.2 allows us to define a bijection between the set of discrete forests and this of coding sequences.

3 Discrete multitype forests

3.1 The space of multitype forests

We will denote by $\mathscr{F}$ the set of plane forests. More specifically, an element $\mathbf{f} \in \mathscr{F}$ is a directed planar graph with no loops on a possibly infinite and non empty set of vertices $\mathbf{v} = \mathbf{v}(\mathbf{f})$, with a *finite* number of connected components, such that each vertex has a finite inner degree and an outer degree equals to 0 or 1. The elements of $\mathscr{F}$ will simply be called forests.The connected components of a forest are called the *trees*. A forest consisting of a single connected component is also called a tree. In a tree $\mathbf{t}$, the only vertex with outer degree equal to 0 is called the *root* of $\mathbf{t}$. It will be denoted by $r(\mathbf{t})$. The roots of the connected components of a forest $\mathbf{f}$ are called the roots of $\mathbf{f}$. For two vertices u and v of a forest $\mathbf{f}$, if (u,v) is a directed edge of $\mathbf{f}$, then we say that u is a *child* of v, or that v is the *parent* of u. The set $\mathbf{v}(\mathbf{f})$ of vertices of each forest $\mathbf{f}$ will be enumerated according to the usual breadth first search order, see Figure 2. We emphasize that we begin by enumerating the roots of the forest from the left to the right. In particular, our enumeration is not performed tree by tree. If a forest $\mathbf{f}$ contains at least n vertices, then the n-th vertex of $\mathbf{f}$ is denoted by $u_n(\mathbf{f})$. When no confusion is possible, we will simply denote the n-th vertex by u_n.

A d-type forest is a couple $(\mathbf{f}, c_{\mathbf{f}})$, where $\mathbf{f} \in \mathscr{F}$ and $c_{\mathbf{f}}$ is an application $c_{\mathbf{f}} : \mathbf{v}(\mathbf{f}) \to [d]$. For $v \in \mathbf{v}(\mathbf{f})$, the integer $c_{\mathbf{f}}(v)$ is called the *type* (or the *color*) of v. The set of finite d-type forests will be denoted by $\mathscr{F}_d$. An element $(\mathbf{f}, c_{\mathbf{f}}) \in \mathscr{F}_d$ will often simply be denoted by $\mathbf{f}$. We assume that for any $\mathbf{f} \in \mathscr{F}_d$, if $u_i, u_{i+1}, \dots, u_{i+j} \in \mathbf{v}(\mathbf{f})$ have the same parent, then $c_{\mathbf{f}}(u_i) \le c_{\mathbf{f}}(u_{i+1}) \le \dots \le c_{\mathbf{f}}(u_{i+j})$. Moreover, if $u_1, \dots, u_k$ are the roots of $\mathbf{f}$, then $c_{\mathbf{f}}(u_1) \le \dots \le c_{\mathbf{f}}(u_k)$. For each $i \in [d]$ we will denote by $u_n^{(i)} = u_n^{(i)}(\mathbf{f})$ the n-th vertex of type i of the forest $\mathbf{f} \in \mathscr{F}_d$, see Figure 2.

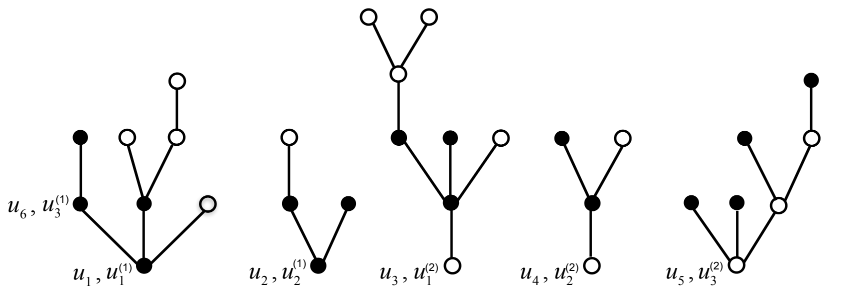

Fig. 2 A two type forest labeled according to the breadth first search order. Vertices of type 1 (resp. 2) are represented in white (resp. black).

3.2 Coding multitype forests

The aim of this subsection is to obtain a bijection between the set of multitype forests and some particular set of integer valued sequences which has been introduced in [5]. This bijection, which will be called a *coding*, depends on the breadth first search ordering defined in the previous subsection. We emphasize that this coding is quite different from the one which is defined in [5].

Definition 2. Let S_d be the set of $\left[\mathbb{Z}^d\right]^d$-valued sequences, $x = (x^{(1)}, x^{(2)}, \dots, x^{(d)})$, such that for all $i \in [d]$, $x^{(i)} = (x^{i,1}, \dots, x^{i,d})$ is a $\mathbb{Z}^d$-valued sequence defined on some interval of integers, $\{0, 1, 2, \dots, n_i\}$, $0 \le n_i \le \infty$, which satisfies $x_0^{(i)} = 0$ and if $n_i \ge 1$, then

(i) for $i \neq j$, the sequence $(x_n^{i,j})_{0 \le n \le n_i}$ is nondecreasing,
$(ii)$for all i, $x_{n+1}^{i,i} - x_n^{i,i} \ge -1$, $0 \le n \le n_i - 1$.

A sequence $x \in S_d$ will sometimes be denoted by $x = (x_k^{i,j}, 0 \le k \le n_i, i, j \in [d])$ and for more convenience, we will sometimes denote $x_k^{i,j}$ by $x^{i,j}(k)$. The vector $\mathrm{n} = (n_1, \dots, n_d) \in \overline{\mathbb{Z}}_+^d$, where $\overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{+\infty\}$ will be called the length of x.

Recall the definition of the order on $\mathbb{R}^d$, from the beginning of Section 2 and let us set $U_{\mathrm{s}} = \{i \in [d] : s_i < \infty\}$, for any $\mathrm{s} \in \overline{\mathbb{Z}}_+^d$. Then the next lemma extends Lemma 2.2 in [5] to the case where the smallest solution of a system such as (r, x) in (7) may have infinite coordinates.

Lemma 1. *Let $x \in S_d$ whose length $\mathrm{n} = (n_1, \dots, n_d)$ is such that $n_i = \infty$ for all i (i.e. $U_{\mathrm{n}} = \emptyset$) and let $\mathrm{r} = (r_1, \dots, r_d) \in \mathbb{Z}_+^d$. Then there exists $\mathrm{s} = (s_1, \dots, s_d) \in \overline{\mathbb{Z}}_+^d$ such that*

$$(\mathrm{r}, x) \qquad r_j + \sum_{i=1}^{d} x^{i,j}(s_i) = 0\,, \quad j \in U_{\mathrm{s}}\,, \tag{7}$$

(we *will say that* s *is a solution of the system* (r,x)) *and such that any other solution* q *of* (r,x) *satisfies* $\mathrm{q}\ge\mathrm{s}$. *Moreover we have* $s_i=\min\{q: x_q^{i,i}=\min_{0\le k\le s_i}x_k^{i,i}\}$, *for all* $i\in U_{\mathrm{s}}$.

The solution s *will be called* the smallest solution *of the system* (r,x). *We emphasize that in* (7), *we may have* $U_{\mathrm{s}}=\emptyset$, *so that according to this definition, the smallest solution of the system* (r,x) *may be infinite, that is* $s_i=\infty$ *for all* $i\in[d]$. *Note also that in* (7) *it is implicit that* $\sum_{i\in[d]\setminus U_{\mathrm{s}}}x^{i,j}(s_i)<\infty$, *for all* $j\in U_{\mathrm{s}}$.

Proof. This proof is based on the simple observation that for fixed j, when at least one of the indices k_i's for $i\neq j$ increases, the term $\sum_{i\neq j}x^{i,j}(k_i)$ may only increase and when k_j increases, the term $x^{j,j}(k_j)$ may decrease only by jumps of amplitude -1.

