University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

A PRIORI BOUNDS FOR ROUGH DIFFERENTIAL EQUATIONS WITH A NON-LINEAR DAMPING TERM

TIMOTHEE BONNEFOI, AJAY CHANDRA, AUGUSTIN MOINAT, AND HENDRIK WEBER

ABSTRACT. We consider a rough differential equation with a non-linear damping drift term:

$$dY(t) = -|Y|^{m-1}Y(t)dt + \sigma(Y(t))dX(t),$$

where $m > 1$, X is a (branched) rough path of arbitrary regularity $\alpha > 0$, and where σ is smooth and satisfies an m and α-dependent growth property. We show a strong a priori bound for Y, which includes the "coming down from infinity" property, i.e. the bound on $Y(t)$ for a fixed $t > 0$ holds uniformly over all choices of initial datum $Y(0)$.

The method of proof builds on recent work on a priori bounds for the ϕ^4 SPDE in arbitrary subcritical dimension [CMW19]. A key new ingredient is an extension of the algebraic framework which permits to derive an estimate on higher order conditions of a coherent controlled rough path in terms of the regularity condition at lowest level.

1. INTRODUCTION

Rough path theory [Lyo94, Lyo98, Gub03, Gub06] was developed from the late 90s as a solution theory for differential equations of the form

$$(1.1) \qquad dY(t) = \sigma(Y(t))dX(t)$$

for irregular drivers $X(t)$, typically of Hölder regularity $\alpha < \frac{1}{2}$. The principal example to have in mind is the case of random drivers. In particular, when X is a Brownian motion, rough path theory provides an alternative approach to the classical theory of Itô stochastic differential equations (SDEs). Since its inception in the late 90s the theory has been highly successful; early applications included the treatment of SDEs for correlated signals, such as fractional Brownian motion, streamlined proofs for several classical results including large deviations, support theorems and the construction of stochastic flows for SDEs, see e.g. the monographs [LQ02, FV10b, FH14]. More recent results include the proof that certain deterministic fast-slow dynamical systems are well-approximated by systems of SDEs [KM16] or the development of a Malliavin calculus for stochastic Navier Stokes equations with multiplicative noise [GH19].

In this article we are concerned with deriving a priori bounds and the related issue of non-explosion for (1.1). Rough path theory primarily addresses the small scale roughness of solutions and the associated problem to define and control the rough integrals $\int \sigma(Y(t))dX(t)$. There are relatively few results on the control of the global behaviour of solutions: in monographs, global existence for (1.1) is typically shown under the restrictive assumptions that σ and all of its derivatives up to order N are bounded (see e.g. [FV10b, Theorem 10.14], [FH14, Theorem 8.4]). Here and throughout the paper we denote by N the unique natural number satisfying

$$N\alpha \leq 1 < (N+1)\alpha.$$

Relaxing these growth conditions turns out to be intricate, and in fact global existence may fail, e.g. for bounded σ with unbounded derivative [LS11]. There are however some results which relax the boundedness assumptions on the full tensors $D^\beta\sigma$, $|\beta| \leq N$, by showing that a control on certain combinations of these derivatives is sufficient [Dav08, Lej12, Wei18].

We are interested in rough differential equations with drift

$$dY(t) = b(Y(t))dt + \sigma(Y(t))dX(t).$$

The striking counter-example [CHJ13, Section 3.3] suggests that even in the case $\sigma = 1$ (additive noise) a standard Lyapunov condition $b(y) \cdot y \leq C(|y|^2 + 1)$ alone may not suffice to exclude finite-time explosion. This issue was addressed in [RS17], where it was shown that if σ and its derivatives up to order N are bounded, then the more restrictive Lyapunov condition

$$(1.2) \qquad b(y) \cdot y \leq C(|y|^2 + 1) \qquad \text{and} \qquad \left| b(y) - \frac{(b(y) \cdot y)y}{|y|^2} \right| \leq C(1 + |y|) \;,$$

suffices to rule out finite-time explosion. In this article we treat the case of super-linear inward drift b. For simplicity, we restrict to b of polynomial form, i.e. we deal with

$$dY(t) = -|Y(t)|^{m-1}Y(t)dt + \sigma(Y(t))dX(t) \qquad t \in (0,1]$$
$$(1.3) \qquad\qquad Y(0) = y_0 \in \mathbb{R}^k \;,$$

where $m > 1$, X is an $\mathbb{R}^d$ valued (branched) rough path of regularity $\alpha \in (0,1)$ in the sense of [Gub06, HK15, BCFP19] and σ is assumed to be a C^N map $\mathbb{R}^k \to \mathbb{R}^{k \times d}$, where $\mathbb{R}^{k \times d}$ denotes the space of $k \times d$ matrices. This drift does satisfy (1.2) and therefore the results of [RS17] imply non-explosion if σ and its derivatives are bounded. In this article, we use the damping effect of the drift to get stronger estimates on solutions and to relax the boundedness conditions on σ. More specifically, we treat coefficients σ satisfying either one of the following two growth assumptions:

Assumption 1.1 (Bounded coefficients). *There exists a constant $C_\sigma < \infty$ such that for each multi-index $\beta = (\beta_1, \ldots, \beta_k) \in \mathbb{N}^k$ of length $|\beta| := \sum_{j=1}^k \beta_j \leq N$ we have*

$$\sup_{x \in \mathbb{R}^d} |\partial^\beta \sigma(x)| \leq C_\sigma.$$

Assumption 1.2 (Coefficients of polynomial growth). *There exists a $\gamma > 0$ and a constant $C_\sigma < \infty$ such that for each multi-index $\beta \in \mathbb{N}^k$ of length $|\beta| \leq N$ and for all $x \in \mathbb{R}^k$ we have*

$$|\partial^\beta \sigma(x)| \leq C_\sigma \langle x \rangle^{\gamma - |\beta|} \;,$$

where above and below, $\langle x \rangle = (1 + |x|^2)^{\frac{1}{2}}$.

Remark 1.3. *Assumption 1.2 is satisfied, for example, in the scalar case $d = k = 1$ and for*

$$\sigma(x) = \langle x \rangle^\gamma \;.$$

Our first main result is the following estimate:

Theorem 1.4 (Coming down from infinity). *Let $\mathbf{X}$ be an α-rough path for some $\alpha \in (0,1)$ and let Y solve (1.3).*

(1) If σ satisfies the growth Assumption 1.1 we have for $t \in (0,1]$

$$(1.4) \qquad\qquad |Y(t)| \leq C \max \left\{ t^{-\frac{1}{m-1}} \;, \; \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{m\alpha|h|}} \right\} \;.$$

(2) If σ satisfies the growth Assumption 1.2 for

$$(1.5) \qquad\qquad 1 \leq \gamma < (m-1)\alpha + 1 \;,$$

then we have for $t \in (0,1]$

$$(1.6) \qquad\qquad |Y(t)| \leq C \max \left\{ t^{-\frac{1}{m-1}} \;, \; \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{((m-1)\alpha - \gamma + 1)|h|}} \right\} \;.$$

In both statements, $C < \infty$ denotes a constant that depends on d, k, m, α and C_σ.

The definition of a (branched) rough path is recalled in Definition 2.5, while the precise notion of solution to a rough differential equation is given in Definition 2.12. We mention right away that the solution $\mathbf{Y}$ of a rough differential equation as in Definition 2.12 has to be viewed as taking values in a larger space than $\mathbb{R}^k$ (a space spanned by decorated forests). Here $Y(t) \in \mathbb{R}^k$ is the "observable solution" which corresponds to the coefficient of $\mathbf{1}$. We

write $|h|$ for the order of a tree, that is the number of its vertices, and $\mathring{\mathcal{T}}_{\leq N}$ is the set of *trees* of order $\leq N$, see Section 2.4. The seminorm $[\mathbf{X} : h]$ controls the order bound for $\mathbf{X}$ at level h - it can be thought of as a Hölder type bound on a (postulated) $|h|$-fold iterated integral of X - see (2.3).

Theorem 1.5 (Small times). *Let Y solve (1.3), for an α-rough path $\mathbf{X}$.*

(1) *Let σ satisfy the growth Assumption 1.1 and for $\varepsilon_1 = \varepsilon_1(m)$ and $\varepsilon_2 = \varepsilon_2(\alpha, d, k, C_\sigma)$ small enough set*

$$(1.7) \qquad T_1 = \varepsilon_1 \frac{1}{\langle y_0 \rangle^{m-1}} , \qquad T_2 = \varepsilon_2 \min \left\{ \frac{1}{[\mathbf{X} : h]^{\frac{1}{|h|\alpha}}} : h \in \mathring{\mathcal{T}}_{\leq N} \right\} .$$

Then for $t \leq \min\{T_1, T_2\}$ we have

$$(1.8) \qquad \langle Y(t) \rangle \leq 2\langle y_0 \rangle .$$

(2) *Let σ satisfy the growth Assumption 1.2 for an exponent γ satisfying (1.5), let T_1 be as in (1.7) and for $\varepsilon_2 = \varepsilon_2(\alpha, d, k, C_\sigma)$ small enough set*

$$(1.9) \qquad T_2 = \varepsilon_2 \left(\frac{1}{\langle y_0 \rangle} \right)^{\frac{\gamma-1}{\alpha}} \min \left\{ \frac{1}{[\mathbf{X} : h]^{\frac{1}{|h|\alpha}}} : h \in \mathring{\mathcal{T}}_{\leq N} \right\} .$$

Then for $t \leq \min\{T_1, T_2\}$ we have

$$(1.10) \qquad \langle Y(t) \rangle \leq 2\langle y_0 \rangle .$$

Theorems 1.4 and 1.5 can be combined into a single bound.

Corollary 1.6. *Let Y solve (1.3), for an α-rough path $\mathbf{X}$ for an $\alpha \in (0, 1)$.*

(1) *If σ satisfies the growth Assumption 1.1 we have for $t \in [0, 1]$*

$$(1.11) \qquad |Y(t)| \leq C \max \left\{ \min\{t^{-\frac{1}{m-1}}, |y_0|\} , \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{(m-1)\alpha|h|}}, 1 \right\} .$$

(2) *If σ satisfies the growth Assumption 1.2 for γ satisfying (1.5), then we have for $t \in [0, 1]$*

$$(1.12) \qquad |Y(t)| \leq C \max \left\{ \min\{t^{-\frac{1}{m-1}}, |y_0|\} , \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{((m-1)\alpha-\gamma+1)|h|}}, 1 \right\} .$$

In both statements, $C < \infty$ denotes a constant that depends on d, k, m, α and C_σ.

Remark 1.7. *Consider the case $\sigma = 0$, where (1.3) reduces to the ordinary differential equation*

$$\dot{Y} = -|Y|^{m-1} Y \qquad t \in (0, 1)$$
$$Y(0) = y_0 .$$

The explicit solution of this equation is given by

$$(1.13) \qquad Y(t) = \frac{y_0}{|y_0|} \left(\frac{1}{(m-1)t + |y_0|^{1-m}} \right)^{\frac{1}{m-1}} ,$$

and the elementary estimate

$$(1.14) \qquad \frac{1}{C(m)} \min\{|y_0|, t^{-\frac{1}{m-1}}\} \leq |RHS \text{ of } (1.13)| \leq C(m) \min\{|y_0|, t^{-\frac{1}{m-1}}\}$$

shows that the first term on the RHS of (1.11) and (1.12) is optimal up to constants.

Remark 1.8 (On the optimality of the growth condition (1.5) on σ). *We compare our result to the case of the Itô stochastic differential equation*

$$dY(t) = -|Y(t)|^{m-1} Y(t) dt + \sigma(Y(t)) dW(t) \qquad t \in (0, 1] ,$$

for a Brownian motion W, for simplicity in the scalar case $k = n = 1$. Itô's formula yields for any $p \geq 2$

$$(1.15) \qquad d\frac{1}{p}|Y(t)|^p = -|Y(t)|^{p-1+m}dt + dM_t + \frac{(p-1)}{2}|Y(t)|^{p-2}\sigma(Y(t))^2 dt \,,$$

for a (local) martingale M_t. Under Assumption 1.2 the Itô correction term on the RHS of (1.15) satisfies

$$\frac{(p-1)}{2}|Y(t)|^{p-2}\sigma(Y(t))^2 \leq \frac{(p-1)}{2}C_\sigma|Y(t)|^{p-2+2\gamma} \,.$$

If the coefficient σ satisfies the upper bound in (1.5) for an $\alpha < \frac{1}{2}$, then the exponent on the RHS satisfies

$$p - 2 + 2\gamma < p - 1 + m \,,$$

which is precisely enough to control this term by the damping term $-|Y(t)|^{p-1+m}$. This calculation suggests the optimality of Assumption 1.2, at least in this case.

Remark 1.9 (More on the growth condition (1.5) on σ). *The reason for the limitation $\gamma \geq 1$ in (1.5) is that in several points in the proof (see e.g. (4.27), (4.29)) the estimate*

$$\sup_{s \in I}|Y(s)|^{(\gamma-1)|h|} \leq \|Y\|_{[t^\star,1]}^{(\gamma-1)|h|},$$

is used crucially. Here I is some (short) interval contained in the (possibly much larger) $[t^\star, 1]$ and $|h|$ is a positive number. Clearly, this estimate only holds for $\gamma \geq 1$. For $\gamma \leq 1$ our proof could be modified slightly to get the estimate

$$|Y(t)| \leq C \max\left\{t^{-\frac{1}{m-1}}, \max_{h \in \mathring{\mathcal{T}}_{\leq N}}[\mathbf{X}:h]^{\frac{1}{((m-1)\alpha)|h|}}\right\} \,.$$

Unfortunately, this bound does not reflect the behaviour of σ and its derivatives at infinity in the exponents of $[\mathbf{X}:h]$.

Remark 1.10 (On optimality of exponents). *A particularly important example, where our result applies is the rough path $\mathbf{X}$ built on top of a fractional Brownian motion X of Hurst parameter $H > \frac{1}{4}$. It is well-known, see e.g. [FV10a], that in this case $\mathbf{X}$ is an α-rough path for $\alpha < H$ and for $h \in \mathring{\mathcal{T}}_{\leq N}$ we have the Gaussian tail estimate*

$$(1.16) \qquad \mathbb{E}\left[\exp\left(\varepsilon[\mathbf{X}:h]^{\frac{2}{|h|}}\right)\right] < \infty,$$

for $\varepsilon > 0$ small enough. In this case, the estimates (1.4) and (1.6) imply stretched exponential tails for the solution Y of the random rough differential equation (1.3). More precisely, (1.4) combined with (1.16) implies that under Assumption 1.1 we have that for $\varepsilon' > 0$ small enough

$$\mathbb{E}\left[\exp\left(\varepsilon' \sup_{t \in [\frac{1}{2},1]}|Y(t)|^{2\alpha m}\right)\right] < \infty \,,$$

while (1.6) combined with (1.16) implies that under Assumption 1.2 and (1.5) we have

$$\mathbb{E}\left[\exp\left(\varepsilon' \sup_{t \in [\frac{1}{2},1]}|Y(t)|^{2(\alpha(m-1)-\gamma+1)}\right)\right] < \infty \,.$$

We do not know if these stretched exponential tails are optimal. The additive noise case, $\sigma = 1$ was considered in [MW20a, Section 2], and there the improved stochastic integrability

$$\mathbb{E}\left[\exp\left(\varepsilon' \sup_{t \in [\frac{1}{2},1]}|Y(t)|^{2+2\alpha(m-1)}\right)\right] < \infty,$$

was obtained. Additionally, in this case comparison with an Itô SDE shows the optimality of exponent, at least in the case of Brownian motion. Martingale estimates suggest that in the case of a Brownian motion W and in the case of bounded σ the stochastic integrability in the multiplicative noise case should coincide with the additive noise case, but our argument does not imply this. The main technical reason is, that in the proof we need to impose the

condition (4.4) *on the parameters and this condition does not have a counterpart in the simpler argument for the additive noise case.*

