Probabilistic representation of ODE solutions with quantitative estimates

Qiao Huang* Nicolas Privault†

February 18, 2025

Abstract

This paper considers the probabilistic representation of the solutions of ordinary differential equations (ODEs) by the generation of random trees. We present sufficient conditions on equation coefficients that ensure the integrability and uniform integrability of the functionals of random trees used in this representation, and yield quantitative estimates of its explosion time. Those conditions rely on the analysis of a marked branching process that controls the growth of random trees, in which marks can be interpreted as mutant types in population genetics models. We also show how branching process explosion is connected to existence and uniqueness of ODE solutions.

Keywords: Branching processes, random trees, ordinary differential equations, Butcher series, weighted progeny.
Mathematics Subject Classification (2020): 65L06, 34A25, 34-04, 05C05, 65C05.

1 Introduction

Consider the d-dimensional autonomous ODE problem

$$\begin{cases} x'(t) = f(x(t)), \quad t \in (0,T], \\ x(0) = x_0 \in \mathbb{R}^d, \end{cases} \tag{1.1}$$

where $f \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and $T > 0$. Since f is smooth, the Picard–Lindelöf theorem guarantees local existence and uniqueness of the solution of the ODE (1.1).

It is well known that the solutions of ordinary differential equations (ODEs) such as (1.1) can be represented using Butcher trees, by combining rooted tree enumeration with Taylor

*School of Mathematics, Southeast University, Nanjing 211189, P.R. China, and School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371. qiao.huang@seu.edu.cn

†School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371. nprivault@ntu.edu.sg

expansions. For this, if $x(t)$ is sufficiently smooth at $t = t_0$, consider the Taylor expansion

$$\begin{aligned} x(t) &= x_0 + (t - t_0) f(x_0) + \frac{(t - t_0)^2}{2} \big(\nabla f(f) \big)(x_0) \\ &\quad + \frac{(t - t_0)^3}{6} \big(\nabla^2 f(f, f) + \nabla f(\nabla f(f)) \big)(x_0) + \cdots , \end{aligned} \tag{1.2}$$

which uses the "elementary differentials"

$$f, \ \nabla f(f), \ \nabla^2 f(f, f), \ \nabla f(\nabla f(f)), \ldots .$$

see (2.2)-(2.4) below for their componentwise expressions. From [But63, But16], it is known that the expansion (1.2) admits the formulation

$$x(t) = \sum_{\tau} \frac{(t - t_0)^{|\tau|}}{\tau! \zeta(\tau)} F(\tau)(x_0), \tag{1.3}$$

as a summation over rooted trees τ, where the functional $F(\tau)$ represents the elementary differentials, and $\zeta(\tau)$, $\tau!$ denote respectively the symmetry and factorial of the tree τ. The series (1.3) can be used to estimate ODE solutions by expanding $x(t)$ into a sum over trees up to a finite order, see e.g. [DB02, Chapters 4-6] and references therein.

In [PP22], stochastic branching processes have been used to express ODE solutions as the expected value of a functional of random trees. In [HP23], this approach has been interpreted as a Monte Carlo random generation of the Butcher trees τ for the numerical estimation of the series (1.3). Monte Carlo estimators represent an alternative to the truncation of series, and they allow for estimates that improve via successive iterations.

Those results also complement other approaches to the use of stochastic processes to provide a diffusion interpretation for the solutions of partial differential equations via the Feynman–Kac formula [FH65], and more generally in the fully nonlinear case via stochastic branching mechanisms or stochastic cascades, see, e.g., [Sko64], [INW69], [McK75], [LS97], [DMTW19], [HLOT$^+$19], [NPP23].

However, the above references generally assume the uniform integrability of random weights and/or the existence of a solution, see [HLOT$^+$19, Theorem 3.5], [NPP23, Theorem 4.2], [PP22, Theorem 4.2]. This paper studies the stability of stochastic branching algorithms for the estimation of ODE solutions by probabilistic methods, without making such assumptions.

On the one hand, in Theorem 4.1 we derive integrability conditions on random weights that ensure the existence of the solution of the ODE (1.1) together with the validity of its probabilistic representation. In Theorem 4.2 we obtain related integrability conditions that ensure uniqueness of this solution. random tree functional.

On the other hand, sufficient conditions for the integrability and uniform integrability of random weights over a time interval $[0, T]$ are provided in Theorems 6.1 and 6.3 under uniform bounds on the derivatives of equation coefficients, provided that T is sufficiently small. In Theorem 7.1 and 7.2 those conditions are then relaxed in order to allow for the growth of derivatives. Our results also include quantitative estimates of explosion times that ensure the integrability of stochastic weights, the existence and uniqueness of solutions, and the validity of the stochastic representations of ODE solutions obtained in [PP22], [HP23].

Starting from the ODE (1.1), our approach is to formulate the infinite ODE systems (3.1a)-(3.1b), (3.2), whose solutions $\{x_g\}$ are indexed by smooth functions $g \in \mathcal{C} := \bigcup_{m\geq 1} \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^m)$. We then provide probabilistic representations for $x_g(t)$, $g \in \mathcal{C}$, as well as for $g(x_{\mathrm{Id}}(t))$, $g \in \mathcal{C}$, where Id denotes the identity mapping on $\mathbb{R}^d$. This is achieved in Proposition 3.5 under uniform integrability conditions on functionals of random trees, which also shows the relation

$$x_g(t) = g(x_{\mathrm{Id}}(t)), \qquad g \in \bigcup_{m\geq 1} \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^m), \quad t \geq 0.$$

This representation is then used in Theorems 4.1 and 4.2 to obtain the existence and uniqueness of ODE solutions represented as the expectation of a random tree functional. Those results rely on the integrability criteria established in Theorems 6.1, 6.3, 7.1 and 7.2, based on the analysis of the stability of the branching process that controls the growth of random trees. This analysis is performed in Proposition 5.5 by controlling the integrability of marked random trees in which marks can be interpreted as mutant types in population genetics models with mutation reversion (or back mutation).

This paper is organized as follows. In Section 2 we introduce the basics of marked branching processes and the related functionals of random trees used for the probabilistic representation (4.1). In Section 3 we reformulate the ODE (1.1) using ODE systems which are solved using probabilistic representations under suitable integrability conditions of random weights. Those results are then applied in Section 4 to the probabilistic representation of ODE solutions. In Section 5 we derive sufficient conditions for the integrability required in Section 3 via the analysis of related branching processes. In Sections 6 and 7 we provide sufficient conditions for the uniform integrability of random weights required in Proposition 3.5 and Theorems 4.1-4.2. Appendix A presents an example of lifetime probability density function satisfying Assumption 5.3. Numerical examples are available at https://github.com/nprivaul/mc-odes/blob/main/mc_odes.ipynb.

2 Marked branching trees

Our probabilistic representation of ODE solutions relies on an age-dependent continuous-time branching chain $(X_t)_{t\in[0,T]}$ generating a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The lifetimes of branches are independent and identically distributed via a common density $\rho : \mathbb{R}^+ \to \mathbb{R}$. In addition, a label in

$$\mathbb{K} := \{\varnothing\} \cup \bigcup_{n\geq 1} \{1, 2\}^n,$$

and a mark in the set of smooth functions

$$\mathcal{C} := \bigcup_{m\geq 1} \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^m)$$

are attached to every branch in the chain. More precisely, based on an initial mark $c \in \mathcal{C}$, and the marking mechanism $\mathcal{M} : \mathcal{C} \to \{f\} \cup (\{f\} \times \mathcal{C})$ defined by

$$\mathcal{M}(\mathrm{Id}) = f, \quad \mathcal{M}(g) = (f, \nabla g), \qquad g \in \mathcal{C} \setminus \{\mathrm{Id}\},$$

we consider the marked and labelled random branching tree $\mathcal{B}^{0,c}$ constructed as follows.

- We start from a single branch with label $\varnothing$ and initial mark $c \in \mathcal{C}$. At the end of its lifetime $T_\varnothing := t_\varnothing$, this branch yields either:
  - a single offspring with label (1) and mark f if $c = \mathrm{Id}$,
  - two independent offsprings with respective labels (1), (2) and mark f, ∇c if $c \neq \mathrm{Id}$.
- At generation $n \geq 1$, a branch having a parent label $\mathbf{k}- := (k_1, \ldots, k_{n-1})$ starts at time $T_{\mathbf{k}-}$ and has the lifetime $t_{\mathbf{k}}$. At time $T_{\mathbf{k}} := T_{\mathbf{k}-} + t_{\mathbf{k}}$, it yields two independent offsprings with the respective labels $(\mathbf{k}, 1) = (k_1, \ldots, k_n, 1)$, $(\mathbf{k}, 2) = (k_1, \ldots, k_n, 2)$ and marks f, $\nabla c_{\mathbf{k}}$.

The set of labels of all branches living in the time interval $[0,t]$ is denoted by $\mathcal{K}_t$, its boundary at time t, i.e. the set of labels of all branches living at time $t>0$, is denoted by $\mathcal{K}_t^\partial$, and its interior in $[0,t)$, i.e. the set of branches that split not later than time t, is denoted by $\mathcal{K}_t^\circ$. Figure 1 presents a sample of $\mathcal{K}_t$ for the random tree $\mathcal{B}^{0,\mathrm{Id}}$ started from the mark $c = \mathrm{Id}$, where Id denotes the identity map on $\mathbb{R}^d$.

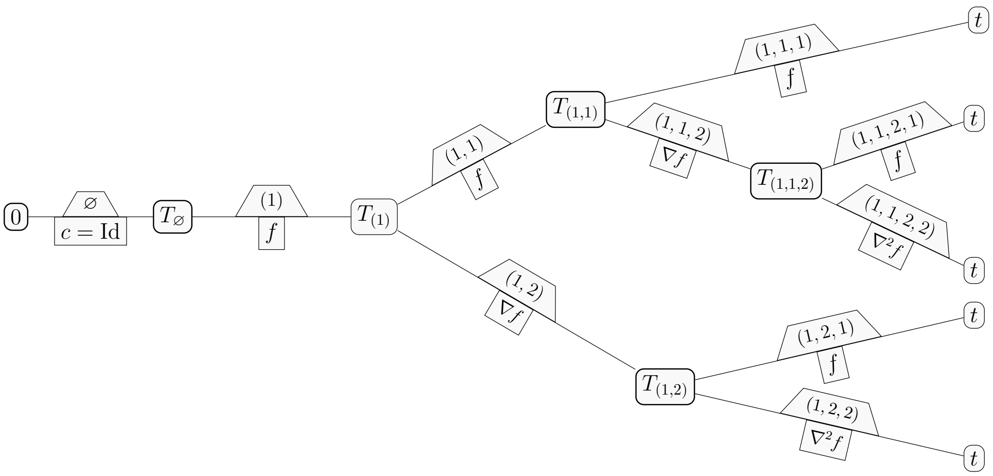


Figure 1: Sample of $\mathcal{K}_t$ for the random tree $\mathcal{B}^{0,c}$ started from $c = \mathrm{Id}$.

Figure 2 presents a sample of $\mathcal{K}_t$ for the random tree $\mathcal{B}^{0,c}$ started from a mark $c \neq \mathrm{Id}$.