First recall the notation U_{s}, for $\mathrm{s}\in\overline{\mathbb{Z}}_+^d$ introduced before Lemma 1 and set $v_j^{(1)}=r_j$ and for $n\ge 1$,

$$k_j^{(n)}=\inf\{k: x_k^{j,j}=-v_j^{(n)}\}\ \text{ and }\ v_j^{(n+1)}=r_j+\sum_{i\neq j}x^{i,j}(k_i^{(n)})\,,$$

where $\inf\emptyset=\infty$. Set also $\mathrm{k}^{(0)}=0$ and $U_{\mathrm{k}^{(0)}}=[d]$. Then note that since for $i\neq j$, the $x^{i,j}$'s are positive and increasing, we have

$$\mathrm{k}^{(n)}\le\mathrm{k}^{(n+1)}\ \text{ and }\ U_{\mathrm{k}^{(n+1)}}\subseteq U_{\mathrm{k}^{(n)}}\,,\quad n\ge 0\,.$$

Moreover, for each $n\ge 0$,

$$r_j+\sum_{i\neq j}x^{i,j}(k_i^{(n)})+x^{j,j}(k_j^{(n)})\ge 0\,,\quad j\in U_{\mathrm{k}^{(n)}}\,. \tag{8}$$

Define

$$n_0=\inf\big\{n\ge 1: r_j+\sum_{i\neq j}x^{i,j}(k_i^{(n)})+x^{j,j}(k_j^{(n)})=0\,,\ j\in U_{\mathrm{k}^{(n)}}\big\}\,,$$

where in this definition, we consider that $r_j+\sum_{i\neq j}x^{i,j}(k_i^{(n)})+x^{j,j}(k_j^{(n)})=0$ is satisfied for all $j\in U_{\mathrm{k}^{(n)}}$ if $U_{\mathrm{k}^{(n)}}=\emptyset$. Note that $\mathrm{k}^{(n)}=\mathrm{k}^{(n_0)}$ and $U_{\mathrm{k}^{(n)}}=U_{\mathrm{k}^{(n_0)}}$, for all $n\ge n_0$. The index n_0 can be infinite and in general, we have $\mathrm{k}^{(n_0)}=\lim_{n\to\infty}\mathrm{k}^{(n)}$. Then the smallest solution of the system (r,x) in the sense which is defined in Lemma 1 is $\mathrm{k}^{(n_0)}$.

Indeed, (7) is clearly satisfied for $\mathrm{s}=\mathrm{k}^{(n_0)}$. Then let $\mathrm{q}\in\overline{\mathbb{Z}}_+^d$ satisfying (7), that is

$$r_j+\sum_{i\neq j}x^{i,j}(q_i)+x^{j,j}(q_j)=0\,,\ j\in U_{\mathrm{q}}\,. \tag{9}$$

We can prove by induction that $\mathrm{q}\ge\mathrm{k}^{(n)}$, for all $n\ge 1$. Firstly for (9) to be satisfied, we should have $q_j\ge\inf\{k: x^{j,j}(k)=-r_j\}$, for all $j\in U_{\mathrm{q}}$, hence $\mathrm{q}\ge\mathrm{k}^{(1)}$.

Now assume that $\mathrm{q} \ge \mathrm{k}^{(n)}$. Then $U_{\mathrm{q}} \subseteq U_{\mathrm{k}^{(n)}}$ and from (8) and (9) for each $j \in U_{\mathrm{q}}$, $q_j \ge \inf\{k : x^{j,j}(k) = -(r_j + \sum_{i\neq j} x^{i,j}(k_i^{(n)}))\}$, hence $\mathrm{q} \ge \mathrm{k}^{(n+1)}$.

Finally the fact that $s_i = \min\{q : x_q^{i,i} = \min_{0\le k\le s_i} x_k^{i,i}\}$, for all $i \in U_{\mathrm{s}}$ readily follows from the above construction of s_i. $\square$

Let $(\mathbf{f}, c_{\mathbf{f}}) \in \mathscr{F}_d$, $u \in \mathbf{v}(\mathbf{f})$ and denote by $p_i(u)$ the number of children of type i of u. For each $i \in [d]$, let $n_i \ge 0$ be the number of vertices of type i in $\mathbf{v}(\mathbf{f})$. Then we define the application Ψ from $\mathscr{F}_d$ to S_d by

$$\Psi(\mathbf{f}, c_{\mathbf{f}}) = x\,, \tag{10}$$

where $x = (x^{(1)}, \dots, x^{(d)})$ and for all $i \in [d]$, $x^{(i)}$ is the d-dimensional chain $x^{(i)} = (x^{i,1}, \dots, x^{i,d})$, with length n_i, whose values belong to the set $\mathbb{Z}^d$, such that $x_0^{(i)} = 0$ and if $n_i \ge 1$, then

$$x_{n+1}^{i,j} - x_n^{i,j} = p_j(u_{n+1}^{(i)})\,, \text{ if } i \neq j \text{ and } x_{n+1}^{i,i} - x_n^{i,i} = p_i(u_{n+1}^{(i)}) - 1\,, \quad 0 \le n \le n_i - 1\,. \tag{11}$$

We recall that $u_n^{(i)}$ is the n-th vertex of type i in the breadth first search order of $\mathbf{f}$. We will see in Theorem 4 that $(n_1, \dots, n_d)$ is actually the smallest solution of the system (r, x), where r_i is the number of roots of type i of the forest $\mathbf{f}$. This leads us to the following definition.

Definition 3. Fix $\mathrm{r} = (r_1, \dots, r_d) \in \mathbb{Z}_+^d$, such that $\mathrm{r} > 0$.

(i) We denote by Σ_d^{r} the subset of S_d of sequences x with length $\mathrm{n} = (n_1, \dots, n_d) \in \overline{\mathbb{Z}}_+^d$ such that n is the smallest solution of the system $(\mathrm{r}, \bar{x})$, where for all $i \in [d]$, $\bar{x}_k^{(i)} = x_k^{(i)}$, if $k \le n_i$ and $\bar{x}_k^{(i)} = x_{k_i}^{(i)}$, if $k \ge n_i$. We will also say that n is the smallest solution of the system (r, x).

(ii) We denote by $\mathscr{F}_d^{\mathrm{r}}$, the subset of $\mathscr{F}_d$ of d-type forests containing exactly r_i roots of type i, for all $i \in [d]$.

The following theorem gives a one to one correspondence between the sets $\mathscr{F}_d^{\mathrm{r}}$ and Σ_d^{r}.

Theorem 4. *Let* $\mathrm{r} = (r_1, \dots, r_d) \in \mathbb{Z}_+^d$*, be such that* $\mathrm{r} > 0$*. Then for all* $(\mathbf{f}, c_{\mathbf{f}}) \in \mathscr{F}_d^{\mathrm{r}}$*, the chain* $x = \Psi(\mathbf{f}, c_{\mathbf{f}})$ *belongs to the set* Σ_d^{r}*. Moreover, the mapping*

$$\begin{aligned} \Psi : \mathscr{F}_d^{\mathrm{r}} &\to \Sigma_d^{\mathrm{r}} \\ (\mathbf{f}, c_{\mathbf{f}}) &\mapsto \Psi(\mathbf{f}, c_{\mathbf{f}}) \end{aligned}$$

is a bijection.

Proof. In this proof, in order to simplify the notation, we will identify the sequence x with its extension $\bar{x}$ introduced in Definition 3.

Let $(\mathbf{f}, c_{\mathbf{f}})$ be a forest of $\mathscr{F}_d^{\mathrm{r}}$ and let $\mathrm{s} = (s_1, \dots, s_d) \in \overline{\mathbb{Z}}_+^d$, where s_i is the number of vertices of type i in $\mathbf{f}$. We define a subtree of type $i \in [d]$ of $(\mathbf{f}, c_{\mathbf{f}})$ as a maximal connected subgraph of $(\mathbf{f},c_{\mathbf{f}})$ whose all vertices are of type i. Formally, $\mathbf{t}$ is a subtree of type i of $(\mathbf{f},c_{\mathbf{f}})$, if it is a connected subgraph whose all vertices are of type i and such that either $r(\mathbf{t})$ has no parent or the type of its parent is different from i. Moreover, if the parent of a vertex $v\in\mathbf{v}(\mathbf{t})^c$ belongs to $\mathbf{v}(\mathbf{t})$, then $c_{\mathbf{f}}(v)\neq i$.