Our argument follows the philosophy developed in [MW20a, MW20b, CMW19] in the context of the dynamic ϕ^4 model. These articles deal primarily with singular stochastic partial differential equations, but as already mentioned above in [MW20a, Section 2] the case of an ordinary differential equation with irregular driver is discussed. More precisely, an estimate for the equation

$$(1.17) \qquad Y(t) = Y(0) - \int_0^t |Y(s)|^{m-1} Y(s) ds + Z(t)$$

for Z of Hölder regularity $\alpha \in (0,1)$ is derived. The argument roughly goes as follows: first the equation is regularised by convolution with a mollification kernel at scale L, resulting in a non-closed equation for the regularised function Y_L,

$$\dot{Y}_L = -|Y_L|^{m-1} Y_L + Z_L + E_L$$

which involves an error term E_L accounting for the fact that non-linearity and smoothing do not commute. Afterwards, a standard ODE estimate (see e.g. [TW18, Lemma 3.8]) is applied to the regularised equation, and finally, the commutator error E_L is controlled by a simple *regularity estimate* for Y. The following lemma is an intermediate step in this analysis and the starting point for our argument:

Lemma 1.11 ([MW20a, Lemma 2.1, Intermediate result in Step 4]). *Let $m > 1$ and let $Y \in C([0,1], \mathbb{R}^d)$ satisfy equation* (1.17) *for some $Z \in C([0,1], \mathbb{R}^d)$ with $Z(0) = 0$. Then for $\alpha \in (0,1)$, $t \in (0,1]$ and $L \in (0,t)$ we have*

$$|Y(t)| \leq C \max \left\{ (t-L)^{-\frac{1}{m-1}}; \left(\sup_{s \in [L,t]} [Z]_{\alpha,[s-L,s]} L^{\alpha-1} \right)^{\frac{1}{m}}; \right.$$

$$(1.18) \qquad \left. \left(L^\alpha \|Y\|_{[0,t]}^{m-1} \sup_{s \in [L,t]} [Y]_{\alpha,[s-L,s]} \right)^{\frac{1}{m}}; \left(L^\alpha [Y]_{\alpha,[t-L,t]} \right) \right\}.$$

Here $C = C(m)$ and

$$\|Y\|_{[0,t]} = \sup_{s \in [0,t]} |Y(s)| \qquad and \qquad [Z]_{\alpha,[s-L,s]} = \sup_{s-L \leq s_1 < s_2 \leq s} \frac{|Z(t) - Z(s)|}{|s_2 - s_1|^\alpha}.$$

In order to apply this Lemma to the rough differential equation (1.3) it thus remains to control α-Hölder norms of Y and of the rough integral

$$Z(t) = \int_0^t \sigma(Y(s)) dX(s) .$$

Gubinelli's Sewing Lemma, see Proposition 2.11 below, yields a control Z in terms of the order bounds $[\mathbf{X} : f]$ as well as the appropriate semi-norms $[\sigma(\mathbf{Y}) : f]$ that control the regularity of the *controlled path* $\sigma(\mathbf{Y})$ at various levels f (f takes values in a set of *decorated forests*, see below). A key step of our argument is the development of an algebraic framework which yields an explicit expression, Corollary 3.9, for the remainder expressions (we note that a similar formula appeared earlier in [Gub06, Theorem 5.2]).

This expression can then be combined with a Taylor remainder formula yielding a control on $[\sigma(\mathbf{Y}) : f]$ in terms of $[Y : \mathbf{1}]$, the supremum norm of Y as well as the various $[\mathbf{X} : h]$. This bound furthermore depends linearly on $[Y : \mathbf{1}]$. The argument can then be closed, because the output of the Sewing Lemma gives a control on $[Y : \mathbf{1}]$ permitting to absorb this term, provided one works on an interval of small enough size. Summarising, this argument can then be used to control the terms $[Y]_\alpha$ and $[Z]_\alpha$ on the right hand side of (1.18) resulting in an estimate that can be iterated.

We finally point out some connections with past and recent work. We first point out that in the setting of stochastic differential equations estimates that show "coming down from infinity" akin to our Theorem 1.4 are known. See, for example, [Cer01, Proposition 1.2.7] for a such a statement, which seems however to impose a slightly more restrictive growth

condition on σ than our (1.5). Turning to the RDEs literature, it was already noticed in [HK15, Proposition 3.8 and Lemma 3.10] that for controlled rough paths which solve RDEs one can control regularity at higher levels via control at the lowest level, but our use of grafting operations allows us to do this more explicitly and in a way that we expect will generalise from branched rough paths to the theory of regularity structures. As this work was nearing completion we also learned of the preprint [BFT20] which develops combinatorial/algebraic methods for comparing remainders quite close to what we use here based on brace algebras [Foi10], [OG08].

1.1. **Outline of paper.** Section 2 starts by reviewing the key elements of the theory of branched rough paths [Gub06, HK15] before closing with a discussion of coherence/elementary differentials - these allow us to describe what the expansions that solve fixed point problems need to look like. Section 3 starts by introducing grafting operations. We then show that this grafting operation behaves well with co-products (Theorem 3.5) and with coherence (Corollary 3.9) - we then finally use these relations to obtain the remainder estimate Corollary 3.12. Finally, Section 4 finishes the proof of our main result, Theorem 1.4, and also proves Theorem 1.5 and Corollary 1.6.

1.2. **Notation.** Given a natural number $m \geq 1$, we write $[m] = \{1, \ldots, m\}$. For any set A, we write $\mathrm{Span}(A)$ for the free vector space over $\mathbb{R}$ generated by A, that is all finite linear combinations of elements of A with real coefficients. Given $p \in \mathbb{N}$ and a p-times differentiable functionable $\sigma : \mathbb{R}^k \to \mathbb{R}^k$, we write $D^p \sigma$ for the p-th derivative which is a p-linear form on $\mathbb{R}^k$ taking values in $\mathbb{R}^k$. We also use standard notation for multi-indices. Given some finite set A, a multi-index $\ell = (\ell_a : a \in A) \in \mathbb{N}^A$, and sufficiently regular function $f : \mathbb{R}^A \to \mathbb{R}^m$, we write

$$\partial^\ell f = \big(\prod_{a \in A} \partial_a^{\ell_a} \big) f \ .$$

where ∂_a is the partial derivative in the a-component. We also write $\ell! = \prod_{a \in A}(\ell_a!)$.

For any $n \geq 1$, $\alpha \in (0, 1)$, interval $I \subseteq [0, 1]$, and function $X : I \to \mathbb{R}^n$ we write

$$\langle X \rangle_I = \sup_{s \in I} \langle X(s) \rangle \ , \qquad \|X\|_I = \sup_{s \in I} |X(s)| \ , \qquad \text{and} \qquad [X]_{\alpha, I} = \sup_{s_1 < s_2 \in I} \frac{|X(t) - X(s)|}{|s_2 - s_1|^\alpha} \ .$$

Throughout the paper we write $\lesssim$ for $\leq C$, where the implicit constant C may depend on d, k, m, α and C_σ. and similarly for any interval $I \subset [0, 1]$ and function $Y : I \to \mathbb{R}^n$ we set $\langle Y \rangle_I = \sup_{s \in I} \langle Y_s \rangle$

2. Branched Rough paths

2.1. **Trees and forests.** We start by inductively defining a set of rooted, decorated trees $\mathring{\mathcal{T}}_{\mathrm{all}}$ along with an associated set of forests $\mathcal{F}_{\mathrm{all}}$.

A forest is a possibly empty, but finite, collection of rooted trees where we allow for multiple instances of the same tree. It is customary to write forests as (commutative) products, that is

$$(2.1) \qquad\qquad\qquad\qquad f = \prod_{i \in I} h_i$$

for some index set I and $h_i \in \mathring{\mathcal{T}}_{\mathrm{all}}$. The empty forest will play a special role for us and we represent it with the symbol $\mathbb{1} \in \mathcal{F}_{\mathrm{all}}$. The set $\mathring{\mathcal{T}}_{\mathrm{all}}$ is generated by taking a forest of trees $f \in \mathcal{F}_{\mathrm{all}}$ and joining them all to a new root with a decoration $\mu \in [d]$, we write the corresponding tree as $[f]_\mu$. Note that we have a natural identification $\mathring{\mathcal{T}}_{\mathrm{all}} \subset \mathcal{F}_{\mathrm{all}}$. We also remark that a forest $f \in \mathcal{F}_{\mathrm{all}}$ is also completely specified by a multi-index $f = (f_h : h \in \mathring{\mathcal{T}}_{\mathrm{all}}) \in \mathbb{N}^{\mathring{\mathcal{T}}_{\mathrm{all}}}$.

We also inductively define a notion of "order" $|\cdot|$ for forests by setting, for $f \in \mathcal{F}_{\mathrm{all}} = \sum_{h \in f} |h|$ and for $h = [f]_\mu \in \mathring{\mathcal{T}}_{\mathrm{all}}$, $|h| = |f| + 1$. Here and in what follows, when we write $\sum_{h \in f}$ or $\prod_{h \in f}$, we are taking the product or sum over all the trees appearing on the right side of (2.1) **with multiplicity**.

The order of a forest is then simply the number of nodes in that forest. The trees with a single node are just of the form $[\mathbb{1}]_j$ for some $j \in [d]$, below we draw such a tree by writing $\bullet_j$. We would then have $[\bullet_j]_i = \overset{\bullet\, j}{\underset{i}{\diagup}}$. We can draw forests by putting trees next to each other, with the order being unimportant, for instance

$$\bullet_k \overset{\bullet\, j}{\underset{i}{\diagup}} = \overset{\bullet\, j}{\underset{i}{\diagup}} \bullet_k \in \mathcal{F}_{\mathrm{all}} \quad \text{and} \quad \left[\bullet_k \overset{\bullet\, j}{\underset{i}{\diagup}} \right]_l = {}_k\overset{\overset{\bullet\, j}{\underset{i}{\bullet}}}{\underset{l}{\bullet}} = \overset{\overset{\bullet\, j}{\underset{i\ k}{\bullet\bullet}}}{\underset{l}{\bullet}} \in \overset{\circ}{\mathcal{T}}_{\mathrm{all}} \ .$$

We write $\# f = \sum_{h \in f} 1$ for the number of trees in f. Given two forests $f, \bar{f}$ we write $f \cdot \bar{f}$ for the forest defined by concatenating the two products of trees. Note that $\mathbb{1}$ plays the role of the unit for the forest product and that both $|\cdot|$ and $\#$ are additive over forest products. By extending the forest product to linear combinations of forests via linearity, $\mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ becomes a unital commutative algebra with unit $\mathbb{1}$. By applying the product component-wise the same is true of $\mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \otimes \mathrm{Span}(\mathcal{F}_{\mathrm{all}})$. We write $\langle \bullet, \bullet \rangle$ for the inner product on $\mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ with $\langle f, f' \rangle = 1\{f = f'\}$ for all $f, f' \in \mathcal{F}_{\mathrm{all}}$. We also denote by $\langle \bullet, \bullet \rangle$ the corresponding inner product on $\mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \otimes \mathrm{Span}(\mathcal{F}_{\mathrm{all}})$. Finally, we write, for any $n \in \mathbb{N}$, the linear projection $P_{\leq n} : \mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \to \mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ that simply annihilate any forest $f \in \mathcal{F}_{\mathrm{all}}$ with $|f| > n$.

2.2. Coproducts.

Definition 2.1. *The coproduct Δ is the map*

$$\Delta : \mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \to \mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \otimes \mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \ ,$$

defined inductively:

- $\Delta\mathbb{1} = \mathbb{1} \otimes \mathbb{1}$.
- *For $[f]_\mu \in \overset{\circ}{\mathcal{T}}_{\mathrm{all}}$ we set $\Delta[f]_\mu = [f]_\mu \otimes \mathbb{1} + (\mathrm{id} \otimes [\bullet]_\mu)\Delta f$.*
- *For $f \in \mathcal{F}_{\mathrm{all}}$, $\Delta f = \prod_{h \in f} \Delta h$.*

Note that this is just the Connes-Kreimer coproduct [CK98]. We note that the co-product preserves the number of nodes in the sense that, for $\bar{f}, \tilde{f}, f \in \mathcal{F}_{\mathrm{all}}$, we have $\langle \Delta \bar{f}, \tilde{f} \otimes f \rangle \neq 0 \Rightarrow |\bar{f}| = |\tilde{f}| + |f|$.

2.3. Iterated integrals.
The set of trees $\overset{\circ}{\mathcal{T}}_{\mathrm{all}}$ will be used to index postulated values of iterated integrals of $X = (X^i)_{i=1}^d$, for instance if X is a smooth path, then all these iterated integrals can be canonically and the natural way to formalise our "canonical" assignment of iterated integrals is to define for $\mu \in [d]$ and $h = [f]_\mu$,

$$\langle \mathbf{X}_{s,t}, h \rangle = \int_s^t \langle \mathbf{X}_{s,r}, f \rangle dX_r^\mu \ .$$

In this case we have

(2.2)
$$\langle \mathbf{X}_{s,t}, \bullet_j \rangle = \int_s^t dX_r^j = X_t^j - X_s^j \ , \quad \langle \mathbf{X}_{s,t}, \overset{\bullet\, j}{\underset{i}{\diagup}} \rangle = \int_s^t \int_s^r dX_u^j dX_r^j \ ,$$

$$\langle \mathbf{X}_{s,t}, \overset{\overset{l\quad n}{\bullet\ \bullet}}{\underset{i}{\bullet}} \rangle = \int_s^t \Big(\int_s^r dX_{u_1}^j \Big) \Big(\int_s^r dX_{u_2}^l \Big) \Big(\int_s^r dX_{u_3}^n \Big) dX_r^i \ .$$

Expansions indexed by such trees/iterated integrals naturally appear when solving differential equations by writing them as integral equations and iterating, for instance, in the

case $k = d = 1$ we can generate a formal expansion for the solution for (1.1) as

$$Y_t - Y_s$$

$$= \int_s^t \sigma(Y_r)dX_r$$

$$= \sigma(Y_s)\int_s^t dX_r + D\sigma(Y_s)\int_s^t (Y_r - Y_s)dX_r + \frac{D^2\sigma(Y_s)}{2}\int_s^t (Y_r - Y_s)^2 dX_r + \cdots$$

$$= \sigma(Y_s)\int_s^t dX_r + D\sigma(Y_s)\int_s^t \int_s^r \sigma(Y_u)dX_u dX_r + \frac{D^2\sigma(Y_s)}{2}\int_s^t \left(\int_s^r \sigma(Y_u)dX_u\right)^2 dX_r + \cdots$$

$$= \sigma(Y_s)\int_s^t dX_r + \left(D\sigma(Y_s)\right)\sigma(Y_s)\int_s^t \int_s^r dX_u dX_r + \frac{D^2\sigma(Y_s)}{2}\left(\sigma(Y_s)\right)^2 \int_s^t \left(\int_s^r dX_u\right)^2 dX_r + \cdots$$

$$= \sigma(Y_s)\langle \mathbf{X}_{s,t}, \bullet \rangle + \left(D\sigma(Y_s)\right)\sigma(Y_s)\langle \mathbf{X}_{s,t}, \text{↗} \rangle + \frac{D^2\sigma(Y_s)}{2}\left(\sigma(Y_s)\right)^2 \langle \mathbf{X}_{s,t}, \text{∨} \rangle + \cdots ,$$

where we suppress the labels here in the one dimensional case - a similar expansion can be generated in the multi-dimensional case. Forests keep track of products of iterated integrals, for instance if f is given as above then $\langle \mathbf{X}_{s,t}, f \rangle = \prod_{h \in f}\langle \mathbf{X}_{s,t}, h \rangle$.

Note that if X is a smooth function, then the Leibniz rule imposes relations on iterated integrals which would allow one to restrict to trees that are given by chains putting one in the setting of *geometric rough paths*. However, the framework of branched rough paths allows us to make sense of this differential equation even when X is too irregular for the right hand sides of (2.2) to be well-defined, the cost is that quantities like those on the left hand sides of (2.2) need to be postulated as part of the well-posedness problem. In practice such data is constructed via stochastic techniques. In particular, the natural choice of values for these iterated integrals may not satisfy Leibniz rule, which is why we allow for trees that branch instead of just chains.