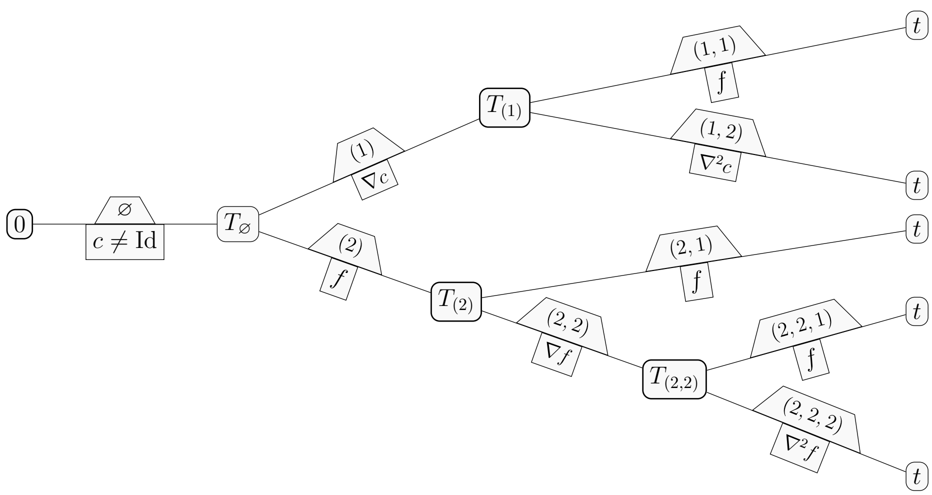


Figure 2: Sample of $\mathcal{K}_t$ for the random tree $\mathcal{B}^{0,c}$ started from $c \neq \mathrm{Id}$.

For $n \geq 1$, we also denote by $\mathcal{K}_t^n$ the label set of the branches generated at the n-th generation, and let

$$\mathcal{K}_t^{\circ,n} := \mathcal{K}_t^\circ \cap \mathcal{K}_t^n, \quad \mathcal{K}_t^{\partial,n} := \mathcal{K}_t^\partial \cap \mathcal{K}_t^n.$$

When the initial mark is $c = g \in \mathcal{C} \setminus \{\mathrm{Id}\}$, we observe that the marks of all branches are of the form either $\nabla^m f$ or $\nabla^m g$, $m \geq 0$. Letting

$$\mathcal{C}_g := \{\nabla^k g\}_{k\geq 0}, \quad g \in \mathcal{C},$$

we also consider the classes of *n-th* generation branches

$$\mathcal{K}_t^n(f) := \{\mathbf{k} \in \mathcal{K}_t^n \ : \ c_{\mathbf{k}} \in \mathcal{C}_f\} \quad \text{and} \quad \mathcal{K}_t^n(g) := \{\mathbf{k} \in \mathcal{K}_t^n \ : \ c_{\mathbf{k}} \in \mathcal{C}_g\}.$$

Finally, we denote by

$$\overline{F}_\rho(t) := \mathbb{P}\left(T_\varnothing > t\right) = \int_t^\infty \rho(r)dr, \qquad t \geq 0,$$

the tail cumulative distribution function of ρ, and we consider the following functional of random trees.

Definition 2.1. *We let $\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)$ denote the functional of $\mathcal{B}^{0,c}$ defined as*

$$\mathcal{H}_t\left(\mathcal{B}^{0,c}\right) := \prod_{\mathbf{k}\in\mathcal{K}_t^\partial} \frac{c_{\mathbf{k}}(x_0)}{\overline{F}_\rho(t - T_{\mathbf{k}-})} \prod_{\mathbf{k}\in\mathcal{K}_t^\circ} \frac{1}{\rho(t_{\mathbf{k}})}, \quad t \in [0, T], \quad c \in \mathcal{C}_f \cup \mathcal{C}_g. \tag{2.1}$$

In (2.1), the product of $c_{\mathbf{k}}(x_0)$ over $k \in \mathcal{K}_t^\partial$ is interpreted as a composition according to tree leaves from bottom to top of the elementary differentials f, $\nabla f(f)$, $\nabla^2 f(f, f)$, $\nabla f(\nabla f(f))$ defined as

$$(\nabla f)(f) := \left(\sum_{i_2=1}^d \frac{\partial f_{i_1}}{\partial x_{i_2}} f_{i_2}\right)_{i_1=1,\ldots,d}, \tag{2.2}$$

$$\nabla^2 f(f, f) := \left(\sum_{i_2,i_3=1}^d \frac{\partial^2 f_{i_1}}{\partial x_{i_2}\partial x_{i_3}} f_{i_2} f_{i_3}\right)_{i_1=1,\ldots,d}, \tag{2.3}$$

$$\nabla f(\nabla f(f)) := \left(\sum_{i_2,i_3=1}^d \frac{\partial f_{i_1}}{\partial x_{i_2}} \frac{\partial f_{i_2}}{\partial x_{i_3}} f_{i_3}\right)_{i_1=1,\ldots,d}, \tag{2.4}$$

etc. For example, in the sample tree of Figure 1 we have

$$\mathcal{K}_t^\circ = \{(1, 1, 1), (1, 1, 2, 1), (1, 1, 2, 2), (1, 2, 1), (1, 2, 2)\}$$

and

$$\prod_{\mathbf{k}\in\mathcal{K}_t^\partial} \frac{1}{\rho(t_{\mathbf{k}})} = \nabla^2 f(f, \nabla^2 f(f, f)),$$

while in the sample tree of Figure 2 we have

$$\mathcal{K}_t^\circ = \{(2,2),(2,1),(1,2,2),(1,2,1),(1,1)\}$$

and

$$\prod_{\mathbf{k}\in\mathcal{K}_t^\partial} \frac{1}{\rho(t_{\mathbf{k}})} = \nabla^2 c(f, \nabla^2 f(f,f)),$$

see Appendix B for details.

3 ODE systems

In this section, we reformulate the ODE (1.1) as an infinite system of differential equations indexed by the set $\mathcal{C}$ of marks.

Lemma 3.1. *Let $T > 0$ such that the ODE (1.1) admits a solution $x \in \mathcal{C}^1([0,T], \mathbb{R}^d)$. Then, the family $(x_c)_{c\in\{\mathrm{Id}\}\cup\{\nabla^k f\}_{k\geq 0}} := (c(x))_{c\in\{\mathrm{Id}\}\cup\{\nabla^k f\}_{k\geq 0}}$, solves the ODE system*

$$\begin{cases} (x_c)' = x_f, \quad c = \mathrm{Id}, & (3.1a) \\ (x_c)' = x_f x_{\nabla c}, \quad c = \nabla^m f,\ m \geq 0, & (3.1b) \end{cases}$$

with the initial conditions $x_c(0) = c(x_0)$, for $c \in \{\mathrm{Id}\} \cup \{\nabla^k f\}_{k\geq 0}$.

Proof. It follows from the time differentiation

$$g(x(t))' = x(t)\nabla g(x(t)) = f(x(t))\nabla g(x(t)), \quad t \in [0,T],$$

that for every test function $g \in \mathcal{C}$, the family $(g(x))_{g\in\mathcal{C}}$ satisfies the system of equations

$$(g(x))' = \begin{cases} f(x), & g = \mathrm{Id}, \\ f(x)\nabla g(x), & g \in \mathcal{C}\setminus\{\mathrm{Id}\}, \end{cases}$$

with initial condition $(x_g(x)(0))_{g\in\mathcal{C}} = (g(x_0))_{g\in\mathcal{C}}$. $\square$

Proposition 3.2. *Assume that*

(C1) $(\nabla^k f(x_0))_{k\geq 0} \in \ell^p(\mathbb{N})$ *for some $p \in [1,\infty]$.*

Then, the subsystem (3.1b) admits a unique local solution in $\ell^p(\mathbb{N})$.

Proof. Using the left shift operator that acts on real sequences as

$$S(x^{(0)}, x^{(1)}, \ldots) = (x^{(1)}, x^{(2)}, \ldots),$$

we rewrite the subsystem (3.1b), as the infinite-dimensional ODE

$$\mathbf{x}' = x^{(0)} \cdot S(\mathbf{x}).$$

Since the nonlinearity $F(\mathbf{x}) := x^{(0)} \cdot S(\mathbf{x})$ is continuous from $\ell^p(\mathbb{N})$ into itself together with its (non-vanishing) Fréchet derivatives

$$DF(\mathbf{x})(\mathbf{u}) = u^{(0)} \cdot S(\mathbf{x}) + x^{(0)} \cdot S(\mathbf{u}) \quad \text{and} \quad D^2F(\mathbf{x})(\mathbf{u}, \mathbf{v}) = u^{(0)} \cdot S(\mathbf{v}) + v^{(0)} \cdot S(\mathbf{u}),$$

local existence follows from the infinite-dimensional version of the Picard–Lindelöf theorem, cf. e.g. [LS94, Theorem 25], and shows the well-posedness of (3.1b). $\square$

The well-posedness of the ODE system (3.1a)-(3.1b) now follows from that of (3.1b). We also note that the local well-posedness of (3.1a)-(3.1b) only relies on the regularity of f at the point x_0, whereas that of the ODE (1.1) relies on the local regularity of f near x_0. In what follows, given $g \in \mathcal{C} \setminus \{\mathrm{Id}\}$ we augment the subsystem (3.1b) as

$$\begin{cases} (x_c)' = x_f x_{\nabla c}, & c = \nabla^m g, \\ (x_c)' = x_f x_{\nabla c}, & c = \nabla^m f, \quad m \geq 0, \end{cases} \tag{3.2}$$

with the initial conditions

$$\begin{cases} x_c(0) = \nabla^m g(x_0), & c = \nabla^m g, \\ x_c(0) = \nabla^m f(x_0), & c = \nabla^m f, \quad m \geq 0. \end{cases}$$

Lemma 3.3. *Assume that*

$$\mathbb{E}\big[\big|\mathcal{H}_t\big(\mathcal{B}^{0,c}\big)\big|\big] < \infty, \qquad t \in [0, T], \quad c \in \mathcal{C}.$$

Then, the functions defined by

$$x_c(t) := \mathbb{E}\big[\mathcal{H}_t\big(\mathcal{B}^{0,c}\big)\big], \quad t \in [0, T], \quad c \in \mathcal{C}, \tag{3.3}$$

satisfy the recursion

$$x_c(t) = c(x_0) + \int_0^t \big(x_f(s)\mathbf{1}_{\{c=\mathrm{Id}\}} + x_{\nabla c}(s) x_f(s) \mathbf{1}_{\{c \neq \mathrm{Id}\}}\big)\, ds, \quad t \in [0, T],\ c \in \mathcal{C}.$$

In particular,

i) The family $(x_c)_{c\in\mathcal{C}_f\cup\mathcal{C}_g}$ *solves the ODE system* (3.2)*, i.e.*

$$\begin{cases} (x_{\nabla^m g})' = x_f x_{\nabla^{m+1}g}, \\ (x_{\nabla^m f})' = x_f x_{\nabla^{m+1}f}, \quad m \geq 0. \end{cases}$$

ii) If $x_f = f(x_{\mathrm{Id}})$*, then* x_{Id} *is a solution of the ODE* (1.1)*.*

Proof. Given $c \in \mathcal{C} \setminus \{\mathrm{Id}\}$, we observe that $\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)$ satisfies

$$\begin{aligned}
\mathcal{H}_t\left(\mathcal{B}^{0,c}\right) &= \mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\left(\mathbf{1}_{\{T_\varnothing>t\}} + \mathbf{1}_{\{T_\varnothing\leq t\}}\right) \\
&= \frac{c(x_0)}{\overline{F}_\rho(t-T_{\varnothing^-})}\mathbf{1}_{\{T_\varnothing>t\}} + \mathbf{1}_{\{T_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}\prod_{\mathbf{k}\in\mathcal{K}_t^\partial}\frac{c_{\mathbf{k}}(x_0)}{\overline{F}_\rho(t-T_{\mathbf{k}^-})}\prod_{\mathbf{k}\in\mathcal{K}_t^\circ\setminus\{\varnothing\}}\frac{1}{\rho(t_{\mathbf{k}})} \\
&= \frac{c(x_0)}{\overline{F}_\rho(t)}\mathbf{1}_{\{T_\varnothing>t\}} + \mathbf{1}_{\{T_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}\left(\mathcal{H}_t\left(\mathcal{B}^{T_\varnothing,f}\right)\mathbf{1}_{\{c=\mathrm{Id}\}} + \mathcal{H}_t\left(\mathcal{B}^{T_\varnothing,\nabla c}\right)\mathcal{H}_t\left(\mathcal{B}^{T_\varnothing,f}\right)\mathbf{1}_{\{c\neq\mathrm{Id}\}}\right).
\end{aligned}$$

