TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Sparse grid approximation of stochastic PDEs: Adaptivity and approximation of the stochastic Landau–Lifshitz–Gilbert equation

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Michael Feischl BSc
E101 – Institut für Analysis und Scientific Computing, TU Wien

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Andrea Scaglioni MSc
Matrikelnummer: 11937141

Diese Dissertation haben begutachtet:

1. **Prof. Dr. Josef Dick**
   School of Mathematics & Statistics, UNSW, Sydney
2. **Dr. Martin Eigel**
   Weierstrass Institute for Applied Analysis and Stochastics, Berlin
3. **Prof. Dr. Michael Feischl**
   Institut für Analysis und Scientific Computing, TU Wien, Vienna

Vienna, 19.09.2024

Abstract

This dissertation tackles the approximation of partial differential equations (PDEs) with random data, focusing on random coefficient PDEs and stochastic PDEs (SPDEs).

As a first result, we present and analyse an adaptive sparse grid-finite element approximation of the random diffusion Poisson equation. The algorithm is based on the reliable a-posteriori error estimator [Guignard, Nobile; SIAM J. Numer. Anal. (2018)]. Firstly, we examine a parametric semidiscrete setting. We consider two possible parametric enrichment strategies and prove plain convergence and convergence with rate (with respect to the number of refinement steps) for both. Secondly, we present the fully discrete algorithm, which additionally incorporates finite element adaptivity. We consider the use of the same or different meshes for distinct collocation nodes in parameter space. We prove rate-optimality of the finite element refinement with tools from [Carstensen, Feischl, Page, Praetorius; Comput. Math. Appl. (2014)]. Finally, combining the convergence results leads to the plain convergence of the fully discrete algorithm.

As a second result, we present a novel numerical scheme for nonlinear SPDEs driven by Gaussian noise. A surrogate of the random field solution is efficiently computed following these steps:

1. Reduce the problem to approximating a parametric coefficient PDE through the Doss-Sussmann transform and the Lévy-Ciesielski expansion of the Wiener process;

2. Prove four regularity properties, which ensure holomorphic regularity and sparsity of the resulting parameter-to-solution map;

3. Use the sparsity information to design an a-priori sparse grid interpolation scheme, which can overcome the curse of dimensionality.

We apply this method to the stochastic Landau–Lifshitz–Gilbert (SLLG) equation, a model for micrometer-scale magnetic bodies whose magnetization is perturbed by heat fluctuations. This SPDE is mathematically and computationally challenging due to its nonlinearity and the presence of Gaussian noise. For SLLG, we prove the four regularity properties in two functional settings: Either Hölder or Lebesgue integrable sample paths in time. The second setting leads to algebraic, dimensions independent convergence of the sparse grid interpolation. Finally, we apply the multilevel sparse grid-finite element scheme [Teckentrup, Jantsch, Webster, Gunzburger; SIAM/ASA J. Uncertain. Quantif. (2015)] and demonstrate its superiority over the single-level method.

The results are validated through numerical experiments, some of which are performed using *SGMethods*, a Python implementation of sparse grid interpolation developed in conjunction with this dissertation.

Kurzfassung

Diese Dissertation befasst sich mit der Approximation von partiellen Differentialgleichungen (PDEs) mit Zufallsdaten, wobei der Schwerpunkt auf PDEs mit Zufallskoeffizienten und stochastischen PDEs (SPDEs) liegt. Als erstes Ergebnis präsentieren und analysieren wir eine adaptive Sparse Grid-Finite Element approximation des random diffusion Poisson Problems Der Algorithmus basiert auf dem zuverlässigen a-posteriori Fehlerschätzer [Guignard, Nobile; SIAM J. Numer. Anal. (2018)]. Zunächst untersuchen wir eine parametrische semidiskrete Einstellung. Wir betrachten zwei mögliche parametrische Anreicherungsstrategien und beweisen für beide einfache Konvergenz und Konvergenz mit Rate (in Bezug auf die Anzahl der Verfeinerungsschritte). Zweitens stellen wir den vollständig diskreten Algorithmus vor, der zusätzlich die adaptive Verfeinerung der finiten Elemente einbezieht. Wir betrachten die Verwendung gleicher oder unterschiedlicher Gitter für verschiedene Kollokationsknoten im Parameterraum. Wir beweisen die Ratenoptimalität der Finite-Elemente-Verfeinerung mit Methoden aus [Carstensen, Feischl, Page, Praetorius; Comput. Math. Appl. (2014)]. Schließlich führt die Kombination der Konvergenzergebnisse zur einfachen Konvergenz des vollständig diskreten Algorithmus.

Als zweites Ergebnis präsentieren wir ein neuartiges numerisches Schema für nichtlineare SPDEs, die durch Gaußsches Rauschen gesteuert werden. Ein Surrogat der Zufallsfeldlösung wird effizient in den folgenden Schritten berechnet:

1. Reduzieren Sie das Problem auf die Approximation einer parametrischen Koeffizienten-PDE durch die Doss-Sussmann-Transformation und die Lévy-Ciesielski-Erweiterung des Wiener-Prozesses;

2. Beweisen Sie vier Regularitätseigenschaften, die holomorphe Regularität und Sparsamkeit der resultierenden Parameter-zu-Lösungs-Abbildung gewährleisten;

3. Verwenden Sie die Informationen über die Sparsamkeit, um eine a-priori Sparse Grid Interpolation zu entwerfen, die den Fluch der Dimensionalität überwinden kann.

Wir wenden die Methode auf die stochastische Landau-Lifshitz-Gilbert-Gleichung (SLLG) an, ein Modell für magnetische Körper im Mikrometerbereich, deren Magnetisierung durch Wärmeschwankungen gestört wird. Die SPDE ist aufgrund ihrer Nichtlinearität und des Gaußschen Rauschens eine mathematische und rechnerische Herausforderung. Für SLLG beweisen wir die vier Regularitätseigenschaften in zwei funktionalen Einstellungen: Entweder Hölder- oder Lebesgue-integrierbare Stichprobenpfade in der Zeit. Die zweite Einstellung führt zu algebraischer, dimensionsunabhängiger Konvergenz der spärlichen Gitterinterpolation. Schließlich wenden wir das multilevel Sparse Grid-Finite Element scheme [Teckentrup, Jantsch, Webster, Gunzburger; SIAM/ASA J. Uncertain. Quantif. (2015)] und zeigen seine Überlegenheit gegenüber der Single-Level-Methode.

Die Ergebnisse werden durch numerische Experimente validiert, von denen einige mit *SG-Methods* durchgeführt werden, einer Python-Implementierung der Sparse-Grid-Interpolation, die in Verbindung mit dieser Dissertation entwickelt wurde.

Riassunto

Questa tesi affronta l'approssimazione delle equazioni differenziali parziali (PDE) con dati casuali, concentrandosi sulle PDE a coefficienti casuali e sulle PDE stocastiche (SPDE).

Come primo risultato, presentiamo e analizziamo un'approssimazione adattiva con sparse grid e elementi finiti dell'equazione di Poisson con diffusione casuale. L'algoritmo si basa sullo stimatore a-posteriori e affidabile [Guignard, Nobile; SIAM J. Numer. Anal. (2018)]. In primo luogo, esaminiamo il problema parametrico semidiscreto. Consideriamo due possibili strategie di arricchimento parametrico e dimostriamo la convergenza semplice e con veloctà di convergenza (rispetto al numero di passi di raffinamento) per entrambe. In secondo luogo, presentiamo l'algoritmo discreto completo, che incorpora inoltre l'adattabilità per elementi finiti. Consideriamo l'uso della stessa o diverse mesh per nodi di collocazione distinti nello spazio dei parametri. Dimostriamo la velocità ottimale del raffinamento degli elementi finiti per mezzo degli strumenti di [Carstensen, Feischl, Page, Praetorius; Comput. Math. Appl. (2014)]. Infine, la combinazione dei risultati di convergenza porta alla convergenza semplice dell'algoritmo discreto completo.

Come secondo risultato, presentiamo un nuovo schema numerico per SPDE non lineari perturbate da rumore gaussiano. Un surrogato della soluzione casuale viene calcolato in modo efficiente seguendo i seguenti passaggi:

1. Ridurre il problema all'approssimazione di una PDE a coefficienti parametrici attraverso la trasformazione di Doss-Sussmann e l'espansione di Lévy-Ciesielski del processo di Wiener;

2. Dimostrare quattro proprietà di regolarità, le quali implicano la regolarità olomorfa e la sparsità della mappa parametri-soluzione risultante dal passaggio precedente;

3. Usare le informazioni sulla sparsità per definire uno schema di interpolazione sparse grid a-priori, in grado di superare il *Curse of Dimensionality*.

Applichiamo questo metodo all'equazione stocastica di Landau–Lifshitz–Gilbert (SLLG), un modello per corpi magnetici di dimensioni micrometriche la cui magnetizzazione è perturbata da fluttuazioni di calore. Questa SPDE è matematicamente e computazionalmente impegnativa a causa della sua non linearità e della presenza di rumore gaussiano. Per la SLLG, dimostriamo le quattro proprietà di regolarità in due impostazioni funzionali: Campioni della soluzione Hölder o Lebesgue integrabili nel tempo. La seconda impostazione porta alla convergenza algebrica, indipendente dalle dimensioni, dell'interpolazione sparse grid. Infine, applichiamo lo schema multilivello sparse grid-elementi finiti [Teckentrup, Jantsch, Webster, Gunzburger; SIAM/ASA J. Uncertain. Quantif. (2015)] e dimostrano la sua superiorità rispetto al metodo a singolo livello.

I risultati sono convalidati da esperimenti numerici, alcuni dei quali sono stati eseguiti utilizzando *SGMethods*, un'implementazione Python dell'interpolazione sparse grid sviluppata insieme a questa tesi.

Acknowledgement

I am grateful for the opportunity to have spent time working on my PhD dissertation, which has proved to be a truly enriching experience, both professionally and personally. I am thankful to all the people I have met along the way, without whom this would not have been possible.

I would like to express my gratitude to my supervisor, Michael Feischl, for his invaluable guidance, collaboration and teaching. He is not only an exemplary researcher and mentor, but also a kind, empathetic and reliable individual.

I am also grateful to my dissertation examiners, Josef Dick and Martin Eigel, for taking the time and having the patience to read and evaluate my work.

I would like to thank my colleagues at ASC, especially Hubert, Amanda, Fabian, and David, and all the colleagues at Freihaus for contributing to a friendly working environment. I would also like to thank the administration and IT personnel at ASC, Andreas, Brigitte, Rebecca, Simon, Sophie, and Ursula, for their invaluable support.

I am grateful to Josef Dick and Thanh Tran for welcoming me and working with me at UNSW Sydney in early 2023 and. It was a memorable time, both in terms of the quality of the scientific exchange and the personal experience I had in Australia.

I am grateful to the TU Wien, the FWF, and the European Research Council for funding my positions. My gratitude also goes to the late Christiana Hörbiger for establishing the Christiana Hörbiger Prize, which partially financed my research stay in Australia.

I am grateful to the CRC Wave Phenomena (KIT Karlsruhe) for welcoming me into their program and for funding a series of German language courses.

Ringrazio i miei genitori Luciano e Margherita, mio fratello Marco, e i miei parenti più cari per avermi sempre sostenuto durante gli studi, sia moralmente che materialmente. Mi siete sempre stati vicino, anche nei momenti più difficili. Grazie.

I am grateful to my beloved girlfriend Aine, who brings me great joy with her love, her humour, and our truly enriching exchanges. I look forward to many more years together, full of photos, travel, delicious food, personal development and love.

Finally, I thank my old and new friends from Mantova, Trento, Lausanne, and Vienna. I have met many generous, selfless, extraordinary people who have made it great to be alive. Whether it was a group project, a German language course, a photography workshop, or simply going out for a meal or a drink, these moments add up to something greater than its parts, which has sustained and energised me over the years.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 19.09.2024

Andrea Scaglioni

Contents

Chapter 1

Introduction

In this first chapter, we introduce results and notation needed in the following chapters as well as additional facts that may be of interest to the reader. The chapter is organized as follows: In Section 1.1, we introduce the theory of random coefficient partial differential equations (PDEs). In particular, we give a rigorous definition of the problem, describe its connection to parametric coefficient PDEs, discuss regularity, and show how some stochastic PDEs (SPDE) are, in a sense, equivalent to random coefficient PDEs. In Section 1.2, we discuss the numerical approximation of parameter-to-solution maps arising from parametric coefficient PDEs. We focus on sparse polynomial approximation methods and in particular on sparse grid interpolation. We give a through description of sparse grid interpolation with examples and review the literature. In Section 1.3, we introduce the Stochastic Landau–Lifshitz–Gilbert equation, a SPDE model for magnetic materials in a thermal bath. We discuss the physics, the mathematical model, space and time approximation algorithm, and equivalence with a random coefficients PDE. Finally, in Section 1.4, we list the original contributions of this thesis.

1.1 Random coefficient PDEs

In the present section we review the mathematical theory of PDEs with random coefficients. We focus on *forward* uncertainty quantification, i.e. the task of producing approximations of the random field solution either through a "surrogate model" or through approximate quantities of interest such as moments and probability of events.

Most deterministic PDE models contain *coefficients*, which may be constants or space and/or time dependent functions (e.g. boundary and initial conditions, material parameters, forcing terms). In practical applications from engineering or physics, these coefficients are often *measured*, a process that entails a *measurement error*, or more generally are not known exactly. It is often important to quantify how the uncertainty on the problem data propagates to the solution. Applications include the study of structural vibrations [Eli99], groundwater flows [GKN$^+$11], and composite materials [BASL99].

The present section is structured as follows: In Section 1.1.1, we describe the problem class and consider the random diffusion Poisson problem as an example. Suitable parametric expansions of the random coefficient, such as the ones described in Section 1.1.2, allow us to turn the problem into a *parametric* coefficient PDE, as described in Section 1.1.3. Since parameters are often high dimensional, their approximation may be affected by the curse of dimensionality (cf. Section 1.2.1). However, as discussed in Section 1.1.4, these maps are often *regular* (they admit a holomorphic extension) and *sparse* (the importance of scalar parameters decreases quickly) which may make the efficient approximation of these functions feasible.

1.1.1 General setting and model problem

In the present section, we introduce random coefficient PDEs and present the random diffusion Poisson problem.

Given appropriate Banach spaces $\mathbb{A}, \mathbb{U}, \mathbb{F}$, a *PDE problem* can be written as: Given $a \in \mathbb{A}$, find $u \in \mathbb{U}$ such that

$$\mathcal{R}(a, u) = 0 \qquad \text{in } \mathbb{F},$$

where we denote by $\mathcal{R} : \mathbb{A} \times \mathbb{U} \to \mathbb{F}$ the *residual*, $a \in \mathbb{A}$ a *problem coefficient*, $u \in \mathbb{U}$ the *solution*, and $\mathbb{F}$ the space of residuals.

Consider the probability triple $(\Omega, \mathcal{E}, \mathcal{P})$, where Ω is a set, $\mathcal{E}$ is a σ-algebra on Ω, and $\mathcal{P}$ is a probability measure on the measurable space $(\Omega, \mathcal{E})$. (from now on, we write only "a.e. in Ω" "a.e. $\omega \in \Omega$"). Substituting a in the problem above, we obtain a *random coefficient PDE problem*: Find a random field u with $u(\omega) \in \mathbb{U}$ for a.e. $\omega \in \Omega$ such that

$$\mathcal{R}(a(\omega), u(\omega)) = 0 \qquad \text{in } \mathbb{F}, \text{ for a.e. } \omega \in \Omega. \tag{1.1}$$

Let us consider, as an example, the most popular model problem in uncertainty quantification of PDEs.

Example 1.1 (Random diffusion Poisson problem)**.** *Consider $d \in \mathbb{N}$, $D \subset \mathbb{R}^d$ a bounded Lipschitz domain, and again the probability triple $(\Omega, \mathcal{E}, \mathcal{P})$. Consider a scalar function $f \in \mathbb{F} = H^{-1}(D)$ and a scalar random field $a : \Omega \times D \to \mathbb{R}$ such that $a(\omega, \cdot) \in \mathbb{A} = L^\infty(D)$ for a.e. $\omega \in \Omega$. The* random diffusion Poisson problem *consists of determining a random field $u : \Omega \to \mathbb{U}$ such that, for a.e. $\omega \in \Omega$,*

$$\Big\{ -\nabla \cdot (a(\omega, \boldsymbol{x}) \nabla u(\omega, \boldsymbol{x})) = f(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in D, u(\boldsymbol{x}) = 0 \qquad \forall \boldsymbol{x} \in \partial D. \tag{1.2}$$

We can consider the problem in weak form for the space variable: Find $u : \Omega \times D \to \mathbb{R}$ such that for a.e. $\omega \in \Omega$, $u(\omega, \cdot) \in H_0^1(D)$ satisfies

$$\int_D a(\omega, \cdot) \nabla u(\omega, \cdot) \cdot \nabla v = \int_D f v \qquad \forall v \in H_0^1(D). \tag{1.3}$$

The precise notion of well-posedness depends on the properties of the random problem data. Let us consider two important examples of random diffusion:

- Uniformly bounded random diffusion*: Assume that there exist $0 < a_{min} \leq a_{max} < \infty$ such that*

  $$a_{min} \leq a(\omega, \boldsymbol{x}) \leq a_{max} \qquad \forall \boldsymbol{x} \in D, \text{for a.e. } \omega \in \Omega. \tag{1.4}$$

  We say that $u \in L^\infty(\Omega; H_0^1(D))$ is a weak solution of (1.2) *if* (1.3) *holds for a.e. $\omega \in \Omega$. The assumptions on $a(\cdot, \cdot)$ allow us to apply the Lax-Milgram theorem for a.e. $\omega \in \Omega$. Therefore, the previous problem has a unique random field solution that satisfies the following stability estimate:*

  $$\|u(\omega, \cdot)\|_{H^1(D)} \leq \frac{C}{a_{min}} \|f\|_{L^2(D)} \qquad \text{for a.e. } \omega \in \Omega, \tag{1.5}$$

  where $C > 0$ depends only on problem data and is independent of ω (see [LPS14, Section 9.1]);

- Non-uniformly bounded diffusion*: The assumption* (1.4) *can be weakened and still obtain a notion of solution. Assume that for all* $\omega \in \Omega$ *there exist* $0 < a_{min}(\omega) \leq a_{max}(\omega) < \infty$ *such that*

$$a_{min}(\omega) \leq a(\omega, \boldsymbol{x}) \leq a_{max}(\omega) \qquad \forall \boldsymbol{x} \in D, \text{ for a.e. } \omega \in \Omega. \tag{1.6}$$

In this case too, the solution $u(\omega, \cdot) \in H_0^1(D)$ *exists for a.e.* $\omega \in \Omega$ *(see [LPS14, Theorem 9.9]), however the uniform stability bound* (1.5) *is not available. Depending on the properties of* $a(\cdot, \cdot)$*, different estimates are available for* u*. A standard example is given by the* log-normal *random field* $a = e^Z$*, where* $Z : \Omega \times D \to \mathbb{R}$ *is a mean-zero Gaussian random field such that there exist* $L, s > 0$ *so that*

$$\mathbb{E}\left[|Z(\omega, \boldsymbol{x}) - Z(\omega, \boldsymbol{y})|^2\right] \leq L \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in D, \text{ for a.e. } \omega \in \Omega.$$

In this case, it can be shown (see [LPS14, Theorem 9.12]) that $a_{min}^{-1} \in L^p(D)$ *for any* $p \geq 1$*,* $u \in L^p(\Omega; H_0^1(D))$*, and*

$$\|u\|_{L^p(\Omega; H_0^1(D))} \leq C \left\|a_{min}^{-1}\right\|_{L^p(\Omega)} \|f\|_{L^2(D)}.$$

Random coefficients need not appear in the differential operator. They may be part of the boundary condition, initial condition, right-hand side, etc. There are special cases in which some UQ tasks can be reduced to simple deterministic problems.

Example 1.2 (Expectation of solution to a linear problem)**.** *Consider the Banach spaces* $\mathbb{U}$ *and* $\mathbb{F}$*, the* linear *and invertible operator* $\mathcal{L} : \mathbb{U} \to \mathbb{F}$*, and the random field* $f : \Omega \to \mathbb{F}$*. Consider then the problem of determining the random field* u *such that, for a.e.* $\omega \in \Omega$*,* $u(\omega) \in \mathbb{U}$ *and*

$$\mathcal{L}[u(\omega)] = f(\omega) \qquad \text{in } \mathbb{F}.$$

Equivalently, the problem can be stated as (1.1) *by setting* $\mathcal{R} = \mathcal{L} - f$*. To compute e.g. the expectation* $\mathbb{E}[u] = \int_\Omega u d\mathcal{P}$*, we can exploit the linearity of* $\mathcal{L}$ *to obtain*

$$\mathbb{E}[f] = \mathbb{E}[\mathcal{L}u] = \mathcal{L}\mathbb{E}[u] \qquad \text{in } \mathbb{F}.$$

In this way, we have reduced the classical UQ task of computing the expectation of the solution to a problem to the approximation of its deterministic counterpart.

1.1.2 Parametric expansions of random fields

In the present section, we present two well-known random field *approximate* representations that aim at splitting its dependence on the "physical" variables (space and time) from the random component. This is done by fixing a basis for the physical variables and writing the random field as a linear combination thereof, with *independent* scalar random coefficients.

The Karhunen-Loève expansion (KLE)

The present section is based on [LPS14, Chapter 5.4].

Let $d \in \mathbb{N}$ and $D \subset \mathbb{R}^d$ a bounded connected domain. Consider a random field $Z : \Omega \times D \to \mathbb{R}$ such that $Z \in L^2(\Omega; L^2(D))$, i.e. a *second order* random field into $L^2(D)$. Denote its mean function by $\mathbb{E}[Z] : D \to \mathbb{R}$ and its covariance function by $c : D \times D \to \mathbb{R}$, which we assume to be symmetric. It holds that $\mathbb{E}[Z] \in L^2(D)$,the realizations $Z(\omega, \cdot) \in L^2(D)$ for a.e. $\omega \in \Omega$ (see

[LPS14, Lemma 5.27]), and $c \in L^2(D \times D)$ is symmetric. Consider the *covariance operator*, i.e. the following integral operator with kernel c:

$$\mathcal{C} : L^2(D) \to L^2(D) \qquad \mathcal{C}[f](\boldsymbol{x}) \coloneqq \int_D c(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}. \tag{1.7}$$

The properties of c imply (Proposition D.5) that $\mathcal{C}$ is a compact, self-adjoint operator (Definition D.4). The spectral theorem D.7 gives the existence of eigenvalues $(\lambda_j)_{j\in\mathbb{N}}$ and eigenvector $(\phi_j)_{j\in\mathbb{N}}$ of $\mathcal{C}$ (Definition D.6) . Moreover, $(\phi_j)_{j\in\mathbb{N}}$ form an orthogonal basis of $L^2(D)$.
The *Karhunen-Loève expansion (KLE)* of the random field Z is by definition:

$$\mathbb{E}[Z](\boldsymbol{x}) + \sum_{j\in\mathbb{N}} Y_j(\omega)\sqrt{\lambda_j}\phi_j(\boldsymbol{x}), \tag{1.8}$$

where the coordinates Y_j are pairwise independent *scalar* random variables taking values each in $\Gamma_j \subset \mathbb{R}$. Moreover, they can be expressed as

$$Y_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \langle Z(\omega, \cdot) - \mathbb{E}[Z], \phi_j \rangle_{L^2(D)} \qquad \forall j \in \mathbb{N}. \tag{1.9}$$

The convergence properties of the KLE are summarized in the following

Theorem 1.3 (see [LPS14, Theorem 5.28, Theorem 5.29])**.** *A second order random field $Z \in L^2(\Omega; L^2(D))$ with mean $\mathbb{E}[Z] : D \to \mathbb{R}$ and covariance $c(\cdot,\cdot) : D \times D \to \mathbb{R}$ is represented by* (1.8)*, where $(\lambda_j, \phi_j)_{j\in\mathbb{N}}$ are eigenpairs of the covariance operator* (1.7)*, $\lambda_1 \geq \lambda_2 \geq ... \geq 0$, and the random variables $(Y_j)_{j\in\mathbb{N}}$, defined by* (1.9)*, have mean zero, unit variables, and are uncorrelated. The series is understood to converge in $L^2(\Omega; L^2(D))$.*
Additionally, for real valued processes, compact $D \subset \mathbb{R}$, and continuous covariance $c(\cdot,\cdot)$ the KLE also converges uniformly*.*
Finally, if Z is a Gaussian random field, then the $(z_j)_{j\in\mathbb{N}}$ are i.i.d. standard normal random variables.

For any $J \in \mathbb{N}$, the *J-terms truncated KLE* is

$$Z_J(\omega, \boldsymbol{x}) = \mathbb{E}[Z](\boldsymbol{x}) + \sum_{j=1}^{J} Y_j(\omega)\sqrt{\lambda_j}\phi_j(\boldsymbol{x}).$$

Since $(\phi_j)_j$ is an orthonormal basis of $L^2(D)$, it is easy to quantify the $L^2(\Omega; L^2(D))$ error of a J-terms KLE:

$$\begin{aligned}\mathbb{E}\left[\|Z - Z_J\|^2_{L^2(D)}\right] &= \mathbb{E}\left\langle \sum_{j>J} Y_j\sqrt{\lambda_j}\phi_j, \sum_{k>J} Y_k\sqrt{\lambda_k}\phi_k \right\rangle = \sum_{j,k>J} \sqrt{\lambda_j}\sqrt{\lambda_k}\mathbb{E}[Y_jY_k]\langle\phi_j, \phi_k\rangle \\ &= \sum_{j>J} \lambda_j,\end{aligned}$$

so that

$$\|Z - Z_J\|_{L^2(\Omega;L^2(D))} = \sqrt{\sum_{j>J} \lambda_j}.$$

Another estimate which does not involve the whole spectrum $(\lambda_j)_{j\in\mathbb{N}}$ is

$$\sqrt{\mathbb{E}\left[\|Z - Z_J\|^2_{L^2(D)}\right]} \leq \sqrt{\mathbb{E}\left[\|Z - \mathbb{E}[Z]\|^2_{L^2(D)}\right]} + \sqrt{\mathbb{E}\left[\|Z_J - \mathbb{E}[Z]\|^2_{L^2(D)}\right]}$$
$$= \sqrt{\int_D \mathrm{Var} Z} + \sqrt{\int_D \mathrm{Var} Z_J} = \sqrt{\int_D \mathrm{Var} Z} + \sqrt{\sum_{j\leq J} \lambda_j}.$$

This time only the first J eigenvalues and variance of Z are needed.

Quantifying the truncation error is an important task in numerical analysis because some methods can handle only a finite number of scalar parameters (see also Remark 1.7 below).

To compute the KLE of a random field, it is in general necessary to approximate eigenvalues and eigenvectors of the covariance operator. The Wiener process (see [Eva13, Chapter 3] for an introduction, in particular Section 3.2.1 for the definition) is an exception for which eigenpairs are known exactly.

Example 1.4 (KLE of the Wiener process)**.** *Simple computations show that the Wiener process $W : \Omega \times [0,1] \to \mathbb{R}$ admits the KLE (1.8) with:*

$$\lambda_j = \left(\pi\left(j - \frac{1}{2}\right)\right)^{-2}, \qquad \phi_j(t) = \sqrt{2}\sin\left(\left(j - \frac{1}{2}\right)\pi t\right) \qquad \forall j \in \mathbb{N}.$$

As stated in the previous theorem, the scalar random variables $(Y_j)_{j\in\mathbb{N}}$ are i.i.d. standard normals.

The KLE is optimal *in the mean-square error sense* as a consequence of it being constructed with the leading eigenvalues of the covariance operator:

Proposition 1.5. *Denote with $(f_j)_{j\in\mathbb{N}}$ an orthonormal basis of $L^2(D)$ and consider, for some $J \in \mathbb{N}$, the J-terms truncation of Z, $Z_{J,f_j} = \sum_{j\leq J} \langle Z, f_j\rangle f_j$. The mean square error $\mathbb{E}\|Z - Z_J\|^2_{L^2(D)}$ is minimized when $f_j = \phi_j$, the j-th eigenvector of the covariance operator $\mathcal{C}$ defined in (1.7), for all $j > J$.*

However, this does not mean that the KLE is the "best" parametric expansion for every application. As explained in [BCDM17], using the Lévy-Ciesielski expansion (see below) can in some cases lead to an improved sparsity. Moreover, being optimal *in the L^2 sense* may just be not be relevant for problems in which other norms are of interest.

