

# Stochastic quantization of Yang–Mills

Cite as: J. Math. Phys. **63**, 091101 (2022); <https://doi.org/10.1063/5.0089431>

Submitted: 25 February 2022 • Accepted: 11 August 2022 • Published Online: 12 September 2022

 Ilya Chevyrev



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[On the development of shocks in fluids](#)

Journal of Mathematical Physics **63**, 091102 (2022); <https://doi.org/10.1063/5.0075767>

[A covariant Stinespring theorem](#)

Journal of Mathematical Physics **63**, 091705 (2022); <https://doi.org/10.1063/5.0071215>

[Quantum Berezinian for a strange Lie superalgebra](#)

Journal of Mathematical Physics **63**, 081702 (2022); <https://doi.org/10.1063/5.0102653>

Journal of  
Mathematical Physics

Young Researcher Award

Recognizing the outstanding work of early career researchers

LEARN  
MORE >>>

# Stochastic quantization of Yang–Mills

Cite as: J. Math. Phys. 63, 091101 (2022); doi: 10.1063/5.0089431

Submitted: 25 February 2022 • Accepted: 11 August 2022 •

Published Online: 12 September 2022



Ilya Chevyrev<sup>a)</sup>

## AFFILIATIONS

School of Mathematics, The University of Edinburgh, Edinburgh, United Kingdom

**Note:** This paper is part of the Special Collection: XX International Congress on Mathematical Physics.

<sup>a)</sup> Author to whom correspondence should be addressed: [ichevyrev@gmail.com](mailto:ichevyrev@gmail.com)

## ABSTRACT

We review two works [Chandra *et al.*, Publ. Math. l’IHÉS (published online, 2022) and Chandra *et al.*, [arXiv:2201.03487](https://arxiv.org/abs/2201.03487) (2022)] that study the stochastic quantization equations of Yang–Mills on two- and three-dimensional Euclidean space with finite volume. The main result of these works is that one can renormalize the 2D and 3D stochastic Yang–Mills heat flow so that the dynamic becomes gauge covariant in law. Furthermore, there is a state space of distributional 1-forms  $\mathcal{S}$  to which gauge equivalence approximately extends and such that the renormalized stochastic Yang–Mills heat flow projects to a Markov process on the quotient space of gauge orbits  $\mathcal{S}/\sim$ . In this Review, we give unified statements of the main results of these works, highlight differences in the methods, and point out a number of open problems.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0089431>

## I. INTRODUCTION

### A. Yang–Mills theory

Yang–Mills (YM) theory plays an important role in the description of force-carrying particles in the standard model. An important unsolved problem in mathematics is to show that YM theory on Minkowski space-time can be rigorously quantized. We refer to Ref. 42 for a description of this problem, together with the surveys<sup>16,48</sup> for related literature and problems.

A basic ingredient in YM theory is a compact Lie group  $G$ , called the *structure group*, with Lie algebra  $\mathfrak{g}$ . Passing to Euclidean space and working in an arbitrary dimension  $d$ , one can reformulate the problem as trying to make sense of the YM probability measure on the space of  $\mathfrak{g}$ -valued 1-forms  $A = (A_1, \dots, A_d) : \mathbf{R}^d \rightarrow \mathfrak{g}^d$ ,

$$\mu(\mathcal{D}A) = \mathcal{Z}^{-1} \exp(-S(A)) \mathcal{D}A,$$

where  $\mathcal{D}A$  on the right-hand side is a formal Lebesgue measure,  $\mathcal{Z}$  is a normalization constant, and  $S(A)$  is the *YM action*,

$$S(A) = \int_{\mathbf{R}^d} |F(A)|^2.$$

Here,  $F(A)$  is the curvature 2-form of  $A$  given by  $F_{ij}(A) = \partial_i A_j - \partial_j A_i + [A_i, A_j]$ , and we equip  $\mathfrak{g}$  with an Ad-invariant inner product  $\langle \cdot, \cdot \rangle$  and norm  $|\cdot|$ .

*Remark 1.1.* Without loss of generality, we can take  $G \subset U(N)$  and  $\mathfrak{g} \subset \mathfrak{u}(N)$  for some  $N \geq 1$ . In this case, the adjoint action is  $\text{Ad}_g A = gAg^{-1}$  and a possible choice for the inner product is  $\langle X, Y \rangle = -\text{Tr}(XY)$ , and one can rewrite  $|F(A)|^2 = -\frac{1}{2} \sum_{i,j=1}^d \text{Tr}(F_{ij}(A)^2)$ .

The case  $d = 4$  corresponds to physical space-time, but the task of constructing the probability measure  $\mu$  makes sense for arbitrary dimension and even for a Riemannian manifold  $M$  in place of  $\mathbf{R}^d$ . The  $\mathfrak{g}$ -valued 1-forms in this case become connections on a principal  $G$ -bundle  $P \rightarrow M$ , and one aims to define the probability measure  $\mu$  on the space of connections on  $P$ .

The cases  $d = 2, 3$  are considered substantially simpler than  $d = 4$  as they correspond to super-renormalizable theories in quantum field theory (vs renormalizable for  $d = 4$  and non-renormalizable for  $d \geq 5$ ). In the remainder of this article, we will focus on these dimensions and further restrict to *finite volume*, replacing  $\mathbf{R}^d$  by the torus  $\mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$ . The underlying principal bundle  $P$  is always assumed trivial and we keep in mind that all geometric objects (connections, curvature forms, etc.) can be written in coordinates ( $\mathfrak{g}$ -valued 1-forms, 2-forms, etc.) once we fix a global section of  $P$  that identifies it with  $\mathbf{T}^d \times G$ . The space of connections is an affine space, with the difference of two connections being a 1-form.

An important postulate in the physics of YM theory (and gauge theories, in general) is that all physical quantities should be invariant under the action of the *gauge group*, i.e., the automorphism group of  $P$ . Since  $P$  is trivial, the gauge group consists of functions  $g : \mathbf{T}^d \rightarrow G$ . These gauge transformations change the global section we use to identify  $P \simeq \mathbf{T}^d \times G$  and act on 1-forms by

$$A \mapsto A^g \stackrel{\text{def}}{=} \text{Ad}_g A - (dg)g^{-1}.$$

One can view  $A^g$  either as a new connection, gauge equivalent to  $A$ , or simply as the same connection  $A$  written in a new coordinate system.

We write  $A \sim B$  if there exists  $g$  such that  $A^g = B$  and write  $[A] = \{B : B \sim A\}$  for the *gauge orbit* of  $A$ . In light of the above, the natural space on which to define the probability measure  $\mu$  is not the space of 1-forms, but rather the quotient space  $\mathcal{O}$  of all gauge orbits. The space  $\mathcal{O}$  is a non-linear space if  $G$  is non-Abelian, which makes non-trivial even the construction of the state space on which the YM measure  $\mu$  should be defined.

A number of works have made contributions to a precise definition of this measure. The most successful case is dimension  $d = 2$ , which includes  $\mathbf{R}^2$  and compact orientable surfaces. The key feature that makes 2D YM special is its exact solvability, which allows one to write down an explicit formula for the joint distribution of Wilson loop observables; this property was observed in the physics literature by Migdal<sup>53</sup> and later developed in mathematics; see, e.g., Refs. 17, 26, 29, 33, 44, 45, and 60.

In the Abelian case  $G = U(1)$  on  $\mathbf{R}^d$ , one can make sense of the measure  $\mu$  for  $d = 3$ <sup>32</sup> and  $d = 4$ .<sup>25</sup> For any structure group  $G$ , a form of ultraviolet stability on  $\mathbf{T}^d$  was demonstrated for  $d = 4$  using a continuum regularization in Ref. 51 and for  $d = 3, 4$  using a renormalization group approach on the lattice in a series of works by Balaban<sup>2-4</sup> (see also Ref. 28). However, a construction of the 3D YM measure and a description of its gauge-invariant observables, even on  $\mathbf{T}^3$ , remains open.

## B. Stochastic quantization

Another approach to the construction of the YM measure was recently initiated in Refs. 11 and 12, which is based on stochastic quantization (see also Ref. 61 that treats scalar QED on  $\mathbf{T}^2$ ). The basic idea behind this approach is to view the measure  $\mu$  as the invariant measure of a Langevin dynamic. By studying this dynamic, one can try to determine properties and even give constructions of  $\mu$ . The method was put forward in the context of gauge theories by Parisi–Wu<sup>37</sup> and has recently been applied to the construction of *scalar* theories (see Refs. 1, 34, 40, and 54).

The Langevin dynamic associated with the YM measure is

$$\partial_t A = -d_A^* F(A) + \xi, \tag{1.1}$$

where  $d_A^*$  is the adjoint of the covariant derivative  $d_A$  and  $\xi$  is a white noise built over  $L^2$ -space of  $\mathfrak{g}$ -valued 1-forms.

We point out right away that a difficulty in solving (1.1), even in the absence of noise, is that the equation  $\partial_t A = -d_A^* F(A)$  is not parabolic. This is a well-known feature of YM theory and is connected with the infinite-dimensional nature of the gauge group: the YM equations  $d_A^* F(A) = 0$  are not elliptic and admit infinitely many solutions because if  $A$  is a solution, then so is every element of the gauge orbit  $[A]$ .

Fortunately, it is possible to bypass this issue by adding a gauge-breaking term to the right-hand side of (1.1), which is tangent to the orbit  $[A]$  at  $A$  and which renders the equation parabolic. This “trick” was used by Zwanziger<sup>64</sup> and Donaldson<sup>24</sup> in YM theory (see also Refs. 5 and 59) and by DeTurck<sup>23</sup> in the context of Ricci flow. The tangent space of  $[A]$  at  $A$  is precisely the set of all  $d_A \omega = \sum_{i=1}^d (\partial_i + [A_i, \cdot]) \omega dx_i$ , where  $\omega$  is a  $\mathfrak{g}$ -valued function. A natural choice for the additional term is  $-d_A d^* A$ , where  $d^* A = -\sum_{i=1}^d \partial_i A_i$ . Therefore, making sense of (1.1) is equivalent to making sense of the parabolic equation

$$\partial_t A = -d_A^* F(A) - d_A d^* A + \xi, \tag{1.2}$$

which is the equation we consider henceforth. In coordinates, (1.2) reads

$$\partial_t A_i = \Delta A_i + [A_j, 2\partial_j A_i - \partial_i A_j + [A_j, A_i]] + \xi_i, \quad i = 1, \dots, d, \tag{1.3}$$

where  $\xi_1, \dots, \xi_d$  are independent and identically distributed (i.i.d.)  $\mathfrak{g}$ -valued white noises and  $[\cdot, \cdot]$  is the Lie bracket of  $\mathfrak{g}$ . Here and below, there is an implicit summation over  $j = 1, \dots, d$ .

### 1. Gauge covariance

The reason why (1.3) is natural is that it is (formally) gauge covariant in law. To see this, it is convenient to work in coordinate-free notation and, for the moment, make a distinction between connections and 1-forms. Recall that the space of connections is affine modeled on the space  $\Omega^1$  of  $\mathfrak{g}$ -valued 1-forms [we are using the global section of  $P$  to identify  $\text{ad}(P)$ -valued forms with  $\mathfrak{g}$ -valued forms]. For a connection  $A$ , recall further that  $d_A : \Omega^k \rightarrow \Omega^{k+1}$  is a linear map from  $\mathfrak{g}$ -valued  $k$ -forms to  $(k+1)$ -forms with adjoint  $d_A^* : \Omega^{k+1} \rightarrow \Omega^k$ . Furthermore, gauge transformations  $g$  act on connections via  $A \mapsto A^g$  and on forms via  $\omega \mapsto \text{Ad}_g \omega$ .

Consider now a time-dependent connection  $A$  and gauge transformation  $g$ . Then,  $B \stackrel{\text{def}}{=} A^g$  satisfies

$$\partial_t B = \text{Ad}_g(\partial_t A) - d_B[(\partial_t g)g^{-1}].$$

Let  $Z$  denote the canonical flat connection associated with our choice of global section, i.e., the connection associated with the 1-form 0. Consider a time-dependent 1-form  $\xi$  and suppose that  $A$  solves (1.3), which we now write as

$$\partial_t A = -d_A^* F(A) - d_A d_A^*(A - Z) + \xi.$$

Observe the covariance properties

$$\text{Ad}_g(A - Z) = A^g - Z^g, \quad \text{Ad}_g(d_A \omega) = d_{A^g}(\text{Ad}_g \omega), \quad \text{Ad}_g[F(A)] = F(A^g)$$

(the whole reason for the above notation is that these properties take on a simple form). It follows that  $B$  solves

$$\partial_t B = -d_B^* F(B) - d_B d_B^*(B - Z^g) + \text{Ad}_g \xi - d_B[(\partial_t g)g^{-1}].$$

To bring  $B$  into a form similar to  $A$ , it is natural to take  $g$  that solves

$$(\partial_t g)g^{-1} = d_B^*(Z^g - Z).$$

In this case, the equation for  $B$  becomes

$$\partial_t B = -d_B^* F(B) - d_B d_B^*(B - Z) + \text{Ad}_g \xi,$$

which is almost the same as the equation for  $A$ , except that  $\xi$  is replaced by  $\text{Ad}_g \xi$ .

If we now assume  $\xi$  is a white noise and that the equations above make sense with the global section in time solutions, then  $\text{Ad}_g \xi$  is equal in law to  $\xi$  by Itô isometry. This formal argument suggests that there is a coupling between two solutions  $A, B$  to (1.3) starting from gauge equivalent initial conditions such that  $A(t) \sim B(t)$  for all  $t \geq 0$ . In particular, the law of the projected process  $[A]$  on gauge orbits is equal to that of  $[B]$ . The projected process on gauge orbits is therefore well-defined and Markov, and its invariant probability measure is a natural candidate for the YM measure.

### C. Main results

The basic objective of Refs. 11 and 12 is to make rigorous the above formal argument in the case of  $\mathbf{T}^d$  for  $d = 2$  and  $d = 3$ , respectively. In particular, they aim to define a natural Markov process on gauge orbits associated with the YM Langevin dynamic (1.3). In this subsection, we give unified statements of the main results in Refs. 11 and 12. We will go into more detail and discuss the differences in proofs in Secs. II and III. We also discuss in detail these results in the simple case that  $G = U(1)$  in Sec. I D.

*Remark 1.2.* In Ref. 12, the more general Yang–Mills–Higgs theory is considered. To simplify our discussion, we restrict here the results of Ref. 12 to pure YM theory.

One of the main difficulties in solving (1.3) is that, unless  $G$  is Abelian, this equation is non-linear and singular. To simplify notation and highlight the nature of the non-linearities, we will henceforth write (1.2) and (1.3) as

$$\partial_t A = \Delta A + A \partial A + A^3 + \xi, \tag{1.4}$$

It is well-known that white noise on  $\mathbf{R} \times \mathbf{T}^d$  can be realized as a random distribution in  $C^{-1-\frac{d}{2}-\kappa}$ , the Hölder–Besov space with parabolic scaling, for arbitrary  $\kappa > 0$ . Furthermore, this regularity is optimal at least in the scale of Besov spaces. By Schauder estimates, we, therefore,

expect that  $A$  is in  $C^{1-\frac{d}{2}-\kappa}$  (and no better), but this renders the non-linear terms  $A\partial A$  and  $A^3$  analytically ill-defined once  $d \geq 2$  because  $1 - \frac{d}{2} - \kappa < 0$ . [We recall here that the bilinear map  $(f, g) \mapsto fg$  defined for smooth functions extends to  $C^\alpha \times C^\beta$  if and only if  $\alpha + \beta > 0$ .]