Hence, by independence of tree branches, we have

$$\begin{aligned}
x_c(t) &= \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right] \\
&= \mathbb{E}\left[\frac{c(x_0)}{\overline{F}_\rho(t)}\mathbf{1}_{\{t_\varnothing>t\}} + \mathbf{1}_{\{t_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}\mathcal{H}_t\left(\mathcal{B}^{t_\varnothing,\nabla c}\right)\mathcal{H}_t\left(\mathcal{B}^{t_\varnothing,f}\right)\right] \\
&= c(x_0)\mathbb{E}\left[\frac{1}{\overline{F}_\rho(t)}\mathbf{1}_{\{t_\varnothing>t\}}\right] + \mathbb{E}\left[\mathbf{1}_{\{t_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}\mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{t_\varnothing,\nabla c}\right)\,\middle|\,t_\varnothing\right]\mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{t_\varnothing,f}\right)\,\middle|\,t_\varnothing\right]\right] \\
&= c(x_0)\mathbb{E}\left[\frac{1}{\overline{F}_\rho(t)}\mathbf{1}_{\{t_\varnothing>t\}}\right] + \mathbb{E}\left[\mathbf{1}_{\{t_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}x_{\nabla c}(t-t_\varnothing)x_f(t-t_\varnothing)\right] \\
&= c(x_0) + \int_0^t x_{\nabla c}(t-s)x_f(t-s)ds \\
&= c(x_0) + \int_0^t x_{\nabla c}(s)x_f(s)ds.
\end{aligned}$$

In case $c = \mathrm{Id}$, we similarly have

$$\begin{aligned}
x_{\mathrm{Id}}(t) &= \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,\mathrm{Id}}\right)\right] \\
&= \mathbb{E}\left[\frac{x_0}{\overline{F}_\rho(t)}\mathbf{1}_{\{t_\varnothing>t\}} + \mathbf{1}_{\{t_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}\mathcal{H}_t\left(\mathcal{B}^{t_\varnothing,f}\right)\right] \\
&= x_0 + \int_0^t \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{s,f}\right)\right]ds \\
&= x_0 + \int_0^t x_f(t-s)ds \\
&= x_0 + \int_0^t x_f(s)ds.
\end{aligned}$$

$\square$

As a consequence of Proposition 3.2 and Lemma 3.3 we have the following proposition, which recovers the conclusion of Theorem 4.2 in [PP22].

Proposition 3.4. *Under the conditions of Lemma 3.3, assume that* $(\nabla^k f)_{k\geq 0}$ *satisfies*

(C2) $(\nabla^k f(y))_{k\geq 0} \in \ell^p(\mathbb{N})$ *for some* $p \in [1,\infty]$ *and all* y *in a neighborhood* U *of* x_0 *in* $\mathbb{R}^d$*.*

Then, for some $T>0$*,* $(x_{\mathrm{Id}}(t))_{t\in[0,T]}$ *defined in (3.3) is a solution of the ODE (1.1).*

Proof. Under Condition (C2) the ODE (1.1) admits a $\mathcal{C}^1$ solution such that for some $T>0$ we have

$$(\nabla^k f(x(t)))_{k\geq 0} \in \ell^p(\mathbb{N}), \qquad t\in[0,T],$$

for some $p\in[1,\infty]$. In addition, from Proposition 3.2 the subsystem (3.1b) admits a unique solution in $\ell^p(\mathbb{N})$ on $[0,T]$, hence by Lemma 3.3 we have $x = x_{\mathrm{Id}}$ and $\nabla^m f(x) = x_{\nabla^m f}$ for all $m\geq 0$. □

In comparison with Condition (C1), Condition (C2) fills the gap of regularity requirement on f between the well-posedness of the ODE system (3.1a)-(3.1b) and that of the ODE (1.1), as observed in Proposition 3.2.

Proposition 3.5 provides a probabilistic interpretation for the solutions of the ODE system (3.1a)-(3.1b). Sufficient conditions on f for the integrability requirements on (3.4) and (3.6) below are provided in Theorems 6.1 and 6.3.

Proposition 3.5. *Let* $g\in\mathcal{C}\setminus\{\mathrm{Id}\}$*.*

i) Assume that the ODE system (3.1a)-(3.1b) admits a solution

$$\{x_c\}_{c\in\{\mathrm{Id}\}\cup\mathcal{C}_f} \subset \bigcup_{m\geq 1}\mathcal{C}^1([0,T],\mathbb{R}^m),$$

and that the sequence of functionals of the tree $\mathcal{B}^{0,g}$ *defined as*

$$\widetilde{\mathcal{H}}_t^{g,n}(\mathcal{B}^{0,g}) := \prod_{\mathbf{k}\in\cup_{i=0}^n\mathcal{K}_t^{\partial,i}} \frac{c_{\mathbf{k}}(x_0)}{\overline{F}_\rho(t-T_{\mathbf{k}^-})} \prod_{\mathbf{k}\in\cup_{i=0}^n\mathcal{K}_t^{\circ,i}} \frac{1}{\rho(t_{\mathbf{k}})} \prod_{\mathbf{k}\in\mathcal{K}_t^{n+1}(f)} x_{c_{\mathbf{k}}}(T_{\mathbf{k}^-}) \prod_{\mathbf{k}\in\mathcal{K}_t^{n+1}(g)} c_{\mathbf{k}}\left(x_{\mathrm{Id}}(T_{\mathbf{k}^-})\right) \tag{3.4}$$

is uniformly integrable in $n\geq 0$ *for all* $t\in[0,T]$*. Then, the following representation holds:*

$$g(x_{\mathrm{Id}}(t)) = \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,g}\right)\right], \qquad t\in[0,T]. \tag{3.5}$$

ii) Assume that the ODE system (3.2) *admits a solution*

$$\{x_c\}_{c\in\mathcal{C}_f\cup\mathcal{C}_g} \subset \bigcup_{m\geq 1} \mathcal{C}^1([0,T],\mathbb{R}^m),$$

and that the sequence

$$\mathcal{H}_t^{g,n}\big(\mathcal{B}^{0,g}\big) := \prod_{\mathbf{k}\in\cup_{i=0}^n \mathcal{K}_t^{\partial,i}} \frac{c_{\mathbf{k}}(x_0)}{\overline{F}_\rho(t-T_{\mathbf{k}^-})} \prod_{\mathbf{k}\in\cup_{i=0}^n \mathcal{K}_t^{\circ,i}} \frac{1}{\rho(t_{\mathbf{k}})} \prod_{\mathbf{k}\in\mathcal{K}_t^{n+1}} x_{c_{\mathbf{k}}}(T_{\mathbf{k}^-}) \tag{3.6}$$

is uniformly integrable in $n \geq 0$ *for all* $t \in [0,T]$. *Then, the solution of the ODE system* (3.2) *is unique and the following representation holds:*

$$x_g(t) = \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,g}\right)\right], \qquad t\in[0,T]. \tag{3.7}$$

Moreover, if all above conditions are satisfied, then for all $g \in \mathcal{C}\setminus\{\mathrm{Id}\}$ *we have*

$$g(x_{\mathrm{Id}}(t)) = x_g(t) = \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,g}\right)\right], \qquad t\in[0,T].$$

Proof. i) It follows from (3.1a)-(3.1b) that

$$\begin{aligned} g(x_{\mathrm{Id}}(t)) &= g(x_0) + \int_0^t \nabla g(x_{\mathrm{Id}}(s))x_f(s)ds \\ &= \mathbb{E}\left[\frac{g(x_0)}{\overline{F}_\rho(t)}\mathbf{1}_{\{T_\varnothing>t\}} + \mathbf{1}_{\{T_\varnothing\leq t\}}\frac{1}{\rho(t_\varnothing)}\nabla g(x_{\mathrm{Id}}(T_\varnothing))x_f(T_\varnothing)\right] \\ &= \mathbb{E}\big[\widetilde{\mathcal{H}}_t^{g,0}\big], \quad t\in[0,T]. \end{aligned} \tag{3.8}$$

Next, repeating the argument leading to (3.8), we expand $\nabla g(x_{\mathrm{Id}}(T_\varnothing))$ for $\mathbf{k}=(1)$ as

$$\begin{aligned} \nabla g(x_{\mathrm{Id}}(T_\varnothing)) &= \nabla g(x_{\mathrm{Id}}(T_{\mathbf{k}-})) \\ &= \mathbb{E}\left[\frac{\nabla g(x_0)}{\overline{F}_\rho(t-T_{\mathbf{k}-})}\mathbf{1}_{\{T_{\mathbf{k}}>t\}} + \mathbf{1}_{\{T_{\mathbf{k}}\leq t\}}\frac{1}{\rho(t_{\mathbf{k}})}\nabla^2 g(x_{\mathrm{Id}}(T_{\mathbf{k}}))x_f(T_{\mathbf{k}}) \,\middle|\, \mathcal{F}_{T_\varnothing}\right], \end{aligned} \tag{3.9}$$

and plugging (3.9) back into (3.8) shows that $g(x_{\mathrm{Id}}(t)) = \mathbb{E}\big[\widetilde{\mathcal{H}}_t^{g,1}\big]$ by conditional independence of branches in $\mathcal{K}_t^1$ given $\mathcal{F}_{T_\varnothing}$. Similarly, we can show by iterations that

$$g(x_{\mathrm{Id}}(t)) = \mathbb{E}\big[\widetilde{\mathcal{H}}_t^{g,n}\big]$$

for all $n \geq 0$. Then (3.5) follows by taking the limit as n tends to infinity, from the uniform integrability of $\big(\widetilde{\mathcal{H}}_t^{g,n}\big)_{n\geq 0}$ and the almost sure convergence of $\widetilde{\mathcal{H}}_t^{g,n}$ to $\mathcal{H}_t(\mathcal{B}^{0,g})$.

ii) By (3.2), we have

$$\begin{aligned} x_g(t) &= x_g(0) + \int_0^t x_{\nabla g}(s) x_f(s) ds \\ &= \mathbb{E}\left[\frac{g(x_0)}{\overline{F}_\rho(t)} \mathbf{1}_{\{T_\varnothing > t\}} + \mathbf{1}_{\{T_\varnothing \le t\}} \frac{1}{\rho(t_\varnothing)} x_{\nabla g}(T_\varnothing) x_f(T_\varnothing) \right] \\ &= \mathbb{E}[\mathcal{H}_t^{g,0}], \quad t \in [0,T]. \end{aligned} \tag{3.10}$$

Next, since each offspring has same dynamics as its parent branch, we can repeat the above argument to the branch $\mathbf{k} = (1) \in \mathcal{K}_t^1$ with mark $g_1 \in \{\nabla g, f\}$, to get

$$\begin{aligned} x_{g_1}(T_\varnothing) &= x_{g_1}(T_{\mathbf{k}-}) \\ &= \mathbb{E}\left[\frac{g_1(x_0)}{\overline{F}_\rho(t - T_{\mathbf{k}-})} \mathbf{1}_{\{T_\mathbf{k} > t\}} + \mathbf{1}_{\{T_\mathbf{k} \le t\}} \frac{1}{\rho(t_\mathbf{k})} x_{\nabla g_1}(T_\mathbf{k}) x_f(T_\mathbf{k}) \,\middle|\, \mathcal{F}_{T_\varnothing} \right], \end{aligned} \tag{3.11}$$

and plugging (3.11) into (3.10) yields $x_g(t) = \mathbb{E}[\mathcal{H}_t^{g,1}]$ by conditional independence of branches in $\mathcal{K}_t^1$ given $\mathcal{F}_{T_\varnothing}$. Similarly, by iterations we find

$$x_g(t) = \mathbb{E}[\mathcal{H}_t^{g,n}], \qquad n \ge 0.$$

As $\mathcal{H}_t^{g,n}$ converges to $\mathcal{H}_t(\mathcal{B}^{0,g})$ almost surely as n tends to infinity, we obtain (3.7) from the uniform integrability of $\left(\mathcal{H}_t^{g,n}\right)_{n\ge 0}$. □

4 Probabilistic solution of ODEs

Theorems 4.1 and 4.2 show the existence and uniqueness of a solution to the ODE (1.1) on a time interval $[0,T]$ by a probabilistic representation argument, under uniform integrability assumptions on the sequences $\left(\widetilde{\mathcal{H}}_t^{f,n}(\mathcal{B}^{0,f})\right)_{n\ge 0}$ and $\left(\mathcal{H}_t^{g,n}(\mathcal{B}^{0,g})\right)_{n\ge 0}$ respectively defined in (3.4) and (3.6). In comparison with Theorem 4.2 of [PP22], the next result does not assume the existence of a solution to the ODE (1.1).