Let us also mention that often there is an intuitive interpretation of the rate of decay of the eigenvalues $(\lambda_n)_{n\in\mathbb{N}}$: The faster the decay of the eigenvalues , the "smoother" the sample paths of the random field look. Conversely, a slower decay gives "rough" sample paths. See Figure 1.1. This intuitive picture is however only valid when the eigenvector $(\boldsymbol{\phi}_n)_{n\in\mathbb{N}}$ of the covariance operator are smooth.

The Lévy-Ciesielski expansion (LCE) of the Wiener process

The present section is based on [Eva13, Section 3.3.3].
The *Haar family* is a sequence of functions $(h_{\ell,j})_{\ell\in\mathbb{N}_0, j=1,\dots,\lceil 2^{\ell-1}\rceil}$ with $h_{\ell,j} : [0,1] \to \mathbb{R}$ defined by using the "mother wavelet" $h : \mathbb{R} \to \mathbb{R}$

$$h(t) := \begin{cases} \frac{1}{2} & \text{for } t \in \left[0, \frac{1}{2}\right] \\ -\frac{1}{2} & \text{for } t \in \left(\frac{1}{2}, 1\right] \\ 0 & \text{otherwise.} \end{cases}$$

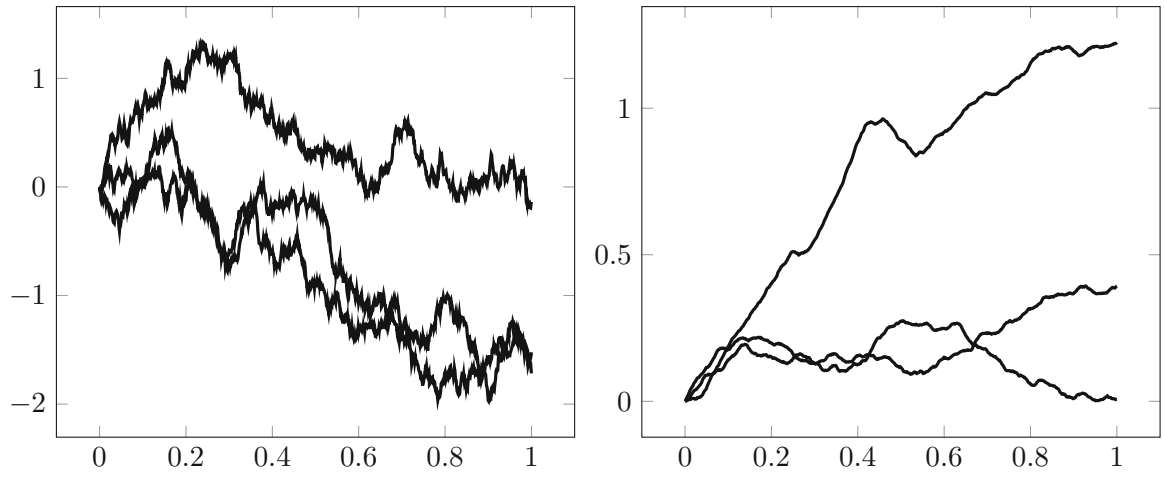

Figure 1.1: Visual comparison of stochastic processes with different decays of absolute value of eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ and same eigenvectors $\phi_j(t) = \sqrt{2}\sin\left((j-\frac{1}{2})\pi t\right)$, $t \in [0,1]$. Left: $\lambda_j = (\pi(j-\frac{1}{2}))^{-2}$, i.e. the Wiener process; Right: $\lambda_j = (\pi(j-\frac{1}{2}))^{-4}$.

In a wavelet fashion, we define:

$$h_{0,1} \equiv 1, \quad \text{and} \quad h_{\ell,j}(t) = 2^{\ell/2} h\left(2^{\ell-1}t - j + 1\right) \qquad \forall \ell \in \mathbb{N}, j = 1, \dots, 2^{\ell-1}.$$

The Haar basis is a complete orthonormal basis of $L^2(0,1)$.

We may use it to write an expansion of *white noise*, i.e. a formal time derivative of the Wiener process (see [Eva13, Section 3.2.3] for more details), as

$$\dot{W}(\omega,t) = \sum_{\ell\in\mathbb{N}_0} \sum_{j=1}^{\lceil 2^{\ell-1}\rceil} Y_{\ell,j}(\omega) h_{\ell,j}(t),$$

where $(Y_{\ell,j})_{\ell,j}$ are i.i.d. standard normal random variables. Since the Wiener process is a formal anti-derivative of the white noise, it is reasonable to express it as the following *Lévy-Ciesielski expansion* (LCE):

$$\begin{aligned} W(\omega,t) &= \sum_{\ell\in\mathbb{N}_0} \sum_{j=1}^{\lceil 2^{\ell-1}\rceil} Y_{\ell,j}(\omega)\eta_{\ell,j}(t), \qquad \text{where} \\ \eta_{\ell,j}(t) &:= \int_0^t h_{\ell,j}(s)\mathrm{d}s, \qquad (Y_{\ell,j})_{\ell,j} \text{ i.i.d. } \mathcal{N}(0,1). \end{aligned} \tag{1.10}$$

The family $(\eta_{\ell,j})_{\ell,j}$ is called the *Faber-Shauder basis* and can be written as follows: First define the "mother wavelet" $\eta : \mathbb{R} \to \mathbb{R}$

$$\eta(t) := \begin{cases} t & t \in [0, \frac{1}{2}] \\ 1-t & t \in (\frac{1}{2}, 1] \\ 0 & \text{otherwise.} \end{cases}$$

In a wavelet fashion, we define:

$$\eta_{0,1}(t) = t, \qquad \eta_{\ell,j} = 2^{-\frac{\ell-1}{2}} \eta(2^{-(\ell-1)}t - j + 1) \qquad \forall \ell \in \mathbb{N}, j = 1, \dots, 2^{\ell-1}.$$

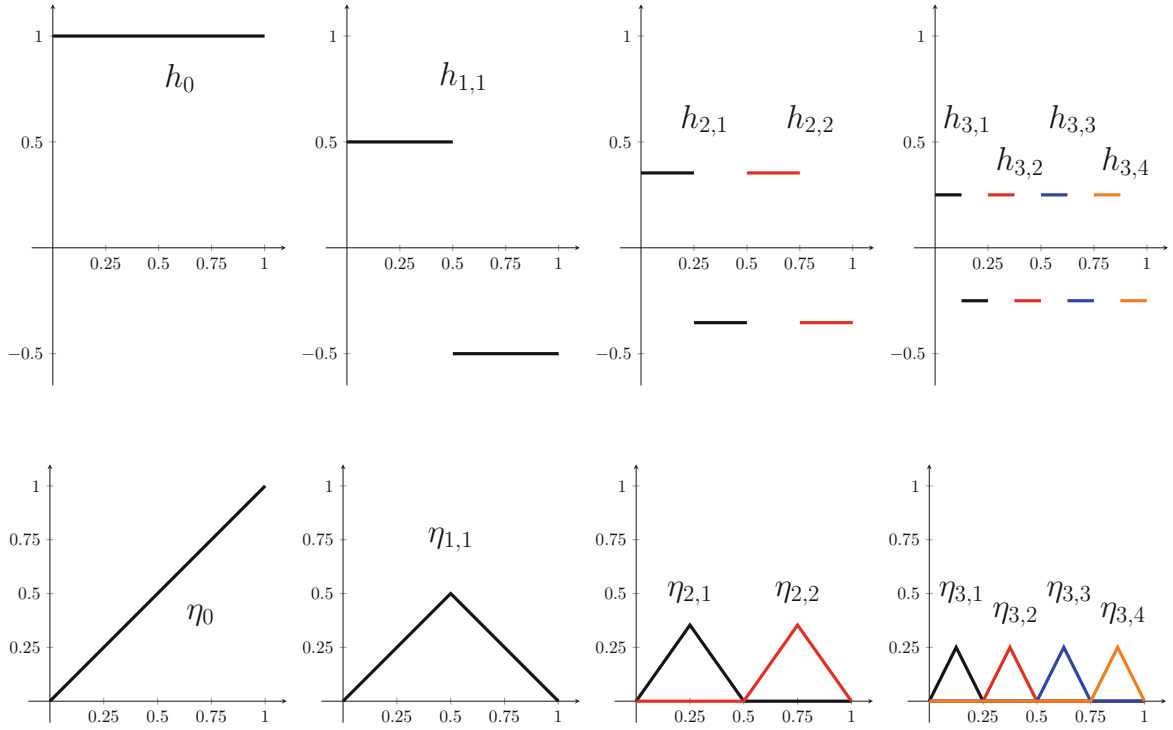

Figure 1.2: Top: The first eight Haar basis functions on $[0, 1]$ split by level. Each function is zero wherever it is not plotted. Bottom: The first eight Faber-Shauder basis functions on $[0, 1]$ split by level.

See Figure 1.2 for a plot of the first Haar and Faber-Schauder basis functions. Observe that $\|\eta_{0,1}\|_{L^\infty(0,1)} = 1$, $\operatorname{supp}(\eta)_{0,1} = (0,1]$ and for all $\ell \in \mathbb{N}, j = 1, \dots, 2^{\ell-1}$, $\|\eta_{\ell,j}\|_{L^\infty(0,1)} = 2^{-(\ell+1)/2}$, $\operatorname{supp}(\eta_{\ell,j}) = \left(\frac{j-1}{2^{\ell-1}}, \frac{j}{2^{\ell-1}}\right)$.

While this derivation above is formal, the LCE converges uniformly in t, almost surely to a continuous function which coincides with the Wiener process everywhere (see [Ste01, Section 3.4]):

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

The main steps of a possible proof are:

1. If there exists $L \in \mathbb{N}$ such that the sequence $(y_{\ell,j})_{\ell,j} \in \mathbb{R}^{\mathbb{N}}$ satisfies: $|y_{\ell,j}| \ll 2^{\ell/2}$ for all $\ell > L$, $j = 1, \dots, 2^{\ell-1}$, then $\sum_{\ell,j} y_{\ell,j}\eta_{\ell,j}$ converges uniformly on $[0,1]$;

2. For $(Y_n)_{n\in\mathbb{N}}$ i.i.d. standard normal random variables, there exists a random variable C such that $|Y_n(\omega)| = \mathcal{O}(\sqrt{\log n})$ for all $n \geq 2$ and $\mathcal{P}(0 < C < \infty) = 1$. The bound can equivalently be written using the hierarchical indexing (see (1.11) below): $|Y_{\ell,j}| \leq C(\ell+1)^{1/2}$ for all $\ell \geq 1$.

3. As a direct consequence of the first point, the series (1.10) converges absolutely and uniformly in $t \in [0,1]$ and for a.e. $\omega \in \Omega$;

4. The resulting function W is a Wiener process, in particular $W(0) = 0$, increments $W(t) - W(s)$ are distributed $\mathcal{N}(0, t-s)$ for any $0 \leq s < t \leq 1$ and disjoint increments are independent. The first fact is trivially true, the second and third can be proved using the notion of characteristic function of a random variable (see Definition C.1);

5. The sample paths $t \mapsto W(\omega,t)$ are continuous for a.e. ω. Indeed, they are uniform limit of continuous functions.

The terms of the KLE can be indexed linearly, i.e. $W(\omega,t) = \sum_{n=1}^{\infty} Y_n(\omega)\eta_n(t)$, as opposed to the hierarchical indexing used so far. The two indexing systems, hierarchical and linear, are related via

$$\eta_{\ell,j} = \eta_n \quad \Longleftrightarrow \quad \ell = \ell(n) := \lceil \log_2(n) \rceil \text{ and } j = n - \ell(n). \tag{1.11}$$

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

Remark 1.6 (Comparison of LCE and KLE of the Wiener process)**.** *The Wiener process, being a second order stochastic process, admits a KLE as seen in Example 1.4. See Figure 1.3 for a comparison of the first KLE and LCE expansion basis functions. An important difference between the two expansions is that the Faber-Schauder basis functions have compact supports. Moreover, basis functions belonging to the same level have disjoint supports.*

It is also interesting to compare the truncation errors introduced by KLE and LCE in different norms in time. Recall that the Faber-Schauder basis is not an orthonormal basis of $L^2(0,T)$ so it could, a priori, lead to a smaller error compared to the KLE with the same number of terms. See in Figure 1.4 for a computational comparison. We see that the KLE performs better, but only by a constant.

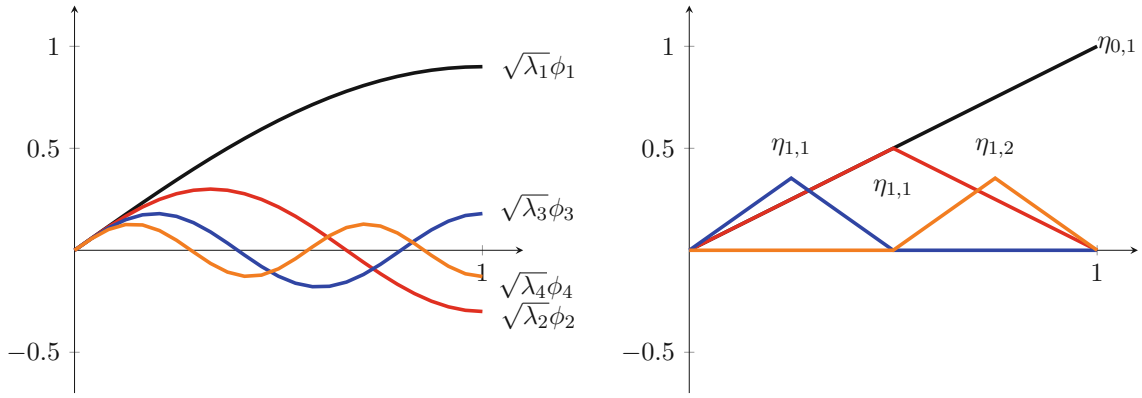

Figure 1.3: Comparison of KLE (left) and LCE (right) expansion basis function. For each, we plot the first four basis functions

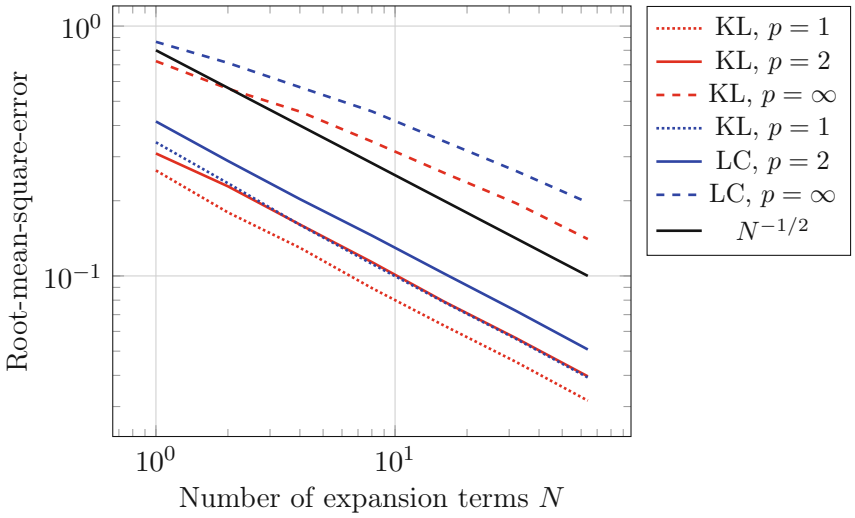

Figure 1.4: Comparison between KLE nad LCE approximation power on the Wiener process. For each method and each $p \in \{1, 2, \infty\}$, we approximate the root mean square error (RMSE) with 128 Monte Carlo samples and the $L^p(0, 1)$ norm in time.

Exact Gaussian random field sampling methods

Let us also briefly also mention the *circulant embedding method* [KTB11]. It allows us to compute the square root of covariance matrices and, in particular, to sample stationary Gaussian random fields exactly at finitely many discrete locations given their covariance there. The complexity with respect to the number of samples is $\mathcal{O}(Nlog(N))$ because the necessary operations can be performed with the *Fast Fourier transform*.

Finally, we also mention [FKS18], where the authors use *H-matrices* to factorize covariance matrices in linear complexity. This method ca be applied to sample non-stationary Gaussian processes, on more general point sets than the ones admissible with circulant embedding.

1.1.3 Reduction of a random coefficient PDE to a parametric coefficient PDE

Recall the setting outlined in the beginning of Section 1.1.1, in particular the random coefficient PDE (1.1). The random field a or a function thereof can be expanded with the KLE (or, alternatively, the LCE if it is the Wiener process). So we may write the random field as $a = a(\boldsymbol{Y}) \in \mathbb{A}$, where $\boldsymbol{Y} = (Y_n)_{n\in\mathbb{N}}$ is a vector of appropriate scalar pairwise independent random variables taking values in $\boldsymbol{\Gamma} = \bigotimes_{n\in\mathbb{N}} \Gamma_n$. The map

$$a : \boldsymbol{\Gamma} \to \mathbb{A}$$

is the *parametric expansion of the random field*. We define the *parametric coefficient PDE problem*: Find $u : \boldsymbol{\Gamma} \to \mathbb{U}$ such that

$$\mathcal{R}(a(\boldsymbol{y}), u(\boldsymbol{y})) = 0 \qquad \text{in } \mathbb{F}, \text{for a.e. } \boldsymbol{y} \in \boldsymbol{\Gamma}. \tag{1.12}$$

The Doob-Dynkin lemma [Øks03, Lemma 2.1.2] proves that the solution u can be written as a function of $\boldsymbol{y}$.

Problem (1.12) is equivalent to the *random* coefficient PDE problem (1.1) in the following sense: If the solution $u : \boldsymbol{\Gamma} \to \mathbb{U}$ of (1.12) is available, a solution of (1.1) is given by $u(\boldsymbol{Y})$, where $\boldsymbol{Y} = (Y_1, Y_2, \dots)$ are i.i.d. scalar random variables as in the KLE (or other parametric expansion). In other words, once the solution u of the parametric coefficients PDE is available, sampling the random field solution of the random coefficients PDE is just a matter of sampling appropriate scalar random variables.

Remark 1.7 (Parametric truncation error)**.** *It is important to observe that the parameter $\boldsymbol{y} = (y_n)_{n\in\mathbb{N}}$ is a real* sequence*. For some numerical schemes, this assumption may be too weak. We can consider the* truncated *parameter $\boldsymbol{y}_N = (y_1, \dots, y_N) \subset \boldsymbol{\Gamma}_N = \bigotimes_{n=1}^{N} \Gamma_n$ for some $N \in \mathbb{N}$ and the corresponding* truncated problem*: Find $u_N : \boldsymbol{\Gamma}_N \to \mathbb{U}$ such that*

$$\mathcal{R}(a(\boldsymbol{y}_N), u_N(\boldsymbol{y}_N)) = 0 \qquad \text{in } \mathbb{F}, \text{for a.e. } \boldsymbol{y}_N \in \boldsymbol{\Gamma}_N.$$

Substituting the original problem with this introduces a truncation error $\|u - u_N\|$*, where the norm depends on the setting. The rate of decay of the truncation error as $N \to \infty$ is determined by the properties of the solution operator, which maps $a(\boldsymbol{y})$ to $u(\boldsymbol{y})$, as well as the parametric expansion (e.g. the KLE or LCE), which maps $\boldsymbol{y}$ to $a(\boldsymbol{y})$.*

Truncated parameters like $\boldsymbol{y}_N \in \boldsymbol{\Gamma}_N$ can be identified with full sequences simply by appending a tail of zeros: $(y_1, \dots, y_N, 0, 0, \dots) \in \boldsymbol{\Gamma}$.

The finite dimensional noise assumption *consists of assuming the existence of $N \in \mathbb{N}$ such that for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$ and $n > N$, a and u are independent of y_n. This effectively allows us to identify the set of sequences $\boldsymbol{\Gamma}$ with a subset of $\mathbb{R}^N$.*

Recall the random diffusion Poisson problem from Example 1.1 and let us now consider its parametric counterpart.

Example 1.8 (Parametric diffusion Poisson problem)**.** *Consider again $d \in \mathbb{N}$, $D \subset \mathbb{R}^d$ a bounded Lipschitz domain, a probability triple $(\Omega, \mathcal{E}, \mathcal{P})$, and a scalar function $f \in L^2(D)$. A parametric expansion of the random diffusion $a(\omega, \boldsymbol{x})$ leads to the* parametric diffusion Poisson problem*: Find $u : \boldsymbol{\Gamma} \times D \to \mathbb{R}$ such that*

$$\begin{cases} -\nabla \cdot (a(\boldsymbol{y}, \boldsymbol{x}) \nabla u(\boldsymbol{y}, \boldsymbol{x})) = f(\boldsymbol{x}) & \forall \boldsymbol{x} \in D, \boldsymbol{y} \in \boldsymbol{\Gamma} \\ u(\boldsymbol{x}) = 0 & \forall \boldsymbol{x} \in \partial D. \end{cases} \tag{1.13}$$

The problem in weak form for the space variable reads: Find $u : \boldsymbol{\Gamma} \times D \to \mathbb{R}$ such that, for a.e. $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $u(\boldsymbol{y}, \cdot) \in H_0^1(D)$ satisfies

$$\int_D a(\boldsymbol{y}, \cdot) \nabla u(\boldsymbol{y}, \cdot) \cdot \nabla v = \int_D f v \qquad \forall v \in H_0^1(D). \tag{1.14}$$

Depending on the type of random model for the diffusion, we obtain different notions of well-posedness:

- Uniformly bounded random diffusion*: An example of parametric diffusion that could arise from a uniformly bounded random diffusion is the following:*

$$a(\boldsymbol{y}, \boldsymbol{x}) = \mathbb{E}[a](\boldsymbol{x}) + \sum_{n \in \mathbb{N}} y_n \phi_n(\boldsymbol{x}) \qquad \boldsymbol{x} \in D, \tag{1.15}$$

  where the sequence of real parameters $\boldsymbol{y} = (y_n)_{n \in \mathbb{N}} \in \boldsymbol{\Gamma} = [-1, 1]^{\mathbb{N}}$ represent uniform i.i.d. random variables on $\Gamma = [0, 1]$ and $\phi_n \in L^\infty(D)$ for all $n \in \mathbb{N}$. For the Poisson problem to be well-posed, we assume uniform ellipticity*: There exist $r > 0$ such that*

$$\sum_{j \in \mathbb{N}} |\psi_n| < \mathbb{E}[a] - r \qquad \text{on } D.$$

  Thus, suitable values of a_{min} and a_{max} in Example 1.1 are $a_{min} = \mathbb{E}[a] - r$ and $a_{max} = \mathbb{E}[a] + r$. The norm of the parametric solution $u \in L^\infty(\boldsymbol{\Gamma}; H_0^1(D))$ is bounded analogously to (1.5).

- Non-uniformly bounded diffusion*: When for example the diffusion follows the log-normal distribution (see Example 1.1), we apply the KLE (see Section 1.1.2) to $\log(a)$ to obtain*

$$a(\omega, \boldsymbol{x}) = \exp\left(\sum_{n \in \mathbb{N}} Y_n(\omega) \phi_n(\boldsymbol{x}) \right)$$

  where $Y_1, Y_2, \ldots$ are i.i.d. standard normal random variables and $\phi_n \in L^\infty(D)$ for all $n \in \mathbb{N}$. This diffusion model, unlike the one in the previous example, is positive for any choice of parameters. The parametric solution $u \in L^p(\boldsymbol{\Gamma}; H_0^1(D))$ satisfies an analogous stability estimate to its random counterpart for any $p \geq 1$.

1.1.4 Regularity and sparsity of parametric PDEs

The optimal rate of approximation of a high dimensional function $u : \boldsymbol{\Gamma} \to \mathbb{R}$ with $\boldsymbol{\Gamma} \subset \mathbb{R}^{\mathbb{N}}$ is a highly problem-dependent quantity. It is not even clear a priori that the function can be approximated with an algebraic rate and it is not clear how an optimal approximation can

be built. As we will see below, the regularity of u can be leveraged to prove that certain approximations do not suffer from the *curse of dimensionality* (see discussion in Section 1.2.1). The precise meaning of "regularity" depends on the chosen approximation method. The efficacy of an approximation method depends on how well it "matches" the regularity of the function u.

Below, we work with *sparse polynomial* approximations. In this case, "regularity" loosely means a *sufficiently fast decay of the partial derivatives of* u. This is because sparse polynomial approximation methods are based on a *partition of unity* decomposition:

$$\mathrm{id} = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}}} \mathfrak{H}_{\boldsymbol{\nu}},$$

where the definition of $\mathfrak{H}_{\boldsymbol{\nu}} : C^0(\boldsymbol{\Gamma}, \mathbb{U}) \to \mathbb{P}(\boldsymbol{\Gamma}, \mathbb{U})$ depends on the specific scheme in use. An instance of the numerical scheme is uniquely determined by choosing an *admissible* (again the concrete definition depends on the numerical scheme) and *finite* multi-index set $\Lambda \subset \mathbb{N}_0^{\mathbb{N}}$ and considering the approximation operator

$$\mathcal{I}_\Lambda \coloneqq \sum_{\boldsymbol{\nu} \in \Lambda} \mathfrak{H}_{\boldsymbol{\nu}} : C^0(\boldsymbol{\Gamma}, \mathbb{U}) \to \mathbb{P}(\boldsymbol{\Gamma}, \mathbb{U}).$$

A prominent example is sparse grid interpolation (see Section 1.2.3).

Because of this structure, the error can in general be estimated, for $1 \leq p \leq \infty$, as

$$\left\| (\mathrm{id} - \mathcal{I}_\Lambda)\, u \right\|_{L^p(\boldsymbol{\Gamma}, \mathbb{U})} \leq \sum_{\boldsymbol{\nu} \notin \Lambda} \left\| \mathfrak{H}_{\boldsymbol{\nu}} u \right\|_{L^p(\boldsymbol{\Gamma}, \mathbb{U})}.$$

If an $L^q(\boldsymbol{\Gamma}, \mathbb{U})$ estimate of $\mathfrak{H}_{\boldsymbol{\nu}}$ for $0 < q < p$ is available, we can apply *Stechkin's lemma* (see Lemma D.1) to estimate the last sum and, in particular, the error.

Several strategies have proved to be effective at estimating the derivatives of the parameter-to-solution map:

- *Proving existence of a sparse holomorphic extension*: This technique was studied in [CDS11] and is also the strategy we adopt in Chapter 3;
- *Induction argument*: One derives equations for the partial derivatives of u with respect tot the scalar parameters inductively. This technique was introduced in [GKN+15];
- *Proving Gevrey regularity*: This approach was recently introduced in [HSS23].

The first strategy often follows this line:

1. Define a complex extension $u : \boldsymbol{\Sigma} \to \mathbb{U}$ of u with $\boldsymbol{\Gamma} \subset \boldsymbol{\Sigma} \subset \mathbb{C}^{\mathbb{N}}$;
2. Prove that the complex extension is holomorphic and uniformly bounded, i.e. there exists $C(\boldsymbol{\Sigma}) > 0$ such that $\sup_{\boldsymbol{z} \in \boldsymbol{\Sigma}} \|u(\boldsymbol{z})\|_{\mathbb{U}} \leq C(\boldsymbol{\Sigma})$;
3. Prove that there exists $\boldsymbol{\rho} = (\rho_n)_{n \in \mathbb{N}}$ such that for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $\boldsymbol{B}(\boldsymbol{y}, \boldsymbol{\rho}) \subset \boldsymbol{\Sigma}$. Here we introduced the *polydisk* $\boldsymbol{B}(\boldsymbol{y}, \boldsymbol{\rho}) = \bigotimes_{n=1}^{N} B(y_n, \rho_n)$, $B(y_n, \rho_n) \coloneqq \{z \in \mathbb{C} : |z - y_n| < \rho_n\} \subset \mathbb{C}$ for all $n \in \mathbb{N}$;
4. Apply Estimate (D.3) (consequence of the Cauchy's integral formula D.8) to obtain, for $\{\boldsymbol{y}_1 \dots, \boldsymbol{y}_n\} \subset \{\boldsymbol{y}_1 \dots, \boldsymbol{y}_N\}$ and $\boldsymbol{p} \in \mathbb{N}_0^n$,

$$\left\| \partial_{y_1}^{p_1} \dots\dots \partial_{y_n}^{p_n} u(\boldsymbol{y}) \right\|_{\mathbb{U}} \leq \boldsymbol{p}! \prod_{i=1}^{n} \rho_n^{-p_n} C(\boldsymbol{\Sigma}) \qquad \forall \boldsymbol{y} \in \boldsymbol{\Gamma}; \tag{1.16}$$

5. The same estimate holds for the Lebesgue norm:

$$\left\|\partial_{y_1}^{p_1} \cdots\cdots \partial_{y_n}^{p_n} u\right\|_{L^p_{\boldsymbol{\mu}}(\boldsymbol{\Gamma},\mathbb{U})} \leq \boldsymbol{p}! \prod_{i=1}^{n} \rho_n^{-p_n} C(\boldsymbol{\Sigma}), \tag{1.17}$$

where $1 \leq p \leq \infty$ and $\boldsymbol{\mu} : \boldsymbol{\Gamma} \to \mathbb{R}$ is a probability measure. This is a trivial consequence of the fact that the last right-hand side is independent of $\boldsymbol{y}$.