2.4. **Important sets of trees and forests.** We define, for $j \in \{N-2, N-1, N\}$,

$$\mathring{\mathcal{T}}_{\leq j} = \{h \in \mathring{\mathcal{T}}_{\text{all}} : |h| \leq j\} , \ \mathcal{T}_{\leq j} = \mathring{\mathcal{T}}_{\leq j} \cup \{\mathbb{1}\}$$

$$\mathcal{F} = \{f \in \mathcal{F}_{\text{all}} : h \in f \Rightarrow h \in \mathring{\mathcal{T}}_{\leq N}\} \subset \mathcal{F}_{\text{all}} , \ \mathcal{F}_{\leq j} = \{f \in \mathcal{F} : |f| \leq j\} .$$

Note that $\text{Span}(\mathcal{F})$ is a sub-algebra of $\text{Span}(\mathcal{F}_{\text{all}})$. We write $\text{Span}(\mathcal{F})^*$ for all linear functionals from $\text{Span}(\mathcal{F})$ to $\mathbb{R}$. Given $u \in \text{Span}(\mathcal{F})^*$ and $a \in \text{Span}(\mathcal{F})$ we overload notation and write $\langle u, a \rangle$ for the corresponding duality pairing. We define $\text{Char} \subset \text{Span}(\mathcal{F})^*$ to be the set of real characters on $\text{Span}(\mathcal{F})$, that is the set of all algebra homomorphisms from $\text{Span}(\mathcal{F})$ to $\mathbb{R}$.

2.5. **Character paths.**

Definition 2.2. *A character path* $\mathbf{X}$ *is a function* $\mathbf{X}_{\bullet,\bullet} : [0,T]^2 \to \text{Char}$, *written* $(s,t) \mapsto \mathbf{X}_{s,t}$.

Remark 2.3. *We can now motivate some of the definitions of sets of trees and forests we introduced earlier.*

The set $\mathcal{T}_{\leq N-1}$ *is used as index set for local expansions of the solution* Y *to* (1.3)*, while* $\mathcal{F}_{\leq N-1}$ *is used as index set for expansions of* $\sigma_\mu(Y(t))$ *for* $\mu \in [d]$.

The crucial data regarding "iterated integrals" of the X *is indexed by the trees* $\mathring{\mathcal{T}}_{\leq N}$. *In particular, while the trees of order* N *never explicitly appear in expansions, the fact that* $\mathbf{X}$ *is defined on those trees and satisfies an appropriate Hölder bound there is crucial - see the Sewing Lemma - Proposition 2.11 (*[Gub03, Gub06]*).*

Extending $\mathbf{X}$ *multiplicatively means that* $\mathcal{F}$ *is the natural set of objects for* $\mathbf{X}$ *to act on. We note that any character path* $\mathbf{X}$ *satisfies* $\langle \mathbf{X}_{s,t}, \mathbb{1} \rangle = 1$.

Definition 2.4. *A character path satisfies Chen's relation if, for every* $s, u, t \in [0,T]$ *and* $h \in \mathcal{T}$,

$$\langle (\mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t}), \Delta h \rangle = \langle \mathbf{X}_{s,t}, h \rangle .$$

Note that above, we overload the notation $\langle \bullet, \bullet \rangle$ for tensor products by setting $\langle u \otimes v, a \otimes b \rangle = \langle u, a \rangle \langle v, b \rangle$ and then extending by linearity. Given a character path $\mathbf{X}$ and $f \in \mathcal{F}$ we define

$$(2.3) \qquad [\mathbf{X} : f] := \sup_{\substack{0 \leq s, t \leq 1 \\ s \neq t}} \frac{|\langle \mathbf{X}_{s,t}, f \rangle|}{|s - t|^{\alpha|f|}} \ .$$

The exponent α above corresponds to the imposed regularity condition on the driving noise. Note that one has the obvious bound $[\mathbf{X} : f] \leq \prod_{h \in \mathcal{F}} [\mathbf{X} : h]$.

Definition 2.5. *A (branched) α-rough path is a character path $\mathbf{X}$ that satisfies Chen's relation and has the property that $[\mathbf{X} : h] < \infty$ for every $h \in \mathring{\mathcal{T}}$.*

We usually drop the word branched below, just calling $\mathbf{X}$ an α-rough path.

Proposition 2.6. *If $\mathbf{X}$ is the canonical lift of a smooth function X then $\mathbf{X}$ is an α-rough path for any $\alpha \in (0, 1]$.*

2.6. Controlled paths.

Definition 2.7. *We call a map $\mathbf{Y} : [0, T] \to \mathrm{Span}(\mathcal{T}_{\leq N-1})^k$ a tree path and a map $\mathbf{Y} : [0, T] \to \mathrm{Span}(\mathcal{F}_{\leq N-1})^k$ a forest path.*

We apply operations to tree/forest paths component wise, writing expressions like $\Delta \mathbf{Y}_s \in (\mathrm{Span}(\mathcal{F}_{\leq N-1}) \otimes \mathrm{Span}(\mathcal{F}_{\leq N-1}))^k$ and so on.

We usually use the removal of boldface to indicate the "$\mathbb{1}$ - component" of a tree path or forest path $\mathbf{Y}$, for instance we write $Y_s = \langle \mathbb{1}, \mathbf{Y}_s \rangle \in \mathbb{R}^k$ and define $\delta \mathbf{Y}_s$ so that

$$(2.4) \qquad \mathbf{Y}_s = Y_s \mathbb{1} + \delta \mathbf{Y}_s \ .$$

Remark 2.8. *Following the thread from Remark 2.3, a tree path $\mathbf{Y}$ is used as an expansion for the solution Y to (1.3) while, for each $\mu \in [d]$, a forest path $\sigma_\mu(\mathbf{Y})$ will be used as an expansion of $\sigma_\mu(Y)$.*

Definition 2.9. *Fix $I \subset [0, 1]$ and let $\mathbf{X}$ be an α-rough path. Given a forest/tree path $\mathbf{Y}$ we define, for any $f \in \mathcal{F}_{\leq N-1}$,*

$$[\mathbf{Y} : f]_I = \sup_{\substack{s, t \in I \\ 0 < |s-t| < 1}} \frac{|R_{s,t}^{\mathbf{Y}, f}|}{|t - s|^{(N-|f|)\alpha}}$$

where the f-remainder of $\mathbf{Y}$ between s and t, denoted $R_{s,t}^{\mathbf{Y}, f}$, is given by

$$R_{s,t}^{\mathbf{Y}, f} := \langle f, \mathbf{Y}_t \rangle - \langle \mathbf{X}_{s,t} \otimes f, \Delta \mathbf{Y}_s \rangle \in \mathbb{R}^k \ .$$

When $I = [0, 1]$ we suppress it from that notation, i..e writing $[\mathbf{Y} : f]_{[0,1]} = [\mathbf{Y} : f]$. We call $\mathbf{Y}$ an $\mathbf{X}$-controlled rough forest/tree path if $\max\{[\mathbf{Y} : f] : f \in \mathcal{F}_{\leq N-1}\} < \infty$.

As in [HK15] we will define the composition of a function $g \in C^{N-1}(\mathbb{R}^k, \mathbb{R}^k)$ with a tree path $\mathbf{Y}$ via appealing to the Taylor expansion of g about $Y_\bullet$. In particular, we define the forest path $g(\mathbf{Y})$ by setting

$$(2.5) \qquad g(\mathbf{Y})_s = g(Y_s)\mathbb{1} + P_{\leq N-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p g(Y_s) \big[\big(\underbrace{\delta \mathbf{Y}_s, \dots, \delta \mathbf{Y}_s}_{p\text{-times}} \big) \big] \ .$$

Remark 2.10. *Note that in (2.5) we naturally view p-linear forms on $\mathbb{R}^k$ as p-linear forms on $\mathrm{Span}(\mathring{\mathcal{T}}_{\leq N-1})^k$, for instance*

$$D^p g(Y_s) \big[\big(\underbrace{\delta \mathbf{Y}_s, \dots, \delta \mathbf{Y}_s}_{p\text{-times}} \big) \big] = \sum_{h_1, \dots, h_p \in \mathring{\mathcal{T}}_{\leq N-1}} D^p g(Y_s) \big[\big(\langle \mathbf{Y}_s, h_1 \rangle, \dots, \langle \mathbf{Y}_s, h_p \rangle \big) \big] \prod_{i=1}^p h_i \ .$$

2.7. **Integration of controlled paths.** We now explain how we define our rough integral that is needed to define our fixed point problem. Given a forest path $\mathbf{U}$, $\mu \in [d]$, and $s, t \in [0, 1]$, we define

$$\Xi_{s,t}^{\mathbf{U},\mu} = \sum_{f \in \mathcal{F}_{\leq N-1}} \langle f, \mathbf{U}_s \rangle \langle \mathbf{X}_{s,t}, [f]_\mu \rangle .$$

The quantity $\Xi_{s,t}$ should be seen as a single term in a Riemann -type approximation for the rough integral. We can now state Gubinelli's sewing lemma [Gub03, Gub06].

Proposition 2.11. *Let* $\mathbf{X}$ *be a branched rough path and* $\mathbf{U}$ *be an* $\mathbf{X}$*-controlled forest path. Then, for any* $0 \leq s \leq t \leq 1$*, we have*

$$(2.6) \qquad \lim_{|\mathcal{P}| \to 0} \sum_{\{s_n, s_{n+1}\} \in \mathcal{P}} \Xi_{s_n, s_{n+1}}^{\mathbf{U},\mu} =: \int_s^t \mathbf{U}_r \mathrm{d}\mathbf{X}_r^{(\mu)}$$

where the limit above is over partitions $\mathcal{P}$ *of* $[s, t]$ *with mesh size going to* 0*.*

   We have the estimate

$$(2.7) \qquad \left| \int_s^t \mathbf{U}_r \mathrm{d}\mathbf{X}_r^{(\mu)} - \Xi_{s,t}^{\mathbf{U},\mu} \right| \leq C(\alpha)|t - s|^{(N+1)\alpha} \sum_{f \in \mathcal{F}_{\leq N-1}} \sum_{\mu \in [d]} [\mathbf{X} : [f]_\mu][\mathbf{U} : f]_{[s,t]} .$$

Note that $\int_s^t \mathbf{U}_r \mathrm{d}\mathbf{X}_r^{(\mu)}$ is an element of $\mathbb{R}^k$, not a tree/forest path. Our notion of a solution to (1.3) will itself be an expansion so we also want to make sense of the integration of an $\mathbf{X}$-controlled forest path as a tree path rather than just an element $\mathbb{R}^k$. Therefore we define, given a branched rough path $\mathbf{X}$, an $\mathbf{X}$-controlled tree path $\mathbf{Y}$, and a tuple of functions $\sigma = (\sigma_\mu : \mu \in [d])$ with $\sigma_\mu \in C^{N-1}(\mathbb{R}^k, \mathbb{R}^k)$, the new tree path

$$(2.8) \qquad \mathbf{Z}_\bullet = \int_0^\bullet \sigma(\mathbf{Y})_r \mathrm{d}\mathbf{X}_r = \sum_{\mu \in [d]} \left(\int_0^\bullet \sigma_\mu(\mathbf{Y})_r \mathrm{d}\mathbf{X}_r^{(\mu)} \right) \mathbb{1} + \sum_{f \in \mathcal{F}_{\leq N-2}} \langle f, \sigma_\mu(\mathbf{Y})_\bullet \rangle [f]_\mu .$$

Note that we typographically distinguish the integral $\int$ appearing in (2.6) from the integral $\int$ appearing in (2.8) since they produce different sorts of objects.

   We can now state a precisely what we mean by a "solution" to (1.3).

Definition 2.12. *Given a branched rough path* $\mathbf{X}$ *we say that an* $\mathbf{X}$*-controlled tree path* $\mathbf{Y}$ *is a solution to* (1.3) *if we have, for* $0 \leq s \leq 1$*,*

$$(2.9) \qquad \mathbf{Y}_s = y_0 \mathbb{1} - \left(\int_0^s |Y_r|^{m-1} Y_r \mathrm{d}r \right) \mathbb{1} + \mathbf{Z}_s .$$

where $\mathbf{Z}_s$ *is defined as in* (2.8).

2.8. **Coherence.** A necessary (but not sufficient) condition for the tree path $\mathbf{Y}$ to be a solution in the sense of Definition 2.12 is that:

$$(2.10) \qquad \forall [f]_\mu \in \mathring{\mathcal{T}}_{\leq N-1}, \ s \in (0, T], \quad \text{one has } \langle f, \sigma_\mu(\mathbf{Y}_s) \rangle = \langle [f]_\mu, \mathbf{Y}_s \rangle .$$

The condition (2.10) is algebraic and completely local in time. In order to be able to conclude that $\mathbf{Y}$ is a solution in the sense of Definition 2.12 one would additionally have to impose that $\mathbf{Y}$ is $\mathbf{X}$-controlled and that, for all $s \in (0, T]$, the $\mathbb{1}$ coefficient on both sides of (2.9) are the same.

   Given Y_s as input, one can use (2.10) to calculate either side of (2.10) inductively in $|f|$. This motivates the following definition.

Definition 2.13. *We define the collection of functions*

$$(\Upsilon_\mu[f](\cdot) : f \in \mathcal{F}_{\leq N-1} , \ \mu \in [d]) , \ \Upsilon_\mu[f](\cdot) : \mathbb{R}^k \to \mathbb{R}^k ,$$

inductively, as follows. Given

$$f = \prod_{i \in I} [f_i]_{\mu_i}$$

we set

$$\Upsilon_\mu[f](\cdot) = D^{|I|}\sigma_\mu(\cdot)\big[\big(\Upsilon_{\mu_i}[f_i](\cdot)\big)_{i\in I}\big] \ . \tag{2.11}$$

The definition given above is inductive in $|f|$ and also fixes the base case $f = \mathbb{1}$ where we have $\Upsilon_\mu[\mathbb{1}](\cdot) = \sigma_\mu(\cdot)$.

Remark 2.14. *The functions $\Upsilon_\mu[f]$ are also known as* elementary differentials *in the analysis of Runge-Kutta methods, see* [Bro00].

These elementary differentials usually appear with symmetry factors. We define $S : \mathcal{F}_{\mathrm{all}} \to \mathbb{N}$ via recursively setting, $S(\mathbb{1}) = 1$, and

$$S(f) = \prod_{h\in\mathring{\mathcal{T}}_{\mathrm{all}}} (f_h!)S(h)^{f_h} \ ,$$

where we recall that $f_h \in \mathbb{N}$ is the number of instances of h in f.

A straightforward inductive proof then gives the following lemma.

Lemma 2.15. *Given a tree path $\mathbf{Y}$, the condition* (2.10) *is equivalent to*

$$\forall [f]_\mu \in \mathring{\mathcal{T}}_{\leq N-1}, \ s \in (0,T], \quad \text{one has } \langle f, \sigma_\mu(\mathbf{Y}_s)\rangle = \langle [f]_\mu, \mathbf{Y}_s\rangle = \frac{1}{S(f)}\Upsilon_\mu[f](Y_s) \ . \tag{2.12}$$

Definition 2.16. *We call a tree path $\mathbf{Y}$ that satisfies* (2.10) *(or equivalently* (2.12)*)* coherent.

Remark 2.17. *We extend $\Upsilon_\mu[\bullet]$ from $\mathcal{F}_{\mathrm{all}}$ to $\mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ by linearity. We also define $\Upsilon[\bullet]$ on $\mathrm{Span}(\mathring{\mathcal{T}}_{\mathrm{all}})$ by setting, for $h = [f]_\mu \in \mathring{\mathcal{T}}_{\mathrm{all}}$, $\Upsilon[h] = \Upsilon_\mu[f]$ and then again extend linearly. We can then write* (2.11) *as*

$$\Upsilon\big[[f]_\mu\big](\cdot) = \Upsilon_\mu[f](\cdot) = D^{\#f}\sigma_\mu(\cdot)\big[\big(\Upsilon[h](\cdot)\big)_{h\in f}\big] \ . \tag{2.13}$$

This duplication of notation (giving equivalent definitions of Υ_μ on forests and Υ on trees) is convenient because trees appear for the expansion of Y while forests appear for the expansion of the right hand side of the equation.