A number of solution theories, including regularity structures<sup>36</sup> and paracontrolled calculus<sup>35</sup> (see also Refs. 27, 43, and 56), have been developed in the last decade, which allow one to make sense of such equations. The key condition that must be satisfied is *subcriticality*, which happens if and only if  $d < 4$  and which parallels the notion of super-renormalizability in quantum field theory (QFT). Subcriticality implies that the solution  $A$  to (1.4) should be a perturbation of the stochastic heat equation (SHE),

$$\partial_t \Psi = \Delta \Psi + \xi,$$

which has the Gaussian free field (GFF) as an invariant measure. The solution theory that Refs. 11 and 12 use and build on is the theory of regularity structures.

Recall that  $A$  represents a geometric object (a principal  $G$ -connection). On the other hand, the solution  $A$  to (1.4) at positive times is expected to be a distribution of the same regularity as the GFF. Therefore, a first natural question is whether there exists a state space  $\mathcal{S}$  large enough to support the GFF while small enough so that gauge equivalence extends to  $\mathcal{S}$ . One of the main results of Refs. 11 and 12 gives an answer to this question, which can be informally stated as follows.

**Theorem 1.3.** *There exists a metric space  $(\mathcal{S}, \Sigma)$  of  $\mathfrak{g}$ -valued distributional 1-forms on  $\mathbf{T}^d$ ,  $d = 2, 3$ , which contains all smooth 1-forms and to which gauge equivalence approximately extends in a canonical way. Furthermore,  $\mathcal{S}$  contains distributions of the same regularity as the GFF on  $\mathbf{T}^d$ .*

The constructions of  $\mathcal{S}$  in Refs. 11 and 12 are rather different. In Ref. 11,  $\mathcal{S}$  (therein denoted by  $\Omega_\alpha^1$ ) is a Banach space defined through line integrals, and gauge equivalence is determined by an action of a gauge group. In Ref. 12,  $\mathcal{S}$  is a non-linear metric space of distributions defined in terms of the effect of the heat flow; gauge equivalence is extended using a gauge-covariant regularizing operator (the deterministic YM flow). We describe these constructions further in Secs. II A and III A.

*Remark 1.4.* A naive construction of  $\mathcal{S}$  would be to take  $C^\eta$  for  $\eta = 1 - \frac{d}{2} - \kappa$  and quotient by the action of the gauge group  $C^{\eta+1}(\mathbf{T}^d, G) \cap C^{-\eta+\kappa}(\mathbf{T}^d, G)$  so that  $\text{Ad}_g A \in C^\eta$  and  $(dg)g^{-1} \in C^\eta$  are well-defined. Such a construction ends up lacking most of the nice properties we discuss in Secs. II A and III A.

*Remark 1.5.* While Theorem 1.3 makes it seem like the 2D and 3D cases are on an equal footing, we actually know much more about  $\mathcal{S}$  in 2D than in 3D, e.g., the space of orbits  $\mathcal{S}/\sim$  in 2D comes with a natural complete metric, and is thus Polish, while we only know that  $\mathcal{S}/\sim$  is completely Hausdorff (and separable) in 3D.

We now turn to the question of solving (1.4). A natural approach is to replace  $\xi$  by a smooth approximation  $\xi^\varepsilon$  and let the mollification parameter  $\varepsilon \downarrow 0$ . The hope then is that the corresponding solutions  $A$  converge. Unfortunately, this is not, in general, the case and the stochastic partial differential equation (SPDE) requires *renormalization* for convergence to take place. The following result ensures that renormalized solutions to (1.4) exist, at least up to a potential finite time blow-up.

A *mollifier* is a smooth compactly supported function  $\chi: \mathbf{R} \times \mathbf{R}^d \rightarrow \mathbf{R}$  such that  $\int \chi = 1$  and which is spatially symmetric and invariant under flipping coordinates  $x_i \mapsto -x_i$  for  $i = 1, \dots, d$ . Denoting  $\chi^\varepsilon(t, x) = \varepsilon^{-2-d} \chi(\varepsilon^{-2}t, \varepsilon^{-1}x)$  and  $*$  for space-time convolution, we define  $\xi^\varepsilon = \chi^\varepsilon * \xi$ . Furthermore, define

$$L_G(\mathfrak{g}, \mathfrak{g}) = \{X \in L(\mathfrak{g}, \mathfrak{g}) : X \text{Ad}_g = \text{Ad}_g X \text{ for all } g \in G\}.$$

**Theorem 1.6.** *For every mollifier  $\chi$ , there exists a family of operators  $\{C_{\text{bphz}}^\varepsilon\}_{\varepsilon \in (0,1)} \subset L_G(\mathfrak{g}, \mathfrak{g})$  such that, for any  $\mathring{C} \in L(\mathfrak{g}, \mathfrak{g})$  and initial condition  $A(0) \in \mathcal{S}$ , the solution to the partial differential equation (PDE),*

$$\partial_t A = \Delta A + A\partial A + A^3 + \xi^\varepsilon + (C_{\text{bphz}}^\varepsilon + \mathring{C})A, \tag{1.5}$$

*converges in probability in  $C(\mathbf{R}_+, \mathcal{S} \sqcup \{\infty\})$  as  $\varepsilon \downarrow 0$ . The limit as  $\varepsilon \downarrow 0$  furthermore does not depend on  $\chi$ .*

In coordinates, (1.5) reads

$$\partial_t A_i = \Delta A_i + [A_j, 2\partial_j A_i - \partial_i A_j + [A_j, A_i]] + \xi_i^\varepsilon + (C_{\text{bphz}}^\varepsilon + \mathring{C})A_i, \quad i = 1, \dots, d.$$

Remark that the operator  $C_{\text{bphz}}^\varepsilon + \mathring{C} \in L(\mathfrak{g}, \mathfrak{g})$  in (1.5) is “block diagonal” meaning that only  $(C_{\text{bphz}}^\varepsilon + \mathring{C})A_i$  appears in the equation for  $A_i$ .

*Definition 1.7.* We call the  $\varepsilon \downarrow 0$  limit of  $A$  as in Theorem 1.6 the *solution to (1.5) driven by  $\xi$  with bare mass  $\hat{C}$* .

The bare mass  $\hat{C}$  is used to parameterize the space of all “reasonable” solutions and is a free parameter at this stage. We will see below (Theorems 1.11 and 1.14) that there does exist a unique choice for  $\hat{C}$ , which selects a distinguished element of this solution space.

*Remark 1.8.* The operators  $C_{\text{bphz}}^\varepsilon$  are called the *Bogoliubov-Parasiuk-Hepp-Zimmermann (BPHZ) constants* and are given by (in principle, explicit) integrals involving  $\chi$  and an arbitrary large scale truncation  $K$  of the heat kernel. While the solution to (1.5) in Definition 1.7 is independent of  $\chi$ , it does, in general, depend on the choice of  $K$  used to define  $C_{\text{bphz}}^\varepsilon$ .

*Remark 1.9.* The point  $\mathfrak{Q}$  is a cemetery state and is added to  $\mathcal{S}$  to handle the possibility of finite time blow-up. Some care is needed to properly define  $C(\mathbf{R}_+, \mathcal{S} \sqcup \{\mathfrak{Q}\})$  and the metric that one equips it with. This is done in Ref. 11 (Sec. 1.5.1), where, for a general metric space  $E$ , a metric space  $E^{\text{sol}}$  of continuous paths with values in  $E \sqcup \{\mathfrak{Q}\}$  is defined in which two paths are close if they track each other until the point when they become large. Our notation  $C(\mathbf{R}_+, \mathcal{S} \sqcup \{\mathfrak{Q}\})$  here really means  $\mathcal{S}^{\text{sol}}$ .

*Remark 1.10.* It turns out that in 2D, due to a cancellation in renormalization constants,  $C_{\text{bphz}}^\varepsilon$  converges to a finite value as  $\varepsilon \downarrow 0$  (see Theorem 2.5). Therefore, Theorem 1.6 in 2D remains true if  $C_{\text{bphz}}^\varepsilon + \hat{C}$  replaced by any fixed  $C \in L(\mathfrak{g}, \mathfrak{g})$ , which is the formulation of Ref. 11 (Theorem 2.4). No such cancellation occurs in 3D and  $C_{\text{bphz}}^\varepsilon$  diverges at order  $\varepsilon^{-1}$ .

We now discuss the way in which solutions to (1.5) are gauge covariant in the sense described in Sec. 1 B 1. Remark that by inserting the counterterm  $(C_{\text{bphz}}^\varepsilon + \hat{C})A$ , we are seemingly breaking the desired gauge covariance property discussed in Sec. 1 B 1 [in the notation of that section,  $(C_{\text{bphz}}^\varepsilon + \hat{C})A$  should be written  $(C_{\text{bphz}}^\varepsilon + \hat{C})(A - Z)$ ]. However, the formal argument in Sec. 1 B 1 also breaks if we replace  $\xi$  by  $\xi^\varepsilon$  because Itô isometry is not true for the latter.

A surprising fact is that if one chooses  $\hat{C}$  carefully, then the broken gauge covariance [due to the counterterm  $(C_{\text{bphz}}^\varepsilon + \hat{C})A$ ] compensates in the  $\varepsilon \downarrow 0$  limit for the broken Itô isometry (due to the mollified noise  $\xi^\varepsilon$ ), and one obtains a solution to (1.5), which is gauge covariant in law.

It is not entirely trivial to make this statement precise, essentially because we do not know if (1.5) (with any bare mass) is global in time solutions. In particular, we do not know how to rule out that solutions to (1.5) with different gauge equivalent initial conditions  $a \sim b$  blow up at different times, and this makes it unclear in what sense we can expect the projected process  $[A]$  on gauge orbits to be Markov.

To address this issue, it is natural to look for a type of process that solves (1.5) on disjoint intervals  $[\zeta_{j-1}, \zeta_j]$  and at time  $\zeta_j$  jumps to a new representative of the gauge orbit  $[\lim_{t \uparrow \zeta_j} A(t)]$ . This should happen in such a way that  $A$  does not blow up unless the entire orbit  $[A]$  “blows up.” This class of processes is defined through *generative probability measures* in Refs. 11 and 12.

We say that a probability measure  $\mu$  on the space of càdlàg functions  $D(\mathbf{R}_+, \mathcal{S} \sqcup \{\mathfrak{Q}\})$  is *generative* with bare mass  $\hat{C}$  and initial condition  $a \in \mathcal{S}$  if there exists a white noise  $\xi$  and a random variable  $A$  with law  $\mu$  such that

- (i)  $A(0) = a$  almost surely;
- (ii) there exists a sequence of stopping times  $\zeta_0 = 0 \leq \zeta_1 \leq \zeta_2 < \dots$  such that  $A$  solves (1.5) driven by  $\xi$  with bare mass  $\hat{C}$  on each interval  $[\zeta_j, \zeta_{j+1})$ ;
- (iii) for every  $j \geq 0$ ,  $A(\zeta_{j+1}) \sim \lim_{t \uparrow \zeta_{j+1}} A(t)$ ; and
- (iv)  $\lim_{j \rightarrow \infty} \zeta_j = T^* \stackrel{\text{def}}{=} \inf\{t \geq 0 : A(t) = \mathfrak{Q}\}$  and, on the event  $T^* < \infty$ ,<sup>65</sup>

$$\lim_{t \uparrow T^*} \inf_{B \sim A(t)} \Sigma(B, 0) = \infty. \tag{1.6}$$

The point of this definition is to give a sufficiently general and natural way in which (1.5) can be restarted along gauge orbits. The following result from Refs. 11 and 12 ensures the existence of a canonical Markov process associated with (1.5) on the quotient space of gauge orbits  $\mathcal{O} \stackrel{\text{def}}{=} \mathcal{S} / \sim$ , provided the bare mass is chosen in a precise way.

**Theorem 1.11.**

- (a) For every  $a \in \mathcal{S}$  and  $\hat{C} \in L(\mathfrak{g}, \mathfrak{g})$ , there exists a generative probability measure  $\mu$  with bare mass  $\hat{C}$  and initial condition  $a$ .
- (b) There exists  $\check{C} \in L_G(\mathfrak{g}, \mathfrak{g})$  with the following properties. For all  $a \sim b \in \mathcal{S}$ , if  $\mu, \nu$  are generative probability measures with initial conditions  $a, b$ , respectively, and bare mass  $\check{C}$ , then the pushforward measures  $\pi_* \mu$  and  $\pi_* \nu$  are equal. In particular, the probability measure  $\mathbf{P}^x = \pi_* \mu$ , where  $\mu$  is generative with bare mass  $\check{C}$  and initial condition  $a \in x \in \mathcal{O}$ , depends only on  $x$ . Finally,  $\{\mathbf{P}^x\}_{x \in \mathcal{O}}$  are the transition functions of a time homogeneous, continuous Markov process on  $\mathcal{O} \sqcup \{\mathfrak{Q}\}$ .

*Remark 1.12.* Theorem 1.11(b) makes no claims about the *uniqueness* of  $\check{C}$ , but we conjecture that  $\check{C}$  is indeed unique (which is not difficult to prove in the Abelian case; see Sec. 1 D).

Finally, we mention a result in Refs. 11 and 12 that is crucial for the Proof of Theorem 1.11(b) and that makes precise the coupling argument outlined in Sec. I B 1. This result makes a stronger statement about the constant  $\check{C}$  for which uniqueness does hold. It can also be seen as a version of the Slavnov–Taylor identities for renormalization schemes that preserve gauge symmetries.

Suppose that  $A$  solves (1.5) and, recalling that  $\Delta A + A\partial A + A^3 + CA$  is shorthand for  $-d_A^*F(A) - d_A d_A^*(A - Z) + C(A - Z)$ , suppose that  $(B, g)$  solves

$$\begin{aligned} \partial_t B &= -d_B^*F(B) - d_B d_B^*(B - Z) + \text{Ad}_g \xi^\varepsilon + (C_{\text{bphz}}^\varepsilon + \check{C})(B - Z^g), \\ (\partial_t g)g^{-1} &= d_B^*(Z^g - Z), \end{aligned} \tag{1.7}$$

where the initial condition of  $B$  is  $B(0) = A(0)^{g(0)}$ . Then, the same computation as in Sec. I B 1 (see also Ref. 11, Sec. 2.2) shows that  $A^g = B$ . In coordinates, (1.7) is written as

$$\begin{aligned} \partial_t B &= \Delta B + B\partial B + B^3 + \text{Ad}_g \xi^\varepsilon + (C_{\text{bphz}}^\varepsilon + \check{C})(B + (dg)g^{-1}), \\ (\partial_t g)g^{-1} &= \partial_j((\partial_j g)g^{-1}) + [B_j, (\partial_j g)g^{-1}]. \end{aligned} \tag{1.8}$$

It is now natural to compare (1.8) to

$$\begin{aligned} \partial_t \bar{A} &= \Delta \bar{A} + \bar{A}\partial \bar{A} + \bar{A}^3 + \chi^\varepsilon * (\text{Ad}_{\bar{g}} \xi) + (C_{\text{bphz}}^\varepsilon + \check{C})\bar{A}, \\ (\partial_t \bar{g})\bar{g}^{-1} &= \partial_j((\partial_j \bar{g})\bar{g}^{-1}) + [\bar{A}_j, (\partial_j \bar{g})\bar{g}^{-1}], \\ \bar{g}(0) &= g(0), \quad \bar{A}(0) = B(0). \end{aligned} \tag{1.9}$$

Remark that  $\chi^\varepsilon * (\text{Ad}_{\bar{g}} \xi)$  is equal in law to  $\xi^\varepsilon$ ; thus,  $\bar{A}$  is equal in law to the solution of (1.5) with initial condition  $B(0)$ , provided we take  $\chi$  non-anticipative in the following sense.