Remark 1.9. *In case $\boldsymbol{\rho}$ and $C(\boldsymbol{\Sigma})$ actually depended on the point $\boldsymbol{y} \in \boldsymbol{\Gamma}$ where* (1.16) *is computed and there is no uniform bound from below (respectively above) of $\boldsymbol{\rho}(\boldsymbol{y})$ or $C(\boldsymbol{\Sigma}, \boldsymbol{y})$, an estimate of the form* (1.17) *is still possible: Assume that $\boldsymbol{y} \mapsto \prod_{i=1}^{n} \alpha_n(\boldsymbol{y})^{-p_n} C(\boldsymbol{\Sigma}, \boldsymbol{y})$ is $L^p_{\boldsymbol{\mu}}(\boldsymbol{\Gamma})$-integrable, where $\alpha_n(\boldsymbol{y}) \in \mathbb{R}$ is such that $\rho_n(\boldsymbol{y}) = \alpha_n(\boldsymbol{y})\hat{\rho}_n$ for a constant (i.e. independent of $\boldsymbol{y}$) sequence $\hat{\boldsymbol{\rho}} \in \mathbb{R}^{\mathbb{N}}_{>0}$. Then,*

$$\left\|\partial_{y_1}^{p_1} \cdots\cdots \partial_{y_n}^{p_n} u\right\|_{L^p_{\boldsymbol{\mu}}(\boldsymbol{\Gamma},\mathbb{U})} \leq \boldsymbol{p}! \prod_{i=1}^{n} \hat{\rho}_n^{-p_n} \int_{\boldsymbol{\Gamma}} \prod_{i=1}^{n} \alpha_n(\boldsymbol{y})^{-p_n} C(\boldsymbol{\Sigma}, \boldsymbol{y}) d\boldsymbol{\mu}(\boldsymbol{y}).$$

This argument is analogous to the one mentioned in Example 1.1 for the well-posedness of the Poisson problem with non-uniformly bounded random diffusion.

The holomorphic regularity of parameter-to-solution maps arising from parametric PDEs has already been extensively studied. In the seminal paper [BNT10, Section 3], bounded and unbounded parameter spaces under the finite dimensional noise assumption are considered. in [CDS11], the authors develop a theory for countably-many parameters taking values on tensor product of bounded. In [BCDM17], the authors extend the theory to the Poisson problem with lognormal coefficients, again with countably many unbounded parameters. Finally in [CCS15], the authors go beyond the linear setting and prove holomorphy for a wider class of operators employing the implicit function theorem.

1.1.5 Reduction of a SPDE to a random coefficient PDE

The present section is based on [ADF+24, Section 2.1].

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 $D_T \coloneqq [0, T] \times D$. Consider the initial condition $U^0 : D \to \mathbb{R}^m$ for $m \in \mathbb{N}$, a drift coefficient $\mathfrak{D} : \mathbb{R}^m \times D_T \to \mathbb{R}^m$ and a noise coefficient $\mathfrak{N} : \mathbb{R}^m \times D \to \mathbb{R}^m$. 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 $(\Omega, \mathcal{E}, \mathcal{P})$, we consider the SPDE problem:
Find a random field $U : \Omega \times D_T \to \mathbb{R}^m$ such that, a.e. in Ω,

$$\begin{cases} \mathrm{d}U = \mathfrak{D}(U, \cdot, \cdot)\mathrm{d}t + \mathfrak{N}(U, \cdot) \circ \mathrm{d}W & \text{on } D_T \\ \partial_n U = 0 & \text{on } [0, T] \times \partial D \\ U(\cdot, 0, \cdot) = U^0 & \text{on } D, \end{cases} \tag{1.18}$$

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

The *Doss-Sussmann transform* [Dos77, Sus78] of U is, by definition,

$$u : \Omega \times D_T \to \mathbb{R}^m, \qquad u = e^{-W\mathfrak{N}} U, \tag{1.19}$$

i.e. the exponential of the operator $-W\mathfrak{N}$ applied to U. Let us now show that the random field u solves a random coefficient PDE, a problem class we already discussed in Section 1.1.1.

We derive the random coefficient PDE formally and refer to the applications in Example 1.10 or Section 3.2.1 below for rigorous derivations in specific settings. First, invert (1.19) to obtain $U = e^{W\mathfrak{N}}u$. Then, by the Stratonovich chain rule Lemma C.3 (formally identical to the deterministic one), we get

$$\mathrm{d}U = e^{W\mathfrak{N}}\left[\mathrm{d}u\right] + \mathfrak{N}\left[e^{W\mathfrak{N}}u\right] \circ \mathrm{d}W.$$

Equation (1.18) yields

$$e^{W\mathfrak{N}}\left[\mathrm{d}u\right] + \mathfrak{N}(e^{W\mathfrak{N}}u, \cdot) \circ \mathrm{d}W = \mathfrak{D}(U, \cdot, \cdot)\mathrm{d}t + \mathfrak{N}(U, \cdot) \circ \mathrm{d}W.$$

The term with the stochastic differential vanishes and we obtain the following random coefficient PDE driven by the Wiener process:

$$\partial_t u = e^{-W\mathfrak{N}}\mathfrak{D}(e^{W\mathfrak{N}}u, \cdot, \cdot) \qquad \text{on } D_T, \text{a.e. in } \Omega.$$

This is clearly a special case of the random coefficient PDE we considered in Section 1.1.1:

$$\mathcal{R}(W(\omega), u(\omega)) = 0 \qquad \text{in } \mathbb{F}, \text{for a.e. } \omega \in \Omega,$$

where $\mathcal{R} : \mathbb{A} \times \mathbb{U} \to \mathbb{F}$ denotes the residual operator of the PDE problem defined for appropriate Banach spaces $\mathbb{A}$, $\mathbb{U}$, $\mathbb{F}$ for coefficients, solutions and residuals sample paths respectively. In the present setting, $\mathcal{R}$ is a differential operator in time and space with respect to $u \in \mathbb{U}$ while it does not contain Itô or Stratonovich differentials of the coefficient (a Wiener process) $W \in \mathbb{A}$.

Given an approximation $u_{\text{approx}} : \Omega \times D_T \to \mathbb{R}^m$ of u, the Doss-Sussmann transform can be inverted to yield an approximation U_{approx} of the random field solution U of (1.18) as

$$U_{\text{approx}} = e^{W\mathfrak{N}}u_{\text{approx}}.$$

Fixed $1 \leq p \leq \infty$, the $L^p(\Omega, \mathbb{U})$-error on the solution to the SPDE is estimated by the $L^p(\Omega, \mathbb{U})$-error on the solution to the random coefficient PDE as long as $e^{W\mathfrak{N}}$ is a linear operator on $L^p(\Omega, \mathbb{U})$ and into $L^p(\Omega, \mathbb{U})$. We have

$$\|U - U_{\text{approx}}\|_{L^p(\Omega,\mathbb{U})} = \left\|e^{W\mathfrak{N}}\left[u - u_{\text{approx}}\right]\right\|_{L^p(\Omega,\mathbb{U})} \leq \left\|e^{W\mathfrak{N}}\right\|_{\mathcal{L}(L^p(\Omega,\mathbb{U}))} \|u - u_{\text{approx}}\|_{L^p(\Omega,\mathbb{U})},$$

where, for a Banach space B, by $\mathcal{L}(B)$ we denote the Banach space of linear bounded operators $L : B \to B$. For example, a sufficient condition for $e^{W\mathfrak{N}}$ to have finite $\mathcal{L}(L^p(\Omega, \mathbb{U}))$-norm is the summability of the sequence $\left(\frac{\|W^n\mathfrak{N}^{(n)}\|_{\mathcal{L}(L^p(\Omega,\mathbb{U}))}}{n!}\right)_{n\in\mathbb{N}}$

Example 1.10 (Example of Doss-Sussmann transform with a simple SDE)**.** *Given a final time $T > 0$, appropriate coefficients a, b (either real constants, real functions on $[0, T]$, or stochastic processes on $[0, T]$), consider the following linear* stochastic differential equation *(SDE) with unknown stochastic process $U : \Omega \times [0, T] \to \mathbb{R}$:*

$$d_t U = aU\,dt + bU \circ dW, \tag{1.20}$$

with initial condition $U(0) = U_0 \in \mathbb{R}$ a.e. in Ω. The Doss-Sussmann transform of U and its inverse read respectively

$$u = e^{-Wb}U, \qquad U = e^{Wb}u.$$

The chain rule gives $d_t U = e^{Wb}\partial_t u + bU \circ d_t W$. *We obtain the following* random coefficient ordinary differential equation *(RODE) with unknown stochastic process* $u : \Omega \times [0,T] \to \mathbb{R}$*:*

$$\partial_t u = -e^{-Wb} a e^{Wb} u.$$

With the corresponding initial condition $u(0) = U_0$ *a.e. in* Ω. *The solution* u*, when it exists, is related to the solution* U *of* (1.20) *through the inverse Doss-Sussmann transform.*

As we will see below (cf. Section 1.3.6), the Doss-Sussmann transform was already applied to much more complex stochastic differential equations such as the the *stochastic LLG equation* (a model in micromagnetics discussed in Section 1.3) in [GLT16].

1.2 Numerical approximation of parametric coefficient PDEs

In the present section, we review the numerical approximation of parameter-to-solution maps arising from parametric coefficient PDEs. We start in Section 1.2.1 by highlighting the difficulties of this problem class, in particular the *curse of dimensionality*. Nevertheless, as shown in Section 1.2.2, a class of functions can be approximated with algebraic rates independently of the number of parameters, i.e. free from then curse of dimensionality. We present the truncated sparse Taylor expansion as a simple (but computationally impractical) example of such an approximation. We then introduce sparse grid interpolation (cf. Section 1.2.3) as a more practical sparse polynomial approximation. We give a detailed definition of the method and present some of its "flavors". We review the history of the use of sparse grid interpolation for parametric PDEs and we discuss the problem of selecting the sparse polynomial space through a "profit maximization" approach. In Section 1.2.4, we discuss adaptive sparse grid interpolation. We close the chapter with additional topics of interest: Alternative approximation methods (Section 1.2.5), and the role of the finite elements approximation (Section 1.2.6).

1.2.1 The problem of high-dimensional approximation

In Section 1.1.3, we have seen how the problem of approximating the solution of a random coefficient PDE (1.1) is equivalent to the approximation of a parameter-to-solution map $u : \boldsymbol{\Gamma} \to \mathbb{U}$ for a parametric coefficient PDE (1.12), where $\boldsymbol{\Gamma} \subset \mathbb{R}^{\mathbb{N}}$. Notice that the domain $\boldsymbol{\Gamma}$ is "high-dimensional" (in general, it can be a space of real sequences). We may attempt to circumvent this issue by considering the truncation $u_N : \boldsymbol{\Gamma}^N \to \mathbb{U}$ for some $N \in \mathbb{N}$ as in Remark 1.7 but in order for the truncation error to decrease, N may need to increase.

The term *curse of dimensionality*, introduced by Bellman [BK59], refers to the exponential dependence of the cost of approximating a function with fixed accuracy $\varepsilon > 0$ on the dimension $N \in \mathbb{N}$ of its domain. In other words, the complexity is $\mathcal{O}(\varepsilon^{-\alpha N})$ for a constant $\alpha > 0$ dependent on the function at hand.

The curse of dimensionality can appear in very simple problems as well, as we see in the following example.

Example 1.11. *Consider the sequence* $\boldsymbol{\alpha} = (\alpha_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{\geq 0}$*, and the function*

$$f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}, \qquad \boldsymbol{y} \to \sin\left(\sum_{n=1}^{N} \alpha_n y_n\right).$$

Imagine now that its law is hidden and that we want to build an approximation based on collocation samples. Depending on the values of $\boldsymbol{\alpha}$*, the function is more or less easy to approximate. For example:*

- $\alpha_n \equiv 0$ for all $n > N$ "small" (e.g. $N = 3$)*: This is a classical low dimensional approximation problem that can be tackled e.g. with tensor product interpolation;*
- $\alpha_n \geq \alpha_0 > 0$ for all $n > N$ and $N \in \mathbb{N}$*: The problem suffers from the curse of dimensionality because, intuitively, each of the infinitely-many variable y_n has some importance;*
- $\alpha_n = \mathcal{O}(n^{-(1+\varepsilon)})$, $\varepsilon > 0$*: Again the domain is infinite-dimensional, but this time the variable y_n becomes less and less "important" at a sufficiently fast rate as $n \to \infty$ because of the corresponding decay of α_n.*

Knowing that a function *can* be approximated (with respect to a certain norm), does not automatically give a construction of an effective approximation scheme. In the next sections, we mention several options and focus on sparse grid interpolation.

Moreover, we stress that, especially in high-dimension, approximation is strictly harder than computing scalar quantities of interest such as the expectation or the probability of an event.

1.2.2 Dimension-independent approximation of sparse maps

In this section, we show how, despite the fact that the parametric dimension may be high or even infinite, it is still possible to approximate parametric problems with an algebraic rate of convergence and, most importantly, independently of the dimension of the parameter domain. This discussion is based on [CDS11].

Recall the residual function $\mathcal{R} : \mathbb{A} \times \mathbb{U} \to \mathbb{F}$ we introduced in Section 1.1.1, where $\mathbb{A}, \mathbb{U}$ and $\mathbb{F}$ are Banach spaces for parameters, solutions, and residuals respectively. The problem consists of approximating the *solution operator* $\mathcal{U} : \mathbb{A} \to \mathbb{U}$ defined by the relation: For $a \in \mathbb{A}$, $\mathcal{U}(a) \in \mathbb{U}$ is such that

$$\mathcal{R}(a, \mathcal{U}(a)) = 0 \qquad \text{in } \mathbb{F}. \tag{1.21}$$

Well-definiteness and appropriate continuity of $\mathcal{U}$ are linked to the well-posedness of the problem (1.21).

The space of coefficients $\mathbb{A}$ is assumed to be *parametrizable*, i.e. that there exists a set of sequences $\boldsymbol{\Gamma} \subset \mathbb{R}^{\mathbb{N}}$ and a *sufficiently regular* map $a : \boldsymbol{\Gamma} \to \mathbb{A}$ such that $\mathbb{A} = \{a(\boldsymbol{y}) : \boldsymbol{y} \in \boldsymbol{\Gamma}\}$. The map

$$u = \mathcal{U} \circ a : \boldsymbol{\Gamma} \to \mathbb{U}$$

is called the *parameter-to-solution map* and may be expressed, analogously to the solution operator, as follows: For any $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $u(\boldsymbol{y})$ is the unique element in $\mathbb{U}$ such that $\mathcal{R}(a(\boldsymbol{y}), u(\boldsymbol{y})) = 0$ in $\mathbb{F}$.

Let us now follow [CDS11] to give an example of approximation that converges algebraically and does not suffer form the curse of dimensionality. The key requirement is that the parameter-to-solution map is sufficiently regular in a specific sense.

Consider the parameter set $\boldsymbol{\Gamma} = [-1, 1]^{\mathbb{N}}$, i.e. sequences taking values in $[-1, 1]$. Consider the set of *finite-support multi-indices*:

$$\mathcal{F} \coloneqq \left\{ \boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} : \exists N = N(\boldsymbol{\nu}) : \nu_n = 0 \quad \forall n > N \right\}. \tag{1.22}$$

For any $\boldsymbol{\nu} \in \mathcal{F}$, define the *support of* $\boldsymbol{\nu}$ as

$$\operatorname{supp}(\boldsymbol{\nu}) \coloneqq \{n \in \mathbb{N} : \nu_n \neq 0\}. \tag{1.23}$$

Denote respectively the factorial (under the usual assumption that $0! = 1$) and the monomial

$$\boldsymbol{\nu}! \coloneqq \prod_{i\in\mathbb{N}} \nu_i!, \qquad \boldsymbol{y}^{\boldsymbol{\nu}} \coloneqq \prod_{i\in\mathbb{N}} y_i^{\nu_i} \qquad \forall \boldsymbol{y} \in \boldsymbol{\Gamma},$$

where both products are finite since $\operatorname{supp}(\boldsymbol{\nu})$ is finite for any $\boldsymbol{\nu} \in \mathcal{F}$. Finally, define the mixed derivative operator $\partial^{\boldsymbol{\nu}} = \partial^{\nu_1} \dots \partial^{\nu_n}$ where $\{\nu_1, \dots, \nu_n\} = \operatorname{supp}(\boldsymbol{\nu})$. The *Taylor expansion of* $u : \boldsymbol{\Gamma} \to \mathbb{U}$ is

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}} t_\nu \boldsymbol{y}^{\boldsymbol{\nu}}, \qquad \text{where } t_{\boldsymbol{\nu}} \coloneqq \frac{\partial^{\boldsymbol{\nu}} u(\mathbf{0})}{\boldsymbol{\nu}!}. \tag{1.24}$$

If there exists $0 < p < 1$ such that

$$(\|t_{\boldsymbol{\nu}}\|_{\mathbb{U}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F}), \tag{1.25}$$

then the series (1.24) converges *unconditionally* in $L^\infty(\boldsymbol{\Gamma}; U)$. By "unconditionally", we mean that for *any* sequence $(\Lambda_n)_{n\in\mathbb{N}}$ that exhausts $\mathcal{F}$ (i.e. for any finite $\Lambda \subset \mathbb{N}_0^N$ there exists $N_0 \in \mathbb{N}$ such that $\Lambda \subset \Lambda_N$ for all $N \geq N_0$), there holds $\lim_{n\to\infty} \left\| u - \sum_{\boldsymbol{\nu}\in\Lambda_n} t_\nu \boldsymbol{y}^{\boldsymbol{\nu}} \right\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} = 0$.

The *best-n-terms truncated Taylor series* is by definition

$$T_n[u](\boldsymbol{y}) \coloneqq \sum_{\boldsymbol{\nu}\in\Lambda_n} t_\nu \boldsymbol{y}^{\boldsymbol{\nu}},$$

where $\Lambda_n \subset \mathcal{F}$ is the set of multi-indices corresponding to the n largest $\|t_{\boldsymbol{\nu}}\|_U$.

The approximation error can be estimated with Stechkin's lemma (see Corollary D.2): If $(\|t_{\boldsymbol{\nu}}\|_{\mathbb{U}})_{\boldsymbol{\nu}\in\mathcal{F}} \in \ell^p(\mathcal{F})$ for $0 < p < 1$, then

$$\|u - T_n[u]\|_{L^\infty(\boldsymbol{\Gamma};\mathbb{U})} = \left\| \sum_{\boldsymbol{\nu}\notin\Lambda_n} t_{\boldsymbol{\nu}} \boldsymbol{y}^{\boldsymbol{\nu}} \right\|_{L^\infty(\boldsymbol{\Gamma};U)} \leq \left\| (\|t_{\boldsymbol{\nu}}\|_U)_{\boldsymbol{\nu}\in\mathcal{F}} \right\|_{\ell^p(\mathcal{F})} (n+1)^{1-\frac{1}{p}}.$$

This is an example of how dimension independent, algebraic convergence can be achieved for a "regular" parameter-to-solution map. In the present case, the specific regularity requirement is ℓ^p- summability of the Taylor coefficients (1.25) for some $0 < p < 1$.

The truncated Taylor expansion is not a very practical approximation method because approximating the derivatives $\partial^{\boldsymbol{\nu}} u$ needed to determine $t_{\boldsymbol{\nu}}$ is not always an easy task in general.

1.2.3 Sparse grid interpolation

In this section, we present sparse grid interpolation. Like the truncated Taylor expansion, sparse grid interpolation produces a sparse polynomial approximation. An important difference is that sparse grid interpolation coefficients are simply collocation samples, much easier to approximate that Taylor coefficients for general parameter-to-solution maps.

A possible first appearance of sparse grid quadrature and interpolation dates back to the 1960s, in a work by Smolyak [Smo63]. After that, the method was rediscovered in different settings (e.g. numerical approximation of high-dimensional PDEs) several times over in the 1990s, when the method gained its current name, for example by Griebel and Zenger [GSZ90]. The main idea behind sparse grid methods is always to write the approximation operator as a telescoping expansion in a tensor product hierarchical basis. The result is a flexible method that can exploit the regularity of the function to accelerate convergence or, in a high-dimensional

setting, circumvent the curse of dimensionality. For a more complete historical note on sparse grids and related concepts, we refer to the next subsection.

Consider $f : \mathbf{\Gamma} \to \mathbb{U}$, defined on $\mathbf{\Gamma} \subset \mathbb{R}^{\mathbb{N}}$ (a set of real valued sequences) and into a Banach space $\mathbb{U}$. The function f is assumed to be continuous, in the sense that for any $\boldsymbol{y} \in \mathbf{\Gamma}$ $f(\boldsymbol{y})$ is a well-defined accessible element of $\mathbb{U}$.

Let us begin with the case in which $y_2 = y_3 = \cdots = 0$ for any $\boldsymbol{y} \in \mathbf{\Gamma}$, i.e. we can assume $\Gamma \coloneqq \mathbf{\Gamma} \subset \mathbb{R}$. A sequence $(\mathcal{Y}_\nu)_{\nu \in \mathbb{N}_0} \subset \Gamma$, is called a *nodes family of* Γ if

- $\mathcal{Y}_0 = \{0\}$;
- The sequence in non-decreasing, in the sense that $\#\mathcal{Y}_\nu \leq \#\mathcal{Y}_{\nu+1}$ for all $\nu \in \mathbb{N}_0$.

We denote by $m : \mathbb{N}_0 \to \mathbb{N}$ the *level-to-knot function* associated to the nodes family $(\mathcal{Y}_\nu)_\nu$, i.e $m(\nu) = \#\mathcal{Y}_\nu$ for all $\nu \in \mathbb{N}_0$.

Consider a corresponding family of *Lagrange basis*, i.e. for any $\nu \in \mathbb{N}_0$ a basis of function $\{L_y\}_{y \in \mathcal{Y}_\nu}$ such that $L_{y_1}(y_2) = \delta_{y_1, y_2}$ for any $y_1, y_2 \in \mathcal{Y}_\nu$. Denote by V_ν the linear space of functions generated by $(L_y)_{y \in \mathcal{Y}_\nu}$. Finally, define the corresponding *interpolation operator* $I_\nu : C^0(\Gamma) \to V_\nu$ as

$$I_\nu[f](\cdot) = \sum_{y \in \mathcal{Y}_\nu} f(y) L_y(\cdot) \qquad \forall f \in C^0(\Gamma). \tag{1.26}$$

This operator can be written as a *telescoping sum* $I_\nu = \sum_{0 \leq j \leq \nu} \Delta_\nu$, where the *difference* $\Delta_j : C^0(\Gamma) \to V_j$ is defined as

$$\begin{aligned} &\Delta_j = I_j - I_{j-1} \qquad \text{if } \nu > 0, \\ &\Delta_0 = I_0. \end{aligned} \tag{1.27}$$

Let us consider now the general case in which $\mathbf{\Gamma} \subset \mathbb{R}^{\mathbb{N}}$ cannot be identified with a subset of $\mathbb{R}$. For any $\boldsymbol{\nu} \in \mathcal{F}$, we define the *tensor product nodes family*

$$\mathcal{Y}_{\boldsymbol{\nu}} \coloneqq \{\boldsymbol{y} \in \mathbf{\Gamma} : y_i \in \mathcal{Y}_{\nu_i} \ \forall i \in \mathbb{N}\}.$$

The corresponding tensor product linear space

$$V_{\boldsymbol{\nu}} = \bigotimes_{i \in \boldsymbol{\nu}} V_i \tag{1.28}$$

is the linear space generated by the tensor product Lagrange basis $\{L_{\boldsymbol{y}}\}_{\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{\nu}}}$, where $L_{\boldsymbol{y}}(\boldsymbol{x}) = \prod_{i \in \mathbb{N}} L_{y_i}(x_i)$ for all $\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{\nu}}$ and $\boldsymbol{x} \in \mathbf{\Gamma}$. The corresponding *tensor product interpolant* $I_{\boldsymbol{\nu}} : C^0(\mathbf{\Gamma}) \to V_{\boldsymbol{\nu}}$ is defined by

$$I_{\boldsymbol{\nu}}[f](\cdot) = \sum_{\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{\nu}}} f(\boldsymbol{y}) L_{\boldsymbol{y}}(\cdot) \qquad \forall f \in C^0(\mathbf{\Gamma}). \tag{1.29}$$

The *hierarchical surplus* $\Delta_{\boldsymbol{\nu}} : C^0(\mathbf{\Gamma}) \to V_{\boldsymbol{\nu}}$ is defined as tensor product of 1D difference operators:

$$\Delta_{\boldsymbol{\nu}} = \bigotimes_{\nu_i \in \boldsymbol{\nu}} \Delta_{\nu_i}. \tag{1.30}$$

The hierarchical surplus operator can be evaluated recursively: Denote $\boldsymbol{\nu} = (\nu_1, \hat{\boldsymbol{\nu}})$ and $\boldsymbol{y} = (y_1, \hat{\boldsymbol{y}})$ with $\nu_1 \in \mathbb{N}_0$, $y_1 \in \Gamma$ and use the definition of difference (1.27) to evaluate

$$\Delta_{\boldsymbol{\nu}}[f](\boldsymbol{y}) = \Delta_{\nu_1}\left[\Delta_{\hat{\boldsymbol{\nu}}}[f(\cdot, \hat{\boldsymbol{y}})]\right](y_1).$$

Consider a *multi-index set* $\Lambda \subset \mathcal{F}$ that is *downward-closed*, i.e.

$$\forall \boldsymbol{\nu} \in \Lambda,\ \exists n \in \operatorname{supp}(\boldsymbol{\nu})\ \ \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda \text{ or } \boldsymbol{\nu} = \mathbf{0}.$$

The *sparse grid space* associated to the downward-closed multi-index set Λ and the 1D-interpolation method $(I_\nu)_{\nu \in \mathbb{N}_0}$ is

$$V_\Lambda \coloneqq \bigoplus_{\boldsymbol{\nu} \in \Lambda} V_{\boldsymbol{\nu}} = \left\{ \sum_{\nu \in \Lambda} \boldsymbol{f}_{\boldsymbol{\nu}} : \boldsymbol{f}_{\boldsymbol{\nu}} \in V_{\boldsymbol{\nu}} \right\} \tag{1.31}$$

and the corresponding *sparse grid interpolant* is defined by

$$\mathcal{I}_\Lambda : C^0(\boldsymbol{\Gamma}) \to V_\Lambda, \qquad \mathcal{I}_\Lambda \coloneqq \sum_{\boldsymbol{\nu} \in \Lambda} \Delta_{\boldsymbol{\nu}}. \tag{1.32}$$

Let us list the most important properties of sparse grid interpolation:

- There exists a *sparse grid*, i.e. a set $\mathcal{H}_\Lambda \subset \bigcup_{\boldsymbol{\nu} \in \Lambda} \mathcal{Y}_{\boldsymbol{\nu}} \subset \boldsymbol{\Gamma}$ such that each function $\boldsymbol{f} \in V_\Lambda$ is uniquely defined by the values $\{\boldsymbol{f}(\boldsymbol{y}) : \boldsymbol{y} \in \mathcal{H}_\Lambda\}$, i.e. $\mathcal{H}_\Lambda$ is *unisolvent* for V_Λ [CCS14, Theorem 2.1];
- As a consequence, there exists a Lagrange basis $(L_{\boldsymbol{y}})_{\boldsymbol{y} \in \mathcal{H}_\Lambda}$ for V_Λ and the sparse grid interpolant can be written as
$$\mathcal{I}_\Lambda[f] = \sum_{\boldsymbol{y} \in \mathcal{H}_\Lambda} f(\boldsymbol{y}) L_{\boldsymbol{y}};$$
- *Inclusion-exclusion formula*: The sparse grid interpolant can be written as a linear combination of tensor product interpolants:
$$\mathcal{I}_\Lambda = \sum_{\boldsymbol{\nu} \in \Lambda} \alpha_{\boldsymbol{\nu}} I_{\boldsymbol{\nu}}, \qquad \text{where } \alpha_{\boldsymbol{\nu}} \coloneqq \sum_{\substack{\boldsymbol{i} \in \{0,1\}^{\mathbb{N}}: \\ \boldsymbol{\nu} + \boldsymbol{i} \in \Lambda}} (-1)^{|\boldsymbol{i}|_1}. \tag{1.33}$$
To prove this, it is sufficient to apply the definition of hierarchical surplus (1.30) and the multi-linearity of the tensor product.
- Assume that I_ν is exact on the linear space V_ν for all $\nu \in \mathbb{N}_0$ and that $V_0 \subset V_1 \subset \ldots$ Then, for a downward-closed $\Lambda \subset \mathcal{F}$, $\mathcal{I}_\Lambda$ is exact on the linear space V_Λ (cf. Definition 1.31).
- The *Lebesgue constant* of the sparse grid interpolant $\mathcal{I}_\Lambda$ is defined as
$$\mathbb{L}_\Lambda \coloneqq \sup_{\substack{f \in C^0(\boldsymbol{\Gamma}) \\ \|f\|_{C^0(\boldsymbol{\Gamma})} = 1}} \|\mathcal{I}_\Lambda[f]\|_{C^0(\boldsymbol{\Gamma})}. \tag{1.34}$$
The relevance of the Lebesgue constant in uniform approximation is that it allows us to relate the interpolation error to the best uniform approximation error in V_Λ:
$$\|f - \mathcal{I}_\Lambda[f]\|_{C^0(\boldsymbol{\Gamma})} \leq (1 + \mathbb{L}_\Lambda) \inf_{g \in V_\Lambda} \|f - g\|_{C^0(\boldsymbol{\Gamma})}.$$
This is a consequence of the fact that $\mathcal{I}_\Lambda$ is exact on V_Λ (cf. previous point). It can be proved (see [CCS14, Lemma 3.1]) that, if the Lebesgue constant of the 1D interpolant I_ν satisfies:
$$\exists \theta \geq 1\ :\ \forall \nu \in \mathbb{N}_0\ \lambda_\nu \coloneqq \|I_\nu\|_{\mathcal{L}(C^0(\boldsymbol{\Gamma}))} \leq (1 + \nu)^\theta,$$
then for the sparse grid interpolant defined with I_ν there holds
$$\mathbb{L}_\Lambda \leq (1 + \#\Lambda)^{\theta+1} \qquad \forall \Lambda \subset \mathcal{F} \text{ downward-closed};$$

- A sparse grid interpolation method (or the corresponding family of sparse grids) is called *nested* if, for downward-closed multi-index sets$\Lambda_1, \Lambda_2 \subset \mathcal{F}$,

$$\Lambda_1 \subset \Lambda_2 \Rightarrow \mathcal{H}_{\Lambda_1} \subset \mathcal{H}_{\Lambda_2}.$$

  We see in Section 1.2.4 that nested sparse grid interpolation methods are particularly suited for *adaptive sparse grid refinement*.