Let $i\in[d]$ and assume first that $s_i<\infty$ (i.e. $i\in U_{\mathrm{s}}$) and let $k_i\le s_i$ be the number of subtrees of type i in $\mathbf{f}$. Then we can check that the length s_i of the sequence $x^{i,i}$ corresponds to its first hitting time of level $-k_i$, i.e.

$$s_i=\inf\{n:x^{i,i}_n=-k_i\}\,. \tag{12}$$

Indeed, let us rank the subtrees of type i in $\mathbf{f}$ according to the breadth first search order of their roots, so that we obtain the subforest of type i: $\mathbf{t}_1,\dots,\mathbf{t}_{k_i}$ and let x' be its Lukasiewicz-Harris coding path, that is $x'_0=0$ and

$$x'_{n+1}-x'_n=p(u_{n+1})-1\,,\quad 0\le n\le s_i-1\,,$$

where u_n is the n-th vertex in the breadth first search of this subforest and $p(u_n)$ is its number of children. We refer to [9] for the coding of forests through their Lukasiewicz-Harris coding path. Then we readily check that both sequences have the same length and end up at the same level, i.e.

$$\inf\{n:x'_n=-k_i\}=\inf\{n:x^{i,i}_n=-k_i\}\,. \tag{13}$$

If $s_i=\infty$ then either $k_i=\infty$ and (12) holds, or $k_i<\infty$ and at least one of the subtrees of type i in the forest is infinite. In this case, we can still compare $x^{i,i}$ to the Lukasiewicz-Harris coding path x' in order to obtain (13), so that (12) also holds.

Now let us check that s is a solution of the system (r,x), that is

$$r_j+\sum_{i=1}^{d}x^{i,j}(s_i)=0\,,\quad j\in U_{\mathrm{s}}\,. \tag{14}$$

Let $j\in U_{\mathrm{s}}$, then $r_j+\sum_{i\neq j}x^{i,j}(s_i)$ clearly represents the total number of vertices of type j in $\mathbf{v}(\mathbf{f})$ which are either a root of type j or whose parent is of a type different from j, i.e. it represents the total number of subtrees of type j in $\mathbf{f}$. On the other hand, from (12), $-x^{j,j}(s_j)\ge 0$ is the number of these subtrees. We conclude that equation (14) is satisfied.

It remains to check that s is the smallest solution of the system (r,x). As in Lemma 1, set $\mathrm{k}^{(0)}=0$ and for all $j\in[d]$,

$$k^{(n)}_j=\inf\big\{k:x^{j,j}_k=-(r_j+\sum_{i\neq j}x^{i,j}(k^{(n-1)}_i))\big\}\,,\ n\ge 1\,. \tag{15}$$

Then from the proof of Lemma 1, we have to prove that $\mathrm{s}=\lim_{\to\infty}\mathrm{k}^{(n)}$. Recall the coding which is defined in (11). For all $j\in[d]$, once we have visited the r_j first vertices of type j which are actually roots of the forest, we have to visit the whole corresponding subtrees, so that, if the total number of vertices of type j in $(\mathbf{f},c_{\mathbf{f}})$ is finite, i.e. $j\in U_{\rm s}$, then the chain $x^{j,j}$ first hits $-r_j$ at time $k_j^{(1)}<\infty$. Then at time $k_j^{(1)}$, an amount of $\sum_{i\neq j}x^{i,j}(k_i^{(1)})$ more subtrees of type j have to be visited. So the chain $x^{j,j}$ has to hit $-(r_j+\sum_{i\neq j}x^{i,j}(k_i^{(1)}))$ at time $k^{(2)}<\infty$. This procedure is iterated until the last vertex of type j is visited, that is until time $s_j=\lim_{n\to\infty}k_j^{(n)}<\infty$ (note that the sequence $k_j^{(n)}$ is constant after some finite index). Besides, from (15), we have

$$s_j=\inf\{k:x_k^{j,j}=-(r_j+\sum_{i\neq j}x^{i,j}(s_i))\}\,,\;\;n\ge 1\,,\;\;j\in U_{\rm s}\,.$$

On the other hand, if the total number of vertices of type j in $(\mathbf{f},c_{\mathbf{f}})$ is infinite, then $k_j^{(n)}$ tends to ∞ (it can be infinite by some rank). So that we also have $s_j=\lim_{n\to\infty}k_j^{(n)}$ in this case. Therefore, s is the smallest solution of $({\rm r},x)$.

Now let $x\in\Sigma_d^{\rm r}$ with length s, then we construct a forest $(\mathbf{f},c_{\mathbf{f}})\in\mathscr{F}_d^{\rm r}$ such that $\Psi(\mathbf{f},c_{\mathbf{f}})=x$, generation by generation, using the definition of Ψ in (10) as follows. At generation $n=1$, for each $i\in[d]$, we take r_i vertices of type i. We rank these $r_1+\ldots+r_d$ vertices as it is defined in Subsection 3.1. Then to the k-th vertex of type i, we give $\Delta x_k^{i,j}:=x_k^{i,j}-x_{k-1}^{i,j}$ children of type $j\in[d]$ if $j\neq i$ and $\Delta x_k^{i,i}+1$ childr en of type i. We rank vertices of generation $n=2$ and to the r_i+k-th vertex of type i, we give $\Delta x_{r_i+k}^{i,j}$ children of type $j\in[d]$, if $j\neq i$ and $\Delta x_{r_i+k}^{i,i}+1$ children of type i, and this procedure is repeated generation by generation. Proceeding this way, until the steps $\Delta x_{s_i}^{i,j}$, $i,j\in[d]$, we have constructed a forest of $\mathscr{F}_d^{\rm r}$. Indeed the total number of children of type j whose parent is of type $i\neq j$ is $x^{i,j}(s_i)$, hence, the total number of children of type j which is a root or whose parent is different from j is $r_j+\sum_{i\neq j}x^{i,j}(s_i)$. Moreover, each branch necessarily ends up with a leaf, since $\Delta x_{s_i+1}^{i,j}=0$, for all $i\neq j$ and $\Delta x_{s_i+1}^{i,i}=-1$. Therefore we have constructed a forest $(\mathbf{f},c_{\mathbf{f}})$ of $\mathscr{F}_d^{\rm r}$ such that $\Psi(\mathbf{f},c_{\mathbf{f}})=x$ and we have proved that Φ is onto.

Finally, let us denote the forest $(\mathbf{f},c_{\mathbf{f}})$ which is reconstructed from x through the above procedure by $\Phi(x)$, then the application Φ is actually the inverse of Ψ. Indeed, we can check from the defintion of Φ above and that of Ψ in (10) that for all $(\mathbf{f},c_{\mathbf{f}})\in\mathscr{F}_d^{\rm r}$, $\Phi(\Psi((\mathbf{f},c_{\mathbf{f}})))=(\mathbf{f},c_{\mathbf{f}})$. Therefore Ψ is one-to-one. $\square$

3.3 Representing the sequence of generation sizes

To each $\mathbf{f}\in\mathscr{F}_d^{\rm r}$ we associate the chain $z=(z^{(1)},\ldots,z^{(d)})$ indexed by the successive generations in $\mathbf{f}$, where for each $i\in[d]$, and $n\ge 1$, $z^{(i)}(n)$ is the size of the population of type i at generation n. More formally, we say that the (index of the) generation of $u\in\mathbf{v}(\mathbf{f})$ is n if $d(r(\mathbf{t}_u),u)=n$, where $\mathbf{t}_u$ is the tree of $\mathbf{f}$ which contains u, $r(\mathbf{t}_u)$ is the root of this tree and d is the usual distance in discrete trees. In order to simplify the notation, we set $|u|=d(r(\mathbf{t}_u),u)$. Let us denote by $h(\mathbf{f})$ the index of the highest generation in $\mathbf{f}$. Then $z^{(i)}$ is defined by

$$z^{(i)}(n) = \begin{cases} \mathrm{Card}\{u \in \mathbf{v}(\mathbf{f}) : c_{\mathbf{f}}(u) = i,\, |u| = n\} & \text{if } n \le h(\mathbf{f}), \\ 0 & \text{if } n \ge h(\mathbf{f}) + 1. \end{cases} \tag{16}$$