*Definition 1.13.* A mollifier  $\chi$  is called *non-anticipative* if it has support in  $(0, \infty) \times \mathbf{R}^d$ .

What we would therefore like to show is that, for a special choice of  $\check{C}$ , the solutions to (1.8) and (1.9) converge as  $\varepsilon \downarrow 0$  to the same limit. The identity  $A^g = B$ , which survives in the limit, would provide a coupling between (1.5) with initial condition  $A(0)$  and initial condition  $A(0)^{g(0)}$  under which the two solutions are gauge equivalent, at least locally in time. It turns out that the following more general result is true.

**Theorem 1.14.** *There exists a unique  $\check{C} \in L_G(\mathfrak{g}, \mathfrak{g})$  such that for all non-anticipative mollifiers  $\chi$ , all  $\check{C} \in L(\mathfrak{g}, \mathfrak{g})$ , and all initial conditions  $(B(0), g(0)) \in \mathcal{S} \times C^\rho(\mathbf{T}^d, G)$ ,  $\rho \in (\frac{1}{2}, 1)$ , the solution  $(B, g)$  to (1.8) converges as  $\varepsilon \downarrow 0$  in probability to the same limit as the solution to*

$$\begin{aligned} \partial_t \bar{A} &= \Delta \bar{A} + \bar{A}\partial \bar{A} + \bar{A}^3 + \chi^\varepsilon * (\text{Ad}_{\bar{g}} \xi) + (C_{\text{bphz}}^\varepsilon + \check{C})\bar{A} + (\check{C} - \check{C})(d\bar{g})\bar{g}^{-1}, \\ (\partial_t \bar{g})\bar{g}^{-1} &= \partial_j((\partial_j \bar{g})\bar{g}^{-1}) + [\bar{A}_j, (\partial_j \bar{g})\bar{g}^{-1}], \\ \bar{g}(0) &= g(0), \quad \bar{A}(0) = B(0). \end{aligned} \tag{1.10}$$

Furthermore, the solution to (1.5) with bare mass  $\check{C}$  is independent of  $\chi$  and of the choice of  $C_{\text{bphz}}^\varepsilon$ .

It follows from Theorem 1.14 that the solutions to (1.8) and (1.9) indeed converge to the same limit as  $\varepsilon \downarrow 0$ , provided we choose  $\check{C} = \check{C}$ . The operator  $\check{C}$  in Theorem 1.14 is exactly the operator appearing in Theorem 1.11(b); in 2D, we can give an explicit expression for  $\check{C}$  [see (2.13)].

*Remark 1.15.* The value of  $\check{C}$  in Theorem 1.14 is determined uniquely after we fix a choice for  $C_{\text{bphz}}^\varepsilon$ . However, recall from Remark 1.8 that  $C_{\text{bphz}}^\varepsilon$  is not unique or canonical—it is determined by  $\chi$  and an arbitrary truncation of the heat kernel (see, e.g., Theorem 2.5). The final part of Theorem 1.14 states that the solution of (1.5) with bare mass  $\check{C}$  is independent of  $\chi$  and this choice of truncation.

## E. Abelian case

We end this section by discussing the above results in the Abelian case, i.e.,  $G = U(1)$  and  $\mathfrak{g} = \mathbf{R}$ . We consider here  $d \geq 1$  arbitrary. We will see in this case that

- the constant in Theorem 1.11(b) is  $\check{C} = 0$  and is *unique*,
- if  $\mathbf{T}^d$  is replaced by  $\mathbf{R}^d$ , then uniqueness of  $\check{C}$  fails (Remark 1.16),
- (1.5) with bare mass  $\check{C} = 0$  is global in time solutions but no invariant probability measure (Remark 1.17)

In the Abelian case, the non-linearities  $A\partial A$  and  $A^3$  as well as the constants  $C_{\text{bphz}}^{\epsilon}$  vanish. Equation (1.5) with bare mass  $\check{C}$ , therefore, becomes linear in  $A$  and converges as  $\epsilon \downarrow 0$  to the solution of the SHE with a mass term

$$\partial_t A = \Delta A + \check{C}A + \xi, \tag{1.11}$$

Treating 1-forms on  $\mathbf{T}^d$  as periodic functions (or distributions) on  $\mathbf{R}^d$  modulo  $\mathbf{Z}^d$ , it is easy to see that  $A \sim B$  if and only if  $A = B + d\omega$  for some  $\omega : \mathbf{R}^d \rightarrow \mathbf{R}$  such that  $e^{i\omega}$  is periodic, where  $i = \sqrt{-1}$ . The tangent space of every gauge orbit is therefore  $\{d\omega : \omega : \mathbf{T}^d \rightarrow i\mathbf{R}\}$ .

Since  $\text{Ad}_g$  is now the identity, it is clear that a possible value for  $\check{C}$  in Theorems 1.11 and 1.14 is  $\check{C} = 0$ . This is because if  $B(0) = A(0) + d\omega(0)$  and  $A, B$  solve (1.11) with  $\check{C} = 0$ , then  $B = A + d\omega$  for all times, where  $\omega$  solves  $\partial_t \omega = \Delta \omega$ .

Furthermore,  $\check{C} = 0$  is the only possible value for  $\check{C}$ . Indeed, consider the two gauge equivalent initial conditions  $A(0) \stackrel{\text{def}}{=} 0$  and  $B(0) = (B_1(0), \dots, B_d(0)) \stackrel{\text{def}}{=} (2\pi, 0, \dots, 0)$ , and suppose that  $A, B$  solve (1.11) with  $\check{C} \neq 0$ . Then,  $B(t) = e^{t\check{C}} B(0) + A(t)$ , where we used that  $B(0)$  is constant on  $\mathbf{T}^d$ . Consider now the gauge-invariant observable  $I[A] \stackrel{\text{def}}{=} e^{i\int_{\mathbf{T}^d} A_1}$ . Then,  $\int_{\mathbf{T}^d} B_1(t) = e^{t\check{C}} 2\pi + \int_{\mathbf{T}^d} A_1(t)$ , and thus,

$$\text{EI}[A(t)] = \exp(i e^{t\check{C}} 2\pi) \text{EI}[B(t)] \neq \text{EI}[B(t)]$$

for all  $t > 0$  sufficiently small since  $\check{C} \neq 0$ . This shows that  $A$  and  $B$  cannot be gauge equivalent in law; thus,  $\check{C} = 0$  is the only possible value in Theorem 1.11(b).

*Remark 1.16.* The above argument relies on the fact that  $\mathbf{T}^d$  is not simply connected: we exploited that  $(dg)g^{-1}$ , which appears in (1.10), is not tangent to gauge orbits, in general (it is tangent if and only if  $g = e^{i\omega}$  for some  $\omega : \mathbf{T}^d \rightarrow \mathbf{R}$ ). If  $\mathbf{T}^d$  is replaced by  $\mathbf{R}^d$ , then  $(dg)g^{-1}$  is tangent to gauge orbits since we can always write  $g = e^{i\omega}$  for some  $\omega : \mathbf{R}^d \rightarrow \mathbf{R}$ . Therefore, on  $\mathbf{R}^d$  in the Abelian case, there is no uniqueness of  $\check{C}$ . Explicitly, working on  $\mathbf{R}^d$ , suppose  $A$  solves (1.11) and consider  $B$  as in (1.8), where  $B(0) = A(0)^{g(0)}$  but now  $g$  satisfies

$$d[(\partial_t g)g^{-1}] = -d(d^*[(dg)g^{-1}]) + \check{C}(dg)g^{-1},$$

i.e.,  $g = e^{i\omega}$ , where  $\omega$  solves  $\partial_t \omega = \Delta \omega + \check{C}\omega$ . Then,  $B(t) = A(t)^{g(t)} = A(t) - d\omega(t)$  for all  $t \geq 0$  and  $B$  solves the same equation (1.11) as  $A$  for any  $\check{C} \in \mathbf{R}$ .

*Remark 1.17.* Clearly (1.11) with  $\check{C} = 0$  does not have an invariant probability measure because  $\int_{\mathbf{T}^d} A_i$  evolves like a Brownian motion. In fact, any gauge equivalent generalization of (1.5) of the form  $\partial_t A = \Delta A + \xi + d\omega$ , where  $\omega$  is adapted, will lack an invariant probability measure because the spatial mean is unaffected by  $d\omega$ . However, we do obtain an invariant probability measure for the projected process  $[A]$  because any 1-form is gauge equivalent to another 1-form  $B$  such that  $\int_{\mathbf{T}^d} B_1, \dots, \int_{\mathbf{T}^d} B_d \in [-\pi, \pi]$ . This remark shows that the Markov process from Theorem 1.11 can have an invariant probability measure, while (1.5) with bare mass  $\check{C}$  does not.

## II. TWO DIMENSIONS

We describe in this section the main results in Ref. 11, i.e., the results of Sec. 1C in 2D.

### A. State space

The definition of the state space  $\mathcal{S}$  (denoted by  $\Omega_\alpha^1$  in Ref. 11) is motivated by the desire to define holonomies and, thus, Wilson loops (for every element  $A \in \mathcal{S}$ ). The construction is a refinement of that introduced in Ref. 17. Let  $\mathcal{X} = \mathbf{T}^2 \times B_{1/4}$ , where  $B_{1/4} = \{v \in \mathbf{R}^2 : |v| \leq \frac{1}{4}\}$ . We think of  $\mathcal{X}$  as the collection of straight line segments  $\ell = (x, v)$  in  $\mathbf{T}^2$  of length at most  $|\ell| \stackrel{\text{def}}{=} |v| < \frac{1}{4}$ . (The starting point of  $\ell$  is  $x$ .)

For  $\alpha \in [0, 1]$  and a smooth 1-form  $A \in C^\infty(\mathbf{T}^2, \mathfrak{g}^2)$ , define the norm

$$|A|_{\alpha\text{-gr}} = \sup_{\ell} \frac{|A(\ell)|}{|\ell|^\alpha},$$

where the supremum is taken over all  $\ell \in \mathcal{X}$  with  $|\ell| > 0$  and where we define the line integral

$$A(\ell) \stackrel{\text{def}}{=} \int_0^1 A(x + vt) v dt = \int_0^1 \sum_{i=1}^2 A_i(x + vt) v_i dt. \tag{2.1}$$

Furthermore, we define the strengthened norm

$$|A|_\alpha = |A|_{\alpha\text{-gr}} + \sup_P \frac{|A(\partial P)|}{|P|^{\alpha/2}},$$

where the supremum is over all oriented triangles  $P = (\ell_1, \ell_2, \ell_3)$  with  $\ell_i \in \mathcal{X}$  and area  $|P| > 0$  and  $A(\partial P) \stackrel{\text{def}}{=} \sum_{i=1}^3 A(\ell_i)$ . We can now define the state space studied in Ref. 11.

*Definition 2.1.* The Banach space  $(\mathcal{S}, |\cdot|_\alpha)$  is defined as the completion of smooth  $\mathfrak{g}$ -valued 1-forms under  $|\cdot|_\alpha$  for some  $\alpha \in (\frac{2}{3}, 1)$ .

The metric  $\Sigma$  in Theorem 1.3 is then the usual metric  $\Sigma(A, B) = |A - B|_\alpha$ .

*Remark 2.2.* To motivate these norms, consider  $A = (A_1, A_2)$  a pair of i.i.d. GFFs. A simple calculation shows that, for all  $\alpha < 1$ ,  $\mathbb{E}|A(\ell)|^2 \lesssim |\ell|^{2\alpha}$ . Furthermore, it follows from the Stokes theorem and the fact that  $dA$  is a white noise and that  $A(\partial P) = \int_P dA$  and, hence,  $\mathbb{E}|A(\partial P)|^2 = |P|$ . A Kolmogorov argument then implies that the GFF has a modification with  $|A|_\alpha < \infty$  almost surely.

To extend gauge equivalence to  $\mathcal{S}$ , let  $\mathcal{G}^{0,\alpha}$  denote the closure of smooth functions in  $C^\alpha(\mathbb{T}^2, G)$ . One can show that the group action  $(A, g) \mapsto A^g$ , defined for smooth 1-forms and gauge transformations, extends to a locally Lipschitz map  $\mathcal{S} \times \mathcal{G}^{0,\alpha} \rightarrow \mathcal{S}$  [see Ref. 11 (Theorem 3.27, Corollary 3.36)]. We then extend gauge equivalence  $\sim$  to  $\mathcal{S}$  by

$$A \sim B \iff A^g = B \text{ for some } g \in \mathcal{G}^{0,\alpha}.$$

With these definitions,  $\mathcal{S}$  has the following desirable properties (see Ref. 11, Theorem 2.1):

- For every  $\ell = (x, v) \in \mathcal{X}$  and  $A \in \mathcal{S}$ , one has  $|\ell_A|_{C^\alpha} \leq |\ell|^\alpha |A|_{\alpha\text{-gr}}$ , where  $\ell_A : [0, 1] \rightarrow \mathfrak{g}$  is the path  $\ell_A(t) = \int_0^t A(x + sv) u ds$ . The holonomy  $\text{hol}(A, \ell) \in G$ , defined by  $\text{hol}(A, \ell) = y_1$ , where  $y$  solves the ordinary differential equation (ODE)

$$dy_t = y_t d\ell_A(t), \quad y_0 = 1,$$

is, therefore, well-defined by Young integration.<sup>31,50,63</sup>

More generally,  $\text{hol}(A, \gamma)$  is well-defined for any  $\gamma \in C^{1,\beta}([0, 1], \mathbb{T}^2)$ , where  $\beta \in (\frac{2}{\alpha} - 1, 1]$ , and the map  $(A, \gamma) \mapsto \text{hol}(A, \gamma)$  is Hölder continuous. In particular, classical Wilson loop observables are well-defined with good stability properties. The relation  $\sim$  can be expressed entirely in terms of holonomies.

*Remark 2.3.* Since  $\text{hol}(A, \gamma)$  is independent of the parameterization of  $\gamma$ , it is natural to also measure the regularity of  $\gamma$  in a parameterization independent way. Such a notion of regularity is introduced in Ref. 11 (Sec. 3.2), which interpolates between  $C^1$  and  $C^2$  (akin to how  $p$ -variation is a parameterization invariant interpolation between  $C^0$  and  $C^1$ ).

- One has the embeddings

$$C^{\alpha/2} \hookrightarrow \mathcal{S} \hookrightarrow \Omega_{\alpha\text{-gr}}^1 \hookrightarrow C^{\alpha-1},$$

where  $\Omega_{\alpha\text{-gr}}^1$  is the completion of smooth functions under  $|\cdot|_\alpha$ . (Only the last of these is non-trivial; see Ref. 17, Proposition 3.21.) These embeddings are furthermore optimal in the sense that  $\alpha/2$  in  $C^{\alpha/2}$  cannot be decreased and  $\alpha - 1$  in  $C^{\alpha-1}$  cannot be increased. Remark also that  $|A|_{1\text{-gr}} \asymp |A|_{L^\infty}$ , while  $\mathcal{S}$ , since  $\alpha < 1$ , contains distributions that cannot be represented by functions, such as the GFF.