As an example, we compute

$$\Upsilon_l\Big[\ \bullet_k \ \ \begin{smallmatrix}\bullet^j\\ \diagup \\ \bullet_i\end{smallmatrix}\ \Big](x) = \Upsilon\Big[{}^k\bullet\!\!\diagdown\!\!\begin{smallmatrix}\bullet^j\\ \bullet_i\end{smallmatrix}\!\!\Big]_l\Big](x) = D^2\sigma_l(x)\big[\big(\sigma_k(x), D\sigma_i(x)[\sigma_k(x)]\big)\big] \ .$$

We close this section with the following remark.

Remark 2.18. *Note that the drift term $-|Y(t)|^{m-1}Y(t)dt$ didn't play much of a role in this section, in particular it has no direct effect on the coherence coefficients $\Upsilon_\mu[\cdot](\cdot)$. However, since$\big|\int_s^t |Y(r)|^{m-1}Y(r)dr\big| \lesssim |s-t|$, this term is negligible compared to the other terms in our local expansion for Y. In particular, by virtue of* (2.9)*, we have*

$$R_{s,t}^{\mathbf{Z},\mathbb{1}} = Z_t - Z_s - \langle \mathbf{X}_{s,t}, \delta\mathbf{Z}_s\rangle = R_{s,t}^{\mathbf{Y},\mathbb{1}} + \int_s^t |Y(r)|^{m-1}Y(r)dr \ , \tag{2.14}$$

where $\delta\mathbf{Z}$ is defined via the same convention as in (2.4) *and we used that, by coherence,*

$$\langle \mathbf{X}_{s,t}, \delta\mathbf{Z}_s\rangle = \sum_{h\in\mathring{\mathcal{T}}_{\leq N-1}} \langle \mathbf{X}_{s,t}, h\rangle\langle h, \mathbf{Z}_s\rangle = \sum_{h\in\mathring{\mathcal{T}}_{\leq N-1}} \langle \mathbf{X}_{s,t}, h\rangle\langle h, \mathbf{Y}_s\rangle \ .$$

The role of the drift becomes important only in the final step of our argument in Section 4.

3. Grafting, coherence, and remainder estimates

3.1. **Grafting operations.** We introduce grafting operations which will later be shown to be "adjoints" of the co-product under a particular inner product we introduce later.

In words, given $\tilde{f}, f \in \mathcal{F}_{\mathrm{all}}$, the grafting of $\tilde{f}$ onto f sums over all the ways of attaching each of the trees in $\tilde{h} \in \tilde{f}$ to a vertex of f with an edge, or concatenating $\tilde{h}$ to the forest f. More precisely, given

$$(3.1) \qquad \tilde{f} = \prod_{j \in J} \tilde{h}_j \in \mathcal{F}_{\mathrm{all}} \text{ and } f = \prod_{i \in I} [f_i]_{\mu_i} \ .$$

we inductively define

$$(3.2) \qquad \tilde{f} \curvearrowright f = \sum_{K \subset J} \Big(\prod_{j \in J \setminus K} \tilde{h}_j \Big) \sum_{\theta : K \to I} \Big(\prod_{i \in I} \Big[\Big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k \Big) \curvearrowright f_i \Big]_{\mu_i} \Big) \in \mathcal{F}_{\mathrm{all}} \ .$$

The definition above is inductive in $|f|$, with the base case corresponding to $f = \mathbb{1}$ for which we have $I = \emptyset$ in (3.1) and so one must choose $K = \emptyset$ in the first sum of (3.2) – in particular we have $\tilde{f} \curvearrowright \mathbb{1} = \tilde{f}$. Additionally, we note that $\langle \tilde{f} \curvearrowright f, \bar{f} \rangle \neq 0 \Rightarrow |\tilde{f}| + |f| = |\bar{f}|$. We give an example grafting computation below:

Remark 3.1. *The operation $\tilde{f} \curvearrowright f$ is a decorated version of the "grafting forests over forests" operation of* [Foi13, Section 1.5]*, this is essentially a version of the Grossman-Larson product* [GL89]*.*

3.2. **Inner products and the adjoint relation.**

Definition 3.2. *We define a new inner product on* $\mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ *by inductively setting*
- $\langle\langle \mathbb{1}, \mathbb{1} \rangle\rangle = 1$
- $\langle\langle [f]_\mu, [g]_{\mu'} \rangle\rangle = 1\{\mu = \mu'\}\langle\langle f, g \rangle\rangle$
- *For $h_i, g_j \in \mathring{\mathcal{T}}_{\mathrm{all}}$,*

$$\langle\langle \prod_{i \in I} h_i, \prod_{j \in J} g_j \rangle\rangle = \sum_{\substack{\theta : I \to J \\ \text{biject.}}} \prod_{i \in I} \langle\langle h_i, g_{\theta(i)} \rangle\rangle \ ,$$

We extend this inner product to $\mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \otimes \mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ *by setting, for $f, g, \bar{f}, \bar{g} \in \mathcal{F}_{\mathrm{all}}$,*

$$\langle\langle f \otimes g, \bar{f} \otimes \bar{g} \rangle\rangle = \langle\langle f, \bar{f} \rangle\rangle \langle\langle g, \bar{g} \rangle\rangle \ .$$

Remark 3.3. *Note that, for $f, g \in \mathcal{F}_{\mathrm{all}}$,*

$$(3.3) \qquad \langle\langle f, g \rangle\rangle = 1\{f = g\} S(f) = S(f)\langle f, g \rangle \ .$$

In particular, (2.12) *can be rewritten as :*

$$\forall [f]_\mu \in \mathring{\mathcal{T}}_{\leq N-1}, \ s \in (0, T], \quad \text{one has } \langle\langle f, \sigma_\mu(\mathbf{Y}_s) \rangle\rangle = \langle\langle [f]_\mu, \mathbf{Y}_s \rangle\rangle = \Upsilon_\mu[f](Y_s) \ .$$

The following lemma is straightforward to prove from our definition of $\langle\langle \bullet, \bullet \rangle\rangle$.

Lemma 3.4. *Given any $\tilde{h}_j, h_i \in \mathring{\mathcal{T}}$ and also $a_n \in \mathrm{Span}(\mathcal{F}_{\mathrm{all}}) \otimes \mathrm{Span}(\mathcal{F}_{\mathrm{all}})$ one has*

$$\Big\langle \Big\langle (\prod_{j \in J} \tilde{h}_j) \otimes (\prod_{i \in I} h_i), \prod_{n \in N} a_n \Big\rangle \Big\rangle = \sum_{\substack{\theta_J : J \to N \\ \theta_I : I \to N}} \prod_{n \in N} \Big\langle \Big\langle (\prod_{j \in \theta_J^{-1}(n)} \tilde{h}_j) \otimes (\prod_{i \in \theta_I^{-1}(n)} h_i), a_n \Big\rangle \Big\rangle$$

Our main purpose for introducing the inner products $\langle\langle \bullet, \bullet \rangle\rangle$ is that it is convenient for stating and proving following theorem.

Theorem 3.5. *For any $\tilde{f}, f, g \in \mathcal{F}_{\mathrm{all}}$, we have*

$$(3.4) \qquad \langle\langle \tilde{f} \frown f, g \rangle\rangle = \langle\langle \tilde{f} \otimes f, \Delta g \rangle\rangle .$$

Proof. We give an inductive proof with the induction being in $|f|$. For the base case we have $f = \mathbb{1}$ in which case (3.4) is immediate. We now prove the inductive step. We write $g = \prod_{n \in N}[g_n]_{\bar{\mu}_n}$ and also write $\tilde{f}$ and f as in (3.1). We then have

$$\langle\langle \tilde{f} \frown f, g \rangle\rangle = \sum_{K \subset J} \sum_{\theta: K \to I} \left\langle\!\left\langle \Big(\prod_{j \in J \setminus K} \tilde{h}_j \Big) \prod_{i \in I} \Big[\Big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k \Big) \frown f_i \Big]_{\mu_i}, \prod_{n \in N}[g_n]_{\bar{\mu}_n} \right\rangle\!\right\rangle$$

$$= \sum_{K \subset J} \sum_{\substack{\sigma: (J \setminus K) \sqcup I \to N \\ \text{biject.}}} \sum_{\theta: K \to I} \Big(\prod_{j \in J \setminus K} \left\langle\!\left\langle \tilde{h}_j, [g_{\sigma(j)}]_{\bar{\mu}_{\sigma(j)}} \right\rangle\!\right\rangle \Big)$$

$$\times \Big(\prod_{i \in I} \left\langle\!\left\langle \Big[\Big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k \Big) \frown f_i \Big]_{\mu_i}, [g_{\sigma(i)}]_{\bar{\mu}_{\sigma(i)}} \right\rangle\!\right\rangle \Big)$$

$$= \sum_{K \subset J} \sum_{\substack{\sigma: (J \setminus K) \sqcup I \to N \\ \text{biject.}}} \sum_{\theta: K \to I} \Big(\prod_{j \in J \setminus K} \left\langle\!\left\langle \tilde{h}_j, [g_{\sigma(j)}]_{\bar{\mu}_{\sigma(j)}} \right\rangle\!\right\rangle \Big)$$

$$\times \Big(\prod_{i \in I} \mathbb{1}\{\mu_i = \bar{\mu}_{\sigma(j)}\} \left\langle\!\left\langle \Big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k \Big) \otimes f_i, \Delta g_{\sigma(i)} \right\rangle\!\right\rangle \Big) .$$

Above, in the second equality we used the third identity of Definition 3.2, and in the third equality we used our the third identity of Definition 3.2 followed by our induction hypothesis.

One the other hand, $\Delta g = \sum_{M \subset N} \Big(\prod_{m \in M}[g_m]_{\bar{\mu}_m} \otimes \mathbb{1} \Big) \Big(\prod_{n \in N \setminus M}(\mathrm{id} \otimes [\bullet]_{\bar{\mu}_n}) \Delta g_n \Big)$ which gives

$$\langle\langle \tilde{f} \otimes f, \Delta g \rangle\rangle$$

$$= \sum_{M \subset N} \left\langle\!\left\langle \tilde{f} \otimes f, \Big(\prod_{m \in M}[g_m]_{\bar{\mu}_m} \otimes \mathbb{1} \Big) \Big(\prod_{n \in N \setminus M}(\mathrm{id} \otimes [\cdot]_{\bar{\mu}_n}) \Delta g_n \Big) \right\rangle\!\right\rangle$$

$$= \sum_{M \subset N} \sum_{\theta_J: J \to N} \sum_{\theta_I: I \to N} \Big(\prod_{m \in M} \left\langle\!\left\langle \Big(\prod_{j \in \theta_J^{-1}(m)} \tilde{h}_j \Big) \otimes \Big(\prod_{i \in \theta_I^{-1}(m)} [f_i]_{\mu_i} \Big), [g_m]_{\bar{\mu}_m} \otimes \mathbb{1} \right\rangle\!\right\rangle \Big)$$

$$\times \Big(\prod_{n \in N \setminus M} \left\langle\!\left\langle \Big(\prod_{j \in \theta_J^{-1}(n)} \tilde{h}_j \Big) \otimes \Big(\prod_{i \in \theta_I^{-1}(n)} [f_i]_{\mu_i} \Big), (\mathrm{id} \otimes [\cdot]_{\bar{\mu}_n}) \Delta g_n \right\rangle\!\right\rangle \Big)$$

$$= \sum_{M \subset N} \sum_{\theta_J: J \to N} \sum_{\substack{\theta_I: I \to N \setminus M \\ \text{biject.}}} \Big(\prod_{m \in M} \left\langle\!\left\langle \Big(\prod_{j \in \theta_J^{-1}(m)} \tilde{h}_j \Big), [g_m]_{\bar{\mu}_m} \right\rangle\!\right\rangle \Big)$$

$$\times \Big(\prod_{i \in I} \mathbb{1}\{\mu_i = \bar{\mu}_{\theta_I(i)}\} \left\langle\!\left\langle \Big(\prod_{j \in \theta_J^{-1}(\theta_I(i))} \tilde{h}_j \Big) \otimes f_i, \Delta g_{\theta_I(i)} \right\rangle\!\right\rangle \Big)$$

$$= \sum_{M \subset N} \sum_{K \subset J} \sum_{\substack{\pi: J \setminus K \to M \\ \text{biject.}}} \sum_{\substack{\theta_I: I \to N \setminus M \\ \text{biject.}}} \sum_{\theta: K \to I} \Big(\prod_{j \in J \setminus K} \left\langle\!\left\langle \tilde{h}_j, [g_{\pi(m)}]_{\bar{\mu}_{\pi(m)}} \right\rangle\!\right\rangle \Big)$$

$$\times \Big(\prod_{i \in I} \mathbb{1}\{\mu_i = \bar{\mu}_{\theta_I(i)}\} \left\langle\!\left\langle \Big(\prod_{j \in \theta^{-1}(i)} \tilde{h}_j \Big) \otimes f_i, \Delta g_{\theta_I(i)} \right\rangle\!\right\rangle \Big)$$

$$= \sum_{K \subset J} \sum_{\substack{\sigma: (J \setminus K) \sqcup I \to N \\ \text{biject.}}} \sum_{\theta: K \to I} \Big(\prod_{j \in J \setminus K} \langle\langle \tilde{h}_j, [g_{\sigma(j)}]_{\bar{\mu}_{\sigma(j)}} \rangle\rangle \Big)$$

$$\times \Big(\prod_{i \in I} \mathbb{1}\{\mu_i = \bar{\mu}_{\sigma(j)}\} \left\langle\!\left\langle \Big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k \Big) \otimes f_i, \Delta g_{\sigma(i)} \right\rangle\!\right\rangle \Big) .$$

In the second equality above we used Lemma 3.4. In the third equality we above we used the fact that the line before vanishes unless θ_I is a bijection from I to $N \setminus M$.

In the fourth equality we used that the line before vanishes unless $|\theta_J^{-1}(m)| = 1$ for every $m \in M$. Therefore the sum over the map θ_J in the line before can be written as a sum over $K = \theta_J^{-1}(N \setminus M)$, $\pi = \theta_J \upharpoonright_{J \setminus K}$ and $\theta = (\theta_J \upharpoonright_K) \circ \theta_I^{-1}$. In the last equality we exchange the sums over π and θ_I, and M for just a single sum over σ. $\qquad\square$

Remark 3.6. *The result of Theorem 3.5 was actually first obtained in* [Hof03, *Proposition 4.4*], *and reflects the duality between the Grossman-Larson and Connes-Kreimer Hopf algebras.*

3.3. **Coherence and grafting.** The next lemma states a generalisation of (2.13) and shows that maps Υ_μ interact well with our grafting procedure.

Lemma 3.7. *Given any* $\tilde{f}, f \in \mathcal{F}$ *with* $|\tilde{f}| + |f| \le N - 1$, *and* $\mu \in [d]$ *one has*
$$\Upsilon_\mu[\tilde{f} \curvearrowright f] = D^{\#\tilde{f}} \Upsilon_\mu[f]\big[\big(\Upsilon[\tilde{h}]\big)_{\tilde{h} \in \tilde{f}}\big] .$$

Proof. We prove the general case by induction in $|f|$ - the base case when $f = \mathbb{1}$ is immediate.

We write f and $\tilde{f}$ as in (3.1). Now we note that, by Leibnitz rule and the definition of Υ, for any $v = (v_j)_{j \in J} \in (\mathbb{R}^k)^J$,

$$
\begin{aligned}
(3.5) \qquad & D^{\#\tilde{f}} \Upsilon_\mu[f][v] \\
& = \sum_{(J_i : i \in I)} D^{\#f + \#\tilde{f} - \sum_{i \in I} |J_i|} \sigma_\mu\Big[\big(D^{|J_i|} \Upsilon_{\mu_i}[f_i][v_{J_i}]\big)_{i \in I} \sqcup v_{J^c}\Big] ,
\end{aligned}
$$

where above the sum is over collections of disjoint subsets $(J_i : i \in I)$, $J_i \subset J$ and we write $v_{J_i} = (v_j)_{j \in J_i}$ and $J^c = J \setminus \big(\bigcup_{i \in I} J_i\big)$.

