Karlsruhe Institute of Technology

Wave phenomena
analysis and numerics

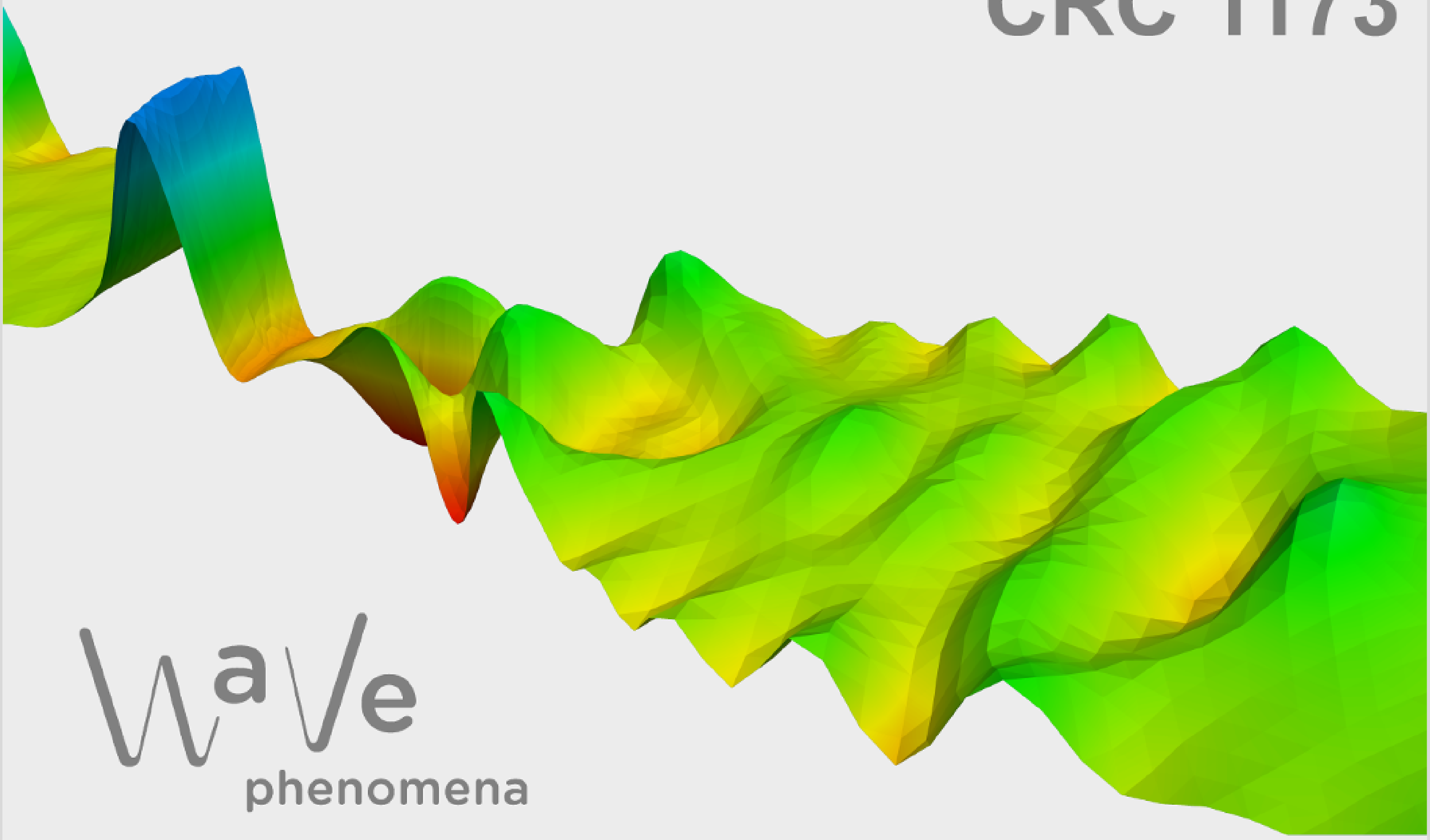
Sparse grid approximation of stochastic parabolic PDEs: The Landau–Lifshitz–Gilbert equation

Xin An, Josef Dick, Michael Feischl, Andrea Scaglioni, Thanh Tran

CRC Preprint 2024/8, March 2024

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

UNIVERSITÄT BONN

Universität Stuttgart

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

TU WIEN

Funded by

DFG

ISSN 2365-662X

Sparse grid approximation of stochastic parabolic PDEs: The Landau–Lifshitz–Gilbert equation *

Xin An † Josef Dick ‡ Michael Feischl § Andrea Scaglioni ‡ Thanh Tran †

10.03.2024

Abstract

We show convergence rates for a sparse grid approximation of the distribution of solutions of the stochastic Landau-Lifshitz-Gilbert equation. Beyond being a frequently studied equation in engineering and physics, the stochastic Landau-Lifshitz-Gilbert equation poses many interesting challenges that do not appear simultaneously in previous works on uncertainty quantification: The equation is strongly non-linear, time-dependent, and has a non-convex side constraint. Moreover, the parametrization of the stochastic noise features countably many unbounded parameters and low regularity compared to other elliptic and parabolic problems studied in uncertainty quantification. We use a novel technique to establish uniform holomorphic regularity of the parameter-to-solution map based on a Gronwall-type estimate and the implicit function theorem. This method is very general and based on a set of abstract assumptions. Thus, it can be applied beyond the Landau-Lifshitz-Gilbert equation as well. We demonstrate numerically the feasibility of approximating with sparse grid and show a clear advantage of a multi-level sparse grid scheme.

Keywords: Stochastic and parametric PDEs, stochastic Landau-Lifshitz-Gilbert problem, Doss-Sussmann transform, Lévy-Ciesielski expansion, regularity of sample paths solution, curse of dimensionality, implicit function theorem, holomorphy and sparsity of parameter-to-solution map, piecewise polynomials, sparse high-dimensional approximation, sparse grid, stochastic collocation, dimension independent convergence, multilevel sparse grid

MSCcodes 35R60, 47H40, 65C30, 60H25, 60H35, 65M15

1 Introduction

While the methods developed in the present work are fairly general and apply to different model problems, we focus on the specific task of approximating the stochastic Landau-Lifshitz-Gilbert equation as it contains many of the difficulties one encounters in nonlinear and stochastic partial differential equations.

The Landau-Lifshitz-Gilbert (LLG) equation is a phenomenological model for the dynamic evolution of the magnetization in ferromagnetic materials. In order to capture heat fluctuations of the magnetization one considers a stochastic extension of the LLG equation driven by stochastic noise, see e.g., [10, 39] for some of the first works devoted to the modelling of magnetic materials under thermal agitation. Following these early works, great interest in the physics community lead to extensive research, see e.g., [9, 31, 37, 42, 52] to name a few examples.

The present work gives a first efficient approximation of the probability distribution of the solution of the stochastic LLG equation. To that end, we employ the Doss-Sussmann transform and discretize the resulting Wiener process via a Lévy-Ciecierski expansion. This leads to a parametrized nonlinear time-dependent PDE with infinite dimensional and unbounded parameter space which can be approximated by using sparse grid techniques. We derive the following main results:

- The first rigorous convergence result for an approximation of a nonlinear and time-dependent parametric coefficient PDE with unbounded parameter space. Precisely, we show convergence of piecewise quadratic sparse grids for the stochastic LLG equation with order 1/2 and dimension dependent constant (see

*XA was fully supported by the Australian Research Council (ARC) grant DP190101197, JD was partially funded by ARCDP190101197, ARCDP220101811, MF and AS were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 258734477 – SFB 1173 as well as the Austrian Science Fund (FWF) under the special research program Taming complexity in PDE systems (grant SFB F65), and TT was partially supported by ARCDP190101197, ARCDP200101866, and ARCDP220101811.

†Cancer Council NSW, Sydney, Australia (xin.an@nswcc.org.au)

‡School of Mathematics & Statistics, UNSW, Sydney, Australia (josef.dick@unsw.edu.au, thanh.tran@unsw.edu.au).

§Institute of Analysis and Scientific Computing, TU Wien, Vienna, Austria (michael.feischl@asc.tuwien.ac.at, andrea.scaglioni@asc.tuwien.ac.at)

Theorem 27). The result assumes that the stochastic LLG equation has uniformly Hölder (in time and space) regular solutions which is true for regular initial conditions which are sufficiently close to constant. Under some reasonable assumptions and simplifications of the stochastic input, we show dimension independent convergence with order 1/2 (see Theorem 31).

- The first result on uniform holomorphic regularity of the parameter-to-solution map for the Landau-Lifshitz-Gilbert equation. To the best of our knowledge, this is also the first uniform holomorphic regularity result for unbounded parameter spaces and strongly nonlinear and time-dependent problems.
- Improved convergence rate of a multi-level version of the stochastic collocation algorithm under natural assumptions on the underlying finite element method.

To achieve the above-mentioned results, we have to overcome several challenges posed by the nonlinear nature of the problem:

- Holomorphic parameter-to-solution map: This is well-understood for linear problems but turns out to be technically challenging for nonlinear problems. While we apply the implicit function theorem as in [18], in our case the parameter space is not compact. To overcome this problem, we control the growth of the extension by means of a Gronwall-like estimate for small imaginary parts. The main challenge here is that there is no canonical complex version of LLG which supports holomorphy. The main reason for this is that any extension of the cross product is either not complex differentiable or loses orthogonality properties which normally ensure L^∞-boundedness of solutions of the LLG equation.
- Lack of parametric regularity: All mentioned works on uncertainty quantification require strong summability of the coefficients which arise in the expansion of the stochastic noise. Typically, ℓ^p-summability with $p < 1$ is required. Even with the holomorphic regularity established, the present problem only provides summability in ℓ^p for $p > 2$. We propose a simplification of the stochastic input which allows us to consider the problem in an L^1-setting in time. This increases the parametric regularity and results in dimension independent estimates.
- Lack of sample path regularity: Regularity results for LLG are sparse even in the deterministic setting. We refer to [16, 17, 19, 41, 44, 43, 45] for partial results in 2D and 3D. Sample path regularity directly influences holomorphic regularity via the implicit function theorem. To that end, we rely on Hölder space regularity results for the stochastic LLG equation (Theorem 8).

1.1 Related work on the numerics of the LLG equation

The nonlinear nature of LLG combined with the stochastic noise attracted a lot of interest in numerical analysis: For the deterministic version of LLG, weak convergence of some time stepping schemes was known since at least 2008 (see, e.g., the midpoint scheme [7] and the tangent-plane scheme [2]). It took another ten years to obtain strong a priori convergence of uniform time stepping schemes that obey physical energy bounds, which has first been proved in [28] and was then extended to higher-order in [1]. The latter two works build on the tangent plane idea first introduced in [2] in order to remove the nonlinear solver required in [7]. This is achieved by solving for the time derivative of the magnetization instead of the magnetization itself.

To study the stochastic version of LLG (SLLG), [12, 13] formulate a rigorous definition of *weak martingale solution* to the SLLG problem, prove existence by means of the Faedo-Galerkin method and discuss regularity even with anisotropy in the effective field and for finite multi-dimensional noise in space. In [11, 14], the authors study the 1D (in space) SLLG problem, which has applications in the manufacturing of nanowires. They prove existence of weak martingale solutions for the problem for a larger class of coefficients compared to previous works in 3D. The works also show pathwise existence and uniqueness of strong solutions and a large deviation principle. This is then used to analyse the transitions between equilibria. The space and time approximation of the SLLG problem was considered in [8]. The authors consider an implicit midpoint scheme that preserves the unit modulus constraint on the magnetization and satisfies relevant discrete energy estimates. Then they prove, by a compactness argument, that the method converges almost surely and weakly to the exact solution, up to extraction of a subsequence. In the follow-up work [6], the scheme is applied to reproduce physically relevant phenomena such as finite-time blow-up of the solution and thermally-activated switching. A different approach is followed in [34], where the authors propose to discretize SLLG in space and time by first applying the Doss-Sussmann transform [24, 54] to the SLLG problem to obtain a random coefficient LLG problem. They then discretize this problem using the tangent-plane scheme [2] and prove convergence (again in the sense of weak convergence of a subsequence), which in particular proves that the random coefficient LLG problem is well posed. A tangent plane scheme is also considered in [3], where the sample paths of the SPDE in Ito form are approximated and stability and convergence results are derived. For the approximation of multi-dimensional (finite) noise, [33] generalizes the approach based on the Doss-Sussmann transform.

1.2 Related work on the approximation of PDEs with random coefficients

Dimension independent approximation of PDEs with random coefficients has first been proposed in [20] and the idea of using a holomorphic extension of the exact solution in order to obtain convergence rates for the parametric approximation goes back to [21].

The works [57] and [4] begin the mathematical study of collocation-type schemes for random coefficient PDEs. Several extensions and improvements of certain aspects of the theory can be found in, e.g., [49], which uses sparse grids to improve the dependence on the number of parametric dimensions and [48], which employs *anisotropic* sparse grids to achieve dimension independent convergence (under tractability assumptions on the problem). In [46], the authors select the sparse grids with a profit-maximization principle, effectively recasting the sparse grid selection problem into a Knapsack problem. They also prove an error bound with explicit dependence on the number of approximated dimensions. In the present work, we use the same principle to build sparse grids and apply the framework to prove dimension independent convergence.

In [58], the authors extend the methodology developed in [4] and the related papers to a linear parabolic problem with random coefficients under the finite-dimensional noise assumption. They prove existence of a holomorphic extension and, based on the ideas in [4], show that this leads to convergence of stochastic collocation schemes for both a space semi-discrete and fully discrete approximations. In [47], the authors study a linear parabolic problem with uncertain diffusion coefficient under the finite dimensional noise assumption. They prove existence of a holomorphic extension by extending the problem to complex parameters and verifying the Cauchy-Riemann equations. They study convergence of stochastic Galerkin and stochastic collocation approximation. In [30], the authors consider a coupled Navier-Stokes and heat equation problem with uncertainty and developed a heuristic adaptive sparse grid scheme based on [32].

In order to discretize the Wiener process in the stochastic LLG equation, one needs to deal with unbounded parameter spaces. This has been done in, e.g., [5], where the authors study the Poisson problem with lognormal diffusion and establish summability results for Hermite coefficients based on *local-in-space* summability of the basis used to expand the logarithm of the diffusion. In [27], the authors approximate functions with this property by means of sparse grids interpolation built using global polynomials with Gauss-Hermite interpolation nodes. They prove algebraic and dimension independent convergence rates.

In the monograph [26], the authors study the regularity of a large class of problems depending on Gaussian random field inputs as well as the convergence of several numerical schemes. Several examples of PDEs with Gaussian random coefficients are given e.g. elliptic and parabolic PDEs with lognormal diffusion. The regularity result implies estimates on the Hermite coefficients of the parameter-to-solution map. These, in turn, can be used to study the convergence of Smolyak-Hermite interpolation and quadrature among other numerical methods.

Beyond linear problems, in [18] the authors deal with infinite-dimensional parametric problems with compact coefficient spaces, but go beyond the setting of affine parametric dependence. They prove the existence of a holomorphic extension of the coefficient-to-solution map without extending the problem to the complex domain (as is usually done for the random Poisson problem). Rather, they employ the implicit function theorem. In [22], the authors use similar techniques in the setting of the stationary Navier-Stokes equation with random domain.

1.3 Structure of the work

In Section 2 we introduce a general framework for the study of the parametric regularity of solutions of SPDEs. We first explain in Section 2.1 how to reduce a SPDE to a parametric coefficients PDE. Then, in Sections 2.2 and 2.3 we prove that the parameter-to-solution map admits a sparse holomorphic extension. The result is based on four main assumptions that have to be proved for each concrete problem. Finally, we estimate the derivatives of the parameter-to-solution map with Cauchy's integral theorem.

We introduce the stochastic version of the LLG equation in Section 3, and, following the general strategy from Section 2.1, transform it into a parametric nonlinear and time-dependent PDE in Section 4. In Section 4.1, we prove that the solution's sample paths are Hölder-continuous under regularity assumption on the problem data. We also prove that they are uniformly bounded with respect to the Wiener process sample paths.

In Section 5, we apply the regularity analysis from Sections 2.2 and 2.3 to the parametric LLG problem and prove that the parameter-to-solution map is holomorphic under the assumptions that sample paths of random coefficients and solutions are Hölder continuous.

In Section 6, we do the same for a simplified version of the parametric LLG problem obtained with additional modelling assumptions. This time, sample paths are assumed to be Lebesgue integrable in time.

The sparsity properties of the parameter-to-solution map in the Hölder setting are weaker than in the Lebesgue setting. This is reflected by the convergence of sparse grid interpolation discussed in Section 7. The results are confirmed by numerical experiments.

The final Section 8 derives the multi-level version of the stochastic collocation method and provides numerical tests.

2 General approach to deriving parametric regularity of an SPDE

In the present section, we outline a fairly general strategy to prove a regularity property of solutions of stochastic partial differential equations (SPDE) driven by the Wiener process. The resulting regularity properties can be used to tailor sparse grid approximation methods to the problem. The arguments presented in this section are formal and need to be verified for each concrete problem. The most important assumptions are listed explicitly below.

2.1 Reduction to a parametric problem

Consider a spatial domain $D \subset \mathbb{R}^d$ of dimension $d \in \mathbb{N}$ and a final time $T > 0$. Denote by ∂D the boundary and by ∂_n the unit exterior normal derivative. The space-time cylinder is denoted by $D_T := [0,T] \times D$. Consider the initial condition $U^0 : D \to \mathbb{R}^m$ for $m \in \mathbb{N}$, a drift coefficient $\mu : \mathbb{R}^m \times [0,T] \times D \to \mathbb{R}^m$ and a noise coefficient $\sigma : \mathbb{R}^m \times D \to \mathbb{R}^m$. While a more general noise coefficient can be treated with analogous techniques, we consider this simple case as it is sufficient for the examples below. Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider the SPDE problem: Find a random field $U : \Omega \times D_T \to \mathbb{R}^m$ such that, $\mathbb{P}$-a.s.

$$\begin{cases} \mathrm{d}U = \mu(U, t, \boldsymbol{x})\mathrm{d}t + \sigma(U, \boldsymbol{x}) \circ \mathrm{d}W(t) & \text{on } D_T \\ \partial_n U = 0 & \text{on } [0,T] \times \partial D \\ U(\cdot, 0, \cdot) = U^0 & \text{on } D, \end{cases}$$

where by $\circ \mathrm{d}W(t)$ we denote the Stratonovich differential applied to a Wiener process W.

The *Doss-Sussmann transform* [24, 54] of U is, by definition,

$$u = e^{-W\sigma} U, \tag{1}$$

i.e. the exponential of the operator $-W\sigma$ applied to U. The result is a random field $u : \Omega \times [0,T] \times D \to \mathbb{R}^m$ that can be shown to satisfy the *random coefficient partial differential equation* (PDE)

$$\mathcal{R}(W(\omega), u(\omega)) = 0 \qquad \text{in } R,\ \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{2}$$

The *residual operator* $\mathcal{R} : \mathbb{W}_{\mathbb{R}} \times \mathbb{U}_{\mathbb{R}} \to R$ is defined for Banach spaces $\mathbb{W}_{\mathbb{R}}, \mathbb{U}_{\mathbb{R}}$ and R. In general, it is a differential operator in time and space with respect to $u \in \mathbb{U}_{\mathbb{R}}$ while it does not contain Itô or Stratonovich differentials of W.

In order to make the distribution of u amenable to *approximation*, we need to parametrize the Brownian motion. It turns out that a local wavelet-type expansion of W is very beneficial as it reduces the number of active basis function at any given moment in time. The Lévy-Ciesielski expansion (LCE) (see e.g. [35, Section 4.2]) of the Brownian motion $W : \Omega \times [0,1] \to \mathbb{R}$ reads

$$W(\omega, t) = \sum_{\ell=0}^{\infty} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} Y_{\ell,j}(\omega) \eta_{\ell,j}(t),$$

where $Y_{\ell,j}$ are independent standard normal random variables and
$\left\{\eta_{\ell,j} : \ell \in \mathbb{N}_0, j = 1, \ldots, \lceil 2^{\ell-1} \rceil\right\}$ is the *Faber-Schauder* hat-function basis (see Figure 1) on $[0,1]$, i.e.,

$$\begin{aligned} &\eta_{0,1}(t) = t, \\ &\eta(t) := \begin{cases} t & t \in [0, \frac{1}{2}] \\ 1 - t & t \in [\frac{1}{2}, 1] \\ 0 & \text{otherwise} \end{cases}, \\ &\eta_{\ell,j}(t) = 2^{-\frac{\ell-1}{2}} \eta\left(2^{\ell-1}t - j + 1\right) \qquad \text{for all } \ell \in \mathbb{N},\ j = 1, \ldots, 2^{\ell-1}. \end{aligned} \tag{3}$$

Observe that $\|\eta_{0,1}\|_{L^\infty(0,1)} = 1$, $\operatorname{supp} \eta_{0,1} = (0,1]$ and $\|\eta_{\ell,j}\|_{L^\infty(0,1)} = 2^{-(\ell+1)/2}$, $\operatorname{supp} \eta_{\ell,j} = \left(\frac{j-1}{2^{\ell-1}}, \frac{j}{2^{\ell-1}}\right)$ for all $\ell \in \mathbb{N}, j = 1, \ldots, 2^{\ell-1}$. The LCE converges uniformly in t, almost surely to a continuous function which coincides with the Brownian motion everywhere (see [53, Section 3.4]):

$$\lim_{L \to \infty} \sup_{t \in [0,1]} \left| W(\omega, t) - \sum_{\ell=0}^{L} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} Y_{\ell,j}(\omega) \eta_{\ell,j}(t) \right| = 0 \qquad \mathbb{P}\text{-a.e. } \omega \in \Omega.$$

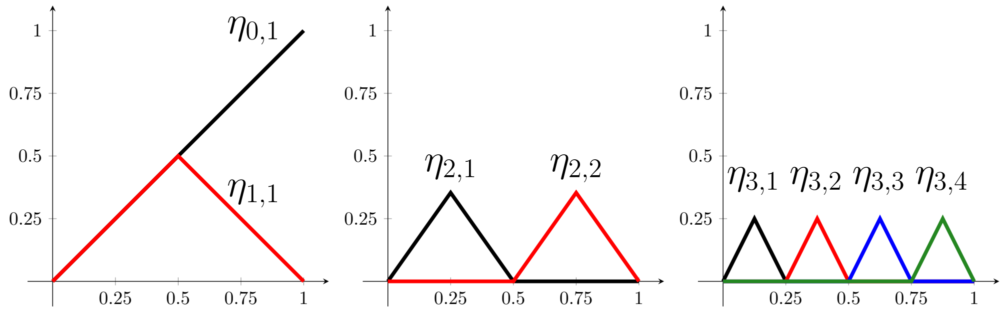


Figure 1: The first eight Faber-Schauder basis functions on $[0,1]$.

By equipping $\mathbb{R}$ with the Gaussian measure μ, we may parametrize W as $W : \mathbb{R}^{\mathbb{N}} \times [0,1] \to \mathbb{R}$ so that

$$W(\boldsymbol{y}, t) = \sum_{\ell=0}^{\infty} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} y_{\ell,j} \eta_{\ell,j}(t), \tag{4}$$

where $y_{\ell,j} \in \mathbb{R}$ for all $\ell \in \mathbb{N}_0, j = 1, \ldots, \lceil 2^{\ell-1} \rceil$. For $L \in \mathbb{N}_0$, we define the *level-L truncation of W* by $W_L(\boldsymbol{y}, t) = \sum_{\ell=0}^{L} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} y_{\ell,j} \eta_{\ell,j}(t)$. We will sometimes also index the same sum as $W_L(\boldsymbol{y}, t) = \sum_{n=0}^{N} y_n \eta_n(t)$. The two indexing systems, hierarchical and linear, are related via

$$\eta_{\ell,j} = \eta_n \quad \iff \quad n = \lfloor 2^{\ell-1} \rfloor + j - 1. \tag{5}$$

We note that the total number of parameters is $N = \sum_{\ell=0}^{L} \lceil 2^{\ell-1} \rceil = 1 + \sum_{\ell=1}^{L} 2^{\ell-1} = 2^L$.

The fact that the parameter domain is unbounded requires the use of appropriate collocation nodes, a topic we treat in Section 7, below.

We denote by $\mathcal{X}_{\mathbb{R}}$ an appropriate space of real sequences such that if $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$, then $W(\boldsymbol{y}, \cdot)$ belongs to a desired Banach space of functions. Denote by $\boldsymbol{\mu}$ the Gaussian measure on $\mathcal{X}_{\mathbb{R}}$.

Example 1. *Consider the Banach space of sequences:*

$$\mathcal{X}_{\mathbb{R}} := \left\{ \boldsymbol{y} = (y_n)_{n \in \mathbb{N}} \subset \mathbb{R} \, : \, \|\boldsymbol{y}\|_{\mathcal{X}_{\mathbb{R}}} < \infty \right\}, \; \|\boldsymbol{y}\|_{\mathcal{X}_{\mathbb{R}}} := |y_{0,1}| + \sum_{\ell \in \mathbb{N}} \max_{j=1,\ldots,2^{\ell-1}} |y_{\ell,j}| \, 2^{-(\ell+1)/2}.$$

Simple computations show that if $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$, then $\|W(\boldsymbol{y})\|_{L^\infty(0,T)} \le \|\boldsymbol{y}\|_{\mathcal{X}_{\mathbb{R}}}$, thus $\mathbb{W} \subset L^\infty(0,T)$.

Assume without loss of generality that $T = 1$. By substituting the random field $W(\omega, t)$ in the random coefficient PDE (2) with the parametric expansion (4), we obtain a *parametric coefficient PDE*: Find $u : \mathcal{X}_{\mathbb{R}} \times D_T \to \mathbb{R}^m$ such that

$$\mathcal{R}(W(\boldsymbol{y}), u(\boldsymbol{y})) = 0 \qquad \text{in } R, \; \boldsymbol{\mu}\text{-a.e. } \boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}. \tag{6}$$

2.2 Holomorphic regularity of the solution operator

While holomorphic parameter regularity of random elliptic equations is well-known by now (see, e.g., [4, Section 3], for the case of bounded or unbounded parameter spaces under the finite dimensional noise assumption, [21] for countably-many parameters taking values on tensor product of bounded intervals, [5], for a discussion of the Poisson problem with lognormal coefficients, in which the authors study countably many unbounded parameters), the literature is much sparser for nonlinear and time-dependent problems. In this section, we follow an approach from [18] which uses the implicit function theorem to obtain analyticity. While the authors in [18] can rely on a compact parameter domain to ensure a non-trivial domain of extension, we have to use intricate bounds on the parametric gradient of the solution. A recent result on the implicit function theorem for Gevrey regularity [36] could also be used to achieve similar results in a less explicit fashion.

We require some assumptions to work in a more general setting.

Assumption 1. *For any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$ there exists $u(\boldsymbol{y}) \in \mathbb{U}_{\mathbb{R}}$ such that $\mathcal{R}(W(\boldsymbol{y}), u(\boldsymbol{y})) = 0$ in R. Moreover, there exists $C_r > 0$ such that, for any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$, $\|u(\boldsymbol{y})\|_{\mathbb{U}_{\mathbb{R}}} \le C_r$.*

Assumption 2. *The residual operator $\mathcal{R} : \mathbb{W}_{\mathbb{R}} \times \mathbb{U}_{\mathbb{R}} \to R$ admits an extension to* complex *Banach spaces $\mathbb{W} \supset \mathbb{W}_{\mathbb{R}}$ and $\mathbb{U} \supset \mathbb{U}_{\mathbb{R}}$. The extended map $\mathcal{R} : \mathbb{W} \times \mathbb{U} \to R$ satisfies the following properties:*

(i) $\mathcal{R}$ *is continuously differentiable;*

(ii) $\partial_u \mathcal{R}(W, u) : \mathbb{U} \to R$ *is a homeomorphism for all* $(W, u) \in \mathbb{W}_\mathbb{R} \times \mathbb{U}_\mathbb{R}$ *such that* $\mathcal{R}(W, u) = 0$.