- There exists a complete metric  $D$  on the quotient space of gauge orbits  $\mathcal{O} = \mathcal{S} / \sim$ , which induces the quotient topology. To define  $D$ , we first define a new (but topologically equivalent) metric  $k$  on  $\mathcal{S}$  by shrinking the usual metric  $\Sigma(\cdot, \cdot) = |\cdot - \cdot|_\alpha$  in such a way that the diameter of every  $R$ -sphere  $S_R \stackrel{\text{def}}{=} \{A \in \mathcal{S} : |A|_\alpha = R\}$  goes to zero as  $R \rightarrow \infty$ , but the distance between  $S_r$  and  $S_R$  for large  $r \leq R$  is of order  $\frac{R}{r} - 1$ , so, in particular, it goes to  $\infty$  as  $R \rightarrow \infty$ . Then,  $D$  is defined as the Hausdorff distance associated with the metric  $k$  on  $\mathcal{S}$ .

*Remark 2.4.* The space  $\mathcal{S}$  strengthens the definition of a Banach space  $\mathcal{S}^{\text{ax}}$  introduced in Ref. 17;  $\mathcal{S}^{\text{ax}}$  is defined in a similar way but with  $\mathcal{X}$  taken as the set of axis-parallel line segments. The main result of Ref. 17 is that if  $G$  is simply connected, then there exists a (non-unique) probability measure on  $\mathcal{S}^{\text{ax}}$  such that the holonomies along all axis-parallel curves agree in distribution with those of the YM measure on  $\mathbb{T}^2$  defined in Refs. 44 and 60. The proof of this result uses a gauge-fixed lattice approximation, which explains the restriction to axis-parallel lines.

## B. Local solutions

It turns out that in 2D we can sharpen the statement of Theorem 1.6 as follows.

**Theorem 2.5.** For every  $\hat{C} \in L(\mathfrak{g}, \mathfrak{g})$ , mollifier  $\chi$ , and initial condition  $A(0) \in \mathcal{S}$ , the solution to

$$\partial_t A = \Delta A + A \partial A + A^3 + (\lambda C_{\text{sym}}^\varepsilon + \hat{C}) A + \xi^\varepsilon \tag{2.2}$$

converges in probability in  $C(\mathbf{R}_+, \mathcal{S} \sqcup \{\infty\})$  as  $\varepsilon \downarrow 0$ . The constant  $C_{\text{sym}}^\varepsilon$  converges to a finite limit as  $\varepsilon \downarrow 0$  and is defined by

$$C_{\text{sym}}^\varepsilon = 4\hat{C}^\varepsilon - \bar{C}^\varepsilon, \quad \hat{C}^\varepsilon = \int \partial_j K^\varepsilon (\partial_j K * K^\varepsilon), \quad \bar{C}^\varepsilon = \int (K^\varepsilon)^2.$$

Here,  $K : \mathbf{R} \times \mathbf{R}^2 \setminus \{0\} \rightarrow \mathbf{R}$  is any spatially symmetric function invariant under flipping coordinates  $x_i \mapsto -x_i$ ,  $i = 1, 2$ , which vanishes for negative times, has bounded support, and is equal to the heat kernel  $(\partial_t - \Delta)^{-1}$  in a neighborhood of the origin. We write  $K^\varepsilon = \chi^\varepsilon * K$ , and  $\partial_j$  is any spatial derivative,  $j = 1, 2$ . The operator  $\lambda \in L_G(\mathfrak{g}, \mathfrak{g})$  is the Casimir element of  $\mathfrak{g}$  in the adjoint representation.

The  $\varepsilon \downarrow 0$  limit of  $A$ , which solves (1.5) with bare mass  $\hat{C}$ , depends on  $K$  and  $\hat{C}$  but not on  $\chi$ .

**Remark 2.6.** If  $\mathfrak{g}$  is simple (which one can assume without loss of generality; see Ref. 11, Remark 2.8), then  $\lambda < 0$  is just a scalar.

Recall from Remark 1.10 that the convergence of  $C_{\text{bphz}}^\varepsilon \stackrel{\text{def}}{=} \lambda C_{\text{sym}}^\varepsilon$  to a finite limit is special to dimension 2 and is due to a cancellation between the diverging constants  $4\hat{C}^\varepsilon$  and  $\bar{C}^\varepsilon$ .

We briefly describe the ingredients in the Proof of Theorem 2.5, which is based on the theory of regularity structures. We only mention the overall strategy behind this theory and refer to Refs. 14, 31, and 37 for an introduction and more details. To solve an SPDE such as (1.4), one constructs a sufficiently large “regularity structure” and lifts the equation to a space of “modeled distributions” with values in the regularity structure. This construction, first introduced in Ref. 36, is done at a high level of generality in Ref. 8. One then constructs a finite number of stochastic objects from the noise called a “model”—these objects are essentially renormalizations of functions of forms (2.3) and (2.4)—the existence of which follows from Ref. 13. The point of construction is that the products  $A\partial A, A^3$  and convolution with the heat kernel become stable operations on modeled distributions, and one can solve a fixed point problem for the “lifted” equation. Finally, one maps the resulting modeled distribution to a distribution on  $\mathbf{R} \times \mathbf{T}^2$  via the “reconstruction operator” and identifies it with a solution to a classical renormalized PDE, at least for  $\varepsilon > 0$ . This final step is carried out systematically in Ref. 6. All these operations are done in a way that is stable as  $\varepsilon \downarrow 0$ , thereby showing the desired convergence.

One of the contributions of Ref. 11 is to develop a framework in which the algebraic results from Refs. 6 and 8 can be transferred to a setting in which the noise and solution are vector-valued. References 6 and 8 provided a general method to compute the renormalized form of a system of scalar SPDEs, which, in principle, does apply to (1.4) by writing it as a system of  $d \times \dim(\mathfrak{g})$  scalar-valued equations using a basis. However, such a procedure is cumbersome and unnatural; it is more desirable to find a framework that preserves the vector-valued nature of the noise and solution, which is the purpose of Ref. 11 (Sec. 5).

The main idea behind the extension in Ref. 11 is to define a category of “symmetric sets” and a functor between this category and the category of vector spaces. This construction allows one to canonically associate partially symmetrized tensor products of vector spaces to combinatorial rooted trees that commonly appear in regularity structures. One of the main outcomes is a procedure to compute the renormalized form of equations like (1.4) without resorting to a basis.

In working out the renormalized equation (2.2), it is easy to deduce by power-counting and symmetry arguments that the only non-vanishing counterterms arise from the three trees



$$(2.3)$$

Before renormalization, these trees correspond, respectively, to the following three functions:

$$[K * (\partial_j K * \xi)] \cdot (\partial_j K * \xi), \quad (K * \xi) \cdot [\partial_j K * (\partial_j K * \xi)], \quad \text{and} \quad (K * \xi)^2. \quad (2.4)$$

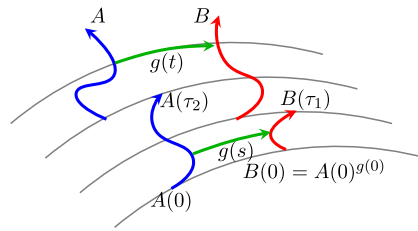
A computation shows that the final counterterm is precisely  $\lambda C_{\text{sym}}^\varepsilon A$  as defined in Theorem 2.5 (see Ref. 11, Lemma 6.14).

It follows from the general theory of regularity structures that (2.2), for any  $\hat{C} \in L(\mathfrak{g}, \mathfrak{g})$ , converges locally in time in  $C^{\alpha-1}$ . To improve this to convergence in  $C(\mathbf{R}_+, \mathcal{S} \sqcup \{\infty\})$ , one decomposes the solution  $A$  into  $A = \Psi + B$ , where  $\Psi$  solves the SHE  $\partial_t \Psi = \Delta \Psi + \xi$  with initial condition  $\Psi(0) = A(0)$  and  $B$  is in  $C^{1-\kappa}$  for all  $\kappa > 0$ . One can then show by hand that  $\Psi \in C(\mathbf{R}_+, \mathcal{S})$  (see Ref. 11, Sec. 4), which, together with the embeddings  $C^{\alpha/2} \hookrightarrow \mathcal{S} \hookrightarrow C^{\alpha-1}$ , shows that  $A$  converges to a maximal solution with values in  $\mathcal{S}$ .

### C. Gauge covariance

Recall that Theorems 1.11 and 1.14 imply a form of gauge covariance for (1.5). Theorem 1.11(a) is a relatively straightforward consequence of Theorem 2.5. One defines the random variable  $A$  by solving (1.5) until the first time that  $|A(t)|_\alpha \geq 2 + \inf_{B \sim A(t)} |B|_\alpha$ , at which point one uses a measurable selection  $S : \mathcal{O} \rightarrow \mathcal{S}$  to jump to a new small representative  $B$  of the gauge orbit  $[A(t)]$  for which  $|B|_\alpha < 1 + \inf_{a \in [A(t)]} |a|_\alpha$ . These jump times define the increasing sequence of stopping times  $\{\tau_j\}_{j \geq 0}$  in item (ii). Items (i)–(iv) all follow readily from the construction.

The Proof of Theorem 1.11(b), which is the main statement of Theorem 1.11, requires more work. The idea is to use Theorem 1.14, which we admit for now, to couple the solutions to (1.5) with bare mass  $\hat{C}$  and initial conditions  $a \sim b$ . Specifically, let  $\nu$  be a generative probability measure with bare mass  $\hat{C}$  and initial condition  $b \in \mathcal{S}$ , and consider any  $a \sim b$ . Letting  $B$  and  $\xi$  denote the random variable and white noise,



**FIG. 1.** Two random variables  $A, B$  associated with generative probability measures  $\mu, \nu$ , respectively. The set of jump times  $\{\tau_1, \tau_2, \dots\}$  is the union of the sets of stopping times  $\{\zeta_1, \zeta_2, \dots\}$  and  $\{\sigma_1, \sigma_2, \dots\}$  associated with  $\mu, \nu$ . Gray lines indicate gauge orbits. There is a coupling such that  $A^g = B$  for a time-dependent gauge transformation  $g$ . In between the jump times,  $A$  and  $B$  solve (1.5) with bare mass  $\check{C}$  driven by  $\xi$  and  $\check{\xi} = \text{Ad}_g \xi$ , respectively.

respectively, corresponding to  $\nu$ , it follows from Theorem 1.14 that, on the same probability space, there exists a càdlàg process  $A$  defined as above with (1.5) driven by  $\xi \stackrel{\text{def}}{=} \text{Ad}_{g^{-1}} \check{\xi}$  and bare mass  $\check{C}$  in such a way that  $B = A^g$ . Here,  $g$  is càdlàg with values in  $\mathcal{G}^{0,\alpha} \sqcup \{\widehat{\varnothing}\}$  and jump times contained in those of  $A$  and  $B$  and solves (1.8) in between these jump times; see Fig. 1. This shows that the pushforward of  $\nu$  to the orbit space  $\mathcal{O}$  is equal to the pushforward of the law of  $A$  from the Proof of Theorem 1.11(a), which proves Theorem 1.11(b).

*Remark 2.7.* A crucial fact used in the above construction is that the solution  $g$  to (1.8) does not blow up before  $B$  or  $A \stackrel{\text{def}}{=} B^{g^{-1}}$  does. This is due to the elementary but important property of  $S$  that

$$|g|_{C^\alpha} \lesssim 1 + |A|_{\alpha\text{-gr}} + |A^g|_{\alpha\text{-gr}}, \tag{2.5}$$

which, in turn, follows from the facts that, for any curve  $\gamma_{xy}$  with  $\gamma_{xy}(0) = x$  to  $\gamma_{xy}(1) = y$ ,

$$g(y) = \text{hol}(A^g, \gamma_{xy})^{-1} g(x) \text{hol}(A, \gamma_{xy}), \tag{2.6}$$

and that if  $\gamma_{xy}$  is the shortest such curve, then the distance of the holonomy  $\text{hol}(A, \gamma_{xy}) \in G$  from the identity in  $G$  is of order  $|x - y|^\alpha |A|_{\alpha\text{-gr}}$  by Young ODE theory (see Ref. 11, Sec. 3.5).

The Proof of Theorem 1.11(b), therefore, boils down to that of Theorem 1.14, which is the main content of Ref. 11 (Sec. 7). The Proof of Theorem 1.14 proceeds by taking a non-anticipative mollifier  $\chi$  and considering the two systems

$$\begin{aligned} \partial_t B &= \Delta B + B \partial B + B^3 + \text{Ad}_g \xi^\varepsilon + \check{C}_1 B + \check{C}_2 (dg) g^{-1}, \\ (\partial_t g) g^{-1} &= \partial_j ((\partial_j g) g^{-1}) + [B_j, (\partial_j g) g^{-1}] \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} \partial_t \bar{A} &= \Delta \bar{A} + \bar{A} \partial \bar{A} + \bar{A}^3 + \chi^\varepsilon * (\text{Ad}_g \xi) + \check{C}_1 \bar{A} + \check{C}_2 (d\bar{g}) \bar{g}^{-1}, \\ (\partial_t \bar{g}) \bar{g}^{-1} &= \partial_j ((\partial_j \bar{g}) \bar{g}^{-1}) + [\bar{A}_j, (\partial_j \bar{g}) \bar{g}^{-1}], \end{aligned} \tag{2.8}$$

where  $\check{C}_1, \check{C}_2 \in L(\mathfrak{g}, \mathfrak{g})$  are new arbitrary bare masses. These two systems represent the unrenormalized versions of (1.8) and (1.10), respectively, with an extra free parameter (the bare mass  $\check{C}_2$ ). [In Ref. 11, one actually writes the above systems using new variables  $U = \text{Ad}_g$  and  $h = (dg) g^{-1}$ , which assists in computing the desired renormalized equations.]

*Remark 2.8.* Well-posedness for systems (2.7) and (2.8) as  $\varepsilon \downarrow 0$  is generally standard. However, a subtlety arises from the multiplicative noise term  $\text{Ad}_g \xi$ , which is in  $C^{-2-\kappa}$  and leads to problems in posing a suitable fixed point problem ( $-2$  is the threshold regularity at which one cannot extend uniquely a distribution from  $\mathbf{R}_+ \times \mathbf{R}^d$  to  $\mathbf{R} \times \mathbf{R}^d$ ). This issue is handled by decomposing  $g = \mathcal{P}g(0) + \hat{g}$ , where  $\mathcal{P}g(0)$  is the harmonic extension of  $g(0)$  to positive times. Then,  $\hat{g}$  vanishes at  $t = 0$ , which makes the product  $\text{Ad}_g \xi$  better behaved, while the product  $\text{Ad}_{\mathcal{P}g(0)} \xi$  is shown to be a well-defined distribution in  $C^{-2-\kappa}(\mathbf{R} \times \mathbf{R}^d)$  using stochastic estimates.