Theorem 4.1 (Existence)**.** *Assume that*

i) the functional $\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)$ is integrable for all $t \in [0,T]$ and $c \in \mathcal{C}_f$, and

ii) the sequence $\left(\widetilde{\mathcal{H}}_t^{f,n}(\mathcal{B}^{0,f})\right)_{n\ge 0}$ defined in (3.4) is uniformly integrable for all $t \in [0,T]$.

Then, the ODE (1.1) has a solution $(x(t))_{t\in[0,T]} \in \mathcal{C}^1([0,T],\mathbb{R}^d)$ which admits the probabilistic representation

$$x(t) = \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,\mathrm{Id}}\right)\right], \quad t \in [0,T]. \tag{4.1}$$

Proof. By taking $c \in \mathcal{C}_f$ in Lemma 3.3, we see that the family $\{x_c\}_{c\in\{\mathrm{Id}\}\cup\mathcal{C}_f}$ defined in (3.3) solves the ODE system (3.1a)-(3.1b). Taking $g = f$ in Proposition 3.5-(i), we have $f(x_{\mathrm{Id}}(t)) = \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{0,f}\right)\right] = x_f(t)$ for all $t \in [0,T]$. It then follows from Lemma 3.3 that x_{Id} is a solution of the ODE (1.1). $\square$

Theorems 6.1 and 6.3 provide sufficient conditions on f for the integrability required in the Theorems 4.1 and 4.2. By Theorems 4.1 and 6.1 and the proof of Lemma 3.3, under the conditions of Theorem 6.3 we also obtain the bound

$$\begin{aligned} x(t) &= x_0 + \int_0^t \mathbb{E}\left[\mathcal{H}_t\left(\mathcal{B}^{s,f}\right)\right] ds \\ &\leq x_0 + \int_0^t \frac{e^{-\lambda s}C_0}{1-(1-e^{-\lambda s})C_0^2} ds \\ &= x_0 + \frac{t}{C_0} + \frac{1}{\lambda C_0}\log\frac{1}{C_0^2-(C_0^2-1)e^{\lambda t}}, \qquad t \in [0,T]. \end{aligned}$$

where C_0 is defined in (6.1).

Theorem 4.2 (Uniqueness)**.** *Assume that the ODE* (1.1) *admits a solution* $(x(t))_{t\in[0,T]} \in \mathcal{C}^1([0,T],\mathbb{R}^d)$*, and that*

i) the functional $\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)$ *is integrable for all* $t \in [0,T]$ *and* $c \in \mathcal{C}_f$*, and*

ii) the sequence $\left(\mathcal{H}_t^{g,n}\left(\mathcal{B}^{0,g}\right)\right)_{n\geq 0}$ *defined in* (3.6) *is uniformly integrable for all* $t \in [0,T]$*.*

Then, $(x(t))_{t\in[0,T]}$ *is the unique solution of the ODE* (1.1)*, and it admits the probabilistic representation* (4.1)*.*

Proof. By taking $c \in \mathcal{C}_f$ in Lemma 3.3, we see that the family $\{x_c\}_{c\in\{\mathrm{Id}\}\cup\mathcal{C}_f}$ defined in (3.3) solves the ODE system (3.1a)-(3.1b). Taking $g = f$ in Proposition 3.5-(ii), we have the uniqueness for the ODE system (3.1a)-(3.1b). By Lemma 3.1, $\{x_c\}_{c\in\{\mathrm{Id}\}\cup\mathcal{C}_f}$ coincides with $\{x, \nabla^m f(x) : m \geq 0\}$. Thus, the solution $(x(t))_{t\in[0,T]}$ admits the probabilistic representation $x = x_{\mathrm{Id}}$, and it is unique. $\square$

In Sections 6 and 7 we will provide sufficient conditions for the uniform integrability of $\left(\widetilde{\mathcal{H}}_t^{f,n}\left(\mathcal{B}^{0,f}\right)\right)_{n\geq 0}$ and $\left(\mathcal{H}_t^{g,n}\left(\mathcal{B}^{0,g}\right)\right)_{n\geq 0}$ required in Proposition 3.5 and Theorems 4.1-4.2, which also imply the integrability of $\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)$ for $t \in [0,T]$ and $c \in \mathcal{C}_f$.

As a consequence, we have the following results under uniform boundedness conditions on the derivatives of f. The existence of a probability density function ρ satisfying the required conditions is shown in Lemma A.1.

Corollary 4.3. *Let $T > 0$, and assume that for some $q > 1$ we have*

$$|\nabla^k f(x_0)| \le \overline{F}_\rho(T) \frac{2^{1/q}(1-e^{-\lambda T})^{1/(2q)}}{(\sqrt{4+e^{-2\lambda T}} - e^{-\lambda T})^{1/q}}, \qquad k \ge 0.$$

Then,

i) the ODE (1.1) *admits at most one solution in* $\mathcal{C}^1([0,T], \mathbb{R}^d)$*;*

ii) if, in addition,

$$\sup_{|x-x_0|<T/(1-e^{-\lambda T})^{1/(2q)}} \left|\nabla^k f(x)\right| < \frac{1}{(1-e^{-\lambda T})^{1/(2q)}}, \qquad k \ge 0, \tag{4.2}$$

then the ODE (1.1) *has a unique solution in* $\mathcal{C}^1([0,T], \mathbb{R}^d)$*, which admits the probabilistic representation* (4.1)*.*

Proof. i) is a consequence of Theorem 4.2, Theorem 6.1, and Theorem 6.3-i); ii) is a consequence of Theorems 4.1-4.2, Theorem 6.1, and Theorem 6.3-ii). □

We note that as q tends to 1, Condition (4.2) is compatible for $m = 0$ with Theorem 2.3 in [CL84], which shows that if f is C-Lipschitz on an interval $[x_0 - b, x_0 + b]$, then (1.1) admits a unique solution $x(t)$ on the time interval $[0, b/\|f\|_\infty]$.

The next result relaxes the existence conditions of Corollary 4.3 by allowing the growth of derivatives of f, as a consequence of Theorems 4.1-4.2 and 7.1-7.2.

Corollary 4.4. *Let $T > 0$, and assume that for some $q > 1$ and $\delta > 0$ we have*

$$|f(x_0)| < \frac{e^{\lambda T}}{(2(1-e^{-\lambda T})\delta)^{1/(2q)}} \quad \text{and} \quad |\nabla^k f(x_0)| \le e^{\lambda T}(k\delta)^{1/(2q)}, \quad k \ge 1.$$

Then,

i) the ODE (1.1) *admits at most one solution in* $\mathcal{C}^1([0,T], \mathbb{R}^d)$*;*

ii) if, in addition,

$$\sup_{|x-x_0|<T/(1-e^{-\lambda T})^{1/(2q)}} \left|\nabla^k f(x)\right|^q < \frac{1}{(1-e^{-\lambda T})^{1/(2q)}}, \quad k \ge 0,$$

then the ODE (1.1) *has a unique solution in* $\mathcal{C}^1([0,T], \mathbb{R}^d)$ *which admits the probabilistic representation* (4.1)*.*

Proof. i) is a consequence of Theorem 4.2, Theorem 7.1, and Theorem 7.2-i); ii) is a consequence of Theorems 4.1-4.2, Theorem 7.1, and Theorem 7.2-ii) with $\delta = 1$ and $\gamma = 0$.

□

5 Stochastic dominance of random binary trees

In this section, we present the stochastic dominance and integrability results for branching processes, that will be used in Section 6.

Definition 5.1. *Let $\big(\widetilde{X}_t\big)_{t\in[0,T]}$ be a continuous-time binary branching chain starting from $\widetilde{X}_0 = 1$, in which the lifetimes of branches are independent and identically distributed via a common exponential density function $\widetilde{\rho}(t) = \lambda e^{-\lambda t}$, $t \geq 0$, with parameter $\lambda > 0$, and progeny process $\widetilde{N}_t$.*

Recall that from e.g. Eq. (8) page 3 of [Ken48], [Har63, Example 13.2 page 112], [AN72, Example 5 page 109], the total progeny $\widetilde{N}_t$ of $\big(\widetilde{X}_t\big)_{t\in\mathbb{R}_+}$ has distribution

$$\mathbb{P}\big(\widetilde{N}_t = m\big) = \begin{cases} e^{-\lambda t}(1-e^{-\lambda t})^m, & n = 2m+1, \\ 0, & \text{otherwise}, \quad m \geq 0, \end{cases}$$

and probability generating function

$$G_t(z) = \mathbb{E}\big[z^{\widetilde{N}_t}\big] = \frac{ze^{-\lambda t}}{1-(1-e^{-\lambda t})z^2}, \quad z < (1-e^{-\lambda t})^{-1/2}. \tag{5.1}$$

Similarly to $\mathcal{B}^{0,c}$, we consider the marked random branching tree $\widetilde{\mathcal{B}}^{0,j}$ constructed by assigning a mark $\tilde{c}_{\mathbf{k}}$ to each branch $\mathbf{k}$ in the set $\widetilde{\mathcal{K}}_t$ of branches of the branching chain $\big(\widetilde{X}_t\big)_{t\in[0,T]}$, in the following way:

a) the initial branch has label $\varnothing$ and is marked by $\tilde{c}_\varnothing = j \in \mathbb{N}$;

b) if a branch $\mathbf{k}$ is marked by $\tilde{c}_{\mathbf{k}} = i \in \mathbb{N}$ and splits, its two children are respectively marked by 0 and $i+1$.

Figure 3 presents a sample of the corresponding random marked tree.

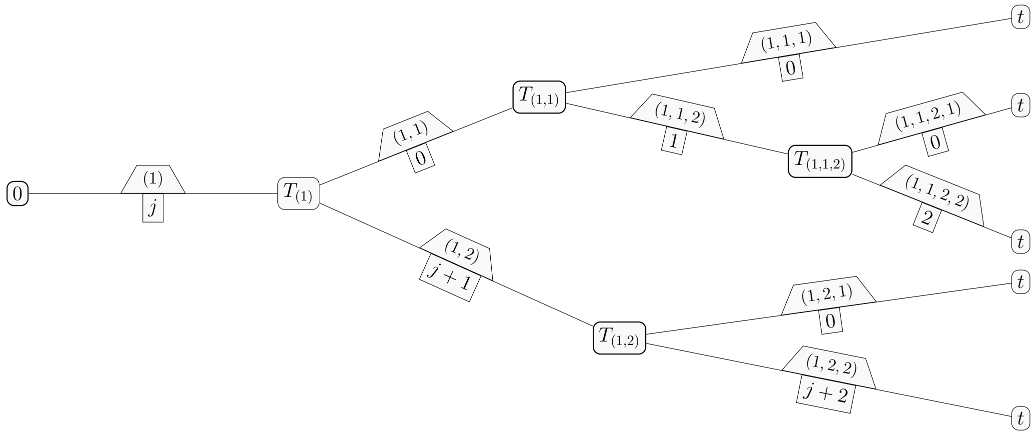


Figure 3: Sample of $\widetilde{\mathcal{K}}_t$ for the random tree $\widetilde{\mathcal{B}}^{0,j}$ started from $j \in \mathbb{N}$.