With this complex extension in mind, in the following for any $W_0 \in \mathbb{W}_\mathbb{R}$ and $u_0 \in \mathbb{U}_\mathbb{R}$ we denote, for $\varrho > 0$,

$$\begin{aligned} B_\varrho(W_0) &:= \{W \in \mathbb{W} : \|W - W_0\|_\mathbb{W} < \varrho\}, \\ B_\varrho(u_0) &:= \{u \in \mathbb{U} : \|u - u_0\|_\mathbb{U} < \varrho\}. \end{aligned} \tag{7}$$

Let us recall the implicit function theorem for maps between Banach spaces (see, e.g., [23, Theorem 10.2.1]).

Theorem 2 (Implicit function). *Let* E, F, G *be Banach spaces,* $A \subset E \times F$ *and* $f : A \to G$ *be a continuously differentiable function. Let* $(x_*, y_*) \in A$ *be such that* $f(x_*, y_*) = 0$ *and the partial derivative* $D_2 f(x_*, y_*)$ *is a linear homeomorphism from* F *onto* G*. Then, there exists a neighbourhood* U_* *of* x_* *in* E *such that, for every open connected neighbourhood* U *of* x_* *in* U_**, there exists a unique continuous mapping* $\mathcal{U} : U \to F$ *such that* $\mathcal{U}(x_*) = y_*$*,* $(x, \mathcal{U}(x)) \in A$ *and* $f(x, \mathcal{U}(x)) = 0$ *for any* x *in* U*. Moreover,* $\mathcal{U}$ *is continuously differentiable in* U *and its derivative is given by*

$$\mathcal{U}'(x) = -\left(D_2 f(x, \mathcal{U}(x))\right)^{-1} \circ \left(D_1 f(x, \mathcal{U}(x))\right) \qquad \text{for all } x \in U. \tag{8}$$

Invoking Theorem 2 for the operator $\mathcal{R} : \mathbb{W} \times \mathbb{U} \to R$, with $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$ and $u(\boldsymbol{y}) \in \mathbb{U}_\mathbb{R}$ satisfying $\mathcal{R}(W(\boldsymbol{y}), u(\boldsymbol{y})) = 0$, there exists $\varepsilon(\boldsymbol{y}) > 0$ and a holomorphic map $\mathcal{U} : B_{\varepsilon(\boldsymbol{y})}(W(\boldsymbol{y})) \to \mathbb{U}$ such that $\mathcal{U}(W(\boldsymbol{y})) = u(\boldsymbol{y})$ and $\mathcal{R}(W, \mathcal{U}(W)) = 0$ for all $W \in B_{\varepsilon(\boldsymbol{y})}(W(\boldsymbol{y}))$ (cf. (7)).

For any $W \in B_{\varepsilon(\boldsymbol{y})}(W(\boldsymbol{y}))$, the differential $\mathcal{U}'(W)$ belongs to $\mathcal{L}(\mathbb{W}, \mathbb{U})$, the set of linear bounded operator from $\mathbb{W}$ into $\mathbb{U}$ equipped with the usual norm.

Recalling definition (7), we make additional assumptions on the regularity of the derivatives of the residual operator $\mathcal{R}$.

Assumption 3. *There exist* $\varepsilon_W, \varepsilon_u > 0$ *such that for any* $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$ *and any* $W \in B_{\varepsilon_W}(W(\boldsymbol{y}))$ *with* $\mathcal{U}(W) \in B_{\varepsilon_u}(\mathcal{U}(W(\boldsymbol{y})))$*, the operator* $\partial_W \mathcal{R}(W, \mathcal{U}(W))$ *is well-defined and* $\partial_u \mathcal{R}(W, \mathcal{U}(W))$ *is homeomorphic with*

$$\begin{aligned} \|\partial_W \mathcal{R}(W, \mathcal{U}(W))\|_{\mathcal{L}(\mathbb{W}, R)} &\le \mathcal{G}_1(\|\mathcal{U}(W)\|_\mathbb{U}), \\ \left\|\partial_u \mathcal{R}(W, \mathcal{U}(W))^{-1}\right\|_{\mathcal{L}(R, \mathbb{U})} &\le \mathcal{G}_2(\|\mathcal{U}(W)\|_\mathbb{U}), \end{aligned}$$

where the functions $\mathcal{G}_1, \mathcal{G}_2$ *are continuous and may depend on problem coefficients and* ε_u*,* ε_W *but depend on* W *and* $\mathcal{U}(W)$ *only through* $\|\mathcal{U}(W)\|_\mathbb{U}$ *and are independent of* $\boldsymbol{y}$*.*

Together with (8) from Theorem 2, this assumption implies the existence of a continuous increasing function $\mathcal{G} = \mathcal{G}(\|\mathcal{U}(W)\|_\mathbb{U}) > 0$ such that

$$\|\mathcal{U}'(W)\|_{\mathcal{L}(\mathbb{W}, \mathbb{U})} \le \mathcal{G}(\|\mathcal{U}(W)\|_\mathbb{U}) \qquad \text{for all } W \in B_{\min(\varepsilon(\boldsymbol{y}), \varepsilon_W)}(W(\boldsymbol{y})). \tag{9}$$

2.3 Uniform holomorphic extension of solution operator

Since we cannot rely on a compact parameter domain, we show existence of a uniformly bounded holomorphic extension through the application of a generalized version of Gronwall's lemma.

As in the previous section, fix $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$. We can assume, without loss of generality, that $\varepsilon(\boldsymbol{y}) \le \varepsilon_W$.

Definition 3. *We consider an open set* $\mathcal{H}(\boldsymbol{y}) \subseteq B_{\varepsilon_W}(W(\boldsymbol{y}))$ *with the following properties:*

- $B_{\varepsilon(\boldsymbol{y})}(W(\boldsymbol{y})) \subseteq \mathcal{H}(\boldsymbol{y})$*,*
- $\mathcal{U}(W) \in B_{\varepsilon_u}(\mathcal{U}(W(\boldsymbol{y})))$ *for all* $W \in \mathcal{H}(\boldsymbol{y})$*,*
- *the solution operator* $\mathcal{U} : B_{\varepsilon(\boldsymbol{y})}(W(\boldsymbol{y})) \to \mathbb{U}$ *extends holomorphically to* $\mathcal{H}(\boldsymbol{y})$*,*
- *for all* $W \in \mathcal{H}(\boldsymbol{y})$ *we have* $\sigma W + (1 - \sigma) W(\boldsymbol{y}) \in \mathcal{H}(\boldsymbol{y})$ *for all* $0 \le \sigma \le 1$*.*

In contrast to [18], this domain of real parameters $\mathbb{W}_\mathbb{R}$ may not be compact. Therefore, $\varepsilon(\boldsymbol{y})$ can be arbitrarily small and hence $\mathcal{H}(\boldsymbol{y})$ might become very small for certain parameters $\boldsymbol{y}$. The goal of the arguments below is to show that there exists $\varepsilon > 0$ such that for all $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$ $\mathcal{H}(\boldsymbol{y}) = B_\varepsilon(W(\boldsymbol{y}))$ is a valid choice. Instead of relying on compactness, we exploit estimate (9) through the following nonlinear generalization of Gronwall's lemma:

Lemma 4 ([25], Theorem 27)**.** *Let $0 \le c \le d < \infty$, $\varphi : [c,d] \to \mathbb{R}$ and $k : [c,d] \to \mathbb{R}$ be positive continuous functions on $[c,d]$ and let a, b be non-negative constants. Further, let $\mathcal{G} : [0,\infty) \to \mathbb{R}$ be a positive non-decreasing function. If*

$$\varphi(t) \le a + b \int_c^t k(s)\mathcal{G}(\varphi(s))\mathrm{d}s \quad \textit{for all } t \in [c,d],$$

then

$$\varphi(t) \le G^{-1}\left(G(a) + b\int_c^t k(s)\mathrm{d}s\right) \quad \textit{for all } c \le t \le d_1 \le d$$

where for $0 < \xi < \lambda$,

$$\mathcal{G}(\lambda) := \int_\xi^\lambda \frac{\mathrm{d}s}{\mathcal{G}(s)} \tag{10}$$

and d_1 is defined such that $G(a) + b\int_c^t k(s)\mathrm{d}s$ belongs to the domain of G^{-1} for $t \in [c, d_1]$.

Theorem 5. *Assume the validity of Assumptions 1, 2, and 3. With $C_r > 0$ given in Assumption 1, choose $0 < \varepsilon < \varepsilon_W$ such that $G(C_r) + \varepsilon$ belongs to the domain of G^{-1} (where G is defined in* (10) *with the corresponding $\mathcal{G}$ given in* (9)*). Then, ε is independent of $\boldsymbol{y}$ and $\mathcal{H}(\boldsymbol{y})$ from Definition 3 can be chosen as $\mathcal{H}(\boldsymbol{y}) = B_\varepsilon(W(\boldsymbol{y}))$ for all $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$. Moreover, $\mathcal{U}$ is uniformly bounded on $B_\varepsilon(W(\boldsymbol{y}))$ by a constant $C_\varepsilon > 0$ that depends only on ε.*

Proof. Fix $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$.

Step 1: We first show that $\mathcal{U}$ is uniformly bounded on $\mathcal{H}(\boldsymbol{y}) \cap B_\varepsilon(W(\boldsymbol{y}))$. To that end, fix $W \in \mathcal{H}(\boldsymbol{y}) \cap B_\varepsilon(W(\boldsymbol{y}))$ and let $W_\sigma := \sigma W + (1-\sigma)W(\boldsymbol{y})$ for any $0 \le \sigma \le 1$. We define $\varphi : [0,1] \to \mathbb{U}$ by $\varphi(\sigma) = \mathcal{U}(W_\sigma)$. Since by definition $\mathcal{U}$ is differentiable in $\mathcal{H}(\boldsymbol{y})$, we may apply the fundamental theorem of calculus to obtain

$$\varphi(t) - \varphi(s) = \int_s^t \mathcal{U}'(W_\sigma)[W - W(\boldsymbol{y})]\mathrm{d}\sigma \quad \text{for all } s,t \in [0,1]. \tag{11}$$

In particular, with $s = 0$, the triangle inequality yields, recalling that $W \in B_\varepsilon(W(\boldsymbol{y}))$,

$$\|\varphi(t)\|_\mathbb{U} \le \|\varphi(0)\|_\mathbb{U} + \varepsilon \int_0^t \|\mathcal{U}'(W_\sigma)\|_{\mathcal{L}(\mathbb{W},\mathbb{U})}\,\mathrm{d}\sigma \quad \text{for all } 0 \le t \le 1.$$

Assumption 1 and estimate (9) (consequence of Assumption 3) imply the estimate

$$\|\varphi(t)\|_\mathbb{U} \le C_r + \varepsilon \int_0^t \mathcal{G}\left(\|\varphi(\sigma)\|_\mathbb{U}\right)\mathrm{d}\sigma \quad \text{for all } 0 \le t \le 1.$$

Apply Lemma 4 to conclude (note that, in the notation of Lemma 4, we have $d_1 = d = 1$ because of the definition of ε as well as $k(s) = 1$)

$$\|\varphi(t)\|_\mathbb{U} \le G^{-1}(G(C_r) + \varepsilon t) \le G^{-1}(G(C_r) + \varepsilon) \quad \text{for all } 0 \le t \le 1. \tag{12}$$

Since $\|\mathcal{U}(W)\|_\mathbb{U} = \|\varphi(1)\|_\mathbb{U} \le C_\varepsilon$, where $C_\varepsilon := G^{-1}(G(C_r) + \varepsilon)$, we derive the uniform boundedness of $\mathcal{U}$ on $\mathcal{H}(\boldsymbol{y})$. Note that this bound is independent of $\boldsymbol{y}$ and $\mathcal{H}(\boldsymbol{y})$.

Step 2: We next show that φ defined in Step 1 is Lipschitz on $[0,1]$. Equation (11) implies, for $0 \le s < t \le 1$,

$$\begin{aligned}\|\varphi(t) - \varphi(s)\|_\mathbb{U} &\le \int_s^t \|\mathcal{U}'(W_\sigma)\|_{\mathcal{L}(\mathbb{W},\mathbb{U})}\|W - W(\boldsymbol{y})\|_\mathbb{W}\,\mathrm{d}\sigma \\ &\le \int_s^t \mathcal{G}\left(\|\varphi(\sigma)\|_\mathbb{U}\right)\|W - W(\boldsymbol{y})\|_\mathbb{W}\,\mathrm{d}\sigma.\end{aligned}$$

The desired results then follows from (12).

Step 3: We can without loss of generality assume that $0 < \varepsilon \le \varepsilon_W$ is such that $W \in B_{2\varepsilon}(W(\boldsymbol{y}))$ implies $\mathcal{U}(W) \in B_{\varepsilon_u}(\mathcal{U}(W(\boldsymbol{y})))$. This is possible due to the Lipschitz continuity of φ proved in the previous step and by possibly making the ε chosen in *Step 1* smaller. We now show that $\mathcal{H}(\boldsymbol{y})$ can be chosen to be $B_\epsilon(W(\boldsymbol{y}))$. Assume by contradiction that the *maximal* $\mathcal{H}(\boldsymbol{y})$ (in the sense as there is no superset of $\mathcal{H}(\boldsymbol{y})$ with the properties specified in Definition 3) is a *proper* subset of $B_\varepsilon(W(\boldsymbol{y}))$, i.e. $\mathcal{H}(\boldsymbol{y}) \subsetneq B_\varepsilon(W(\boldsymbol{y}))$. Let $W \in \partial\mathcal{H}(\boldsymbol{y}) \cap B_\varepsilon(W(\boldsymbol{y})) \neq \emptyset$. Lipschitz continuity of φ in Step 2 shows that $\mathcal{U}$ can be extended continuously to $\overline{\mathcal{H}(\boldsymbol{y})}$. Consequently, $\mathcal{U}(W)$ is well-defined and equals $\lim_{\sigma\to 1^-}\mathcal{U}(\sigma W + (1-\sigma)W(\boldsymbol{y})) \in \mathbb{U}$. Since $\mathcal{R}$ is continuous, $\mathcal{R}(W, \mathcal{U}(W)) = 0$. By

Assumption 3, $\partial_u \mathcal{R}(\overline{W}, \mathcal{U}(\overline{W}))$ is a homeomorphism for any $\overline{W}$ in a neighbourhood of W in $\mathbb{W}$. We may therefore apply the implicit function theorem in W to show that the domain of existence of a holomorphic extension of $\mathcal{U}$ can be further extended to an open neighbourhood $B \supsetneq \mathcal{H}(\boldsymbol{y})$ of W in $\mathbb{W}$. Clearly, the neighbourhood can be chosen such that $\mathcal{U}(\overline{W}) \in B_{\varepsilon_u}(\mathcal{U}(W(\boldsymbol{y})))$ for all $\overline{W} \in B$. Since B can be chosen star shaped with respect to $W(\boldsymbol{y})$, this contradicts the maximality of $\mathcal{H}(\boldsymbol{y})$. Thus, we proved that $B_\varepsilon(W(\boldsymbol{y})) = \mathcal{H}(\boldsymbol{y})$. The argument used in Step 1 immediately implies the uniform boundedness. □

Theorem 5 provides all the tools to estimate parametric regularity through Cauchy's integral theorem. The Lévy-Ciesielski expansion (4) can be (formally) extended to the complex parameters $\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}}$. Thus, in view of Theorem 5, $\boldsymbol{z} \mapsto \mathcal{U}(W(\boldsymbol{z}))$ is a holomorphic extension of the parameter-to-solution map in $\boldsymbol{y}$ for all $\boldsymbol{z}$ such that $W(\boldsymbol{z})$ belongs to the domain of holomorphy of $\mathcal{U}$, which in Theorem 5 was proved to contain $B_\varepsilon(W(\boldsymbol{y}))$ (recall that ε is independent of $\boldsymbol{y}$). Such a set of parameters can be defined as follows: Let $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ be a sequence of non-negative real numbers, and consider the polydisk

$$\boldsymbol{B_\rho}(\boldsymbol{y}) \coloneqq \{\boldsymbol{z} \in \mathcal{X} : |z_n - y_n| < \rho_n \quad \text{for all } n \in \mathbb{N}\}. \tag{13}$$

Assumption 4. *For $\varepsilon > 0$, $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$, there exists a real positive sequence $\boldsymbol{\rho} = \boldsymbol{\rho}(\varepsilon) = (\rho_n)_{n\in\mathbb{N}}$ such that,*

$$\boldsymbol{z} \in \boldsymbol{B_\rho}(\boldsymbol{y}) \Rightarrow W(\boldsymbol{z}) \in B_\varepsilon(W(\boldsymbol{y})),$$

In conclusion, for any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$, $\mathcal{U} \circ W : \boldsymbol{B_\rho}(\boldsymbol{y}) \to \mathbb{U}$ is holomorphic because it is a composition of holomorphic functions. Moreover, $\mathcal{U} \circ W$ is uniformly bounded by C_ε (see Theorem 5) independently of $\boldsymbol{y}$.

Consider a multi-index $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n) \in \mathbb{N}_0^n$ and denote by $\partial^{\boldsymbol{\nu}}$ the mixed derivative $\partial_1^{\nu_1} \ldots \partial_n^{\nu_n}$ where $\partial_j^{\nu_j}$ denotes the partial derivative of order ν_j with respect to y_j (if $\nu_j = 0$, the j-th partial derivative is omitted). Cauchy's integral theorem implies:

Theorem 6. *Consider $u : \mathcal{X}_{\mathbb{R}} \to \mathbb{U}$, the parameter-to-solution map that solves the parametric PDE (6). Let Assumptions 1, 2, 3 hold and fix $\varepsilon > 0$ as in Theorem 5. Finally, consider a real positive sequence $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ as in Assumption 4. Then, for any $n \in \mathbb{N}$, $\boldsymbol{\nu} = (\nu_i)_{i=1}^n \in \mathbb{N}_0^n$, it holds that*

$$\|\partial^{\boldsymbol{\nu}} u(\boldsymbol{y})\|_{\mathbb{U}} \leq \prod_{j=1}^{n} \nu_j! \rho_j^{-\nu_j} C_\varepsilon \qquad \text{for all } \boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}, \tag{14}$$

where $C_\varepsilon > 0$ from Theorem 5 is independent of $\boldsymbol{\nu}$ or $\boldsymbol{y}$. The same bound holds for $\|\partial^{\boldsymbol{\nu}} u\|_{L^2_{\boldsymbol{\mu}}(\mathcal{X}_{\mathbb{R}};\mathbb{U})}$, where $\boldsymbol{\mu}$ denotes a probability measure on $\mathcal{X}_{\mathbb{R}}$.

Note that Theorem 6 contains the crucial bound on the derivatives which justifies many high-dimensional approximation (and quadrature) methods such as, e.g., sparse grids, polynomial chaos, quasi-Monte Carlo.

3 The Stochastic Landau–Lifshitz–Gilbert equation

In the present section, we introduce the stochastic Landau-Lifshitz-Gilbert problem and we show that it fits the general theory described in the previous section. We consider a bounded Lipschitz domain $D \subset \mathbb{R}^3$ representing a ferromagnetic body in the time interval $[0, T]$. By $D_T \coloneqq [0, T] \times D$ we denote the space-time cylinder and by ∂_n the outward pointing normal derivative on ∂D. Given $\boldsymbol{M}_0 : D \to \mathbb{S}^2 \coloneqq \{\boldsymbol{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$ (the magnetization of the magnetic body at initial time), $\lambda > 0$ (called the *Gilbert damping parameter*), the deterministic version of the problem (the *LLG equation*) consists of determining the time evolution of the magnetization: Find $\boldsymbol{M}(t, \boldsymbol{x}) : D_T \to \mathbb{S}^2$ such that

$$\begin{cases} \partial_t \boldsymbol{M} = \lambda_1 \boldsymbol{M} \times \Delta \boldsymbol{M} - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \Delta \boldsymbol{M}) & \text{in } D_T, \\ \partial_n \boldsymbol{M} = \boldsymbol{0} & \text{on } \partial D \times [0, T], \\ \boldsymbol{M}(0) = \boldsymbol{M}_0 & \text{on } D, \end{cases} \tag{15}$$

where $\lambda_1 = \frac{1}{1+\lambda^2}$, $\lambda_2 = \frac{\lambda}{1+\lambda^2}$. The solution has constant magnitude in space and time (this follows immediately from scalar multiplication of (15) with $\boldsymbol{M}$). This implies that, assuming a normalized initial condition $|\boldsymbol{M}_0| \equiv 1$ on D, that

$$|\boldsymbol{M}(t, \boldsymbol{x})| = 1 \qquad \text{for all } (t, \boldsymbol{x}) \in D_T.$$

In (15), the *exchange term* $\Delta \boldsymbol{M}$ can be substituted by a more general *effective field* $\boldsymbol{H}_{\text{eff}}(\boldsymbol{M})$ containing $\Delta \boldsymbol{M}$ and additional lower order contributions modelling additional physical effects like material anisotropy, magnetostatic energy, external magnetic fields or the more involved Dzyaloshinskii-Moriya interaction (DMI) (see e.g. [50, Section 1.2]).

The effect of heat fluctuations on the systems is described with a random model. Denote by $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and let $\mathrm{d}\boldsymbol{W} : \Omega \times D_T \to \mathbb{R}^3$ be a suitable space-time noise (note that the exact form of this noise is subject of research and below we consider a simple one-dimensional model). Consider the following formal equation for $\boldsymbol{M} : \Omega \times D_T \to \mathbb{S}^2$:

$$\partial_t \boldsymbol{M} = \lambda_1 \boldsymbol{M} \times (\Delta \boldsymbol{M} + \mathrm{d}\boldsymbol{W}) - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \Delta \boldsymbol{M}) \quad \text{in } D_T, \mathbb{P}\text{-a.s.}$$

with the same initial and boundary conditions as in (15). It is customary not to include a noise in the second term of the right-hand-side because of the smallness of λ_2 compared to λ_1 (see, e.g., [12, page 3]). For simplicity, we additionally assume one-dimensional noise $\boldsymbol{W}(\omega, t, \boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x}) W(\omega, t)$ for all $\omega \in \Omega, (t, \boldsymbol{x}) \in D_T$, where $\boldsymbol{g} : D \to \mathbb{R}^3$ is given and $W : \Omega \times [0, T] \to \mathbb{R}$ denotes a (scalar) Wiener process.

The previous formal equation corresponds to the following *stochastic* partial differential equation called the *stochastic LLG problem*: Find $\boldsymbol{M} : \Omega \times D_T \to \mathbb{S}^2$ such that

$$\mathrm{d}\boldsymbol{M} = (\lambda_1 \boldsymbol{M} \times \Delta \boldsymbol{M} - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \Delta \boldsymbol{M}))\, \mathrm{d}t + (\lambda_1 \boldsymbol{M} \times \boldsymbol{g}) \circ \mathrm{d}W \quad \text{in } D_T,\ \mathbb{P}\text{-a.s.} \tag{16}$$

again with initial and boundary conditions as in (15). By $\circ \mathrm{d}W$ we denote the Stratonovich differential. We define a weak solution of this problem following [34].

Definition 7. *A weak martingale solution of* (16) *is a tuple* $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}, W, \boldsymbol{M}\right)$ *where*

- $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}\right)$ *is a filtered probability space;*
- $W : \Omega \times [0, T] \to \mathbb{R}$ *is a scalar Wiener process adapted to* $(\mathcal{F}_t)_{t\in[0,T]}$*;*
- $\boldsymbol{M} : \Omega \times [0, T] \to L^2(D)^3$ *is a progressively measurable stochastic process;*

such that the following properties hold:

- $\boldsymbol{M}(\omega, \cdot) \in C^0([0, T], H^{-1}(D))$ *$\mathbb{P}$-a.e. $\omega \in \Omega$;*
- $\mathbb{E}\left(\operatorname{esssup}_{t\in[0,T]} \|\nabla \boldsymbol{M}\|^2_{L^2(D)}\right) < \infty$*;*
- $|\boldsymbol{M}(\omega, t, \boldsymbol{x})| = 1$ *$\mathbb{P}$-a.e. $\omega \in \Omega$, for all $t \in [0, T]$, for a.e. $\boldsymbol{x} \in D$;*
- *For all $t \in [0, T]$ and all $\boldsymbol{\phi} \in C_0^\infty(D)^3$, $\mathbb{P}$-a.s. there holds*

$$\begin{aligned}\langle \boldsymbol{M}(t), \boldsymbol{\phi}\rangle - \langle \boldsymbol{M}_0, \boldsymbol{\phi}\rangle = -\lambda_1 \int_0^t \langle \boldsymbol{M} \times \nabla \boldsymbol{M}, \nabla \boldsymbol{\phi}\rangle \,\mathrm{d}s - \lambda_2 \int_0^t \langle \boldsymbol{M} \times \nabla \boldsymbol{M}, \nabla (\boldsymbol{M} \times \boldsymbol{\phi})\rangle \,\mathrm{d}s \\ + \lambda_1 \int_0^t \langle \boldsymbol{M} \times \boldsymbol{g}, \boldsymbol{\phi}\rangle \circ \mathrm{d}W(s),\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2(D)^3$ scalar product.

Existence of solutions to (16) in this sense was first established in [12], while uniqueness of weak solutions is still an open question. An alternative existence proof was given in [34]. Here the authors use the Doss-Sussman transform to obtain a PDE with random coefficients instead of the stochastic differential as explained in the previous section.