To show that some renormalized forms of (2.7) and (2.8) converge to the *same* limit as  $\varepsilon \downarrow 0$ , the strategy taken in Ref. 11 is to introduce  $\varepsilon$ -dependent norms on the underlying regularity structure. The idea behind these norms is that they allow one to lift the basic estimate

$$|f - \chi^\varepsilon * f|_{C^\ell} \lesssim \varepsilon^\theta |f|_{C^{\ell+\theta}}$$

to the level of modeled distributions. It is then possible to show that the modeled distributions that solve systems (2.7) and (2.8) are, for small times, at distance of order  $\varepsilon^\theta$  for some  $\theta > 0$  sufficiently small. By continuity of the reconstruction operator, it follows that some renormalized forms of (2.7) and (2.8) converge to the same limit as  $\varepsilon \downarrow 0$ .

We next identify the renormalized forms of (2.7) and (2.8). We drop for the moment the assumption that  $\chi$  is non-anticipative since this step can be done without it. It follows from a direct computation using the algebraic theory developed in Ref. 6 (see also Ref. 11, Sec. 5) that the renormalized equations are

$$\partial_t B = \Delta B + B\partial B + B^3 + (\lambda C_{\text{sym}}^\varepsilon + \check{C}_1)B + \text{Ad}_g(\xi^\varepsilon) + (\lambda \check{C}^\varepsilon + \check{C}_2)(dg)g^{-1} \tag{2.9}$$

and

$$\partial_t \bar{A} = -\Delta \bar{A} + \bar{A}\partial \bar{A} + \bar{A}^3 + (\lambda C_{\text{sym}}^\varepsilon + \check{C}_1)\bar{A} + \chi^\varepsilon * (\text{Ad}_g \xi) + (\lambda \check{C}^{0,\varepsilon} + \check{C}_2)(d\bar{g})\bar{g}^{-1}, \tag{2.10}$$

where

$$\check{C}^\varepsilon = \int \chi^\varepsilon(K * K^\varepsilon), \quad \check{C}^{0,\varepsilon} = \lim_{\delta \downarrow 0} \int \chi^\delta(K * \chi^\delta * K^\varepsilon) = (K * K^\varepsilon)(0),$$

and  $C_{\text{sym}}^\varepsilon$  is the same constant appearing in Theorem 2.5. Remark that  $\check{C}^\varepsilon$  and  $\check{C}^{0,\varepsilon}$  both converge to finite limits as  $\varepsilon \downarrow 0$  because if  $G$  is the heat kernel, then  $(G * G)(t, \cdot) = tG(t, \cdot)$ , which is a bounded function because  $d = 2$ . The components  $g$  and  $\bar{g}$  do not require renormalization and solve the same equations as in (2.7) and (2.8).

The values of  $\check{C}^\varepsilon$  and  $\check{C}^{0,\varepsilon}$  are, respectively, the BPHZ constants associated with the two trees

$$\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} \quad \text{and} \quad \begin{array}{c} \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array}, \tag{2.11}$$

which correspond, before renormalization, to the two functions

$$\xi^\varepsilon \cdot (K * K * \xi^\varepsilon) \quad \text{and} \quad \xi \cdot (K * K^\varepsilon * \xi).$$

*Remark 2.9.* To derive and solve the equation for  $\bar{A}$ , one substitutes  $\xi$  by  $\xi^\delta$  and takes the limit  $\delta \downarrow 0$  with  $\varepsilon > 0$  fixed; this ensures that all objects are smooth for  $\varepsilon, \delta > 0$ . This also explains the definition of  $\check{C}^{0,\varepsilon}$  as a limit in  $\delta \downarrow 0$ .

To summarize, for any  $\check{C}_1, \check{C}_2 \in L(\mathfrak{g}, \mathfrak{g})$  and any mollifier  $\chi$ , the solutions  $B$  and  $\bar{A}$  to (2.9) and (2.10), respectively, converge to the same limit as  $\varepsilon \downarrow 0$  over a short random time interval [the argument in Ref. 11 that  $(B, g)$  and  $(\bar{A}, \bar{g})$  converge as *maximal* solutions uses non-anticipativity of  $\chi$ ].

To make the identification with the equations in Theorem 1.14, we take  $\check{C}_1 = \check{C}$ . We then set  $\check{C}_2 = \lim_{\varepsilon \downarrow 0} \lambda C_{\text{sym}}^\varepsilon + \check{C} - \lambda \check{C}^\varepsilon$  so that  $\lambda \check{C}^\varepsilon + \check{C}_2 = \lambda C_{\text{sym}}^\varepsilon + \check{C} + o(1)$  as  $\varepsilon \downarrow 0$ . Finally, we want to find  $C$  (playing the role of  $\check{C}$ ) so that  $\lambda \check{C}^{0,\varepsilon} + \check{C}_2 = \check{C}_1 - C + o(1)$ . This desired operator is

$$C = \lim_{\varepsilon \downarrow 0} \lambda (\check{C}^\varepsilon - C_{\text{sym}}^\varepsilon - \check{C}^{0,\varepsilon}). \tag{2.12}$$

We have now argued that the solutions  $B$  and  $\bar{A}$  as in Theorem 1.14 with  $\check{C} = C$  converge to the same limit over a short time interval as  $\varepsilon \downarrow 0$  for *any* mollifier  $\chi$ .

Next, we claim that if  $\chi$  is *non-anticipative*, then  $C$  is independent of  $\chi$ . Indeed,  $\check{C}^{0,\varepsilon} = 0$  for non-anticipative  $\chi$ . Furthermore, it follows from the identity  $(\partial_t - \Delta)K = \delta + Q$ , where  $\delta$  is the Dirac delta and  $Q$  is smooth and supported away from the origin, and a computation with integration by parts (see Ref. 11, Lemma 6.9), that  $C$  in this case is equal to

$$\check{C} = \lim_{\varepsilon \downarrow 0} \lambda (\check{C}^\varepsilon - C_{\text{sym}}^\varepsilon) = \lim_{\varepsilon \downarrow 0} \lambda \left( - \int (K * K^\varepsilon)(Q * \chi^\varepsilon) - \int (Q * K^\varepsilon)K^\varepsilon \right). \tag{2.13}$$

Since  $Q$  is supported away from the origin, the final limit is independent of  $\chi$ .

To conclude the Proof of Theorem 1.14, it remains to show that (1.5) with bare mass  $\check{C}$  defined by (2.13) is independent of  $\chi$  and of  $K$ . Remark that  $\chi$  can now be *any* mollifier, not necessarily non-anticipative. Independence of  $\chi$  follows from the final part of Theorem 2.5 since  $\check{C}$  is independent of  $\chi$ . Independence of  $K$  follows from the fact that  $\lim_{\varepsilon \downarrow 0} (\lambda C_{\text{sym}}^\varepsilon + \check{C}) = \lim_{\varepsilon \downarrow 0} \lambda \check{C}^\varepsilon$ , which does not depend on the choice of  $K$  since  $K$  is always equal to the heat kernel near the origin.

*Remark 2.10.* The final mass renormalization one takes in (1.5) to obtain gauge covariance in 2D is, therefore,

$$\lim_{\epsilon \downarrow 0} (\lambda C_{\text{sym}}^\epsilon + \check{C}) = \lim_{\epsilon \downarrow 0} \lambda \check{C}^\epsilon = \lim_{\epsilon \downarrow 0} \lambda \int \chi^\epsilon (K * \chi^\epsilon * K), \tag{2.14}$$

which is the constant “ $\check{C}$ ” appearing in Ref. 11 (Theorem 2.9).

*Remark 2.11.* The existence of  $\check{C}$  with the above properties may appear as a bit of a miracle. Indeed, the fact that  $\check{C}^\epsilon$  and  $\check{C}^{0,\epsilon}$  converge to finite limits was easy to see because  $(G * G)(t, \cdot) = tG(t, \cdot)$  is a bounded function for the heat kernel  $G$  in 2D. On the other hand, the fact that  $C_{\text{sym}}^\epsilon$  converges to a finite limit, and, thus, that  $\check{C}$  exists, is not *a priori* obvious because it relies on a cancellation between diverging BPHZ constants  $4\check{C}^\epsilon$  and  $\check{C}^\epsilon$  in Theorem 2.5. The fact that  $\check{C}$  is furthermore independent of  $\chi$  relies on a cancellation between  $\check{C}^\epsilon$  and  $C_{\text{sym}}^\epsilon$  and that (1.5) with bare mass  $\check{C}$  is independent of  $K$  relies on the expression for  $\check{C}^\epsilon$ .

These cancellations, convergences, and independencies are shown in Ref. 11 using explicit computations. We will see in Sec. III C that there is another argument that allows us to see the existence of  $\check{C}$  by instead exploiting symmetries of the equation.

### III. THREE DIMENSIONS

We now discuss the main results of Ref. 12, which deals with the 3D theory.

#### A. State space

The idea of defining a state space of 1-forms in terms of line integrals no longer works in 3D because already the GFF  $A$  is too singular to be restricted to lines. To see this, recall that the correlation function of  $A$  in 3D behaves like  $C(x, y) \asymp \frac{1}{|x-y|}$  for  $x, y$  close [vs  $C(x, y) \asymp -\log|x-y|$  in 2D]. Therefore, for  $\ell = (x, v)$ ,

$$\mathbf{E}|A(\ell)|^2 = \int_0^1 dt \int_0^1 ds |\ell|^2 C(x + tv, x + sv) \asymp \int_0^1 dt \int_0^1 ds \frac{|\ell|}{|t-s|} = \infty.$$

This strongly suggests that it is hopeless to find a state space for 3D quantum Yang–Mills such that holonomies are well-defined as we did in 2D.

The construction of the state space  $\mathcal{S}$  in Ref. 12 proceeds in two steps. The first step is to define a space  $\mathcal{I}$  of initial conditions for a gauge-covariant regularizing operator. Abstractly, we will find a metric space of distributional 1-forms  $(\mathcal{I}, \Theta)$  and a family of operators  $\{\mathcal{F}_t\}_{t>0}$ ,  $\mathcal{F}_t : \mathcal{I} \rightarrow C^\infty$  such that

(a) for smooth  $A, B$ ,

$$A \sim B \iff \mathcal{F}_t(A) \sim \mathcal{F}_t(B) \text{ for some } t > 0, \tag{3.1}$$

(b)  $\mathcal{F}_t : \mathcal{I} \rightarrow C^\infty$  is continuous for every  $t > 0$ .

If we can find such  $\mathcal{I}$  and  $\{\mathcal{F}_t\}_{t>0}$ , then we can extend gauge equivalence  $\sim$  to  $\mathcal{I}$  by using (3.1) as a *definition*. Finally, we want  $\mathcal{I}$  to be sufficiently large to contain distributions as rough as the GFF.

Our concrete choice for  $\mathcal{F}_t(a)$  is the solution  $A(t)$  to the deterministic YM flow (with the DeTurck term),<sup>66</sup>

$$\partial_t A = -d_A^* F(A) - d_A d^* A = \Delta A + A \partial A + A^3, \quad A(0) = a. \tag{3.2}$$

This choice is natural because we ultimately want to start (1.5), the stochastic version of (3.2), from the initial data as rough as the GFF, which is at least as hard as solving (3.2).

*Remark 3.1.* The fact that (3.1) holds for  $\mathcal{F}$  was likely already known in the literature, but a proof is given in Ref. 12 (Sec. 2.2) based on the analytic continuation and global existence of the YM flow without the DeTurck term.<sup>41,58</sup>

The second step in the construction of  $\mathcal{S}$  is to augment  $\mathcal{I}$  with an additional norm, which ensures that a form of the bound (2.5) holds. This turns out to be critical in several places of the construction for the Markov process in Theorem 1.11(b).

We mention that the idea to use the YM flow to define a suitable space of distributional 1-forms was already suggested in Ref. 15 (see also Refs. 22, 30, 49, and 55 for related ideas in physics). We also point out that another state space that bears close similarity to  $\mathcal{I}$  was independently defined in Refs. 9 and 10 and was shown to support the GFF.

### 1. The first half

Before defining  $\mathcal{I}$ , we briefly illustrate why classical Hölder–Besov spaces  $C^\eta$  are not suitable for our purposes. The standard strategy to solve (3.2) is to rewrite the equation in mild formulation,

$$A(t) = \mathcal{M}_t(A) \stackrel{\text{def}}{=} \mathcal{P}_t a + \int_0^t \mathcal{P}_{t-s} [A(s) \partial A(s) + A(s)^3] ds,$$

where  $\mathcal{P}_t = e^{t\Delta}$  is the heat flow, and show that the map  $\mathcal{M} : A \mapsto \mathcal{M}(A)$  is a contraction on a ball in a suitable Banach space. This Banach space should at least contain  $\mathcal{M}(0) = \mathcal{P}a$  and  $\mathcal{M}(\mathcal{P}a)$  restricted to a short time interval, which are the first and second Picard iterates, respectively. However, for generic  $a \in C^\eta$ , the best we can do to handle the product  $\mathcal{P}_{t-s}[\mathcal{P}_s a \cdot \partial \mathcal{P}_s a]$  in  $\mathcal{M}(\mathcal{P}a)$  is to estimate for  $s > 0$ ,

$$|\mathcal{P}_s a|_{L^\infty} \lesssim s^{\frac{\eta}{2}} |a|_{C^\eta}, \quad |\partial \mathcal{P}_s a|_{L^\infty} \lesssim s^{\frac{\eta}{2} - \frac{1}{2}} |a|_{C^\eta} \tag{3.3}$$

$$\Rightarrow |\mathcal{P}_s a \cdot \partial \mathcal{P}_s a|_{L^\infty} \lesssim s^{\eta - \frac{1}{2}} |a|_{C^\eta}. \tag{3.4}$$

We now recall that the GFF takes values in  $C^\eta$  for any  $\eta < -\frac{1}{2}$  (and not for  $\eta = -\frac{1}{2}$ ). Therefore, trying to estimate the term  $\int_0^t \mathcal{P}_{t-s}[\mathcal{P}_s a \cdot \partial \mathcal{P}_s a] ds$  in the second Picard iterate leads to a non-integrable singularity  $\int_0^t s^{\eta - \frac{1}{2}} ds = \infty$ . The above estimates are, in general, sharp, which suggests that one cannot start the YM flow (3.2) from the initial data in  $C^\eta$  with  $\eta < -\frac{1}{2}$  (and even in  $C^{-1/2}$ ).

To circumvent this issue and motivate the definition of  $\mathcal{I}$ , remark that the GFF  $a$  is a highly non-generic element of  $C^\eta$  and  $\mathcal{P}_s a \cdot \partial \mathcal{P}_s a$  behaves better than the naive bound (3.4) would suggest. Indeed, a second moment estimate shows that for any  $\beta \in (-1, 0)$  and  $\delta > 1 + \frac{\beta}{2}$ , almost surely uniformly in  $s \in (0, 1)$ ,

$$|\mathcal{P}_s a \cdot \partial \mathcal{P}_s a|_{C^\beta} \lesssim s^{-\delta} \tag{3.5}$$

[The way to guess the bound (3.5) is to use the estimates  $|\mathcal{P}_s a|_{C^{\beta/2}} \lesssim s^{-\frac{1}{4} - \kappa - \frac{\beta}{4}}$  and  $|\partial \mathcal{P}_s a|_{C^{\beta/2}} \lesssim s^{-\frac{3}{4} - \kappa - \frac{\beta}{4}}$  for any  $\kappa > 0$  and then pretend that multiplication is a bounded operator  $C^{\beta/2} \times C^{\beta/2} \rightarrow C^\beta$ —this is clearly false since  $\beta < 0$ , but ends up working for  $(\mathcal{P}_s a, \partial \mathcal{P}_s a)$  due to probabilistic cancellations.]