On the other hand, using the definition of the grafting procedure followed by that of Υ_μ

$$
\begin{aligned}
& \Upsilon_\mu[\tilde{f} \curvearrowright f] \\
& = \sum_{K \subset J} \sum_{\theta : K \to I} \Upsilon_\mu\Big[\Big(\prod_{j \in J \setminus K} \tilde{h}_j\Big)\Big(\prod_{i \in I}\big[\big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k\big) \curvearrowright f_i\big]_{\mu_i}\big)\Big] \\
& = \sum_{K \subset J} \sum_{\theta : K \to I} D^{|J \setminus K| + \#f} \sigma_\mu\Big[\big(\Upsilon[\tilde{h}_j]\big)_{j \in J \setminus K} \sqcup \Big(\Upsilon_{\mu_i}\big[\big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k\big) \curvearrowright f_i\big]\Big)_{i \in I}\Big] .
\end{aligned}
$$

Now, by our induction hypothesis, for any $\theta : K \to I$ and $i \in I$ we have
$$\Upsilon_{\mu_i}\Big[\big(\prod_{k \in \theta^{-1}(i)} \tilde{h}_k\big) \curvearrowright f_i\Big] = D^{|\theta^{-1}(i)|} \Upsilon_{\mu_i}[f_i]\Big[\big(\Upsilon[\tilde{h}_k]\big)_{k \in \theta^{-1}(i)}\Big] .$$

Inserting the above calculation into the previous equation gives
$$
\begin{aligned}
& \Upsilon_\mu[\tilde{f} \curvearrowright f] \\
& = \sum_{K \subset J} \sum_{\theta : K \to I} D^{|J \setminus K| + \#f} \sigma_\mu\Big[\big(D^{|\theta^{-1}(i)|} \Upsilon_{\mu_i}[f_i][(\Upsilon[\tilde{h}_j])_{j \in \theta^{-1}(i)}]\big)_{i \in I} \sqcup (\Upsilon[\tilde{h}_j])_{j \in J \setminus K}\Big] ,
\end{aligned}
$$

which finishes the proof by comparing with (3.5), identifying $J_i \leftrightarrow \theta^{-1}(i)$, and setting $v = (\Upsilon[\tilde{h}_j])_{j \in J}$. $\qquad\square$

In what follows we define, for $g \in \mathcal{F}_{\le N-1}$ and $\mu \in [d]$, $\boldsymbol{\Upsilon}_\mu[g](\cdot) = \frac{1}{S(g)} \Upsilon_\mu[g](\cdot)$ and, for $h \in \mathring{\mathcal{T}}_{\le N-1}$, $\boldsymbol{\Upsilon}[h](\cdot) = \frac{1}{S(h)} \Upsilon[h](\cdot)$. We also write
$$\boldsymbol{\Upsilon}_\mu(\cdot) = \sum_{g \in \mathcal{F}_{\le N-1}} \boldsymbol{\Upsilon}_\mu[g](\cdot) g .$$

In particular, $\boldsymbol{\Upsilon}_\mu : \mathbb{R}^k \to \mathrm{Span}(\mathcal{F}_{\le N-1})^k$ and for any $g \in \mathcal{F}_{\le N-1}$ we have $\langle\langle \boldsymbol{\Upsilon}_\mu, g \rangle\rangle = \Upsilon_\mu[g]$. We also define
$$\boldsymbol{\Upsilon}(\cdot) = \sum_{h \in \mathring{\mathcal{T}}_{\le N-1}} \boldsymbol{\Upsilon}[h](\cdot) h .$$

We see that if a tree path $\mathbf{Y}$ is coherent then, for any $s \in (0, 1]$, we have $\delta \mathbf{Y}_s = \mathbf{\Upsilon}(Y_s)$ and, for every $\mu \in [d]$, $\sigma_\mu(\mathbf{Y}_s) = \mathbf{\Upsilon}_\mu(Y_s)$.

Lemma 3.8. *Given any $f \in \mathcal{F}_{\leq N-1}$ and $\mu \in [d]$,*

$$\sum_{\tilde{f} \in \mathcal{F}_{\leq N-1}} \langle \tilde{f} \otimes f, \Delta \mathbf{\Upsilon}_\mu \rangle \tilde{f} = P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f] \Big[\underbrace{(\mathbf{\Upsilon}, \ldots, \mathbf{\Upsilon})}_{p\text{-times}} \Big] .$$

Proof. We observe that

$$\begin{aligned}
\langle \tilde{f} \otimes f, \Delta \mathbf{\Upsilon}_\mu \rangle \tilde{f} &= \frac{1}{S(\tilde{f}) S(f)} \langle\langle \tilde{f} \otimes f, \Delta \mathbf{\Upsilon}_\mu \rangle\rangle \tilde{f} \\
&= \frac{1}{S(\tilde{f}) S(f)} \langle\langle \tilde{f} \frown f, \mathbf{\Upsilon}_\mu \rangle\rangle \tilde{f} = \frac{1\{|f| + |\tilde{f}| < N\}}{S(\tilde{f}) S(f)} \mathbf{\Upsilon}_\mu[\tilde{f} \frown f] \tilde{f} \\
&= \frac{1\{|f| + |\tilde{f}| < N\}}{S(\tilde{f}) S(f)} D^{\#\tilde{f}} \mathbf{\Upsilon}_\mu[f] \Big[\big(\mathbf{\Upsilon}[\tilde{h}] \big)_{\tilde{h} \in \tilde{f}} \Big] \tilde{f} \\
&= 1\{|f| + |\tilde{f}| < N\} \frac{\mathrm{Multi}(f)}{\#\tilde{f}!} D^{\#\tilde{f}} \mathbf{\Upsilon}_\mu[f] \Big[\big(\mathbf{\Upsilon}[\tilde{h}] \tilde{h} \big)_{\tilde{h} \in \tilde{f}} \Big] .
\end{aligned}$$

In the first equality above we are using (3.3) and in the second equality we are applying Theorem 3.5. In the third equality above we are using Lemma 3.7, and for the third equality we define the multinomial coefficient

$$\mathrm{Multi}(\tilde{f}) = \frac{\#\tilde{f}!}{\prod_{\tilde{h} \in \mathring{\mathcal{T}}_{\leq N-1}} \tilde{f}_{\tilde{h}}!}$$

where $\tilde{f}_{\tilde{h}}$ in again the number of instances of $\tilde{h}$ in $\tilde{f}$, so that we have

$$S(\tilde{f}) = \frac{\#\tilde{f}!}{\mathrm{Multi}(\tilde{f})} \prod_{\tilde{h} \in \tilde{f}} S(\tilde{h}) .$$

It follows that

$$\begin{aligned}
\sum_{\tilde{f} \in \mathcal{F}_{\leq N-1}} \langle \tilde{f} \otimes f, \Delta \mathbf{\Upsilon}_\mu \rangle \tilde{f} &= \sum_{\substack{\tilde{f} \in \mathcal{F}_{\leq N-1} \\ |\tilde{f}| < N-|f|}} \frac{\mathrm{Multi}(\tilde{f})}{\#\tilde{f}!} D^{\#\tilde{f}} \mathbf{\Upsilon}_\mu[f] \Big[\big(\mathbf{\Upsilon}[\tilde{h}] \tilde{h} \big)_{\tilde{h} \in \tilde{f}} \Big] \\
&= \sum_{p=0}^{N-1} \frac{1}{p!} \sum_{\substack{\tilde{f} \in \mathcal{F}_{\leq N-1} \\ |\tilde{f}| < N-|f|}} 1\{\#f = p\} \mathrm{Multi}(\tilde{f}) D^p \mathbf{\Upsilon}_\mu[f] \Big[\big(\mathbf{\Upsilon}[\tilde{h}] \big)_{\tilde{h} \in \tilde{f}} \Big] \\
&= \sum_{p=0}^{N-1} \frac{1}{p!} \sum_{h_1, \ldots, h_p \in \mathring{\mathcal{T}}_{\leq N-1}} D^p \mathbf{\Upsilon}_\mu[f] \Big[\big(\mathbf{\Upsilon}[h_i] h_i \big)_{i=1}^p \Big] \\
&= P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f] \Big[\underbrace{(\mathbf{\Upsilon}, \ldots, \mathbf{\Upsilon})}_{p\text{-times}} \Big] .
\end{aligned}$$

(3.6)

$\square$

The above lemma gives us the following corollary, the main result of this section. It expresses remainders at higher levels in terms of remainders at level 0.

Corollary 3.9. *For any coherent tree path $\mathbf{Y}$, $f \in \mathcal{F}_{\leq N-1}$, and $\mu \in [d]$, we have*

(3.7)

$$R_{s,t}^{\sigma_\mu(\mathbf{Y}), f} = \mathbf{\Upsilon}_\mu[f](Y_t) - \Big\langle \mathbf{X}_{s,t}, P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f](Y_s) \Big[\underbrace{(\mathbf{\Upsilon}(Y_s), \ldots, \mathbf{\Upsilon}(Y_s))}_{p\text{-times}} \Big] \Big\rangle .$$

Proof. Using the definition of coherence we can write

$$
\begin{aligned}
R_{s,t}^{\sigma_\mu(\mathbf{Y}),f} &= \langle f, \sigma_\mu(\mathbf{Y}_t) \rangle - \langle \mathbf{X}_{s,t} \otimes f, \Delta\sigma_\mu(\mathbf{Y}_t) \rangle \\
&= \mathbf{\Upsilon}_\mu[f](Y_t) - \langle \mathbf{X}_{s,t} \otimes f, \Delta\mathbf{\Upsilon}_\mu(Y_s) \rangle.
\end{aligned}
$$

We then use Lemma 3.8 to see that

$$
\begin{aligned}
\langle \mathbf{X}_{s,t} \otimes f, \Delta\mathbf{\Upsilon}_\mu(Y_s) \rangle &= \Big\langle \mathbf{X}_{s,t}, \sum_{\tilde f \in \mathcal{F}_{\leq N-1}} \langle \tilde f \otimes f, \Delta\mathbf{\Upsilon}_\mu(Y_s) \rangle \tilde f \Big\rangle \\
&= \Big\langle \mathbf{X}_{s,t}, P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f](Y_s) \Big[\underbrace{(\mathbf{\Upsilon}(Y_s), \ldots, \mathbf{\Upsilon}(Y_s))}_{p\text{-times}} \Big] \Big\rangle.
\end{aligned}
$$

$\square$

Remark 3.10. *Note that if we know that the functions $(\sigma_\mu)_{\mu=1}^d$ are actually smooth then the $\mathbf{\Upsilon}_\mu$ make sense on $\mathcal{F}_{\mathrm{all}}$ and it easy to check that the various lemmas and corollaries of this subsection hold without any constraint on the orders of the participating forests.*

3.4. **The main remainder estimate.** In order to estimate (3.7) we will use the Taylor expansion/remainder formula of [Hai14, Proposition A.1].

We will want to impose a stopping rule on generating our Taylor expansion that takes into the consideration the order forests that are produced, a Taylor expansion like [Hai14, Proposition A.1] is convenient because it lets us enforce a rule on how the expansion is iteratively generated.

We say $A \subset \mathcal{F}$ is *full* if it is non-empty, finite, and, for any $f \in A$, we have

$$
\{ \tilde f \in \mathcal{F} : \tilde f_h \leq f_h \text{ for every } h \in \mathcal{F} \} \subset A.
$$

Note that, for any $M \in \mathbb{N}$, $\mathcal{F}_{\leq M}$ is full.

We now fix some arbitrary total order on $\mathring{\mathcal{T}}_{\leq N-1}$. For any $A \subset \mathcal{F}$ we define a "boundary" of A by setting

$$
(3.8) \qquad \partial A = \Big\{ f \in \mathcal{F} \setminus (A \cup \mathbb{1}) : (f_h - \mathbb{1}\{h = \mathfrak{m}(f)\} : h \in \mathring{\mathcal{T}}_{\leq N-1}) \in A \Big\},
$$

where

$$
\mathfrak{m}(f) = \min\{ h \in \mathring{\mathcal{T}}_{\leq N-1} : f_h \neq 0 \}
$$

is the minimal tree appearing in the non-empty forest f. In words, $f \in \partial A$ if f is non-empty, does not belong to A, but if one reduces the number of instances of the minimal tree appearing in f by 1 then the resulting forest does belong to A.

We can then state the formulation of [Hai14, Proposition A.1] we will use.

Proposition 3.11. *Let $A \subset \mathcal{F}$ be full and $F \in C^{K+1}([0,1]^{\mathring{\mathcal{T}}_{\leq N-1}})$ where $K = \max_{f \in A} \# f$, then*

$$
F(\underline{1}) = \sum_{f \in A} \frac{\partial^f F(0)}{f!} + \sum_{f \in \partial A} \int_{[0,1]^{\mathring{\mathcal{T}}_{\leq N-1}}} \partial^f F(y) Q^f(dy)
$$

where $\underline{1} = (1 : h \in \mathcal{F}) \in [0,1]^{\mathring{\mathcal{T}}_{\leq N-1}}$ is vector of all ones and we are using standard notation for derivatives and monomials indexed by multi-indices $f = (f_h : h \in \mathring{\mathcal{T}}_{\leq N-1}) \in \mathcal{F}$, and for each $f \in \mathcal{F}$, $Q^f(dy)$ is measure of total mass 1 on $[0,1]^{\mathring{\mathcal{T}}_{\leq N-1}}$.

We can now state the key lemma for this section.

Corollary 3.12. *Suppose the tree path $\mathbf{Y}$ is coherent, then, for any $\mu \in [d]$ and $f \in \mathcal{F}_{\leq N-1}$,*

$$
|R_{s,t}^{\sigma_\mu(\mathbf{Y}),f}| \lesssim \sup \Big\{ \big| D\Upsilon_\mu[f](Y_s + a) \big| \cdot |R_{s,t}^{\mathbf{Y},\mathbb{1}}| : a \in \mathbb{R}^k, |a| \leq E_{s,t}^{\mathbf{Y}} + |Y_t - Y_s| \Big\}
$$

$$
(3.9)
$$

$$
+ \sup \left\{ \Big| D^{\#\tilde f} \Upsilon_\mu[f](Y_s + z) \prod_{h \in \tilde f} \Upsilon[h](Y_s) \langle \mathbf{X}_{s,t}, h \rangle \Big| : \begin{array}{l} \tilde f \in \partial\mathcal{F}_{\leq N-|f|-1} \\ z \in \mathbb{R}^k, |z| \leq E_{s,t}^{\mathbf{Y}} \end{array} \right\},
$$

where

$$(3.10) \qquad E_{s,t}^{\mathbf{Y}} = \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} |\Upsilon[h](Y_s)\langle \mathbf{X}_{s,t}, h \rangle| \ .$$

Before turning to the proof, we point out that (3.9) estimates $R_{s,t}^{\sigma_\mu(\mathbf{Y}),f}$ by treating it as a multivariable Taylor remainder - this is why we see suprema of derivatives on the right hand side above. Being careful about the constraints on the arguments of these derivatives will be important, in particular when working with coefficients σ that are allowed to grow at infinity, see Assumption 1.2.

Proof. Given any $x = (x_h) \in [0,1]^{\mathring{\mathcal{T}}_{\leq N-1}}$ we define

$$\mathbf{Y}_{s,t}(x) = \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} x_h \Upsilon[h](Y_s)\langle \mathbf{X}_{s,t}, h \rangle \ .$$

We then have

$$Y_s + \mathbf{Y}_{s,t}(\underline{1}) + R_{s,t}^{\mathbf{Y},\mathbb{1}} = Y_t$$

We now define the function $F : [0,1]^{\mathring{\mathcal{T}}_{\leq N-1}} \to \mathbb{R}^k$ given by

$$F(x) = \mathbf{\Upsilon}_\mu[f]\Big(Y_s + \mathbf{Y}_{s,t}(x)\Big).$$

Now, by Corollary 3.9 and adding and subtracting $F(\underline{1})$, we have

$$(3.11) \qquad \begin{aligned} R_{s,t}^{\sigma_\mu(\mathbf{Y}),f} \leq & \left| \mathbf{\Upsilon}_\mu[f](Y_s + \mathbf{Y}_{s,t}(\underline{1}) + R_{s,t}^{\mathbf{Y},\mathbb{1}}) - \mathbf{\Upsilon}_\mu[f](Y_s + \mathbf{Y}_{s,t}(\underline{1})) \right| \\ & + \left| F(\underline{1}) - \Big\langle \mathbf{X}_{s,t}, P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f]\big[\underbrace{(\mathbf{\Upsilon}, \ldots, \mathbf{\Upsilon})}_{p\text{-times}} \big] \Big\rangle \right| \ . \end{aligned}$$

By the mean value theorem the first term on the right hand side of (3.11) is bounded by

$$\sup_{\lambda \in [0,1]} \left| D\mathbf{\Upsilon}_\mu[f](Y_s + \mathbf{Y}_{s,t}(\underline{1}) + \lambda R_{s,t}^{\mathbf{Y},\mathbb{1}}) \right| \cdot |R_{s,t}^{\mathbf{Y},\mathbb{1}}| \ ,$$

which in turn is bounded by the first term on the right hand side of (3.9).