Definition 5.2. *Let*

$$X_t := |\mathcal{K}_t^{\partial}| = 1 + \sum_{T_{\mathbf{k}} \leq t} \Delta X_{T_{\mathbf{k}}} \tag{5.2}$$

denote the binary branching chain starting from $X_0 = 1$ and formed by the number of branches of $\mathcal{B}^{0,c}$ living at time $t > 0$, with offspring count $1 + \Delta X_{T_{\mathbf{k}}} = 2$ at any splitting time $T_{\mathbf{k}}$, with lifetimes distribution ρ, and total progeny process

$$N_t := |\mathcal{K}_t| = 1 + \sum_{T_{\mathbf{k}} \leq t} (1 + \Delta X_{T_{\mathbf{k}}}) = X_t + \sharp\{\mathbf{k} \in \mathbb{K} : T_{\mathbf{k}} \leq t\}, \quad t \in [0, T]. \tag{5.3}$$

Lemma 5.4 states that the condition $\mathbb{P}\big(\widetilde{N}_t = 1\big) \leq \mathbb{P}(N_t = 1)$ implies

$$\mathbb{P}\big(\widetilde{N}_t \leq n\big) \leq \mathbb{P}(N_t \leq n), \qquad n \geq 1.$$

Assumption 5.3. *There exists $\lambda > 0$ such that ρ dominates the exponential distribution with parameter λ in the sense that*

$$\overline{F}_\rho(r) \geq e^{-\lambda r} = \overline{F}_{\widetilde{\rho}}(r), \qquad r \geq 0. \tag{5.4}$$

An example of probability density function ρ satisfying Assumption 5.3 is provided in Appendix A.

Lemma 5.4. *Under Assumption 5.3, the processes $(X_t)_{t\in[0,T]}$ and $(N_t)_{t\in[0,T]}$ are stochastically dominated by $\big(\widetilde{X}_t\big)_{t\in[0,T]}$ and $\big(\widetilde{N}_t\big)_{t\in[0,T]}$ respectively, i.e.*

$$\mathbb{P}\left(X_t \geq n\right) \leq \mathbb{P}\big(\widetilde{X}_t \geq n\big) \quad \textit{and} \quad \mathbb{P}\left(N_t \geq n\right) \leq \mathbb{P}\big(\widetilde{N}_t \geq n\big), \quad n \geq 0, \ t \in [0, T]. \tag{5.5}$$

Proof. By (5.2) we have

$$\mathbb{P}\left(X_t \geq n\right) = \mathbb{P}\left(1 + \sharp\{\mathbf{k} \in \mathbb{K} \ : \ T_{\mathbf{k}} \leq t\} \geq n\right) = \mathbb{E}[\Phi((t_{\mathbf{k}} : \mathbf{k} \in \mathbb{K}))], \qquad n \geq 0,$$

where the function

$$\Phi : (t_{\mathbf{k}} \ : \ \mathbf{k} \in \mathbb{K}) \mapsto \mathbf{1}_{\{1+\sharp\{\mathbf{k}\in\mathbb{K} \ : \ T_{\mathbf{k}}\leq t\}\geq n\}}$$

is non-increasing in every variable $t_{\mathbf{k}}$, $k \in \mathbb{K}$. Assumption 5.4 implies that for every $\mathbf{k} \in \mathbb{K}$ and $s \geq 0$,

$$\mathbb{P}\left(T_{\mathbf{k}} - T_{\mathbf{k}-} \leq s\right) \leq 1 - e^{-\lambda s} = \mathbb{P}\big(\widetilde{T}_{\mathbf{k}} - \widetilde{T}_{\mathbf{k}-} \leq s\big),$$

i.e., $t_{\mathbf{k}} = T_{\mathbf{k}} - T_{\mathbf{k}-}$ stochastically dominates $\tilde{t}_{\mathbf{k}} = \widetilde{T}_{\mathbf{k}} - \widetilde{T}_{\mathbf{k}-}$, hence from [SS07, Theorem 6.B.3], the random vector $(t_{\mathbf{k}} : \mathbf{k} \in \mathbb{K})$ stochastically dominates $(\tilde{t}_{\mathbf{k}} : \mathbf{k} \in \mathbb{K})$. Thus, using the definition of stochastic dominance for random vectors [SS07, Eq. (6.B.4)], we get

$$\mathbb{P}\left(X_t \geq n\right) = \mathbb{E}\left[\Phi((t_{\mathbf{k}} \ : \ \mathbf{k} \in \mathbb{K}))\right]$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[\Phi((\tilde{t}_{\mathbf{k}} \ : \ \mathbf{k} \in \mathbb{K}))\right] \\
&= \mathbb{P}\big(1 + \sharp\{\mathbf{k} \in \mathbb{K} \ : \ \widetilde{T}_{\mathbf{k}} \leq t\} \geq n\big) \\
&= \mathbb{P}\big(\widetilde{X}_t \geq n\big),
\end{aligned}
$$

and (5.5) follows similarly from (5.3) and its counterpart for $\widetilde{N}_t$. □

Finally, from [HP25] we have the following integrability result for the multiplicative progeny of the random tree $\widetilde{\mathcal{B}}^{0,j}$ started from $j \geq 1$.

Proposition 5.5. *(Corollary* **??** *in [HP25]). Let $t \in [0, T]$, $j \geq 0$, $\delta > 0$, $\gamma > 1$, and let $(\sigma(k))_{k \geq 0}$ be a real sequence such that*

$$
0 \leq \sigma(0) < \frac{1}{(1 - e^{-\lambda t})\gamma\delta} \quad \textit{and} \quad 0 \leq \sigma(k) < (k - 2 + \gamma)\delta, \quad k \geq 1.
$$

Then, we have the bound

$$
\mathbb{E}_j\left[\prod_{\mathbf{k} \in \widetilde{\mathcal{K}}_t} \sigma(c_{\mathbf{k}})\right] \leq \frac{e^{-\lambda t}\sigma(j)}{(1 - (1 - e^{-\lambda t})\gamma\delta\sigma(0))^{1+(j-1)/\gamma}} < \infty.
$$

Note that in the special case $\sigma \equiv 1$ we have $\widetilde{N}_t^j \equiv \widetilde{N}_t$ for all $j \geq 0$.

6 Integrability - bounded marks

The goal of this section and the next one is to derive the integrability results needed for the application of Proposition 3.5 and Theorems 4.1-4.2. We fix an initial mark $c = g \in \mathcal{C} \setminus \{\mathrm{Id}\}$, in which case $\mathcal{B}^{0,g}$ is a binary branching chain bearing marks of the form either $\nabla^m f$ or $\nabla^m g$, $m \geq 0$. In Theorems 6.1 and 6.3 we improve on [PP22, Proposition 4.3], see also [BM10, Theorem 3.3], via a detailed study of the integrability of $\mathcal{H}_t(\mathcal{B}^{0,g})$ by investigating the probability generating function of the cardinality $|\mathcal{K}_t|$. In comparison with e.g. Theorem 4.2 of [PP22], Theorems 6.1 and 6.3 provide sufficient conditions on equation coefficients for the (uniform) integrability of the random weights involved in the probabilistic representation of ODE solutions.

Theorem 6.1 (Integrability I)**.** *Let $q \geq 1$ and $g \in \mathcal{C} \setminus \{\mathrm{Id}\}$. Under Assumption 5.3, suppose in addition that $\rho_*(T) := \min_{s \in [0,T]} \rho(s) > 0$ and*

$$
C_0 := \max\left(\sup_{k \geq 0} \frac{|\nabla^k f(x_0)|}{\overline{F}_\rho(T)}, \sup_{k \geq 0} \frac{|\nabla^k g(x_0)|}{\overline{F}_\rho(T)}, \frac{1}{\rho_*(T)}\right) < \frac{1}{\left(1 - e^{-\lambda T}\right)^{1/(2q)}}. \tag{6.1}
$$

Then, we have the L^q bound

$$\mathbb{E}\left[\left|\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right|^q\right] \leq \frac{e^{-\lambda t}C_0^q}{1-(1-e^{-\lambda t})C_0^{2q}}, \quad t \in [0,T],\ c \in \mathcal{C}_f \cup \mathcal{C}_g.$$

Proof. Under Assumption 5.3, $(N_t)_{t\in[0,T]} := (|\mathcal{K}_t|)_{t\in[0,T]}$ is stochastically dominated by $\big(\widetilde{N}_t\big)_{t\in[0,T]}$ by Lemma 5.4. Hence, since $x \mapsto C_0^{qx}$ is increasing, we have

$$\mathbb{E}\left[\left|\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right|^q\right] \leq \mathbb{E}\big[(C_0^q)^{|\mathcal{K}_t|}\big] \leq \mathbb{E}\big[(C_0^q)^{\widetilde{N}_t}\big],$$

see Section 1.A.1 of [SS07]. We conclude from (5.1), which shows that $\mathbb{E}\big[(C_0^q)^{\widetilde{N}_t}\big] < \infty$ since $C_0 < (1-e^{-\lambda T})^{-1/(2q)}$ under (6.1), for $t \in [0,T]$. □

Condition (6.1) involves constraints on both the nonlinearity f and the lifetime density ρ.

Remark 6.2. *Let $\mu > 0$. For the rescaled ODE*

$$x_\mu'(t) = \frac{1}{\mu} f(x_\mu(t)), \quad t \in (0, \mu T],$$

with solution $x_\mu(t) := x(t/\mu)$, C_0 in (6.1) is replaced with

$$C_{0,\mu} := \max\left(\sup_{k\geq 0}\frac{|\nabla^k f(x_0)|}{\mu \overline{F}_\rho(\mu T)}, \sup_{k\geq 0}\frac{|\nabla^k g(x_0)|}{\overline{F}_\rho(\mu T)}, \frac{1}{\rho_*(\mu T)}\right).$$

In this case, the constraint on ρ is now realized as an upper bound on μT, hence a smaller μ close to zero yields a looser constraint on T, but can translate into a stricter constraint on f.

In Lemma A.1 we show the existence of a probability density function $\rho : [0,\infty) \to \mathbb{R}$ satisfying both Assumption 5.3 and Condition (6.1), resp. (6.2), using an upper bound on the existence time T of the solution.

Theorem 6.3 (Uniform integrability I)**.** *Let $q \geq 1$ and $g \in \mathcal{C} \setminus \{\mathrm{Id}\}$. Under Assumption 5.3 and (6.1), suppose in addition that*

$$C_0^q < \frac{\sqrt{4+e^{-2\lambda T}} - e^{-\lambda T}}{2\sqrt{1-e^{-\lambda T}}}, \tag{6.2}$$

where C_0 is defined in (6.1). Then, $\mathcal{H}_t^{g,n}$ defined in (3.6) is L^q-integrable, uniformly in $t \in [0,T]$ and $n \geq 0$. Moreover,

i) if $q > 1$, then $\big(\mathcal{H}_t^{g,n}\big)_{(n,t)\in\mathbb{N}\times\mathbb{R}_+}$ is uniformly integrable;

ii) if

$$\sup_{|x-x_0|<T(1-e^{-\lambda T})^{-1/(2q)}} |\nabla^m g(x)|^q < \frac{1}{\sqrt{1-e^{-\lambda T}}}, \quad m \geq 0, \tag{6.3}$$

then $\widetilde{\mathcal{H}}_t^{g,n}$ *defined in* (3.4) *is* L^q*-integrable, uniformly in* $t \in [0,T]$ *and* $n \geq 0$.