Since we can take  $\delta < 1$  in (3.5) [vs  $-\eta + \frac{1}{2} > 1$  in (3.4)], this improved regularity ends up being enough to show that every Picard iterate of  $\mathcal{M}$  is well-defined when  $a$  is the GFF. It is, therefore, natural to make the following definition.

*Definition 3.2.* For  $\eta < -\frac{1}{2}$ ,  $\beta < 0$ , and  $\delta \in (1 + \frac{\beta}{2}, 1)$ , let  $\mathcal{I}$  be the completion of smooth 1-forms under the metric

$$\Theta(A, B) \stackrel{\text{def}}{=} |A - B|_{C^\eta} + \sup_{t \in (0, 1)} t^\delta |\mathcal{P}_t A \cdot \partial \mathcal{P}_t A - \mathcal{P}_t B \cdot \partial \mathcal{P}_t B|_{C^\beta}.$$

*Remark 3.3.* The space  $\mathcal{I}$  can be identified with a subset of  $C^{0, \eta} \stackrel{\text{def}}{=} \overline{C^\infty C^\eta}$  because the map  $C^\eta \ni A \mapsto \mathcal{P}_t A \cdot \partial \mathcal{P}_t A \in C^\infty((0, 1), C^\beta)$  has a closed graph.

A standard argument with Young’s product theorem and estimates of the type (3.3) shows that the YM flow extends to  $\mathcal{I}$  in the following sense.

*Proposition 3.4.* For every ball  $\mathcal{B}$  in  $(\mathcal{I}, \Theta)$  centered at 0, there exists  $T > 0$  such that for all  $t \in (0, T)$ , the YM flow (3.2) extends to a continuous function  $\mathcal{F}_t : \mathcal{B} \rightarrow C^\infty$  (which is Lipschitz for any norm on  $C^\infty$ ).

We can therefore extend gauge equivalence  $\sim$  to  $\mathcal{I}$  by using (3.1) as a definition. We emphasize that  $\mathcal{I}$  is not a vector space and that this is unavoidable: there is no Banach space that carries the GFF on  $\mathbf{T}^3$  and to which the YM flow (3.2) extends as a continuous function (locally in time); see Ref. 18.

### 2. The second half

We now refine the space  $\mathcal{I}$  in order to obtain control on gauge transformations of the type (2.5), which proves crucial in the construction of the Markov process on gauge orbits associated with (1.5).

*Definition 3.5.* For  $\alpha \in (0, \frac{1}{2})$  and  $\theta > 0$ , define the norm  $||| \cdot |||_{\alpha, \theta}$  on smooth 1-forms

$$|||A|||_{\alpha, \theta} \stackrel{\text{def}}{=} \sup_{t \in (0, 1)} \sup_{|\ell| < t^\theta} \frac{|(\mathcal{P}_t A)(\ell)|}{|\ell|^\alpha},$$

where the second supremum is over all line segments  $\ell = (x, v) \in \mathbf{T}^3 \times \{v \in \mathbf{R}^3 : |v| < \frac{1}{4}\}$  with  $|\ell| \stackrel{\text{def}}{=} |v| < t^\theta$ , and  $(\mathcal{P}_t A)(\ell)$  is the line integral of  $\mathcal{P}_t A$  along  $\ell$  defined analogously to (2.1). Define further the metric

$$\Sigma(A, B) = \Theta(A, B) + \|A - B\|_{\alpha, \theta},$$

and let  $\mathcal{S}$  by the completion of smooth 1-forms under  $\Sigma$ .

To motivate the norm  $\|\cdot\|_{\alpha, \theta}$ , recall that the estimate (2.5) relies on the identity (2.6), which, in turn, requires that line integrals and holonomies of  $A$  are well-defined. However, we saw that the GFF  $A$  cannot even be restricted to lines.

The idea above is to consider, instead of  $A$ , the heat flow regularization  $\{\mathcal{P}_t A\}_{t \in (0,1)}$ . A quick computation shows that, uniformly in  $0 < \sqrt{t} < |\ell| < \frac{1}{4}$ ,

$$\mathbf{E}|(\mathcal{P}_t A)(\ell)|^2 \asymp |\ell| \log(|\ell| t^{-1})$$

(with even better bounds for  $|\ell| \leq \sqrt{t}$ ). Therefore, as expected,  $\mathbf{E}|(\mathcal{P}_t A)(\ell)|^2$  blows up for fixed  $\ell$  as  $t \downarrow 0$ , but rather slowly. Furthermore, restricting to short length scales, say,  $|\ell| < t^\theta$  for any  $\theta > 0$ ,

$$\mathbf{E}|(\mathcal{P}_t A)(\ell)|^2 \lesssim -|\ell| \log |\ell|$$

uniformly in  $t \in (0, 1)$  and  $|\ell| < t^\theta$ . Combined with a Kolmogorov argument, this shows that  $\|A\|_{\alpha, \theta} < \infty$  almost surely. The restriction to  $\alpha < \frac{1}{2}$  is natural because the GFF in 3D has  $\frac{1}{2}$  less regularity than in 2D (e.g.,  $C^{-\kappa-1/2}$  in 3D vs  $C^{-\kappa}$  in 2D for the Hölder–Besov regularity), and we saw that  $|A|_{\alpha-\text{gr}} < \infty$  for  $\alpha < 1$  for the GFF  $A$  in 2D.

*Remark 3.6.* The metric space  $\mathcal{S}$  can be identified with a subset of  $\mathcal{I} \subset C^{0, \eta}$  and comes with the parameters  $(\eta, \beta, \delta, \alpha, \theta)$ , the possible range of which is given in Ref. 12 (Sec. 5);  $(\mathcal{S}, \Sigma)$  is the space appearing in Theorem 1.3 for  $d = 3$ .

With these definitions, one can show that the group of sufficiently smooth gauge transformations acts continuously on  $\mathcal{S}$ . Specifically, there exists  $\rho \in (\frac{1}{2}, 1)$  such that  $(A, g) \mapsto A^g$  extends continuously to a map  $\mathcal{S} \times \mathcal{G}^\rho \rightarrow \mathcal{S}$ , where

$$\mathcal{G}^\rho \stackrel{\text{def}}{=} C^\rho(\mathbf{T}^3, G).$$

Furthermore, this action preserves  $\sim$  defined by (3.1), i.e.,  $A \sim A^g$  for all  $(A, g) \in \mathcal{S} \times \mathcal{G}^\rho$ .

We can now state Ref. 12 (Theorem 2.39), which is one of the main results of (Ref. 12, Sec. 2) and the motivation behind the norm  $\|\cdot\|_{\alpha, \theta}$ .

**Theorem 3.7.** *There exist constants  $C, q > 0$  and  $\nu \in (0, \frac{1}{2})$  such that, for all  $g \in \mathcal{G}^\rho$  and  $A \in \mathcal{S}$ ,*

$$|g|_{C^\nu} \leq C(1 + \Sigma(A, 0) + \Sigma(A^g, 0))^q.$$

The Proof of Theorem 3.7 relies on two estimates: (i) the estimate (2.5) used in the 2D case (which, of course, holds in arbitrary dimension) and (ii) a “backward estimate” that controls the initial condition of a parabolic PDE in terms of its behavior for positive times [see Ref. 12, Lemma 2.46(b)]—this estimate is applied to the harmonic map flow-type PDE solved by  $h$  for which  $h(0) = g$  and  $\mathcal{F}_t(A)^{h(t)} = \mathcal{F}_t(A^g)$  for all  $t > 0$ . Theorem 3.7 is then obtained by suitably interpolating between estimates (i) and (ii).

*Remark 3.8.* One can show that mollifications of the SHE converge in probability in the space  $C(\mathbf{R}_+, \mathcal{S})$  (see Ref. 12, Corollary 3.14). In particular, the SHE admits a modification with sample paths in  $C(\mathbf{R}_+, \mathcal{S})$ .

*Remark 3.9.* Unlike in 2D, the action of  $\mathcal{G}^\rho$  on  $\mathcal{S}$  is not transitive over the orbits, and it is unclear if  $\sim$ , or some variant of it, is determined by the action of a group. This lack of a gauge group is responsible for the gap in our understanding of the quotient space  $\mathcal{S}/\sim$  in 3D vs 2D; see Remark 1.5.

## B. Local solutions

We next explain how one proves Theorem 1.6 in 3D, which is done in Ref. 12 (Sec. 5). We do not restate the result here like we did in Sec. II B since we cannot make it substantially more precise.

Though primarily using the theory of regularity structures as before, there are two main additional challenges on top of the 2D case. The first is purely algebraic and concerns showing that the renormalization counterterms are of the form  $C_{\text{bphz}}^\epsilon A$ . The difficulty is that there are dozens of trees that potentially contribute to renormalization (vs just nine trees in 2D; see Ref. 11, Sec. 6.2.3).

By power counting, one can deduce that the renormalization is *linear* in  $A$ . To argue that one sees precisely  $C_{\text{bphz}}^\epsilon \in L_G(\mathfrak{g}, \mathfrak{g})$  requires a systematic approach to symmetry arguments, which is developed in Ref. 12 (Sec. 4) and which could of independent interest in other contexts.

To give an example of how this works, we argue that the renormalization is “block diagonal,” i.e., if the counterterm  $cA_j$  appears in the  $A_i$  equation for  $j \neq i$ , then  $c = 0$ . Indeed, if we flip the coordinate  $x_i \mapsto -x_i$  and, thus,  $\partial_i \mapsto -\partial_i$ , together with  $A_i \mapsto -A_i$  and  $\xi_i^\epsilon \mapsto -\xi_i^\epsilon$ , while keeping all terms with indices  $j \neq i$  the same, it is immediate that all the terms in the  $A_i$  equation (1.3) flip sign. Using the symmetry of the noise  $\xi_i^{\text{law}} = -\xi_i^\epsilon$  and invariance under the flip  $x_i \mapsto -x_i$  of the kernel  $K$  used to define  $C_{\text{bphz}}^\epsilon$ , one can show that renormalized equation must possess the same symmetry, namely, all terms in the renormalized  $A_i$  equation must flip sign. Since we kept  $A_j$  the same, this shows that any factor  $cA_j$  in the renormalized  $A_i$  equation must have  $c = 0$ .

The way one argues that the *same*  $C_{\text{bphz}}^\epsilon$  appears for all  $i \in \{1, 2, 3\}$  and that  $C_{\text{bphz}}^\epsilon \in L_G(\mathfrak{g}, \mathfrak{g})$  is similar: one exploits symmetry under reflections  $x_i \leftrightarrow x_j$  and  $\xi_i^{\text{law}} = \xi_j^\epsilon$  for the former and symmetry under constant gauge transformations  $A_i \mapsto \text{Ad}_g A_i$  and  $\xi_i^{\text{law}} = \text{Ad}_g \xi_i^\epsilon$ , where  $g \in G$  for the latter.

The second challenge is analytic and comes from the singularity of the initial condition in  $C^{-\frac{1}{2}-\kappa}$  for  $\kappa > 0$  (this was already encountered in 2D in a more mild form; see Remark 2.8). This singularity means, for example, that  $\mathcal{P}A(0)\partial\Psi$  is ill-defined for generic distributions  $\Psi \in C^{-\frac{1}{2}-\kappa}(\mathbf{R} \times \mathbf{T}^3)$  and  $A(0) \in C^{-\frac{1}{2}-\kappa}(\mathbf{T}^3)$ . Similar to the discussion in Sec. III A 1, this type of product appears in the Picard iteration used to solve (1.5). As in Remark 2.8, this issue is addressed by decomposing  $A = \mathcal{P}A(0) + \Psi + \hat{A}$ , where  $\Psi$  solves the SHE  $\partial_t \Psi = \Delta \Psi + \xi^\epsilon$ , and solving for the “remainder”  $\hat{A}$ . One then shows with separate stochastic bounds that  $\mathcal{P}A(0)\partial\Psi$  and  $\Psi\partial\mathcal{P}A(0)$  converge in  $C^{-2-\kappa}(\mathbf{R} \times \mathbf{T}^3)$  as  $\epsilon \downarrow 0$ .

A closely related issue *not present* in 2D is that of restarting the equation at some positive time  $\tau > 0$  to obtain maximal solutions. This is because, for  $\epsilon > 0$ ,  $A(\tau)$  and  $\xi^\epsilon \upharpoonright_{[\tau, \infty)}$  see each other on a time interval of order  $\epsilon^2$ . Since the regularities of  $A(\tau)$  and  $\partial\Psi$  add up to  $< -2$ , this breaks the argument used to show that  $\mathcal{P}A(0)\partial\Psi$  and  $\Psi\partial\mathcal{P}A(0)$  converge as  $\epsilon \downarrow 0$  when  $A(0)$  is independent of  $\xi$ . To restart the equation, one instead leverages that  $A(\tau)$  for  $\tau > 0$  is not a generic element of  $\mathcal{S}$  but takes the form  $A(\tau) = \Psi(\tau) + R(\tau)$ , where  $R(\tau) \in C^{-\kappa}(\mathbf{T}^3)$ . Since  $\Psi$  is defined globally in time, this decomposition allows one to restart the equation using the “generalized Da Prato–Debussche trick” from Ref. 6.

### C. Gauge covariance

Finally, we describe the Proof of Theorem 1.11 in 3D. The Proof of Theorem 1.11(a) is similar to its 2D counterpart. The only appreciable difference is that the measurable selection  $S : \mathcal{O} \rightarrow \mathcal{S}$  is replaced by a Borel map  $S : \mathcal{S} \rightarrow \mathcal{S}$ , which preserves gauge orbits and such that  $\tilde{\Sigma}(S(X)) \leq 2\text{inf}_{Y \sim X} \tilde{\Sigma}(Y)$  whenever the right-hand side is finite. Here,  $\tilde{\Sigma} \geq \Sigma$  is defined analogously to  $\Sigma$  but with a stronger set of parameters  $(\eta, \beta, \delta, \alpha, \theta)$ . This complication is due to a lack of any nice known properties of  $\mathcal{O} = \mathcal{S}/\sim$  in 3D (e.g., polishness) (see Remarks 1.5 and 3.9), and we instead leverage compactness of the embedding  $(\tilde{\mathcal{S}}, \tilde{\Sigma}) \hookrightarrow (\mathcal{S}, \Sigma)$ .

The Proof of Theorem 1.11(b) is where we start to see a difference with the 2D case. First, admitting for now Theorem 1.14, we aim to prove that solutions to (1.5) with bare mass  $\check{C}$  and gauge equivalent initial conditions can be suitably coupled. Specifically, one has the following result.

**Lemma 3.10** (Coupling). Suppose  $A$  solves (1.5) with bare mass  $\check{C}$  and initial condition  $a$ . Then, for any  $b \sim a$ , there exists on the same probability space a white noise  $\tilde{\xi}$  and a process  $(B, g)$  such that  $g(t) \in \mathcal{G}^{\frac{3}{2}-\kappa}$  and  $A(t)^{g(t)} = B(t)$  for all  $t > 0$  (before blow-up of  $A, B$ ) and such that  $B$  solves (1.5) driven by  $\tilde{\xi}$  with bare mass  $\check{C}$ .