4 Random LLG equation by Doss-Sussmann transform and parametric LLG equation by Lévy-Ciesielski expansion

In the present section, we apply the strategy outlined in Section 2.1 to the SLLG problem (16) in order to obtain a random coefficient PDE. While this was done in [34] for technical reasons, we are mainly interested in obtaining an equivalent problem that is more amenable to collocation-type approximation. Another advantage is (formally) gaining a full order of differentiability of the solution. Given $\boldsymbol{g} : D \to \mathbb{R}^3$, $s \in \mathbb{R}$ and $\boldsymbol{v} : D \to \mathbb{R}^3$ with suitable regularity, consider the following operators:

$$\begin{aligned}
G\boldsymbol{v} &= \boldsymbol{v} \times \boldsymbol{g} \\
\mathcal{C}\boldsymbol{v} &= \boldsymbol{v} \times \Delta \boldsymbol{g} + 2\nabla \boldsymbol{v} \times \nabla \boldsymbol{g} \\
e^{sG}\boldsymbol{v} &= \boldsymbol{v} + \sin(s) G\boldsymbol{v} + (1 - \cos s) G^2 \boldsymbol{v} \\
\mathcal{E}(s, \boldsymbol{v}) &= \sin(s)\mathcal{C}\boldsymbol{v} + (1 - \cos(s))(\mathcal{C}G + G\mathcal{C})\boldsymbol{v} \\
\hat{\mathcal{C}}(s, \boldsymbol{v}) &= e^{-sG}\mathcal{E}(s, \boldsymbol{v}) = \mathcal{E}(s, \boldsymbol{v}) - \sin(s) G\mathcal{E}(s, \boldsymbol{v}) + (1 - \cos(s)) G^2 \mathcal{E}(s, \boldsymbol{v}),
\end{aligned} \tag{17}$$

where we define $\nabla \boldsymbol{v} \times \nabla \boldsymbol{g} := \sum_{j=1}^{3} \frac{\partial \boldsymbol{v}}{\partial x_j} \times \frac{\partial \boldsymbol{g}}{\partial x_j}$. Note that e^{sG} is the exponential of the operator sG. The fact $G \circ G \circ G\boldsymbol{v} = -\boldsymbol{v}$ simplifies the expression. Expanding some of the definitions, the last operator can be written as

$$\begin{aligned}\hat{\mathcal{C}}(s, \boldsymbol{v}) = &\sin(s)\mathcal{C}\boldsymbol{v} + (1-\cos(s))\left(\mathcal{C}G + G\mathcal{C}\right)\boldsymbol{u} - \sin(s)^2 G\mathcal{C}\boldsymbol{u} + \\ &- \sin(s)(1-\cos(s))G\left(\mathcal{C}G + G\mathcal{C}\right)\boldsymbol{u} + (1-\cos(s))\sin(s)G^2\mathcal{C}\boldsymbol{u} + \\ &+ (1-\cos(s))^2 G^2\left(\mathcal{C}G + G\mathcal{C}\right)\boldsymbol{u}\end{aligned}$$

or, in compact form, as

$$\hat{\mathcal{C}}(s, \boldsymbol{v}) = \sum_{i=1}^{6} b_i(s) F_i(\boldsymbol{v}), \tag{18}$$

where b_i are uniformly bounded with bounded derivatives (let $0 < \beta < \infty$ be a uniform bound for both, which depends only on $\boldsymbol{g}$) and the F_i are linear and globally Lipschitz with the Lipschitz constant $0 < L < \infty$ depending only on $\boldsymbol{g}$, i.e., for any $I = 1, \ldots, 6$,

$$\begin{aligned}&\|b_i(W)\|_{L^\infty(\mathbb{R})} \le \beta, \quad \|b_i'(W)\|_{L^\infty(\mathbb{R})} \le \beta && \text{for all } W \in C^0([0,T]),\\ &\|F_i(\boldsymbol{u}) - F_i(\boldsymbol{v})\|_{L^2(D)} \le L\,\|\boldsymbol{u} - \boldsymbol{v}\|_{H^1(D)} && \text{for all } \boldsymbol{u}, \boldsymbol{v} \in H^1(D)^3.\end{aligned}$$

In the present setting, the *Doss-Sussmann transform* (1) reads

$$\boldsymbol{m} = e^{-WG}\boldsymbol{M}.$$

We obtain the *random coefficients LLG problem*: Given $\boldsymbol{M}^0 : D \to \mathbb{S}^2$, find $\boldsymbol{m} : \Omega \times D_T \to \mathbb{S}^2$ such that for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$\begin{cases}\partial_t \boldsymbol{m}(\omega) & = \lambda_1 \boldsymbol{m}(\omega) \times \Big(\Delta \boldsymbol{m}(\omega) + \hat{\mathcal{C}}(W(\omega), \boldsymbol{m}(\omega))\Big) \\ & \quad -\lambda_2 \boldsymbol{m}(\omega) \times \Big(\boldsymbol{m}(\omega) \times \Big(\Delta \boldsymbol{m}(\omega) + \hat{\mathcal{C}}(W(\omega), \boldsymbol{m}(\omega))\Big)\Big) \qquad \text{in } D_T,\\ \partial_n \boldsymbol{m}(\omega) & = 0 \qquad \text{on } [0,T] \times \partial D,\\ \boldsymbol{m}(\omega, 0, \cdot) & = \boldsymbol{M}_0 \qquad \text{on } D.\end{cases} \tag{19}$$

It is shown in [34, Lemma 4.6] that any weak solution $\boldsymbol{m}$ of (19) corresponds to a weak martingale solution $\boldsymbol{M} = e^{WG}\boldsymbol{m}$ of (16) through the *inverse Doss-Sussmann transform*. Existence of solutions to (19) is shown in [34], but again uniqueness is open.

Following Section 2.1, we derive a parametric PDE problem using the Lévy-Ciesielski expansion of the Wiener process. The *parametric LLG problem* reads: Given $\boldsymbol{M}_0 : D \to \mathbb{S}^2$, find $\boldsymbol{m} : \mathcal{X}_\mathbb{R} \times D_T \to \mathbb{S}^2$ such that for a.a. $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$

$$\begin{cases}\partial_t \boldsymbol{m}(\boldsymbol{y}) & = \boldsymbol{m}(\boldsymbol{y}) \times \Big(\Delta \boldsymbol{m}(\boldsymbol{y}) + \hat{\mathcal{C}}(W(\boldsymbol{y}), \boldsymbol{m}(\boldsymbol{y}))\Big) \\ & \quad -\boldsymbol{m}(\boldsymbol{y}) \times \Big(\boldsymbol{m}(\boldsymbol{y}) \times \Big(\Delta \boldsymbol{m}(\boldsymbol{y}) + \hat{\mathcal{C}}(W(\boldsymbol{y}), \boldsymbol{m}(\boldsymbol{y}))\Big)\Big) \qquad \text{in } D_T,\\ \partial_n \boldsymbol{m}(\boldsymbol{y}) & = 0 \qquad \text{on } [0,T] \times \partial D,\\ \boldsymbol{m}(\boldsymbol{y}, 0, \cdot) & = \boldsymbol{M}_0 \qquad \text{on } D,\end{cases} \tag{20}$$

where we set $\lambda_1 = \lambda_2 = 1$ for simplicity. The precise definition of $\mathcal{X}_\mathbb{R}$ will be given below in (33).

Applying the triple cross-product formula $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b}(\boldsymbol{a} \cdot \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a} \cdot \boldsymbol{b})$ on $\boldsymbol{m}(\boldsymbol{y}) \times (\boldsymbol{m}(\boldsymbol{y}) \times (\Delta \boldsymbol{m}(\boldsymbol{y})))$, together with the fact that $|\boldsymbol{m}| \equiv 1$, gives an equivalent equation valid again for a.a. $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$:

$$\partial_t \boldsymbol{m}(\boldsymbol{y}) = \Delta \boldsymbol{m}(\boldsymbol{y}) + \boldsymbol{m}(\boldsymbol{y}) \times \Delta \boldsymbol{m}(\boldsymbol{y}) - (\nabla \boldsymbol{m}(\boldsymbol{y}) : \nabla \boldsymbol{m}(\boldsymbol{y}))\,\boldsymbol{m}(\boldsymbol{y}) + \tag{21}$$

$$+ \boldsymbol{m}(\boldsymbol{y}) \times \hat{\mathcal{C}}(W, \boldsymbol{m}(\boldsymbol{y})) - \boldsymbol{m}(\boldsymbol{y}) \times \Big(\boldsymbol{m}(\boldsymbol{y}) \times \hat{\mathcal{C}}(W, \boldsymbol{m}(\boldsymbol{y}))\Big) \quad \text{in } D_T. \tag{22}$$

4.1 Space and time Hölder regularity of sample paths of the random LLG problem

In the present section, we prove that the sample paths of solutions of the random LLG problem (19) are Hölder regular.

We recall basic definitions and important facts about Hölder spaces. Let $n \in \mathbb{N}$, $D \subset \mathbb{R}^n$, $\alpha \in (0,1)$, $v : D \to \mathbb{C}$. The Hölder-seminorm reads $|v|_{C^\alpha(D)} := \sup_{\boldsymbol{x}, \boldsymbol{y} \in D, \boldsymbol{x} \neq \boldsymbol{y}} \frac{|v(\boldsymbol{x}) - v(\boldsymbol{y})|}{|\boldsymbol{x} - \boldsymbol{y}|^\alpha}$ and by $C^\alpha(D)$, we denote the Banach space of functions with finite Hölder-norm $\|v\|_{C^\alpha(D)} := \|v\|_{C^0(D)} + |v|_{C^\alpha(D)}$. Clearly, $u, v \in C^\alpha(D)$

implies $uv \in C^\alpha(D)$. Higher Hölder regularity of order $k \in \mathbb{N}$ is characterized via the seminorm $|v|_{C^{k+\alpha}(D)} := \sum_{j=1}^k |D^j v|_{C^\alpha(D)}$ and the corresponding Banach space $C^{k+\alpha}(D) := \{v : D \to \mathbb{C} \,:\, D^j v \in C^\alpha(D) \text{ for all } j = 0, \ldots, k\}$ with the norm $\|v\|_{C^{k+\alpha}(D)} := \sum_{j=0}^k \|D^j v\|_{C^\alpha(D)}$. Again $u, v \in C^{k+\alpha}(D)$ immediately implies $uv \in C^{k+\alpha}(D)$. In the parabolic setting, it is useful to define the *parabolic distance* between $P = (t, \boldsymbol{x})$, $Q = (s, \boldsymbol{y}) \in D_T$ by

$$d(P,Q) := \left(|t-s| + |\boldsymbol{x}-\boldsymbol{y}|^2\right)^{1/2}.$$

For a function $v : D_T \to \mathbb{C}$, define the seminorm $|v|_{C^{\alpha/2,\alpha}(D_T)} := \sup_{\substack{P,Q \in D_T \\ P \neq Q}} \frac{|v(P)-v(Q)|}{d(P,Q)^\alpha}$. Define the Banach spaces $C^{\alpha/2,\alpha}(D_T) := \{v : D_T \to \mathbb{C} : v \in C^0(D_T) \text{ and } |v|_{C^{\alpha/2,\alpha}(D)} < \infty\}$ with the norm (see [56, Section 1.2.3] for details) $\|v\|_{C^{\alpha/2,\alpha}(D_T)} := \|v\|_{C^0(D_T)} + |v|_{C^{\alpha/2,\alpha}(D_T)}$. Finally,

$$C^{1+\alpha/2,2+\alpha}(D_T) := \left\{v : D_T \to \mathbb{C} : \partial_t v \text{ and } D^j v \in C^{\alpha/2,\alpha}(D_T), j = 0,1,2\right\} \tag{23}$$

is a Banach space when endowed with the norm

$$\|v\|_{C^{1+\alpha/2,2+\alpha}(D_T)} := \sum_{j=0}^2 \|D^j v\|_{C^{\alpha/2,\alpha}(D_T)} + \|\partial_t v\|_{C^{\alpha/2,\alpha}(D_T)}.$$

In what follows, we work with the Hölder seminorm

$$|v|_{C^{1+\alpha/2,2+\alpha}(D_T)} := |v|_{C^{\alpha/2,\alpha}(D_T)} + \sum_{j=1}^2 \|D^j v\|_{C^{\alpha/2,\alpha}(D_T)} + \|\partial_t v\|_{C^{\alpha/2,\alpha}(D_T)}. \tag{24}$$

As above, if $u, v \in C^{1+\alpha/2,2+\alpha}(D_T)$ then also $uv \in C^{1+\alpha/2,2+\alpha}(D_T)$. In particular, it can be proved that $\|uv\|_{C^{\alpha/2,\alpha}(D_T)} \le \|u\|_{C^{\alpha/2,\alpha}(D_T)} \|v\|_{C^{\alpha/2,\alpha}(D_T)}$.

Definitions generalize to vector fields in the usual way. We use the same symbols for scalar and vector spaces. In the remainder of this section, we adopt the short notation $\|\cdot\|_\alpha = \|\cdot\|_{C^\alpha(D)}$, $\|\cdot\|_{1+\alpha/2,2+\alpha} = \|\cdot\|_{C^{1+\alpha/2,2+\alpha}(D_T)}$, and analogously for all other norms and seminorms.

To prove Hölder regularity of sample paths, we work with the following equivalent form of (19), obtained using algebraic manipulations including the triple product expansion and the fact that $|\boldsymbol{m}| = 1$ for all $t \in [0,T]$ and a.a. $\boldsymbol{x} \in D$:

$$\lambda \partial_t \boldsymbol{m} + \boldsymbol{m} \times \partial_t \boldsymbol{m} = \Delta \boldsymbol{m} + |\nabla \boldsymbol{m}|^2 \boldsymbol{m} - \boldsymbol{m} \times \left(\boldsymbol{m} \times \hat{\mathcal{C}}(W, \boldsymbol{m})\right), \tag{25}$$

where we recall that $\lambda > 0$ is the Gilbert damping parameter and $\hat{\mathcal{C}}$ was defined in (17).

The main result of this section is summarized in the following theorem.

Theorem 8. *Let $0 < \alpha < 1$. Assume that $W \in C^{\alpha/2}([0,T])$, $\boldsymbol{M}_0 \in C^{2+\alpha}(D)$ and $\boldsymbol{g} \in C^{2+\alpha}(D)$. There exists $\varepsilon > 0$ such that if $\|\boldsymbol{M}_0\|_{2+\alpha} \le \varepsilon$, $\|\Delta \boldsymbol{g}\|_\alpha \le \varepsilon$, and $\|\nabla \boldsymbol{g}\|_\alpha \le \varepsilon$, then the solution $\boldsymbol{m}$ of equation (25) with initial condition $\boldsymbol{m}(0) = \boldsymbol{M}_0$ and homogeneous Neumann boundary conditions belongs to $C^{1+\alpha/2,2+\alpha}(D_T)$. Moreover,*

$$\|\boldsymbol{m}\|_{1+\alpha/2,2+\alpha} \le C_r, \tag{26}$$

where $C_r > 0$ depends on $\|\boldsymbol{g}\|_{2+\alpha}$, $\|\boldsymbol{M}_0\|_{2+\alpha}$, λ, D and T but is independent of W.

The proof of the theorem is inspired by [29]. The proofs in the mentioned work require higher temporal regularity than is available for stochastic LLG, which we circumvent by the use of Hölder spaces instead of Sobolev spaces. In the following, we will require some notation:

$$\begin{aligned}
H(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) &:= \boldsymbol{u} \times (\boldsymbol{v} \times \hat{\mathcal{C}}(W, \boldsymbol{w})) \quad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in C^{\alpha/2,\alpha}(D_T), \boldsymbol{w} \in C^{\alpha/2,1+\alpha}(D_T),\\
\mathcal{R}_a(\boldsymbol{v}) &:= \lambda \partial_t \boldsymbol{v} + \boldsymbol{v} \times \partial_t \boldsymbol{v} - |\boldsymbol{v}|^2 \Delta \boldsymbol{v} - |\nabla \boldsymbol{v}|^2 \boldsymbol{v} + H(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in C^{1+\alpha/2,2+\alpha}(D_T),\\
L\boldsymbol{v} &:= L_{\boldsymbol{x}_0} \boldsymbol{v} := \lambda \boldsymbol{v} + \boldsymbol{M}_0(\boldsymbol{x}_0) \times \boldsymbol{v} \quad \text{for all } \boldsymbol{x}_0 \in D, \boldsymbol{v} \in C^{\alpha/2,\alpha}(D_T).
\end{aligned}$$

We note that $\mathcal{R}_a$ is the residual defined from the alternative form (25) of the LLG equation; confer (6).

We will require a couple of technical results.

Lemma 9 (Continuity of the trilinear form H and of the LLG residual $\mathcal{R}_a$)**.** *If $\boldsymbol{u}$, $\boldsymbol{v} \in C^{\alpha/2,\alpha}(D_T)$ and $\boldsymbol{w} \in C^{\alpha/2,1+\alpha}(D_T)$, then $H(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in C^{\alpha/2,\alpha}(D_T)$ and*

$$\|H(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})\|_{\alpha/2,\alpha} \le C_{\boldsymbol{g}} \|\boldsymbol{u}\|_{\alpha/2,\alpha} \|\boldsymbol{v}\|_{\alpha/2,\alpha} \|\boldsymbol{w}\|_{\alpha/2,1+\alpha}, \tag{27}$$

where $C_{\boldsymbol{g}} := \left(1 + \|\boldsymbol{g}\|_{1+\alpha}\right)^3 (\|\nabla \boldsymbol{g}\|_\alpha + \|\Delta \boldsymbol{g}\|_\alpha)$*. Moreover, if* $\boldsymbol{v} \in C^{1+\alpha/2,2+\alpha}(D_T)$*, then* $\mathcal{R}_a(\boldsymbol{v}) \in C^{\alpha/2,\alpha}(D_T)$ *and*

$$\|\mathcal{R}_a(\boldsymbol{v})\|_{\alpha/2,\alpha} \leq \left(|\boldsymbol{v}|_{1+\alpha/2,2+\alpha} + |\boldsymbol{v}|^2_{1+\alpha/2,2+\alpha}\right)\left(\lambda + \|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha}\right)^2 + C_{\boldsymbol{g}} \|\boldsymbol{v}\|^2_{\alpha/2,\alpha} \|\boldsymbol{v}\|_{\alpha/2,1+\alpha} . \tag{28}$$

In particular, $\|\mathcal{R}_a(\boldsymbol{v})\|_{\alpha/2,\alpha}$ *vanishes when* $|\boldsymbol{v}|_{1+\alpha/2,2+\alpha}$ *and* $\|\nabla \boldsymbol{g}\|_{\alpha/2,\alpha} + \|\Delta \boldsymbol{g}\|_{\alpha/2,\alpha}$ *all vanish.*

Proof. To prove (27), note the following elementary estimates

$$\begin{aligned}
\|\mathcal{C}\boldsymbol{v}\|_{\alpha/2,\alpha} &\leq 2\|\nabla \boldsymbol{v}\|_{\alpha/2,\alpha} \|\nabla \boldsymbol{g}\|_\alpha + \|\boldsymbol{v}\|_{\alpha/2,\alpha} \|\Delta \boldsymbol{g}\|_\alpha , \\
\|\mathcal{C}G\boldsymbol{v}\|_{\alpha/2,\alpha} &\leq \|\boldsymbol{v}\|_{\alpha/2,\alpha} \|\boldsymbol{g}\|_\alpha \|\Delta \boldsymbol{g}\|_\alpha + 2(\|\nabla \boldsymbol{v}\|_{\alpha/2,\alpha} \|\boldsymbol{g}\|_\alpha + \|\boldsymbol{v}\|_{\alpha/2,\alpha} \|\nabla \boldsymbol{g}\|_\alpha) \|\nabla \boldsymbol{g}\|_\alpha , \\
\|\mathcal{E}(s,\boldsymbol{v})\|_{\alpha/2,\alpha} &\leq \|\mathcal{C}\boldsymbol{v}\|_{\alpha/2,\alpha} + \|\mathcal{C}G\boldsymbol{v}\|_{\alpha/2,\alpha} + \|\boldsymbol{g}\|_\alpha \|\mathcal{C}\boldsymbol{v}\|_{\alpha/2,\alpha} , \\
\left\|\hat{\mathcal{C}}(s,\boldsymbol{v})\right\|_{\alpha/2,\alpha} &\leq \left(1 + \|\boldsymbol{g}\|_\alpha + \|\boldsymbol{g}\|^2_\alpha\right) \|\mathcal{E}(s,\boldsymbol{v})\|_{\alpha/2,\alpha} , \\
\|H(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\|_{\alpha/2,\alpha}) &\leq \|\boldsymbol{u}\|_{\alpha/2,\alpha} \|\boldsymbol{v}\|_{\alpha/2,\alpha} \left\|\hat{\mathcal{C}}(W,\boldsymbol{w})\right\|_{\alpha/2,\alpha} .
\end{aligned}$$

Putting these facts together, one obtains (27). To get the second inequality (28), estimate

$$\begin{aligned}
\|\mathcal{R}_a(\boldsymbol{v})\|_{\alpha/2,\alpha} &\leq \lambda |\boldsymbol{v}|_{1+\alpha/2,2+\alpha} + \|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha} |\boldsymbol{v}|_{1+\alpha/2,2+\alpha} \\
&\quad + \|\boldsymbol{v}\|^2_{1+\alpha/2,2+\alpha} |\boldsymbol{v}|_{1+\alpha/2,2+\alpha} + |\boldsymbol{v}|^2_{1+\alpha/2,2+\alpha} \|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha} + \|H(\boldsymbol{v},\boldsymbol{v},\boldsymbol{v})\|_{\alpha/2,\alpha} \\
&\leq \left(|\boldsymbol{v}|_{1+\alpha/2,2+\alpha} + |\boldsymbol{v}|^2_{1+\alpha/2,2+\alpha}\right)\left(\lambda + \|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha}\right)^2 + \|H(\boldsymbol{v},\boldsymbol{v},\boldsymbol{v})\|_{\alpha/2,\alpha} .
\end{aligned}$$

Using (27) to estimate the last term yields (28).

□

Additionally, we need some finer control over the boundedness of $\mathcal{R}_a$. The point of the following result is that all terms apart from the first one on the right-hand side of the estimate in Lemma 10 below are either at least quadratic in $\boldsymbol{w}$ or can be made small by choosing $\boldsymbol{v}$ close to a constant function. This will allow us to treat the nonlinear parts as perturbations of the heat equation.

Lemma 10. *For* $\boldsymbol{v}, \boldsymbol{w} \in C^{1+\alpha/2,2+\alpha}(D_T)$ *and* $\boldsymbol{x}_0 \in D$*, there holds*

$$\begin{aligned}
\|\mathcal{R}_a(\boldsymbol{v}-\boldsymbol{w})\|_{\alpha/2,\alpha} &\leq \|\mathcal{R}_a(\boldsymbol{v}) - (L\partial_t - \Delta)\boldsymbol{w}\|_{\alpha/2,\alpha} + \|\boldsymbol{v} - \boldsymbol{M}_0(\boldsymbol{x}_0)\|_{\alpha/2,\alpha} \|\boldsymbol{w}\|_{1+\alpha/2,2+\alpha} + \\
&\quad + \left\|\left(1 - |\boldsymbol{v}|^2\right)\Delta \boldsymbol{w}\right\|_{\alpha/2,\alpha} + \\
&\quad + \|\boldsymbol{w}\|_{1+\alpha/2,2+\alpha}\left(|\boldsymbol{v}|_{1+\alpha/2,2+\alpha} + C_{\boldsymbol{g}}\right)\left(1 + \|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha}\right)^2 + \\
&\quad + \|\boldsymbol{w}\|^2_{1+\alpha/2,2+\alpha}\left(1 + (1+C_{\boldsymbol{g}})\|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha}\right) + \|\boldsymbol{w}\|^3_{1+\alpha/2,2+\alpha}(1 + C_{\boldsymbol{g}}),
\end{aligned}$$

where $C_{\boldsymbol{g}} > 0$ *is defined in Lemma 9.*

Proof. All but the last term in the definition of $\mathcal{R}_a$ are estimated as in [29]. As for the last term, observe that

$$\begin{aligned}
H(\boldsymbol{v}-\boldsymbol{w},\boldsymbol{v}-\boldsymbol{w},\boldsymbol{v}-\boldsymbol{w}) &= H(\boldsymbol{v},\boldsymbol{v},\boldsymbol{v}) - H(\boldsymbol{w},\boldsymbol{w},\boldsymbol{w}) \\
&\quad - H(\boldsymbol{w},\boldsymbol{v},\boldsymbol{v}) - H(\boldsymbol{v},\boldsymbol{w},\boldsymbol{v}) - H(\boldsymbol{v},\boldsymbol{v},\boldsymbol{w}) \\
&\quad + H(\boldsymbol{v},\boldsymbol{w},\boldsymbol{w}) + H(\boldsymbol{w},\boldsymbol{v},\boldsymbol{w}) + H(\boldsymbol{w},\boldsymbol{w},\boldsymbol{v}).
\end{aligned}$$

The term $H(\boldsymbol{v},\boldsymbol{v},\boldsymbol{v})$ is absorbed in $\mathcal{R}_a(\boldsymbol{v})$. Then, by the previous lemma:

$$\begin{aligned}
\|-H(\boldsymbol{w},\boldsymbol{v},\boldsymbol{v}) - H(\boldsymbol{v},\boldsymbol{w},\boldsymbol{v}) - H(\boldsymbol{v},\boldsymbol{v},\boldsymbol{w})\|_{\alpha/2,\alpha} &\lesssim C_{\boldsymbol{g}} \|\boldsymbol{w}\|_{\alpha/2,1+\alpha} \|\boldsymbol{v}\|^2_{\alpha/2,1+\alpha} , \\
\|H(\boldsymbol{v},\boldsymbol{w},\boldsymbol{w}) + H(\boldsymbol{w},\boldsymbol{v},\boldsymbol{w}) + H(\boldsymbol{w},\boldsymbol{w},\boldsymbol{v})\|_{\alpha/2,\alpha} &\lesssim C_{\boldsymbol{g}} \|\boldsymbol{w}\|^2_{\alpha/2,1+\alpha} \|\boldsymbol{v}\|_{\alpha/2,1+\alpha} , \\
\|-H(\boldsymbol{w},\boldsymbol{w},\boldsymbol{w})\|_{\alpha/2,\alpha} &\lesssim C_{\boldsymbol{g}} \|\boldsymbol{w}\|^3_{\alpha/2,1+\alpha} .
\end{aligned}$$

Altogether, we obtain the required result. □

To prove Theorem 8, we use a fixed point iteration.

Proof of Theorem 8. Consider the initial guess $\boldsymbol{m}_0(t,\boldsymbol{x}) = \boldsymbol{M}_0(\boldsymbol{x})$ for all $t \in [0,T]$, $\boldsymbol{x} \in D$, and fix one $\boldsymbol{x}_0 \in D$ (for the definition of $L = L_{\boldsymbol{x}_0}$). Define the sequence $(\boldsymbol{m}_\ell)_\ell$ as follows: For $\ell = 0, 1, \ldots$

1. Define $\boldsymbol{r}_\ell := \mathcal{R}_a(\boldsymbol{m}_\ell)$
2. Solve

$$\begin{cases} L\partial_t \boldsymbol{R}_\ell - \Delta \boldsymbol{R}_\ell &= \boldsymbol{r}_\ell \quad \text{in } D_T, \\ \partial_n \boldsymbol{R}_\ell &= 0 \quad \text{on } \partial D \times [0,T], \\ \boldsymbol{R}_\ell(0) &= 0 \quad \text{on } D. \end{cases}$$

3. Update $\boldsymbol{m}_{\ell+1} := \boldsymbol{m}_\ell - \boldsymbol{R}_\ell$.

Step 1 (Well-posedness): By definition, we have $\boldsymbol{m}_0 \in C^{1+\alpha/2,2+\alpha}(D_T)$ as well as $\partial_n \boldsymbol{m}_0 = 0$. Assume that $\boldsymbol{m}_\ell \in C^{1+\alpha/2,2+\alpha}(D_T)$ and $\partial_n \boldsymbol{m}_\ell = 0$. Then, Lemma 9 implies that $\boldsymbol{r}_\ell \in C^{\alpha/2,\alpha}(D_T)$. The parabolic regularity result [40, Theorem 10.4,§10, VII] yields $\boldsymbol{R}_\ell \in C^{1+\alpha/2,2+\alpha}(D_T)$.