We now claim that

$$(3.12)$$

$$\sum_{\tilde{f} \in \mathcal{F}_{\leq N-|f|=1}} \frac{\partial^{\tilde{f}} F(0)}{\tilde{f}!} = \Big\langle \mathbf{X}_{s,t}, P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f](Y_s)\big[\underbrace{(\mathbf{\Upsilon}(Y_s), \ldots, \mathbf{\Upsilon}(Y_s))}_{p\text{-times}} \big] \Big\rangle \ .$$

To verify that (3.12) holds, note that we can substitute into (3.12) the right hand side of the first line of (3.6) and also use that

$$\frac{\text{Multi}(\tilde{f})}{\#\tilde{f}!} = \frac{1}{\tilde{f}!} \quad \text{and} \quad \Big\langle \mathbf{X}_{s,t}, D^{\#\tilde{f}} \mathbf{\Upsilon}_\mu[f](Y_s)\big[\big(\mathbf{\Upsilon}[\tilde{h}](Y_s)\tilde{h}\big)_{\tilde{h} \in \tilde{f}} \big] \Big\rangle = \partial^{\tilde{f}} F(0) \ .$$

Applying Proposition 3.11 to $F(\underline{1})$ with $A = \mathcal{F}_{\leq N-1-|f|}$ then gives us

$$(3.13) \qquad \begin{aligned} & F(\underline{1}) - \Big\langle \mathbf{X}_{s,t}, P_{\leq N-|f|-1} \sum_{p=0}^{N-1} \frac{1}{p!} D^p \mathbf{\Upsilon}_\mu[f](Y_s)\big[\underbrace{(\mathbf{\Upsilon}(Y_s), \ldots, \mathbf{\Upsilon}(Y_s))}_{p\text{-times}} \big] \Big\rangle \\ & = \sum_{\tilde{f} \in \partial \mathcal{F}_{\leq N-1-|f|}} \int_{[0,1]^{\mathring{\mathcal{T}}_{\leq N-1}}} (\partial^{\tilde{f}} F)(y) Q^{\tilde{f}}(dy) \ . \end{aligned}$$

We then see the second term on the right hand side of (3.11) is bounded by the second term on the right hand side of (3.9). $\qquad \square$

4. Proof of main Theorems

Before proceeding to the details of the proof we give a brief overview. The key idea is contained in the *interior regularity estimate*, Lemma 4.2, which in turn builds on the Sewing Lemma, Proposition 2.11, and Corollary 3.12. More precisely, the Sewing Lemma provides an estimate of order $(N+1)\alpha$ on the rough integral Z in terms of the various order bounds $[\mathbf{X} : [f]_\mu]$ and the semi-norms $[\sigma_\mu(\mathbf{Y}) : f]$ which measure the quality of local approximations of all coefficients. According to Corollary 3.12 the semi-norms $[\sigma_\mu(\mathbf{Y}) : f]$ can be controlled in terms of

- the size of the coefficients Υ on the relevant interval,
- the order bounds $[\mathbf{X} : [f]_\mu]$, as well as
- the remainder $R_{s,t}^{\mathbf{Y},\mathbb{1}}$.

It is important to note the relevant estimate (3.9) is *linear* in $|R_{s,t}^{\mathbf{Y},\mathbb{1}}|$ and that $|R_{s,t}^{\mathbf{Y},\mathbb{1}}|$ describes an error of order $N\alpha$ in the local description of Y. Furthermore, the quantity $|R_{s,t}^{\mathbf{Y},\mathbb{1}}|$ can be easily estimated in terms of the corresponding quantity for Z. Combining all of these estimates one obtains an estimate on the error of the local approximation of order $N+1$ of Z in terms of (the coefficients Υ, the order bounds $[\mathbf{X} : [f]_\mu]$ and) the quantities $|R_{s,t}^{\mathbf{Z},\mathbb{1}}|$ which measure the error in the approximation of order $N\alpha$. This can be seen as a regularity improvement from $N\alpha$ to $(N+1)\alpha$, which in turn allows to absorb the term $|R_{s,t}^{\mathbf{Z},\mathbb{1}}|$ provided one works on a small-enough interval. The proof of Theorem 1.4 then mostly consists of plugging (a truncated version of) the *interior regularity estimate*, Lemma 4.2, into Lemma 1.11, which captures the damping of the non-linear term as discussed in the introduction, and choosing appropriate parameters depending on the growth assumption on the coefficient σ. The final iteration argument (see Step 3 of the Proof of Theorem 1.4 1)) is an adaptation of the *additive noise case*, [MW20a, Proof of Lemma 2.1, Step 6].

We start the argument with the following simple lemma that translates Assumptions 1.1 and 1.2 into estimates on $\Upsilon[h]$ for $h \in \mathring{\mathcal{T}}_{\leq N}$.

Lemma 4.1. *Let $h \in \mathring{\mathcal{T}}_{\leq N}$ and $0 \leq p \leq N - |h| + 1$.*
Suppose that Assumption 1.1 holds, then

$$\sup_{y \in \mathbb{R}^k} |D^p \Upsilon[h](y)| \lesssim C_\sigma^{|h|} . \tag{4.1}$$

Suppose instead that Assumption 1.2 holds, then

$$|D^p \Upsilon[h](y)| \lesssim C_\sigma^{|h|} \langle y \rangle^{(\gamma-1)|h|+1-p} . \tag{4.2}$$

Proof. We treat the case $p = 0$, the case of $p \neq 0$ can be treated similarly - the constraint on p comes from the fact that we control only N derivatives of the $(\sigma_\mu)_{\mu \in d}$.

A straightforward inductive argument shows that

$$|\Upsilon[h](y)| \lesssim \prod_{u \in N_h} |D^{c(u)} \sigma_{\mu(u)}(y)| ,$$

where on the right hand side we are taking an appropriate operator norm, N_h denotes the set of nodes of the tree h and, given $u \in N_h$, we write $c(u)$ for the number of children nodes of u in h (that is, the number of nodes directly attached to u that are further from the root) and $\mu(u) \in [d]$ for the value of the node decoration that u carries.

Using the fact that the cardinality of N_h is $|h|$, the bound in the case of Assumption 1.1 follows immediately.

In the case of Assumption 1.2, the first bound follows from the fact that $\sum_{u \in N_h} c(u)$ is just the number of edges of h which is $|h| - 1$. □

We can alternatively formulate the above lemma by saying that, for any $f \in \mathcal{F}_{\leq N-1}$ and $\mu \in [d]$, we have, uniformly over $y \in \mathbb{R}^k$ and $0 \leq p \leq N$,

$$|D^p \Upsilon_\mu[f](y)| \lesssim \begin{cases} C_\sigma^{|f|+1} & \text{under the Assumption 1.1,} \\ C_\sigma^{|f|+1} \langle y \rangle^{(\gamma-1)(|f|+1)+1-p} & \text{under Assumption 1.2.} \end{cases}$$

We are now ready to pass to the proofs of Theorems 1.4 and Theorems 1.5. The following "interior regularity estimate" which gives a control of the C^α norm of Z and Y in terms of its lower regularity L^∞ norm is key to both arguments. To keep the length of some formulas within reason we use the following shorthand notation combining some of the quantities appearing on the right hand side of (3.9)

$$U_\mu(f, \mathbf{Y}, I) := \sup\left\{\left|D\Upsilon_\mu[f](Y_s + a)\right| : s, t \in I, |a| \le E_{s,t}^{\mathbf{Y}} + |Y_t - Y_s|\right\},$$

(4.3)

$$U_\mu(f, \bar{f}, \mathbf{Y}, I) := \sup\left\{\left|D^{\#\bar{f}}\Upsilon_\mu[f](Y_s + z)\prod_{h \in \bar{f}}\Upsilon[h](Y_s)\right| : \begin{array}{l} s, t \in I, \\ |z| \le E_{s,t}^{\mathbf{Y}} \end{array}\right\}.$$

Below, for any j, we denote by $\mathcal{F}_{=j}$ the set $\{f \in \mathcal{F} : |f| = j\}$.

Lemma 4.2 (Interior regularity). *Let $\mathbf{Y}$ solve* (1.3) *on* $[0,1]$ *in the sense of Definition 2.12 and let $\mathbf{Z}$ be given by* (2.8). *Let $I \subseteq [0,1]$ be a closed interval of length L and assume that L is small enough to ensure that for all forests $f \in \mathcal{F}_{\le N-1}$ and all $\mu \in [d]$*

(4.4)
$$L^{(|f|+1)\alpha}\left[\mathbf{X} : [f]_\mu\right]U_\mu(f, \mathbf{Y}, I) \le \varepsilon,$$

for a suitably small $\varepsilon = \varepsilon(\alpha, d, k) \le 1$. Then we have the estimates

$$L^{N\alpha}[\mathbf{Z} : \mathbb{1}]_I \lesssim \sum_{f \in \mathcal{F}_{\le N-1}}\sum_{\mu \in [d]}[\mathbf{X} : [f]_\mu]$$

$$\cdot\left\{L^{(|f|+1)\alpha+1}U_\mu(f, \mathbf{Y}, I)\|Y\|_I^m + \max_{\bar{f} \in \partial\mathcal{F}_{\le N-|f|-1}}L^{(|\bar{f}|+|f|+1)\alpha}U_\mu(f, \bar{f}, \mathbf{Y}, I)[\mathbf{X} : \bar{f}]\right\}$$

(4.5)
$$+ \sum_{f \in \mathcal{F}_{=N-1}}\sum_{\mu \in [d]}L^{N\alpha}[\mathbf{X} : [f]_\mu]\|\Upsilon_\mu[f](Y_\bullet)\|_I.$$

Proof. **Part 1: Application of Sewing Lemma.** For $0 \le s \le t \le 1$, the estimate (2.7) implies that

$$\left|Z_t - Z_s - \Xi_{s,t}^{\sigma(\mathbf{Y})}\right| \lesssim |t - s|^{(N+1)\alpha}\sum_{f \in \mathcal{F}_{\le N-1}}\sum_{\mu \in [d]}[\mathbf{X} : [f]_\mu]\left[\sigma_\mu(\mathbf{Y}) : f\right]_{[s,t]},$$

where

$$\Xi_{s,t}^{\sigma(\mathbf{Y})} := \sum_{\mu \in [d]}\Xi_{s,t}^{\sigma_\mu(\mathbf{Y})} = \sum_{\mu \in [d]}\sum_{f \in \mathcal{F}_{\le N-1}}\langle\mathbf{X}_{s,t}, [f]_\mu\rangle\langle f, \sigma_\mu(\mathbf{Y})_s\rangle \in \mathbb{R}^k.$$

By coherence of $\mathbf{Y}$, i.e. by (2.10), we get,

$$\Xi_{s,t}^{\sigma(\mathbf{Y})} = \sum_{h \in \mathring{\mathcal{T}}_{\le N-1}}\langle\mathbf{X}_{s,t}, h\rangle\langle h, \mathbf{Y}_s\rangle + \sum_{f \in \mathcal{F}_{=N-1}}\sum_{\mu \in [d]}\langle\mathbf{X}_{s,t}, [f]_\mu\rangle\langle f, \sigma_\mu(\mathbf{Y})_s\rangle.$$

Recalling that $\delta\mathbf{Y} = \delta\mathbf{Z}$ this leads to

$$\begin{aligned}|R_{s,t}^{\mathbf{Z},\mathbb{1}}| &= \left|Z_t - Z_s - \langle\mathbf{X}_{s,t}, \delta\mathbf{Y}_s\rangle\right| \\ &\le \left|Z_t - Z_s - \Xi_{s,t}^{\sigma(\mathbf{Y})}\right| + \left|\Xi_{s,t}^{\sigma(\mathbf{Y})} - \langle\mathbf{X}_{s,t}, \delta\mathbf{Y}_s\rangle\right| \\ &\lesssim |t - s|^{(N+1)\alpha}\sum_{f \in \mathcal{F}_{\le N-1}}\sum_{\mu \in [d]}[\mathbf{X} : [f]_\mu]\left[\sigma_\mu(\mathbf{Y}) : f\right]_{[s,t]} \\ &\quad + |t - s|^{N\alpha}\sum_{\mathcal{F}_{=N-1}}\sum_{\mu \in [d]}[\mathbf{X} : [f]_\mu] \cdot |\langle f, \sigma_\mu(\mathbf{Y})_s\rangle|.\end{aligned}$$

Dividing by $|t - s|^{N\alpha}$ and taking the supremum over all $s < t$ in I we arrive at

$$[\mathbf{Z} : \mathbb{1}]_I \lesssim L^\alpha\sum_{f \in \mathcal{F}_{\le N-1}}\sum_{\mu \in [d]}[\mathbf{X} : [f]_\mu]\left[\sigma_\mu(\mathbf{Y}) : f\right]_I$$

(4.6)
$$+ \sum_{f \in \mathcal{F}_{=N-1}}\sum_{\mu \in [d]}[\mathbf{X} : [f]_\mu] \cdot \|\Upsilon_\mu[f](Y_\bullet)\|_I.$$

Part 2: Application of Corollary 3.12. We use the estimate (3.9) in Corollary 3.12 in the form

(4.7)
$$[\sigma_\mu(\mathbf{Y}):f]_I \lesssim L^{|f|\alpha} U_\mu(f,\mathbf{Y},I)[\mathbf{Y}:\mathbb{1}]_I + \max_{\bar{f} \in \partial \mathcal{F}_{\leq N-|f|-1}} L^{(|\bar{f}|+|f|-N)\alpha} U_\mu(f,\bar{f},\mathbf{Y},I)[\mathbf{X}:\bar{f}] \ .$$

Note that $|\bar{f}| \geq N - |f|$, so that the exponent $(|\bar{f}| + |f| - N)\alpha$ in the second term on the RHS of (4.7) is always non-negative.