We also define the chains $z^{i,j}$, for $i, j \in [d]$, as follows: $z^{i,j}(0) = 0$ if $i \neq j$, $z^{i,i}(0) = r_i$, and for $n \ge 1$,

$$z^{i,j}(n) = r_i \mathbf{1}_{i=j} + \sum_{|u| \le n-1} \mathbf{1}_{c_{\mathbf{f}}(u)=i} \left(p_j(u) - \mathbf{1}_{i=j}\right). \tag{17}$$

In words, if $i \neq j$ then $z^{i,j}(n)$ is the total number of vertices of type j whose parent is of type i in the first n generations of the forest $\mathbf{f}$. If $i = j$ then we only count the number of vertices of type i with at least one younger brother of type i and whose parent is of type i in the n first generations. To this number, we subtract the number of vertices of type i with no children of type i, whose generation is less or equal than $n-1$. Then it is not difficult to check the following relation:

$$z^{(j)}(n) = \sum_{i=1}^{d} z^{i,j}(n)\,, n \ge 0\,, \quad j \in [d]. \tag{18}$$

We end this subsection by a lemma which provides a relationship between the chains $x^{i,j}$ and $z^{i,j}$, where $x = \Psi(\mathbf{f}, c_{\mathbf{f}})$. This result is the deterministic expression of the Lamperti representation of Theorem 5 below and its continuous time counterpart in Theorems 2 and 3.

Lemma 2. *The chain $z^{i,j}$ may be obtained as the following time change of the chain $x^{i,j}$:*

$$z^{i,j}(n) = x^{i,j}\big(\textstyle\sum_{k=0}^{n-1} z^{(i)}(k)\big)\,, \quad n \ge 1\,, \quad i, j \in [d]\,,\ i \neq j\,, \tag{19}$$

$$z^{i,i}(n) = r_i + x^{i,i}\big(\textstyle\sum_{k=0}^{n-1} z^{(i)}(k)\big)\,, \quad n \ge 1\,, \quad i \in [d]\,. \tag{20}$$

In particular, we have

$$z^{(j)}(n) = r_j + \sum_{i=1}^{d} x^{i,j}\big(\textstyle\sum_{k=0}^{n-1} z^{(i)}(k)\big)\,, \quad n \ge 1\,, \quad j \in [d]\,. \tag{21}$$

Moreover, given $x^{i,j}$, $i, j \in [d]$, the chains $z^{i,j}$, $i, j \in [d]$ are uniquely determined by equations (18)*,* (19) *and* (20)*.*

Proof. It suffices to check relations (19) and (20). Then (21) will follow from (18). But (19) and (20) are direct consequences of the definition of the chains $x^{i,j}$ and $z^{i,j}$ in (11) and (17) respectively. $\square$

3.4 Application to discrete time branching processes

Recall that $\nu = (\nu_1, \dots, \nu_d)$ is a progeny distribution, such that the ν_i's are any probability measures on $\mathbb{Z}_+^d$, such that $\nu_i(e_i) < 1$. Let $(\Omega, \mathscr{G}, P)$ be some measurable space on which, for any $\mathrm{r} \in \mathbb{Z}_+^d$ such that $\mathrm{r} > 0$, we can define a probability measure $\mathbb{P}_\mathrm{r}$ and a random variable $(F, c_F) : (\Omega, \mathscr{G}, \mathbb{P}_\mathrm{r}) \to \mathscr{F}_d^\mathrm{r}$ whose law under $\mathbb{P}_\mathrm{r}$ is this of a branching forest with progeny law ν. Then we construct from (F, c_F) the random chains, $X = (X^{(1)}, \dots, X^{(d)})$, $Z = (Z^{(1)}, \dots, Z^{(d)})$ and $Z^{i,j}$, $i, j \in [d]$, exactly as in (11), (16) and (17), respectively. In particular, $X = \Psi(F, c_F)$. We can check that under $\mathbb{P}_\mathrm{r}$, Z is a branching process with progeny distribution ν. More specifically, recall from (2) the definition of $\tilde{\nu}_i$, then the random processes X and Z satisfy the following result.

Theorem 5. *Let* $\mathrm{r} \in \mathbb{Z}_+^d$ *be such that* $\mathrm{r} > 0$ *and let* (F, c_F) *be an* $\mathscr{F}_d^\mathrm{r}$*-valued branching forest with progeny law* ν *under* $\mathbb{P}_\mathrm{r}$*. Let* $N = (N_1, \dots, N_d) \in \overline{\mathbb{Z}}_+^d$*, where* N_i *is the number of vertices of type* i *in* F*.Then,*

1. *The random variable* N *is almost surely the smallest solution of the system* (r, X) *in the sense of Definition* 3*. If* ν *is irreducible, non degenerate and* (sub)*critical, then almost surely,* $N_i < \infty$*, for all* $i \in [d]$*. If* ν *is irreducible, non degenerate and supercritical, then with some probability* $p > 0$*,* $N_i = \infty$*, for all* $i \in [d]$ *and with probability* $1 - p$*,* $N_i < \infty$*, for all* $i \in [d]$*.*
2. *On the space* $(\Omega, \mathscr{G}, P)$*, we can define independent random walks* $\tilde{X}^{(i)}$*,* $i \in [d]$*, with respective step distributions* $\tilde{\nu}_i$*,* $i \in [d]$*, such that* $\tilde{X}_0^{(i)} = 0$ *and if* $\tilde{N} = (\tilde{N}_1, \dots \tilde{N}_d) \in \overline{\mathbb{Z}}_+^d$ *is the smallest solution of the system* $(\mathrm{r}, \tilde{X})$*, then the following identity in law*
$$(X_k^{(i)}, 0 \le k \le N_i, i \in [d]) \stackrel{(d)}{=} (\tilde{X}_k^{(i)}, 0 \le k \le \tilde{N}_i, i \in [d])$$
holds.
3. *The joint law of* X_N *and* N *is given as follows: for any integers* n_i *and* k_{ij}*,* $i, j \in [d]$*, such that* $n_i > 0$*,* $k_{ij} \in \mathbb{Z}_+$*, for* $i \neq j$*,* $-k_{jj} = r_j + \sum_{i \neq j} k_{ij}$ *and* $n_i \ge -k_{ii}$*,*
$$\mathbb{P}_\mathrm{r}\Big(X_{n_i}^{i,j} = k_{ij}, \, i, j \in [d] \text{ and } N = \mathrm{n}\Big) = \\ \frac{\det(-k_{ij})}{n_1 n_2 \dots n_d} \prod_{i=1}^d \nu_i^{*n_i}\Big(k_{i1}, \dots, k_{i(i-1)}, n_i + k_{ii}, k_{i(i+1)}, \dots, k_{id}\Big).$$
4. *The random process* Z *is a branching process with progeny law* ν*, which is related to* X *through the time change:*
$$Z^{(i)}(n) = \sum_{i=1}^d X^{i,j}\big(\textstyle\sum_{k=1}^{n-1} Z^{(i)}(k)\big), \quad n \ge 1. \tag{22}$$

Proof. The fact that N is the smallest solution of the system (r,X) is a direct consequence of Theorem 4 and the definition of X, that is $\Psi(F,c_F)=X$. Assume that ν is irreducible, non degenerate and (sub)critical. Then since the forest F contains almost surely a finite number of vertices, all coordinates of N must be finite from Theorem 4. If ν is irreducible, non degenerate and supercritical, then with probability $p>0$ the forest F contains an infinite number of vertices of type i, for all $i\in[d]$ and with probability $1-p$ its total population is finite. Then the result also follows from Theorem 4.