Step 2 (Convergence): We show the Cauchy property of the sequence $(\boldsymbol{m}_\ell)_\ell$: Fix $0 \le \ell' < \ell < \infty$ and observe that $\|\boldsymbol{m}_\ell - \boldsymbol{m}_{\ell'}\|_{1+\alpha/2,2+\alpha} \le \sum_{j=\ell'}^{\ell-1} \|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}$. By the previous lemmata, we have

$$\|\boldsymbol{R}_{j+1}\|_{1+\alpha/2,2+\alpha} \le C_s \|\boldsymbol{r}_{j+1}\|_{\alpha/2,\alpha} = C_s \|\mathcal{R}_a(\boldsymbol{m}_{j+1})\|_{\alpha/2,\alpha} = C_s \|\mathcal{R}_a(\boldsymbol{m}_j - \boldsymbol{R}_j)\|_{\alpha/2,\alpha}, \tag{29}$$

where $C_s > 0$ is the stability constant from [40, Theorem 10.4,§10, VII], which only depends on D_T and L (particularly, it is independent of ℓ). We invoke Lemma 10 with $\boldsymbol{v} = \boldsymbol{m}_j$ and $\boldsymbol{w} = \boldsymbol{R}_j$. By construction, $\mathcal{R}_a(\boldsymbol{m}_j) - (L\partial_t - \Delta)\boldsymbol{R}_j = 0$. What remains is estimated as

$$\begin{aligned} \|\mathcal{R}_a(\boldsymbol{m}_j - \boldsymbol{R}_j)\|_{\alpha/2,\alpha} \le{}& \|\boldsymbol{m}_j - \boldsymbol{M}_0(\boldsymbol{x}_0)\|_{\alpha/2,\alpha} \|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha} + \left\|\left(1 - |\boldsymbol{m}_j|^2\right)\Delta \boldsymbol{R}_j\right\|_{\alpha/2,\alpha} \\ &+ \|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}\left(|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha} + C_{\boldsymbol{g}}\right)\left(1 + \|\boldsymbol{m}_j\|_{1+\alpha/2,2+\alpha}\right)^2 \\ &+ \|\boldsymbol{R}_j\|^2_{1+\alpha/2,2+\alpha}\left(1 + (1 + C_{\boldsymbol{g}})\,|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}\right) + \\ &+ \|\boldsymbol{R}_j\|^3_{1+\alpha/2,2+\alpha}(1 + C_{\boldsymbol{g}}). \end{aligned} \tag{30}$$

Let us estimate the first term in (30). For any $(t,\boldsymbol{x}) \in D_T$, the fundamental theorem of calculus yields $|\boldsymbol{m}_j(\boldsymbol{x},t) - \boldsymbol{M}_0(\boldsymbol{x}_0)| \lesssim \|(\partial_t,\nabla)\boldsymbol{m}_j\|_{C^0(D_T)} \le |\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}$. Analogously, we get

$$\|\boldsymbol{m}_j - \boldsymbol{M}_0(\boldsymbol{x}_0)\|_{\alpha/2,\alpha} \le 2\,|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}.$$

Let us estimate the second term in (30). Since $\boldsymbol{m}_j = \boldsymbol{m}_0 + \sum_{i=0}^{j-1}\boldsymbol{R}_i$ and $|\boldsymbol{m}_0| = 1$ a.e., we have $|\boldsymbol{m}_j|^2 = 1 + 2\boldsymbol{m}_0 \cdot \sum_{i=0}^{j-1}\boldsymbol{R}_i + \left(\sum_{i=0}^{j-1}\boldsymbol{R}_i\right)^2$. Thus, the fact that Hölder spaces are closed under multiplication and the triangle inequality imply

$$\left\|1 - |\boldsymbol{m}_j|^2\right\|_{\alpha/2,\alpha} \le 2\,\|\boldsymbol{m}_0\|_{\alpha/2,\alpha}\left\|\sum_{i=0}^{j-1}\boldsymbol{R}_i\right\|_{\alpha/2,\alpha} + \left(\sum_{i=0}^{j-1}\|\boldsymbol{R}_i\|_{\alpha/2,\alpha}\right)^2.$$

All in all, we obtain

$$\|\boldsymbol{R}_{j+1}\|_{1+\alpha/2,2+\alpha} \le \tilde{C}Q_j\,\|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}, \tag{31}$$

where $\tilde{C} > 0$ is independent of j and

$$\begin{aligned} Q_j :={}& |\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha} + \|\boldsymbol{m}_0\|_{\alpha/2,\alpha}\left\|\sum_{i=0}^{j-1}\boldsymbol{R}_i\right\|_{\alpha/2,\alpha} + \left(\sum_{i=0}^{j-1}\|\boldsymbol{R}_i\|_{\alpha/2,\alpha}\right)^2 \\ &+ \left(|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha} + C_{\boldsymbol{g}}\right)\left(1 + \|\boldsymbol{m}_j\|_{1+\alpha/2,2+\alpha}\right)^2 \\ &+ \|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}\left(1 + (1 + C_{\boldsymbol{g}})\,|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}\right) + \|\boldsymbol{R}_j\|^2_{1+\alpha/2,2+\alpha}(1 + C_{\boldsymbol{g}}). \end{aligned}$$

It can be proved that for any $q \in (0,1)$ there exists $\varepsilon > 0$ such that $\tilde{C}Q_j < q$ for all $j \in \mathbb{N}$. One proceeds by induction, as done in [29], using additionally the assumption on the smallness of $\nabla \boldsymbol{g}$ and $\Delta \boldsymbol{g}$. Therefore, $\|\boldsymbol{R}_{j+1}\|_{1+\alpha/2,2+\alpha} \le q\,\|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}$, which implies that $(\boldsymbol{m}_\ell)_\ell$ is a Cauchy sequence in $C^{1+\alpha/2,2+\alpha}(D_T)$. Hence, we find a limit $\boldsymbol{m} \in C^{1+\alpha/2,2+\alpha}(D_T)$ and the arguments above already imply the estimate in Theorem 8.

Step 3 ($\boldsymbol{m}$ solves (25)): $\boldsymbol{m}$ fulfils the initial condition $\boldsymbol{m}(0) = \boldsymbol{M}_0$ (and thus $|\boldsymbol{m}(0)| = 1$) and boundary condition $\partial_n \boldsymbol{m} = 0$ on $[0,T] \times \partial D$ by the continuity of the trace operator. The continuity of $\mathcal{R}_a$ and the contraction (31) imply

$$\|\mathcal{R}_a(\boldsymbol{m})\|_{\alpha/2,\alpha} = \lim_{\ell} \|\mathcal{R}_a(\boldsymbol{m}_\ell)\|_{\alpha/2,\alpha} \lesssim \lim_{\ell} \|\boldsymbol{R}_\ell\|_{1+\alpha/2,2+\alpha} \leq \lim_{\ell} q^\ell \|\boldsymbol{R}_0\|_{1+\alpha/2,2+\alpha} = 0$$

The arguments of the proof of [29, Lemma 4.8] show that $\mathcal{R}_a(\boldsymbol{m}) = 0$ implies that $\boldsymbol{m}$ solves (25) and hence concludes the proof. □

5 Holomorphic regularity of parameter-to-solution map with Hölder sample paths

In this section we frequently work with complex-valued functions. If not mentioned otherwise, Banach spaces of functions such as $L^2(D)$ are understood to contain complex valued functions. To denote the codomain explicitly, we write e.g. $L^2(D;\mathbb{C})$ or $L^2(D;\mathbb{R})$.

We specify a possible choice of Banach spaces used in Section 2 for the case of the SLLG problem. Fix $0 < \alpha < 1$ and consider the parameter set

$$\mathcal{X} = \mathcal{X}(\alpha) := \left\{ \boldsymbol{z} \in \mathbb{C}^{\mathbb{N}} \,:\, \|\boldsymbol{z}\|_{\mathcal{X},\alpha} < \infty \right\}, \quad \text{where } \|\boldsymbol{z}\|_{\mathcal{X},\alpha} := \sum_{\ell \in \mathbb{N}_0} \max_{j=1,\dots,\lceil 2^{\ell-1} \rceil} |z_{\ell,j}|\, 2^{-(1-\alpha)\ell/2}. \tag{32}$$

where we used the hierarchical indexing (5). For real parameters consider

$$\mathcal{X}_{\mathbb{R}} := \mathcal{X} \cap \mathbb{R}^{\mathbb{N}}. \tag{33}$$

The definition of $\mathbb{W}$, $\mathbb{W}_{\mathbb{R}}$ follows from the Lévy-Ciesielski expansion (4). It is however interesting to identify classical spaces to which they belong.

Remark 11. *In the regularity results used below, we have to work in Hölder spaces with $\alpha \in (0,1)$. For the Faber-Schauder basis functions on $[0,1]$ (see Section 2.1) we have*

$$\|\eta_{\ell,j}\|_{L^\infty([0,1])} \leq 2^{-\ell/2}, \quad |\eta_{\ell,j}|_{C^1([0,1])} \leq 2^{\ell/2}, \quad \textit{and} \quad \|\eta_{\ell,j}\|_{C^\alpha([0,1])} \leq 2 \cdot 2^{-(1/2-\alpha)\ell}.$$

Only for $\alpha \ll 1$, we obtain a decay of $\|\eta_{\ell,j}\|_{C^\alpha([0,1])}$ close to $2^{-\ell/2}$, which is what we expect for a truncated Brownian motion. Hence, in the following we will assume that $\alpha > 0$ is arbitrarily small.

It can be proved that

$$\mathbb{W}_{\mathbb{R}} \subset C^{\alpha/2}([0,T];\mathbb{R}) \tag{34}$$
$$\mathbb{W} \subset C^{\alpha/2}([0,T]) \tag{35}$$

with the same techniques used in the proof of Lemma 15 below. This choice of parameter space is motivated by the fact that the sample paths of the Wiener process belong to $C^{1/2-\varepsilon}([0,T])$ almost surely for any $\varepsilon > 0$. To define the space of solutions $\mathbb{U}$, write the magnetizations as

$$\boldsymbol{m}(\omega,t,\boldsymbol{x}) = \boldsymbol{M}_0(\boldsymbol{x}) + u(\omega,t,\boldsymbol{x}) \qquad \text{for a.a. } \omega \in \Omega, (t,\boldsymbol{x}) \in D_T,$$

where we recall $\boldsymbol{M}_0$ is the given initial condition, which we assume to belong to $C^{2+\alpha}(D)$. Consider then

$$u \in \mathbb{U} = C_0^{1+\alpha/2,2+\alpha}(D_T) := \left\{ \boldsymbol{v} \in C^{1+\alpha/2,2+\alpha}(D_T) : \boldsymbol{v}(0) = \boldsymbol{0} \text{ in } D,\ \partial_n \boldsymbol{v} = \boldsymbol{0} \text{ on } \partial D \right\}, \tag{36}$$
$$\mathbb{U}_{\mathbb{R}} = \left\{ \boldsymbol{v} : D_T \to \mathbb{S}^2 \,:\, \boldsymbol{v} \in \mathbb{U} \right\}, \tag{37}$$

where $\mathbb{S}^2$ is the unit sphere in $\mathbb{R}^3$. See Section 4.1 for the definition of the relevant Hölder spaces. Given a noise coefficient $\boldsymbol{g} \in C^{2+\alpha}(D)$, we define the residual as:

$$\begin{aligned} \mathcal{R}(W,u) &:= \widetilde{\mathcal{R}}(W, \boldsymbol{M}_0 + u), \qquad \text{where} \\ \widetilde{\mathcal{R}}(W,\boldsymbol{m}) &:= \partial_t \boldsymbol{m} - \Delta \boldsymbol{m} - \boldsymbol{m} \times \Delta \boldsymbol{m} + (\nabla \boldsymbol{m} : \nabla \boldsymbol{m})\, \boldsymbol{m} - \boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{m}) + \\ &\quad + \boldsymbol{m} \times \left(\boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{m}) \right). \end{aligned} \tag{38}$$

Here, the cross product $\times$ is defined as in the real setting by

$$\boldsymbol{a} \times \boldsymbol{b} = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1) \qquad \text{for all } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^3.$$

Note that due to the sesquilinear complex scalar product this implies that $\langle \boldsymbol{a} \times \boldsymbol{b}, \boldsymbol{a} \rangle$ might not vanish for complex valued vector fields $\boldsymbol{a}, \boldsymbol{b}$. Finally, the space of residuals is

$$R = C^{\alpha/2,\alpha}(D_T),$$

so that $\mathcal{R}$ is understood as a function between Banach spaces:

$$\mathcal{R} : \mathbb{W} \times \mathbb{U} \to R, \qquad (W, \boldsymbol{m}) \mapsto \mathcal{R}(W, \boldsymbol{m}). \tag{39}$$

Observe that we already proved Assumption 1 in Theorem 8.

5.1 Proof of Assumptions 2 and 3

In order to apply the general strategy outlined in Section 2, we need to prove Assumption 2 and 3 for the problem defined by (38).

To this end, we apply the following lemma found in much more general form, e.g., in [40, Chapter VII, § 10, Theorem 10.3].

Lemma 12 (Well posedness of linear parabolic systems with Hölder coefficients)**.** *Consider $d \in \mathbb{N}$, $0 < \alpha < 1$, $D \subset \mathbb{R}^d$ bounded with $\partial D \in C^{2+\alpha/2}$, $T > 0$ and define $D_T := [0,T] \times D$. Denote by a_{ij}, a_i, a for $i, j = 1, \ldots, d$ real scalar functions in $C^{1+\alpha/2,2+\alpha}(D_T)$. Let $\mathcal{L} = \sum_{i,j=1}^{3} a_{i,j} D_i D_j + \sum_{i=1}^{3} a_i D_i + a\mathrm{id}$ denote a vector-valued, linear second-order operator. Assume moreover that the system $\partial_t + \mathcal{L}$ is strongly parabolic in the usual sense (see, e.g., [40, Chapter VII, § 8, Definition 7]). Consider $f \in C^{\alpha/2,\alpha}(D_T)$. Then, the problem*

$$\begin{cases} \partial_t u + \mathcal{L} u = f & \text{in } D_T, \\ u(0,\cdot) = 0 & \text{on } D, \\ \partial_n u = 0 & \text{on } [0,T] \times \partial D \end{cases}$$

has a unique solution $u \in C^{1+\alpha/2,2+\alpha}(D_T)$ with $\|u\|_{1+\alpha/2,2+\alpha} \leq C_{\mathrm{stab}} \|f\|_{\alpha/2,\alpha}$. The constant C_{stab} depends on the respective norms of the coefficients a_{ij}, a_i, a as well as on the ellipticity constant.

Remark 13. *Note that the compatibility conditions in [40, Chapter VII, § 10, Theorem 10.3] of order zero ($\alpha < 1$) are automatically satisfied in our case. This also takes care of the fact that [40, Chapter VII, § 10, Theorem 10.3] only works for small end times $0 < \widetilde{T} \leq T$ as we can restart the estimate at any time $\widetilde{T}$ and get the estimate for the full time interval. Moreover, while not stated explicitly, analysing the proof of [40, Chapter VII, § 10, Theorem 10.3] gives the dependence of C_{stab} on the coefficients of the problem.*

Lemma 14. *Let $\alpha \in (0,1)$, $\boldsymbol{g} \in C^{2+\alpha}(D)$ and $\boldsymbol{M}_0 \in C^{2+\alpha}(D)$. Consider the spaces $\mathbb{W}, \mathbb{W}_{\mathbb{R}}, \mathbb{U}, \mathbb{U}_{\mathbb{R}}, R$ defined at the beginning of the present section. Then, the residual $\mathcal{R}$ (cf. (38), (39)) is a well-defined function and Assumptions 2 holds true. More generally, it can be proved that $\partial_u \mathcal{R}(W,u)$ is a homeomorphism between $\mathbb{U}$ and R if*

$$W \in \mathbb{W} \quad \text{and} \quad u \in \mathbb{U} \text{ satisfies } \|\mathfrak{Im}\,(u)\|_{L^\infty(D_T)} \leq \frac{1}{4}. \tag{40}$$

Finally, Assumption 3 also holds true with $\varepsilon_W > 0$, $\varepsilon_u = \frac{1}{4}$, and

$$\begin{aligned} \mathcal{G}_1(s) &= (1 + e^{\varepsilon_W}(1+\varepsilon_W))^2 \left(1 + \|\boldsymbol{g}\|_{C^{2+\alpha}(D)}\right)^4 \left(1 + \|\boldsymbol{M}_0\|_{C^{2+\alpha}(D)} + s\right)^3, \\ \mathcal{G}_2(s) &= C_{\mathrm{stab}}(s) \qquad \text{for all } s \geq 0, \end{aligned}$$

and $C_{\mathrm{stab}} = C_{\mathrm{stab}}(\|u\|_{\mathbb{U}}) > 0$ is as in Lemma 12, i.e. it guarantees that $\left\|(\partial_u \mathcal{R}(W,u))^{-1} f\right\|_{\mathbb{U}} \leq C_{\mathrm{stab}}(\|u\|_{\mathbb{U}}) \|f\|_R$ for any $f \in R$, $W \in \mathbb{W}$, $u \in \mathbb{U}$.

Proof that $\mathcal{R}$ is well-defined. Let us first show that the residual $\mathcal{R}$ is a well-defined function. Clearly, $\boldsymbol{M}_0 + u \in C^{1+\alpha/2,2+\alpha}(D_T)$ if $u \in C_0^{1+\alpha/2,2+\alpha}(D_T)$. Observe that

$$G\colon C^{1+\alpha/2,2+\alpha}(D_T) \to C^{1+\alpha/2,2+\alpha}(D_T) \quad \text{and} \quad \mathcal{C}\colon C^{\alpha/2,1+\alpha}(D_T) \to C^{\alpha/2,\alpha}(D_T),$$

so $\hat{\mathcal{C}}(W, \boldsymbol{m}) \in C^{\alpha/2,\alpha}(D_T)$. Thus, $\mathcal{R}(W,u)$ is a sum of functions belonging to $C^{\alpha/2,\alpha}(D_T)$. The fact that $\mathcal{R}$ is continuous can be easily verified by checking that each term (cf. (38)) is continuous. □

Proof of (i) *in Assumption 2.* The residual $\mathcal{R}$ is differentiable because it is a linear combination of differentiable functions. We now prove that each partial derivative is continuous. For $\omega \in C^{\alpha/2}([0,T])$,

$$\partial_W \mathcal{E}(W,\boldsymbol{m})[\omega] = (\cos(W)\mathcal{C}\boldsymbol{m} + \sin(W)(G\mathcal{C}+\mathcal{C}G)\boldsymbol{m})\,\omega, \tag{41}$$

$$\partial_W \hat{\mathcal{C}}(W,\boldsymbol{m})[\omega] = e^{WG}\partial_1\mathcal{E}(W,\boldsymbol{m})[\omega] + \left(\cos(W)G\mathcal{E}(W,\boldsymbol{m}) + \sin(W)G^2\mathcal{E}(W,\boldsymbol{m})\right)\omega, \tag{42}$$

$$\partial_W \widetilde{\mathcal{R}}(W,\boldsymbol{m})[\omega] = -\boldsymbol{m}\times\partial\hat{\mathcal{C}}(W,\boldsymbol{m})[\omega] + \boldsymbol{m}\times\left(\boldsymbol{m}\times\partial\hat{\mathcal{C}}(W,\boldsymbol{m})[\omega]\right). \tag{43}$$

Formally estimating the linear operator $\partial_W\mathcal{R}(W,u)$ gives that for all $\omega\in C^{\alpha/2}([0,T])$

$$\begin{aligned}\|\partial_W\mathcal{R}(W,u)[\omega]\|_{C^{\alpha/2,\alpha}(D_T)} \le \left(1+\left\|e^{\mathfrak{Im}(W)}\right\|_{C^{\alpha/2}([0,T])}\right)^2\left(1+\|\boldsymbol{g}\|_{C^{2+\alpha}(D)}\right)^4\\ \left(1+\|\boldsymbol{M}_0+u\|_{C^{\alpha/2,1+\alpha}(D_T)}\right)^3\|\omega\|_{C^{\alpha/2}([0,T])}.\end{aligned} \tag{44}$$

The exponential dependence on $\mathfrak{Im}\,(W)$ comes from the exponential behaviour of sine and cosine in imaginary direction. The right-hand-side is finite because $\|\mathfrak{Im}\,(W)\|_{\mathbb{W}}\le\varepsilon$ implies $\left\|e^{\mathfrak{Im}(W)}\right\|_{C^{\alpha/2}([0,T])}\lesssim e^{\varepsilon}(1+\varepsilon)$. Indeed, $\left\|e^{\mathfrak{Im}(W)}\right\|_{L^\infty([0,T])} = e^{\|W\|_{L^\infty([0,T])}}\le e^{\varepsilon}$ and

$$\begin{aligned}\left|e^{\mathfrak{Im}(W)}\right|_{C^{\alpha/2}([0,T])} &= \sup_{\substack{s,t\in[0,T]\\ s\neq t}}\frac{\left|e^{\mathfrak{Im}(W(s))}-e^{\mathfrak{Im}(W(t))}\right|}{|s-t|^{\alpha/2}}\\ &\le \sup_{\substack{s,t\in[0,T]\\ s\neq t}}\frac{\left|e^{\mathfrak{Im}(W(s))}-e^{\mathfrak{Im}(W(t))}\right|}{|\mathfrak{Im}\,(W(s))-\mathfrak{Im}\,(W(t))|}\sup_{\substack{s,t\in[0,T]\\ s\neq t}}\frac{|\mathfrak{Im}\,(W(s))-\mathfrak{Im}\,(W(t))|}{|s-t|^{\alpha/2}}.\end{aligned}$$

Because of the assumption on $\mathfrak{Im}\,(W)$, we have that $|\mathfrak{Im}\,(W(t))|\le\varepsilon$ for all $t\in[0,T]$ and $\sup_{\substack{s,t\in[0,T]\\ s\neq t}}\frac{|\mathfrak{Im}(W(s))-\mathfrak{Im}(W(t))|}{|s-t|^{\alpha/2}}\le\varepsilon$. Thus,

$$\sup_{s,t\in[0,T],s\neq t}\frac{\left|e^{\mathfrak{Im}(W(s))}-e^{\mathfrak{Im}(W(t))}\right|}{|\mathfrak{Im}\,(W(s))-\mathfrak{Im}\,(W(t))|} = \sup_{\substack{-\varepsilon\le a,b\le\varepsilon\\ a\neq b}}\frac{\left|e^a-e^b\right|}{|a-b|}\lesssim e^b,$$

where the last inequality is a consequence of the Taylor expansion $e^a = e^b + e^b(a-b) + \mathcal{O}(|a-b|^2)$. All in all, we obtain $\left|e^{\mathfrak{Im}(W)}\right|_{C^{\alpha/2}([0,T])}\lesssim e^{\varepsilon}\varepsilon$ and

$$\left\|e^{\mathfrak{Im}(W)}\right\|_{C^{\alpha/2}([0,T])}\lesssim e^{\varepsilon}(1+\varepsilon). \tag{45}$$

For $\boldsymbol{v}\in C^{1+\alpha/2,2+\alpha}(D_T)$, we get

$$\begin{aligned}\partial_{\boldsymbol{m}}\widetilde{\mathcal{R}}(W,\boldsymbol{m})[\boldsymbol{v}] &= \partial_t\boldsymbol{v} - \Delta\boldsymbol{v} - \boldsymbol{v}\times\Delta\boldsymbol{m} - \boldsymbol{m}\times\Delta\boldsymbol{v} + 2(\nabla\boldsymbol{v}:\nabla\boldsymbol{m})\boldsymbol{m} + (\nabla\boldsymbol{m}:\nabla\boldsymbol{m})\,\boldsymbol{v}\\ &\quad-\left(\boldsymbol{v}\times\hat{\mathcal{C}}(W,\boldsymbol{m}) + \boldsymbol{m}\times\hat{\mathcal{C}}(W,\boldsymbol{v})\right)\\ &\quad-\left(\boldsymbol{v}\times\left(\boldsymbol{m}\times\hat{\mathcal{C}}(W,\boldsymbol{m})\right) + \boldsymbol{m}\times\left(\boldsymbol{v}\times\hat{\mathcal{C}}(W,\boldsymbol{m}) + \boldsymbol{m}\times\hat{\mathcal{C}}(W,\boldsymbol{v})\right)\right),\end{aligned} \tag{46}$$

and continuity of $\partial_u\mathcal{R}(W,u) = \partial_{\boldsymbol{m}}\widetilde{\mathcal{R}}(W,\boldsymbol{M}_0+u)$ follows by the same arguments used for $\partial_W\mathcal{R}(W,\boldsymbol{m})$. □

Proof of (ii) *in Assumption 2.* While we are only interested in the case $W\in\mathbb{W}_{\mathbb{R}}$, $u\in\mathbb{U}_{\mathbb{R}}$ such that $\mathcal{R}(W,u)=0$, let us consider the more general case (40) for future use. Consider $\boldsymbol{f}\in R$ (the residuals space) and the problem

$$\begin{cases}\partial_u\mathcal{R}(W_*,u_*)[\boldsymbol{v}] &= \boldsymbol{f} \quad \text{in } D_T,\\ \partial_n\boldsymbol{v} &= 0 \quad \text{on } [0,T]\times\partial D,\\ \boldsymbol{v}(0,\cdot) &= \boldsymbol{0} \quad \text{on } D.\end{cases}$$

With the aim of applying Lemma 12, we note that the principal part of $\partial_2\mathcal{R}(W,u)[\boldsymbol{v}]$ is $-\Delta\boldsymbol{v} - u\times\Delta\boldsymbol{v}$. We now show that for any $(t,\boldsymbol{x})\in D_T$ and $\boldsymbol{w}\in\mathbb{C}^3$,

$$\mathfrak{Re}\left(\langle\boldsymbol{w}+u(t,\boldsymbol{x})\times\boldsymbol{w},\boldsymbol{w}\rangle\right)\ge\frac{1}{2}\|\boldsymbol{w}\|^2, \tag{47}$$

where here $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote respectively the standard norm and scalar product on $\mathbb{C}^3$. Indeed,

$$\mathfrak{Re}\left(\langle \boldsymbol{w} + u(t,\boldsymbol{x}) \times \boldsymbol{w}, \boldsymbol{w}\rangle\right) = \|\boldsymbol{w}\|^2 + \mathfrak{Re}\left(\langle u(t,\boldsymbol{x}) \times \boldsymbol{w}, \boldsymbol{w}\rangle\right)$$

and algebraic manipulations lead to the identity

$$\mathfrak{Re}\left(\langle u(t,\boldsymbol{x}) \times \boldsymbol{w}, \boldsymbol{w}\rangle\right) = 2\left\langle \mathfrak{Im}\left(\boldsymbol{w}\right) \times \mathfrak{Re}\left(\boldsymbol{w}\right), \mathfrak{Im}\left(u(t,\boldsymbol{x})\right)\right\rangle,$$

which implies the estimate

$$\left|\mathfrak{Re}\left(\langle u(t,\boldsymbol{x}) \times \boldsymbol{w}, \boldsymbol{w}\rangle\right)\right| \leq 2\left\|\mathfrak{Im}\left(u(t,\boldsymbol{x})\right)\right\|_{L^\infty(D_T)} \|\boldsymbol{w}\|^2.$$

Thus, by virtue of Assumption (40), we obtain (47). This shows that $\partial_2\mathcal{R}(W,u)$ is parabolic in the sense of Lemma 12 and hence, we obtain that $\partial_2\mathcal{R}(W,u)$ admits a continuous inverse. Together with its continuity, this implies that it is a homeomorphism. The norm of the inverse can be estimated as

$$\left\|\partial_u\mathcal{R}(W,u)^{-1}[\boldsymbol{f}]\right\|_{C^{1+\alpha/2,2+\alpha}(D_T)} \leq C_{\mathrm{stab}}(W,u)\,\|\boldsymbol{f}\|_{C^{\alpha/2,\alpha}(D_T)}, \tag{48}$$

where $C_{\mathrm{stab}}(W,u) > 0$ is independent of $\boldsymbol{f}$ (but depends on W and u). ☐

Proof of Assumption 3. The continuity bound for $\partial_W\mathcal{R}(W,u)$ follows from (44) and (45) with

$$\mathcal{G}_1(s) = \left(1 + e^{\varepsilon_W}(1+\varepsilon_W)\right)^2\left(1 + \|\boldsymbol{g}\|_{C^{2+\alpha}(D)}\right)^4\left(1 + \|\boldsymbol{M}_0\|_{C^{2+\alpha}(D)} + s\right)^3.$$

where $\varepsilon_W > 0$. The bound on $(\partial_u\mathcal{R}(W,u))^{-1}$ is already proved in (48) with $\varepsilon_u = \frac{1}{4}$ and $\mathcal{G}_2 = C_{\mathrm{stab}}$. The fact that C_{stab} depends on $\mathcal{U}(W)$ only through $\|\mathcal{U}(W)\|_{\mathbb{U}}$ is implied by the sufficient condition for well posedness in (40). ☐

We recall that, as shown in Section 2.2, the implicit function theorem and Theorem 5 prove the existence of $\varepsilon > 0$ such that for any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$ there exists a holomorphic map $\mathcal{U} : B_\varepsilon(W(\boldsymbol{y}) \to \mathbb{U}$ such that $\mathcal{R}(W,\mathcal{U}(W)) = 0$ for all $W \in B_\varepsilon(W(\boldsymbol{y}))$. The function $\mathcal{U}$ is bounded by a constant $C_\varepsilon > 0$ again independent of $\boldsymbol{y}$.