Proof. *i*) One can check that Condition (6.1) is satisfied under (6.2), hence by Theorem 6.1, $\mathcal{H}_t(\mathcal{B}^{0,c})$ is L^q-integrable, uniformly in $t \in [0,T]$ and $c \in \mathcal{C}_f \cup \mathcal{C}_g$. It also follows from Lemma 3.3 that the family

$$\{x_c\}_{c\in\mathcal{C}_f\cup\mathcal{C}_g} = \{\mathbb{E}[\mathcal{H}_t(\mathcal{B}^{0,c})]\}_{c\in\mathcal{C}_f\cup\mathcal{C}_g}$$

solves the ODE system (3.2). On the other hand, by Theorem 6.1 we have

$$\begin{aligned} |x_c(t)|^q &\leq \mathbb{E}\left[\left|\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right|^q\right] \\ &\leq \mathbf{1}_{\{C_0\leq 1\}} + \frac{C_0^q e^{-\lambda T}}{1-(1-e^{-\lambda T})C_0^{2q}}\mathbf{1}_{\{C_0>1\}}. \end{aligned} \tag{6.4}$$

Under (6.2) this yields

$$|x_c(t)|^q < \frac{1}{\sqrt{1-e^{-\lambda T}}}, \qquad c \in \mathcal{C}_f \cup \mathcal{C}_g, \ t \in [0,T],$$

hence it follows from (3.6) that

$$\left|\mathcal{H}_t^{g,n}\left(\mathcal{B}^{0,g}\right)\right|^q \leq (C_1^q)^{|\mathcal{K}_t^{n+1}|} \leq (C_1^q)^{|\mathcal{K}_t|},$$

where

$$C_1^q := \max\left(C_0^q, \sup_{c\in\mathcal{C}_f\cup\mathcal{C}_g,\, t\in[0,T]} |x_c(t)|^q\right) < \frac{1}{\sqrt{1-e^{-\lambda T}}}.$$

From (5.1), we conclude that $\mathcal{H}_t^{g,n}$ is L^q-integrable, uniformly in $t \in [0,T]$ and $n \geq 0$, as in the proof of Theorem 6.1.

ii) By point *i*) above, the family

$$\{x_c\}_{c\in\mathcal{C}_f\cup\{\mathrm{Id}\}} = \{\mathbb{E}[\mathcal{H}_t(\mathcal{B}^{0,c})]\}_{c\in\mathcal{C}_f\cup\{\mathrm{Id}\}}$$

solves the ODE system (3.1a)-(3.1b), and by (6.4) and the Cauchy–Schwartz inequality we have

$$|x_{\mathrm{Id}}(t) - x_0| \leq t^{1-1/q}\left(\int_0^t |x_f(s)|^q\, ds\right)^{1/q} < \frac{T}{(1-e^{-\lambda T})^{1/(2q)}}, \qquad t \in [0,T]. \tag{6.5}$$

In addition, by (3.4) we have

$$\left|\widetilde{\mathcal{H}}_t^{g,n}\left(\mathcal{B}^{0,g}\right)\right|^q \leq (C_2^q)^{|\mathcal{K}_t^{n+1}|} \leq (C_2^q)^{|\mathcal{K}_t|},$$

where

$$C_2^q := \max\left(C_0^q, \sup_{c\in\mathcal{C}_f,\, t\in[0,T]} |x_c(t)|^q, \sup_{c\in\mathcal{C}_g,\, t\in[0,T]} |c\left(x_{\mathrm{Id}}(t)\right)|^q\right) < \frac{1}{\sqrt{1-e^{-\lambda T}}},$$

due to (6.3) and (6.5). As above, we conclude from (5.1). □

Remark 6.4. *When $g = f$, Condition (6.3) provides a quantitative estimate for the neighborhood U in Condition (C2).*

7 Integrability - unbounded marks

In this section we consider a weight function $\sigma(k)$ that depends only on the order k of derivatives in $\nabla^k f$, $\nabla^k g$, $k \geq 0$.

Theorem 7.1 (Integrability II)**.** *Let $T > 0$, $q \geq 1$, $\delta > 0$, and $\gamma \geq 2$. Under Assumption 5.3, suppose that*

$$\rho_*(T) > \frac{1}{\left(1-e^{-\lambda T}\right)^{1/(2q)}}, \tag{7.1}$$

and that

$$\sigma(k) := \frac{|\nabla^k f(x_0)|}{\overline{F}_\rho(T)} \vee \frac{|\nabla^k g(x_0)|}{\overline{F}_\rho(T)}, \qquad k \geq 0, \tag{7.2}$$

satisfies

$$\sigma(0)^{2q} < \frac{1}{(1-e^{-\lambda T})\gamma\delta} \quad \text{and} \quad \sigma(k)^{2q} \leq (k-2+\gamma)\delta, \quad k \geq 1. \tag{7.3}$$

Then, we have the L^q bound

$$\mathbb{E}\left[\left|\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right|^q\right] \leq e^{-\lambda t}\sigma(j)^q \frac{\left(1-(1-e^{-\lambda t})\sigma(0)^{2q}\gamma\delta\right)^{-1/2-(j-1)/(2\gamma)}}{\left(1-(1-e^{-\lambda t})/\rho_*(T)^{2q}\right)^{1/2}}, \quad c \in \{\nabla^j f, \nabla^j g\},$$

$j \geq 0$, $t \in [0,T]$.

Proof. As in the proof of Theorem 6.1, the process $(N_t)_{t\in[0,T]} = (|\mathcal{K}_t|)_{t\in[0,T]}$ is stochastically dominated by $\left(\widetilde{N}_t\right)_{t\in[0,T]}$ under Assumption 5.3, and since $\mathcal{B}^{0,g}$ is a binary tree, from (2.1), for $c \in \{\nabla^j f, \nabla^j g\}_{j\geq 0}$ we have

$$\mathbb{E}_j\left[\left|\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right|^q\right] \leq \mathbb{E}\left[\rho_*(T)^{-(|\mathcal{K}_t|-1)/2} \prod_{\mathbf{k}\in\mathcal{K}_t} \sigma(c^{\mathbf{k}})\right]$$

$$\leq \mathbb{E}\left[\rho_*(T)^{-(|\mathcal{K}_t|-1)q/2}\left(\prod_{\mathbf{k}\in\mathcal{K}_t}\sigma(c^{\mathbf{k}})\right)^q\right]$$

$$\leq \left(\rho_*(T)^q\mathbb{E}\left[\rho_*(T)^{-q|\mathcal{K}_t|}\right]\right)^{1/2}\left(\mathbb{E}\left[\prod_{\mathbf{k}\in\mathcal{K}_t}\sigma(c^{\mathbf{k}})^{2q}\right]\right)^{1/2}.$$

The conclusion then follows from (5.1) and Proposition 5.5. □

An example of a nonlinearity satisfying the growth condition (7.3) is given by $f(x) = x\cos x$.

Theorem 7.2 (Uniform integrability II)**.** *Let $T > 0$, $q \geq 1$, $\delta > 0$, $\gamma \geq 2$, and $g \in \mathcal{C}\setminus\{\mathrm{Id}\}$. Under Assumption 5.3, suppose that (7.1) holds and that the weight function σ in (7.2) satisfies*

$$\sigma(0)^q < \frac{1}{(1-e^{-\lambda T})\gamma\delta}$$

and

$$\sigma(k)^q < e^{\lambda T}\sqrt{\frac{1-(1-e^{-\lambda T})/\rho_*(T)^{2q}}{1-e^{-\lambda T}}}\left(1-(1-e^{-\lambda T})\sigma(0)^q\gamma\delta\right)^{1/2+(k-1)/(2\gamma)}, \quad k \geq 1. \tag{7.4}$$

Then, $\mathcal{H}_t^{g,n}$ defined in (3.6) is L^q-integrable, uniformly in $t \in [0,T]$ and $n \geq 0$. Morover,

i) if $q > 1$, then $\left(\mathcal{H}_t^{g,n}\right)_{(n,t)\in\mathbb{N}\times\mathbb{R}_+}$ is uniformly integrable;

ii) if

$$\sup_{|x-x_0|<T/(1-e^{-\lambda T})^{1/(2q)}}|\nabla^m g(x)| < \frac{1}{(1-e^{-\lambda T})^{1/(2q)}}, \quad m \geq 0, \tag{7.5}$$

then $\widetilde{\mathcal{H}}_t^{g,n}$ defined in (3.4) is L^q-integrable, uniformly in $t \in [0,T]$ and $n \geq 0$.

Proof. *i*) We note that Condition (7.4) implies (7.3), hence (*i*) follows from Theorem 7.1. *ii*) Next, Lemma 3.3 and Theorem 7.1 imply that the family $(x_c = \mathbb{E}[\mathcal{H}_t(\mathcal{B}^{0,c})] : c \in \mathcal{C}_f \cup \{\mathrm{Id}\})$ solves the ODE system (3.1a)-(3.1b), and the family $(x_c = \mathbb{E}[\mathcal{H}_t(\mathcal{B}^{0,c})] : c \in \mathcal{C}_f \cup \mathcal{C}_g)$ solves the ODE system (3.2). Moreover, by Theorem 7.1, for $c \in \{\nabla^k f, \nabla^k g\}$, $k \geq 0$, and all $t \in [0,T]$, we have

$$|x_c(t)|^q \leq \mathbb{E}\left[\left|\mathcal{H}_t\left(\mathcal{B}^{0,c}\right)\right|^q\right] \leq e^{-\lambda T}\sigma(k)^q\frac{\left(1-(1-e^{-\lambda T})\sigma(0)^q\gamma\delta\right)^{-1/2-(k-1)/(2\gamma)}}{\left(1-(1-e^{-\lambda T})/\rho_*(T)^{2q}\right)^{1/2}},$$

hence it follows from (7.4) that

$$|x_c(t)|^q < \left(1-e^{-\lambda T}\right)^{-1/2}, \quad c \in \{\nabla^k f, \nabla^k g\}, \quad k \geq 0,$$

hence (6.5) and (7.5) hold. Hence, as in the proof of Theorem 7.1 we have

$$\left|\widetilde{\mathcal{H}}_t^{g,n}\right|^q \leq (C_2^q)^{|\mathcal{K}_t^{n+1}|} \leq (C_2^q)^{|\mathcal{K}_t|},$$

with

$$\begin{aligned} C_2^q &:= \max\left(\sup_{k\geq 1}\sigma(k)^q, \sup_{c\in\{\nabla^k f,\nabla^k g\},\, t\in[0,T]} |x_c(t)|^q, \sup_{c=\nabla^k g,\, t\in[0,T]} |c\left(x_{\mathrm{Id}}(t)\right)|^q\right) \\ &< (1-e^{-\lambda T})^{-1/2}, \end{aligned}$$

and we conclude from (5.1). □

Remark 7.3. *It can be verified that an explicit and sufficient condition for (7.4) to hold is*

$$\begin{aligned} \sigma(0)^q &\leq \frac{\left(1-e^{-\lambda T}\right)^{-1/2} e^{\lambda T}(1-\left(1-e^{-\lambda T}\right)/\rho_*(T)^{2q})^{1/2}}{1+(1-e^{-\lambda T})^{1/2}\gamma\delta e^{\lambda T}\left(1-(1-e^{-\lambda T})/\rho_*(T)^{2q}\right)^{1/2}}, \\ \sigma(k)^q &\leq \frac{\left(1-e^{-\lambda T}\right)^{-1/2} e^{\lambda T}(1-\left(1-e^{-\lambda T}\right)/\rho_*(T)^{2q})^{1/2}}{\left(1+(1-e^{-\lambda T})^{1/2}e^{\lambda T}\gamma\delta(1-\left(1-e^{-\lambda T}\right)/\rho_*(T)^{2q})^{1/2}\right)^{-1/2-(k-1)/(2\gamma)}}, \quad k\geq 1. \end{aligned}$$

A Existence of ρ

In this section we construct an example of probability density function satisfying Assumption 5.3. In what follows, we let

$$C_1(q;T) := \frac{1}{(1-e^{-\lambda T})^{1/(2q)}} \quad \text{and} \quad C_2(q;T) := \frac{(\sqrt{4+e^{-\lambda T}}-e^{-\lambda T})^{1/q}}{2^{1/q}(1-e^{-\lambda T})^{1/(2q)}}.$$

The inequality (5.4), together with Conditions (6.1) and (6.2), imply

$$1 = \overline{F}_\rho(0) = \int_0^T \rho(r)dr + \overline{F}_\rho(T) > \frac{T}{C_i(q;T)} + e^{-\lambda T},$$

hence

$$T < (1-e^{-\lambda T})C_i(q;T), \quad i=1,2.$$

Given that for any $q\geq 1$ the functions $(0,\infty)\ni T\mapsto (1-e^{-\lambda T})C_i(q;T)$, $i=1,2$, take values in the whole interval $(0,1)$, we have the following.