In order to prove part 2, let (F_n,c_{F_n}) with $(F_1,c_{F_1})=(F,c_F)$, be a sequence of independent and identically distributed forests. Let us define $X^n=\Psi(F_n,c_{F_n})$ and then let us concatenate the processes $X^n=(X^{n,(1)},\dots,X^{n,(d)})$ in a process that we denote $\tilde X$. More specifically, let us denote by N_i^n the length of $X^{n,(i)}$, then the process obtained from this concatenation is $\tilde X=(\tilde X^{(1)},\dots,\tilde X^{(d)})$, where $\tilde X_0^{(i)}=0$, $N_i^0=0$ and

$$\tilde X_k^{(i)}=\tilde X^{(i)}_{N_i^0+\dots+N_i^{n-1}}+X^{n,(i)}_{k-(N_i^0+\dots+N_i^{n-1})}\,,\qquad \text{if } N_i^0+\ldots+N_i^{n-1}\le k\le N_i^0+\ldots+N_i^n\,,\ \ n\ge 1\,.$$

Note that $\tilde X$ is obtained by coding the forests (F_n,c_{F_n}), $n\ge1$ successively. Then it readily follows from the construction of $\tilde X$ and the branching property that the coordinates $\tilde X^{(i)}$ are independent random walks with step distribution $\tilde\nu_i$. Moreover N is a solution of the system $(\mathrm{r},\tilde X)$, so its smallest solution, say N', is necessarily smaller than N. This means that N' is a solution of the system (r,X), hence $N'=N$.

The third part is a direct consequence of the first part and the multivariate ballot Theorem which is proved in [5], see Theorem 3.4 therein.

Then part 4, directly follows from the definition of Z and Lemma 2. $\square$

Conversely, from any random walk whose step distribution is given by (2), we can construct a unique branching forest with law $\mathbb{P}_\mathrm{r}$. The following result is a direct consequence of Theorems 4 and 5.

Theorem 6. *Let $\tilde X^{(i)}$, $i\in[d]$ be d independent random walks defined on $(\Omega,\mathscr{G},P)$, whose respective step distributions are $\tilde\nu_i$, and set $\tilde X=(\tilde X^{(1)},\dots,\tilde X^{(d)})$. Let $\mathrm{r}\in\mathbb{Z}_+^d$ such that $\mathrm{r}>0$ and let $\tilde N$ be the smallest solution of the system $(\mathrm{r},\tilde X)$. We define the Σ_d^r-valued process $X=(X^{(1)},\dots,X^{(d)})$ by $(X_k^{(i)},0\le k\le\tilde N_i)=(\tilde X_k^{(i)},0\le k\le\tilde N_i)$.*

Then $(F,c_F):=\Psi^{-1}(X)$ is a $\mathscr{F}_d^\mathrm{r}$-valued branching forest (F,c_F) with progeny distribution ν.

4 The continuous time setting

4.1 Proofs of Theorem 1 and Proposition 1.

Let $Y=(Y^{(1)},\dots,Y^{(d)})$ be the underlying random walk of the compound Poisson process X, that is the random walk such that there are independent Poisson processes $N^{(i)}$, with parameters λ_i such that $X^{(i)}_t=Y^{(i)}(N^{(i)}_t)$ and $(N^{(i)},Y^{(j)},i,j\in[d])$ are independent. Then from Lemma 1, there is a random time $\mathrm{s}\in\overline{\mathbb{Z}}^d_+$, such that almost surely, for all $j\in U_{\mathrm{s}}$, $x_j+\sum_{i=1}^d Y^{i,j}(s_i)=0$. Moreover, if s' is any time satisfying this property, then $\mathrm{s}'\ge\mathrm{s}$. For $i\in[d]\setminus U_{\mathrm{s}}$, the latter equality implies that $Y^{i,j}(\infty)<\infty$. Since $Y^{i,j}$ is a renewal process, it is possible only if $Y^{i,j}\equiv 0$, a.s., so that we can write: almost surely,

$$x_j+\sum_{i,i\in U_{\mathrm{s}}}Y^{i,j}(s_i)=0\,,\quad\text{for all } j\in U_{\mathrm{s}}\,.$$

Then the first part of the Theorem follows from the construction of X as a time change of Y. More formally, the coordinates of T_x are given by $T^{(i)}_x=\inf\{t: N^{(i)}(t)=\mathrm{s}_i\}$, so that in particular, $s_i=N^{(i)}(T^{(i)}_x)$.

Additivity property of T_x is a consequence of Lemma 1 and time change. From this lemma, we can deduce that for all $x,y\in\mathbb{Z}^d_+$, if s is the smallest solution of $(x+y,Y)$, then conditionally to $s_i<\infty$, for all $i\in[d]$, the smallest solution $\mathrm{s}_1=(s_{1,1},\dots,s_{1,d})$ of (x,Y), and satisfies $\mathrm{s}_1\le\mathrm{s}$ and $s-s_1$ is the smallest solution of the system $(y,\tilde Y)$, where $\tilde Y^{(i)}_k=Y^{(i)}_{s_{1,i}+k}-Y^{(i)}_{s_{1,i}}$, $k\ge 0$. Moreover, $\tilde Y=(\tilde Y^{(i)},i\in[d])$ has the same law as Y and is independent of $(Y^{(i)}_k,0\le k\le s_{1,i})$. Using the time change, we obtain,

$$L_{x+y}\overset{(d)}{=}L_x+\tilde L_y\,,$$

where L_x has the law of T_x conditionally on $T^{(i)}_x<\infty$, for all $i\in[d]$ and $\tilde L_y$ is an independent copy of L_y. Then identity (4) follows.

The law of T_x on $\mathbb{R}^d_+$ follows from time change and the same result in the discrete time case obtained in [5], see Theorem 3.4 therein.

Proposition 1 is a direct consequence of part 1 of Theorem 1 and the time change.

4.2 Multitype forests with edge lengths

A d type forest with edge lengths is an element $(\mathbf{f},c_{\mathbf{f}},\ell_{\mathbf{f}})$, where $(\mathbf{f},c_{\mathbf{f}})\in\mathscr{F}_d$ and $\ell_{\mathbf{f}}$ is some application $\ell_{\mathbf{f}}:\mathbf{v}(\mathbf{f})\to(0,\infty)$. For $u\in\mathbf{v}(\mathbf{f})$, the quantity $\ell_{\mathbf{f}}(u)$ will be called the life time of u. It corresponds to the length of an edge incident to u in $(\mathbf{f},c_{\mathbf{f}},\ell_{\mathbf{f}})$ whose color is this of u. This edge is a segment which is closed at the extremity corresponding to u and open at the other extremity. If u is not a leaf of

$(\mathbf{f},c_\mathbf{f})$ then $\ell_\mathbf{f}(u)$ corresponds to the length of the edge between u and its children in the continuous forest $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$. To each tree of $(\mathbf{f},c_\mathbf{f})$ corresponds a tree of $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ which is considered as a continuous metric space, the distance being given by the Lebesgue measure along the branches. To each forest $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ we associate a time scale such that a vertex u is born at time t if the distance between u and the root of its tree in $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ is t. Time t is called the birth time of u in $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ and it is denoted by $b_\mathbf{f}(u)$. The death time of u in $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ is then $b_\mathbf{f}(u)+\ell_\mathbf{f}(u)$. If $s\in[b_\mathbf{f}(u),b_\mathbf{f}(u)+\ell_\mathbf{f}(u))$ then we say that u is alive at time s in the forest $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$. We denote by $\mathrm{h}_\mathbf{f}$ the smallest time when no individual is alive in $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$.

The set of d type forests with edge lengths will be denoted by F_d. The subset of F_d of elements $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ such that $(\mathbf{f},c_\mathbf{f})\in\mathscr{F}_d^{\mathrm{r}}$ will be denoted by F_d^{r}. Elements of F_d will be represented as in Figure 1.