Remark 1.12 (Sparse grid interpolation *method* vs sparse grid *interpolant*)**.** *In this section, we showed how a 1D interpolation method can be "upgraded" to an arbitrary number of dimensions through the sparse grid construction. Let us be more precise on the vocabulary we use in the rest of this work:*

- *A sparse grid interpolation* method *is uniquely defined by a 1D nodes family* $(\mathcal{Y}_\nu)_{\nu\in\mathbb{N}_0}$ *(that satisfied the assumptions listed in the beginning of this section) and a linear space* V_ν *of functions uniquely determined by their values on* $\#\mathcal{Y}_\nu$ *(usually piecewise or global polynomials as seen above);*

- *A sparse grid* interpolant *is an instance of a sparse grid interpolation method once a downward-closed multi-index set* Λ *is selected. It is an* operator $\mathcal{I}_\Lambda : C^0(\mathbb{R}^{\mathbb{N}}) \to V_\Lambda$*, where* V_Λ *is defined in* (1.31) *based on the 1D space* V_ν *through* (1.28)*.*

History of sparse grids

The present short historical note is partially inspired by [BG04] and [Du23].

Already in the 1930s, Sobol studied functions $f : [-1,1]^N \to \mathbb{R}$ for $N \in \mathbb{N}$ and under the regularity condition

$$\|\partial^{\boldsymbol{a}} f\|_{L^p([-1,1]^N)} \leq 1,$$

where $\boldsymbol{a} = (a_1 \dots, a_N)$ is such that $|\boldsymbol{a}|_1 \leq A \in \mathbb{N}_0$. He obtained quadrature error estimates for $m \in \mathbb{N}$ nodes formulas of order $m^{-A/N}$.

In the end of the 1950s Korobov [Kor59] considered the class of functions satisfying the previous condition for all mixed derivatives corresponding to multi-indices with uniformly bounded ℓ^1 norm, i.e.

$$\left\{ f : [-1,1]^N \to \mathbb{R} : \|f^{\boldsymbol{a}}\|_{L^p([-1,1]^N)} < 1 \qquad \forall \boldsymbol{a} = (a_1, \dots, a_N) \in \mathbb{N}_0^N : |\boldsymbol{a}|_1 \leq A \right\},$$

again for a given $A \in \mathbb{N}_0$. Korobov constructed a cubature method with accuracy $m^{-A} \log(m)^{AN}$ with $m \in \mathbb{N}$ nodes.

Babenko (cf. [Bab60]) introduced in the 1960s the class of *dominating mixed derivative*

$$W^{r,p} = \left\{ f \in L^p([-1,1]^N) : \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^N : |\boldsymbol{\alpha}|_1 \leq r} \|\partial^{\boldsymbol{\alpha}} f\|_{L^p([-1,1]^N)} \leq 1 \right\}$$

and studied their approximation with hyperbolic cross polynomials.

Sparse grid interpolation was first considered by Smolyak in the 1960s [Smo63]. Here, the author considers only *isotropic* sparse grids, which is an improvement over classical tensor product interpolation. In this work, the author already identified the now classical interpolation error bound $\log(N)^\beta (\#\mathcal{H})^{-\alpha}$ with $\#\mathcal{H}$ collocation nodes in $N \in \mathbb{N}$ dimensions.

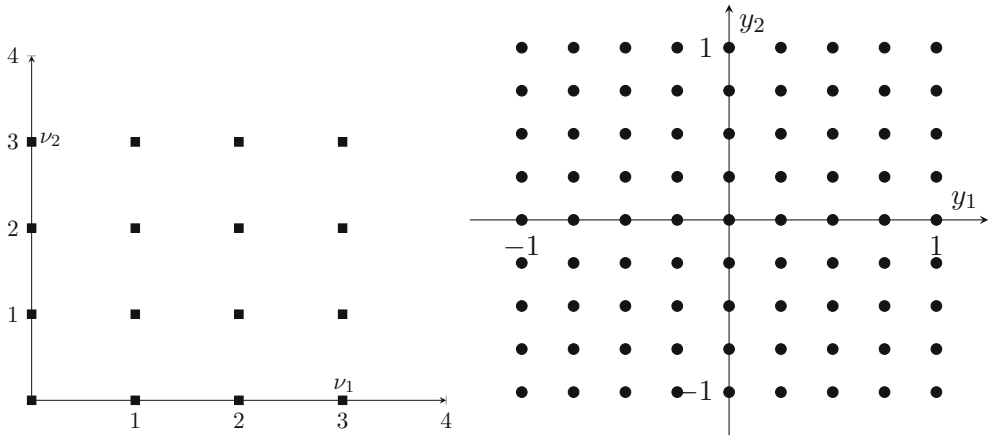

Figure 1.5: Left: Tensor product multi-index set $\Lambda_{\mathrm{TP}}(3)$. Right: Corresponding collocation nodes based on equispaced 1D nodes on $[-1, 1]$.

Sparse grid interpolation is only an instance of the more general technique introduced in [Zen91] that can combine a much wider class of operators than 1D interpolation, such as discretization in time and space (e.g. finite element or finite difference methods).

Multilevel methods, such as multilevel Monte Carlo [Gil15], can also be understood as simple sparse grid methods over two discretization schemes indexed by a scalar parameter. For example, for the multilevel Monte Carlo method applied to a random coefficient elliptic problem, the two schemes are: Monte Carlo quadrature (indexed by the number of sample); and the space/time discretization (indexed by the number of mesh elements and/or timesteps).

In the review paper [BG04] a thorough description of the state of the art at the time of the publication is given.

Examples of sparse grid interpolation methods

Let us give some examples of sparse grid interpolation methods.

Example 1.13 (Tensor product interpolation)**.** *Consider, given* $\nu \in \mathbb{N}_0$*, the* level-to-knot function

$$m(\nu) = \begin{cases} 1 & \text{if } \nu = 0, \\ 2^\nu + 1 & \text{if } \nu \in \mathbb{N} \end{cases}, \tag{1.35}$$

often called the doubling rule*. Consider the family of* $m(\nu)$ *equispaced nodes on* $[-1, 1]$*:*

$$\begin{aligned} \mathcal{Y}_0 &= \{0\}, \\ \mathcal{Y}_\nu &= \left\{ -1 + \frac{2(i-1)}{m(\nu)-1} : i = 1, \dots, m(\nu) \right\} \qquad \forall \nu \in \mathbb{N}. \end{aligned} \tag{1.36}$$

Observe that this nodes family is nested. tensor product interpolation on $[-1, 1]^N$*, for* $N \in \mathbb{N}$ *can be understood as a sparse grid interpolation scheme with multi-index set*

$$\Lambda_{TP}(w) = \left\{ \boldsymbol{\nu} \in \mathbb{N}_0^N : \max_{n=1,\dots,N} \nu_n \leq w \right\} \qquad \forall w \in \mathbb{N}_0.$$

See Figure 1.5 for an example of tensor product multi-index set and corresponding sparse grid.

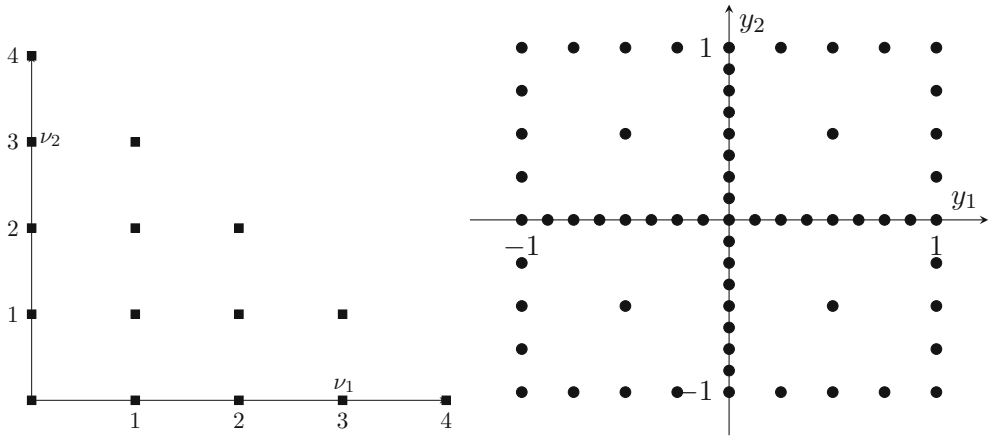

Figure 1.6: Left: Total degree multi-index set $\Lambda_{\mathrm{TD}}(4)$. Right: Corresponding collocation nodes based on equispaced 1D nodes.

If $u \in C^2([-1,1]^N)$, we can estimate the error using Taylor's theorem. We obtain:

$$\left\| u - \mathcal{I}_{\Lambda_{TP(w)}} u \right\|_{L^\infty([-1,1]^N)} \leq \frac{N}{2} 2^{-2\nu} \left\| u \right\|_{C^2([-1,1]^N)} \quad \forall w \geq 0.$$

Observing that the method uses $m(w)^N$ nodes, the error estimate with respect to the number of nodes reads:

$$\left\| u - \mathcal{I}_{\Lambda_{TP(w)}} u \right\|_{L^\infty([-1,1]^N)} \leq 2N \left(\#\mathcal{H}_{\Lambda_{TP(w)}} \right)^{-\frac{2}{N}} \left\| u \right\|_{C^2([-1,1]^N)} \quad \forall w \geq 0,$$

The convergence is algebraic with respect to the number of collocation nodes, but quickly degrades as $N \to \infty$.

Example 1.14 (Total degree)**.** *Consider again the 1D equispaced nodes* (1.36) *and the doubling rule* (1.35) *as level-to-knot function. This time, we adopt the $N \in \mathbb{N}$ dimensional* total degree *multi-index set*

$$\Lambda_{TD}(w) = \left\{ \boldsymbol{\nu} \in \mathbb{N}_0^N : \sum_{n=1}^{N} \nu_i \leq w \right\} \qquad \forall w \geq 0. \tag{1.37}$$

See Figure 1.6 for an example of multi-index set and corresponding sparse grid. For this choice of multi-index set, (1.33) *yields the following simplified formula for $w \geq N$:*

$$\mathcal{I}_{\Lambda_{TD}(w)} = \sum_{w-N+1 \leq |\boldsymbol{\nu}|_1 \leq w} (-1)^{w-|\boldsymbol{\nu}|_1} \binom{N-1}{w-|\boldsymbol{\nu}|_1} I_{\boldsymbol{\nu}}.$$

The corresponding sparse grid is $\bigcup_{w-N+1 \leq |\boldsymbol{\nu}|_1 \leq w} \mathcal{Y}_{\boldsymbol{\nu}}$.

This choice of sparse grid parameters was studied e.g. in [BG04]. Consider, for $q \in \{2, \infty\}$ and $r \in \mathbb{N}$, the space

$$W_{mix}^{q,r}([-1,1]^N) := \left\{ f : [-1,1]^N \to \mathbb{R} : \partial^{\boldsymbol{a}} u \in L^q([-1,1]^N) \ \forall \boldsymbol{a} \in \mathbb{N}_0^N : |\boldsymbol{a}|_\infty < r \right\}.$$

The sparse grid interpolation error is bounded as (see [BG04, Lemma 3.13])

$$\left\| u - \mathcal{I}_{\Lambda_{TD}(w)} u \right\|_{L^2([-1,1]^N)} \leq C \log_2 \left(\#\mathcal{H}_{\Lambda_{TD}(w)} \right)^{3(N-1)} \left(\#\mathcal{H}_{\Lambda_{TD}(w)} \right)^{-2} |u|_{W_{mix}^{2,2}([-1,1]^N)},$$

$$\left\| u - \mathcal{I}_{\Lambda_{TD}(w)} u \right\|_{L^\infty([-1,1]^N)} \leq C \log_2 \left(\#\mathcal{H}_{\Lambda_{TD}(w)} \right)^{3(N-1)} \left(\#\mathcal{H}_{\Lambda_{TD}(w)} \right)^{-2} |u|_{W_{mix,}^{2,\infty}([-1,1]^N)}$$

where the constant $C > 0$ is independent of N or Λ. Thus, convergence is algebraic in the number of collocation nodes but degrades as N increases (although not as quickly as in the case of tensor product interpolation).

Example 1.15 (Total degree with Clenshaw-Curtis nodes)**.** *For $\nu \in \mathbb{N}_0$, the set of $m(\nu)$ (defined with the doubling rule* (1.35)*) Clenshaw-Curtis nodes is given by the extrema of the degree $m(\nu)-1$ Chebyshev polynomial, i.e.*

$$\begin{aligned} \mathcal{Y}_0 &= \{1\} \\ \mathcal{Y}_\nu &= \left\{ -\cos\left(\frac{\pi(i-1)}{m(\nu)-1}\right) : i = 1, \dots, m(\nu) \right\} \qquad \forall \nu \in \mathbb{N}. \end{aligned} \tag{1.38}$$

This nodes family is again nested. Observe that these nodes differ from Chebyshev nodes, which are roots *of Chebyshev polynomials. For $N \in \mathbb{N}$, we consider the total degree multi-index set $\Lambda_{TD}(w)$* (1.37) *for $w \geq 0$.*

In [NTW08b], the sparse grid interpolation scheme based on these parameters is used to approximate maps $u : [-1,1]^N \to \mathbb{U}$ into a Banach space of functions that admit a holomorphic extension to $\Sigma_\tau^N = \bigotimes_{n=1}^N \Sigma_\tau$, where $\Sigma_\tau := \left\{ z \in \mathbb{C} : \min_{x \in [-1,1]} |z - x| < \tau \right\}$ for some $\tau > 0$. An example of such map is given by the parameter-to-solution operator of the random diffusion Poisson problem for suitable choices of the random diffusion. The authors derive the following convergence estimate:

$$\left\| u - \mathcal{I}_{\Lambda_{TD}(w)} u \right\|_{C^0([-1,1]^N;\mathbb{U})} \leq C_1 \left(\# \mathcal{H}_{\Lambda_{TD}(w)} \right)^{-\mu_1},$$

where $\mu_1 = \frac{\sigma}{1+\log(2N)}$ and $\sigma = \log\left(\tau + \sqrt{1+\tau^2}\right) > 0$. $C_1 = C_1(\sigma, N)$ is independent of $\# \mathcal{H}_{\Lambda_{TD}(W)}$ and diverges as $N \to \infty$.

The follow-up work [NTW08a] investigates the more general case of a map $u : [-1,1]^N \to \mathbb{U}$ admitting a holomorphic extension to $\bigotimes_{n=1}^N \Sigma_{\tau_n} \subset \mathbb{C}^N$ for a vector $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N) \in \mathbb{R}_{>0}^N$. The regularity is allowed to differ from parameter to parameter and less regularity with respect to one parameters corresponds to a smaller size of the domain of extension in that direction (τ_n smaller). Consequently, the same direction requires more collocation nodes. The authors use again Clenshaw-Curtis nodes (1.38) *and the doubling rule* (1.35)*. However this time they employ an* anisotropic *multi-index set:*

$$\Lambda_{TD}(\boldsymbol{\rho}, w) = \left\{ \boldsymbol{\nu} \in \mathbb{N}_0^N : \sum_{n=1}^N \rho_n \nu_n \leq w \right\} \qquad \forall w \geq 0,$$

where $\boldsymbol{\rho} = (\rho_n)_{n=1}^N \subset \mathbb{R}_{>0}^N$ and $\rho_n = \frac{1}{2}\log(\tau_n + \sqrt{1+\tau_n^2})$. See Figure 1.7 for examples of multi-index sets and corresponding sparse grids. The authors derive again algebraic convergence rates of the error with respect to the number of collocation nodes:

$$\left\| u - \mathcal{I}_{\Lambda_{TD}(\boldsymbol{\rho},w)} u \right\|_{C^0([-1,1]^N;\mathbb{U})} \leq C_2 \left(\# \mathcal{H}_{\Lambda_{TD}(W)} \right)^{-\mu_2}$$

where $\mu_2 = \frac{\min(\boldsymbol{\rho})(\log(2)e - 1/2)}{\log(2) + \min(\boldsymbol{\rho}) \sum_{n=1}^N \rho_n^{-1}}$ and $C_2 = C_2(\boldsymbol{\rho}, N)$ in general diverge as $N \to \infty$. However, if $\sum_{n \in \mathbb{N}} \rho_n^{-1} < \infty$, then C_2 and μ_2 are uniformly bounded with respect to N and convergence is dimension-independent. As a consequence, this also allows us to treat parametric functions defined on sequence domains (formally, $N = \infty$).

Example 1.16 (Hermite interpolation for non-compact parameter domains)**.** *Consider, for $\nu \in \mathbb{N}_0$, $\mathcal{Y}_\nu = \{y_0, \dots, y_{m(\nu)}\} \subset \mathbb{R}$, the set of $m(\nu) := \nu + 1$ zeros of the Hermite polynomial of degree $m(\nu)$. Observe that this nodes family is not nested. In [DNSZ23a], the authors*

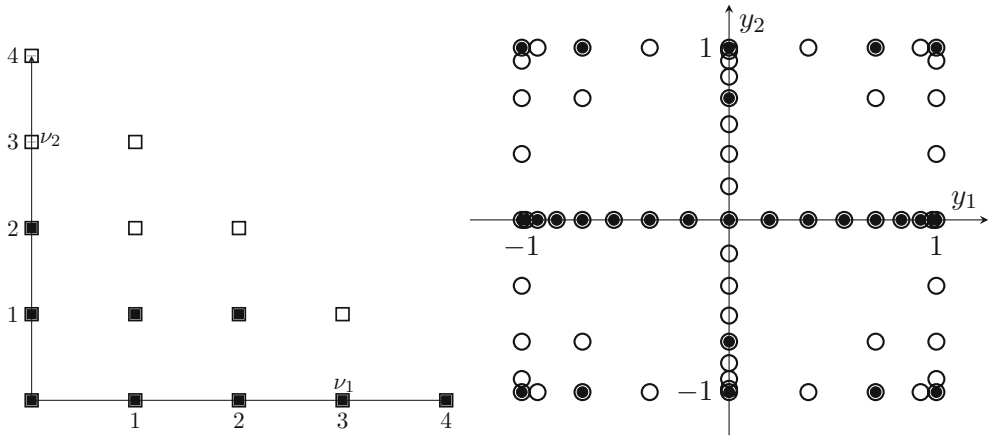

Figure 1.7: Left: Total degree multi-index set $\Lambda_{\mathrm{TD}}(4)$ (hollow squares) and anisotropic total degree multi-index set $\Lambda_{\mathrm{TD}}((\frac{1}{2},1),4)$ (solid squares). Right: Corresponding collocation nodes based on Clenshaw-Curtis 1D nodes, respectively in hollow and solid circles.

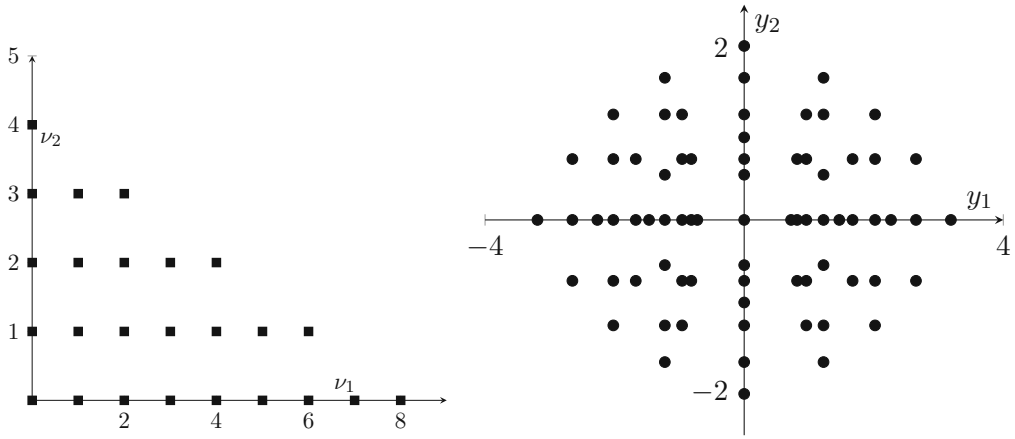

Figure 1.8: Left: Total degree multi-index set $\Lambda_{\mathrm{TD}}((\frac{1}{2},1),8)$. Right: Corresponding collocation nodes based on Hermite 1D nodes.

apply this sparse grid interpolation method to a class of problem including the parametric log-normal diffusion Poisson problem. There, it is assumed that the parameter-to-solution map u is $(\boldsymbol{b},\xi,\delta,\mathbb{U})$-holomorphic *(see [DNSZ23a, Definition 4.1]). By definition, u admits a holomorphic extension to $\mathcal{S}(\boldsymbol{\rho}) = \{\boldsymbol{z}\in\mathbb{C}^{\mathbb{N}} : |\Im z_j| < \rho_j \ \forall j\in\mathbb{N}\}$ and the $\mathbb{U}$-norm of the holomorphic extension is uniformly bounded by $\delta > 0$. Here $\boldsymbol{\rho} > 0$ is such that $\sum_{j\in\mathbb{N}} b_j\rho_j < \xi$. The multi-index set is defined as*

$$\Lambda(\varepsilon) = \left\{\boldsymbol{\nu}\in\mathcal{F} : c_{\boldsymbol{\nu}}^{-1} > \varepsilon\right\} \qquad \forall \varepsilon > 0,$$

where $c_{\boldsymbol{\nu}} = \prod_{j\in supp(\boldsymbol{\nu})} \max\{1, K\rho_j\}^2 \nu_j^{r-\tau}$ *and in turn* $K > 0$, $\tau > 0$, $r > \max\{\tau, 1\}$, $\rho_j = \frac{b_j^{p-1}\xi}{4\sqrt{r!}\|\boldsymbol{b}\|_{\ell^p}}$ *for all* $j\in\mathbb{N}$. *See Figure 1.8 for examples of multi-index sets and corresponding sparse grids. If* $\boldsymbol{b}\in\ell^p(\mathbb{N})$ *for* $0 < p < \frac{2}{3}$, *then we have the following convergence (see [DNSZ23a, Theorem 6.13])*

$$\|u - \mathcal{I}_\Lambda u\|_{L^2(\mathbb{R}^{\mathbb{N}},\mathbb{U})} \le C_3 \left(\#\mathcal{H}_{\Lambda(\varepsilon)}\right)^{-\frac{1}{p}+\frac{3}{2}},$$

where $C_3 > 0$ *is independent of* $\#\mathcal{H}_{\Lambda(\varepsilon)}$.

Example 1.17 (Leja sequence)**.** Leja sequences *on* $[-1,1]$*, denoted here as* $\mathcal{Y}_n = (y_i)_{i=1}^n$ *for* $n \in \mathbb{N}$*, are given by* $y_0 \in [-1,1]$ *arbitrary, and the following* $n-1$ *nodes recursively defined as*

$$y_i \in \arg\max_{z\in[-1,1]} \prod_{j=1}^{i-1} |z - y_j|$$

(this optimization problem may not have a unique solution, therefore the "$\in$*" rather than simple equality). For Leja sequences, the Lebesgue constant grows sub-exponentially with respect to the number of nodes:* $\lim_{n\to\infty} (\lambda_n)^{1/n} = 0$ *[TT10]. This can be intuitively understood as a result of the fact that Leja sequences are a greedy approximation of* Fekete nodes *(see [Tre19, Page 116]), i.e. nodes with optimal stability properties (but hard to construct). Leja sequences can be proved to be asymptotically distributed as Fekete nodes.*

Leja sequences can be generalized to any measure and to non-compact domains (see [JWZ18] and references within): Given a weight $w : \mathbb{R} \to \mathbb{R}$*, define* w-Leja sequences *on* $\mathbb{R}$ *as follows: After choosing again* $y_0 \in \mathbb{R}$ *arbitrarily, let*

$$z_i \in \arg\max_{z\in\mathbb{R}} w(z) \prod_{j=1}^{i-1} |z - y_j| \qquad \forall i \in \mathbb{N}.$$

It was proved in [JWZ18] that when the weight is in the form $w(z) = e^{|z|^\alpha}$*, the Lebesgue constant* λ_n *grows sub-exponentially in* $n \in \mathbb{N}$*.*

Leja sequences are not often used for a priori sparse grid approximation and their a priori convergence theory seems not to be well understood. They are more popular in adaptive sparse grid interpolation (see e.g. [CCS14] and Section 1.2.4 below for more details).

A review of sparse grid interpolation of parametric PDEs

The collocation type approximation of random coefficients PDE started at the end of the 20th century and satisfactory results for the model problem (the random diffusion Poisson problem presented in Section 1.1.1) were obtained at the beginning of the 21st century.

The work [XH05] is one of the first in which the authors suggest using sparse grid interpolation to approximate PDEs with random coefficients. They use Clenshaw-Curtis nodes and a total degree multi-index set as in Example 1.15. They discuss fundamental properties of the scheme such as exactness over certain polynomial spaces and estimates on the cardinality of the sparse grid as a function of the dimension $N \in \mathbb{N}$ and the scalar parameter $w \geq 0$ which determines the cardinality of the sparse grid. The error, estimated in the uniform norm, displays a convergence behavior that depends weakly exponentially on the number of parameters. This was still a large improvement over tensor product grids.