Moreover, Assumption 3 implies the bound (9) on the differential $\mathcal{U}'(W)$ as a function of $\mathcal{U}(W)$ through $\|\mathcal{U}(W)\|_{\mathbb{U}}$ under the assumption that $W \in B_\varepsilon(W(\boldsymbol{y}))$ in $\mathbb{W}$.

5.2 Proof of Assumption 4 and estimates of derivatives of parameter-to-solution map

Let us now estimate the derivatives of the parameter-to-solution map. While this is a standard technique established already in [21], it turns out this will not be quite sharp enough to obtain dimension independent convergence of the sparse grid approximation. In Section 6, we present a possible way to resolve this in the future.

Let us show that Assumption 4 holds for the present problem. Recall the definitions of parameter spaces in (32) and (33).

Lemma 15. *Assumption 4 holds in the present setting. In particular, it is sufficient to choose $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ such that*

$$\|\boldsymbol{\rho}\|_{\mathcal{X}} \leq \frac{\varepsilon}{2} \tag{49}$$

.

Proof. Fix $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$ and $\boldsymbol{z} \in \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y})$ (i.e.$|z_n - y_n| < \rho_n$ for all $n \in \mathbb{N}$). Let us prove that $W(\boldsymbol{z}) \in B_\varepsilon(W(\boldsymbol{y}))$. By linearity, $W(\boldsymbol{z},\cdot) - W(\boldsymbol{y},\cdot) = \sum_{n\in\mathbb{N}}(z_n - y_n)\eta_n(\cdot)$. Recalling the hierarchical indexing (5) and by a triangle inequality, we obtain

$$\|W(\boldsymbol{z},\cdot) - W(\boldsymbol{y},\cdot)\|_{C^{\alpha/2}([0,T])} \leq \sum_{\ell\in\mathbb{N}_0}\left\|\sum_{j=1}^{\lceil 2^{\ell-1}\rceil}(z_{\ell,j} - y_{\ell,j})\eta_{\ell,j}\right\|_{C^{\alpha/2}([0,T])}.$$

The terms on the right-hand side can be estimated by Banach space interpolation and the fact that all basis functions $\eta_{\ell,j}$ on the same level have disjoint supports, i.e.,

$$\begin{aligned}
\left\|\sum_j(z_{\ell,j} - y_{\ell,j})\eta_{\ell,j}\right\|_{C^{\alpha/2}([0,T])} &\leq \left\|\sum_j(z_{\ell,j} - y_{\ell,j})\eta_{\ell,j}\right\|_{C^0([0,T])}^{1-\alpha/2}\left\|\sum_j(z_{\ell,j} - y_{\ell,j})\eta_{\ell,j}\right\|_{C^1([0,T])}^{\alpha/2} \\
\leq &\Big(\max_j|z_{\ell,j} - y_{\ell,j}|\,\|\eta_{\ell,j}\|_{C^0([0,T])}\Big)^{1-\alpha/2}\Big(\max_j|z_{\ell,j} - y_{\ell,j}|\,\|\eta_{\ell,j}\|_{C^0([0,T])} + \max_j|z_{\ell,j} - y_{\ell,j}|\,|\eta_{\ell,j}|_{C^1([0,T])}\Big)^{\alpha/2}.
\end{aligned}$$

Recalling that $\|\eta_{i(\ell)}\|_{C^0([0,T])} \leq 2^{-\ell/2}$ and $|\eta_{i(\ell)}|_{C^1([0,T])} \leq 2^{\ell/2}$ (see Remark 11), we find

$$\left\| \sum_j (z_{\ell,j} - y_{\ell,j})\eta_{\ell,j} \right\|_{C^{\alpha/2}([0,T])} \leq \max_j |z_{\ell,j} - y_{\ell,j}| \left(2^{-\ell/2} + 2^{-(1-\alpha)\ell/2}\right).$$

With $\boldsymbol{z} \in \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y})$, we obtain $\|W(\boldsymbol{z},\cdot) - W(\boldsymbol{y},\cdot)\|_{C^{\alpha/2}([0,T])} < \varepsilon$, which gives the statement. □

An example of valid sequence of holomorphy radii is

$$\rho_n = \varepsilon 2^{\frac{(1-\alpha)\lceil \log_2(n) \rceil}{2}} \qquad \text{for all } n \in \mathbb{N}. \tag{50}$$

Having so concluded that for any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$ the parameter-to-solution map $\mathcal{M} \circ W : \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y}) \to \mathbb{U}$ is holomorphic and uniformly bounded, we can estimate its derivatives as done in Theorem 6.

Proposition 16. *Consider $\boldsymbol{m} = \boldsymbol{M}_0 + u : \mathcal{X}_{\mathbb{R}} \to C^{1+\alpha/2,2+\alpha}(D_T)$, the parameter-to-solution map of the parametric LLG problem with Hölder spaces ($\mathcal{X}_{\mathbb{R}}$ and $C^{1+\alpha/2,2+\alpha}(D_T)$ defined in* (33) *and* (23) *respectively). Fix $\varepsilon > 0$ as in Theorem 5 and let $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ a positive sequence that satisfies* (49)*. Then, for any $n \in \mathbb{N}$, $\boldsymbol{\nu} = (\nu_i)_{i=1}^n \subset \mathbb{N}^n$, it holds that*

$$\|\partial^{\boldsymbol{\nu}} \boldsymbol{m}(\boldsymbol{y})\|_{C^{1+\alpha/2,2+\alpha}(D_T)} \leq \prod_{j=1}^n \nu_j! \rho_j^{-\nu_j} C_\varepsilon \qquad \text{for all } \boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}, \tag{51}$$

where $C_\varepsilon > 0$ from Theorem 5 is independent of $\boldsymbol{\nu}$ or $\boldsymbol{y}$.

Remark 17. *Note that we essentially proved "$(\boldsymbol{b}, \xi, \delta, X)$-holomorphy" [26, Definition 4.1] for the Stochastic LLG problem in the case of a Hölder-valued parameter-to-solution map. However, this regularity is not sufficient to apply the theory in [26], as the summability coefficient is $p = 2$, which lies out of the range $(0, \frac{2}{3})$ considered in [26]. This fact is analogous to what happens in our analysis.*

6 Holomorphy of a simplified parameter-to-solution map with Lebesgue sample paths

In the present section, we aim at proving stronger regularity and sparsity properties of the random LLG parameter-to-solution map again based on the general strategy outlined in Section 2. A key observation is that these properties depend on the Banach spaces chosen for the sample paths of the random coefficients (in our case, the Wiener process) and the sample paths of the solutions (in our case, the magnetizations). In this case, we show that using *Lebesgue* spaces for the time variable is superior to using Hölder spaces.

Because of the nonlinear nature of the random LLG problem, the results hold only for a simplified version of the stochastic input. We make the following modelling assumptions:

- The sample paths of the Wiener process W are "small". This is justified e.g. for small final times $T \ll 1$ with high probability;
- The gradient $\nabla \boldsymbol{g}$ is "small", meaning that the stochastic noise is spatially uniform. This is justified for small domain sizes (samples in real-world applications are often in the nano- and micrometer range).

This leads to the following simplifications in the random LLG residual (defined in (38)):

$$\begin{aligned} \nabla \boldsymbol{m} \times \nabla \boldsymbol{g} &\approx 0, \\ \sin(W) &\approx W, \\ 1 - \cos(W) &\approx \frac{W^2}{2} \approx 0. \end{aligned}$$

Consequently, we approximate $\hat{\mathcal{C}}(W, \boldsymbol{m})$ defined in (17) with the first order expansion

$$\tilde{\mathcal{C}}(W, \boldsymbol{m}) \coloneqq W\boldsymbol{m} \times \Delta \boldsymbol{g},$$

where $\boldsymbol{g} \in C^{2+\alpha}(D)$. This term appears in the *simplified random LLG residual*

$$\begin{aligned} \mathcal{R}_s(W, u) &\coloneqq \widetilde{\mathcal{R}}_s(W, \boldsymbol{M}_0 + u), \qquad \text{where} \\ \widetilde{\mathcal{R}}_s(W, \boldsymbol{m}) &\coloneqq \partial_t \boldsymbol{m} - \Delta \boldsymbol{m} - \boldsymbol{m} \times \Delta \boldsymbol{m} + (\nabla \boldsymbol{m} : \nabla \boldsymbol{m})\, \boldsymbol{m} - \boldsymbol{m} \times \tilde{\mathcal{C}}(W, \boldsymbol{m}) + \\ &\quad + \boldsymbol{m} \times \left(\boldsymbol{m} \times \tilde{\mathcal{C}}(W, \boldsymbol{m})\right). \end{aligned} \tag{52}$$

Observe that the magnetization corresponding to $W(\omega,\cdot)$ is $\boldsymbol{m}(\omega)=\boldsymbol{M}_0+u(\omega)$ for any $\omega\in\Omega$.

In order to define the space for the coefficients, we again start from the parameters: Define, for $1<q<\infty$,

$$\mathcal{X}=\mathcal{X}^q:=\left\{\boldsymbol{z}\in\mathbb{C}^{\mathbb{N}}\,:\,\|\boldsymbol{z}\|_{\mathcal{X}^q}<\infty\right\},\qquad \text{where } \|\boldsymbol{z}\|_{\mathcal{X}^q}:=\sum_{\ell\in\mathbb{N}_0}|\boldsymbol{y}_\ell|_{\ell^q}\,2^{-\ell(1/2+1/q)}.$$

and we denoted $\boldsymbol{y}_\ell=(y_{\ell,1},\dots,y_{\ell,\lceil 2^{\ell-1}\rceil})$. We then define the space of (complex) coefficients through the Lévy-Ciesielski expansion (4): $\mathbb{W}=\{W(\boldsymbol{z},\cdot):[0,T]\to\mathbb{C}\,:\,\boldsymbol{z}\in\mathcal{X}\}$. For real parameters, we fix $\theta>0$ and let

$$\mathcal{X}_{\mathbb{R}}=\mathcal{X}(\alpha,\theta):=\left\{\boldsymbol{y}\in\mathbb{R}^{\mathbb{N}}\,:\,\|\boldsymbol{y}\|_{\mathcal{X}(\alpha)}<\theta\right\},\tag{53}$$

where $\mathcal{X}(\alpha)$ was defined in (32).

Lemma 18. *For fixed $1<q<\infty$ and $\theta>0$, there holds,*

$$\mathbb{W}\subset L^q([0,T])\quad\textit{and}\quad \mathbb{W}_{\mathbb{R}}\subset\left\{W\in C^\alpha([0,T];\mathbb{R})\,:\,\|W\|_{C^\alpha([0,T])}<\theta\right\}.$$

Proof. To prove the first inclusion, fix $\boldsymbol{z}\in\mathcal{X}$ and estimate

$$\|W(\boldsymbol{z})\|_{L^q([0,T])}=\left\|\sum_{\ell\in\mathbb{N}_0}\sum_{j=1}^{\lceil 2^{\ell-1}\rceil}z_{\ell,j}\eta_{\ell,j}\right\|_{L^q([0,T])}\le\sum_{\ell\in\mathbb{N}_0}\left\|\sum_{j=1}^{\lceil 2^{\ell-1}\rceil}z_{\ell,j}\eta_{\ell,j}\right\|_{L^q([0,T])}.$$

Examine one summand at a time to get, using the fact that Faber-Schauder basis functions of same level have disjoint supports,

$$\left\|\sum_{j=1}^{\lceil 2^{\ell-1}\rceil}y_{\ell,j}\eta_{\ell,j}\right\|_{L^q([0,T])}^q=\int_0^T\sum_{j=1}^{\lceil 2^{\ell-1}\rceil}y_{\ell,j}^q\eta_{\ell,j}^q=\sum_{j=1}^{\lceil 2^{\ell-1}\rceil}y_{\ell,j}^q\int_0^T\eta_{\ell,j}^q=|\boldsymbol{y}_\ell|_{\ell^q}^q\,\|\eta_{\ell,1}\|_{L^q([0,T])}^q.$$

Finally, simple computations reveal that (cf. (3)) $\|\eta_{\ell,j}\|_{L^q([0,T])}=2^{-\ell(1/2+1/q)}2^{-1/2}\left(\frac{2}{q+1}\right)^{1/q}$, so we get $\|W(\boldsymbol{z})\|_{L^q([0,T])}\le\|\boldsymbol{z}\|_{\mathcal{X}^q}$, which implies the first inclusion. The second inclusion follows with the methods of the proof of Lemma 15. □

Intuitively, $\mathbb{W}_{\mathbb{R}}$ can be understood as the set of “small” real valued Wiener processes.
The space of solutions is chosen as

$$\mathbb{U}=\{u:D_T\to\mathbb{C}^3: u\in L^q([0,T],C^{2+\alpha}(D)),\ \partial_t u\in L^q([0,T],C^\alpha(D)),\tag{54}$$

$$u(0,\cdot)=\mathbf{0}\text{ on }D,\ \partial_n u=\mathbf{0}\text{ on }[0,T]\times\partial D\},\tag{55}$$

$$\mathbb{U}_{\mathbb{R}}=\left\{u:D_T\to\mathbb{S}^2\,:\,u\in\mathbb{U}\right\}.\tag{56}$$

Finally the space of residuals is chosen as $R:=L^q([0,T],C^\alpha(D))$. The map $\mathcal{R}_s$ is understood as a function between Banach spaces:

$$\mathcal{R}_s:\mathbb{W}\times\mathbb{U}\to R,\tag{57}$$

Observe that if $u\in\mathbb{U}$ for $q>1$, then $\|u(t)\|_{C^\alpha(D)}\le\|\partial_t u\|_{L^1([0,T],C^\alpha(D))}$ for all $t\in[0,T]$. This implies that $u\in C^0([0,T],C^\alpha(D))$ and $\|u\|_{C^0([0,T],C^\alpha(D))}\le\|u\|_{\mathbb{U}}$. In particular, interpolation shows that for any $U\in\mathbb{U}$, $\|u\|_{L^\infty(D_T)}+\|u\|_{L^2([0,T],C^1(D))}\le\|u\|_{\mathbb{U}}$. Note that $\tilde{\mathcal{C}}$ is bounded and linear in both arguments: For all $W\in\mathbb{W},\boldsymbol{m}\in C^0([0,T],C^\alpha(D))$ it holds

$$\left\|\tilde{\mathcal{C}}(W,\boldsymbol{m})\right\|_{L^q([0,T],C^\alpha(D))}\le\|W\|_{L^q([0,T])}\,\|\boldsymbol{m}\|_{C^0([0,T],C^\alpha(D))}\,\|\boldsymbol{g}\|_{C^{2+\alpha}(D)}.\tag{58}$$

The proof of Theorem 8 can be transferred to this simplified version of LLG and hence we have that there exists $\overline{C}_r=\overline{C}_r(\theta)>0$ such that

$$\|\mathcal{U}(W)\|_{\mathbb{U}}\le\overline{C}_r\qquad\text{for all }W\in\mathbb{W}_{\mathbb{R}}.\tag{59}$$

This gives the validity of Assumption 1 with $C_r=\overline{C}_r$ for the present problem.

6.1 Proof of Assumptions 2 and 3

In order to apply the general strategy outlined in Section 2.2, we need to prove Assumptions 2 and 3 for the spaces and residual chosen at the beginning of this section.

Remark 19. *The proof of* ii. *in Assumption 2 requires the use of a L^q-regularity result for the linear parabolic problem given by the operator $\partial_u \mathcal{R}_s(W,u) : \mathbb{U} \to R$ which coincides with* (46) *but $\hat{\mathcal{C}}$ replaced by $\tilde{\mathcal{C}}$. For scalar problems, this can be found in [51, Section 4]. Strictly speaking, however, Lemma 20 only holds under the assumption that [51] can be generalized to the vector valued case.*

We can prove, analogously to Lemma 14, the following result:

Lemma 20. *Let $\alpha \in (0,1)$, $\boldsymbol{g} \in C^{2+\alpha}(D)$, $\boldsymbol{M}_0 \in C^{2+\alpha}(D)$ and $0 < \theta < \infty$. Consider the spaces $\mathbb{W}, \mathbb{W}_{\mathbb{R}}, \mathbb{U}, \mathbb{U}_{\mathbb{R}}$ defined at the beginning of the present section. Then, the residual $\mathcal{R}_s$ (cf.* (52), (57)*) is a well-defined function and Assumption 2 holds true. More generally, it can be proved that $\partial_u \mathcal{R}_s(W,u)$ is a homeomorphism between $\mathbb{U}$ and R if*

$$W \in \mathbb{W}, u \in \mathbb{U} : \|\mathfrak{Im}\,(u)\|_{L^\infty(D_T)} \leq \frac{1}{4}. \tag{60}$$

Finally, Assumption 3 holds true with $\varepsilon_W > 0$ and $\varepsilon_u = \frac{1}{4}$ and

$$\begin{aligned} \mathcal{G}_1(s) &= \|\boldsymbol{g}\|_{C^{2+\alpha}(D)} \left(1 + \|\boldsymbol{M}_0\|_{\mathbb{U}} + s\right)^3 \\ \mathcal{G}_2(s) &= C_{\mathrm{stab}}(\varepsilon + \theta, s) \qquad \text{for all } s \geq 0, \end{aligned}$$

where $C_{\mathrm{stab}}(\|W\|_{\mathbb{W}}, \|u\|_{\mathbb{U}}) > 0$ is as c_p in [51, Theorem 2.5], i.e. it guarantees that $\left\|(\partial_u \mathcal{R}(W,u))^{-1} f\right\|_{\mathbb{U}} \leq C_{\mathrm{stab}}(\|W\|_{\mathbb{W}}, \|u\|_{\mathbb{U}}) \|f\|_R$ for any $f \in R$, $W \in \mathbb{W}$, $u \in \mathbb{U}$.

We recall that, as shown in Section 2.2, the implicit function theorem and Theorem 5 prove the existence of $\varepsilon > 0$ such that for any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$ there exists a holomorphic map $\mathcal{U} : B_\varepsilon(W(\boldsymbol{y})) \to \mathbb{U}$ such that $\mathcal{R}(W, \mathcal{U}(W)) = 0$ for all $W \in B_\varepsilon(W(\boldsymbol{y}))$. The function $\mathcal{U}$ is bounded by a constant $C_\varepsilon > 0$ again independent of $\boldsymbol{y}$. Moreover, Assumption 3 implies the bound (9) on the differential $\mathcal{U}'(W)$ as a function of $\mathcal{U}(W)$ through $\|\mathcal{U}(W)\|_{\mathbb{U}}$ for all those $W \in B_\varepsilon(W(\boldsymbol{y}))$ in $\mathbb{W}$.

6.2 Proof of Assumption 4 and estimates of derivatives of parameter-to-solution map

Let us now estimate the derivatives of the parameter-to-solution map. To this end, let us find a real positive sequence $\boldsymbol{\rho} = (\rho_n)_n$ that verifies Assumption 4. Contrary to Section 5.2, here $\boldsymbol{\rho}$ depends on which mixed derivative $\partial^{\boldsymbol{\nu}}$ is considered: Given a multi-index $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n) \in \mathbb{N}_0^n$, $0 < \delta < \frac{1}{2}$ and $0 < \gamma < 1$ consider a sequence of positive numbers $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\nu}, \delta, \gamma)$ defined as follows:

$$\rho_{\ell,j} \coloneqq \gamma \begin{cases} 1 & \text{if } \nu_{\ell,j} = 0 \\ 2^{\left(\frac{3}{2}-\delta\right)\ell} \frac{1}{r_\ell(\boldsymbol{\nu})} & \text{if } \nu_{\ell,j} = 1 \\ 2^{\left(\frac{1}{2}-\delta\right)\ell} & \text{otherwise} \end{cases} \qquad \text{for all } \ell \in \mathbb{N}_0, j = 1, \dots, \lceil 2^{\ell-1} \rceil, \tag{61}$$

where we used the hierarchical indexing (5) and $r_\ell(\boldsymbol{\nu}) \coloneqq \#\left\{j \in 1, \dots, \lceil 2^{\ell-1} \rceil : \nu_{\ell,j} = 1\right\}$.

Lemma 21. *Consider a multi-index $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n) \in \mathbb{N}_0^n$, $\delta > 0$ and $1 < q < \frac{1}{1-\delta/2}$. There exists $0 < \gamma < 1$ such that defining $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\nu}, \delta, \gamma)$ as in* (61) *verifies Assumption 4.*

Proof. Let $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$ and $\boldsymbol{z} \in \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y})$. (i.e. $|z_n - y_n| \leq \rho_n$ for all $n \in \mathbb{N}$). A triangle inequality yields: $\|W(\boldsymbol{z}) - W(\boldsymbol{y})\|_{L^q(0,T)} \leq \sum_{\ell \in \mathbb{N}_0} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} |z_{\ell,j} - y_{\ell,j}| \, \|\eta_{\ell,j}\|_{L^q(0,T)}$. For the Faber-Schauder basis functions (3), $\|\eta_{\ell,j}\|_{L^q([0,T])} \leq 2^{-(1/q+1/2)\ell}$ for any $\ell \in \mathbb{N}_0$ and $j = 1, \dots, \lceil 2^{\ell-1} \rceil$. Together with the fact that $\boldsymbol{z} \in \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y})$, this gives

$$\|W(\boldsymbol{z}) - W(\boldsymbol{y})\|_{L^q([0,T])} \leq \sum_{\ell \in \mathbb{N}_0} 2^{-(1/q+1/2)\ell} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} \rho_{\ell,j}. \tag{62}$$

By the definition of $\boldsymbol{\rho}$, we may write

$$\sum_{j=1}^{\lceil 2^{\ell-1} \rceil} \rho_{\ell,j} = \gamma \left(\#\{i : \nu_{\ell,i} = 0\} + 2^{\left(\frac{3}{2}-\delta\right)\ell} \frac{1}{r_\ell(\boldsymbol{\nu})} r_\ell(\boldsymbol{\nu}) + 2^{\left(\frac{1}{2}-\delta\right)\ell} \#\{i : \nu_{\ell,i} > 1\} \right). \tag{63}$$

Trivially, $\#\{i : \nu_{\ell,i} = 0\} \le 2^\ell$ and $\#\{i : \nu_{\ell,i} > 1\} \le 2^\ell$. This, together with (62) and (63) yields

$$\|W(\boldsymbol{z}) - W(\boldsymbol{y})\|_{L^q([0,T])} \le \gamma \sum_{\ell\in\mathbb{N}_0} \left(2^{-(1/q-1/2)\ell} + 2^{-\delta\ell/2} + 2^{-\delta\ell/2}\right).$$

Which is finite as long as $1 < q < \frac{1}{1-\delta/2}$. This implies that there exists $\gamma > 0$ such that $W(\boldsymbol{z}) \in B_\varepsilon(W(\boldsymbol{y}))$. □

Having so concluded that for any $\boldsymbol{y} \in \mathcal{X}_\mathbb{R}$ the parameter-to-solution map $\mathcal{M} \circ W : \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y}) \to \mathbb{U}$ is holomorphic and uniformly bounded, we can estimate its derivatives as done in Theorem 6.

Proposition 22. *Consider $\boldsymbol{m} = \boldsymbol{M}_0 + u : \mathcal{X}_\mathbb{R} \to \boldsymbol{M}_0 + \mathbb{U}_\mathbb{R}$, the parameter-to-solution map of the parametric LLG problem defined in the beginning of this section, where $\mathcal{X}_\mathbb{R}$ and $\mathbb{U}_\mathbb{R}$ are defined in* (53), (56) *respectively. Fix $\varepsilon > 0$ as in Theorem 5, let $\delta > 0$, $1 < q < \frac{1}{1-\delta/2}$. Fix a multi-index $\boldsymbol{\nu} = (\nu_i)_{i=1}^n \in \mathbb{N}_0^n$ for $n \in \mathbb{N}$. Define the positive sequence $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ as in* (61) *and choose $0 < \gamma < 1$ such that Assumption 4 holds. Then, it holds that*

$$\|\partial^{\boldsymbol{\nu}} \boldsymbol{m}(\boldsymbol{y})\|_\mathbb{U} \le \prod_{j=1}^n \nu_j! \rho_j^{-\nu_j} C_\varepsilon \qquad \text{for all } \boldsymbol{y} \in \mathcal{X}_\mathbb{R}, \tag{64}$$

where $C_\varepsilon > 0$ from Theorem 5 is independent of $\boldsymbol{\nu}$ or $\boldsymbol{y}$.

7 Sparse grid approximation of the parameter-to-solution map

We briefly recall the sparse grid construction. A more complete discussion can be found e.g. in [15] or [48]. Consider the family of distinct nodes $\mathcal{Y}^m = (y_i^m)_{i=1}^m \subset \mathbb{R}$ for any $m \in \mathbb{N}$ such that $y_1^1 = 0$ (the reason for this requirement will be clarified below). We shall write y_i rather than y_i^m when the context makes it unambiguous. Let $I_m : C^0(\mathbb{R}) \to P_m$ denote an *interpolation operator over* $\mathcal{Y}^m$ into a suitable m dimensional space P_m, i.e., $I_m[u](y) = u(y)$ for all $y \in \mathcal{Y}^m$ and any $u \in C^0(\mathbb{R})$.

Consider a *level-to-knot function* $m : \mathbb{N}_0 \to \mathbb{N}$, i.e. a strictly increasing function such that $m(0) = 1$. For any $i \in \mathbb{N}_0$, the *detail operator* is by definition

$$\begin{aligned} &\Delta_i : C^0(\mathbb{R}) \to P_{m(i)} \\ &\Delta_i u = I_{m(i)} u - I_{m(i-1)} u \qquad \text{for all } u \in C^0(\mathbb{R}), \end{aligned}$$

where $I_{m(-1)} \equiv 0$ so that $\Delta_0 u = I_1 u \equiv u(0)$, the constant interpolant in the origin.

In order to discuss interpolation schemes in more than one dimension, denote by $\mathcal{F}$ the set of integer-valued sequences (also called *multi-indices*) with finite support, i.e. $\boldsymbol{\nu} \in \mathcal{F}$ if and only if $\operatorname{supp}(\boldsymbol{\nu}) \coloneqq \{i \in \mathbb{N} : \nu_i \neq 0\}$ is finite.