If  $a^{g(0)} = b$  for some  $g(0) \in \mathcal{G}^\rho$  for  $\rho > \frac{1}{2}$  as in Sec. III A 2, then this result follows almost immediately from Theorem 1.14. However, unlike the 2D case, it is now possible that  $b \sim a$ , but no  $g(0)$  exists such that  $a^{g(0)} = b$ , which leads to trouble in applying Theorem 1.14—we effectively have no initial condition for  $g$  in the PDE (1.8).

To circumvent the issue, we regularize  $a$  and  $b$  using the YM flow  $\mathcal{F}_t$  so that  $\mathcal{F}_t(a)^{g_t} = \mathcal{F}_t(b)$  for all  $t$  sufficiently small and some smooth  $g_t$ . Then,  $\mathcal{F}_t(a) \rightarrow a$  and  $\mathcal{F}_t(b) \rightarrow b$  in  $\mathcal{S}$  as  $t \downarrow 0$ , which, in particular, implies that  $\limsup_{t \downarrow 0} \|g_t\|_{C^\nu} < \infty$  for some  $\nu \in (0, \frac{1}{2})$  due to Theorem 3.7. Therefore, there exists  $g(0) \in \mathcal{G}^\nu$  such that  $g_t \rightarrow g(0)$  in  $\mathcal{G}^{\nu/2}$  along a subsequence. We can then rewrite the equation for  $g$  in (1.8) in terms of  $A = B^{g^{-1}}$ , namely,

$$g^{-1}\partial_t g = \partial_j(g^{-1}\partial_j g) + [A_j, g^{-1}\partial_j g], \tag{3.6}$$

and show that the system for  $(A, g)$ , where  $A$  solves (1.5), is well-posed for any initial condition in  $\mathcal{S} \times \mathcal{G}^\nu$  [vs  $\mathcal{S} \times \mathcal{G}^\rho$  with  $\rho > \frac{1}{2}$  for (1.8) due to the multiplicative noise  $\text{Ad}_g \xi$ ]. One can then use continuity of  $(A, g)$  with respect to the initial conditions, the fact that  $g$  takes values in  $\mathcal{G}^{\frac{3}{2}-\kappa}$  for positive times and the joint continuity of the group action  $\mathcal{G}^{\frac{3}{2}-\kappa} \times \mathcal{S} \rightarrow \mathcal{S}$  to prove Lemma 3.10. The  $g$  in Lemma 3.10 solves precisely (3.6) with initial condition  $g(0)$ .

With Lemma 3.10 in hand, together with the fact that  $g$  in its statement cannot blow up before  $\Sigma(A, 0) + \Sigma(B, 0)$  blows up (again due to Theorem 3.7), it is relatively straightforward to prove Theorem 1.11(b) like we did in the 2D case. See, in particular, the discussion around Fig. 1.

It remains to prove Theorem 1.14 to complete the above argument. Fix henceforth a non-anticipative mollifier  $\chi$ . Like in Sec. II C, it is natural consider systems (2.7) and (2.8) and show that their renormalized versions converge to the same limit. To do this, we show that the equations for  $g$  and  $\bar{g}$ , which are now singular, do not require renormalization and that the renormalized equations for  $B$  and  $\bar{A}$  are of exactly the same form as in Sec. II C, namely,

$$\partial_t B = \Delta B + B\partial B + B^3 + \text{Ad}_g \xi^\varepsilon + (C_{\text{bphz}}^\varepsilon + \hat{C}_1)B + (\tilde{C}^\varepsilon + \hat{C}_2)(dg)g^{-1} \quad (3.7)$$

and

$$\partial_t \bar{A} = \Delta \bar{A} + \bar{A}\partial \bar{A} + \bar{A}^3 + \chi^\varepsilon * (\text{Ad}_{\bar{g}} \xi) + (C_{\text{bphz}}^\varepsilon + \hat{C}_1)\bar{A} + (\tilde{C}^{0,\varepsilon} + \hat{C}_2)(d\bar{g})\bar{g}^{-1}. \quad (3.8)$$

The constants  $\tilde{C}^\varepsilon, \tilde{C}^{0,\varepsilon}$  are, respectively, called  $C_{\text{Gauge}}^\varepsilon, C_{\text{Gauge}}^{0,\varepsilon}$  in Ref. 12 and correspond to  $\lambda\tilde{C}^\varepsilon, \lambda\tilde{C}^{0,\varepsilon}$  in Sec. II C;  $C_{\text{bphz}}^\varepsilon$  is the BPHZ constant in Theorem 1.6 and corresponds to  $\lambda C_{\text{sym}}^\varepsilon$  in Sec. II.

Both of these facts are again proven by introducing the new variables  $h = (dg)g^{-1}$  and  $U = \text{Ad}_g$  and writing the corresponding equations for  $h, U$ . This time, instead of a direction computation with just two trees as in (2.11), a more complicated strategy relying on power counting and symmetry arguments is necessary. The main insight, which helps with this argument, is that the trees appearing in the “lifted”  $B$  and  $\bar{A}$  equations are obtained by attaching (or grafting) the trees appearing in the  $U$  equation onto the trees appearing in (1.5) (see Ref. 12, Sec. 6.2.1).

The analytic theory for  $B$  and  $\bar{A}$  equations is somewhat more involved than what we saw in Secs. II C and III B, but the general strategy is the same: we decompose the solution into the initial condition, globally defined singular terms, and a better behaved remainder and solve for the last of these. Restarting these equations, specifically the equation for  $\bar{A}$ , is also not straightforward because, unlike in Sec. III B, we are outside the scope of the generalized Da Prato–Debussche trick of Ref. 6 since the multiplicative noise means that the most singular part of the solution is not just the SHE. Therefore, an entirely separate fixed point problem needs to be written for the restarted equation (see Ref. 12, Sec. 6.6), which leverages that the new initial condition comes from a modeled distribution defined for earlier times. The proof that  $B$  and  $\bar{A}$  converge to the same limit uses the same  $\varepsilon$ -dependent norms on regularity structures introduced in Ref. 11.

To conclude the Proof of Theorem 1.14, like in Sec. II C, it suffices to show that

$$\tilde{C} \stackrel{\text{def}}{=} \lim_{\varepsilon \downarrow 0} (\tilde{C}^\varepsilon - C_{\text{bphz}}^\varepsilon - \tilde{C}^{0,\varepsilon}) \text{ exists and is independent of non-anticipative } \chi. \quad (3.9)$$

The difficulty compared to 2D is that we have no explicit formulas for the constants on the right-hand side. Instead, the argument in Ref. 12 proceeds in two steps.

First, one shows that

$$\limsup_{\varepsilon \downarrow 0} |\tilde{C}^\varepsilon - C_{\text{bphz}}^\varepsilon| < \infty \quad \text{and} \quad \limsup_{\varepsilon \downarrow 0} |\tilde{C}^{0,\varepsilon}| < \infty. \quad (3.10)$$

Arguing by contradiction, suppose that

$$|\tilde{C}^\varepsilon - C_{\text{bphz}}^\varepsilon| \rightarrow \infty \quad (3.11)$$

along a subsequence of  $\varepsilon \downarrow 0$ . The idea, inspired by Ref. 7, is to introduce a parameter  $\sigma^\varepsilon \in \mathbf{R}$  in front of the noise that we take to zero as  $\varepsilon \downarrow 0$  in a precise way.

Letting  $\tilde{C}_\sigma^\varepsilon, C_{\text{bphz},\sigma}^\varepsilon$ , etc., denote the renormalization constants associated with the noise  $\sigma\xi$ , it is not difficult to see that these constants depend polynomially on  $\sigma$  and converge to zero as  $\sigma \downarrow 0$ . Therefore, by assumption (3.11), we can send  $\sigma^\varepsilon \downarrow 0$  as  $\varepsilon \downarrow 0$  in such a way that, after passing to another subsequence, we can find  $\hat{C} \neq 0$  such that

$$\hat{C}^\varepsilon \stackrel{\text{def}}{=} C_{\text{bphz},\sigma^\varepsilon}^\varepsilon - \tilde{C}_{\sigma^\varepsilon}^\varepsilon \rightarrow \hat{C}.$$

Now, we consider the equation for  $B$  with bare masses  $\hat{C}_1 = 0$  and  $\hat{C}_2 = \hat{C}^\varepsilon$ , that is,

$$\partial_t B = \Delta B + B\partial B + B^3 + C_{\text{bphz},\sigma^\varepsilon}^\varepsilon B + \sigma^\varepsilon \text{Ad}_g \xi^\varepsilon + (\tilde{C}_{\sigma^\varepsilon}^\varepsilon + \hat{C}^\varepsilon)(dg)g^{-1}.$$

Remark that  $\tilde{C}_{\sigma^\varepsilon}^\varepsilon + \hat{C}^\varepsilon = C_{\text{bphz},\sigma^\varepsilon}^\varepsilon$  by definition; thus,  $B = A^\varepsilon$  pathwise, where  $A$  solves (1.5) with bare mass zero, i.e.,

$$\partial_t A = \Delta A + A\partial A + A^3 + C_{\text{bphz},\sigma^\varepsilon}^\varepsilon A + \sigma^\varepsilon \xi^\varepsilon.$$

Since  $\sigma^\varepsilon \downarrow 0$ ,  $A$  converges as  $\varepsilon \downarrow 0$  to solution of the deterministic YM flow (with the DeTurck term)

$$\partial_t A = \Delta A + A \partial A + A^3 \tag{3.12}$$

with the same initial condition. Furthermore, because  $\hat{C}^\varepsilon \rightarrow \hat{C}$ , we can use joint continuity in the bare mass and the noise to see that  $B$  converges as  $\varepsilon \downarrow 0$  to the solution of

$$\partial_t B = \Delta B + B \partial B + B^3 + \hat{C}(dg)g^{-1}. \tag{3.13}$$

On the other hand, the equality  $A^\varepsilon = B$  is preserved under the  $\varepsilon \downarrow 0$  limit, where  $g$  solves (1.8), but this implies by the calculation in Sec. I B 1 with  $\xi \equiv 0$  that  $B$  should also solve

$$\partial_t B = \Delta B + B \partial B + B^3. \tag{3.14}$$

This is clearly a contradiction because, for any  $\hat{C} \neq 0$ , we can find the initial conditions  $(B(0), g(0))$  for which the solutions to (3.13) and (3.14) are different.

The argument that  $\limsup_{\varepsilon \downarrow 0} |\tilde{C}^{0,\varepsilon}| < \infty$  is similar. Specifically, if  $|\tilde{C}^{0,\varepsilon}| \rightarrow \infty$  along a subsequence, then we can find  $\sigma^\varepsilon \downarrow 0$  such that  $\tilde{C}_{\sigma^\varepsilon}^{0,\varepsilon} \rightarrow \hat{C} \neq 0$  along another subsequence. Now, we consider

$$\partial \bar{A} = \Delta \bar{A} + \bar{A} \partial \bar{A} + \bar{A}^3 + (C_{\text{bphz},\sigma^\varepsilon}^\varepsilon + \hat{C}_1) \bar{A} + (\tilde{C}_{\sigma^\varepsilon}^{0,\varepsilon} + \hat{C}_2)(d\bar{g})\bar{g}^{-1} + \chi^\varepsilon * (\text{Ad}_{\bar{g}} \sigma^\varepsilon \xi)$$

with bare masses  $\hat{C}_1 = 0$  and  $\hat{C}_2 = -\tilde{C}_{\sigma^\varepsilon}^{0,\varepsilon}$ . With this choice, the term  $(d\bar{g})\bar{g}^{-1}$  vanishes and the solution is, by Itô isometry (which requires  $\chi$  non-anticipative), equal in law to the solution of

$$\partial \bar{A} = \Delta \bar{A} + \bar{A} \partial \bar{A} + \bar{A}^3 + C_{\text{bphz},\sigma^\varepsilon}^\varepsilon \bar{A} + \sigma^\varepsilon \xi^\varepsilon.$$

However, now, since  $\sigma^\varepsilon \downarrow 0$ ,  $\bar{A}$  converges in law again to the solution of the deterministic YM flow (3.12), while  $\bar{A}$  converges to the solution of

$$\partial \bar{A} = \Delta \bar{A} + \bar{A} \partial \bar{A} + \bar{A}^3 - \hat{C}(d\bar{g})\bar{g}^{-1},$$

which are not equal for generic initial conditions, bringing us to a contradiction.

*Remark 3.11.* The bounds (3.10) are actually used in the short time analysis of (3.7) and (3.8) since they allow us to relate  $B$  and  $\bar{A}$  to a simpler equation with *additive* noise through the above mechanism ( $B$  is related through pathwise gauge transformations and  $\bar{A}$  is related through equality in law; see Ref. 12, Sec. 6.6).

To finish the proof of (3.9), it is left to show that

$$\lim_{\varepsilon \downarrow 0} C_{\text{bphz}}^\varepsilon - \tilde{C}^\varepsilon \text{ exists and is independent of non-anticipative } \chi \tag{3.15}$$

and similarly for  $\tilde{C}^{0,\varepsilon}$ . Arguing again by contradiction, suppose we have two subsequences  $\varepsilon \downarrow 0$  along which  $C_{\text{bphz}}^\varepsilon - \tilde{C}^\varepsilon$  converges to distinct limits  $\hat{C}_1$  and  $\hat{C}_2$ . We then consider the  $B$  equation with bare masses  $\hat{C}_1 = 0$  and  $\hat{C}_2 = C_{\text{bphz}}^\varepsilon - \tilde{C}^\varepsilon$ . With this choice,  $B = A^\varepsilon$  for all  $\varepsilon > 0$  where  $A$  solves (1.5) with zero bare mass and  $g$  solves (3.6). Since the limit of  $(A, g)$  is the same along the two subsequences, it follows that the limit of  $B$  along the two subsequence is also the same almost surely. However, it is possible to show that the limiting bare mass constants  $\hat{C}_1$  and  $\hat{C}_2$  can be recovered from the  $\varepsilon \downarrow 0$  limit of  $(B, g)$  (see Ref. 12, Appendix D), which leads to a contradiction since we assumed  $\hat{C}_1 \neq \hat{C}_2$ . The same argument shows that the limit in (3.15) is independent of  $\chi$  because the solution  $(A, g)$  to (1.5) and (3.6) is independent of  $\chi$ .

Finally, to prove that  $\lim_{\varepsilon \downarrow 0} \tilde{C}^{0,\varepsilon}$  exists and is independent of the non-anticipative mollifier  $\chi$ , one argues in a similar way except, as earlier, we appeal to equality with (1.5) *in law* and use instead that two limiting solutions  $(\bar{A}, \bar{g})$  with different bare masses cannot be equal *in law*.

*Remark 3.12.* These final statements mimic exactly what we saw in Sec. II C where  $\tilde{C}^{0,\varepsilon} = 0$  for non-anticipative  $\chi$  and  $\lim_{\varepsilon \downarrow 0} (\tilde{C}^\varepsilon - C_{\text{bphz}}^\varepsilon)$  is given by (2.13). By analogy, (3.15) should hold for any mollifier, not necessarily non-anticipative, but this is not necessarily the case for  $\lim_{\varepsilon \downarrow 0} \tilde{C}^{0,\varepsilon}$ . Furthermore, as in 2D, we expect that  $\lim_{\varepsilon \downarrow 0} \tilde{C}^{0,\varepsilon} = 0$  for non-anticipative  $\chi$ .

#### IV. OPEN PROBLEMS

We close with several open problems that we believe to be of interest.