In [BNT10], the authors study a Poisson problem with random diffusion and forcing term under the finite dimensional noise assumption (see Remark 1.7) for both compact and non compact parameter spaces. They prove holomorphic regularity of the parameter-to-solution map and derive exponential convergence estimates for tensor product approximation. The convergence rate depends however on the number of parameters N, which allows them to treat only problems with a moderate number of dimensions.

In the subsequent work [NTW08b], the authors consider the scheme we presented in Example 1.14 based on Clenshaw-Curtis nodes and a total-degree multi-index set. Under the finite dimensional noise assumption and when the parameter-to-solution map has holomorphic regularity, the scheme converges exponentially. Convergence depends only mildly on the number of dimensions, a clear improvement over tensor product interpolation.

Another improvement was presented in [NTW08a], in which the second scheme we presented in Example 1.15 was introduced. A stronger sparsity assumption on the diffusion coefficient allows for a truly dimension-independent approximation under the condition that the multi-index set, of anisotropic total degree form, is appropriately tuned.

In [NTT16], the approach presented in [NTW08a] is generalized further. The selection of the sparse grid is based on a *profit* attached to each hierarchical surplus, so that the sparse grid selection problem is recast into a Knapsack problem following the earlier work by Griebel and co authors (see e.g. [BG04]). A convergence criterion is introduced to quantify the rate of algebraic convergence with respect to the number of collocation nodes and the dependence on the dimension of the parameter space. The approach is very general, as it makes few assumptions on the structure of the problem or the 1D interpolation scheme on which sparse grid interpolation is based. As an example, the authors apply the method to a Hilbert space valued function with the classical sparsity structure found in solutions of the random elliptic PDEs and discuss both a priori and adaptive methods. We will dive further into this result in Section 1.2.3.

Beyond the linear elliptic model problem, some works tackled time-dependent problems. In [ZG12], the authors extend the methodology developed in [BNT10] and the related papers to a liner parabolic problem with random coefficients under the finite-dimensional noise assumption. They prove existence of a holomorphic extension and convergence of stochastic collocation schemes for both a space semi-discrete and fully discrete approximations. In this work, holomorphy follows from estimates on the derivatives of the parameter-to-solution map together with the linearity of the problem.

In [NT09], the authors study a linear parabolic problem with random diffusion 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 approximations. The convergence of the latter family of schemes is an application of the convergence proofs found in previous works.

In [GZ07], the authors consider a coupled Navier-Stokes and heat equation problem with uncertainty. They opt for the adaptive sparse grid interpolation scheme from [GG03]. The performance of the scheme is showcased in several examples, e.g. the Rayleigh-Bernard advection under shape uncertainty of the bottom boundary.

Another important development was considering problems with unbounded parameters that arise e.g. from Gaussian random field coefficients. In [BCDM17], the authors study the Poisson problem with lognormal diffusion, the benchmark problem in UQ with unbounded random variables. They establish summability results for Hermite coefficients based on *local-in-space* summability of the basis used to expand the logarithm of the diffusion.

In [EST18], the authors approximate functions with this property by means of sparse grid interpolation built using global polynomials with Gauss-Hermite interpolate nodes, like the ones presented in Example 1.16. They prove algebraic and dimension independent convergence rates (Theorem 3.18) based on a certain summability properties of the Hermite coefficients (Assumption 3.9). It is important to notice that this property is stronger than the estimates on the derivatives one would obtain with the Cauchy's integral formula. Therefore, generalization to problems other than the lognormal diffusion Poisson looks problematic.

In the book [DNSZ23a], 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 main regularity assumption, "$(\boldsymbol{b}, \xi, \delta, X)$-holomorphy" (Definition 4.1), 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 (see also Example 1.16).

Finally, another important research path is going beyond the realm of linear PDEs. In [CCS15], 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 for maps between Banach spaces (see Theorem 3.2). This allowed them to prove parametric regularity for several non-standard problem. based on this regularity, they prove convergence of Legendre polynomial chaos expansion and sparse grid approximations.

In [CSZ18], the authors use similar techniques in the setting of the stationary Navier-Stokes equation with random domain. The random domain is described by a parametric transformation of a reference domain. The countably-many parameters span a bounded domain, so that the parameter space is compact. They then proceed to apply the result to prove convergence with rate of approximation with truncated Legendre expansions and sparse grid.

Sparse grid selection through profit maximization and convergence

Above we presented the sparse grid interpolation method and remarked that the interpolation operator requires a multi-index set $\Lambda \subset \mathcal{F}$. Several explicit examples were also given. However the choice of the multi-index set Λ is in general not obvious because of its high-dimensionality, which equals the number of problem parameters. We now present a useful method to select Λ based on attaching a *profit*, or "benefit-cost" ratio, to each multi-index and defining Λ as a suitable set containing the profit maximizing multi-indices. The definition of "benefit" is tightly bound to the regularity properties of the parameter-to-solution map u. This method was originally introduce by Griebel and co authors (see e.g. [BG04]) and later used e.g. in [GG03, BNTT11a, BTNT12] (for sparse grid quadrature) and in [EEST22, GN18] (for adaptive sparse grid interpolation). The following discussion is based on the more recent application in [NTT16].

Fix a function $u \in C^0(\mathbf{\Gamma})$, a sparse grid interpolation method with 1D-nodes family $\mathcal{Y}_{\nu\nu\in\mathbb{N}_0}$, and a *computational budget* $Q \in \mathbb{N}$, i.e. the maximum number of collocation samples allowed. We look for a *(quasi) optimal approximation*, or equivalently for a downward-closed multi-index set $\Lambda \subset \mathcal{F}$ with $\#\mathcal{H}_\Lambda \leq Q$ such that

$$\|u - \mathcal{I}_\Lambda u\| \lesssim \min\left\{\|u - \mathcal{I}_{\tilde{\Lambda}} u\| : \tilde{\Lambda} \subset \mathcal{F} \text{ downward closed such that } \#\mathcal{H}_{\tilde{\Lambda}} \leq Q\right\}, \tag{1.39}$$

where the hidden constant is independent of Λ and $\|\cdot\|$ is an appropriate norm for parametric maps $u : \mathbf{\Gamma} \to \mathbb{U}$.

In [NTT16], the authors suggest to proceed as follows. A sequence $(v_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \subset \mathbb{R}_{\geq 0}$ is called a *value* for the sparse grid interpolation method applied to the function u if

$$\|\Delta_{\boldsymbol{\nu}}[u]\| \lesssim v_{\boldsymbol{\nu}} \qquad \forall \boldsymbol{\nu} \in \mathcal{F}. \tag{1.40}$$

A sequence $(w_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}} \subset \mathbb{R}_{>0}$ is called a *work* for the sparse grid interpolation method if

$$\sum_{\boldsymbol{\nu}\in\Lambda} w_{\boldsymbol{\nu}} \lesssim \#\mathcal{H}_\Lambda \lesssim \sum_{\boldsymbol{\nu}\in\Lambda} w_{\boldsymbol{\nu}} \qquad \forall \Lambda \subset \mathcal{F} \text{ downward-closed}, \tag{1.41}$$

where the symbol "$\lesssim$" denotes an estimate up to a fixed constant independent of $\boldsymbol{\nu} \in \mathcal{F}$ (respectively $\Lambda \subset \mathcal{F}$). The multi-index Λ as in (1.39) solves the following *knapsack problem*:

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

Since the Knapsack problem is NP-hard, we consider its linear relaxation. Its solution is computed as follows:

1. Define the *profit*

$$\mathcal{P}_{\boldsymbol{\nu}} \coloneqq \frac{v_{\boldsymbol{\nu}}}{w_{\boldsymbol{\nu}}} \qquad \forall \boldsymbol{\nu} \in \mathcal{F}. \tag{1.42}$$

   We can always assume that profits are *monotone*, i.e. $\mathcal{P}_{\boldsymbol{\nu}+\boldsymbol{e}_n} \leq \mathcal{P}_{\boldsymbol{\nu}}$ for all $\boldsymbol{\nu} \in \mathcal{F}, n \in \mathbb{N}$. This is because, in case they are not, we can always work with *modified profits* $\tilde{\mathcal{P}}_{\boldsymbol{\nu}}$ defined as $\tilde{\mathcal{P}}_{\boldsymbol{\nu}} = \max_{\boldsymbol{\mu} \geq \boldsymbol{\nu}} \mathcal{P}_{\boldsymbol{\mu}}$. However, if the profit is estimated a-posteriori, this quantity may be difficult to compute. Therefore the decay of the profits may need to be assumed (*saturation assumption*).

2. Introduce the partial ordering of $\mathcal{F}$ induced by $\mathcal{P}_{\boldsymbol{\nu}}$:

$$\boldsymbol{\nu}_i \geq \boldsymbol{\nu}_j \Leftrightarrow \mathcal{P}_{\boldsymbol{\nu}_i} \geq \mathcal{P}_{\boldsymbol{\nu}_j} \qquad \forall i, j \in \mathbb{N}.$$

   If two multi-indices have the same profit, sort them in lexicographic order.

3. Let

$$\Lambda_n \coloneqq \{\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_n\} \qquad \forall n \in \mathbb{N}, \tag{1.43}$$

   and find the largest $M \in \mathbb{N}$ such that $\#\mathcal{H}_{\Lambda_M} \leq Q$, and define $\Lambda = \Lambda_M$.

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

Theorem 1.18. *[NTT16, Theorem 1] If there exists $\tau \in (0, 1]$ such that*

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

then

$$\|u - \mathcal{I}_{\Lambda_n} u\| \leq C_\tau \#\mathcal{H}_{\Lambda_n}^{1-1/\tau}.$$

The previous theorem generalizes Stechkin's lemma (cf. Corollary D.2), which corresponds to the case $w_{\boldsymbol{\nu}} = 1$ for all $\boldsymbol{\nu} \in \mathcal{F}$. In the original reference, the result is proved only under the finite dimensional noise assumption (see Remark 1.7). However, the proof applies verbatim to the infinite dimensional case since $\mathcal{F}$ is countable.

1.2.4 Adaptive sparse grid interpolation

In the previous section, we assumed we knew regularity and sparsity of the parametric map. We used this information to allocate a suitable number of collocation nodes per tuple of parametric direction in order to resolve the dependence of the solution on the corresponding scalar parameters. If however the regularity and sparsity of the parametric map are unknown, it may a priori not even be possible to produce an approximation algorithm that converges with a dimension-independent rate. This is what often happens in practical applications, in which the random coefficient PDE problem may be too complicated or beyond the known techniques to conduct a regularity analysis. Adaptive approximation methods offer a possible solution to this problem. Adaptive algorithms iteratively refine an approximation of increasing dimension and use it to

choose with respect to which parameter (or parameter tuple) to refine it in order to improve accuracy.

In general, adaptive numerical methods work by generating a sequence of numerical approximation $u_\ell \in V_\ell$, where V_ℓ is a linear space of dimension $d(\ell) \in \mathbb{N}$, which in turn is an increasing function of $\ell \in \mathbb{N}_0$. The sequence of approximate solutions achieves smaller and smaller errors $\|u - u_\ell\|$ so that, for any fixed tolerance $\varepsilon > 0$ an $\ell^* = \ell^*(\varepsilon) \in \mathbb{N}_0$ exists so that $\|u - u_\ell\| \leq \varepsilon$. Adaptive algorithms often require a "refinement parameter" $\theta \in (0, 1)$. They often consist of the following four steps, which are iterated in a loop for $\ell = 0, 1, \ldots, L$:

1. `COMPUTE` the discrete solution $u_\ell \in V_\ell$. It is assumed that this can be computed based on the knowledge of the discrete space V_ℓ. This space is either given as an initial space V_0 or computed based on information obtained in the previous loop;

2. `ESTIMATE` the error: $\|u - u_\ell\| \leq \zeta(u_\ell)$. The *a-posteriori error estimator* $\zeta(u_\ell) \geq 0$ must be computable given u_ℓ. Often finding a suitable definition of $\zeta(u_\ell)$ is not trivial;

3. $\theta -$ `MARK` the "features" of V_ℓ that need to be refined;

4. `REFINE` the space V_ℓ based on the previous step to obtain $V_{\ell+1}$.

The spaces V_ℓ and the precise definition of `MARK` and `REFINE` depend on the concrete problem at hand. Often, the a-posteriori estimator $\zeta(u_\ell)$ has a richer structure that is used in the steps `MARK` and `REFINE`. The fact that adaptive algorithms produce (quasi)-optimal solutions and their convergence is non-trivial and must be proved for the specific problem at hand (see [CFPP14] for a fairly general theory that however can be applied mostly to Galerkin-type methods).

The adaptive approach to sparse grid selection has a number of advantages: There is no need to derive a priori estimates to determine a suitable index set. Moreover, adaptive algorithms tend to select a discrete space that is tailored to the function at hand, while a priori selection is usually optimal for a whole *regularity class* of functions. This means that adaptive algorithms often deliver more accurate approximations for a comparable computational cost.
On the other hand, when possible, selecting the sparse grid a priori also has advantages: All collocation samples can be computed in parallel (if appropriate hardware is available). Moreover, the use of non-nested nodes makes adaptive computation possibly redundant, while for a priori sparse grid non-nested nodes are just as effective as nested ones. Finally, the theoretical analysis of adaptive algorithms is more complex and fewer results are available.

To describe some popular adaptive methods we have to introduce:

Definition 1.19. *Given a multi-index set $\Lambda \subset \mathcal{F}$ the* margin of Λ *is by definition*

$$\mathcal{M}_\Lambda := \{\boldsymbol{\nu} \in \mathcal{F} \setminus \Lambda : \textit{there exists } n \in \textit{supp}\,(\boldsymbol{\nu}) : \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda\}\,.$$

The reduced margin of Λ *is by definition*

$$\mathcal{RM}_\Lambda := \{\boldsymbol{\nu} \in \mathcal{F} \setminus \Lambda : \textit{ for all } n \in \textit{supp}\,(\boldsymbol{\nu}) : \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda\}\,.$$

Let us now briefly present some adaptive sparse grid interpolation methods.

In [GG03], the authors propose one of the first dimension-adaptive sparse grid quadrature methods. The algorithm automatically selects the most relevant dimensions needed to achieve an approximation error within the desired tolerance and neglects the others. Starting from the trivial multi-index set, a nested sequence $(\Lambda_\ell)_{\ell \in \mathbb{N}}$ of multi-index sets is defined as

$$\Lambda_0 = \{\mathbf{0}\}\,, \qquad \Lambda_{\ell+1} = \Lambda_\ell \cup \{\overline{\boldsymbol{\nu}}\} \qquad \forall \ell \in \mathbb{N},$$

where $\overline{\boldsymbol{\nu}} \in \mathcal{RM}_{\Lambda_\ell}$ (see Definition 1.19) is the maximizer of a given *error indicator* $g_{\boldsymbol{\nu}}$:

$$\overline{\boldsymbol{\nu}} = \operatorname*{arg\,max}_{\boldsymbol{\nu} \in \mathcal{RM}_{\Lambda_\ell}} g_{\boldsymbol{\nu}}$$

The error indicator $g_{\boldsymbol{\nu}}$ is defined based on $\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty(\mathbf{\Gamma}^N)}$, which can be interpreted as a computable estimate of the *value* (1.40), and the *work* (1.41) (always easy to compute) associated to $\boldsymbol{\nu}$. For example, $g_{\boldsymbol{\nu}}$ can be defined as the ratio of a-posteriori value and work, analogously to the profit (1.42).

A heuristic error estimator is given by $\sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} g_{\boldsymbol{\nu}}$. However, the actual error is only bounded by $\sum_{\boldsymbol{\nu} \notin \Lambda} \|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty(\mathbf{\Gamma})}$ and there is in general no guarantee that the former sum bounds the latter, even up to an Λ_ℓ-independent constant. This may be more easily proved if an appropriate *saturation assumption* holds, i.e. if $\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty(\mathbf{\Gamma})}$ decreases sufficiently fast with the magnitude of the multi-index. This assumption excludes e.g. functions u with local peaks or discontinuities. Nevertheless, this adaptive method almost always works well in concrete examples, which suggests that a saturation assumption indeed holds true. Finally, let use mention that computing $g_{\boldsymbol{\nu}}$ requires additional samples of the parameter-to-solution map, therefore entailing a relatively large computational cost of this step.

In [CCS14], an adaptive sparse grid interpolation method is defined. The authors use Leja sequences (see also Example 1.17) to obtain a maximally granular sparse grid interpolant, i.e. sparse grid and multi-index set have the same cardinality (each collocation node corresponds to one multi-index). The multi-index set is selected greedily by looking at the multi-index in the reduced margin that corresponds to the point with maximum pointwise interpolation error, thus leading to naturally nested sparse grids. There is no guarantee of convergence because of data oscillations, and lack of additional assumptions that would come from considering a more specific problem. In Theorem 4.4 it is proved that, for the parametric affine diffusion Poisson problem (Example 1.8 with diffusion 1.15) and $0 < p < 1$ such that $\left(\|\phi_j\|_{L^\infty(D)}\right)_{j\in\mathbb{N}} \in \ell^p$ (decay of the affine diffusion basis functions), a sequence of sparse grid interpolants leading to optimal approximation rates $\mathcal{O}\left((\#\mathcal{H})^{1-1/p}\right)$, indeed exists. The adaptive sparse grid algorithm behaves optimally in numerical experiments but convergence is guaranteed only after a possibly sub-optimal modification of the algorithm.

In [GN18], the authors tackle the problem of sparse grid–finite element approximation of the parametric affine diffusion Poisson problem (i.e. Example 1.8 with diffusion (1.15)). They first define a reliable a-posteriori estimator for both the sparse grid and finite element approximation error. They use it to define an adaptive algorithm under the assumption of a negligible finite element error. $u : \mathbf{\Gamma} \to \mathbb{U}$ denotes the (exact) parameter-to-solution map from Problem (1.14), where $\mathbf{\Gamma} = [-1, 1]^N$ for $N \in \mathbb{N}$ (i.e. they work under the finite dimensional noise assumption, see Remark 1.7) and $\mathbb{U} = H_0^1(D)$. The fully discrete sparse grid-finite element solution is $\mathcal{I}_\Lambda[U] = \sum_{\boldsymbol{y} \in \mathcal{H}_\Lambda} U_{\boldsymbol{y}} L_{\boldsymbol{y}}$, where $\mathcal{I}_\Lambda$ denotes sparse grid interpolation with respect to a given downward-closed multi-index set $\Lambda \subset \mathbb{N}_0^N$ and $U : \mathbf{\Gamma} \to \mathbb{U}$ denotes the *parameter-to-finite-element-solution map* which, to the parameter $\boldsymbol{y} \in \mathbf{\Gamma}$, associates $U(\boldsymbol{y}) = U_{\boldsymbol{y}} \in \mathbb{U}_{\boldsymbol{y}} \coloneqq \mathcal{S}_0^1(\mathcal{T}_{\boldsymbol{y}})$, i.e. a conforming finite element approximation of $u(\boldsymbol{y})$ in the space of piecewise linear functions with respect to the shape-regular mesh $\mathcal{T}_{\boldsymbol{y}}$. The total estimator is composed of a *parametric estimator*

$$\zeta_{SC,\Lambda} \coloneqq \sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \zeta_{\boldsymbol{\nu},\Lambda}, \qquad \zeta_{\boldsymbol{\nu},\Lambda} \coloneqq \|\Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[U]\right)\|_{L^\infty(\mathbf{\Gamma}, L^2(D))} \tag{1.44}$$

(the gradient ∇ here acts exclusively on the space variable $\boldsymbol{x} \in D$) and a *finite element estimator*

$$\begin{aligned} \eta_{\mathrm{FE},\Lambda} &\coloneqq \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} \eta_{\boldsymbol{y}} \left\|L_{\boldsymbol{y}}\right\|_{L^\infty(\boldsymbol{\Gamma})}, \qquad \eta_{\boldsymbol{y}} \coloneqq \left(\sum_{T\in\mathcal{T}_{\boldsymbol{y}}} \eta_{\boldsymbol{y},T}^2\right)^{\frac{1}{2}}, \\ \eta_{\boldsymbol{y},T}^2 &\coloneqq h_T^2 \left\|f + \nabla\cdot(a(\boldsymbol{y})\nabla U_{\boldsymbol{y}})\right\|_{L^2(T)}^2 + \sum_{e\subset\partial T} h_e \left\|\frac{1}{2}\left[a(\boldsymbol{y})\nabla U_{\boldsymbol{y}}\cdot\boldsymbol{n}_e\right]_{\boldsymbol{n}_e}\right\|_{L^2(e)}^2, \end{aligned} \tag{1.45}$$

where $[\cdot]_{\boldsymbol{n}_e}$ denotes the jump over the edge (face) in normal direction $\boldsymbol{n}_e$. Their combination yields a reliable upper bound (see [GN18, Proposition 4.3]) i.e.

$$\left\|u - \mathcal{I}_\Lambda[U]\right\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} \leq \frac{1}{a_{\min}}\left(C\eta_{\mathrm{FE},\Lambda} + \zeta_{SC,\Lambda}\right), \tag{1.46}$$

where $a_{\min} > 0$ appears in the equivalence relation between $H_0^1(D)$ and energy norm

$$a_{\min}^{1/2}\left\|v\right\|_{H_0^1(D)} \leq \left\|a(\boldsymbol{y})^{\frac{1}{2}}\nabla v\right\|_{L^2(D)} \leq a_{\max}^{1/2}\left\|v\right\|_{H_0^1(D)} \qquad \text{for a.e. } \boldsymbol{y}\in\boldsymbol{\Gamma} \text{ for all } v\in H_0^1(D)$$

and $C > 0$ depends only on the shape regularity of $\mathcal{T}_{\mathrm{init}}$. The authors also discuss an adaptive algorithm in the case of a negligible finite element error and support results with numerical experiments.

In the sequence of papers [BSX22a, BS23, BS24], the authors suggest an alternative to [GN18] that can tackle the non-affine case (but still compact parameter domain and under the finite dimensional noise assumption). They employ a *hierarchical a-posteriori error estimator*, meaning that the error is estimated by a difference of a refinement of the current approximation and the approximation itself. Also this error estimator splits in a parametric and space contributions. Reliability depends on the validity of a *saturation assumption* i.e., loosely speaking, the approximation error (in parametric and physical space) contracts when both parametric and space discretizations are refined. As for the adaptive algorithm, either parametric or space approximation is carried out at each iteration, based a on comparison of the estimators. A Dörfler-type refinement is carried out on the selected feature, based on a parameter $\theta \in (0,1)$. Numerical examples suggest a convergence rate of order $\frac{1}{3}$ for the single-level method (i.e. same finite element space in all collocation nodes). A more general strategy consists of allowing for different meshes in different collocation nodes, called a "multilevel" strategy in this work. Again, the method approximates the solution of the affine diffusion Poisson problem with order $\frac{1}{3}$. Only on a "single peak" problem with parametric right-hand side the multilevel method has an improved convergence order $\frac{1}{2}$ compared to again order $\frac{1}{3}$ for the single-level. Plain convergence of the multilevel algorithm is proved under summation properties of the Taylor coefficients of a semidiscrete approximation, implied for the Poisson problem by suitable growth conditions on derivatives of the diffusion. The authors stress the difference between error *estimators*, used exclusively to estimate the approximation error, and error *indicators*, used to guide refinement. While a sum of error indicators bounds the estimator, it is not an efficient estimate itself, i.e. the error decays at a faster rate. More recently, the related preprint [BPRS24] on goal oriented sparse grid adaptivity appeared.

In the recent work [GS24], the authors approximate quantities of interest of the the lognormal diffusion Poisson problem using an adaptive sparse grid technique that combines parametric, spatial and truncation approximations. They define difference operators for the truncation, finite element approximation (on a sequence of uniform meshes indexed by a scalar parameter),

and quadrature (through sparse grid Gauss-Hermite quadrature) and combine them to define hierarchical surpluses, which in turn form the sparse grid approximation when appropriately summed. The error can be heuristically estimated with a sum of norms of hierarchical surpluses over the margin: $\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \|\Delta u\|$. This is as in [GG03] not a reliable estimator and requires additional samples. The sparse grid is enlarged greedily by iteratively enriching Λ with the multi-index that maximizes a *value-cost ratio* $\|\Delta_{\boldsymbol{\nu}}\| / c_{\boldsymbol{\nu}}$, where $c_{\boldsymbol{\nu}}$ quantifies the added cost if evaluating the refined sparse grid approximation. Convergence follows from a saturation assumption, the proof of which is however out of reach in this setting. While convergence and optimality are not proved, a number of numerical examples suggest that the bottleneck of convergence is the space discretization.

Remark 1.20 (Comparison of adaptive methods)**.**

- *The estimator from [GN18] and [BSX22a, BS23, BS24] are reliable, while the one from [GG03] and [GS24] may not be. Remember however that the first two algorithm are tailored to the affine diffusion Poisson problem, while the last two are, in different aspects, more general. The reliable estimator can be used to stop the adaptive procedure and ensure that the error is below a given tolerance. Reliability also allows us to prove convergence of the discrete solution by proving that the* estimator *vanishes;*
- *The estimator in [GN18] can be computed without further sampling of the parameter-to-solution map, while the one in [GG03], [GS24], and [BSX22a, BS23, BS24] require additional collocation samples;*
- *The fact that the reliable parametric estimator* (1.44) *is a* finite *sum (it is a sum over the margin of the multi-index set $\mathcal{M}_\Lambda$, which is finite if the finite-dimensional noise assumption holds) is a consequence of the choice of an* affine *random diffusion. For different random models, such as the Poisson problem with lognormal diffusion, reliability is in general possible only with an infinite sum of pointwise estimators. However, if a* saturation assumption *holds, the infinite sum may again be bounded, up to a constant, by a finite sum (e.g. on the margin);*
- *Neither [GG03] nor [GN18] or [GS24] prove convergence of the* parametric *adaptive procedures. In [EEST22], the authors prove plain convergence for the methods studied in the first two papers. In [FS21], we prove plain convergence of the algorithm from [GN18] (see also Theorem 2.15). We then define a fully discrete adaptive algorithm by adding finite element refinement and prove convergence for this as well. As already mentioned, [BS24] proves the same for the adaptive algorithms hierarchical estimators. These three proofs are based on a technique introduced in [BPRR19] in the context of the stochastic Galerkin method;*
- *Dimensional adaptivity is considered only in [GG03], [CCS14] and [GS24]. The other algoirhtms rely on the finite dimensional noise assumption;*
- *An important aspect to consider to obtain optimal adaptive sparse grid methods is that error* indicators *should not be directly used as error* estimators*. Efficient error estimators are often bounded from above by sums of error indicators, an the last ones should only be used to guide refinement (often in a greedy way, based on a cost-work ratio;*

1.2.5 Other high-dimensional approximation and integration methods

In the present section, we mention a few (among the many) other high dimensional approximation and integration algorithms and compare them to sparse grid interpolation.

Quasi-Monte Carlo quadrature

Quasi-Monte Carlo quadrature (QMC) (see the review [DKS13], the monograph [LP14] or the upcoming monograph [Owe23] with draft available online) is an equal weights quadrature formula with deterministic collocation nodes aimed at improving the $\mathcal{O}(N^{-1/2})$ convergence rate of Monte Carlo quadrature. Given a function $f : \mathbf{\Gamma} \to \mathbb{R}$, where $\mathbf{\Gamma} \subset \mathbb{R}^N$ with $N \in \mathbb{N}$ (or, as above, $\mathbf{\Gamma}$ may come from truncating $\mathcal{F}$, the space of sequences with finite support defined in Equation (1.22)), a QMC formula with $M \in \mathbb{N}$ nodes $(\boldsymbol{y}_i)_{i=1}^M \subset \mathbf{\Gamma}$ reads simply $Q_M(f) = \frac{1}{M}\sum_{i=1}^M f(\boldsymbol{y}_i)$.

The quadrature error of any equal weight formula is bounded by the product of the *discrepancy* of the points set and a norm of the derivatives of the integrand (the *Koksma-Hlawka inequality* [Hic14]). For this reason, collocation nodes are chosen as *low discrepancy sequences*. This means that they are well distributed on the integration domain without leaving large empty areas. Some examples are: Sobol sequences [Sob67], lattice rules [DKP22], and digital nets [DP10].