Given $\boldsymbol{\nu} \in \mathcal{F}$, the corresponding *hierarchical surplus* operator is $\Delta_{\boldsymbol{\nu}} \coloneqq \bigotimes_{n\in\mathbb{N}} \Delta_{\nu_n}$, where the tensor product is finite as $\boldsymbol{\nu} \in \mathcal{F}$ and $\Delta_0 u \equiv u(0)$. In particular, $\Delta_{\boldsymbol{\nu}} = \bigotimes_{n\in\operatorname{supp}(\boldsymbol{\nu})} \Delta_{\nu_n}$. The hierarchical surplus operator can be applied to any function $u \in C^0(\mathbb{R}^\mathbb{N})$, i.e. a continuous function defined on the space of real valued sequences, by considering only the components of the independent variable in $\operatorname{supp}(\boldsymbol{\nu})$ and fixing the remaining ones to 0. Clearly, $\Delta_{\boldsymbol{\nu}} u \in P_{\boldsymbol{\nu}} \coloneqq \bigotimes_{n\in\operatorname{supp}(\boldsymbol{\nu})} P_{m(\nu_n)}$. Consider now $\Lambda \subset \mathcal{F}$ *downward-closed*, i.e.

$$\boldsymbol{\nu} \in \Lambda \;\Rightarrow\; \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda \quad \text{for all } n \in \operatorname{supp}(\boldsymbol{\nu}),$$

where $\boldsymbol{e}_n$ is the n-th coordinate unit vector. The *sparse grid interpolant* is by definition

$$\mathcal{I}_\Lambda \coloneqq \sum_{\boldsymbol{\nu}\in\Lambda} \Delta_{\boldsymbol{\nu}} : C^0(\mathbb{R}^\mathbb{N}) \to P_\Lambda, \tag{65}$$

where $P_\Lambda \coloneqq \bigoplus_{\boldsymbol{\nu}\in\Lambda} P_{\boldsymbol{\nu}}$. It can be proved that the fact that Λ is downward-closed implies that there exists a finite set $\mathcal{H}_\Lambda \subset \mathbb{R}^\mathbb{N}$, the *sparse grid*, such that $\mathcal{I}_\Lambda u(\boldsymbol{y}) = u(\boldsymbol{y})$ for any $u \in C^0(\mathbb{R}^\mathbb{N})$ and $\boldsymbol{y} \in \mathcal{H}_\Lambda$ and $\mathcal{I}_\Lambda u$ is the unique element of P_Λ with this property. Equivalently, there exists a *Lagrange basis* $(L_{\boldsymbol{y}})_{\boldsymbol{y}\in\mathcal{H}_\Lambda}$ of P_Λ, i.e. $L_{\boldsymbol{y}}(\boldsymbol{z}) = \delta_{\boldsymbol{y},\boldsymbol{z}}$ for any $\boldsymbol{y}, \boldsymbol{z} \in \mathcal{H}_\Lambda$. As a consequence, $\mathcal{I}_\Lambda u(\boldsymbol{z}) = \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} u(\boldsymbol{y}) L_{\boldsymbol{y}}(\boldsymbol{z})$ for any $u \in C^0(\mathbb{R}^\mathbb{N})$ and all $\boldsymbol{z} \in \mathbb{R}^\mathbb{N}$.

An important question is the one of (quasi) optimal approximation: Assume that $u \in C^0(\mathbb{R}^\mathbb{N})$ belongs to a given function class and given a *computational budget* $Q \in \mathbb{N}$, we look for $\Lambda \subset \mathcal{F}$ downward-closed with $\#\mathcal{H}_\Lambda \le Q$ such that

$$\|u - \mathcal{I}_\Lambda u\|_{L^2_\mu(\mathbb{R}^\mathbb{N})} \lesssim \min\left\{\|u - \mathcal{I}_{\tilde\Lambda} u\|_{L^2_\mu(\mathbb{R}^\mathbb{N})} : \tilde\Lambda \subset \mathcal{F} \text{ downward closed such that } \#\mathcal{H}_{\tilde\Lambda} \le Q\right\}, \tag{66}$$

where the hidden constant is independent of Λ. Following [46], we reformulate the problem of selecting Λ as a (linear programming relaxation of a) *knapsack problem.* For a generic multi-index $\boldsymbol{\nu} \in \mathcal{F}$, consider a *value* $v_{\boldsymbol{\nu}} \geq 0$ and *work* $w_{\boldsymbol{\nu}} > 0$ that satisfy

$$\|\Delta^{\boldsymbol{\nu}} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}})} \lesssim v_{\boldsymbol{\nu}} \qquad \text{for all } \boldsymbol{\nu} \in \mathcal{F},$$
$$\#\mathcal{H}_\Lambda \lesssim \sum_{\boldsymbol{\nu}\in\Lambda} w_{\boldsymbol{\nu}} \qquad \text{for all } \Lambda \subset \mathcal{F} \text{ downward-closed.}$$

The optimal multi-index set selection problem (66) is substituted by the following *knapsack problem*:

$$\max\left\{\sum_{\boldsymbol{\nu}\in\Lambda} v_{\boldsymbol{\nu}} \ : \ \Lambda \subset \mathcal{F} \text{ downward-closed and } \sum_{\boldsymbol{\nu}\in\Lambda} w_{\boldsymbol{\nu}} \leq Q\right\}.$$

The solution to (the linear relaxation of) this problem can be found as follows: Define the *profit*

$$\mathcal{P}_{\boldsymbol{\nu}} := \frac{v_{\boldsymbol{\nu}}}{w_{\boldsymbol{\nu}}} \qquad \text{for all } \boldsymbol{\nu} \in \mathcal{F}.$$

Induce a partial ordering on $\mathcal{F}$ as:

$$\boldsymbol{\nu}_1 \preceq \boldsymbol{\nu}_2 \Leftrightarrow \mathcal{P}_{\boldsymbol{\nu}_1} \geq \mathcal{P}_{\boldsymbol{\nu}_2} \qquad \text{for all } \boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \in \mathcal{F}$$

and sort its elements accordingly as $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots$ In case $\mathcal{P}_{\boldsymbol{\nu}_1} = \mathcal{P}_{\boldsymbol{\nu}_2}$ sort the two multi-indices in lexicographic order. Define for any $n \in \mathbb{N}$ the *n-elements quasi-optimal multi-index set*

$$\Lambda_n := \{\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n\} \subset \mathcal{F}. \tag{67}$$

We assume that value and work are monotone, i.e. $v_{\boldsymbol{\nu}+\boldsymbol{e}_n} \leq v_{\boldsymbol{\nu}}$, $w_{\boldsymbol{\nu}+\boldsymbol{e}_n} \geq w_{\boldsymbol{\nu}}$ for all $\boldsymbol{\nu} \in \mathcal{F}$, $n \in \mathbb{N}$. This implies that Λ_n is downward-closed for any $n \in \mathbb{N}$.

The approximation error of the corresponding sparse grid interpolation is estimated as follows.

Theorem 23. *[46, Theorem 1] If there exists $\tau \in (0, 1]$ such that*

$$C_\tau := \left(\sum_{\boldsymbol{\nu}\in\mathcal{F}} \mathcal{P}_{\boldsymbol{\nu}}^\tau w_{\boldsymbol{\nu}}\right)^{1/\tau} < \infty,$$

then

$$\|u - \mathcal{I}_{\Lambda_n} u\|_{L^2_\mu(\mathbb{R}^{\mathbb{N}})} \leq C_\tau \#\mathcal{H}_{\Lambda_n}^{1-1/\tau}.$$

In the next two sections we discuss the sparse grid methods defined using piecewise polynomial interpolation.

7.1 1D piecewise polynomial interpolation on $\mathbb{R}$

Let $\mu(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$ denote the normal density with mean zero and variance $\sigma^2 > 0$. Let $\mu(x) = \mu(x; 1)$ and $\tilde{\mu}(x) = \mu(x; \sigma^2)$ for some fixed $\sigma^2 > 1$. Consider the error function $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \mathrm{d}t$.

For $m \in \mathbb{N}$ odd, define $\mathcal{Y}^m = \{y_1, \ldots, y_m\} \subset \mathbb{R}$ by

$$y_i = \phi\left(-1 + \frac{i}{m+1}\right) \qquad i = 1, \ldots, m, \tag{68}$$

where

$$\phi(x) := \alpha \operatorname{erf}^{-1}(x) \qquad \text{for all } x \in (-1, 1), \tag{69}$$
$$\alpha = \alpha(p, \sigma^2) := \sqrt{\frac{4p}{1 - \frac{1}{\sigma^2}}}. \tag{70}$$

The m nodes define $m + 1$ intervals (the first and last are unbounded). See Figure 2.

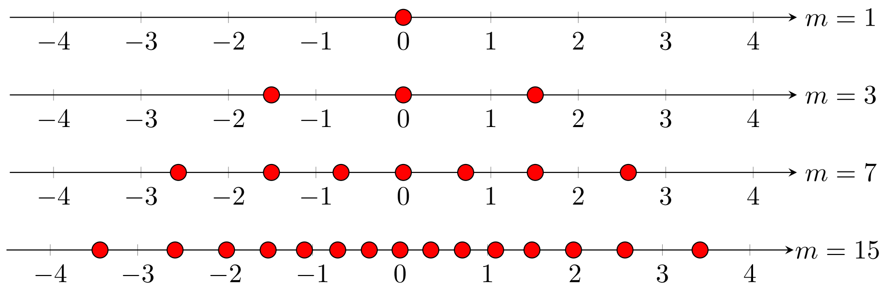


Figure 2: Examples of nodes (68) for $p = 2$ on $\mathbb{R}$. It can be seen that the nodes span a wider and wider portion of the real line and, at the same time, become denser. If the number of nodes is suitably increased (for example using (72)), the nodes family is nested.

We define a 1D piecewise polynomial interpolation operator as follows. When $m = 1$, let for any $u \in C^0(\mathbb{R})$ and any $p \geq 2$

$$I_1^p[u] = I_1[u] \equiv u(0),$$

i.e. the constant interpolation. When $m \geq 3$, $I_m^p[u]$ is the piecewise polynomial function of degree $p-1$ over the intervals defined by $\mathcal{Y}^m$. More precisely, for any $u \in C^0(\mathbb{R})$,

$$\begin{aligned}
&I_m^p[u](y_i) = u(y_i) && \text{for all } i = 1, \ldots, m,\\
&I_m^p[u]|_{[y_i, y_{i+1}]} \in \mathbb{P}_{p-1} && \text{for all } i = 1, \ldots, m-1,\\
&I_m^p[u](y) \text{ polynomial extension of } I_m^p[u]|_{[y_1, y_2]} && \text{if } y \leq y_1,\\
&I_m^p[u](y) \text{ polynomial extension of } I_m^p[u]|_{[y_{m-1}, y_m]} && \text{if } y \geq y_m.
\end{aligned}$$

We assume that for each $i = 1, \ldots, m-1$, the interval (y_i, y_{i+1}) contains additional $p-2$ distinct interpolation nodes so that $I_m^p[u]$ is uniquely defined.

The function ϕ is such that $(\phi'(x))^{2p}\, \tilde{\mu}^{-1}(\phi(x))\mu(\phi(x))$ is constant and equals

$$C_\phi = \sqrt{\sigma^2}\left(\frac{\alpha\sqrt{\pi}}{2}\right)^{2p}, \tag{71}$$

where α was defined in (70).

The following result is a standard interpolation error estimate on weighted spaces which, in this precise form, we could not find in the literature.

Lemma 24. *Consider $u : \mathbb{R} \to \mathbb{R}$ with $\partial u \in L^2_{\tilde{\mu}}(\mathbb{R})$. Then,*

$$\|u - I_1[u]\|_{L^2_\mu(\mathbb{R})} \leq \tilde{C}_1 \|\partial u\|_{L^2_{\tilde{\mu}}(\mathbb{R})},$$

where $\tilde{C}_1 = \sqrt{\int_{\mathbb{R}} |y|\, \tilde{\mu}^{-1}(y) \mathrm{d}\mu(y)}$. If additionally, $\partial^p u \in L^2_{\tilde{\mu}}(\mathbb{R})$ for $p \geq 2$, then

$$\|u - I_m^p[u]\|_{L^2_\mu(\mathbb{R})} \leq \tilde{C}_2 (m+1)^{-p} \frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}(\mathbb{R})}}{p!} \qquad \textit{for all } m \geq 3 \textit{ odd},$$

where $\tilde{C}_2 = \sqrt{C_\phi \frac{p}{2}\left(m - 1 + 2^{2p+1}\right)}$ and C_ϕ was defined in (71).

Proof. For the first estimate, the fundamental theorem of calculus and Cauchy-Schwarz inequality yield $u(y) - u(0) = \int_0^y \partial u \leq \|\partial u\|_{L^2_{\tilde{\mu}}(\mathbb{R})} \sqrt{\int_0^y \tilde{\mu}^{-1}}$. Substitute this in $\|u - u(0)\|_{L^2_\mu(\mathbb{R})}$ to obtain the first estimate.

For the second estimate, let $i \in \{2, \ldots, m-2\}$. Apply the fundamental theorem of calculus p times and recall that $I_m^p[u] \in \mathbb{P}_{p-1}([y_i, y_{i+1}])$ to obtain: $\left(u - I_m^p[u]\right)(y) = \int_{y_i}^y \int_{\xi_1}^{z_1} \cdots \int_{\xi_{p-1}}^{z_{p-1}} \partial^p u$, for all $y \in [y_i, y_{i+1}]$, where $\xi_i \in [y_i, y_{i+1}]$ is such that $\partial^i(u - I_m^p[u])(\xi_i) = 0$. The Cauchy-Schwarz inequality applied to the last integral gives $\int_{\xi_{p-1}}^{z_{p-1}} \partial^p u \leq \|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])} \sqrt{\int_{y_i}^{y_{i+1}} \tilde{\mu}^{-1}}$. We can then estimate with elementary facts

$$\begin{aligned}(u - I^p_m[u])(y) &\leq \|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])} \int_{y_i}^{y} \int_{y_i}^{z_1} \cdots \int_{y_i}^{z_{p-2}} \sqrt{\int_{y_i}^{y_{i+1}} \tilde{\mu}^{-1}} \\ &\leq \|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])} \tilde{\mu}^{-1/2}(y) \int_{y_i}^{z_1} \cdots \int_{y_i}^{z_{p-2}} \sqrt{|z_{p-1} - y_i|} \\ &\leq \|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])} \tilde{\mu}^{-1/2}(y) \frac{|y-y_i|^{p-1+\frac{1}{2}}}{(p-1)!}.\end{aligned}$$

Consider now only $i \in \left\{\frac{m+1}{2}, \dots, m-2\right\}$. Then,

$$\int_{y_i}^{y_{i+1}} |u - I^p_m[u]|^2 (y) \mathrm{d}\mu(y) \leq \frac{\|\partial^p u\|^2_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])}}{(p-1)!} \int_{y_i}^{y_{i+1}} |y - y_i|^{2p-1} \tilde{\mu}^{-1}(y) \mathrm{d}\mu(y).$$

In order to estimate the last integral, change variables using ϕ defined in (69). We get

$$\int_{y_i}^{y_{i+1}} |y - y_i|^{2p-1} \tilde{\mu}^{-1}(y) \mathrm{d}\mu(y) \leq \int_{x_i}^{x_{i+1}} |\phi(x) - \phi(x_i)|^{2p-1} \tilde{\mu}^{-1}(\phi(x)) \mu(\phi(x)) \phi'(x) \mathrm{d}x.$$

A Taylor expansion and the fact that ϕ' is increasing give $\phi(x) - \phi(x_i) \leq \phi'(x)(x - x_i)$. So we get

$$\int_{y_i}^{y_{i+1}} |y - y_i|^{2p-1} \tilde{\mu}^{-1}(y) \mathrm{d}\mu(y) \leq \int_{x_i}^{x_{i+1}} (x - x_i)^{2p-1} (\phi'(x))^{2p} \tilde{\mu}^{-1}(\phi(x)) \mu(\phi(x)) \mathrm{d}x.$$

Recall now that $(\phi'(x))^{2p} \tilde{\mu}^{-1}(\phi(x))\mu(\phi(x)) \equiv C_\phi$. Integration yields

$$\int_{y_i}^{y_{i+1}} |y - y_i|^{2p-1} \tilde{\mu}^{-1}(y) \mathrm{d}\mu(y) \leq \frac{(m+1)^{-2p}}{2p} C_\phi.$$

For the original quantity, we get

$$\int_{y_i}^{y_{i+1}} |u - I^p_m[u]|^2 (y) \mathrm{d}\mu(y) \leq C_\phi \frac{p^2}{2p} (m+1)^{-2p} \left(\frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])}}{p!} \right)^2.$$

For $i = m-1, m$, recall that I^p_m is defined in $[y_m, +\infty)$ as the polynomial extension from the previous interval. Analogous estimates give

$$\int_{y_{m-1}}^{+\infty} |u - I^p_m[u]|^2 (y) \mathrm{d}\mu(y) \leq C_\phi \frac{p^2}{2p} 2^{2p} (m+1)^{-2p} \left(\frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])}}{p!} \right)^2.$$

Analogous estimates for $i \leq \frac{m+1}{2}$ give the second estimates. □

We define the following *level-to-knots function*:

$$m(\nu) := 2^{\nu+1} - 1 \qquad \text{for all } \nu \in \mathbb{N}_0 \tag{72}$$

and observe that $\left(\mathcal{Y}^{m(i)}\right)_{i \in \mathbb{N}_0}$ are nested, i.e. $\mathcal{Y}^{m(i)} \subset \mathcal{Y}^{m(i+1)}$ for all $i \in \mathbb{N}_0$. The level-to-knot function is used to define detail operators Δ_ν and hierarchical surpluses as explained in the beginning of the section. We now apply the previous results to estimate 1D detail operators.

Lemma 25. *Consider $u : \mathbb{R} \to \mathbb{R}$, a continuous function with $\partial u \in L^2_{\tilde{\mu}}(\mathbb{R})$ and $p \geq 2$. There holds*

$$\|\Delta_1[u]\|_{L^2_\mu(\mathbb{R})} \leq C_1 \|\partial u\|_{L^2_{\tilde{\mu}}(\mathbb{R})},$$

where $C_1 = 2^{3/2} \tilde{C}_1 \sqrt{\int_0^\infty \sum_{j=1}^p |l_j'|^2 \mathrm{d}\tilde{\mu}} \sqrt[4]{\int_0^{y_3} \tilde{\mu}^{-1}}$, $\tilde{C}_1 > 0$ was defined in the previous lemma, y_1, y_2, y_3 delimit the intervals of definition of the piecewise polynomial $I^p_3[u]$ and $(l_j)_{j=1}^p$ denote the Lagrange basis of $\mathbb{P}_{p-1}([y_2, y_3])$ with respect to y_2, y_3 and other $p-2$ district points in (y_2, y_3).
If additionally $\partial^p u \in L^2_{\tilde{\mu}}(\mathbb{R})$, then we have

$$\|\Delta_\nu[u]\|_{L^2_\mu(\mathbb{R})} \leq C_2 2^{-p\nu} \frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}(\mathbb{R})}}{p!} \qquad \text{for all } \nu \geq 1,$$

where $C_2 = \tilde{C}_2(1 + 2^{-p})$ and $\tilde{C}_2 > 0$ was defined in the previous lemma.

Proof. To prove the first estimate, recall that nodes are nested so $I_1[u] = I_1\left[I_3^p[u]\right]$. This implies that

$$\Delta_1[u] = I_3^p[u] - I_1[u] = I_3^p[u] - I_1\left[I_3^p[u]\right] = (1 - I_1)\left[I_3^p[u]\right].$$

The previous lemma gives

$$\|\Delta_1[u]\|_{L^2\mu(\mathbb{R})} \leq \tilde{C}_1 \left\|\partial I_3^p[u]\right\|_{L^2_{\tilde{\mu}}(\mathbb{R})}.$$

To estimate the last integral, consider $x_1 = y_2 < x_2 < \cdots < x_p = y_3$ the interpolation nodes in the interval $[y_2, y_3]$. Observe that $\partial I^p[u] = \partial I^p[u - u(0)]$ and estimate

$$\begin{aligned}\int_0^\infty |\partial I^p[u]|^2 \,\mathrm{d}\tilde{\mu} &= \int_0^\infty |\partial I^p[u - u(0)]|^2 \,\mathrm{d}\tilde{\mu} = \int_0^\infty \left|\sum_{j=1}^p (u(x_j) - u(0)) l_j'\right|^2 \mathrm{d}\tilde{\mu} \\ &\leq 2 \max_{j=1,\dots,n} |u(x_j) - u(0)|^2 \int_0^\infty \sum_{j=1}^p \left|l_j'\right|^2 \mathrm{d}\tilde{\mu}\end{aligned}$$

Since the second term is bounded for fixed p, let us focus on the first one. Simple computations give

$$\max_{j=1,\dots,n} |u(x_j) - u(0)| \leq \int_0^{y_3} |\partial u| \leq \|\partial u\|_{L^2_{\tilde{\mu}}([0,y_3])} \sqrt{\int_0^{y_3} \tilde{\mu}^{-1}}.$$

This, together with analogous computations on $(-\infty, 0]$, gives the first estimate.

To prove the second estimate, observe that

$$\|\Delta_\nu[u]\|_{L^2_{\mu}(\mathbb{R})} = \left\|I^p_{m(\nu)}[u] - I^p_{m(\nu-1)}[u]\right\|_{L^2_{\mu}(\mathbb{R})} \leq \left\|u - I^p_{m(\nu)}[u]\right\|_{L^2_{\mu}(\mathbb{R})} + \left\|u - I^p_{m(\nu-1)}[u]\right\|_{L^2_{\mu}(\mathbb{R})}$$

The previous lemma and simple computations imply the second estimate. □

We can finally estimate hierarchical surpluses as follows.

Proposition 26. *Let $u : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, $p \geq 2$ and $\boldsymbol{\nu} \in \mathcal{F}$. Then*

$$\|\Delta_{\boldsymbol{\nu}}[u]\|_{L^2_{\mu}(\mathbb{R}^{\mathbb{N}})} \leq \left(\prod_{i:\nu_i=1} C_1\right) \prod_{i:\nu_i>1} \left(\frac{C_2 2^{-p\nu_i}}{p!}\right) \left\|\partial_{\{i:\nu_i=1\}} \partial^p_{\{i:\nu_i>1\}} u\right\|_{L^2_{\tilde{\mu}}(\mathbb{R}^{\mathbb{N}})},$$

where u is understood to be sufficiently regular for the right-hand-side to be well defined and $C_1, C_2 > 0$ are constants defined in the previous lemma.

Proof. Assume without loss of generality that all components of $\boldsymbol{\nu}$ are non-zero except the first $N \in \mathbb{N}$. Then,

$$\|\Delta_{\boldsymbol{\nu}}[u]\|^2_{L^2_{\mu}(\mathbb{R}^{\mathbb{N}})} = \int_{\mathbb{R}^N} |\Delta_{\boldsymbol{\nu}}[u]|^2 \,\mathrm{d}\mu = \int_{\mathbb{R}^{N-1}} \int_{\mathbb{R}} |\Delta_1[y_1 \mapsto \Delta_{\hat{\boldsymbol{\nu}}_1} u]|^2 \,\mathrm{d}\mu_1 \mathrm{d}\hat{\mu}_1,$$

where we denoted $\hat{\boldsymbol{\nu}}_1 = (\nu_2, \dots, \nu_N)$ and $\hat{\mu}_1$ the $N-1$-dimensional Gaussian measure. We apply the previous estimate (assume that $\nu_1 = 1$, the other case is analogous) to get

$$\|\Delta_{\boldsymbol{\nu}}[u]\|^2_{L^2_{\mu}(\mathbb{R}^{\mathbb{N}})} \leq \int_{\mathbb{R}^{N-1}} C_1^2 \int_{\mathbb{R}} |\partial_1 \Delta_{\hat{\boldsymbol{\nu}}_1}[u]|^2 \,\mathrm{d}\tilde{\mu}_1 \mathrm{d}\hat{\mu}_1.$$

We now exchange the integrals as well as the operators acting on u to get

$$\|\Delta_{\boldsymbol{\nu}}[u]\|^2_{L^2_{\mu}(\mathbb{R}^{\mathbb{N}})} \leq C_1^2 \int_{\mathbb{R}} \int_{\mathbb{R}^{N-1}} |\Delta_{\hat{\boldsymbol{\nu}}_1}[\partial_1 u]|^2 \,\mathrm{d}\hat{\mu}_1 \mathrm{d}\tilde{\mu}_1.$$

We can iterate this procedure $N-1$ additional times to obtain the statement. □

7.2 Basic profits and dimension dependent convergence

In the present section, we discuss the convergence of sparse grid approximation when the sample paths of Wiener processes and magnetizations are assumed to be Hölder-continuous. To this end, we apply the results found in Section 5.

Let us apply Proposition 6 to estimate the derivatives appearing in the estimate we found in Proposition 26. We find

$$\|\Delta_{\boldsymbol{\nu}}[u]\|_{L^2_\mu(\mathbb{R}^\mathbb{N})} \leq \prod_{i\in\operatorname{supp}(\boldsymbol{\nu})} \tilde{v}_{\nu_i}, \quad \text{where } \tilde{v}_{\nu_i} = \begin{cases} C_1\rho_i^{-1} & \text{if } \nu_i = 1 \\ C_2\left(2^{\nu_i}\rho_i\right)^{-p} & \text{if } \nu_i > 1 \end{cases}$$

and $\rho_i = \varepsilon 2^{\frac{(1-\alpha)\lceil\log_2(i)\rceil}{2}}$. Recall the framework presented at the beginning of Section 7. Given a multi-index $\boldsymbol{\nu} \in \mathcal{F}$, we define as its value and work respectively

$$\tilde{v}_{\boldsymbol{\nu}} = \prod_{i\in\operatorname{supp}(\boldsymbol{\nu})} \tilde{v}_{\nu_i}, \tag{73}$$

$$w_{\boldsymbol{\nu}} = \prod_{i\in\operatorname{supp}(\boldsymbol{\nu})} p2^{\nu_i}. \tag{74}$$

The definition of work is justified as follows: From the definition of 1D nodes (68) and level-to-knots function (72), each time a multi-index is added to the multi-index set, the sparse grid gains $(2^{\nu_i+1}-2)(p-1)+1$ new nodes in the i-th coordinate.

Recall that the profit is the ratio of value and work. In this case, it reads

$$\tilde{\mathcal{P}}_{\boldsymbol{\nu}} = \frac{\tilde{v}_{\boldsymbol{\nu}}}{w_{\boldsymbol{\nu}}}. \tag{75}$$

We apply the convergence Theorem 23 to obtain a convergence rate that depends root-exponentially on the number of approximated parameters. We skip the computations because they are a simplified version of the ones presented in the next section.

Theorem 27. *Let $N \in \mathbb{N}$ and denote by $\boldsymbol{m}_N : \mathbb{R}^N \to C^{1+\alpha/2,2+\alpha}(D_T)$ the parameter-to-solution map of the parametric LLG problem under assumption that $W(\boldsymbol{y},t) = \sum_{i\in\mathbb{N}} y_i\eta_i(t) = \sum_{i=1}^N y_i\eta_i(t)$ for all $t \in [0,T]$ and all $\boldsymbol{y} \in \mathbb{R}^\mathbb{N}$. Let $\Lambda_N \subset \mathbb{N}^N$ denote the multi-index set (67) defined using $\tilde{\mathcal{P}}_{\boldsymbol{\nu}}$ as in (75). Let $\mathcal{I}_{\Lambda_N}$ denote the corresponding piecewise polynomial sparse grids interpolant of degree $p-1$ with nodes (68) and $p \geq 2$. Denote $\mathcal{H}_{\Lambda_n} \subset \mathbb{R}^N$ the corresponding sparse grid. Under the assumptions of Theorem 8, for any $\frac{2}{(1+\alpha)p} < \tau < 1$,*

$$\|\boldsymbol{m}_N - \mathcal{I}_{\Lambda_N}\boldsymbol{m}_N\|_{L^2_\mu(\mathbb{R}^N, C^{1+\alpha/2,2+\alpha}(D_T))} \leq C_{\tau,p}(N)\left(\#\mathcal{H}_{\Lambda_n}\right)^{1-1/\tau}, \tag{76}$$

where $C_{\tau,p}(N)$ is a function of τ, p, N. In particular,

$$C_{\tau,p}(N) = (1+P_0)^{1/\tau} \exp\frac{1}{\tau}\left(\frac{C_1^\tau\,(2p)^{1-\tau}}{2}\frac{1-N^{(1-(1-\alpha)\tau/2)}}{1-2^{1-(1-\alpha)\tau/2}} + \frac{C_2^\tau\sigma(p,\tau)}{2}\frac{1}{1-2^{1-(1-\alpha)p\tau/2}}\right),$$

where $P_0 = C_1^\tau(2p)^{1-\tau} + C_2^\tau p^{1-\tau}\sigma(p,\tau)$, $\sigma(p,\tau) = \frac{2^{2(1-\tau(p+1))}}{1-2^{1-\tau(p+1)}}$ and C_1, C_2 were defined in Lemma 25. In particular, the bound grows root-exponentially in the number of dimensions.