- Does the Markov process on gauge orbits in Theorem 1.11 possess a unique invariant measure? The existence of the invariant measure should imply uniqueness due to the strong Feller property<sup>38</sup> and full support theorem for SPDEs.<sup>39</sup> Furthermore, the invariant measure for  $d = 2$  is expected to be the YM measure associated with the trivial principal  $G$ -bundle on  $\mathbf{T}^2$  constructed in Refs. 44, 45, and 60. For  $d = 3$ , this would provide the first construction of the YM measure in 3D, even in finite volume.
- Can the analysis in Refs. 11 and 12 be extended to infinite volume  $\mathbf{T}^d \rightsquigarrow \mathbf{R}^d$ ? This is non-trivial even for  $d = 2$ , although the YM measure on  $\mathbf{R}^2$  is arguably simpler.<sup>26</sup>
- Can one extend these results beyond the case that the underlying principal bundle  $P \rightarrow \mathbf{T}^d$  is trivial? For non-trivial principal bundles, one can no longer write connections as globally defined 1-forms, which complicates the solution theory.
- For  $d = 3$ , can one modify the construction of the state space  $\mathcal{S}$  in Sec. III A so that the gauge equivalence  $\sim$  is determined by a gauge group or a similar structure? This would yield a notion of gauge equivalence conceptually closer to the classical space of gauge orbits and would carry a number of technical advantages (see Remarks 1.5 and 3.9 and the start of Sec. III C).
- Taking  $G$  to be one of the classical groups, e.g.,  $G = U(N)$ , what is the behavior of the dynamic as  $N \rightarrow \infty$ ? In 2D, the associated YM measure is known to converge to a deterministic object called the master field,<sup>19–21,46</sup> which is governed by the Makeenko–Migdal equations;<sup>52</sup> see Ref. 47 for a survey. No such result is rigorously known in 3D (the measure at finite  $N$  has not been constructed). It would be interesting if one can use stochastic quantization to recover some of the known results in 2D and obtain new results in 3D; see Ref. 62 where the Langevin dynamic is used to derive the finite  $N$  master loop equation on the lattice.

## ACKNOWLEDGMENTS

It is a great pleasure to thank Ajay Chandra, Martin Hairer, and Hao Shen for many interesting discussions and explorations while carrying out the reviewed works.

## AUTHOR DECLARATIONS

### Conflict of Interest

The author has no conflicts to disclose.

## Author Contributions

Ilya Chevyrev: Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## REFERENCES

- <sup>1</sup>Albeverio, S. and Kusuoka, S., “The invariant measure and the flow associated to the  $\Phi_3^4$ -quantum field model,” *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **20**(4), 1359–1427 (2020).
- <sup>2</sup>Balaban, T., “Ultraviolet stability of three-dimensional lattice pure gauge field theories,” *Commun. Math. Phys.* **102**(2), 255–275 (1985).
- <sup>3</sup>Balaban, T., “Renormalization group approach to lattice gauge field theories. I. Generation of effective actions in a small field approximation and a coupling constant renormalization in four dimensions,” *Commun. Math. Phys.* **109**(2), 249–301 (1987).
- <sup>4</sup>Balaban, T., “Large field renormalization. II. Localization, exponentiation, and bounds for the R operation,” *Commun. Math. Phys.* **122**(3), 355–392 (1989).
- <sup>5</sup>Bern, Z., Halpern, M. B., Sadun, L., and Taubes, C., “Continuum regularization of quantum field theory. II. Gauge theory,” *Nucl. Phys. B* **284**(1), 35–91 (1987).
- <sup>6</sup>Bruned, Y., Chandra, A., Chevyrev, I., and Hairer, M., “Renormalising SPDEs in regularity structures,” *J. Eur. Math. Soc.* **23**(3), 869–947 (2021).
- <sup>7</sup>Bruned, Y., Gabriel, F., Hairer, M., and Zambotti, L., “Geometric stochastic heat equations,” *J. Am. Math. Soc.* **35**(1), 1–80 (2021).
- <sup>8</sup>Bruned, Y., Hairer, M., and Zambotti, L., “Algebraic renormalisation of regularity structures,” *Inventiones Math.* **215**(3), 1039–1156 (2019).
- <sup>9</sup>Cao, S. and Chatterjee, S., “A state space for 3D Euclidean Yang–Mills theories,” [arXiv:2111.12813](https://arxiv.org/abs/2111.12813) (2021).
- <sup>10</sup>Cao, S. and Chatterjee, S., “The Yang–Mills heat flow with random distributional initial data,” [arXiv:2111.10652](https://arxiv.org/abs/2111.10652) (2021).
- <sup>11</sup>Chandra, A., Chevyrev, I., Hairer, M., and Shen, H., “Langevin dynamic for the 2D Yang–Mills measure,” *Publ. Math. l’IHÉS* (published online, 2022).
- <sup>12</sup>Chandra, A., Chevyrev, I., Hairer, M., and Shen, H., “Stochastic quantisation of Yang–Mills–Higgs in 3D,” [arXiv:2201.03487](https://arxiv.org/abs/2201.03487) (2022).
- <sup>13</sup>Chandra, A. and Hairer, M., “An analytic BPHZ theorem for regularity structures,” [arXiv:1612.08138](https://arxiv.org/abs/1612.08138) (2016).
- <sup>14</sup>Chandra, A. and Weber, H., “Stochastic PDEs, regularity structures, and interacting particle systems,” *Ann. Fac. Sci. Toulouse: Math.* **26**(4), 847–909 (2017).
- <sup>15</sup>Charalambous, N. and Gross, L., “The Yang–Mills heat semigroup on three-manifolds with boundary,” *Commun. Math. Phys.* **317**(3), 727–785 (2013).
- <sup>16</sup>Chatterjee, S., “Yang–Mills for probabilists,” in *Probability and Analysis in Interacting Physical Systems*, Springer Proceedings in Mathematics and Statistics Vol. 283 (Springer, Cham, 2019), pp. 1–16.
- <sup>17</sup>Chevyrev, I., “Yang–Mills measure on the two-dimensional torus as a random distribution,” *Commun. Math. Phys.* **372**(3), 1027–1058 (2019).
- <sup>18</sup>Chevyrev, I., “Norm inflation for a non-linear heat equation with Gaussian initial conditions,” [arXiv:2205.14350](https://arxiv.org/abs/2205.14350) (2022).
- <sup>19</sup>Dahlqvist, A. and Lemoine, T., “Large  $N$  limit of the Yang–Mills measure on compact surfaces II: Makeenko–Migdal equations and planar master field,” [arXiv:2201.05886](https://arxiv.org/abs/2201.05886) (2022).

- <sup>20</sup>Dahlqvist, A. and Lemoine, T., “Large N limit of Yang-Mills partition function and Wilson loops on compact surfaces,” [arXiv:2201.05882](#) (2022).
- <sup>21</sup>Dahlqvist, A. and Norris, J. R., “Yang-Mills measure and the master field on the sphere,” *Commun. Math. Phys.* **377**(2), 1163–1226 (2020).
- <sup>22</sup>Del Debbio, L., Patella, A., and Rago, A., “Space-time symmetries and the Yang-Mills gradient flow,” *J. High Energy Phys.* **2013**, 212.
- <sup>23</sup>DeTurck, D. M., “Deforming metrics in the direction of their Ricci tensors,” *J. Differ. Geom.* **18**(1), 157–162 (1983).
- <sup>24</sup>Donaldson, S. K., “Anti self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles,” *Proc. London Math. Soc.* **s3–50**(1), 1–26 (1985).
- <sup>25</sup>Driver, B. K., “Convergence of the  $U(1)_4$  lattice gauge theory to its continuum limit,” *Commun. Math. Phys.* **110**(3), 479–501 (1987).
- <sup>26</sup>Driver, B. K., “YM<sub>2</sub>: Continuum expectations, lattice convergence, and lassos,” *Commun. Math. Phys.* **123**(4), 575–616 (1989).
- <sup>27</sup>Duch, P., “Flow equation approach to singular stochastic PDEs,” [arXiv:2109.11380](#) (2021).
- <sup>28</sup>Federbush, P., “A phase cell approach to Yang-Mills theory. I. Modes, lattice-continuum duality,” *Commun. Math. Phys.* **107**(2), 319–329 (1986).
- <sup>29</sup>Fine, D. S., “Quantum Yang-Mills on a Riemann surface,” *Commun. Math. Phys.* **140**(2), 321–338 (1991).
- <sup>30</sup>Fodor, Z., Holland, K., Kuti, J., Negradi, D., and Wong, C. H., “The Yang-Mills gradient flow in finite volume,” *J. High Energy Phys.* **2012**(11), 7.
- <sup>31</sup>Friz, P. K. and Hairer, M., *A Course on Rough Paths: With an Introduction to Regularity Structures*, 2nd ed., Universitext (Springer, Cham, 2020).
- <sup>32</sup>Gross, L., “Convergence of  $U(1)_3$  lattice gauge theory to its continuum limit,” *Commun. Math. Phys.* **92**(2), 137–162 (1983).
- <sup>33</sup>Gross, L., King, C., and Sengupta, A., “Two-dimensional Yang-Mills theory via stochastic differential equations,” *Ann. Phys.* **194**(1), 65–112 (1989).
- <sup>34</sup>Gubinelli, M. and Hofmanová, M., “A PDE construction of the Euclidean  $\phi_3^4$  quantum field theory,” *Commun. Math. Phys.* **384**(1), 1–75 (2021).
- <sup>35</sup>Gubinelli, M., Imkeller, P., and Perkowski, N., “Paracontrolled distributions and singular PDEs,” *Forum Math., Pi* **3**(e6), 75 (2015).
- <sup>36</sup>Hairer, M., “A theory of regularity structures,” *Inventiones Math.* **198**(2), 269–504 (2014).
- <sup>37</sup>Hairer, M., “Regularity structures and the dynamical  $\Phi_3^4$  model,” in *Current Developments in Mathematics* (International Press, Somerville, MA, 2016), Vol. 2014, pp. 1–49.
- <sup>38</sup>Hairer, M. and Mattingly, J., “The strong Feller property for singular stochastic PDEs,” *Ann. Inst. Henri Poincaré, Probab. Stat.* **54**(3), 1314–1340 (2018).
- <sup>39</sup>Hairer, M. and Schönbauer, P., “The support of singular stochastic partial differential equations,” *Forum Math., Pi* **10**(e1), 1 (2022).
- <sup>40</sup>Hairer, M. and Steele, R., “The  $\Phi_3^4$  measure has sub-Gaussian tails,” *J. Stat. Phys.* **186**(3), 38 (2022).
- <sup>41</sup>Hong, M.-C. and Tian, G., “Global existence of the  $m$ -equivariant Yang-Mills flow in four dimensional spaces,” *Commun. Anal. Geom.* **12**(1), 183–211 (2004).
- <sup>42</sup>Jaffe, A. and Witten, E., “Quantum Yang-Mills theory,” in *The Millennium Prize Problems* (Clay Mathematics Institute, Cambridge, MA, 2006), pp. 129–152.
- <sup>43</sup>Kupiainen, A., “Renormalization group and stochastic PDEs,” *Ann. Henri Poincaré* **17**(3), 497–535 (2016).
- <sup>44</sup>Lévy, T., “Yang-Mills measure on compact surfaces,” *Mem. Am. Math. Soc.* **166**(790), xiv+122 (2003).
- <sup>45</sup>Lévy, T., “Discrete and continuous Yang-Mills measure for non-trivial bundles over compact surfaces,” *Probab. Theory Relat. Fields* **136**(2), 171–202 (2006).
- <sup>46</sup>Lévy, T., “The master field on the plane,” *Astérisque* **388**, ix+201 (2017).
- <sup>47</sup>Lévy, T., “Two-dimensional quantum Yang-Mills theory and the Makeenko-Migdal equations,” in *Frontiers in Analysis and Probability: In the Spirit of the Strasbourg-Zürich Meetings* (Springer, Cham, 2020), pp. 275–325.
- <sup>48</sup>Lévy, T. and Sengupta, A., “Four chapters on low-dimensional gauge theories,” in *Stochastic Geometric Mechanics*, Springer Proceedings in Mathematics and Statistics Vol. 202 (Springer, Cham, 2017), pp. 115–167.
- <sup>49</sup>Lüscher, M., “Properties and uses of the Wilson flow in lattice QCD,” *J. High Energy Phys.* **2010**(8), 71.
- <sup>50</sup>Lyons, T., “Differential equations driven by rough signals (I): An extension of an inequality of L. C. Young,” *Math. Res. Lett.* **1**(4), 451–464 (1994).
- <sup>51</sup>Magnen, J., Rivasseau, V., and Sénéor, R., “Construction of YM<sub>4</sub> with an infrared cutoff,” *Commun. Math. Phys.* **155**(2), 325–383 (1993).
- <sup>52</sup>Makeenko, Y. M. and Migdal, A. A., “Exact equation for the loop average in multicolor QCD,” *Phys. Lett. B* **88**, 135 (1979); Erratum **89**, 437 (1980).
- <sup>53</sup>Migdal, A. A., “Recursion equations in gauge theories,” *Sov. Phys. JETP* **42**, 413 (1975).
- <sup>54</sup>Moinat, A. and Weber, H., “Space-time localisation for the dynamic  $\Phi_3^4$  model,” *Commun. Pure Appl. Math.* **73**(12), 2519–2555 (2020).
- <sup>55</sup>Narayanan, R. and Neuberger, H., “Infinite N phase transitions in continuum Wilson loop operators,” *J. High Energy Phys.* **2006**(03), 064.
- <sup>56</sup>Otto, F. and Weber, H., “Quasilinear SPDEs via rough paths,” *Arch. Ration. Mech. Anal.* **232**(2), 873–950 (2019).
- <sup>57</sup>Parisi, G. and Wu, Y. S., “Perturbation theory without gauge fixing,” *Sci. Sin.* **24**(4), 483–496 (1981).
- <sup>58</sup>Rade, J., “On the Yang-Mills heat equation in two and three dimensions,” *J. Reine Angew. Math.* **1992**(431), 123–164.
- <sup>59</sup>Sadun, L. A., “Continuum regularized Yang-Mills theory,” Ph.D. thesis, University of California, Berkeley, CA, 1987.
- <sup>60</sup>Sengupta, A., “Gauge theory on compact surfaces,” *Mem. Am. Math. Soc.* **126**(600), viii+85 (1997).
- <sup>61</sup>Shen, H., “Stochastic quantization of an Abelian gauge theory,” *Commun. Math. Phys.* **384**(3), 1445–1512 (2021).
- <sup>62</sup>Shen, H., Smith, S. A., and Zhu, R., “A new derivation of the finite N master loop equation for lattice Yang-Mills,” [arXiv:2202.00880](#) (2022).
- <sup>63</sup>Young, L. C., “An inequality of the Hölder type, connected with Stieltjes integration,” *Acta Math.* **67**(1), 251–282 (1936).
- <sup>64</sup>Zwanziger, D., “Covariant quantization of gauge fields without Gribov ambiguity,” *Nucl. Phys. B* **192**(1), 259–269 (1981).
- <sup>65</sup>In 3D, one needs to work with a slightly stronger metric  $\tilde{\Sigma}$  in (1.6); see Sec. III C.
- <sup>66</sup>To avoid arguing that (3.2) is global in time solutions, the definition of  $\mathcal{F}_t$  in Ref. 12 is restricted to short intervals  $t \in (0, T)$ —see Proposition 3.4—and items (a) and (b) should also be understood locally in time. To simplify the exposition, we ignore this detail here.