When no assumption on the sparsity of the integrand is made, usual results show dimension-dependent convergence with estimates that present e.g. a multiplicative term $\log(M)^N$. Quadrature methods can be informed about the sparsity of the integrand through *weighted Sobolev spaces*. Weights $\{\gamma_{\mathfrak{u}} : \mathfrak{u} \subset \{1, \dots, N\}\} \subset \mathbb{R}_{\geq 0}$ quantify the relative importance of a parameter or set of parameters. The basic setting is the one of a *Reproducing Kernel Hilbert Space* (RKHS) defined via a weighted kernel. Such a kernel $K_{N,\boldsymbol{\gamma}}(\boldsymbol{x}, \boldsymbol{y}) : \mathbf{\Gamma} \times \mathbf{\Gamma} \to \mathbb{R}$ may be defined by

$$K_{N,\boldsymbol{\gamma}}(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \sum_{\mathfrak{u} \subset \{1,\dots,N\}} \gamma_{\mathfrak{u}} \prod_{j \in \mathfrak{u}} \eta(x_j, y_j),$$

where $\eta(x, y)$ is an appropriate 1D kernel of $x, y \in \mathbb{R}$. It can be proved that the corresponding scalar product reads

$$\langle f, g \rangle_{N,\boldsymbol{\gamma}} = \sum_{\mathfrak{u} \subset \{1,\dots,N\}} \gamma_{\mathfrak{u}}^{-1} \int_{[0,1]^{|\mathfrak{u}|}} \left(\int_{[0,1]^{N-|\mathfrak{u}|}} \partial_{\mathfrak{u}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}_{\mathfrak{u}^c} \right) \left(\int_{[0,1]^{N-|\mathfrak{u}|}} \partial_{\mathfrak{u}} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}_{\mathfrak{u}^c} \right) \mathrm{d}x_{\mathfrak{u}}, \quad (1.47)$$

where $\mathfrak{u}^c = \{1, \dots, s\} \setminus \mathfrak{u}$. If the weights vanish sufficiently fast (equivalently, they have suitable summability properties), then a dimension independent convergence of the QMC formula is possible for an appropriate choice of quadrature nodes which necessarily have low discrepancy and depend on the weights $\boldsymbol{\gamma}$. The choice of weights is crucial to describe the sparsity of the integrand. An important example in the setting of random coefficient PDEs are *Component by component* (CBC) weights [KSS12].

QMC has applications in several fields of science and engineering (especially in computational finance [AG07]). In applied mathematics, QMC is a popular method to approximate integral quantities of interest (such as moments or probabilities of events) of random coefficient PDEs (see, among many other works, [KN16, GKN$^+$11, KSS12]). The use of different low discrepancy sequences has been explored and dimension-independent linear, polynomial, or exponential convergence rates have been obtained depending on the regularity of the parametric map and choice of quadrature nodes.

Remark 1.21 (Comparison of analytical properties and techniques for QMC quadrature and sparse grid interpolation)**.** *While QMC quadrature and sparse grid interpolation are numerical methods for different problems, the high-dimensionality setting and how both can circumvent the curse of dimensionality allows for some interesting comparison:*

- *Sparse grid nodes are very far from being low discrepancy sequences (it is clear looking at e.g. Figure 1.6). Moreover, sparse grid quadrature is not an equal weights formula;*

- *When no assumption on the sparsity of the function is available, error estimates for both methods present dimension-dependent factors;*

- *Sparsity, the assumption needed to obtain dimension-independent convergence, is understood in two formally different but morally similar ways in QMC quadrature and sparse grid interpolation:*

  - *In QMC, the integrand f is assumed to belong to a RKHS with weighted scalar product, where the weights vanish (in an appropriate sense). A finite RKHS norm (e.g. the one associated to (1.47)) roughly means summability of appropriate integral quantities of derivatives of the integrand function;*
  - *In sparse grid interpolation, sparsity can ultimately be understood as the finiteness of the constant C_τ (for some $0 < \tau < 1$) in Theorem 1.18. This in turn means weighted summability of integral quantities of derivatives of the function to interpolate; A different method is analysed in [EST18], where convergence of a spectral-type sparse grid interpolation is implied (again) by a weighted summability of integrals of derivatives (see Assumption 3.9);*

  Therefore, the two concepts of sparsity are formally different but qualitatively similar. Indeed, sparsity is often proved with similar methods, such as the existence of a bounded holomorphic extension (with a sufficiently large domain of holomorphy) or through a recursive argument (see the list in Section 1.1.4);

- *QMC often has a dimension-independent convergence in a wider range of cases compared to sparse grid. This can be intuitively understood since a surrogate model contains more information than quadrature;*

- *Stability is an issue for sparse grid interpolation of "spectral type", as well as for high-order QMC quadrature;*

- *Several different strategies are available with both methods when dealing with unbounded domains, e.g. $\mathbb{R}^N$ for $N \in \mathbb{N}$:*

  - *The problem can be reduced to interpolation or approximation on the unit cube $[-1, 1]^N$ by a change of variable. In QMC, this is done for example in [NS23], in which the authors truncate the domain of integration $\mathbb{R}^N$ to a high-dimensional cube $[\boldsymbol{a}, \boldsymbol{b}]$ for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$ and perform quadrature on $[\boldsymbol{a}, \boldsymbol{b}]$ with scaled lattice rules on $[0, 1]^N$. The error is split in a quadrature and a truncation term. They so achieve high-order convergence limited only by the integrand smoothness. As for sparse grid interpolation, our approach in Chapter 3 is under some aspects similar: We build a sparse grid interpolation method based on piecewise polynomial interpolation on an interval, and extend outward using the polynomial defined on the first and last subinterval. See Section 3.5 for additional details;*
  - *When using high-order methods, a scaling is not always possible because of stability issues. Hermite nodes, i.e. zeros of hermite orthogonal polynomials are a popular choice for high order sparse grid approximation of functions of Gaussian random variables, see e.g. [EST18, DNSZ23b] as well as Example 1.16 above. For QMC integration, an approach in which the affine scaling strategy is still viable is given in [DILP18]. Here, the authors approximate integrals in the form $\int_{\mathbb{R}^N} f(\boldsymbol{y})\mu(\boldsymbol{x})d\boldsymbol{x}$, where f belongs to a RKHS based on Hermite polynomials, and μ is the N-dimensional standard Gaussian density. High-order digital nets are mapped from the unit cube*

$[0,1]^N$ to an appropriate cube in $\mathbb{R}^s$ to obtain optimal rate $\mathcal{O}(N^{-\alpha})$ up to a logarithmic term, where $\alpha > 0$ corresponds to the regularity of the integrand.

Other high-dimensional approximation methods

Moving to the domain of *approximation* (also called *reduced order modeling*), methods that produce a sparse polynomial approximation are presented in the review paper [CD15] and the monograph [ABW22]. In particular, we have

- *Truncated sparse Taylor polynomial*: Already discussed in Section 1.2.2;
- *Generalized polynomial chaos (gPC).* The polynomial approximations is expressed in a basis of orthogonal polynomials with respect to a fixed measure. The approximation is often quasi-optimal in the least-square sense with respect to the product of these measures. See [EMSU12, WK06, XE02];
- *Stochastic Galerkin*: As in the generalized polynomial chaos expansion (see above), the parametric dependence is modeled with orthogonal polynomials, while the physical variables dependence is discretized with the classical finite element method. The problem is so reduced to the solution of a large linear system with a block structure where each block is comparable to a deterministic finite element matrix and the number of blocks is determined by the cardinality of the parametric polynomial basis. Stochastic Galerkin is an intrusive method since the problem can only be solved "all at once". The method requires the solution of large linear systems, whose size is a multiple of the size of the corresponding deterministic finite element system dependent on the number of parametric basis functions. On the positive side, being a projection-type method, Stochastic Galerkin offers quasi-optimality properties. See [FST05, GS91, Gha99, MK05, XK03] for a priori approximation and [BPR21, BPR22, EGSZ14, EGSZ15] for adaptive methods;
- *Least Squares*: A flexible method with good stability properties that achieves a quasi-optimal approximation in the L^2 sense. It allows for approximation with noisy samples. Samples can (but need not to) be random Monte Carlo samples. The number of samples needed for a stable approximation scales like the approximation space dimension plus a log-dependent term. stability is only log-dependent on the number of parametric dimensions. While the quasi-optimality leaves freedom in the choice of discrete space, sparse polynomial spaces defined though downward-closed multi-index sets are often the preferred choice. See [CDL13, CCM$^+$15, CM17, CM23];
- *Compressed sensing*: A method to approximate sparse solutions of under-determined linear systems. By *sparse* here we mean small ℓ^0 norm (number of nonzero components). Since this problem is NP hard, it is often relaxed to an ℓ^1 minimization problem. Compressed sensing does not require exact knowledge of the multi-index set to use, rather identifies it itself. Compressed sensing algorithms, such as *Quadratically Constrained Basis Pursuit* (QCBP) or *Least Absolute Shrinking and Selection Operator* (LASSO), can be constrained to look for solutions on lower sets, to deal with measurement errors, and close-to-sparse exact solution. The method is non-intrusive (requires only collocation samples). See [BBRS15, BMP15, MG12].

Other methods that use sparsity but do not necessarily produce a sparse polynomial approximation, are

- *Reduced basis*: The aim is to find an n-dimensional linear subspace $V_n \subset V$ of the linear space to which solutions belong. The dimension n should be as small as possible, but V_n should also "approximates well" the *solutions manifold* (in the sense of *Kolmogorov widths*, see [CD15, Definition 1.41]). V_n is build with linear combinations of *snapshots* $u(a_1), \dots, u(a_n)$ form the solution manifold. While the computation of snapshots (*offline phase*) may be time consuming, once V_n is available, it can be used (*online phase*) to carry out fast queries of the surrogate model. The performance of the algorithm depends on how the snapshots are chosen. Popular strategies belong to the family of *greedy selection algorithms* (see [CD15, Section 8]). See also[CQR17] for a study of the method applied to the classical random coefficient elliptic problem, and the monograph [QMN15];

- *Low rank tensor representations*: A family of nonlinear methods that can be understood as a generalization of the classical separation of variables techniques. They provide a compressed approximation of functions and allow for memory save and speed-up of linear algebra operations. See [DKLM15, EHL$^+$14, BCD17, Bac23]. Dynamical low-rank approximation methods proved to be very effective to approximate evolutionary problems with high-dimensional parameters. See [Mus17];

- *Artificial Neural Networks (ANN)*: A surrogate model of the parameter-to-solution map can be given by an ANN. If appropriately trained, it can quickly provide accurate approximate samples of the random PDE solutions. ANNs have been shown to be a powerful approximation tool in a number of fields such as image classification, speech recognition, time series analysis. The current evidence suggests that they can overcome the curse of dimensionality (see Section 1.2.1). Despite the ever increasing number of successful applications, a comprehensive theoretical understanding is still missing and theoretical aspects of neural network approximation is an active field of research. Let us mention a few results in this direction: In [GHJvW23], the authors show that ANNs can effectively approximate the solutions of Back-Scholes PDEs and circumvent the curse of dimension. In particular, the number of ANN parameters scales linearly with the inverse of the prescribed approximation accuracy and also linearly with the number of PDE dimensions. In [OPS20], the authors use Deep Neural Networks to approximate functions on the interval and they prove they achieve convergence rates comparable to "best in class" approximation schemes for functions of the corresponding regularity. In the follow-up work [OS24], the authors prove expressivity and stability results with respect to Sobolev norms for piecewise-polynomial functions defined on a partition of the interval. They achieve this by a new construction based on Chebyshev coefficients, easy to compute given samples of the function over Clenshaw-Curtis nodes. Again exponential expression rates are proved for analytic function. However, one should also mention negative results such as [GV23], which shows that high rates of convergence is impossible in practice.

1.2.6 Finite element and fully discrete approximations

Consider $u : \mathbf{\Gamma} \to \mathbb{U}$ solution to a parametric boundary value problem. Recall that $\mathbf{\Gamma} \subset \mathbb{R}^{\mathbb{N}}$ is the parameter set, $\mathbb{U}$ is a Banach space of functions, and $\mathcal{I}_\Lambda : C^0(\mathbf{\Gamma}) \to V_\Lambda$ denotes the sparse grid interpolant defined by the downward-closed (and finite) multi-index set $\Lambda \subset \mathbb{N}_0^{\mathbb{N}}$ and appropriate 1D interpolation operator. However, $\mathcal{I}_\Lambda u$ is not a *computable* approximation of u, since the collocation samples $u(\boldsymbol{y})$ for $\boldsymbol{y} \in \mathbf{\Gamma}$ are in general not explicitly available.

To this end, we introduce the parameter-to-*finite-element-solution* map $U : \mathbf{\Gamma} \to \mathbb{U}_h$, i.e. for any $\boldsymbol{y} \in \mathbf{\Gamma}$, $U(\boldsymbol{y})$ is the finite element approximation of $u(\boldsymbol{y})$ in the finite element space $\mathbb{U}_h$, a piecewise polynomial space on a triangulation of the domain with suitable boundary condition.

We can then substitute the collocation samples $u(\boldsymbol{y})$ for $\boldsymbol{y} \in \mathcal{H}_\Lambda$ with their finite element approximations $U(\boldsymbol{y})$ and obtain a *sparse grid finite element approximation*

$$\mathcal{I}_\Lambda U = \sum_{\boldsymbol{y} \in \mathcal{H}_\Lambda} U(\boldsymbol{y}) L_{\boldsymbol{y}}.$$

We estimate the approximation error by the triangle inequality:

$$\|u - \mathcal{I}_\Lambda U\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})} \leq \|u - \mathcal{I}_\Lambda u\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})} + \|\mathcal{I}_\Lambda [u - U]\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})}, \tag{1.48}$$

where the first term accounts for the sparse grid error, while the second one for the finite element approximation. To obtain a small approximation error, the multi-index set must be chosen to capture the major features of the parameter-to-solution map, while the finite element space should resolve the geometric features and related singularities in space. Observe that for $p = \infty$ this second contribution can in turn be estimated as

$$\|\mathcal{I}_\Lambda [u - U]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} \leq \mathbb{L}_\Lambda \|u - U\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})},$$

where we recall that $\mathbb{L}_\Lambda > 0$ denotes the Lebesgue constant of the sparse grid interpolant $\mathcal{I}_\Lambda$ (cf. (1.34)).

Observe that an alternative splitting is possible:

$$\|u - \mathcal{I}_\Lambda U\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})} \leq \|u - U\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})} + \|U - \mathcal{I}_\Lambda U\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})}.$$

However, this expansion is harder to analyze than (1.48) since bounding $\|U - \mathcal{I}_\Lambda U\|_{L^p(\boldsymbol{\Gamma},\mathbb{U})}$ requires knowledge of the regularity of the parameter-to-finite element solution map, which in general is not easier to study that the "exact" parameter-to-solution map. As we have seen in Section 1.1.4, the parameter-to-solution map arising from parametric PDEs is often analytic (admits a holomorphic extension) and sparse. This needs not be the case for the parameter-to-finite-element-solution map. A notable exception is the Poisson problem when the same finite element space corresponds to each collocation node. The proof of holomorphy can be carried out verbatim from the one of the exact solution. The main reason is that the finite element solution is also defined as the solution to a variational problem.

When instead different finite element spaces are used for different collocation nodes, the regularity may "break" because the variational spaces corresponding to each collocation point is not the same. A strategy that lies "in between" using different finite element spaces for each collocation node and always using the same is using a *multilevel method*. This idea was introduced by Giles (see e.g. [Gil15]) in he context of Monte Carlo quadrature. The high flexibility and generality of the multilevel framework allows to apply it to a number of other methods such as quasi-Monte Carlo quadrature [HS19, KSS15] and sparse grid interpolation [TJWG15, LSS20], as well as different space and time discretization methods.

1.3 The stochastic Landau–Lifshitz–Gilbert equation (SLLG)

In the present section, we introduce the stochastic Landau–Lifshitz–Gilbert equation, a random model in *micromagnetics*, i.e. the study of magnetic bodies at sub-micrometer scale. We start in Section 1.3.1 by giving a physical description of such materials. We then move to the deterministic LLG equation (cf. Section 1.3.2) and only afterwards (cf. Section 1.3.4) extending the model to the stochastic case. In Sections 1.3.3 and 1.3.5, we present numerical methods for the space and time approximation in the deterministic and stochastic problem respectively. In Section 1.3.6, we show how the SLLG equation can be reduced to a *random coefficient* PDE and in Section 1.3.7 we discuss a numerical scheme for its space and time approximation. Finally, in Section 1.3.8 we list possible physics and engineering applications of a reduced order model of the random field solution.

1.3.1 Physical observations on magnetic materials

The following short introduction to the physics of magnetic materials is based on [Rug20]. See Appendix A for a list of physical units.
Simple experiments reveal that bodies made of particular materials react when subject to an external magnetic field $\boldsymbol{B}$ (in $\boldsymbol{T}$). Such materials are called magnetic. They are subject to a torque $\boldsymbol{\tau}$ (in $\boldsymbol{J}$) that causes the magnetic body to align with the magnetic field. The *magnetic moment* $\boldsymbol{\mu}$ (in $\frac{\boldsymbol{J}}{\boldsymbol{T}}$ or equivalently Am^2) of a body relates the torque to the external magnetic field:

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \boldsymbol{B}.$$

The quantity of magnetic moment per unit volume is called *magnetization* $\tilde{\boldsymbol{m}}$ (in $\frac{A}{m}$). In micromagnetics, one neglects the atomistic nature of the body and rather describes it in the framework of continuum physics.
A body made of a *magnetic material* possesses a (non-zero) intrinsic magnetic moment.
Permanent (or spontaneous) magnets are bodies made of magnetic materials that can be *magnetized* and, as a consequence, spontaneously generate a magnetic field.
Permanent magnets can be either *ferromagnetic* or *ferrimagnetic*. The two categories differ in their microstructure: For ferromagnets, the atomic magnetic poles are aligned in the same direction. For ferrimagnets, neighbouring magnetic poles point in opposite directions, but their contribution is different and the net magnetic moment is non zero.
A ferromagnet that easily demagnetizes after being magnetized is called magnetically *soft*. By contrast, if it tends to remain magnetized it is called magnetically *hard*. Ferromagnets can be demagnetized when subject to external magnetic fields. This susceptibility is called *coercivity* of the ferromagnet. Soft materials have low coercivity while hard materials have high coercivity.
For a magnetic material, we have then the following local relation:

$$\boldsymbol{B} = \mu_0(\boldsymbol{H} + \tilde{\boldsymbol{m}}),$$

where $\boldsymbol{B}$ (in $\boldsymbol{T}$) is the *magnetic flux density*, μ_0 is the vacuum permeability (in $\frac{N}{A^2}$) and $\boldsymbol{H}$ is the *magnetic field* (in $\frac{A}{m}$). Magnetic materials can be classified based on the dependence of $\boldsymbol{m}$ on $\boldsymbol{H}$.
In *linear* magnetic materials there is a constant *magnetic susceptibility* $\chi_m \in \mathbb{R}^{3\times 3}$ (dimensionless) such that $\tilde{\boldsymbol{m}} = \chi_m \boldsymbol{H}$. If χ_m is positive definite, the material is *paramagnetic*. If χ_m is negative definite, the material is *diamagnetic*. Paramagnetic and diamagnetic materials are (unintuitively) called *non-magnetic* because they have no magnetization without an external magnetic field.

Beyond linear materials, it may even be impossible to write the magnetization as a function of the magnetic field. This is the case of *magnetic hysteresis*, in which the magnetization depends on the past time evolution of the magnetic field [BM05]. Concrete specimens of magnetic materials often present a domain pattern: Each domain has a relatively uniform magnetization and they are separated by interfaces called *walls*. These structures have a width of order $10nm$ and are classified based on how the magnetization transitions. For example, in *Bloch walls*, transition proceeds perpendicularly to the transition axis. In *Néel walls*, the rotation proceeds in the plane spanned by the transition axis. More complex wall structures are also known [KP06].

1.3.2 The (deterministic) Landau–Lifshitz–Gilbert equation (LLG)

Each magnetic material has a *Curie temperature* (in K), below which the magnetization has constant modulus $|\tilde{\boldsymbol{m}}|_2 = \sqrt{\tilde{m}_1^2 + \tilde{m}_2^2 + \tilde{m}_3^2} = M_s$ called the *saturation magnetization*. We consider materials below their Curie temperature only and study the *normalized magnetization* (adimensional) $\boldsymbol{m} = \frac{\tilde{\boldsymbol{m}}}{M_s}$.

Consider a bounded, connected domain with Lipschitz boundary $D \subset \mathbb{R}^3$ representing the magnetic body, and $T > 0$ the final observation time (not to be confused with $T \in \mathcal{T}_h$, i.e. a mesh element, as will appear often in Chapter 2). We denote $D_T \coloneqq [0, T] \times D$ the *space-time cylinder* and $\partial D \subset \mathbb{R}^3$ the boundary of D. The (normalized) magnetization of the magnetic body is represented by the vector field

$$\boldsymbol{m} : D_T \to \mathbb{S}^2, \tag{1.49}$$

where $\mathbb{S}^2 \coloneqq \left\{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}|_2 = 1\right\}$ denotes the unit sphere in 3D.

The behavior of the magnetic body D is governed by a *Gibbs free energy* $\mathcal{E} = \mathcal{E}(\boldsymbol{m})$ (in J) in the form:

$$\mathcal{E} = \mathcal{E}_{\text{ex}} + \mathcal{E}_{\text{ani}} + \mathcal{E}_{\text{ms}} + \mathcal{E}_{\text{ext}}, \tag{1.50}$$

where each summand accounts for a physical effect:

- *Exchange energy*: $\mathcal{E}_{\text{ex}} = C_{\text{ex}} \int_D |\nabla \boldsymbol{m}|^2$, where $C_{\text{ex}} > 0$ (in $\frac{J}{m}$) is the *exchange stiffness constant*. This energy contribution penalizes non-constant magnetization configurations;

- *Anisotropy energy*: $\mathcal{E}_{\text{ani}} = \int_D \phi(\boldsymbol{m})$, which is related to the crystalline structure of the magnetic body. $\phi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ (in $\frac{J}{m^3}$) attains its minimum in a zero measure subset of $\S^2$. This energy contribution penalizes magnetization configurations that are not aligned with one or several preferred directions called *easy axis*;

- *Magnetostatic energy*: $\mathcal{E}_{\text{ms}} = \frac{\mu_0}{2} \int_{\mathbb{R}^3} |\nabla u|^2$, where ∇u is the magnetic field solution of the magnetostatic Maxwell equations:

$$\operatorname{div}(-\mu_0 \nabla u + \boldsymbol{m}\chi_D) = 0 \qquad \text{on } \mathbb{R}^3,$$

  where χ_D is the characteristic function of D. Observe that $\nabla \times \nabla u = 0$ by construction if u is twice differentiable. In this setting, ∇u is called *stray field* or *demagnetizing field* because it tends to reduce the total magnetic moment;

- *Zeeman energy*: $\mathcal{E}_{\text{ext}} = -\mu_0 M_s \int_D \boldsymbol{H}_{\text{ext}} \cdot \boldsymbol{m}$, where $\boldsymbol{H}_{\text{ext}}$ (in $\frac{A}{m}$) is an external magnetic field independent of the magnetization;

- *Antisymmetric exchange energy, or Dzyaloshinskii–Moriya interaction (DMI)*: $\mathcal{E}_{\mathrm{DM}} = \int_D \boldsymbol{D} : (\nabla \boldsymbol{m} \times \boldsymbol{m})$ where $\boldsymbol{D} : D \to \mathbb{R}^{3\times 3}$ is the *spiralization tensor* (in $\frac{J}{m^2}$). This energy contribution explains the formation of *magnetic skyrmions*, i.e. topologically protected vortex-like magnetization configurations (see [BY89] for their theoretical prediction and [MBJ+09, RHM+13] for their observations in experiments).

The *magnetostatics* problem consists of determining a magnetization within an appropriate functions class $\mathcal{M}$ of unit-modulus vector fields which minimizes the Gibbs free energy:

$$\min_{\boldsymbol{m}\in\mathcal{M}} \mathcal{E}(\boldsymbol{m}).$$

Recall that this model is valid at *microscopic* length scale, i.e. from 1 μm down to $10nm$. For the smaller atomic length scale, statistical physics models are used to describe the interaction of spins. For the larger mesoscopic scale (up to mm size), the unknown averaged magnetization is the solution of a degenerate convex, non-local, variational problem [KP06].

Another important task is the one of determining the time evolution of the magnetization. If the Gibbs free energy presents several minima, the only way to predict the stationary configuration reached by a magnetic system with a given initial condition is to simulate the dynamics. In analogy to the equations of motions of the electron, in micromagnetics one may write

$$\partial_t \boldsymbol{m} = -\gamma_0 \boldsymbol{m} \times \boldsymbol{H}_{\mathrm{eff}}(\boldsymbol{m}).$$

Here, $\gamma_0 > 0$ (in $\frac{m}{As}$) is called *rescaled gyromagnetic ratio* and $\boldsymbol{H}_{\mathrm{eff}}$ (in $\frac{A}{m}$) is called *effective field* and is defined by the relation

$$\mu_0 M_s \boldsymbol{H}_{\mathrm{eff}}(\boldsymbol{m}) = \frac{\delta \mathcal{E}(\boldsymbol{m})}{\delta \boldsymbol{m}}, \tag{1.51}$$

i.e., up to a constant, the Frechet derivative of the Gibbs free energy (1.50). In the simplest form of the equation, the *small-particle limit*, one considers $\mathcal{E} = \mathcal{E}_{\mathrm{ex}}$ and obtains $\boldsymbol{H}_{\mathrm{eff}}(\boldsymbol{m}) = 2C_{\mathrm{ex}}\Delta \boldsymbol{m}$.

However, this model does not describe the dissipation observed in experiments. In 1935, Landau and Lifshitz introduced an additional dissipation term and defined what we now call *Landau-Lifshitz equation* (LL) (cf. [LL35])

$$\partial_t \boldsymbol{m} = -\gamma_0 \boldsymbol{m} \times \boldsymbol{H}_{\mathrm{eff}}(\boldsymbol{m}) - \gamma_0 \tilde{\lambda} \boldsymbol{m} \times (\boldsymbol{m} \times \boldsymbol{H}_{\mathrm{eff}}(\boldsymbol{m})), \tag{LL}$$

where $\tilde{\lambda} > 0$ is a non-dimensional phenomenological parameter. The additional term with the double cross product is chosen for purely phenomenological reasons: The dissipation should drive the magnetization towards the direction of the effective field (a configuration of minimal magnetization energy) while the magnetization magnitude stays constant. One should note that the last term has the physical behavior of a damping only if $\lambda < 1$, otherwise the relaxation becomes faster as λ increases, which is in strong contradiction with the physical idea of damping. Again, (LL) is an appropriate model for length scales from $10nm$ to $1\mu m$ (see [KP06, Section 1, p. 444]).

The LL equation is equipped with the following boundary conditions and initial condition

$$\begin{aligned} \partial_n \boldsymbol{m} &= 0 \qquad \text{on } [0,T]\times \partial D, \\ \boldsymbol{m}(0) &= \boldsymbol{m}^0 \qquad \text{on } D. \end{aligned}$$

The initial condition can be assumed without loos of generality to have unit modulus: For a.e. $\boldsymbol{x} \in D$, $\left|\boldsymbol{m}^0(\boldsymbol{x})\right|_2 = 1$.

Elementary proprieties of the cross product imply that $\partial_t \boldsymbol{m}(t,\boldsymbol{x})$ and $\boldsymbol{m}(t,\boldsymbol{x})$ are orthogonal, i.e. $\partial_t \boldsymbol{m}(t,\boldsymbol{x}) \cdot \boldsymbol{m}(t,\boldsymbol{x}) = 0$ for all $(t,\boldsymbol{x}) \in D_T$. This implies that

$$\frac{\mathrm{d}}{\mathrm{d}t} \left|\boldsymbol{m}(t,\boldsymbol{x})\right|_2 = 2\boldsymbol{m} \cdot \partial_t \boldsymbol{m} = 0,$$

so that for all $0 \leq t \leq T$, for a.e. $\boldsymbol{x} \in D$, $\left|\boldsymbol{m}(t,\boldsymbol{x})\right|_2 = \left|\boldsymbol{m}^0(\boldsymbol{x})\right|_2 = 1$. This justifies $\boldsymbol{m}$ taking values in $\mathbb{S}^2$ like $\boldsymbol{m}^0$. Therefore, the model is consistent with condition (1.49).