7.3 Improved profits and dimension independent convergence

In the previous section, we could prove only a dimension-*dependent* convergence. This may be attributed to the slow growth of the holomorphy radii $\rho_i \lesssim 2^{\frac{(1-\alpha)\ell(i)}{2}}$. Let us consider the setting from Section 6, in which we assumed small Wiener processes and a coefficient $\boldsymbol{g}$ with small gradient. With these modelling assumptions, we proved that the holomorphy radii can be chosen as (61). This will be sufficient to obtain dimension-*independent* convergence.

Again we work within the framework described at the beginning of Section 7.

We need to define *values* that, for any $\boldsymbol{\nu} \in \mathcal{F}$, bound $\|\Delta_{\boldsymbol{\nu}}u\|_{L^2_\mu(\mathbb{R}^\mathbb{N},\mathbb{U})}$ from above. The estimates from Proposition 26 and the estimate on the derivatives from Proposition 22 motivate the following choice of values:

$$v_{\boldsymbol{\nu}} = \prod_{i\in\operatorname{supp}(\boldsymbol{\nu})} v_{\nu_i}, \qquad \text{where } v_{\nu_i} = \begin{cases} C_1\rho_i^{-1} & \text{if } \nu_i = 1 \\ C_2\left(2^{\nu_i}\rho_i\right)^{-p} & \text{if } \nu_i > 1 \end{cases}$$

and

$$\rho_i = \rho_{\ell,j} := \gamma\begin{cases} 2^{\left(\frac{3}{2}-\delta\right)\ell}\frac{1}{r_\ell(\boldsymbol{\nu})} & \text{if } \nu_{\ell,j} = 1 \\ 2^{\left(\frac{1}{2}-\delta\right)\ell} & \text{otherwise} \end{cases}.$$

Here, i and (ℓ, j) are related through the hierarchical indexing (5), $\delta > 0$ is small and for any $\ell \in \mathbb{N}_0$, $\boldsymbol{\nu} \in \mathcal{F}$, $r_\ell(\boldsymbol{\nu}) = \#\left\{j \in \left\{1, \dots, \lceil 2^{\ell-1} \rceil\right\} : \nu_{\ell,j} = 1\right\}$. With the *work* defined as in (74), the profits now read

$$\mathcal{P}_{\boldsymbol{\nu}} = \frac{v_{\boldsymbol{\nu}}}{w_{\boldsymbol{\nu}}}. \tag{77}$$

Let us determine for which $\tau \in (0,1)$ the sum $\sum_{\boldsymbol{\nu} \in \mathcal{F}} v_{\boldsymbol{\nu}}^\tau w_{\boldsymbol{\nu}}^{1-\tau}$ is finite. This setting is more complex than the one in the previous section because the factors v_{ν_i} that define the values $v_{\boldsymbol{\nu}}$ depend in general on $\boldsymbol{\nu}$ rather than ν_i alone. Define

$$\mathcal{F}^* \coloneqq \{\boldsymbol{\nu} \in \mathcal{F} : \nu_i \neq 1 \ \text{ for all } i \in \mathbb{N}\}$$

and for any $\boldsymbol{\nu} \in \mathcal{F}^*$

$$K_{\boldsymbol{\nu}} \coloneqq \{\hat{\boldsymbol{\nu}} \in \mathcal{F} \, : \, \hat{\nu}_i = \nu_i \text{ if } \nu_i > 1 \text{ and } \hat{\nu}_i \in \{0,1\} \ \text{ if } \nu_i = 0\}\,.$$

The family $\{K_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathcal{F}^*}$ is a partition of $\mathcal{F}$. As a consequence,

$$\begin{aligned}
\sum_{\boldsymbol{\nu} \in \mathcal{F}} v_{\boldsymbol{\nu}}^\tau w_{\boldsymbol{\nu}}^{1-\tau} &= \sum_{\boldsymbol{\nu} \in \mathcal{F}^*} \sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} v_{\hat{\boldsymbol{\nu}}}^\tau w_{\hat{\boldsymbol{\nu}}}^{1-\tau} \\
&= \sum_{\boldsymbol{\nu} \in \mathcal{F}^*} \sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} \prod_{i : \hat{\nu}_i \le 1} \left(v_{\hat{\nu}_i}^\tau w_{\hat{\nu}_i}^{1-\tau}\right) \prod_{i : \hat{\nu}_i > 1} \left(v_{\hat{\nu}_i}^\tau w_{\hat{\nu}_i}^{1-\tau}\right) \\
&= \sum_{\boldsymbol{\nu} \in \mathcal{F}^*} \prod_{i : \nu_i > 1} \left(v_{\nu_i}^\tau w_{\nu_i}^{1-\tau}\right) \sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} \prod_{i : \hat{\nu}_i \le 1} \left(v_{\hat{\nu}_i}^\tau w_{\hat{\nu}_i}^{1-\tau}\right).
\end{aligned} \tag{78}$$

Consider the following subset of $\mathcal{F}$:

$$\mathcal{F}\{0,1\} \coloneqq K_{\mathbf{0}} = \{\boldsymbol{\nu} \in \mathcal{F} \, : \, \nu_i \in \{0,1\} \ \text{ for all } i \in \mathbb{N}\}\,.$$

Lemma 28. *Let $0 < p < 1$, $p < q < \infty$, and the sequence $\boldsymbol{a} = (a_j)_{j \in \mathbb{N}} \in \ell^p(\mathbb{N})$. Then,*

$$\left(|\boldsymbol{\nu}|_1! \, \boldsymbol{a}^{\boldsymbol{\nu}}\right)_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \in \ell^q(\mathcal{F}\{0,1\}).$$

Proof. Choose $\varepsilon > 0$ such that $\frac{1}{1+\varepsilon} \ge p$ and $q > p(1+\varepsilon)$. We consider $\alpha > |\boldsymbol{a}|_{1/(1+\varepsilon)}$ and write

$$\sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(|\boldsymbol{\nu}|_1! \, \boldsymbol{a}^{\boldsymbol{\nu}}\right)^q = \sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(|\boldsymbol{\nu}|_1! \, \alpha^{|\boldsymbol{\nu}|_1} \left(\frac{\boldsymbol{a}}{\alpha}\right)^{\boldsymbol{\nu}}\right)^q.$$

There exists $C_\varepsilon > 0$ such that $\alpha^{|\boldsymbol{\nu}|_1} \le C_\varepsilon \left(|\boldsymbol{\nu}|_1!\right)^\varepsilon$ for all $\boldsymbol{\nu} \in \mathcal{F}\{0,1\}$. Thus,

$$\sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(|\boldsymbol{\nu}|_1! \, \boldsymbol{a}^{\boldsymbol{\nu}}\right)^q \lesssim \sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(\left(|\boldsymbol{\nu}|_1!\right)^{1+\varepsilon} \left(\frac{\boldsymbol{a}}{\alpha}\right)^{\boldsymbol{\nu}}\right)^q.$$

Factorizing out the $1+\varepsilon$ yields

$$\sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(|\boldsymbol{\nu}|_1! \, \boldsymbol{a}^{\boldsymbol{\nu}}\right)^q \lesssim \sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(|\boldsymbol{\nu}|_1! \left(\frac{\boldsymbol{a}}{\alpha}\right)^{\frac{1}{1+\varepsilon}\boldsymbol{\nu}}\right)^{(1+\varepsilon)q}.$$

Since $\boldsymbol{\nu}! = 1$ for all $\boldsymbol{\nu} \in \mathcal{F}\{0,1\}$, we can write

$$\sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(|\boldsymbol{\nu}|_1! \, \boldsymbol{a}^{\boldsymbol{\nu}}\right)^q \lesssim \sum_{\boldsymbol{\nu} \in \mathcal{F}\{0,1\}} \left(\frac{|\boldsymbol{\nu}|_1!}{\boldsymbol{\nu}!} \left(\frac{\boldsymbol{a}}{\alpha}\right)^{\frac{1}{1+\varepsilon}\boldsymbol{\nu}}\right)^{(1+\varepsilon)q}. \tag{79}$$

Observe that $\sum_j \left(\frac{a_j}{\alpha}\right)^{\frac{1}{1+\varepsilon}} < 1$ because of the definition of α. Moreover, from the assumption on $\boldsymbol{a}$ we have $\left(\frac{\boldsymbol{a}}{\alpha}\right)^{\frac{1}{1+\varepsilon}} \in \ell^r(\mathbb{N})$ for any $r \ge p(1+\varepsilon)$. Then, [21, Theorem 1] implies that the second sum in (79) is finite, thus proving the statement. ∎

Lemma 29. *If $\tau > \frac{1}{\frac{3}{2}-\delta}$, there exists $C > 0$ such that for any $\boldsymbol{\nu} \in \mathcal{F}^*$,*

$$\sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} \prod_{i : \hat{\nu}_i \le 1} \left(v_{\hat{\nu}_i}^\tau w_{\hat{\nu}_i}^{1-\tau}\right) \le C.$$

Proof. For this proof, we denote the level of i by $\ell(i)$. First observe that, from the definitions of value and work, we may write

$$\prod_{i:\hat{\nu}_i\leq 1}\left(v_{\hat{\nu}_i}^{\tau}w_{\hat{\nu}_i}^{1-\tau}\right)=\prod_{i:\hat{\nu}_i=1}\left(C_1 2^{-\left(\frac{3}{2}-\delta\right)\ell(i)}r_{\ell(i)}(\boldsymbol{\nu})\right)^{\tau}(2p)^{1-\tau}.$$

The factors in the right-hand-side are independent of the components of $\boldsymbol{\nu}$ for which $\nu_i\neq 1$. Thus, we define

$$D_{\boldsymbol{\nu}}=\left\{\boldsymbol{d}\in\mathcal{F}:\begin{cases}d_i=0 & \text{if } \nu_i>1\\ d_i\in\{0,1\} & \text{otherwise}\end{cases}\right\}\subset\mathcal{F}\{0,1\}$$

and substitute

$$\sum_{\hat{\boldsymbol{\nu}}\in K_{\boldsymbol{\nu}}}\prod_{i:\hat{\nu}_i=1}\left(C_1 2^{-\left(\frac{3}{2}-\delta\right)\ell(i)}r_{\ell(i)}(\hat{\boldsymbol{\nu}})\right)^{\tau}(2p)^{1-\tau}=\sum_{\boldsymbol{d}\in D_{\boldsymbol{\nu}}}\prod_{i:d_i=1}\left(C_1 2^{-\left(\frac{3}{2}-\delta\right)\ell(i)}r_{\ell(i)}(\boldsymbol{d})\right)^{\tau}(2p)^{1-\tau}.$$

From the definition of $r_{\ell(i)}(\boldsymbol{d})$, we estimate $\prod_{i:d_i=1}r_{\ell(i)}(\boldsymbol{d})\leq\prod_{\ell:\exists j:d_{\ell,j}=1}r_\ell(\boldsymbol{d})^{r_\ell(\boldsymbol{d})}$. Stirling's formula gives $r_\ell(\boldsymbol{d})^{r_\ell(\boldsymbol{d})}\leq r_\ell(\boldsymbol{d})!e^{r_\ell(\boldsymbol{d})}$. Denote $\boldsymbol{d}_\ell=\left(d_{\ell,1},\ldots,d_{\ell,\lceil 2^{\ell-1}\rceil}\right)$ for any $\ell\in\mathbb{N}_0$ and observe that $r_\ell(\boldsymbol{d})\leq|\boldsymbol{d}_\ell|_1$. Together with an elementary property of the factorial, this gives $\prod_{\ell:\exists j:d_{\ell,j}=1}(r_\ell(\boldsymbol{d}))!\leq\prod_{\ell:\exists j:d_{\ell,j}=1}|\boldsymbol{d}_\ell|_1!\leq\left(\sum_{\ell\in\mathbb{N}}|\boldsymbol{d}_\ell|_1\right)!=|\boldsymbol{d}|_1!$. To summarize, we have estimated

$$\sum_{\hat{\boldsymbol{\nu}}\in K_{\boldsymbol{\nu}}}\prod_{i:\hat{\nu}_i\leq 1}\left(v_{\hat{\nu}_i}^{\tau}w_{\hat{\nu}_i}^{1-\tau}\right)\leq\sum_{\boldsymbol{d}\in D_{\boldsymbol{\nu}}}(|\boldsymbol{d}|_1!)^{\tau}\prod_{i:d_i=1}\left(C_1^{\tau}2^{-\left(\frac{3}{2}-\delta\right)\ell(i)\tau}(2p)^{1-\tau}e^{\tau}\right).$$

Define $c_j:=C_1 2^{-\left(\frac{3}{2}-\delta\right)\ell(j)}(2p)^{(1-\tau)/\tau}e$ for all $j\in\mathbb{N}$ to obtain $\sum_{\hat{\boldsymbol{\nu}}\in K_{\boldsymbol{\nu}}}\prod_{i:\hat{\nu}_i\leq 1}\left(v_{\hat{\nu}_i}^{\tau}w_{\hat{\nu}_i}^{1-\tau}\right)\leq\sum_{\boldsymbol{d}\in D_{\boldsymbol{\nu}}}\left(|\boldsymbol{d}|_1!\boldsymbol{c}^{\boldsymbol{d}}\right)^{\tau}$. Simple computations reveal that $\boldsymbol{c}=(c_j)_j\in\ell^{\tau}(\mathbb{N})$ for all $\tau>(\frac{3}{2}-\delta)^{-1}$. We apply the previous lemma and conclude the proof. □

Going back to (78), we are left with determining for which parameters $p\geq 3,\tau>\frac{1}{\frac{3}{2}-\delta}$ the series $\sum_{\boldsymbol{\nu}\in\mathcal{F}^*}\prod_{i:\nu_i>1}\left(v_{\nu_i}^{\tau}w_{\nu_i}^{1-\tau}\right)$ is summable. By means of the product structure of the summands, we can write

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}^*}\prod_{i:\nu_i>1}\left(v_{\nu_i}^{\tau}w_{\nu_i}^{1-\tau}\right)=\prod_{i\in\mathbb{N}}\sum_{\nu_i\in\mathbb{N}\setminus\{1\}}v_{\nu_i}^{\tau}w_{\nu_i}^{1-\tau}=\prod_{i\in\mathbb{N}}\left(1+\sum_{\nu_i\geq 2}\left(C_2 2^{-p\left(\left(\frac{1}{2}-\delta\right)\ell(i)+\nu_i\right)}\right)^{\tau}\left(p2^{\nu_i}\right)^{1-\tau}\right).$$

Observe that the sum is finite if $\tau\geq\frac{1}{p+1}$ and in this case

$$\sum_{\nu_i\geq 2}\left(C_2 2^{-p\left(\left(\frac{1}{2}-\delta\right)\ell(i)+\nu_i\right)}\right)^{\tau}\left(p2^{\nu_i}\right)^{1-\tau}=C_2^{\tau}2^{-p\left(\frac{1}{2}-\delta\right)\ell(i)\tau}p^{1-\tau}\sigma,$$

where $\sigma=\sigma(p,\tau)=\frac{2^{2(-(p+1)\tau+1)}}{1-2^{-(p+1)\tau+1}}$. To summarize, denoting $F_\ell:=C_2^{\tau}2^{-p\left(\frac{1}{2}-\delta\right)\ell\tau}p^{1-\tau}\sigma$, so far we have estimated $\sum_{\boldsymbol{\nu}\in\mathcal{F}^*}\prod_{i:\nu_i>1}\left(v_{\nu_i}^{\tau}w_{\nu_i}^{1-\tau}\right)\leq\prod_{i\in\mathbb{N}}\left(1+F_{\ell(i)}\right)$. We can further estimate, recalling the hierarchical indexing (5),

$$\prod_{i\in\mathbb{N}}\left(1+F_{\ell(i)}\right)\leq\exp\left(\sum_{i\in\mathbb{N}}\log\left(1+F_{\ell(i)}\right)\right)\leq\exp\left(\sum_{\ell\in\mathbb{N}_0}2^{\ell}\log\left(1+F_\ell\right)\right)\leq\exp\left(\sum_{\ell\in\mathbb{N}_0}2^{\ell}F_\ell\right).$$

The last sum can be written as $\sum_{\ell\in\mathbb{N}_0}2^{\ell}F_\ell=C_2^{\tau}p^{1-\tau}\sigma\sum_{\ell\in\mathbb{N}_0}2^{\left(1-\left(\frac{1}{2}-\delta\right)p\tau\right)\ell}$, which is finite for $\tau>\frac{1}{p\left(\frac{1}{2}-\delta\right)}$ and in this case equals $C_2^{\tau}p^{1-\tau}\sigma\left(1-2^{1-\left(\frac{1}{2}-\delta\right)p\tau}\right)^{-1}$.

Remark 30. *When $p=2$ the condition $\tau>\frac{1}{p\left(\frac{1}{2}-\delta\right)}$ just above gives $\tau>1$ for any $\delta>0$. This means that we are not able to show that piecewise* linear *sparse grids converges independently of the number of dimensions (although we see it in the numerical experiments below). Conversely, if $p\geq 3$ there exists $\frac{2}{3}<\tau<1$ that satisfies all the conditions (remember that while δ cannot be 0, it can be chosen arbitrarily small).*

Finally Theorem 23 implies the following convergence.

Theorem 31. *Consider the parameter-to-solution map of the random LLG problem $\boldsymbol{m}=\boldsymbol{M}_0+u$ as in Section 6. Recall that $\boldsymbol{M}_0\in C^{2+\alpha}(D)$ is the initial condition and $u:\mathcal{X}_{\mathbb{R}}\to\mathbb{U}_{\mathbb{R}}$ with $\mathcal{X}_{\mathbb{R}}$ and $\mathbb{U}_{\mathbb{R}}$ defined in (53) and (56) respectively. Let $\Lambda_n\subset\mathcal{F}$ denote the multi-index set (67) defined using the profits $\mathcal{P}_\nu$ (77). Let $\mathcal{I}_{\Lambda_n}$ denote the* *corresponding piecewise polynomial sparse grids interpolant of degree $p-1$ for $p \geq 3$ with nodes* (68)*. Assume that the corresponding sparse grid satisfies $\mathcal{H}_{\Lambda_n} \subset \mathcal{X}_{\mathbb{R}}$. Under the assumptions of Theorem 8, for any $\frac{2}{3} < \tau < 1$,*

$$\|\boldsymbol{m} - \mathcal{I}_{\Lambda_n} \boldsymbol{m}\|_{L^2_\mu(\mathcal{X}_{\mathbb{R}};\mathbb{U})} \leq C_{\tau,p} \left(\#\mathcal{H}_{\Lambda_n}\right)^{1-1/\tau}.$$

where $C_{\tau,p}$ is a function of τ, p but is dimension independent. In particular,

$$C_{\tau,p} = C^{\frac{1}{\tau}} \exp\left(\frac{1}{\tau} C_2^\tau p^{1-\tau} \frac{2^{2(-(p+1)\tau+1)}}{1-2^{-(p+1)\tau+1}} \frac{1}{1-2^{1-(\frac{1}{2}-\delta)p\tau}}\right),$$

where in turn C is defined in Lemma 29 and C_2 is defined in Lemma 25.

Remark 32 (On optimality of the convergence rate $-\frac{1}{2}$)**.** *The best convergence rate with respect to the number of collocation nodes predicted by the theorem is $-\frac{1}{2}$ and corresponds to $\tau = \frac{2}{3}$. This is the same as the convergence rate of the parametric truncation with respect to the number of parameters: Denoting $\boldsymbol{m}(\boldsymbol{y})$ the parametric solution for $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$ and by $\boldsymbol{m}_N(\boldsymbol{y}) \coloneqq \boldsymbol{m}((y_1, \dots y_N, 0, 0, \dots))$, for any $N \in \mathbb{N}$, its N-dimensional truncation, one can show that*

$$\|\boldsymbol{m} - \boldsymbol{m}_N\|_{L^2_\mu(\mathcal{X}_{\mathbb{R}}, L^2([0,T], H^1(D)))} \lesssim N^{-1/2}.$$

Since it is not possible to have less than 1 collocation node per dimension, the sparse grid algorithm achieves the optimal approximation rate.

In particular, piecewise quadratic approximation ($p=3$) has optimal convergence rate and using $p > 3$ does not improve the convergence rate (but may improve the constant $C_{\tau,p}$). For the same reason, sparse grid interpolation based on other 1D interpolations schemes (e.g. global polynomials) cannot give a better convergence rate (but may improve the constant).

Remark 33. *Given an approximation $\boldsymbol{m}_\Lambda(\boldsymbol{y}) = \mathcal{I}_\Lambda[\boldsymbol{m}](\boldsymbol{y})$ of the solution to the parametric LLG problem* (20)*, it is easy to sample an approximate random solution of the* random *LLG problem* (19) *too. One has to sample i.i.d. standard normal random variables $\boldsymbol{Y} = (Y_i)_{i=1}^{N_\Lambda}$ and evaluate $\boldsymbol{m}_\Lambda(Y_1, \dots, Y_{N_\Lambda})$. Here $N_\Lambda \in \mathbb{N}$ is the support size of the multi-index set Λ: $N_\Lambda \coloneqq \min\{n \in \mathbb{N} : \text{ for all } \boldsymbol{\nu} \in \Lambda \ \operatorname{supp}(\boldsymbol{\nu}) \subset \{1, \dots, n\}\}$. Equivalently, N_Λ is the number of active parameters in the sparse grid interpolant $\mathcal{I}_\Lambda$. The root-mean-square error is naturally the same as the one we estimated in the previous theorem*

$$\sqrt{\mathbb{E}_{\boldsymbol{Y}} \|\boldsymbol{m}(\boldsymbol{Y}) - \boldsymbol{m}_\Lambda(\boldsymbol{Y})\|_{\mathbb{U}}} = \|\boldsymbol{m} - \mathcal{I}_\Lambda \boldsymbol{m}\|_{L^2_\mu(\mathcal{X}_{\mathbb{R}}, \mathbb{U})}.$$

We can also draw approximate samples from the random solution of the stochastic *PDE* (16)*:*

1. *Sample a Wiener process $W(\omega, \cdot)$*
2. *Compute the first N_Λ coordinates $\boldsymbol{Y} = (Y_1, \dots, Y_{N_\Lambda}) \in \mathbb{R}^N_\Lambda$ of its Lévy-Ciesielski expansion $W(\omega, \cdot) = \sum_{i=1}^\infty Y_i \eta_i(\cdot)$*
3. *Compute $\boldsymbol{m}_\Lambda(\boldsymbol{Y})$, the approximate solution to the random LLG problem* (19)
4. *Finally compute the inverse Doss-Sussmann transform $\boldsymbol{M}_\Lambda \coloneqq e^{WG} \boldsymbol{m}_\Lambda$. (Recall the convenient expression for e^{WG} in the third line of* (17)*).*

The approximation error is again comparable to the one found in the previous theorem. Indeed, denoting $\|\cdot\|$ the root-mean-square error, the Doss-Sussmann transform implies $\sqrt{\mathbb{E}_W \|\boldsymbol{M} - \boldsymbol{M}_\Lambda\|_{\mathbb{U}}} = \left\|e^{WG}(\boldsymbol{m} - \boldsymbol{m}_\Lambda)\right\|$. The third line of (17) *followed by a triangle inequality then give*

$$\|\boldsymbol{M} - \boldsymbol{M}_\Lambda\| \leq \left(1 + \|\boldsymbol{g}\|_{L^\infty(D)} + \|\boldsymbol{g}\|^2_{L^\infty(D)}\right) \|\boldsymbol{m} - \boldsymbol{m}_\Lambda\|.$$

7.4 Numerical tests

We numerically test the convergence of the sparse grid method defined above. Since no exact sample path of the solution is available, we approximate them with the space and time approximation from [1]. This method is based on the *tangent plane scheme* and has the advantage of solving one *linear* elliptic problem per time step with finite elements. The time-stepping is based on a BDF formula. The method is high-order for both the finite elements and BDF discretization.

We consider the problem on the 2D domain $D = [0,1]^2$ with $z = 0$. The final time is $T = 1$. The noise coefficient is defined as

$$\boldsymbol{g}(\boldsymbol{x}) = \left(-\frac{1}{2}\cos(\pi x_1), -\frac{1}{2}\cos(\pi x_2), \sqrt{1 - \left(\frac{1}{2}\cos(\pi x_1)\right)^2 - \left(\frac{1}{2}\cos(\pi x_2)\right)^2}\right). \tag{80}$$

Observe that $\partial_n \boldsymbol{g} = 0$ on ∂D and that $|\boldsymbol{g}| = 1$ on D. The initial condition is $\boldsymbol{M}_0 = (0, 0, 1)$.

The space discretization is order 1 on a structured triangular mesh with 2048 elements and mesh-size $h > 0$. The time discretization is order 1 on 256 equispaced time steps of size $\tau > 0$. We use piecewise affine sparse grid, corresponding to $p = 2$. As for the multi-index selection, we compare two strategies:

- The basic profit from Section 7.2. We recall that

$$\tilde{\mathcal{P}}_{\boldsymbol{\nu}} = \prod_{i:\nu_i=1} 2^{-\frac{1}{2}\ell(i)} \prod_{i:\nu_i>1} \left(2^{\nu_i + \frac{1}{2}\ell(i)}\right)^{-p} \left(\prod_{i:\nu_i \geq 1} p2^{\nu_i}\right)^{-1} \qquad \text{for all } \boldsymbol{\nu} \in \mathcal{F},$$

  where $\ell(i) = \lceil \log_2(i) \rceil$. Compared to (75), here we have set $C_1 = C_2 = \varepsilon = 1$ and $\alpha = 0$ for simplicity.

- A modified version of the improved profit from Section 7.3, namely

$$\mathcal{P}_{\boldsymbol{\nu}} = \prod_{i:\nu_i=1} 2^{-\frac{3}{2}\ell(i)} \prod_{i:\nu_i>1} \left(2^{\nu_i + \frac{1}{2}\ell(i)}\right)^{-p} \left(\prod_{i:\nu_i \geq 1} p2^{\nu_i}\right)^{-1} \qquad \text{for all } \boldsymbol{\nu} \in \mathcal{F}, \tag{81}$$

  where again $\ell(i) = \lceil \log_2(i) \rceil$. Compared to Section 7.3, we have set $C_1 = C_2 = \gamma = 1$ and neglected the factor $r_\ell(\boldsymbol{\nu})$.