Lemma A.1. *Let $T>0$, $q\geq 1$, and $i=1$, resp. $i=2$. If $T<1$, then there exists a probability density function $\rho:[0,\infty)\to\mathbb{R}$ satisfying Assumption 5.3 and $1/\rho_*(T) < C_i(q;T)$ for $i=1$, resp. $i=2$.*

Proof. *i*) Case 1: $C_i(q;T) > e^{\lambda T}/\lambda$. In this case, we let

$$\rho(t) := \lambda e^{-\lambda t}, \qquad t \geq 0,$$

for some $\lambda > 0$. Then Assumption 5.3 trivially holds, and $1/\rho_*(T) = e^{\lambda T}/\lambda < C_i(q;T)$.
ii) Case 2: $C_i(q;T) \leq e^{\lambda T}/\lambda$. In this case, we let

$$\rho(t) := \begin{cases} \dfrac{1}{C_i(q;T) - \varepsilon}, & 0 \leq t \leq T, \\ \lambda_2 e^{-\lambda_1 t}, & t > T, \end{cases}$$

for some $\lambda, \lambda_1, \lambda_2 > 0$ and $\varepsilon > 0$. Then the condition $1/\rho_*(T) < C_i(q;T)$ trivially holds. Since ρ is a probability density, we have

$$1 = \int_0^T \rho(t)dt + \int_T^\infty \rho(t)dt = \frac{T}{C_i(q;T) - \varepsilon} + \frac{\lambda_2}{\lambda_1} e^{-\lambda_1 T}, \tag{A.1}$$

hence

$$\frac{\lambda_2}{\lambda_1} e^{-\lambda_1 T} = 1 - \frac{T}{C_i(q;T) - \varepsilon}.$$

Since

$$\overline{F}_\rho(t) = \begin{cases} \dfrac{T - t}{C_i(q;T) - \varepsilon} + \dfrac{\lambda_2}{\lambda_1} e^{-\lambda_1 T}, & 0 \leq t \leq T, \\ \dfrac{\lambda_2}{\lambda_1} e^{-\lambda_1 t}, & t > T, \end{cases}$$

Assumption 5.3 amounts to

$$\lambda_1 \leq \lambda, \quad \frac{\lambda_2}{\lambda_1} e^{-\lambda_1 T} \geq e^{-\lambda T}, \tag{A.2}$$

and

$$\frac{T - t}{C_i(q;T) - \varepsilon} + \frac{\lambda_2}{\lambda_1} e^{-\lambda_1 T} \geq e^{-\lambda t}, \qquad 0 \leq t \leq T. \tag{A.3}$$

By applying (A.1) and the second inequality of (A.2), we get

$$T \leq \left(1 - e^{-\lambda T}\right) \left(C_i(q;T) - \varepsilon\right). \tag{A.4}$$

By applying (A.1) and (A.3), we get

$$\frac{t}{1 - e^{-\lambda t}} \leq C_i(q;T) - \varepsilon, \quad \text{for all } 0 \leq t \leq T,$$

which amounts again to (A.4). In summary, the restriction that ρ is a probability density and Assumption 5.3 together are equivalent to the three conditions: (A.1), the first inequality of (A.2), and (A.4). Since $(1 - e^{-\lambda T})C_i(q;T)$, $i = 1, 2$ take values in the whole interval $(0, 1)$, and recalling the assumption $T<1$, we deduce that there exist $\varepsilon>0$ and $\lambda>0$ such that (A.4) holds. Finally, one can choose appropriate $\lambda_1, \lambda_2 > 0$ satisfying the identity (A.1). These define a probability density ρ satisfying Assumption 5.3 and $1/\rho_*(T) < C_i(q;T)$, as required. $\square$

Remark A.2. *i) Note that in the first case of the above proof, i.e. $C_i(q;T) > e^{\lambda T}/\lambda$, we have*

$$T < \lambda T e^{-\lambda T} C_i(q;T), \tag{A.5}$$

and the suprema of the above right hand sides belong to $(0,1/2)$ for $i=1,2$. Hence, when $T\in(0,1/2)$ there exists $\lambda>0$ such that (A.5) holds, and one can thereby choose ρ as the exponential density

$$\rho(t) := \lambda e^{-\lambda t}, \qquad t \geq 0.$$

In this case, conditions (6.1) and (6.2) reduce respectively to

$$\max\left(\sup_{m\geq 0}|\nabla^m f(x_0)|, \sup_{m\geq 0}|\nabla^m g(x_0)|\right) < e^{-\lambda T} C_i(q;T), \quad i=1,2,$$

where the above right hand sides are both increasing in $e^{-\lambda T}\in(0,1)$. Thus, if T is fixed in $(0,1/2)$ then a smaller $\lambda>0$ yields looser constraints on the nonlinearity, while if $\lambda>0$ is fixed, then a larger $T\in(0,1/2)$ requires stricter constraints on the nonlinearity.

ii) As seen in Remark 6.2, the upper bounds on T can be adjusted by a rescaling. In this case, the condition on the nonlinearity f will be adjusted accordingly.

B Branching trees *vs.* Butcher trees

Recall that a rooted tree $\tau = (V, E, \bullet)$ is a nonempty set V of vertices and a set of edges E between some of the pairs of vertices, with a specific vertex $\bullet$ called the root, such that the graph (V,E) is connected with no loops. We call the two special trees $\emptyset$ and $\bullet$ empty tree and dot tree respectively. We denote by $\mathbf{T}$ the set of all rooted trees, and by $\mathbf{T}_n$, $n\geq 0$, the subset of $\mathbf{T}$ consisting of all trees with order n. The following notion of grafting product is a generalization of the notion of beta product from unlabeled trees, see [But21, Section 2.1], to labeled trees.

Definition B.1 (Grafting product). *i) Given τ_1, τ_2 two labeled trees and $l \in \{1, \ldots, |\tau_1|\}$, the grafting-product with label l of τ_1 and τ_2, denoted by $\tau_1 *_l \tau_2$, is the tree of order $|\tau_1| + |\tau_2|$ formed by grafting (attaching) τ_2 from its root to the vertex l of τ_1, so that the whole τ_2 are descendants of vertex l.*

ii) The new tree is labeled by keeping all labels of τ_1, and adding all labels of τ_2 by $|\tau_1|$.

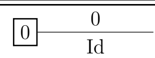
iii) For any labeled tree τ, we let $\emptyset *_0 \tau = \tau *_l \emptyset = \tau$ for all $0 \le l \le |\tau|$, and keep the labels of τ.

For $t \geq 0$ we consider the finite random tree $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ obtained by killing all offsprings of $\mathcal{B}^{0,\mathrm{Id}}$ that are born beyond time t. The (random) leaves of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ coincide with the set $\mathcal{K}_t^\partial$. Associated to every sample tree of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$, we draw a labeled and marked Butcher tree denoted by $\mathcal{T}^{\mathrm{Id}}$ via the following recursive algorithm based on the sequence of splitting times of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ sorted in increasing order.[1]

[1] This is possible since the splitting times of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ are a.s. distinct.

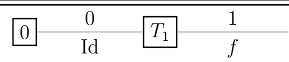
Algorithm 1 Recursive construction of the labeled tree $\mathcal{T}^{\mathrm{Id}}$.

Initialization: Starting from an initial branch with label 0 and mark Id, we initialize an empty tree $\emptyset$ with label 0 and mark Id.

Splitting	Branching tree $\mathcal{B}^{0,\mathrm{Id}}\|_{[0,0]}$	Labeled tree $\mathcal{T}^{\mathrm{Id}}$	$\prod_{\mathbf{k}\in\mathcal{K}_0^{\partial}} c^{\mathbf{k}}$
0-th		${}_0\varnothing_{\mathrm{Id}}$	1

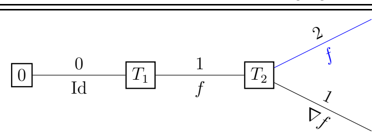
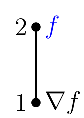
If only one splitting time occurred before time t,

- the initial branch yields one offspring with label 1 and mark f, and
- we update the initial labeled tree $\emptyset$ to $\emptyset *_0 \bullet = \bullet$, and mark this root with f.

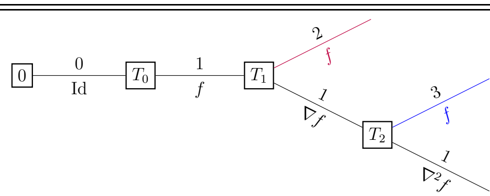
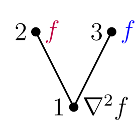
Splitting	Branching tree $\mathcal{B}^{0,\mathrm{Id}}\|_{[0,t]}$	Labeled tree $\mathcal{T}^{\mathrm{Id}}$	$\prod_{\mathbf{k}\in\mathcal{K}_t^{\partial}} c^{\mathbf{k}}$
1st		$1\bullet$f	f

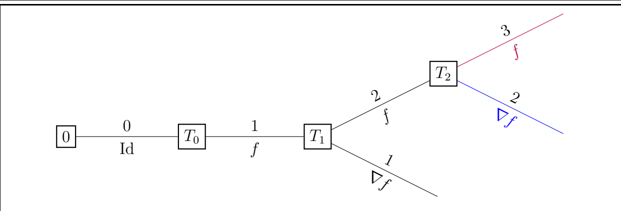
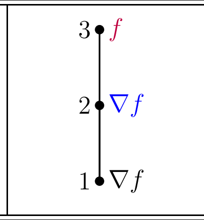
Recursion: Assume that a marked labeled tree τ with order $|\tau| = i \geq 1$ has been constructed, with *i-th* splitting time not later than t.