As example of exact solution to the LLG equation is given in [KP06, example 3.4] in the form of a flat domain wall traveling at constant velocity along the x axis.

Equation (LL) has several useful equivalent forms. The *Landau-Lifshitz-Gilbert equation* (LLG) was derived by Gilbert in 1955 (cf. [Gil55])

$$\partial_t \boldsymbol{m} = -\boldsymbol{m} \times \boldsymbol{H}_{\text{eff}}(\boldsymbol{m}) + \lambda \boldsymbol{m} \times \partial_t \boldsymbol{m}, \tag{LLG}$$

where $\lambda > 0$ is called *Gilbert damping*. For most materials of interest, $\lambda < 10^{-2}$ [dSC+06]. This formulation describes more effectively the damping effect by writing the corresponding term as a function of $\partial_t \boldsymbol{m}$

The *alternative form* (LLA), important in numerical analysis, reads

$$\lambda \partial_t \boldsymbol{m} + \boldsymbol{m} \times \partial_t \boldsymbol{m} = \boldsymbol{H}_{\text{eff}}(\boldsymbol{m}) - \left(\boldsymbol{H}_{\text{eff}}(\boldsymbol{m}) \cdot \boldsymbol{m}\right) \boldsymbol{m} \tag{LLA}$$

The equivalence of the three formulations, LL, LLG, and LLA, is proved by the following implications:

- $\boldsymbol{m}$ *solves* (LLA) $\Rightarrow$ $\boldsymbol{m}$ *solves* (LLG):
  Apply the operator $\boldsymbol{m} \times \cdot$ to both sides of (LLA) to obtain

  $$\lambda \boldsymbol{m} \times \partial_t \boldsymbol{m} + \boldsymbol{m} \times (\boldsymbol{m} \times \partial_t \boldsymbol{m}) = \boldsymbol{m} \times \boldsymbol{H}_{\text{eff}}.$$

  Then, use the triple cross product formula (B.2), the unit-modulus of $\boldsymbol{m}$, and orthogonality of $\boldsymbol{m}$ and $\partial_t \boldsymbol{m}$ to get:

  $$\boldsymbol{m} \times (\boldsymbol{m} \times \partial_t \boldsymbol{m}) = -\partial_t \boldsymbol{m}. \tag{1.52}$$

  Substitution in the previous equation gives (LLG);

- $\boldsymbol{m}$ *solves* (LL) $\Rightarrow$ $\boldsymbol{m}$ *solves* (LLA):
  Apply the triple cross product formula (B.2) and the unit modulus of $\boldsymbol{m}$ to the last term of (LL) to obtain

  $$\boldsymbol{m} \times (\boldsymbol{m} \times \boldsymbol{H}_{\text{eff}}) = \boldsymbol{m} \left(\boldsymbol{m} \cdot \boldsymbol{H}_{\text{eff}}\right) - \boldsymbol{H}_{\text{eff}}. \tag{1.53}$$

  For (LL), this implies

  $$\partial_t \boldsymbol{m} = -\gamma_0 \boldsymbol{m} \times \boldsymbol{H}_{\text{eff}}(\boldsymbol{m}) - \gamma_0 \tilde{\lambda} \left(\boldsymbol{m} \left(\boldsymbol{m} \cdot \boldsymbol{H}_{\text{eff}}\right) - \boldsymbol{H}_{\text{eff}}\right).$$

  Applying the operator $\boldsymbol{m} \times \cdot$ gives

  $$\boldsymbol{m} \times \partial_t \boldsymbol{m} = \gamma_0 \boldsymbol{m} \times (\boldsymbol{m} \times \boldsymbol{H}_{\text{eff}}) + \gamma_0 \tilde{\lambda} \boldsymbol{m} \times \boldsymbol{H}_{\text{eff}}.$$

  Finally, multiply the second-to-last equation by λ and sum it to the last one to obtain (LLA) with $\tilde{\lambda} = \lambda$ and $\gamma_0 = \frac{1}{1+\alpha^2}$;

- $\boldsymbol{m}$ *solves* (LLG) $\Rightarrow$ $\boldsymbol{m}$ *solves* (LL):
  Apply the operator $\boldsymbol{m}\times\cdot$ to (LLG), followed by (1.52) to obtain

$$\boldsymbol{m}\times\partial_t\boldsymbol{m} = -\boldsymbol{m}\times(\boldsymbol{m}\times\boldsymbol{H}_{\text{eff}}) - \alpha\partial_t\boldsymbol{m}.$$

  Then, multiply it by λ and sum it with (LLG) itself. We obtain (LL) with $\gamma_0 = \frac{1}{1+\alpha^2}$ and $\tilde{\lambda} = \lambda$.

Let us also rewrite the Landau-Lifshitz equation (LL) expressing the scalar coefficients as a function of the Gilbert damping parameter λ:

$$\partial_t\boldsymbol{m} = \frac{1}{1+\lambda^2}\boldsymbol{m}\times\boldsymbol{H}_{\text{eff}}(\boldsymbol{m}) - \frac{\lambda}{1+\lambda^2}\boldsymbol{m}\times(\boldsymbol{m}\times\boldsymbol{H}_{\text{eff}}(\boldsymbol{m}))\,.$$

We now obtain a *weak formulation* of the problem. From now on, Lebesgue and Sobolev spaces of vector-valued functions are denoted e.g. $\mathbb{L}^2(D) = L^2(D)^3$ and $\mathbb{H}^1(D) = H^1(D)^3$. Consider a test function $\boldsymbol{\psi}\in C_0^\infty(D_T)^3$, i.e. an infinitely-differentiable vector field with compact support. Testing equation (LLG) with $\boldsymbol{\psi}$, we obtain

$$\int_0^T \langle\partial_t\boldsymbol{m},\boldsymbol{\psi}\rangle = \int_0^T \langle -\boldsymbol{m}\times\boldsymbol{H}_{\text{eff}} + \lambda\boldsymbol{m}\times\partial_t\boldsymbol{m},\boldsymbol{\psi}\rangle\,.$$

where by $\langle\cdot,\cdot\rangle$ we denote the $\mathbb{L}^2(D)$ scalar product: $\langle\boldsymbol{f},\boldsymbol{g}\rangle = \int_D \boldsymbol{f}\cdot\boldsymbol{g}$ for all $\boldsymbol{f},\boldsymbol{g}\in\mathbb{L}^2(D)$.

Depending on the definition of $\boldsymbol{H}_{\text{eff}}$, different weak formulations can be considered. Let us consider the so called *small particle limit* in which $\boldsymbol{H}_{\text{eff}} = \Delta\boldsymbol{m}$. The scalar triple product (B.1) gives $\langle\boldsymbol{m}\times\Delta\boldsymbol{m},\boldsymbol{\psi}\rangle = \langle\Delta\boldsymbol{m},\boldsymbol{\psi}\times\boldsymbol{m}\rangle$. Integration by parts, recalling the homogeneous Neumann boundary condition, gives $\langle\Delta\boldsymbol{m},\boldsymbol{\psi}\times\boldsymbol{m}\rangle = \langle\nabla\boldsymbol{m},\nabla\boldsymbol{\psi}\times\boldsymbol{m}\rangle$. Another application of (B.1) finally yields

$$\int_0^T \langle\partial_t\boldsymbol{m},\boldsymbol{\psi}\rangle = -\int_0^T \langle\boldsymbol{m}\times\nabla\boldsymbol{m},\nabla\boldsymbol{\psi}\rangle + \lambda\langle\boldsymbol{m}\times\partial_t\boldsymbol{m},\boldsymbol{\psi}\rangle\,. \tag{1.54}$$

Once a space for $\boldsymbol{m}$ is chosen, a density argument can be used to select $\boldsymbol{\psi}$ in a larger space than the one of test-functions, as long as the previous quantities are finite. Adherence to initial and boundary conditions must also be suitably defined. The following is a natural choice, as given in [AS92].

Definition 1.22 (Weak solution of LLG in the small-particle limit)**.** *Given* $\boldsymbol{m}^0\in\mathbb{H}^1(D)$ *such that* $|\boldsymbol{m}^0| = 1$ *a.e. in* D*, a function* $\boldsymbol{m}: D_T\to\mathbb{R}^3$ *is a* weak solution of the LLG equation *if and only if*

- $\boldsymbol{m}\in\mathbb{H}^1(D_T)$ *(i.e.* $\boldsymbol{m}\in L^2(0,T,\mathbb{H}^1(D))$*,* $\partial_t\boldsymbol{m}\in L^2(0,T,\mathbb{L}^2(D))$*) and* $|\boldsymbol{m}| = 1$ *a.e. in* D_T*;*
- $\boldsymbol{m}$ *solves* (1.54) *for all* $\boldsymbol{\psi}\in L^2(0,T,\mathbb{H}^1(D))$*;*
- $\boldsymbol{m}(0,\cdot) = \boldsymbol{m}^0$ *on* D *and* $\partial_n\boldsymbol{m} = 0$ *on* $[0,T]\times\partial D$ *both in the sense of traces;*
- *The following energy bound holds:*

$$\frac{1}{2}\|\nabla\boldsymbol{m}(t)\|^2_{\mathbb{L}^2(D)} + \frac{\lambda}{1+\lambda^2}\int_0^t \|\partial_t\boldsymbol{m}\|_{\mathbb{L}^2(D)} \le \frac{1}{2}\|\nabla\boldsymbol{m}^0\|^2_{\mathbb{L}^2(D)} \qquad \forall t\in[0,T]. \tag{1.55}$$

The well-posedness of the weak LLG equation has been discussed in several works and is in some aspects not closed. Let us review some of the most important results. In the seminal paper [AS92, Theorem 1.5], the authors prove existence of weak solutions in the small-particle limit. In the same work (Theorem 1.6), a *non-uniqueness* result is also proved: There exists $\boldsymbol{m}^0 \in \mathbb{H}^1(D)$ with $\left|\boldsymbol{m}^0\right| \equiv 1$ a.e. in D such that there exist infinitely many (a continuum indexed by a real parameter) distinct solutions in the sense of Definition 1.22. This fact and its proof are analogous to the case of weak solutions of *harmonic maps heat flows into the sphere* (i.e. equations in the form (LLG) up to the term $\boldsymbol{m} \times \Delta \boldsymbol{m}$). Non-uniqueness limits the available convergence results of approximation schemes. Often finite elements approximations converge only weakly and up to a subsequence.

For the LLG equation, singularities may arise at times $t \in [0, T]$ when $\|\nabla \boldsymbol{m}(t)\|_{L^\infty(D)^3}$ is not finite (recall that by definition a weak solution is only L^2 integrable in time). In physical terms, finite time blow-ups translate into spatial energy concentrations and defects (see [KP06]).

From the mathematical point of view, the phenomenon of finite time blowup was studied e.g. in [WZZ22]. Here, the authors construct a solution to the 2D LLG equation that blows up at pre-determined points in space. In particular, they prove an asymptotic behavior $\|\nabla \boldsymbol{m}\|_{\mathbb{L}^\infty(D)} \approx \frac{|\log(T-t)|^2}{T-t}$ for times approaching the singular time $T > 0$.

In the work [CF01a], the authors prove existence and uniqueness of *strong* solutions of LLG in
$C^0(0, T, \mathbb{H}^2(D)) \cap L^2(0, T, \mathbb{H}^3(D))$ for small times. In 2 dimensions and under an additional smallness assumption on the initial condition, the result holds globally in time.

In [FT17b], the authors prove existence and uniqueness of arbitrarily regular strong solution of LLG (in the small particle limit) under the assumption that the initial condition is also regular and sufficiently close to a constant. Solutions belong to parabolic Sobolev spaces $\mathbb{H}^{k,2k}(D_T)$ for $k \geq 3$ and the their norm depends continuously on the $\mathbb{H}^{2k}(D)$ norm of the initial condition.

A weak-strong uniqueness principle was proved in [DFIP20, DS14]. This means that when a problem admits both a weak and strong solution, then they coincide (for any time for which both are defined).

Other well-posedness results are available for $D = \mathbb{R}^d$ and $d \in \{2, 3\}$ or periodic boundary conditions. See [CDG98, Mel05, Mos05, LLW15, Mel12, Cim07].

1.3.3 Numerical methods for LLG

Consider a mesh $\mathcal{T}_h$ of the open, connected, and Lipschitz domain $D \subset \mathbb{R}^3$ with element size $h > 0$. Denote by $\mathcal{N}_h \subset \overline{D}$ the (finite) set its vertices. Denote by $\boldsymbol{V}_h \coloneqq \mathcal{S}^1(\mathcal{T}_h)^3$, the space of continuous piecewise affine functions on $\mathcal{T}_h$. To approximate a function $\boldsymbol{v} \in C^0(D)^3$ in $\boldsymbol{V}_h$ we can consider:

- *Nodal interpolation* $\mathfrak{I}_h : C^0(D)^3 \to \boldsymbol{V}_h$: $\mathfrak{I}_h[\boldsymbol{v}] \in \boldsymbol{V}_h$ is the unique function such that
$$\mathfrak{I}_h[\boldsymbol{v}](x) = \boldsymbol{v}(x) \qquad \forall \boldsymbol{x} \in \mathcal{N}_h; \tag{1.56}$$
- *L^2-projection* $\Pi_h : C^0(D)^3 \to \boldsymbol{V}_h$: $\Pi_h[\boldsymbol{v}] \in \boldsymbol{V}_h$ is the unique function such that
$$\langle \boldsymbol{v} - \Pi_h[\boldsymbol{v}], \boldsymbol{\phi}_h \rangle_{L^2(D)^3} = 0 \qquad \forall \boldsymbol{\phi}_h \in \boldsymbol{V}_h. \tag{1.57}$$

Let $J \in \mathbb{N}$ and $\tau = \frac{T}{J}$ to define timesteps $t_j \coloneqq j\tau$ for $j = 0, \dots, J$.
We discuss all the following schemes for LLG in the small participle limit, i.e. $\boldsymbol{H}_{\text{eff}} = \Delta \boldsymbol{m}$. This is done for simplicity and can always be generalized.

Midpoint rule

We present the *midpoint scheme* as defined and analysed in [BP06].
Denote the *mass-lumped scalar product*

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_h \coloneqq \int_D \mathfrak{I}_h [\boldsymbol{u}\boldsymbol{v}] \tag{1.58}$$

and the *discrete Laplace operator* $\Delta_h : \mathbb{H}^1(D) \to \boldsymbol{V}_h$ such that

$$-\langle \Delta_h \boldsymbol{u}, \boldsymbol{v}_h \rangle_h = \langle \nabla \boldsymbol{u}, \nabla \boldsymbol{v}_h \rangle \qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_h. \tag{1.59}$$

Consider additionally the *discrete time derivative* and *midpoint* operators respectively

$$\mathrm{d}_t \boldsymbol{v}^j \coloneqq \frac{\boldsymbol{v}^j - \boldsymbol{v}^{j-1}}{\tau} \qquad \forall j = 1, \dots, J, \tag{1.60}$$

$$\boldsymbol{v}^{j+\frac{1}{2}} = \frac{\boldsymbol{v}^{j+1} + \boldsymbol{v}^j}{2} \qquad \forall j = 0, \dots, J-1. \tag{1.61}$$

The midpoint rule is based on the Gilbert formulation (LLG) and reads:

Algorithm 1 $\left(\boldsymbol{m}_h^j\right)_{j=0}^J \leftarrow$ `Midpoint rule`$(\mathcal{T}_h, \boldsymbol{m}^0, J, \lambda)$

1: Compute $\boldsymbol{m}_h^0 = \Pi_h \boldsymbol{m}^0$
2: **for** $j = 0, 1, ..., J-1$ **do**
3: Find $\boldsymbol{m}_h^{j+1} \in \boldsymbol{V}_h$ such that

$$\left\langle \mathrm{d}_t \boldsymbol{m}_h^{j+1} + \lambda \boldsymbol{m}_h^j \times \mathrm{d}_t \boldsymbol{m}_h^{j+1}, \boldsymbol{\phi}_h \right\rangle_h = \left\langle \boldsymbol{m}_h^{j+\frac{1}{2}} \times \Delta_h \boldsymbol{m}_h^{j+\frac{1}{2}}, \boldsymbol{\phi}_h \right\rangle_h \qquad \forall \boldsymbol{\phi}_n \in \boldsymbol{V}_h.$$

4: **end for**

Observe that the second term is linear i $\boldsymbol{m}_h^{j+1}$ and motivated by the following identity:

$$\boldsymbol{m}_h^j \times \mathrm{d}_t \boldsymbol{m}_h^{j+1} = \left(\boldsymbol{m}_h^{j+\frac{1}{2}} - \frac{\tau}{2} \mathrm{d}_t \boldsymbol{m}_h^{j+1} \right) \times \mathrm{d}_t \boldsymbol{m}_h^{j+1} = \boldsymbol{m}_h^{j+\frac{1}{2}} \times \mathrm{d}_t \boldsymbol{m}_h^{j+1}.$$

The method requires the solution of a nonlinear system at each timestep. This can be done e.g. with a fixed-point iteration but convergence is guaranteed only under the CFL-type condition $\tau = \mathcal{O}(h^2)$.

By testing with $\boldsymbol{\phi}_h = \boldsymbol{m}_h^{j+1}(\boldsymbol{x})\boldsymbol{\phi}_{\boldsymbol{x}}$, where $\boldsymbol{\phi}_{\boldsymbol{x}}$ denotes the hat basis function of $\mathcal{S}^1(\mathcal{T}_h)$ corresponding to $\boldsymbol{x} \in \mathcal{N}_h$, one sees that $\left|\boldsymbol{m}_h^{j+1}(\boldsymbol{x})\right| = \left|\boldsymbol{m}_h^j(\boldsymbol{x})\right|$ for all $\boldsymbol{x} \in \mathcal{N}_h$ and all $0 \le j < J$, so that the scheme automatically produces numerical solution with unit modulus in the vertices.

Moreover, testing with $\boldsymbol{\phi}_h = -\Delta_h \boldsymbol{m}_h^{j+\frac{1}{2}} + \lambda \mathrm{d}_t \boldsymbol{m}_h^{j+1}$ and summing over $j = 0, \dots, J-1$ one obtains the following discrete energy conservation:

$$\frac{1}{2} \left\| \nabla \boldsymbol{m}_h^J \right\|^2 + \lambda \tau \sum_{j=0}^{J-1} \left\| \mathrm{d}_t \boldsymbol{m}_h^{j+1} \right\|_h^2 = \frac{1}{2} \left\| \nabla \boldsymbol{m}_h^0 \right\|^2,$$

which mimics the continuous energy estimate in Equation (1.55).

Tangent plane scheme

We present the *tangent plane scheme*, proposed and analysed in [AJ06].
Consider the set of normalized discrete fields

$$\mathcal{M}_h \coloneqq \{\boldsymbol{\phi}_h \in \boldsymbol{V}_h : |\boldsymbol{\phi}_h(\boldsymbol{x})|_2 = 1 \ \forall \boldsymbol{x} \in \mathcal{N}_h\}$$

and the *discrete tangent plane* to $\boldsymbol{m}_h \in \mathcal{M}_h$ defined imposing orthogonality on the mesh vertices only:

$$\mathcal{K}_h(\boldsymbol{m}_h) \coloneqq \{\boldsymbol{\phi}_h \in \boldsymbol{V}_h : \boldsymbol{\phi}_h(\boldsymbol{x}) \cdot \boldsymbol{m}_h(\boldsymbol{x}) = 0 \ \forall \boldsymbol{x} \in \mathcal{N}_h\}\,. \tag{1.62}$$

We start from the alternative formulation (LLA) and consider test functions that are in a sense orthogonal to the magnetization at current time. Thus, it makes sense to ignore the nonlinear term $(\Delta \boldsymbol{m} \cdot \boldsymbol{m})\boldsymbol{m}$.

The algorithm reads:

Algorithm 2 $\left(\boldsymbol{m}_h^j\right)_{j=0}^{J} \leftarrow$ `Tangent plane scheme`$(\mathcal{T}_h, \boldsymbol{m}^0, J, \lambda)$

1: Compute $\boldsymbol{m}_h^0 = \Pi_h \boldsymbol{m}^0$
2: **for** For $j = 0, 1, ..., J-1$ **do**
3: Find $\boldsymbol{v}_h^j \in \mathcal{K}_h(\boldsymbol{m}_h^j)$ such that

$$\lambda \left\langle \boldsymbol{v}_h^j + \boldsymbol{m}_h^j \times \boldsymbol{v}_h^j, \boldsymbol{\phi}_h \right\rangle = \left\langle \nabla \left(\boldsymbol{m}_h^j + \theta\tau \boldsymbol{v}_h^j \right), \nabla \boldsymbol{\phi}_h^j \right\rangle \qquad \forall \boldsymbol{\phi}_n \in \mathcal{K}_h(\boldsymbol{m}_h^j)$$

4: $\boldsymbol{m}_h^{j+1} = \mathfrak{I}_h \left[\frac{\boldsymbol{m}_h^j + \tau \boldsymbol{v}_h^j}{|\boldsymbol{m}_h^j + \tau \boldsymbol{v}_h^j|} \right]$
5: **end for**

Observe that at each timestep only one linear problem in $\boldsymbol{v}_h^j$ must be solved. The final step is a renormalization (a nonlinear operation) to guarantee that the discrete solution satisfies a discrete equivalent of the unit modulus condition: $\left|\boldsymbol{m}_h^j(\boldsymbol{x})\right| = 1$ for all $\boldsymbol{x} \in \mathcal{N}_h$. However, this step can be omitted and the numerical solution will have unit modulus in the limit for vanishing h and τ. If the normalization is included, then it was proved in [Bar05, Lemma 13] that convergence holds under an *angle condition* on the mesh, i.e. if the off-diagonal terms of the stiffness matrix are non-positive.

The scheme produces a sequence of approximations $\boldsymbol{m}_h^i$ of $\boldsymbol{m}(t_i, \cdot)$ for $i = 0, \ldots, J$. These can be interpolated to obtain an approximation for any time, thus yielding an approximation $\boldsymbol{m}_{\tau,h} : D_T \to \mathbb{R}^3$.

The sequence of the obtained solutions $(\boldsymbol{m}_{\tau,h})_{\tau,h}$ converges weakly in $\mathbb{H}^1(D)$, up to extraction of a subsequence, to the weak solution of the LLG equation (cf. Definition 1.22) when both h and τ vanish under the condition (see [Alo08, Theorem 2,3]):

$$\begin{cases} \tau = \mathcal{O}(h^2) & \text{if } \theta \leq \frac{1}{2}, \\ \text{no condition} & \text{if } \theta > \frac{1}{2}. \end{cases}$$

In [FT17a], the authors prove convergence with rates of the whole sequence of discrete solutions under additional regularity assumptions on the exact solution. In particular, they prove first order convergence in h and τ for the $L^2(0, T; \mathbb{H}^1(D))$ error. Results are also extended to the Maxwell-LLG system as well.

High-order linearly implicit BDF scheme

We present the method from [AFKL21] as an example of one of few known high-order schemes for LLG with a comprehensive convergence analysis (See the end of the section for another possible scheme).

Consider the "projection-type" tangent plane to $\boldsymbol{m}_n \in \boldsymbol{V}_h$ defined using the L^2 projection (cf. (1.57)).

$$\hat{\mathcal{K}}_h(\boldsymbol{m}_h) \coloneqq \{\boldsymbol{\phi}_h \in \boldsymbol{V}_h : \Pi_h [\boldsymbol{\phi}_h \cdot \boldsymbol{m}_h] = 0\}. \tag{1.63}$$

Observe that the orthogonality constraint is imposed weakly, as opposed to the condition used to define $\mathcal{K}_h(\boldsymbol{m}_h)$ in (1.62). Define, for $k \in \mathbb{N}$, $k \le J$, $k \le j \le J$, the following k-steps BDF-type approximations of time derivative and extrapolation respectively:

$$\dot{\boldsymbol{m}}_h^j \coloneqq \frac{1}{\tau} \sum_{\ell=0}^{k} \delta_\ell \boldsymbol{m}_h^{j-\ell}$$
$$\hat{\boldsymbol{m}}_h^j \coloneqq \frac{\sum_{\ell=0}^{k-1} \gamma_j \boldsymbol{m}_h^{j-\ell-1}}{\left|\sum_{\ell=0}^{k-1} \gamma_j \boldsymbol{m}_h^{j-\ell-1}\right|}.$$

The constant $\delta_1, \dots, \delta_k$ and $\gamma_0, \dots, \gamma_{k-1}$ are defined by

$$\sum_{j=0}^{k} \delta_j x^j = \sum_{\ell=1}^{k} \frac{1}{\ell}(1-x)^\ell, \quad \sum_{i=0}^{k} \gamma_i x^i = \frac{1}{x}(1-(1-x)^k).$$

The algorithm is similar to the tangent plane scheme and reads:

Algorithm 3 $\left(\boldsymbol{m}_h^j\right)_{j=0}^{J} \leftarrow$ `HOLIBDF`$(\mathcal{T}_h, \boldsymbol{m}^0, J, \lambda)$

1: Compute $\boldsymbol{m}_h^0 = \Pi_h \boldsymbol{m}^0$
2: Compute $\boldsymbol{m}_h^1, \dots, \boldsymbol{m}_h^{k-1}$
3: **for** For $j = k, k+1, J$ **do**
4: Find $\boldsymbol{v}_h^j \in \hat{\mathcal{K}}(\boldsymbol{m}_h^j)$ such that for all $\boldsymbol{\phi}_n \in \mathcal{K}(\hat{\boldsymbol{m}}_h^j)$

$$\lambda \left\langle \boldsymbol{v}_h^j + \hat{\boldsymbol{m}}_h^j \times \boldsymbol{v}_h^j, \boldsymbol{\phi}_h \right\rangle + \frac{\tau}{\delta_0} \left\langle \nabla \boldsymbol{v}_h^j, \nabla \boldsymbol{\phi}_h \right\rangle = \frac{1}{\delta_0} \left\langle \nabla \left(\sum_{\ell=1}^{k} \delta_\ell \boldsymbol{m}_h^{j-\ell} \right), \nabla \boldsymbol{\phi}_h \right\rangle \tag{1.64}$$

5: Compute $\boldsymbol{m}_h^j = \frac{\tau}{\delta_0} \boldsymbol{v}_h^j - \frac{1}{\delta_0} \sum_{\ell=1}^{k} \delta_\ell \boldsymbol{m}_h^{j-\ell}$
6: **end for**

The computation of $\boldsymbol{m}_h^1, \dots, \boldsymbol{m}_h^{k-1}$ must be carried out with a lower order scheme. The problem (1.64) is indeed a problem in $\boldsymbol{v}_h^j$ since the occurrence of $\boldsymbol{m}_h^j$ can be substituted using $\boldsymbol{v}_h^j = \dot{\boldsymbol{m}}_h^j$. The same consideration justifies the final step of the loop above. Convergence in the $C^0([0,T], \mathbb{H}^1(D))$ norm is proved under regularity assumptions on the exact solution. For $k = 1, 2$ the following discrete energy estimate is available because of the use of BDF time integration:

$$\gamma_k^- \left\| \nabla \boldsymbol{m}_h^j \right\|^2 + \frac{\lambda\tau}{2} \sum_{\ell=k}^{j} \left\| \dot{\boldsymbol{m}}_h^\ell \right\|^2 \le \gamma_k^+ \sum_{\ell=0}^{k-1} \left\| \nabla \boldsymbol{m}_h^\ell \right\|^2.$$

with $\gamma_1^{\pm} = 1$ and $\gamma_2^{\pm} = \frac{3\pm 2\sqrt{2}}{4}$. For $k = 3, 4, 5$, convergence is guaranteed only under the (mild) CFL-type condition $\tau = \mathcal{O}(h)$ as well as restrictions to how small the Gilbert damping λ is allowed to be depending on the value of k. No discrete energy estimate is available in this case.