We estimate the approximation error of the sparse grid approximations with the following computable quantity: $\frac{1}{N}\sum_{i=1}^N \|\boldsymbol{m}_{\tau h}(\boldsymbol{y}_i) - \mathcal{I}_\Lambda[\boldsymbol{m}_{\tau h}](\boldsymbol{y}_i)\|_{L^2([0,T],H^1(D))}$, where $N = 1024$, $(\boldsymbol{y}_i)_{i=1}^N$ are i.i.d. standard normal samples of dimension 2^{10} each and $\boldsymbol{m}_{hk}(\boldsymbol{y}_i)$ denotes the corresponding space and time approximation of the sample paths.

Observe that if the time step is $\tau = 2^{-n}$, then the parameter-to-finite element solution map depends only on the first $n+1$ levels of the Lévy-Ciesielski expansion. In our case, $n = 8$, so the maximum relevant level is $L = 9$, i.e. 512 dimensions. In the following numerical examples we always approximate fewer dimensions, which means that the time-discretization error is negligible compared to the parametric approximation error.

The results are displayed in Figure 3. In the top plot, we observe that using basic profits leads to a sub-algebraic convergence rate which decreases as the number of approximated dimensions increases. Conversely, improved profits leads to a robust algebraic convergence of order about $\frac{1}{2}$. Piecewise *quadratic* interpolation is optimal as predicted in Section 7.3 and it delivers the same convergence rate as piecewise linear interpolation. Hence, the restriction in Theorem 31 is possibly an artefact of the proof. In view of Remark 32, it seems unnecessary to test higher polynomial degrees. In the bottom left plot, we observe that the number of active dimensions (i.e., those dimension which are seen by the sparse-grid algorithm) grows similarly for all methods, with the basic profit having a slightly higher value. Finally, we verify numerically that the approximation power of the method does not degrade when space and time approximations are refined, see the bottom right plot in Figure 3.

8 Multi-level sparse grid collocation

In the present section, we show how the sparse grid scheme defined and studied in this work can be combined with a method for space and time approximation to define a fully discrete approximation scheme. Here we employ again the linearly implicit BDF-finite element scheme from [1].

Given $\tau > 0$, consider $N_\tau = \frac{T}{\tau}$ equispaced time steps on $[0,T]$. Given $h > 0$, define a quasi-uniform triangulation $\mathcal{T}_h$ of the domain $D \in \mathbb{R}^d$ for $d \in \mathbb{N}$ with mesh-spacing h. Denote, for any $\boldsymbol{y} \in \mathcal{X}_{\mathbb{R}}$, $\boldsymbol{m}_{\tau h}(\boldsymbol{y})$ the space and time approximation of $\boldsymbol{m}(\boldsymbol{y})$. Assume that there exists a constant $C_{\mathrm{FE}} > 0$ independent of h or τ such that

$$\|\boldsymbol{m} - \boldsymbol{m}_{\tau,h}\|_{L^2_\mu(\mathcal{X}_{\mathbb{R}};\mathbb{U})} \leq C_{\mathrm{FE}}(\tau + h).$$

Moreover, we assume that the computational cost (number of floating-point operations) of computing a single $\boldsymbol{m}_{\tau h}(\boldsymbol{y})$ is proportional to

$$C_{\mathrm{sample}}(\tau, h) = \tau^{-1} h^{-d}.$$

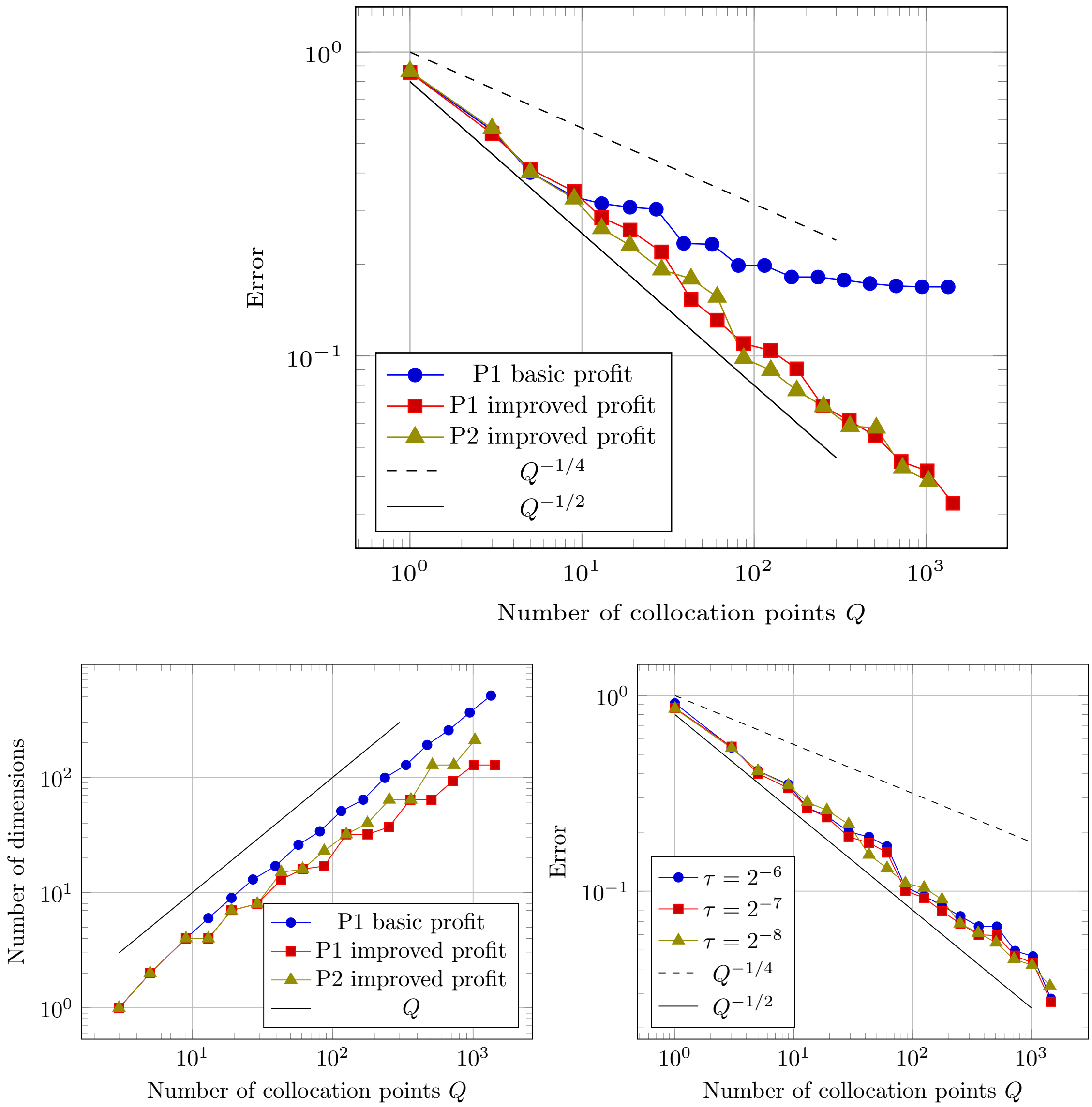


Figure 3: Approximation of $\boldsymbol{y} \mapsto \boldsymbol{m}(\boldsymbol{y})$. Top: Error vs. number of collocation nodes; Bottom left: Number of effective dimensions vs. number of collocation nodes. Bottom right: Comparison of convergence of the sparse grid approximation ($p = 3$, i.e. piecewise quadratic) for different space and time discretization parameters. In all cases time step τ and mesh size h are related by $h = 8\tau$.

Indeed, the numerical scheme requires, at each time step, solving a linear system of size proportional to the number of elements of $\mathcal{T}_h$, which in turn is proportional to h^{-d}. The latter operation can be executed with empirical linear complexity using GMRES with multigrid preconditioning. See [38], for a mathematically rigorous preconditioning strategies for LLG.

Theorem 31 shows that there exists $C_{\mathrm{SG}} > 0$ and $0 < r < \frac{1}{2}$ such that, denoting $\mathcal{I}_\Lambda$ the sparse grid interpolant and $\mathcal{H}_\Lambda$ the corresponding sparse grid,

$$\|\boldsymbol{m} - \mathcal{I}_\Lambda[\boldsymbol{m}]\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} \leq C_{\mathrm{SG}} \left(\#\mathcal{H}_\Lambda\right)^{-r}.$$

A *Single-Level* approximation of $\boldsymbol{m}$ can be defined as

$$\boldsymbol{m}^{\mathrm{SL}}_{\Lambda,\tau,h} \coloneqq \mathcal{I}_\Lambda\left[\boldsymbol{m}_{\tau,h}\right].$$

The cost of computing the single-level approximation is $C^{\mathrm{SL}}_{\Lambda,\tau,h} \coloneqq \#\mathcal{H}_\Lambda C_{\mathrm{sample}}(\tau,h)$. The approximation accuracy can be estimated as

$$\begin{aligned}\left\|\boldsymbol{m} - \boldsymbol{m}^{\mathrm{SL}}_{\Lambda,\tau,h}\right\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} &\leq \|\boldsymbol{m} - \mathcal{I}_\Lambda[\boldsymbol{m}]\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} + \|\mathcal{I}_\Lambda[\boldsymbol{m} - \boldsymbol{m}_{\tau,h}]\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} \\ &\leq C_{\mathrm{SG}} \left(\#\mathcal{H}_\Lambda\right)^{-r} + C_{\mathrm{stab}} C_{\mathrm{FE}}(h+\tau),\end{aligned}$$

where $C_{\mathrm{stab}} = C_{\mathrm{stab}}(p) > 0$ is the stability constant of the sparse grid interpolation operator, which depends on the degree $p-1$ of piecewise interpolation. A quasi-optimal single-level approximation requires balancing the three approximation parameters Λ, τ and h so that the summands in the previous estimate have similar values. This choice leads to, as it can be proved with simple computations, the following error estimate with respect to the cost $C^{\mathrm{SL}}_{\Lambda,\tau,h}$

$$\left\|\boldsymbol{m} - \boldsymbol{m}^{\mathrm{SL}}_{\Lambda,\tau,h}\right\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} \lesssim \left(C^{\mathrm{SL}}_{\Lambda,\tau,h}\right)^{-\frac{1}{\frac{1}{r}+(d+1)}}. \tag{82}$$

A *Multilevel* approximation of $\boldsymbol{m}$ can be defined following [55], to which we refer to for further details. Let $K \geq 0$ and consider a sequence of approximation parameters $(\Lambda_k)_{k=0}^K$, $(\tau_k)_{k=0}^K$ and $(h_k)_{k=0}^K$. Denote $\boldsymbol{m}_k = \boldsymbol{m}_{\tau_k,h_k}$ for $0 \leq k \leq K$ and $\boldsymbol{m}_{-1} \equiv 0$. Define the multilevel approximation as

$$\boldsymbol{m}^{\mathrm{ML}}_K \coloneqq \sum_{k=0}^{K} \mathcal{I}_{\Lambda_k}\left[\boldsymbol{m}_{K-k} - \boldsymbol{m}_{K-k-1}\right].$$

The computational cost is proportional to $C^{\mathrm{ML}}_K = \sum_{k=0}^K \#\mathcal{H}_{\Lambda_k} C_{\mathrm{sample}}(\tau_{K-k}, h_{K-k})$. To guarantee approximation, we require the following assumption on the sparse grid approximation of differences: For any $0 \leq k \leq K$,

$$\|\boldsymbol{m}_k - \boldsymbol{m}_{k-1} - \mathcal{I}_\Lambda[\boldsymbol{m}_k - \boldsymbol{m}_{k-1}]\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} \leq C_{\mathrm{SG}} \left(\#\mathcal{H}_\Lambda\right)^{-r}(h_k + \tau_k).$$

For the multilevel-approximation to be quasi-optimal, all terms in the multilevel expansion shall have similar magnitude to the K-th (finest) time and space approximation. To this end, we choose the multi-index sets Λ_k so that

$$\left(\#\mathcal{H}_{\Lambda_{K-k}}\right)^{-r} \leq C_{\mathrm{FE}} \left(C_{\mathrm{SG}}(K+1)\right)^{-1} \frac{\tau_K + h_K}{\tau_k + h_k}. \tag{83}$$

As a consequence, the multilevel error with optimal sparse grids sizes (83) can be estimated as $\left\|\boldsymbol{m} - \boldsymbol{m}^{\mathrm{ML}}_K\right\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} \leq 2C_{\mathrm{FE}}(\tau_K + h_K)$. The error can be related to the computational cost as done in [55]. We obtain the improved error-to-cost relation

$$\left\|\boldsymbol{m} - \boldsymbol{m}^{\mathrm{ML}}_K\right\|_{L^2_\mu(\mathcal{X}_\mathbb{R};\mathbb{U})} \lesssim \left(C^{\mathrm{ML}}_K\right)^{-\frac{1}{d+1}}. \tag{84}$$

We compare numerically single- and multilevel schemes on the following example of relaxation dynamics with thermal noise. The domain is $D = [0,1]^2$ with $z = 0$. The final time is $T = 1$. The noise coefficient $\boldsymbol{g}$ is set to one fifth of the coefficient defined in (80). The initial condition $\boldsymbol{M}^0$ coincides with (80). The time and space approximations are both of order 1. The sparse grid scheme is piecewise linear and the multi-index sets are built using the improved profit (81) from the previous numerical experiments. Observe that, in the following convergence tests, refinement leads automatically to an increase of the number of approximated parameters and a reduction of the parametric *truncation* error. We consider $0 \leq K \leq 5$ and define $\tau_k = 2^{-k-2}$, $h_k = 2^{-k}$, and Λ_k using the same profit-maximization as in the previous section.

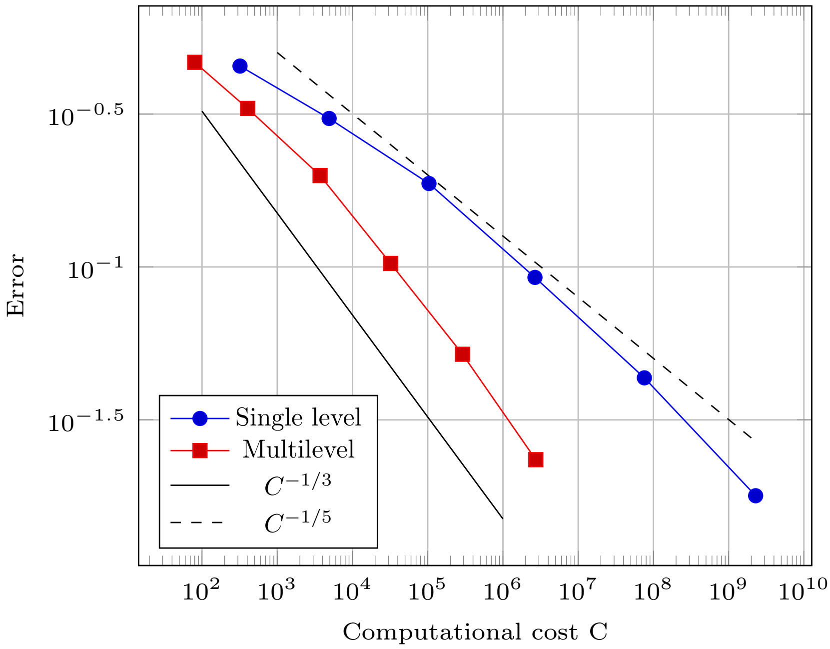


Figure 4: Comparison of single- and multilevel approximations based on piecewise polynomial sparse grids for the parametric approximation and the linearly implicit BDF-finite elements methods from [1] for time and space approximation.

For the single-level approximation, we choose Λ_k minimal such that $\#\mathcal{H}_{\Lambda_k} > 2^{2k}$. The last choice corresponds to assuming that the sparse grid approximation converges with order $r = \frac{1}{2}$ with respect to the number of collocation nodes. We compute a sequence of single-level approximations $\boldsymbol{m}^{\mathrm{SL}}_{\Lambda_k,\tau_k,h_k}$ for $k = 0, \dots K$ and report the results in Figure 4.

For the multilevel approximation, we follow formula (83). The constants $C_{\mathrm{FE}} \approx 0.7510$, $C_{\mathrm{SG}} \approx 0.1721$ and $r \approx 0.4703$ are determined with short sparse grid and finite element convergence tests. We obtain

K	$\#\mathcal{H}_{\Lambda_0}$	$\#\mathcal{H}_{\Lambda_1}$	$\#\mathcal{H}_{\Lambda_2}$	$\#\mathcal{H}_{\Lambda_3}$	$\#\mathcal{H}_{\Lambda_4}$	$\#\mathcal{H}_{\Lambda_5}$
0	1					
1	1	3				
2	1	3	10			
3	1	4	18	82		
4	2	7	27	131	602	
5	2	10	42	193	887	1500

The last figure 1500 is chosen smaller than the one required by the formula (4082) to guarantee reasonable computational times. Again results are reported in Figure 4.

Since the solution in closed form is not available, we approximate it with a reference solution. We consider 128 Monte Carlo samples of W and approximate the corresponding sample paths in space and time with time step $\tau_{\mathrm{ref}} = 2^{-9}$ and mesh size $h_{\mathrm{ref}} = 2^{-7}$. Computing the error for the single- and multilevel approximation requires first sampling the interpolants on the Monte Carlo sample parameters and then interpolating in the reference space.

The convergence test confirms that the multilevel method is superior to the single-level method.

References

[1] Georgios Akrivis, Michael Feischl, Balázs Kovács, and Christian Lubich. Higher-order linearly implicit full discretization of the Landau-Lifshitz-Gilbert equation. *Math. Comp.*, 90(329):995–1038, 2021.

[2] François Alouges. A new finite element scheme for Landau–Lifchitz equations. *Discrete Contin. Dyn. Syst. Ser. S*, 1(2):187–196, 2008.

[3] François Alouges, Anne de Bouard, and Antoine Hocquet. A semi-discrete scheme for the stochastic Landau-Lifshitz equation. *Stoch. Partial Differ. Equ. Anal. Comput.*, 2(3):281–315, 2014.

[4] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Rev.*, 52(2):317–355, 2010.

[5] Markus Bachmayr, Albert Cohen, Ronald DeVore, and Giovanni Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part II: Lognormal coefficients. *ESAIM Math. Model. Numer. Anal.*, 51(1):341–363, 2017.

[6] Ľubomír Baňas, Zdzisław Brzeźniak, and Andreas Prohl. Computational studies for the stochastic Landau-Lifshitz-Gilbert equation. *SIAM J. Sci. Comput.*, 35(1):B62–B81, 2013.

[7] Sören Bartels and Andreas Prohl. Convergence of an implicit finite element method for the Landau–Lifshitz–Gilbert equation. *SIAM J. Numer. Anal.*, 44(4):1405–1419, 2006.

[8] Ľubomír Baňas, Zdzisław Brzeźniak, Mikhail Neklyudov, and Andreas Prohl. A convergent finite-element-based discretization of the stochastic Landau-Lifshitz-Gilbert equation. *IMA J. Numer. Anal.*, 34(2):502–549, 2014.

[9] Dmitri V Berkov. Magnetization dynamics including thermal fluctuations: Basic phenomenology, fast remagnetization processes and transitions over high-energy barriers. *Handbook of Magnetism and Advanced Magnetic Materials*, 2007.

[10] William Fuller Brown Jr. Thermal fluctuations of a single-domain particle. *Physical review*, 130(5):1677, 1963.

[11] Zdzisław Brzeźniak, Ben Goldys, and Terence Jegaraj. Large deviations and transitions between equilibria for stochastic Landau-Lifshitz-Gilbert equation. *Arch. Ration. Mech. Anal.*, 226(2):497–558, 2017.

[12] Zdzisław Brzeźniak, Beniamin Goldys, and Terence Jegaraj. Weak solutions of a stochastic Landau-Lifshitz-Gilbert equation. *Appl. Math. Res. Express. AMRX*, (1):1–33, 2013.

[13] Zdzisław Brzeźniak and Liang Li. Weak solutions of the stochastic Landau-Lifshitz-Gilbert equations with nonzero anisotrophy energy. *Appl. Math. Res. Express. AMRX*, (2):334–375, 2016.

[14] Zdzisław Brzeźniak, Utpal Manna, and Debopriya Mukherjee. Wong-Zakai approximation for the stochastic Landau-Lifshitz-Gilbert equations. *J. Differential Equations*, 267(2):776–825, 2019.

[15] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[16] Gilles Carbou and Pierre Fabrie. Regular solutions for Landau-Lifschitz equation in $\mathbb{R}^3$. *Commun. Appl. Anal.*, 5(1):17–30, 2001.

[17] Yunmei Chen, Shijin Ding, and Boling Guo. Partial regularity for two-dimensional Landau-Lifshitz equations. *Acta Math. Sinica (N.S.)*, 14(3):423–432, 1998.

[18] Abdellah Chkifa, Albert Cohen, and Christoph Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *J. Math. Pures Appl. (9)*, 103(2):400–428, 2015.

[19] Ivan Cimrák. Existence, regularity and local uniqueness of the solutions to the Maxwell–Landau–Lifshitz system in three dimensions. *J. Math. Anal. Appl.*, 329:1080–1093, 2007.

[20] Albert Cohen, Ronald DeVore, and Christoph Schwab. Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.

[21] Albert Cohen, Ronald Devore, and Christoph Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.

[22] Albert Cohen, Christoph Schwab, and Jakob Zech. Shape holomorphy of the stationary Navier-Stokes equations. *SIAM J. Math. Anal.*, 50(2):1720–1752, 2018.

[23] Jean Dieudonné. *Foundations of modern analysis.* Pure and Applied Mathematics, Vol. 10-I. Academic Press, New York-London, 1969. Enlarged and corrected printing.

[24] Halim Doss. Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, (no. 2,):99–125, 1977.

[25] Sever Silvestru Dragomir. *Some Gronwall type inequalities and applications.* Nova Science Publishers, Inc., Hauppauge, NY, 2003.

[26] Dinh Dũng, Van Kien Nguyen, Christoph Schwab, and Jakob Zech. *Analyticity and Sparsity in Uncertainty Quantification for PDEs with Gaussian Random Field Inputs.* Lecture Notes in Mathematics. Springer Cham, 2023.

[27] Oliver G. Ernst, Björn Sprungk, and Lorenzo Tamellini. Convergence of sparse collocation for functions of countably many Gaussian random variables (with application to elliptic PDEs). *SIAM J. Numer. Anal.*, 56(2):877–905, 2018.

[28] Michael Feischl and Thanh Tran. The eddy current–LLG equations: FEM-BEM coupling and a priori error estimates. *SIAM J. Numer. Anal.*, 55(4):1786–1819, 2017.

[29] Michael Feischl and Thanh Tran. Existence of regular solutions of the Landau-Lifshitz-Gilbert equation in 3D with natural boundary conditions. *SIAM J. Math. Anal.*, 49(6):4470–4490, 2017.

[30] Baskar Ganapathysubramanian and Nicholas Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *Journal of Computational Physics*, 225(1):652–685, 2007.

[31] José Luis García-Palacios and Francisco J Lázaro. Langevin-dynamics study of the dynamical properties of small magnetic particles. *Physical Review B*, 58(22):14937, 1998.

[32] Thomas Gerstner and Michael Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, 2003.

[33] Beniamin Goldys, Joseph F. Grotowski, and Kim-Ngan Le. Weak martingale solutions to the stochastic Landau-Lifshitz-Gilbert equation with multi-dimensional noise via a convergent finite-element scheme. *Stochastic Process. Appl.*, 130(1):232–261, 2020.

[34] Beniamin Goldys, Kim-Ngan Le, and Thanh Tran. A finite element approximation for the stochastic Landau-Lifshitz-Gilbert equation. *J. Differential Equations*, 260(2):937–970, 2016.

[35] Michael Griebel, Frances Y. Kuo, and Ian H. Sloan. The ANOVA decomposition of a non-smooth function of infinitely many variables can have every term smooth. *Math. Comp.*, 86(306):1855–1876, 2017.

[36] Helmut Harbrecht, Marc Schmidlin, and Christoph Schwab. The gevrey class implicit mapping theorem with application to uq of semilinear elliptic pdes, 2023.

[37] Robert V Kohn, Maria G Reznikoff, and Eric Vanden-Eijnden. Magnetic elements at finite temperature and large deviation theory. *Journal of nonlinear science*, 15:223–253, 2005.

[38] Johannes Kraus, Carl-Martin Pfeiler, Dirk Praetorius, Michele Ruggeri, and Bernhard Stiftner. Iterative solution and preconditioning for the tangent plane scheme in computational micromagnetics. *J. Comput. Phys.*, 398:108866, 27, 2019.

[39] Ryogo Kubo and Natsuki Hashitsume. Brownian motion of spins. *Progress of Theoretical Physics Supplement*, 46:210–220, 1970.

[40] Olga Ladyženskaja, Vsevolod. Solonnikov, and Nina. N. Ural'ceva. *Linear and quasilinear equations of parabolic type.* Translations of Mathematical Monographs, Vol. 23. American Mathematical Society, Providence, R.I., 1968. Translated from the Russian by S. Smith.

[41] Junyu Lin, Baishun Lai, and Changyou Wang. Global well-posedness of the Landau-Lifshitz-Gilbert equation for initial data in Morrey spaces. *Calc. Var. Partial Differential Equations*, 54(1):665–692, 2015.

[42] Isaak D Mayergoyz, Giorgio Bertotti, and Claudio Serpico. *Nonlinear magnetization dynamics in nanosystems.* Elsevier, 2009.

[43] C. Melcher. Global solvability of the Cauchy problem for the Landau-Lifshitz-Gilbert equation in higher dimensions. *Indiana Univ. Math. J.*, 61(3):1175–1200, 2012.

[44] Christof Melcher. Existence of partially regular solutions for Landau-Lifshitz equations in $\mathbb{R}^3$. *Comm. Partial Differential Equations*, 30(4-6):567–587, 2005.

[45] Roger Moser. *Partial regularity for harmonic maps and related problems.* World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[46] Fabio Nobile, Lorenzo Tamellini, and Raul Tempone. Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs. *Numer. Math.*, 134(2):343–388, 2016.

[47] Fabio Nobile and Raul Tempone. Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients. *International Journal for Numerical Methods in Engineering*, 80(6-7):979–1006, 2009.

[48] Fabio Nobile, Raul Tempone, and Clayton G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.

[49] Fabio Nobile, Raul Tempone, and Clayton G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[50] Carl-Martin Pfeiler. *Numerical analysis and efficient simulation of micromagnetic phenomena*. Phd thesis, TU Wien, 2022. Available at `https://repositum.tuwien.at/handle/20.500.12708/19849?mode=simple`.

[51] Jan Prüss and Roland Schnaubelt. Solvability and maximal regularity of parabolic evolution equations with coefficients continuous in time. *J. Math. Anal. Appl.*, 256(2):405–430, 2001.

[52] Werner Scholz, Thomas Schrefl, and Josef Fidler. Micromagnetic simulation of thermally activated switching in fine particles. *Journal of Magnetism and Magnetic Materials*, 233(3):296–304, 2001.

[53] J. Michael Steele. *Stochastic calculus and financial applications*, volume 45 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001.

[54] Héctor J. Sussmann. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probability*, 6(1):19–41, 1978.

[55] Aretha L. Teckentrup, Peter Jantsch, Clayton G. Webster, and Max Gunzburger. A multilevel stochastic collocation method for partial differential equations with random input data. *SIAM/ASA J. Uncertain. Quantif.*, 3(1):1046–1074, 2015.

[56] Zhuoqun Wu, Jingxue Yin, and Chunpeng Wang. *Elliptic & parabolic equations*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006.

[57] Dongbin Xiu and Jan S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

[58] Guannan Zhang and Max Gunzburger. Error analysis of a stochastic collocation method for parabolic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 50(4):1922–1940, 2012.