We plug this estimate into the RHS of (4.6) resulting in

$$[\mathbf{Z}:\mathbb{1}]_I \lesssim L^\alpha \sum_{f \in \mathcal{F}_{\leq N-1}} \sum_{\mu \in [d]} [\mathbf{X}:[f]_\mu]$$

$$\cdot \left\{ L^{|f|\alpha} U_\mu(f,\mathbf{Y},I)[\mathbf{Y}:\mathbb{1}]_I + \max_{\bar{f} \in \partial \mathcal{F}_{\leq N-|f|-1}} L^{(|\bar{f}|+|f|-N)\alpha} U_\mu(f,\bar{f},\mathbf{Y},I)[\mathbf{X}:\bar{f}] \right\}$$

(4.8) $$+ \sum_{\mathcal{F}_{=N-1}} \sum_{\mu \in [d]} [\mathbf{X}:[f]_\mu] \cdot \|\Upsilon_\mu[f](Y_\bullet)\|_I \ .$$

The key observation at this point is that the terms $[\mathbf{Y}:\mathbb{1}]_I$ on the RHS can be replaced by $[\mathbf{Z}:\mathbb{1}]_I$. Indeed, equation (2.14) which relates the tree paths $\mathbf{Y}$ and $\mathbf{Z}$ immediately implies the bound

$$[\mathbf{Y}:\mathbb{1}]_I \leq [\mathbf{Z}:\mathbb{1}]_I + L^{1-N\alpha}\|Y\|_I^m \ .$$

Plugging this into the RHS of (4.8), the resulting terms involving $[\mathbf{Z}:\mathbb{1}]_I$ can be absorbed into the LHS, provided their pre-factors are small. More precisely,

$$[\mathbf{Z}:\mathbb{1}]_I \lesssim L^\alpha \sum_{f \in \mathcal{F}_{\leq N-1}} \sum_{\mu \in [d]} [\mathbf{X}:[f]_\mu]$$

$$\cdot \left\{ L^{|f|\alpha} U_\mu(f,\mathbf{Y},I) L^{1-N\alpha}\|Y\|_I^m + \max_{\bar{f} \in \partial \mathcal{F}_{\leq N-|f|-1}} L^{(|\bar{f}|+|f|-N)\alpha} U_\mu(f,\bar{f},\mathbf{Y},I)[\mathbf{X}:\bar{f}] \right\}$$

$$+ \sum_{|f| \in \mathcal{F}_{=N-1}} \sum_{\mu \in [d]} [\mathbf{X}:[f]_\mu] \cdot \|\Upsilon_\mu[f](Y_\bullet)\|_I \ ,$$

provided for all forests $f \in \mathcal{F}_{\leq N-1}$ and all $\mu \in [d]$

$$L^{(|f|+1)\alpha} [\mathbf{X}:[f]_\mu] \, U_\mu(f,\mathbf{Y},I) \leq \varepsilon \ ,$$

for a suitably small $\varepsilon = \varepsilon(\alpha,d,k) \leq 1$. This is the desired estimate (4.9).   □

It turns out, that for the proof of Theorems 1.4 and 1.5 only the following truncated version of the estimate (4.5) is required

Corollary 4.3. *Let* $\mathbf{Y}$ *solve* (1.3) *on* $[0,1]$ *in the sense of Definition 2.12 and let* $\mathbf{Z}$ *be given by* (2.8). *Let* $I \subseteq [0,1]$ *be a closed interval of length* L *satisfying* (4.4). *Then we have the estimates*

$$L^\alpha[Z]_{\alpha,I} \lesssim \sum_{f \in \mathcal{F}_{\leq N-1}} \sum_{\mu \in [d]} [\mathbf{X}:[f]_\mu]$$

$$\cdot \left\{ L^{(|f|+1)\alpha+1} U_\mu(f,\mathbf{Y},I)\|Y\|_I^m + \max_{\bar{f} \in \partial \mathcal{F}_{\leq N-|f|-1}} L^{(|\bar{f}|+|f|+1)\alpha} U_\mu(f,\bar{f},\mathbf{Y},I)[\mathbf{X}:\bar{f}] \right\}$$

(4.9) $$+ \sum_{h \in \mathcal{T}_{\leq N}} \sum_{\mu \in [d]} L^{|h|\alpha}[\mathbf{X}:h]\|\Upsilon[h](Y_\bullet)\|_I$$

and

(4.10) $$L^\alpha[Y]_{\alpha,I} \lesssim \ RHS \ of \ (4.9) + L\|Y\|_I^m \ .$$

Proof. The estimate (4.9) follows from (4.5) combined with

$$L^\alpha[Z]_{\alpha,I} \leq \sum_{h\in\mathring{\mathcal{T}}_{\leq N-1}} L^{|h|\alpha}[\mathbf{X}:h]\|\mathbf{Y}:h\|_I + L^{N\alpha}[\mathbf{Z}:\mathbb{1}]_I \,,$$

as well as the identification

$$\sum_{\mathcal{F}_{=N-1}} \sum_{\mu\in[d]} L^{N\alpha}[\mathbf{X}:[f]_\mu] \cdot \|\Upsilon_\mu[f](Y_\bullet)\|_I = \sum_{h\in\mathcal{T}_{=N}} L^{|h|\alpha}[\mathbf{X}:h] \cdot \|\Upsilon[h](Y_\bullet)\|_I \,.$$

The estimate (4.10) is an immediate consequence of the defining relationship $Y_t = y_0 - \int_0^t |Y_r|^{m-1}Y_r dr + Z_t$. $\qquad\square$

Proof of Theorem 1.4 (1) - The case of bounded coefficients. **Step 1: Simplification of interior regularity estimate.** We now work under Assumption 1.1, i.e. we assume that σ and its derivatives up to order N are bounded. Using the boundedness of coefficients, and estimate (4.1) in Lemma 4.1, this the estimate (4.9) simplifies to

$$
\begin{aligned}
L^\alpha[Z]_{\alpha,I} \lesssim &\sum_{h\in\mathring{\mathcal{T}}_{\leq N}} L^{|h|\alpha}[\mathbf{X}:h] \\
&+ \sum_{f\in\mathcal{F}_{\leq N-1}} \sum_{\mu\in[d]} L^{(|f|+1)\alpha+1}[\mathbf{X}:[f]_\mu]\|Y\|_I^m \\
(4.11) \qquad &+ \sum_{f\in\mathcal{F}_{\leq N-1}} \sum_{\mu\in[d]} \max_{\bar{f}\in\partial\mathcal{F}_{\leq N-|f|-1}} L^{(|\bar{f}|+|f|+1)\alpha}[\mathbf{X}:[f]_\mu][\mathbf{X}:\bar{f}]
\end{aligned}
$$

provided I is an interval of length L satisfying

$$(4.12) \qquad L^{(|f|+1)\alpha}[\mathbf{X}:[f]_\mu] \leq \varepsilon \,,$$

for all forests $f \in \mathcal{F}_{\leq N-1}$ and for a potentially smaller value of $\varepsilon = \varepsilon(\alpha,d,k,C_\sigma) \leq 1$. In order to simplify these expressions further, we make additional assumptions: first, we fix a $t^\star \in [0,1]$ and for the remainder of this step we assume that the interval I satisfies $I \subseteq [t^\star,1)$. Next, we work under the assumption that for all trees $h \in \mathring{\mathcal{T}}_{\leq N}$ we have

$$(4.13) \qquad [\mathbf{X}:h] \leq \varepsilon\|Y\|_{[t^\star,1]}^{|h|\delta\alpha} \,,$$

for $\delta = m$. Under (4.13) the condition (4.12) is guaranteed as soon as

$$(4.14) \qquad L \leq \|Y\|_{[t^\star,1]}^{-\delta} \,.$$

Under these assumptions the estimate (4.11) simplifies further to

$$(4.15) \qquad L^\alpha[Z]_{\alpha,I} \lesssim \varepsilon(1+L\|Y\|_{[t^\star,1]}^m + 1 + 1) \lesssim \varepsilon + \varepsilon L\|Y\|_{[t^\star,1]}^m \,.$$

The estimate (4.10) turns into

$$(4.16) \qquad L^\alpha[Y]_{\alpha,I} \lesssim \varepsilon + L\|Y\|_{[t^\star,1]}^m \,.$$

Step 2: Application of Lemma 1.11 We continue working under the boundedness Assumption 1.1 on the coefficients, and assuming the bound (4.13) on the trees $[\mathbf{X}:h]$. We apply Lemma 1.11 on the interval $[t_\star,1]$ and for

$$L = \|Y\|_{[t^\star,1]}^{-\delta} = \|Y\|_{[t^\star,1]}^{-m} \,.$$

We assume throughout that all times involved are contained in $[0,1]$ which implies in particular that $L \in (0,1)$ so that (4.14) implies

$$(4.17) \qquad \|Y\|_{[t^\star,1]} \geq 1.$$

Under these assumptions, combining Lemma 1.11 with (4.15), (4.16) we obtain for $t > t_\star + L$

$$|Y(t)| \lesssim \max \left\{ (t - t^\star - L)^{-\frac{1}{m-1}} ; \left(\varepsilon L^{-1} + \varepsilon \|Y\|_{[t^\star,1]}^m \right)^{\frac{1}{m}} ; \right.$$
$$\left. \left(\|Y\|_{[t^\star,1]}^{m-1} \left(\varepsilon + L\|Y\|_{[t^\star,1]}^m \right) \right)^{\frac{1}{m}} ; \left(\varepsilon + L\|Y\|_{[t^\star,1]}^m \right) \right\}$$
$$\lesssim \max \left\{ (t - t^\star - L)^{-\frac{1}{m-1}} ; \varepsilon^{\frac{1}{m}} \|Y\|_{[t^\star,1]} ; \|Y\|_{[t^\star,1]}^{\frac{m-1}{m}} ; 1 \right\} .$$

We bound using Young's inequality

$$\|Y\|_{[t^\star,1]}^{\frac{m-1}{m}} \leq \varepsilon \|Y\|_{[t^\star,1]} + C(\varepsilon, m) ,$$

and then choose first ε small and absorb all constants into the term $(t - t^\star - L)^{-\frac{1}{m-1}} \geq 1$ and conclude that if t is large enough to ensure that

$$(4.18) \qquad\qquad (t - t^\star - L)^{-\frac{1}{m-1}} \leq \bar{\varepsilon} \|Y\|_{[t^\star,1]},$$

for $\bar{\varepsilon} = \bar{\varepsilon}(\alpha, d, k, C_\sigma)$ small enough, we get

$$(4.19) \qquad\qquad |Y(t)| \leq \frac{1}{2} \|Y\|_{[t^\star,1]}.$$

We observe that (4.18) is guaranteed as soon as

$$t - t^\star \geq \bar{\varepsilon}^{-(m-1)} \|Y\|_{[t^\star,1]}^{-(m-1)} + \|Y\|_{[t^\star,1]}^{-m}.$$

which, taking into account (4.17) follows from the simpler condition

$$(4.20) \qquad\qquad t - t^\star \geq (\bar{\varepsilon}^{-(m-1)} + 1) \|Y\|_{[t^\star,1]}^{-(m-1)}.$$

Step 3: Coming down from infinity for bounded coefficients. We continue working under the boundedness Assumption 1.1 and aim to establish the estimate (1.4). We define a finite sequence $0 = t_0 < t_1 < t_2 < \ldots < t_N = 1$ by setting

$$(4.21) \qquad\qquad t_{n+1} - t_n = A\|Y\|_{[t_n,1]}^{-(m-1)},$$

where $A = \bar{\varepsilon}^{-(m-1)} + 1$ (see (4.20)), as long as the resulting t_{n+1} satisfies $t_{n+1} < 1$. As soon as t_{n+1} defined by this rule is ≥ 1 we set $t_{n+1} = t_N = 1$ and terminate the sequence. Note that the sequence $t_{n+1} - t_n$ is increasing by definition (at least before passing 1), so that this algorithm must terminate after a finite number of steps.

The time t is contained in precisely one interval $(t_n, t_{n+1}]$ and precisely one of the following must hold:

   (1) either there exists a $k \leq n$ and a tree h such that

$$\varepsilon \|Y\|_{[t_k,1]}^{|h|m\alpha} < [\mathbf{X} : h] ,$$

   (2) or for all $k \leq n$ all trees h satisfy condition (4.13) for $\delta = m$.

In case (1), the desired estimate (1.4) follows immediately from

$$|Y(t)| \leq \|Y\|_{[t_k,1]} \leq \frac{1}{\varepsilon} [\mathbf{X} : h]^{\frac{1}{|h|m\alpha}} .$$

In case (2) invoking (4.19) repeatedly for $t_* = t_k$ yields

$$(4.22) \qquad\qquad \|Y\|_{[t_n,1]} \leq 2^{k-n} \|Y\|_{[t_k,1]} \qquad \text{for } k = 0, 1, 2, \ldots n.$$

We split (2) further into the following three cases:

   (2a) $n + 1 < N$.
   (2b) $n + 1 = N$ and $t_n \leq \frac{1}{2}$.
   (2c) $n + 1 = N$ and $t_n > \frac{1}{2}$.

In case (2a) we write using (4.22)

$$t_{n+1} = \sum_{k=0}^{n} t_{k+1} - t_k = A \sum_{k=0}^{n} \|Y\|_{[t_k,1]}^{-(m-1)} \leq A\|Y\|_{[t_n,1]}^{-(m-1)} \sum_{k=0}^{n} 2^{-(m-1)(n-k)}$$
$$\lesssim \|Y\|_{[t_n,1]}^{-(m-1)},$$

which implies

$$|Y(t)| \leq \|Y\|_{[t_n,1]} \lesssim t_{n+1}^{-\frac{1}{m-1}} \leq t^{-\frac{1}{m-1}},$$

as desired.

For (2b) to hold by definition (4.21) we have

$$A\|Y\|_{[t_n,1]}^{-(m-1)} \geq \frac{1}{2},$$

and in particular

$$|Y(t)| \leq \|Y\|_{[t_n,1]} \leq \left(\frac{1}{2A}\right)^{-\frac{1}{m-1}},$$

which implies (1.4).

Finally, in case (2c) we can apply case (2a) to t_n to get

$$|Y(t)| \leq \|Y\|_{[t_n,1]} \leq t_n^{-\frac{1}{m-1}} \leq \left(\frac{1}{2}\right)^{-\frac{1}{m-1}},$$

completing the proof of (1.4). $\qquad\square$

Proof of Theorem 1.4 (2) - The case of polynomial coefficients. The proof follows the same general lines as the proof in the case of bounded coefficients and we only explain the differences. Combining, their definition (4.3) with the estimate on Υ provided by Lemma 4.1, yields

(4.23)
$$U_\mu(f, \mathbf{Y}, I) \lesssim \sup \left\{ \langle Y_s + a \rangle^{(\gamma-1)(|f|+1)} : s, t \in I, |a| \leq E_{s,t}^{\mathbf{Y}} + |Y_t - Y_s| \right\},$$
$$U_\mu(f, \bar{f}, \mathbf{Y}, I) \lesssim \sup \left\{ \langle Y_s + z \rangle^{(\gamma-1)(|f|+1)+1-\#\bar{f}} \langle Y_s \rangle^{(\gamma-1)|\bar{f}|+\#\bar{f}} : \begin{array}{l} s, t \in I, \\ |z| \leq E_{s,t}^{\mathbf{Y}} \end{array} \right\}.$$

We follow the same approach as Step 1 of the proof in the case of bounded coefficients, fixing $t^\star \in [0,1)$, but here the condition (4.14) is replaced by

(4.24)
$$L \leq \varepsilon_2 \langle Y \rangle_{[t^\star,1]}^{-(m-1)}$$

where we have introduced an auxiliary parameter $\varepsilon_2 \leq 1$ that will ultimately be chosen small but still large relative to ε and we have changed values of exponents from $\delta = m$ to $\delta = m - 1$. Instead of (4.13) we assume that for $h \in \mathring{\mathcal{T}}_{\leq N}$

(4.25)
$$[\mathbf{X} : h] \leq \varepsilon_1 \|Y\|_{[t^\star,1]}^{|h|(\alpha(m-1)-(\gamma-1))}.$$

Note that the exponent $|h|(\alpha(m-1)-(\gamma-1))$ appearing on the right hand side is positive if and only if the upper bound $\gamma < (m-1)\alpha + 1$ in Assumption 1.5 holds. Under (4.24) and (4.25) we have

$$E_{s,t}^{\mathbf{Y}} = \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} |\langle \mathbf{X}_{s,t}, h \rangle \Upsilon[h](Y_s)|$$
$$\lesssim \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} \left(\varepsilon_2 \langle Y \rangle_{[t^\star,1]}^{-(m-1)} \right)^{\alpha|h|} \varepsilon_1 \|Y\|_{[t^\star,1]}^{|h|(\alpha(m-1)-(\gamma-1))} \langle Y \rangle_{[t^\star,1]}^{(\gamma-1)|h|+1}$$
(4.26)
$$\lesssim \varepsilon_2^\alpha \varepsilon_1 \langle Y \rangle_{[t^\star,1]}.$$

Note that in the first inequality, we have made use of (4.2) and the assumption $\gamma \geq 1$ in (1.5) which ensures that the exponent $(\gamma-1)|h|+1$ in the last term is non-negative and permits to use the estimate

(4.27)
$$\left| \Upsilon[h](Y_s) \right| \lesssim \langle Y_s \rangle^{(\gamma-1)|h|+1} \leq \langle Y \rangle_{[t^\star,1]}^{(\gamma-1)|h|+1}.$$

We now enforce that

$$\varepsilon_1 \varepsilon_2^\alpha \leq \frac{\varepsilon}{C} \, ,$$
(4.28)