To each forest $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})\in F_d^{\mathrm{r}}$, we associate the multidimensional the step functions, $(z^{(i)}(t),t\ge 0)$ that are defined as follows:

$$z^{(i)}(t)=\mathrm{Card}\{u\in\mathbf{v}(\mathbf{f}):c_\mathbf{f}(u)=i,\,u\text{ is alive at time }t\text{ in }(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f}).\}\tag{23}$$

Then the process $(z^{i,j}(t),t\ge 0)$ introduced in Definition 1 is formally defined by $z^{i,j}(0)=0$ if $i\neq j$, $z^{i,i}(0)=r_i$, and for $t>0$,

$$z^{i,j}(t)=r_i\mathbf{1}_{i=j}+\sum_{b_\mathbf{f}(u)+\ell_\mathbf{f}(u)<t}\mathbf{1}_{c_\mathbf{f}(u)=i}\,(p_j(u)-\mathbf{1}_{i=j})\,.\tag{24}$$

It readily follows from these definitions that

$$z^{(j)}(t)=\sum_{i=1}^d z_t^{i,j}\,,\quad t\ge 0\,.$$

We now define the discretisation of forests of F_d, with some span $\delta>0$. Let $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})\in F_d$, then on each tree of $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})\in F_d$, we place new vertices at distance $n\delta$, $n\in\mathbb{Z}_+$ of the root along all the branches. A vertex which is placed along an edge with color i has also color i. Then we define a forest in $\mathscr{F}_d$ as follows. A new vertex v is the child of a new vertex u if and only if both vertices belong to the same branch of $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$, and there is $n\ge 0$ such that u and v are respectively at distance $n\delta$ and $(n+1)\delta$ from the root. This transformation defines an application which we will denote by

$$\begin{aligned}D_\delta:F_d&\to\mathscr{F}_d\\(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})&\mapsto D_\delta(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})\,.\end{aligned}$$

Note that with this definition, the roots of the three forests $(\mathbf{f},c_\mathbf{f})$, $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ and $D_\delta(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ are the same and more generally, a vertex of $D_\delta(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ corresponds to a vertex u of $(\mathbf{f},c_\mathbf{f},\ell_\mathbf{f})$ if and only if u is at a distance equal to $n\delta$ from the root, for some integer $n\ge 0$. It is also equivalent to the fact that u is at generation n in

$D_\delta(\mathbf{f}, c_\mathbf{f}, \ell_\mathbf{f})$. The definition of the discretisation of a forest with edge lengths should be obvious from Figure 3.

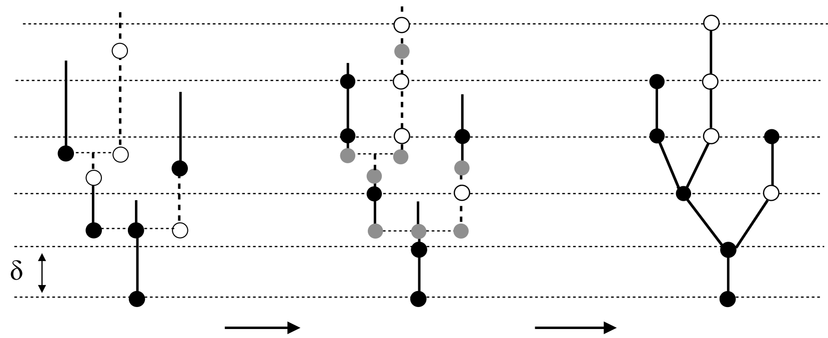

Fig. 3 Discretisation of a two type tree with edge lengths: from $(\mathbf{f}, c_\mathbf{f}, \ell_\mathbf{f})$ to $D_\delta(\mathbf{f}, c_\mathbf{f}, \ell_\mathbf{f})$.

4.3 Multitype branching forests with edge lengths.

Recall from Section 2 that $\lambda_1, \dots, \lambda_d$ are positive, finite and constant rates, and that $\nu = (\nu_1, \dots, \nu_d)$, where ν_i, $i \in [d]$ are any distributions in $\mathbb{Z}_+^d$ such that $\nu_i(e_i) < 1$. From the setting established in Subsection 4.2, we can define a branching forest with edge lengths as a random variable $(F, c_F, \ell_F) : (\Omega, \mathscr{G}, \mathbb{P}_\mathrm{r}) \to (F_d, \mathscr{H}_d)$, where $\mathscr{H}_d$ is the sigma field of the Borel sets of F_d endowed with the Gromov-Hausdorff topology (see Section 2.1 in [9]) and where the law of $(F, c_F) : (\Omega, \mathscr{G}, \mathbb{P}_\mathrm{r}) \to \mathscr{F}_d^\mathrm{r}$ under $\mathbb{P}_\mathrm{r}$ is this of a discrete branching forest with progeny distribution ν, as defined in Subsection 3.4. Besides, let N_i be the number of vertices of type i in (F, c_F), then for all $\mathrm{n} = (n_1, \dots, n_d) \in \overline{\mathbb{Z}}_+^d$, conditionally on $N_i = n_i$, $i \in [d]$, $(\ell_F(u_n^{(i)}))_{0 \le n \le n_i}$ are sequences of i.i.d. exponentially distributed random variables with respective parameters λ_i, and $[(\ell_F(u_n^{(i)}))_{0 \le n \le n_i}, i \in [d], (F, c_F)]$ are independent.

Then we have the following result which is straightforward from the above definitions.

Proposition 2. *Let (F, c_F, ℓ_F) be a branching forest with edge lengths with progeny distribution $\nu = (\nu_1, \dots, \nu_d)$ on $\mathbb{Z}_+^d$ and reproduction rates $\lambda_1, \dots, \lambda_d \in (0, \infty)$. Then for all* r*, under $\mathbb{P}_\mathrm{r}$, the process $Z = (Z^{(i)}(t), t \ge 0, i \in [d])$ which is defined as in* (23) *with respect to (F, c_F, ℓ_F) is a continuous time, $\mathbb{Z}_+^d$-valued branching process starting at $Z_0 = \mathrm{r}$, with progeny distribution $\nu = (\nu_1, \dots, \nu_d)$ and reproduction rates $\lambda_1, \dots, \lambda_d$.*

The law of a discretised branching forest with edge lenghts is given by the following lemma.

Lemma 3. *Let (F, c_F, ℓ_F) be a branching forest with edge lengths, with progeny distribution $\nu = (\nu_1, \dots, \nu_d)$ on $\mathbb{Z}_+^d$ and life time rates $\lambda_1, \dots, \lambda_d \in (0, \infty)$. For $\delta > 0$, the forest $D_\delta(F, c_F, \ell_F)$ is a (discrete) branching forest with progeny distribution:*

$$\nu_i^{(\delta)}(\mathrm{k}) = \mathbb{P}_{e_i}(Z_\delta = \mathrm{k}), \quad \mathrm{k} = (k_1, \dots, k_d) \in \mathbb{Z}_+^d,$$

where e_i is the i-th unit vector of $\mathbb{Z}_+^d$.

Proof. The fact that $D_\delta(F, c_F, \ell_F)$ is a discrete branching forest is a direct consequence of its construction. Indeed, at generation n, that is at time $n\delta$, the vertices of this forest inherits from the (time homogeneous) branching property of (F, c_F, ℓ_F) the fact they will give birth to some progeny, independently of each other and with some distribution which only depends on their type and δ. Then it remains to determine the progeny distribution $\nu^{(\delta)}$. But it is obvious from the construction of $D_\delta(F, c_F, \ell_F)$ that $\nu_i^{(\delta)}$ is the law of the total offspring at time δ of a root of type i in the forest (F, c_F, ℓ_F). This is precisely the expression which is given in the statement. $\square$

4.4 Proof of Theorem 2.

Let (F, c_F, ℓ_F) be a branching forest issued from x, with edge lengths, with progeny distribution $\nu = (\nu_1, \dots, \nu_d)$ on $\mathbb{Z}_+^d$ and life time rates $\lambda_1, \dots, \lambda_d \in (0, \infty)$. Then from Proposition 2, the process $Z = (Z^{(i)}(t), t \geq 0, i \in [d])$ which is defined as in (23) with respect to (F, c_F, ℓ_F) is a continuous time, $\mathbb{Z}_+$-valued branching process with progeny distribution $\nu = (\nu_1, \dots, \nu_d)$ and life time rates $\lambda_1, \dots, \lambda_d$.