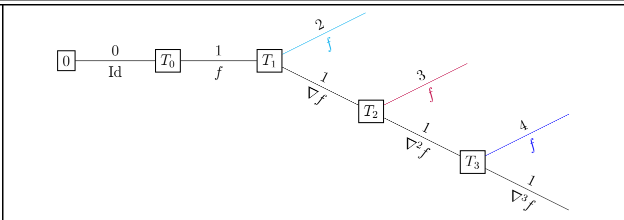
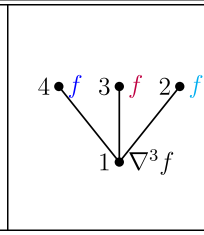
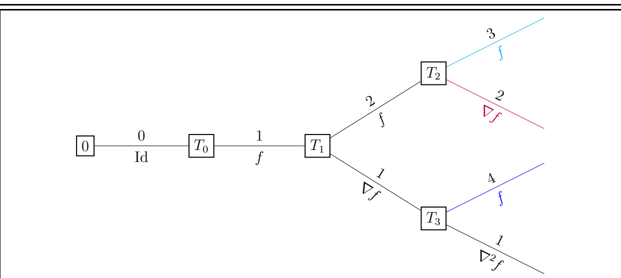
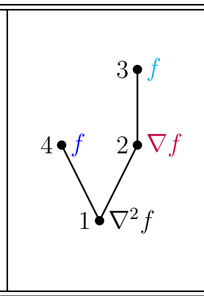
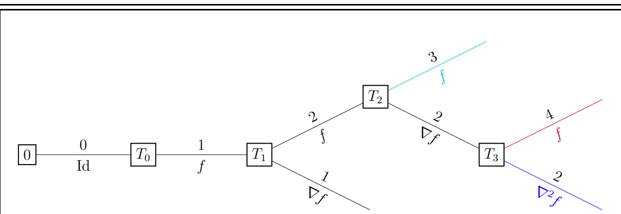
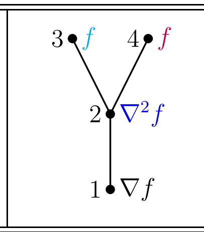
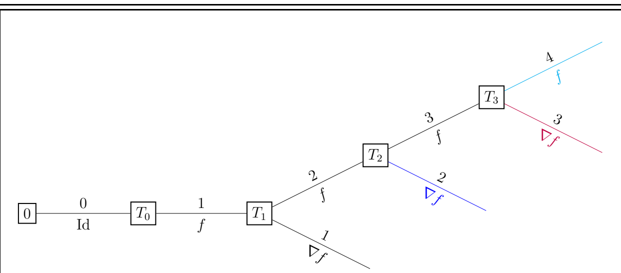
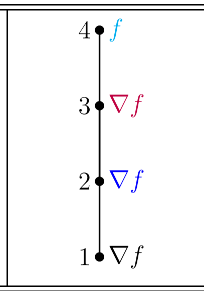
- If the $(i+1)$-*th* splitting time of a branch with label $l \geq 1$ and mark $\nabla^m f$, $m \geq 0$, is not later than t it then gives two offsprings, respectively with label $i+1$ and mark f, and with label l and mark $\nabla^{m+1} f$.
- We update the labeled tree τ to $\tau *_l \bullet$. We also update the mark of the vertex l to $\nabla^{m+1} f$ and mark the new vertex $i+1$ with f.

Splitting	Branching tree $\mathcal{B}^{0,\mathrm{Id}}\|_{[0,t]}$	Labeled tree $\mathcal{T}^{\mathrm{Id}}$	$\prod_{\mathbf{k}\in\mathcal{K}_t^{\partial}} c^{\mathbf{k}}$
2nd			$(\nabla f)f$

Since $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ has finite splitting, the above induction will end in finite steps. The following graphs further illustrate this recursive construction.

Splitting	Branching tree $\mathcal{B}^{0,\mathrm{Id}}\|_{[0,t]}$	Labeled tree $\mathcal{T}^{\mathrm{Id}}$	$\prod_{\mathbf{k}\in\mathcal{K}_t^{\partial}} c^{\mathbf{k}}$
3rd			$(\nabla^2 f)f^2$

3rd			$(\nabla f)^2 f$

Splitting	Branching tree $\mathcal{B}^{0,\mathrm{Id}}\vert_{[0,t]}$	Labeled tree $\mathcal{T}^{\mathrm{Id}}$	$\prod_{\mathbf{k}\in\mathcal{K}_t^\partial} c^{\mathbf{k}}$
4-th			$(\nabla^3 f) f^3$
4-th			$(\nabla^2 f)(\nabla f) f^2$
4-th			$(\nabla f)(\nabla^2 f) f^2$
4-th			$(\nabla f)^3 f$

The above construction relies on two inductive assumptions, which are proved in the next lemma by induction on splitting times.

Lemma B.2. *i) The labeled tree* τ *drawn at the* i*-th splitting of* $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ *has order* $|\tau| = i \geq 1$*, its vertex labels form permutation of* $\{0, \ldots, |\tau|\}$*, and each vertex with* m *descendants is marked by* $\nabla^m f$*, with* $m \leq i - 1$*.*

ii) Right after the i-th splitting of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$, $i \geq 1$, the labels of all living branches form a permutation of $\{0, \ldots, i\}$. If the label of a living branch coincides with the label of a vertex of the tree τ that we have drawn, then their marks also coincide.

Proof. i) The statements clearly hold for the first splitting. Suppose that they hold for the tree τ drawn at the i-th splitting, $i \geq 1$. In the $(i+1)$-th splitting, the new tree $\tau *_l \bullet$ is a labeled tree with order $|\tau| + 1 = i + 1$. Compared to τ, the only vertex whose descendant number changes in the new tree $\tau *_l \bullet$ is the vertex l. This vertex has mark $\nabla^m f$ in τ, so by the inductive assumption, it has m descendants with $m \leq i$. In the new tree $\tau *_l \bullet$, the vertex l has one more descendant than in τ, and its mark is updated to $\nabla^{m+1} f$.

ii) The statements hold for the first splitting, since the only living branch right after this splitting must be the only child of the initial branch, which has label 1 and mark f, which is the same as the mark of the root we have drawn. Suppose that the statements hold for the i-th splitting, $i \geq 1$. Then by construction, at the $(i+1)$-th splitting, the branch with label l dies and gives two offsprings are labeled by $i+1$ and l, while all other living branches have distinct labels forming a permutation of $\{0, \ldots, i\} \setminus \{l\}$. Thus, the labels of all these living branches have distinct labels forming a permutation of $\{0, \ldots, i+1\}$. Again by construction, the two offsprings with labels $i+1$ and l have marks f and $\nabla^{m+1} f$, which are the same as those of the vertex $i+1$ and l in the new drawn tree, while all other living branches also share the same marks with the vertices whose labels equal to their labels, by the inductive assumption. This completes the proof. □

The next result is consequence of Lemma B.2.

Corollary B.3. *i) Each vertex of $\mathcal{T}^{\mathrm{Id}}$ having m descendants has the mark $\nabla^m f$, for $m = 1, \ldots, |\mathcal{T}^{\mathrm{Id}}| - 1$.*

ii) The labels of the leaves of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ are distinct and they constitute a permutation of $\{0, \cdots, |\mathcal{K}_t^\partial|\}$. If the label of a leave of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ coincides with the label of a vertex of $\mathcal{T}^{\mathrm{Id}}$, then their marks also coincide.

Next is the definition of elementary differentials, cf. [But10, § 3.3].

Definition B.4. *The elementary differential of f is the mapping $F : \mathbf{T} \to \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$ defined recursively by $F(\emptyset) = \mathrm{Id}$, $F(\bullet) = f$, and*

$$F(\tau) = \nabla^m f(F(\tau_1), \ldots, F(\tau_m))$$

for $\tau = [\tau_1, \ldots, \tau_m]$.

When $|\tau| = n$, we have

$$F(\tau) = \nabla^{m_1} f(\nabla^{m_2} f(\cdots), \cdots, \cdots(\cdots, f)\cdots) \tag{B.1}$$

for a sequence $(m_i)_{i=1,\ldots,n}$ of integers satisfying $m_n = 0$ and $\sum_{i=1}^n m_i = n-1$. For a given tree $\tau \in \mathbf{T}$, the map F also provides a way to mark each vertex of τ by f or its derivatives: each vertex with no descendants is marked by f; the vertices with m descendants is marked by $\nabla^m f$, for $m \geq 0$.

Lemma B.5. *Letting* $\mathcal{T}^{\mathrm{Id}}$ *be the random labeled tree defined in Algorithm 1, we have*

$$\prod_{\mathbf{k} \in \mathcal{K}_t^\partial} c^{\mathbf{k}} = F(\mathcal{T}^{\mathrm{Id}}). \tag{B.2}$$

Proof. By *ii*) of Corollary B.3, the composition of marks over the leaves of $\mathcal{B}^{0,\mathrm{Id}}|_{[0,t]}$ coincides with the composition of marks of all vertices of $\mathcal{T}^{\mathrm{Id}}$. From the construction of the map F and *i*) of Corollary B.3, we conclude that the product $\prod_{\mathbf{k}\in\mathcal{K}_t^\partial} c^{\mathbf{k}}$ coincides with $F(\mathcal{T}^{\mathrm{Id}})$. $\square$

Acknowledgement. This research is supported by the National Research Foundation, Singapore. The work of Q. Huang is supported by the Start-Up Research Fund of Southeast University under Grant No. RF1028624194 and the Jiangsu Provincial Scientific Research Center of Applied Mathematics under Grant No. BK20233002.

References

[AN72] K.B. Athreya and P.E. Ney. *Branching processes*, volume Band 196 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, New York-Heidelberg, 1972.

[BM10] Y. Bakhtin and C. Mueller. Solutions of semilinear wave equation via stochastic cascades. *Commun. Stoch. Anal.*, 4(3):425–431, 2010.

[But63] J.C. Butcher. Coefficients for the study of Runge-Kutta integration processes. *J. Austral. Math. Soc.*, 3:185–201, 1963.

[But10] J.C. Butcher. Trees and numerical methods for ordinary differential equations. *Numerical Algorithms*, 53:153–170, 2010.

[But16] J.C. Butcher. *Numerical methods for ordinary differential equations.* John Wiley & Sons, Ltd., Chichester, third edition, 2016.

[But21] J.C. Butcher. *B-Series: Algebraic Analysis of Numerical Methods*, volume 55 of *Springer Series in Computational Mathematics.* Springer, Cham, 2021.

[CL84] E. A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations.* Krieger Publishing Company, Florida, 1984.

[DB02] P. Deuflhard and F. Bornemann. *Scientific Computing with Ordinary Differential Equations*, volume 42 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2002.

[DMTW19] R. Dascaliuc, N. Michalowski, E. Thomann, and E.C. Waymire. Complex Burgers equation: a probabilistic perspective. In *Sojourns in probability theory and statistical physics. I. Spin glasses and statistical mechanics, a Festschrift for Charles M. Newman*, volume 298 of *Springer Proc. Math. Stat.*, pages 138–170. Springer, Singapore, 2019.

[FH65] R. Feynman and A. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw-Hill, 1965.

[Har63] T.E. Harris. *The theory of branching processes*, volume Band 119 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin; Prentice Hall, Inc., Englewood Cliffs, NJ, 1963.

[HLOT+19] P. Henry-Labordère, N. Oudjane, X. Tan, N. Touzi, and X. Warin. Branching diffusion representation of semilinear PDEs and Monte Carlo approximation. *Ann. Inst. H. Poincaré Probab. Statist.*, 55(1):184–210, 2019.

[HP23] Q. Huang and N. Privault. On the random generation of Butcher trees. Preprint arXiv:2404.05969, 2023.

[HP25] Q. Huang and N. Privault. Binary Galton–Watson trees with mutations. Preprint arXiv:2501.10951, 2025.

[INW69] N. Ikeda, M. Nagasawa, and S. Watanabe. Branching Markov processes I, II, III. *J. Math. Kyoto Univ.*, 8-9:233–278, 365–410, 95–160, 1968-1969.

[Ken48] D.G. Kendall. On the generalized "birth-and-death" process. *Ann. Math. Statistics*, 19:1–15, 1948.

[LS94] S.G. Lobanov and O.G. Smolyanov. Ordinary differential equations in locally convex spaces. *Russian Mathematical Surveys*, 49(3):97, 1994.

[LS97] Y. Le Jan and A. S. Sznitman. Stochastic cascades and 3-dimensional Navier-Stokes equations. *Probab. Theory Related Fields*, 109(3):343–366, 1997.

[McK75] H.P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Comm. Pure Appl. Math.*, 28(3):323–331, 1975.

[NPP23] J.Y. Nguwi, G. Penent, and N. Privault. A fully nonlinear Feynman-Kac formula with derivatives of arbitrary orders. *Journal of Evolution Equations*, 23:Paper No. 22, 29pp., 2023.

[PP22] G. Penent and N. Privault. Numerical evaluation of ODE solutions by Monte Carlo enumeration of Butcher series. *BIT Numerical Mathematics*, 62:1921–1944, 2022.

[Sko64] A.V. Skorokhod. Branching diffusion processes. *Teor. Verojatnost. i. Primenen.*, 9:492–497, 1964.

[SS07] M. Shaked and J.G. Shanthikumar. *Stochastic orders*. Springer Science+Business Media, LLC, 2007.