Using this, (4.26), and once more the fact that $(\gamma - 1) \geq 1$ the estimate (4.23) simplifies to

$$U_\mu(f, \mathbf{Y}, I) \lesssim \langle Y \rangle_{[t^\star, 1]}^{(\gamma-1)(|f|+1)},$$

$$U_\mu(f, \bar{f}, \mathbf{Y}, I) \lesssim \langle Y \rangle_{[t^\star, 1]}^{(\gamma-1)(|f|+\bar{f}+1)+1} \, .$$
(4.29)

Note that these estimates in conjunction with (4.28), (4.24), and Assumption (4.25) imply (4.4) and in this case (4.9) and (4.10) turn into

$$L^\alpha[Z]_{\alpha,I} \lesssim \sum_{f \in \mathcal{F}_{\leq N-1}} \sum_{\mu \in [d]} [\mathbf{X} : [f]_\mu]$$

$$\cdot \left\{ L^{(|f|+1)\alpha+1} \langle Y \rangle_{[t^\star,1]}^{(|f|+1)(\gamma-1)} \|Y\|_I^m + \max_{\bar{f} \in \partial \mathcal{F}_{\leq N-|f|-1}} L^{(|\bar{f}|+|f|+1)\alpha} \langle Y \rangle_{[t^\star,1]}^{(|\bar{f}|+|f|+1)(\gamma-1)+1} [\mathbf{X} : \bar{f}] \right\}$$

$$+ \sum_{h \in \mathring{\mathcal{T}}_{\leq N}} L^{|h|\alpha} [\mathbf{X} : h] \langle Y \rangle_{[t^\star,1]}^{|h|(\gamma-1)+1}$$

(4.30)

$$\lesssim \varepsilon (L \langle Y \rangle_{[t^\star,1]}^m + \langle Y \rangle_{[t^\star,1]} + \langle Y \rangle_{[t^\star,1]}) \lesssim \varepsilon \langle Y \rangle_{[t^\star,1]} \, ,$$

and

$$L^\alpha[Y]_{\alpha,I} \lesssim \varepsilon \langle Y \rangle_{[t^\star,1]} + L \|Y\|_{[t^\star,1]}^m \lesssim \varepsilon_2 \langle Y \rangle_{[t^\star,1]}^m \, .$$

As above in Step 2 we next apply Lemma 1.11 on the interval $[t_\star, 1]$ yielding

$$\langle Y(t) \rangle \lesssim \max \left\{ (t - t^\star - L)^{-\frac{1}{m-1}}; \left(\frac{\varepsilon}{\varepsilon_2} \right)^{\frac{1}{m}} \|Y\|_{[t^\star,1]}; \right.$$

$$\left. \left(\|Y\|_{[t^\star,1]}^{m-1} \left(\varepsilon_2 \langle Y \rangle_{[t^\star,1]} \right) \right)^{\frac{1}{m}}; \varepsilon_2 \langle Y \rangle_{[t^\star,1]} \right\}.$$

Then choosing first ε_2 small to control the pre-factors in the second line and then also ε small to also make the pre-factors in the first line small (this amounts to choosing ε_1 in (4.25) small) we obtain

$$\langle Y(t) \rangle \leq \max \left\{ C(t - t^\star - L)^{-\frac{1}{m-1}}; \frac{1}{2} \langle Y \rangle_{[t^\star,1]} \right\}.$$

The proof concludes as in the case of bounded coefficients, Step 3 above, where in (4.21) and in what follows, the role of $\|Y\|_{[t_k,1]}$ is played by $\langle Y \rangle_{[t_k,1]}$. □

Proof of Theorem 1.5. We first show (1.8) under Assumption 1.1 (i.e. for bounded coefficients). We first define the auxiliary time

$$T_0 = \inf \left\{ t \in [0,1] : \langle y(t) \rangle \geq 3 \langle y_0 \rangle \right\} ,$$
(4.31)

with the convention that $T_0 = 1$ if the set is empty. Continuity of $|y(t)|$ in t implies that if $T_0 < 1$ we have

$$\langle y(T_0) \rangle = 3 \langle y_0 \rangle \, .$$
(4.32)

Furthermore, let T_1 and T_2 be as in (1.7), where ε_2 is assumed to be small enough to play the role of ε in the absorption condition (4.12) above and ε_1 is fixed below, and set

$$T^\star = \min\{T_0, T_1, T_2\} \, .$$

Note that by choice of T_2 condition (4.12) is automatically satisfied for $I = [0, T^\star]$ and (4.11) turns into

$$(T^\star)^\alpha [Z]_{\alpha,[0,T^\star]} \lesssim \varepsilon_2 + \varepsilon_2 \varepsilon_1 3^m \langle y_0 \rangle \, ,$$

where we have used the fact that by definition

$$T^\star \langle Y \rangle_{[0,T^\star]}^m \leq \varepsilon_1 \frac{1}{\langle y_0 \rangle^{m-1}} 3^m \langle y_0 \rangle^m \leq \varepsilon_1 3^m \langle y_0 \rangle \ .$$

Hence we get from (1.3), choosing $\varepsilon_1 = \frac{1}{2} 3^{-m}$ and possibly making ε_2 smaller

$$\langle Y \rangle_{[0,T^\star]} \leq \|Y\|_{[0,T^\star]} + 1 \leq |y_0| + T^\star \|Y\|_{[0,T^\star]}^m + (T^\star)^\alpha [Z]_\alpha$$

$$\leq |y_0| + \varepsilon_1 3^m \langle y_0 \rangle + C\Big(\varepsilon_2 + \varepsilon_2 \varepsilon_1 3^m \langle y_0 \rangle \Big)$$

(4.33)
$$\leq \frac{3}{2} \langle y_0 \rangle + 1 \leq \frac{5}{2} \langle y_0 \rangle \ .$$

To conclude that (1.8) holds indeed, it remains to see that $T^\star = \min\{T_0, T_1, T_2\} = \min\{T_1, T_2\}$. This is indeed the case, because otherwise we would have $\|Y\|_{[0,T^\star]} = \|Y\|_{[0,T_0]}$ and $T_0 < 1$, but then (4.33) would contradict (4.32).

We next pass to the case of polynomially bounded σ, i.e. to establishing (1.10) under Assumption 1.2. Let T_0 be as in (4.31) and set

$$T^\star = \min\{T_0, T_1, T_2\} \ ,$$

where this time T_2 is defined in (1.9). We claim that on $I = [0, T^\star]$ satisfies the Assumption (4.4). Indeed, just as in (4.26) above one can see that for s, t we have

$$E_{s,t}^{\mathbf{Y}} = \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} |\langle \mathbf{X}_{s,t}, h \rangle \Upsilon[h](Y_s)|$$

$$\lesssim \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} (T^\star)^{|h|\alpha} [\mathbf{X} : h] \|\Upsilon[h](Y_\bullet)\|_{[0,T^\star]}$$

$$\lesssim \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} \varepsilon_2^{|h|\alpha} \frac{1}{[\mathbf{X} : h]} \Big(\frac{1}{\langle y_0 \rangle}\Big)^{(\gamma-1)|h|} [\mathbf{X} : h] \langle Y \rangle_{[0,T^\star]}^{(\gamma-1)|h|+1}$$

$$\lesssim \sum_{h \in \mathring{\mathcal{T}}_{\leq N-1}} \varepsilon_2^{|h|\alpha} \frac{1}{[\mathbf{X} : h]} \Big(\frac{1}{\langle y_0 \rangle}\Big)^{(\gamma-1)|h|} [\mathbf{X} : h] \langle Y \rangle_{[0,T^\star]}^{(\gamma-1)|h|+1}$$

$$\lesssim \varepsilon_2^\alpha \langle Y \rangle_{[0,T^\star]} \ ,$$

where in the first inequality the fact $T^\star \leq T_2$ and in the last inequality $T^\star \leq T_0$ was used. From this one can conclude just as in (4.29) that

$$U_\mu(f, \mathbf{Y}, [0, T^\star]) \lesssim \langle Y \rangle_{[0,T^\star]}^{(\gamma-1)(|f|+1)},$$

$$U_\mu(f, \bar{f}, \mathbf{Y}, [0, T^\star]) \lesssim \langle Y \rangle_{[0,T^\star]}^{(\gamma-1)(|f|+\bar{f}+1)+1} \ ,$$

which in turn leads to

$$L^\alpha [Z]_{\alpha, [0,T^\star]} \lesssim \varepsilon \langle Y \rangle_{[0,T^\star]} \lesssim \varepsilon \langle y_0 \rangle$$

where the first inequality follows like (4.30) and the second follows from $T^\star \leq T_0$. The rest of the argument follows as in the case of bounded coefficients. □

Proof of Corollary 1.6 . We treat the case of bounded coefficients first. If we have $t \geq T_1 = \varepsilon_1 \frac{1}{\langle y_0 \rangle^{m-1}}$ or if $t \geq T_2 = \varepsilon_2 \min\left\{\frac{1}{[\mathbf{X} : h]^{\frac{1}{|h|\alpha}}} : h \in \mathring{\mathcal{T}}_{\leq N}\right\}$ defined in (1.7) we use the estimate (1.4)

$$|Y(t)| \lesssim \max\left\{t^{-\frac{1}{m-1}} , \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{m\alpha|h|}}\right\} \ .$$

In the second term on the right-hand-side, we can replace the exponent $\frac{1}{m\alpha|h|}$ by the slightly larger $\frac{1}{(m-1)\alpha|h|}$ at the expense of adding the additional 1 into the maximum in (1.11). For

the first term, we use the lower bound on t provided by T_1 and T_2. Indeed, for $t \geq T_1$ we get

$$(4.34) \qquad t^{-\frac{1}{m-1}} \leq \varepsilon_1^{-\frac{1}{m-1}} \langle y_0 \rangle$$

and for $t \geq T_2$ we have

$$t^{-\frac{1}{m-1}} \leq \varepsilon_2^{-\frac{1}{m-1}} \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} \colon h]^{\frac{1}{(m-1)|h|\alpha}}$$

in each case establishing (1.11). Finally, if $t \leq \min\{T_1, T_2\}$ (1.11) follows from (1.8).

The argument for polynomial coefficients is very similar. If $t \geq T_1$ or if $t \geq T_2$, where this time T_2 is defined by (1.9), we use the estimate (1.6)

$$|Y(t)| \lesssim \max \left\{ t^{-\frac{1}{m-1}} , \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{((m-1)\alpha-\gamma+1)|h|}} \right\} .$$

For $t \geq T_1$ (4.34) gives the desired (1.12), while for $t \geq T_2$ we write

$$t^{-\frac{1}{m-1}} \leq \varepsilon_2^{-\frac{1}{m-1}} \langle y_0 \rangle^{\frac{\gamma-1}{(m-1)\alpha}} \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} \colon h]^{\frac{1}{(m-1)|h|\alpha}}$$

$$\lesssim \langle y_0 \rangle + \max_{h \in \mathring{\mathcal{T}}_{\leq N}} [\mathbf{X} : h]^{\frac{1}{((m-1)\alpha-\gamma+1)|h|}} ,$$

where in the second inequality we have used the estimate $ab \leq a^p + b^q$ for $p = \frac{(m-1)\alpha}{\gamma-1}$ and $q = \frac{(m-1)\alpha}{(m-1)\alpha-\gamma+1}$. As above, if $t \leq \min\{T_1, T_2\}$ (1.12) follows from (1.10). $\qquad \square$

References

[BCFP19] Y. Bruned, I. Chevyrev, P. K. Friz, and R. Preiß. A rough path perspective on renormalization. *J. Funct. Anal.*, 277(11):108283, 60, 2019.

[BFT20] Christian Bayer, Peter Friz, and Nikolas Tapia. Stability of deep neural networks via discrete rough paths. *WIAS Preprint*, 2732, 2020.

[Bro00] Ch. Brouder. Runge Kutta methods and renormalization. *European Physical Journal C*, 12(3):521–534, February 2000.

[Cer01] Sandra Cerrai. *Second order PDE's in finite and infinite dimension*, volume 1762 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001. A probabilistic approach.

[CHJ13] Sonja Cox, Martin Hutzenthaler, and Arnulf Jentzen. Local lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations. *arXiv preprint arXiv:1309.5595*, 2013.

[CK98] Alain Connes and Dirk Kreimer. Hopf Algebras, Renormalization and Noncommutative Geometry. *Communications in Mathematical Physics*, 199(1):203–242, January 1998.

[CMW19] Ajay Chandra, Augustin Moinat, and Hendrik Weber. A priori bounds for the ϕ^4 equation in the full sub-critical regime, 2019.

[Dav08] A. M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX*, pages Art. ID abm009, 40, 2008. [Issue information previously given as no. 2 (2007)].

[FH14] Peter K. Friz and Martin Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

[Foi10] Loïc Foissy. Free brace algebras are free pre-Lie algebras. *Comm. Algebra*, 38(9):3358–3369, 2010.

[Foi13] Loïc Foissy. An introduction to hopf algebras of trees. *preprint*, 2013.

[FV10a] Peter Friz and Nicolas Victoir. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(2):369–413, 2010.

[FV10b] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[GH19] Andris Gerasimovičs and Martin Hairer. Hörmander's theorem for semilinear spdes. *Electronic Journal of Probability*, 24, 2019.

[GL89] Robert Grossman and Richard G Larson. Hopf-algebraic structure of families of trees. *Journal of Algebra*, 126(1):184–210, 1989.

[Gub03] Massimiliano Gubinelli. Controlling Rough Paths. *arXiv:math/0306433*, June 2003. arXiv: math/0306433.

[Gub06] M. Gubinelli. Ramification of rough paths. *arXiv:math/0610300*, October 2006. arXiv: math/0610300.

[Hai14]   Martin Hairer. A theory of regularity structures. *Inventiones mathematicae*, 198(2):269–504, 2014.

[HK15]   Martin Hairer and David Kelly. Geometric versus non-geometric rough paths. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 51(1):207–251, February 2015.

[Hof03]   Michael Hoffman. Combinatorics of rooted trees and hopf algebras. *Transactions of the American Mathematical Society*, 355(9):3795–3811, 2003.

[KM16]   David Kelly and Ian Melbourne. Smooth approximation of stochastic differential equations. *The Annals of Probability*, 44(1):479–520, 2016.

[Lej12]   Antoine Lejay. *Global Solutions to Rough Differential Equations with Unbounded Vector Fields*, pages 215–246. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[LQ02]   Terry Lyons and Zhongmin Qian. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.

[LS11]   X.-M. Li and M. Scheutzow. Lack of strong completeness for stochastic flows. *Annals of Probability*, 39(4):1407–1421, 2011.

[Lyo94]   Terry Lyons. Differential equations driven by rough signals (I): an extension of an inequality of L. C. Young. *Mathematical Research Letters*, 1(4):451–464, 1994.

[Lyo98]   Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[MW20a]   Augustin Moinat and Hendrik Weber. Local bounds for stochastic reaction diffusion equations. *Electron. J. Probab.*, 25:26 pp., 2020.

[MW20b]   Augustin Moinat and Hendrik Weber. Space-time localisation for the dynamic model. *Communications on Pure and Applied Mathematics*, 73(12):2519–2555, 2020.

[OG08]   J.-M. Oudom and D. Guin. On the Lie enveloping algebra of a pre-Lie algebra. *Journal of K-Theory*, 2(1):147–167, 2008.

[RS17]   S. Riedel and M. Scheutzow. Rough differential equations with unbounded drift term. *J. Differential Equations*, 262(1):283–312, 2017.

[TW18]   Pavlos Tsatsoulis and Hendrik Weber. Spectral gap for the stochastic quantization equation on the 2-dimensional torus. *Ann. Inst. H. Poincaré Probab. Statist.*, 54(3):1204–1249, 08 2018.

[Wei18]   Martin Philipp Weidner. *A geometric view on rough differential equations*. PhD thesis, Imperial College London, 2018.

TIMOTHEE BONNEFOI, ÉCOLE NORMALE SUPÉRIEURE DE LYON
Email address: `timothee.bonnefoi@ens-lyon.fr`

AJAY CHANDRA, IMPERIAL COLLEGE LONDON
Email address: `a.chandra@imperial.ac.uk`

AUGUSTIN MOINAT, RADRiGS SLACKLINES
Email address: `augustin.moinat@gmail.com`

HENDRIK WEBER, UNIVERSITY OF BATH
Email address: `h.weber@bath.ac.uk`