Let $\delta > 0$ and consider the discrete forest $D_\delta(F, c_F, \ell_F)$ whose progeny distribution is given by Lemma 3. Let $\mathrm{Z}^\delta = (\mathrm{Z}^{\delta,(1)}, \dots, \mathrm{Z}^{\delta,(d)})$ be the associated (discrete time) branching process and let $\mathrm{Z}^{\delta,i} := (\mathrm{Z}^{\delta,i,1}, \dots, \mathrm{Z}^{\delta,i,d})$, $i \in [d]$ be the decomposition of Z^δ, as it is defined in (17). Then it is straightforward that

$$(\mathrm{Z}^{\delta,i,j}([\delta^{-1}t]), t \geq 0) \to (Z_t^{i,j}, t \geq 0), \tag{25}$$

almost surely on the Skohorod's space of càdlàg paths toward the process $Z^{i,j}$, as δ tends to 0, for all $i, j \in [d]$, where $Z^{i,j}$ is the decomposition of Z as it is defined in (24).

Now, let $\mathrm{X}^\delta = (\mathrm{X}^{\delta,(i)}, i \in [d])$ be the coding random walk associated to $D_\delta(F, c_F, \ell_F)$, as in Theorem 5. We will first assume that X^δ is actually the coding random walk of a sequence of i.i.d. discrete forests with the same distribution as $D_\delta(F, c_F, \ell_F)$ in the same manner as in the proof of part 2 of Theorem 5, so that in particular, X^δ is not a stopped random walk.

For $i \in [d]$, let $\tau_{i,n}^{\mathrm{X}^\delta}$ and $\tau_{i,n}^{\mathrm{Z}^\delta}$ be the sequences of jump times of the discrete time processes $\mathrm{X}^{\delta,(i)}$ and $\mathrm{Z}^{\delta,i}$. That is the ordered sequences of times such that $\tau_{i,0}^{\mathrm{X}^\delta} = \tau_{i,0}^{\mathrm{Z}^\delta} = 0$ and $\Delta \mathrm{X}_n^{\delta,(i)} := \mathrm{X}^{\delta,(i)}(\tau_{i,n}^{\mathrm{X}^\delta}) - \mathrm{X}^{\delta,(i)}(\tau_{i,n-1}^{\mathrm{X}^\delta}) \neq 0$ and $\Delta \mathrm{Z}_n^{\delta,i} := \mathrm{Z}^{\delta,i}(\tau_{i,n}^{\mathrm{Z}^\delta}) - \mathrm{Z}^{\delta,i}(\tau_{i,n-1}^{\mathrm{Z}^\delta}) \neq 0$, for $n \geq 1$. Fix $k \geq 1$, then since two jumps of the process Z almost surely never occur at the same time, there is δ_0, sufficiently small such that for all $0 < \delta \leq \delta_0$, the sequences $(\Delta \mathrm{X}_n^{\delta,(i)}, 0 \leq n \leq k)$ and $(\Delta \mathrm{Z}_n^{\delta,i}, 0 \leq n \leq k)$ are a.s. identical, for all $i \in [d]$. Moreover, from Lemma 3 and Theorem 5, the jumps $\Delta \mathrm{X}_n^{\delta,(i)}$ have law

$$\mu_i^{(\delta)}(\mathrm{k}) = \frac{\tilde{\nu}_i^{(\delta)}(\mathrm{k})}{1 - \tilde{\nu}_i^{(\delta)}(0)}\,, \;\; \mathrm{k} \neq 0\,, \;\; \mu_i^{(\delta)}(0) = 0\,,$$

where $\tilde{\nu}_i^{(\delta)}(\mathrm{k}) = \mathbb{P}_{e_i}(Z_\delta = (k_1, \dots, k_{i-1}, k_i + 1, k_{i+1}, \dots, k_d))$. The measure $\mu_i^{(\delta)}$ converges weakly to μ_i defined in (1), as $\delta \to 0$. Hence from (25), the sequence $(\Delta \mathrm{Z}_n^{\delta,i}, n \geq 0)$ converges almost surely toward the sequence $(\Delta Z_n^i, n \geq 0)$ of jumps of the process $(Z_t^i, t \geq 0) := (Z_t^{i,j}, j \in [d], t \geq 0)$, which is therefore a sequence of i.i.d. random variables, with law μ_i.

On the other hand, it follows from (19) and (20) in Lemma 2 and the fact that two jumps of the process Z almost surely never occur at the same time, that for all n_1, there is $\delta_1 > 0$ sufficiently small, such that for all $n \leq n_1$ and $0 < \delta \leq \delta_1$,

$$\tau_{i,n}^{\mathrm{X}^\delta} - \tau_{i,n-1}^{\mathrm{X}^\delta} = \sum_{k=\tau_{i,n-1}^{\mathrm{Z}^\delta}}^{\tau_{i,n}^{\mathrm{Z}^\delta}} \mathrm{Z}_k^{\delta,(i)}\,, \quad n \geq 1\,.$$

From Lemma 3 , the latter is a sequence of i.i.d. geometrically distributed random variables with parameter $1 - \mathbb{P}_{e_i}(Z_\delta = e_i)$. Hence from (25), the sequence

$$\delta \cdot \sum_{k=\delta^{-1}\tau_{i,n-1}^{\mathrm{Z}^\delta}}^{\delta^{-1}\tau_{i,n}^{\mathrm{Z}^\delta}} \mathrm{Z}_k^{\delta,(i)}\,, \quad n \geq 1$$

converges almost surely toward the sequence

$$\int_{\tau_{i,n-1}^Z}^{\tau_{i,n}^Z} Z_t^{(i)}\,dt\,, \quad n \geq 1\,,$$

as $\delta \to 0$, where $(\tau_{i,n}^Z)$ is the sequence of jump times of Z^i. Moreover the variables of this sequence are i.i.d. and exponentially distributed with parameter $\lim_{\delta \to 0} \delta^{-1}(1 - \mathbb{P}_{e_i}(Z_\delta = e_i)) = \lambda_i$.

Then since $(\mathrm{X}^{\delta,(i)}_n)$ is a random walk, the sequences $(\Delta\mathrm{X}^{\delta,(i)}_n)_{n\ge0}$, $(\tau^{\mathrm{X}^\delta}_{i,n})_{n\ge0}$, $i\in[d]$ are independent. Therefore, from the convergences proved above, the sequences $(\Delta Z^i_n, n\ge0)$ and $(\int_{\tau^Z_{i,n-1}}^{\tau^Z_{i,n}} Z^{(i)}_t\,dt\,, n\ge1)$ are independent. Then we have proved that the process

$$X^{(i)} := (Z^i(\tau^{(i)}_t), t\ge0)\,,\quad \text{where } \tau^{(i)}_t = \inf\{s : \textstyle\int_0^s Z^{(i)}_u\,du > t\}, \tag{26}$$

is a compound Poisson process with parameter λ_i and jump distribution μ_i. Moreover, it follows from the independence between the random walks $(\mathrm{X}^{\delta,(i)}_n)$, $i\in[d]$ that the processes $(Z^i(\tau^{(i)}_t), t\ge0)$, $i\in[d]$ are independent.

Now from part 1 of Theorem 5, if N^δ_i is the total population of type $i\in[d]$ in the forest $D_\delta(F,c_F,\ell_F)$, then $N^\delta=(N^\delta_1,\dots,N^\delta_d)$ is the smallest solution of the system (x,X^δ). Moreover it follows from the construction of $D_\delta(F,c_F,\ell_F)$, that $\lim_{\delta\to0}\delta N^\delta_i = \int_0^\infty Z^{(i)}_t\,dt := T^{(i)}_x$, almost surely. Note that $T^{(i)}_x$ represents the total length of the branches of type i in the forest (F,c_F,ℓ_F). Then from the definition of $X^{(i)}$ in (26) it appears the these compound Poisson processes are stopped at $T^{(i)}_x$ and that $T_x=(T^{(1)}_x,\dots,T^{(d)}_x)$ satisfies (3).

The fact that (5) is invertible is a direct consequence the first part of the following lemma.

Lemma 4. *Let $x=(x_1,\dots,x_d)\in\mathbb{Z}^d_+$ and $\{(x^{i,j}_t, t\ge0),\ i,j\in[d]\}$ be a family of càdlàg $\mathbb{Z}$-valued, step functions such that for $i\ne j$, $x^{i,j}=0$, $x^{i,j}$ are increasing, $x^{i,i}_0=x_i\ge0$ and $x^{i,i}$ are downward skip free, i.e. $x^{i,i}_t-x^{i,i}_{t-}\ge-1$, for all $t\ge0$, with $x^{i,i}_{0-}=x_i$. Then there exists a (unique) time $t_x=(t^{(1)}_x\dots,t^{(d)}_x)\in\overline{\mathbb{R}}^d_+$ such that*

$$x_j+\sum_{i=1}^d x^{i,j}(t^{(i)}_x)=0\,,\quad \textit{for all } j \textit{ such that } t^{(j)}_x<\infty, \tag{27}$$

and if t'_x is any time satisfying (27), then $t'_x\ge t_x$.

Moreover, the system

$$z^{i,j}_t=\begin{cases} x^{i,j}\left(\int_0^t z^{(i)}_s\,ds\right)\,,\ t\ge0\,,\ \text{if } i\ne j\\ x_i+x^{i,i}\left(\int_0^t z^{(i)}_s\,ds\right)\,,\ t\ge0\,,\ \text{if } i=j\end{cases}$$

admits a unique solution $\{(z^{i,j}_t, t\ge0),\ i,j\in[d]\}$ and the system

$$z^{(j)}_t=x_j+\sum_{i=1}^d x^{i,j}\left(\int_0^t z^{(i)}_s\,ds\right)\,,\ t\ge0,\ j\in[d]$$

admits a unique solution $(z_t^{(i)}, t \geq 0, i \in [d])$*. These solutions are càdlàg step functions which are functionals of the stopped functions* $\{(x_t^{i,j}, 0 \leq t \leq t_x^{(i)}),\ i,j \in [d]\}$*.*

Proof. The proof of the first part of the lemma can be done by applying Lemma 1 to the discrete time skeleton of the functions $\{(x_t^{i,j}, t \geq 0),\ i,j \in [d]\}$, exactly as for the proof of the first part of Theorem 1, see Subsection 4.1.

Then the proof of the existence and uniqueness of the solutions of both systems is straightforward. Let τ_n, $n \geq 1$ be the discrete, ordered sequence of jump times of the processes $\{(x_t^{i,j}, t \geq 0),\ i,j \in [d]\}$ (note that two functions $x^{i,j}$ can jump simultaneously). Then for each of these two systems the solution can be constructed in between the times τ_n and τ_{n+1} in a unique way. □

4.5 Proof of Theorem 3.

Let $(\theta_{k,n}, k,n \geq 1)$ be a family of independent random variables, such that for each k, $\theta_{k,n}$ is uniformly distributed on $[n]$. We assume moreover that the family $(\theta_{k,n}, k,n \geq 1)$ is independent of the compound Poisson process $X = (X^{(1)}, \dots, X^{(d)})$.

Then let us construct a multitype branching forest with edge lengths in the following way. Let $\tau_n^{(i)}$, $n \geq 1$ be the sequence of ordered jump times of the process $X^{(i)}$. We first start with x_i vertices of type $i \in [d]$. Let i be such that $x_i^{-1}\tau_1^{(i)} = \min\{x_j^{-1}\tau_1^{(j)} : x_j^{-1}\tau_1^{(j)} \leq T_x^{(j)},\ j \in [d]\}$. Then we grow the branches issued from each of the $x_1 + \dots + x_d$ vertices in the same time scale and at time $x_i^{-1}\tau_1^{(i)}$, we choose among the x_i vertices of type i according to θ_{1,x_i} the vertex who gives birth.

Then the construction is done recursively. Let y_j be the number of vertices of type $j \in [d]$ in the forest at time $x_i^{-1}\tau_1^{(i)}$ and let $Y = (Y^{(1)}, \dots, Y^{(d)})$ be such that $Y^{(j)}$ corresponds to $X^{(j)}$ shifted at time $s_j = x_j x_i^{-1}\tau_1^{(i)}$, i.e. $Y_t^{(j)} = X_{s_j+t}^{(j)}$. Then let $\tau_{Y,n}^{(j)}$, $n \geq 1$ be the sequence of ordered jump times of the process $Y^{(j)}$, and let k such that $y_k^{-1}\tau_{Y,1}^{(k)} = \min\{y_j^{-1}\tau_{Y,1}^{(j)} : x_i^{-1}\tau_1^{(i)} + y_j^{-1}\tau_{Y,1}^{(j)} \leq T_x^{(j)},\ j \in [d]\}$. Then we continue the construction of the branches issued from each vertex of type $j \in [d]$ and at time $x_i^{-1}\tau_1^{(i)} + y_k^{-1}\tau_{Y,1}^{(k)}$, we choose among the y_k vertices of type k according to θ_{2,y_k} the vertex who gives birth. This construction is performed until all processes $X^{(i)}$ are stopped at time $T_x^{(i)}$.

It is clear from this construction that the forest which is obtained is a multitype branching forest with edge lengths, with the required distribution and such that the processes $Z^{i,j}$ defined as in Definition 1 with respect to this forest satisfy equation (5).

Finally, the fact that equation,

$$(Z_t^{(1)}, \dots, Z_t^{(d)}) = x + \left(\sum_{i=1}^{d} X^{i,1}_{\int_0^t Z_s^{(i)}\,ds}, \dots, \sum_{i=1}^{d} X^{i,d}_{\int_0^t Z_s^{(i)}\,ds}\right), \quad t \geq 0$$

admits a unique solution is a direct consequence of the second part of Lemma 4. □

References

1. K.B. ATHREYA AND P. E. NEY: *Branching processes.* Die Grundlehren der mathematischen Wissenschaften, Band 196. Springer-Verlag, New York-Heidelberg, 1972.
2. M.E. CABALLERO, A. LAMBERT AND G. URIBE BRAVO: Proof(s) of the Lamperti representation of continuous-state branching processes. *Probab. Surv.* 6, 62–89, (2009).
3. M.E. CABALLERO, J.L. PÉREZ AND G. URIBE BRAVO: Affine processes on $\mathbb{R}_+^m \times \mathbb{R}^n$ and multiparameter time changes. *Preprint*, arXiv:1501.03122, (2015).
4. M.E. CABALLERO, J.L. PÉREZ AND G. URIBE BRAVO: A Lamperti-type representation of continuous-state branching processes with immigration. *Ann. Probab.*, Vol. 41, No. 3A, 1585–1627, (2013).
5. L. CHAUMONT AND R. LIU: Coding multitype forests: application to the law of the total population of branching forests. To appear in *Transactions of the American Mathematical Society*, (2015).
6. N. GABRIELLI AND J. TEICHMANN: Pathwise construction of affine processes. *Preprint*, arXiv:1412.7837, (2014).
7. T.E. HARRIS: *The theory of branching processes.* Dover Phoenix Editions. Dover Publications, Inc., Mineola, NY, 2002.
8. J. LAMPERTI Continuous state branching processes. *Bull. Amer. Math. Soc.* 73, 382–386, (1967).
9. J.F. LE GALL: Random trees and applications. *Probab. Surv.*, **2**, 245–311, (2005).
10. J. KALLSEN: A didactic note on affine stochastic volatility models. From stochastic calculus to mathematical finance, 343–368, *Springer, Berlin*, 2006.
11. D. KHOSHNEVISAN AND Y. XIAO: Harmonic analysis of additive Lévy processes. *Probab. Theory Related Fields*, 145, no. 3-4, 459–515, (2009).
12. C.J. MODE: *Multitype branching processes. Theory and applications.* Modern Analytic and Computational Methods in Science and Mathematics, No. 34 American Elsevier Publishing Co., Inc., New York 1971.
13. S. WATANABE: On two dimensional Markov processes with branching property. *Trans. Amer. Math. Soc.* 136, 447–466, (1969).