We also briefly mention the recent work [BKW23], in which harmonic map heat flows into the sphere (a simplification of LLG with $\boldsymbol{H}_{\text{eff}} = -\Delta \boldsymbol{m}$ obtained neglecting the term $\boldsymbol{m} \times \Delta \boldsymbol{m}$) are discretized by an analogous method. An important difference is that the orthogonality constraint is not imposed weakly through the projection-type tangent plane $\hat{\mathcal{K}}_h$, but in a pointwise fashion , i.e. using $\mathcal{K}_h(\cdot)$ as in the "classical" tangent plane scheme (see Section 1.3.3). Analysis only for order one in space and time, but convergence for order up to five in space and time (sung BDF in time) can be proved.

1.3.4 Thermal noise perturbations the stochastic LLG equation (SLLG)

When a magnetic body is placed in a heat bath with temperature below the Curie temperature of the magnetic material, the magnetization exhibits random fluctuations while preserving a constant modulus. As a consequence, the magnetization may transition between equilibrium states without any external force (e.g. external magnetic fields). Interest in modeling this phenomenon increased in the physics community during the second half of the twentieth century, see e.g. [BJ63, KH70] for some of the first works on the topic and [GPL98, KRVE05, SSF01, Ber07, MBS09] for more recent treatments.

The relevance of modeling small-scale (sub-micrometer) magnetic bodies comes from engineering applications e.g. magnetic storage devices (see also Section 1.3.8).

Let us introduce a mathematical model for micromagnetics with thermal noise. The following is based on [BBNP14b, Chapter 2] and [BGJ12].
Denote again by $D \subset \mathbb{R}^3$ a bounded Lipschitz domain representing the magnetic body and $T > 0$ a final observation time. Consider a probability triple $(\Omega, \mathcal{E}, \mathcal{P})$. The contribution of random noise to the governing equation is given by a space-time white noise $\mathrm{d}\boldsymbol{W}$, where $\boldsymbol{W} = \boldsymbol{W}(\omega, t, \boldsymbol{x})$ denotes a space-time Wiener process and differentiation is understood in a formal sense. The magnetization

$$\boldsymbol{M} : \Omega \times D_T \to \mathbb{S}^2$$

is now a *random field* because of the thermal noise appearing in the governing SLLG equation:

$$\partial_t \boldsymbol{M} = \lambda_1 \boldsymbol{M} \times (\boldsymbol{H}_{\text{eff}} + \mathrm{d}\boldsymbol{W}) - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{H}_{\text{eff}}), \tag{1.65}$$

obtained from the Landau-Lifshitz equation (LL) by adding $\mathrm{d}\boldsymbol{W}$ to the first occurrence of $\boldsymbol{H}_{\text{eff}}$ (this is not done for the second because λ_2 is usually smaller than λ_1). As in the LLG equation (LLG), $\boldsymbol{H}_{\text{eff}} = \boldsymbol{H}_{\text{eff}}(\boldsymbol{m})$ is the effective field defined as a sum of functions of $\boldsymbol{m}$ with leading order term $\Delta \boldsymbol{m}$.

Appropriate boundary and initial conditions also have to be enforced:

$$\begin{aligned} \partial_{\boldsymbol{n}} \boldsymbol{M} &= 0 \qquad \text{on } [0,T] \times \partial D, \\ \boldsymbol{M}(0) &= \boldsymbol{M}^0 \qquad \text{on } D. \end{aligned}$$

We highlight the fact that the solution $\boldsymbol{M}$ of (1.65) still satisfies the unit modulus condition $|\boldsymbol{M}(\omega, t, x)| = 1$ $\mathcal{P}$-a.s. and for all $(t, \boldsymbol{x}) \in D_T$.

The precise mathematical model for the noise, i.e. the specific definition of $\boldsymbol{W}$, is an ongoing topic of research (see Remark 1.23 below). For simplicity, in what follows we use

$$\boldsymbol{W}(\omega, t, \boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x}) W(\omega, t), \tag{1.66}$$

where $\boldsymbol{g} : D \to \mathbb{R}^3$ is given and is assumed to be of unit modulus a.e. in D and $W : \Omega \times [0,T] \to \mathbb{R}$ is a (scalar) Wiener process in t only. The regularity we require of $\boldsymbol{g}$ is:

$$\begin{aligned} \boldsymbol{g} &\in \mathbb{W}^{2,\infty}(D), \\ \partial_{\boldsymbol{n}} \boldsymbol{g} &= 0 \quad \text{on } \partial D, \\ |\boldsymbol{g}| &= 1 \qquad \text{on } D. \end{aligned} \tag{1.67}$$

Remark 1.23 (On noise modeling)**.** *The ansatz* (1.66) *is called* scalar *or* 1-dimensional *as opposed to the* multi-dimensional *noise* $\boldsymbol{W} = \sum_{n\in\mathbb{N}} a_n W_n(t) \boldsymbol{g}_n(\boldsymbol{x})$ *where* $(W_n(t))_{n\in\mathbb{N}}$ *are independent Wiener processes,* $(a_n)_{n\in\mathbb{N}}$ *is a real sequence and* $(\boldsymbol{g}_n)_{n\in\mathbb{N}}$ *is a sequence of functions of* $\boldsymbol{x} \in D$*. If for example* $\boldsymbol{W}(\omega, t, \cdot) \in \mathbb{L}^2(D)$*, then* $(\boldsymbol{g}_n)_{n\in\mathbb{N}}$ *could be chosen as an orthonormal basis of* $\mathbb{L}^2(D)$*.*
It was argued in [Ber07, Sections 3.2, 3.3] that space-time white noise leads to a model that best describes physical phenomenology. This corresponds to a noise that is uncorrelated in both space and time, i.e. $\mathbb{E}[\boldsymbol{W}(s,x)\boldsymbol{W}(t,y)] = \delta(s-t)\delta(x-y)$ *or, equivalently, to a sequence* $a_n = 1$ *for all* $n \in \mathbb{N}$*. However, the mathematical analysis relies on the sequence* $(a_n)_{n\in\mathbb{N}}$ *vanishing sufficiency fast. Negative examples are given in [RNT12, HRW12], where it is shown that even* additive *space-time white noise with* $a_n \equiv 1$ *can cause a stochastic parabolic problem to have no solution.*

We would like to give to the formal equation (1.65) a more precise meaning. Let us first write it as a stochastic partial differential equation:

$$\mathrm{d}\boldsymbol{M} = [\lambda_1 \boldsymbol{M} \times \boldsymbol{H}_{\text{eff}} - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{H}_{\text{eff}})]\,\mathrm{d}t + [\boldsymbol{M} \times \boldsymbol{g}] \circ \mathrm{d}W. \tag{1.68}$$

Observe that we used the *Stratonovich* integral (see Appendix C). This choice is dictated by the *modeling* requirement that the solution $\boldsymbol{M}$ shall satisfy the unit-modulus condition $|\boldsymbol{M}(\omega, t, \boldsymbol{x})| = 1$ for all $t \in [0,T]$, for a.e. $\boldsymbol{x} \in D$ and for a.e. $\omega \in \Omega$. Equation (1.68), but with an Itô integral, would not satisfy this condition because of the Itô chain rule (see Lemma C.2).

Remark 1.24 (Itô formulation of SLLG)**.** *An appropriate modification of the drift term allows us to write an Itô problem that is equivalent to* (1.68)*. As shown in [AdBH14], the Itô-Stratonovich conversion formula (see Lemma C.4) gives*

$$[\boldsymbol{M} \times \boldsymbol{g}] \circ dW = [\boldsymbol{M} \times \boldsymbol{g}]\, dW + \frac{1}{2} (\boldsymbol{M} \times \boldsymbol{g}) \times \boldsymbol{g}\, dt.$$

Together with the triple product expansion B.2 and the unit modulus condition, it gives the following Itô SPDE

$$d\boldsymbol{M} = \left[\lambda_1 \boldsymbol{M} \times \Delta \boldsymbol{M} + \lambda_2 |\nabla \boldsymbol{M}|^2 \boldsymbol{M} + \lambda_2 \Delta \boldsymbol{M} + \frac{\lambda_2}{2} (\boldsymbol{M} \times \boldsymbol{g}) \times \boldsymbol{g}\right] dt + [\lambda_1 \boldsymbol{M} \times \boldsymbol{g}]\, dW.$$

The Gilbert form of the equation is obtained by (formally) applying the operator $(Id - \frac{1}{\lambda}\boldsymbol{M} \times \cdot)$ *to the previous equation*

$$\begin{aligned} d\boldsymbol{M} - \frac{1}{\lambda} \boldsymbol{M} \times d\boldsymbol{M} = &\left[\frac{1}{\lambda}\left(\Delta \boldsymbol{M} + |\nabla \boldsymbol{m}|^2 \boldsymbol{M}\right) + \frac{\lambda_2}{2}\left(Id - \frac{1}{\lambda}\boldsymbol{M} \times \cdot\right)((\boldsymbol{M} \times \boldsymbol{g}) \times \boldsymbol{g})\right] dt \\ &+ \left[\left(Id - \frac{1}{\lambda}\boldsymbol{M} \times \cdot\right)(\boldsymbol{M} \times \boldsymbol{g})\right] dW \end{aligned} \tag{1.69}$$

(setting $\lambda = \lambda_1 = \lambda_2 = 1$ *gives the equation in [AdBH14]).*

Following [BGJ12], we give a definition of weak solution of the SLLG equation.

Definition 1.25 (Weak martingale solution of the SLLG equation)**.** *A weak martingale solution to the SLLG equation is a tuple* $(\Omega, (\mathcal{E}_t)_{t\in[0,T]}, \mathcal{E}, \mathcal{P}, W, \boldsymbol{M})$*, where*

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

such that:

- $\boldsymbol{M}(\omega) \in C^0([0,T], \mathbb{H}^{-1}(D))$ $\mathcal{P}$*-a.s.;*
- $\mathbb{E}\left[esssup_{t\in[0,T]} \|\nabla \boldsymbol{M}\|^2\right] < \infty$*;*
- $|\boldsymbol{M}| = 1$ *a.e. in* Ω *and a.e. in* D_T*;*
- *for all* $t \in [0,T]$ *and* $\boldsymbol{\phi} \in \mathbb{C}_0^\infty(D;\mathbb{R})^3$ *there holds, a.e. in* Ω*,*

$$\begin{aligned}\langle \boldsymbol{M}(t), \boldsymbol{\phi}\rangle - \langle \boldsymbol{M}^0, \boldsymbol{\phi}\rangle \; &= -\lambda_1 \int_0^t \langle \boldsymbol{M} \times \nabla \boldsymbol{M}, \nabla \boldsymbol{\phi}\rangle \, ds \\ &\quad -\lambda_2 \int_0^t \langle \boldsymbol{M} \times \nabla \boldsymbol{M}, \nabla (\boldsymbol{M} \times \boldsymbol{\phi})\rangle \, ds + \int_0^t \langle \boldsymbol{M} \times \boldsymbol{g}, \boldsymbol{\phi}\rangle \circ dW.\end{aligned}$$

It was proved in [BGJ12] by means of the Faedo-Galerkin method that weak solutions to the SLLG equation indeed exist. The result was extended to the case including the anisotropy contribution in the effective field in [BL16].

Additional well-posedness results were obtained in the 1D case, i.e. the magnetic body is assumed to have one dominant dimension (this has applications in the manufacturing of nanowires). In [BGJ17], the authors prove existence of weak martingale solutions for the problem for a larger class of coefficients compared to the previous works in 3D. They prove pathwise existence and uniqueness of strong solutions. The also prove a *large deviation principle* and use it to analyze the transitions between equilibria. In [BMM19] the 1D problem is treated using the Doss-Sussmann transformation (see Section 1.1.5). They then prove existence, uniqueness and regularity of solution of the last problem with a Faedo-Galerkin method based on the Wong-Zakai approximation of the Wiener process [WZ65].

1.3.5 Space and time integration of the SLLG equation

In the present section, we list numerical methods for the approximation of sample paths of the SLLG equation. For simplicity we consider the problem in the small particle limit, i.e. $\boldsymbol{H}_{\text{eff}} = \Delta \boldsymbol{M}$.

Midpoint rule

We apply the midpoint rule proposed in [BBNP14a] (we already discussed this method for the deterministic LLG equation in Section 1.3.3). Recall the definitions of mass-lumped scalar product (1.58), discrete Laplace operator (1.59) and midpoint (1.61). The method reads:

Algorithm 4 $\left(\boldsymbol{M}_h^j\right)_{j=0}^J \leftarrow$ `Random midpoint rule`$(\mathcal{T}_h, \boldsymbol{m}^0, J, \tau\ \lambda_1, \lambda_2)$

1: Let $\boldsymbol{M}_h^0 = \Pi_h \boldsymbol{M}^0$
2: **for** For $j = 0, 1, ..., J$ **do**
3: Sample $\delta_j W = W(t_{j+1}) - W(t_j) \sim \mathcal{N}(0, \tau)$
4: Find $\boldsymbol{M}_h^{j+1} \in \boldsymbol{V}_h$ such that

$$\begin{aligned}\left\langle \frac{\boldsymbol{M}_h^{j+1} - \boldsymbol{M}_h^j}{\tau}, \boldsymbol{\phi}_h \right\rangle_h =& \lambda_1 \left\langle \boldsymbol{M}_h^{j+\frac{1}{2}} \times \Delta_h \boldsymbol{M}_h^{j+1}, \boldsymbol{\phi}_h \right\rangle_h \\ &- \lambda_2 \left\langle \boldsymbol{M}_h^{j+\frac{1}{2}} \times \left(\boldsymbol{M}_h^{j+\frac{1}{2}} \times \Delta_h \boldsymbol{M}_h^{j+1} \right), \boldsymbol{\phi}_h \right\rangle_h \\ &+ \lambda_1 \left\langle \boldsymbol{M}_h^{j+\frac{1}{2}} \times \boldsymbol{g} \delta_j W, \boldsymbol{\phi} \right\rangle_h \qquad \forall \boldsymbol{\phi}_n \in \boldsymbol{V}_h.\end{aligned}$$

5: **end for**

This method is defined starting from the Landau-Lifshitz formulation (1.68) rather that the Gilbert formulation as done in the deterministic case (see Section 1.3.3).

In [BBNP14a], the method is shown to produce discrete solutions that satisfy a discrete energy bound and that preserve the modulus of the initial condition at the mesh nodes $\mathcal{N}_h$.

Moreover, the sequence of discrete solutions obtained as the discretization parameters τ, h vanish is shown to converge (up to extraction of a subsequence) to a weak solution of SLLG in the sense of Definition 1.25.

In the follow-up work [BBP13], the scheme is applied to reproduce numerically relevant phenomena such as finite-time blow-up of the solution and thermally-activated switching. A method to solve the nonlinear system at each timestep is chosen and the overall numerical feasibility of the scheme is demonstrated in the numerical experiments.

Tangent plane scheme

In [AdBH14], the authors consider a time semi-discrete approximation based on the tangent plane scheme. We already discussed the application of this method to the deterministic LLG equation in Section 1.3.3), where we also considered the finite element approximation in space. Here, the sample paths of the SPDE *in Itô form* are approximated and stability and convergence results analogous to the previous works are given.

The numerical scheme is defined starting from the SLLG equation in Itô form (see Remark 1.24. Consider the following infinite-dimensional tangent plane, a natural analogue of (1.63):

$$\mathcal{K}(\boldsymbol{M}^j) \coloneqq \left\{ \boldsymbol{v} \in \mathbb{H}^1(D) : \boldsymbol{M}^j \cdot \boldsymbol{v} = \boldsymbol{0} \text{ a.e. in } D \right\}.$$

The time-semidiscrete algorithm reads:

Algorithm 5 $\left(\boldsymbol{M}^j\right)_{j=0}^J \leftarrow$ `Random tangent plane scheme`$(J,\, \tau,\, \boldsymbol{M}^0,\, \theta,\, \lambda_1,\, \lambda_2)$

1: **for** For $j = 0, 1, ..., J-1$ **do**
2: Sample $\delta_j W = W(t_{j+1}) - W(t_j) \sim \mathcal{N}(0, \tau)$
3: Find $\boldsymbol{v}^j \in \mathcal{K}(\boldsymbol{M}^j)$ such that

$$\begin{aligned}\left\langle \boldsymbol{v}^j - \boldsymbol{M}^j \times \boldsymbol{v}^j, \boldsymbol{\phi} \right\rangle + 2\theta\tau \left\langle \nabla \boldsymbol{v}^j, \boldsymbol{\phi} \right\rangle = &-2 \left\langle \nabla \boldsymbol{M}^j, \boldsymbol{\phi} \right\rangle + \left\langle \left(\mathrm{Id} - \boldsymbol{M}^j \times \cdot\right)\left(\boldsymbol{M}^j \times \boldsymbol{g}\right) \delta_j W, \boldsymbol{\phi} \right\rangle \\ &+ \frac{1}{2} \left\langle \left(\mathrm{Id} - \boldsymbol{M}^j \times \cdot\right)\left(\left(\boldsymbol{M} \times \boldsymbol{g}\right) \times \boldsymbol{g}\right), \boldsymbol{\phi} \right\rangle \quad \forall \boldsymbol{\phi} \in \mathcal{K}(\boldsymbol{M}^j)\end{aligned}$$

4: $\boldsymbol{M}^{j+1} = \frac{\boldsymbol{M}^j + \tau \boldsymbol{v}^j}{\left|\boldsymbol{M}^j + \tau \boldsymbol{v}^j\right|}$
5: **end for**

In [AdBH14], the authors prove that, as the timestep τ vanishes, the sequence of discrete solutions converges, up to extraction of a subsequence, to a weak martingale solution of the SLLG equation. Convergence is understood *in law*, more precisely in the norm $L^2(\Omega; \mathbb{L}^2(D_T))$. They actually consider a more general ∞-dimensional noise but increase its regularity by applying to each term a Hilbert-Schmidt operator $G_i : \mathbb{L}^2(D) \to \mathbb{H}^2(D)$.

1.3.6 Reduction of the SLLG equation to a random coefficient LLG equation

The present section is based on [GLT16], in which it was proved that the SLLG Equation (1.68) (a SPDE) can be reduced to a *random coefficient PDE*.

The *Doss-Sussmann transform* of $\boldsymbol{M}$, solution of (1.68), is the random field $\boldsymbol{m} : \Omega \times D_T \to \mathbb{S}^2$ defined as

$$\boldsymbol{m}(\omega, t, \boldsymbol{x}) \coloneqq e^{-W(\omega,t)G(\boldsymbol{x})} \boldsymbol{M}(\omega, t, \boldsymbol{x}) \tag{1.70}$$

(we use the symbol $\boldsymbol{m}$ as done for the (deterministic) LLG equation for reasons that will become clear soon). Here, G denotes the function $G\boldsymbol{u} \coloneqq \boldsymbol{u} \times \boldsymbol{g}$ for any $\boldsymbol{u} \in \mathbb{R}^3$ and by e^{-WG} we denote the exponential of the operator $-WG$, defined as the matrix exponential: $e^{-WG}\boldsymbol{u} = \sum_{i=0}^{\infty} \frac{(-W(t))^i}{i!} G^{(i)}\boldsymbol{u}$, where $G^{(i)}$ denotes the composition of G with itself i times (see also (D.2)).

It can be proved (see [GLT16, Lemma 3.1, Lemma 3.2]) that the operators G and e^{sG} satisfy the following properties:

$$e^{sG} e^{-sG} = \mathrm{id} = e^{-sG} e^{sG} \tag{1.71}$$

$$\left(e^{WG}\right)^* = e^{-WG} \tag{1.72}$$

$$e^{sG}(\boldsymbol{a} \times \boldsymbol{b}) = e^{sG}\boldsymbol{a} \times e^{sG}\boldsymbol{b} \tag{1.73}$$

$$e^{sG}\boldsymbol{u} = \boldsymbol{u} + \sin(s) G\boldsymbol{u} + (1 - \cos(s)) G^2\boldsymbol{u}. \tag{1.74}$$

The field $\boldsymbol{m}$ is itself solution of a PDE related to the SLLG equation. Let us first show this formally. For $\boldsymbol{M} = e^{WG}\boldsymbol{m}$, there holds

$$\partial_t \boldsymbol{M} = \partial_t e^{WG}\boldsymbol{m} = G\left(e^{WG}\boldsymbol{m}\right) \partial_t W + e^{WG} \partial_t \boldsymbol{m} = \left(\boldsymbol{M} \times \boldsymbol{g}\right) \partial_t W + e^{WG} \partial_t \boldsymbol{m}.$$

Recall that $\boldsymbol{M}$ solves (1.68), which we write here in the case $\boldsymbol{H}_{\mathrm{eff}} = \Delta \boldsymbol{M}$:

$$\mathrm{d}\boldsymbol{M} = \left[\lambda_1 \boldsymbol{M} \times \Delta \boldsymbol{M} - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \Delta \boldsymbol{M})\right] \mathrm{d}t + \left[\lambda_1 \boldsymbol{M} \times \boldsymbol{g}\right] \circ \mathrm{d}W.$$

We identify $\partial_t \boldsymbol{M}$ with $\mathrm{d}\boldsymbol{M}$ as well as $(\boldsymbol{M} \times \boldsymbol{g})\, \partial_t W$ with $[\boldsymbol{M} \times \boldsymbol{g}] \circ \mathrm{d}W$ in order to write

$$e^{WG}\partial_t \boldsymbol{m} = \lambda_1 e^{WG}\boldsymbol{m} \times \Delta e^{WG}\boldsymbol{m} - \lambda_2 e^{WG}\boldsymbol{m} \times \left(e^{WG}\boldsymbol{m} \times \Delta e^{WG}\boldsymbol{m}\right).$$

We simplify by applying (1.71) and (1.73) to obtain

$$\partial_t \boldsymbol{m} = \lambda_1 \boldsymbol{m} \times e^{-WG}\Delta e^{WG}\boldsymbol{m} - \lambda_2 \boldsymbol{m} \times \left(\boldsymbol{m} \times e^{-WG}\Delta e^{WG}\boldsymbol{m}\right).$$

We denote

$$\hat{\mathcal{C}}(W, \boldsymbol{m}) \coloneqq e^{-WG}\Delta e^{WG}\boldsymbol{m} - \Delta \boldsymbol{m} \tag{1.75}$$

in order to finally write

$$\partial_t \boldsymbol{m} = \lambda_1 \boldsymbol{m} \times \left(\Delta \boldsymbol{m} + \hat{\mathcal{C}}(W, \boldsymbol{m})\right) - \lambda_2 \boldsymbol{m} \times \left(\boldsymbol{m} \times \left(\Delta \boldsymbol{m} + \hat{\mathcal{C}}(W, \boldsymbol{m})\right)\right). \tag{1.76}$$

The last equation may be viewed as a LLG equation with effective field $\Delta \boldsymbol{m} + \hat{\mathcal{C}}(W, \boldsymbol{m})$. However, this is not a deterministic PDE, rather a random coefficient PDE since the Wiener process W appears as problem data. The solution $\boldsymbol{m}$ is therefore a random field.

Integration by parts for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{H}^1(D)$ with $\partial_{\boldsymbol{n}} \boldsymbol{u} = 0$ gives (recalling the assumptions $\partial_{\boldsymbol{n}} \boldsymbol{g} = 0$)

$$\begin{aligned}
\left\langle \hat{\mathcal{C}}(s, \boldsymbol{u}), \boldsymbol{v} \right\rangle &= \left\langle e^{-WG}\Delta e^{WG}\boldsymbol{u} - \Delta \boldsymbol{u}, \boldsymbol{v} \right\rangle \\
&= \int_{\partial D} \partial_{\boldsymbol{n}} e^{WG}\boldsymbol{u} \cdot e^{WG}\boldsymbol{v} - \left\langle \nabla e^{WG}\boldsymbol{u}, \nabla e^{WG}\boldsymbol{v} \right\rangle - \left(\int_{\partial D} \partial_{\boldsymbol{n}} \boldsymbol{u} \cdot e^{WG}\boldsymbol{v} - \langle \nabla \boldsymbol{u}, \nabla \boldsymbol{v} \rangle \right) \\
&= \langle \nabla \boldsymbol{u}, \nabla \boldsymbol{v} \rangle - \left\langle \nabla e^{WG}\boldsymbol{u}, \nabla e^{WG}\boldsymbol{v} \right\rangle.
\end{aligned} \tag{1.77}$$

Another way to write $\hat{\mathcal{C}}(s, \boldsymbol{u})$ is the following: Start from (1.75) and use (1.74):

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

By the chain rule, $\Delta\left(G\boldsymbol{u}\right) = \Delta\left(\boldsymbol{u} \times \boldsymbol{g}\right) = \Delta \boldsymbol{u} \times \boldsymbol{g} + 2\nabla \boldsymbol{u} \times \nabla \boldsymbol{g} + \boldsymbol{u} \times \Delta \boldsymbol{g}$, so that

$$\Delta G\boldsymbol{u} - G\Delta \boldsymbol{u} = \boldsymbol{u} \times \Delta \boldsymbol{g} + 2\nabla \boldsymbol{u} \times \nabla \boldsymbol{g}.$$

The previous formula implies $\Delta G^2\boldsymbol{u} = \Delta(G\boldsymbol{u} \times \boldsymbol{g}) = \Delta G\boldsymbol{u} \times \boldsymbol{g} + 2\nabla G\boldsymbol{u} \times \nabla \boldsymbol{g} + G\boldsymbol{u} \times \Delta \boldsymbol{g}$ and

$$\Delta G^2\boldsymbol{u} - G^2\Delta \boldsymbol{u} = \Delta G^2\boldsymbol{u} - G\Delta G\boldsymbol{u} + G\Delta G\boldsymbol{u} - G^2\Delta \boldsymbol{u} = \mathcal{C}G\boldsymbol{u} + G\mathcal{C}\boldsymbol{u}.$$

All in all, we obtained

$$\begin{aligned}
\mathcal{C}\boldsymbol{u} &\coloneqq \boldsymbol{u} \times \Delta \boldsymbol{g} + 2\nabla \boldsymbol{u} \times \nabla \boldsymbol{g} \\
\mathcal{E}(s, \boldsymbol{u}) &\coloneqq \sin(s)\mathcal{C}\boldsymbol{u} + (1 - \cos(s))\left(\mathcal{C}G + G\mathcal{C}\right)\boldsymbol{u} \\
\hat{\mathcal{C}}(s, \boldsymbol{u}) &= e^{-WG}\mathcal{E}(s, \boldsymbol{u}).
\end{aligned} \tag{1.78}$$

This makes it evident that $\hat{\mathcal{C}}$ is actually a first order differential operator in $\boldsymbol{u}$.

In [GLT16], an appropriate definition of weak solution to (1.76) is given.

Definition 1.26 (Weak solution of the random LLG equation)**.** *A weak solution of the random LLG equation* (1.76) *is a tuple* $(\Omega, (\mathcal{E}_t)_{t\in[0,T]}, \mathcal{E}, \mathcal{P}, W, \boldsymbol{m})$*, where*

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

such that:

- $\boldsymbol{m}(\omega) \in \mathbb{H}^1(D_T)$ *a.e. in* Ω*;*
- $\mathbb{E}\left[esssup_{t\in[0,T]} \|\nabla \boldsymbol{m}\|^2\right] < \infty$*;*
- $|\boldsymbol{m}| = 1$ *in* D_T *and a.e. in* Ω*;*
- $\boldsymbol{m}(0,\cdot) = \boldsymbol{M}^0$ *in* D *and a.e. in* Ω*;*
- *for each* $t \in [0,T]$ *and* $\boldsymbol{\phi} \in L^2(0,T,\mathbb{H}^1(D))$ *there holds, a.e. in* Ω*,*

$$\begin{aligned}\lambda_1 \int_0^t \langle \partial_t \boldsymbol{m}, \boldsymbol{\phi}\rangle + \lambda_2 \int_0^t \langle \boldsymbol{m}\times\partial_t\boldsymbol{m}, \boldsymbol{\phi}\rangle &= \mu \int_0^t \langle \nabla \boldsymbol{m}, \nabla(\boldsymbol{m}\times\boldsymbol{\phi})\rangle \\ &\quad + \int_0^t \langle \lambda_1 F(t,\boldsymbol{m}) + \lambda_2 \boldsymbol{m}\times F(t,\boldsymbol{m}), \boldsymbol{\phi}\rangle\,.\end{aligned} \tag{1.79}$$

where $\mu = \lambda_1^2 + \lambda_2^2$ *and* $F(t,\boldsymbol{m}) = \lambda_1 \boldsymbol{m}\times\hat{\mathcal{C}}(W(t),\boldsymbol{m}) - \lambda_2 \boldsymbol{m}\times\Big(\boldsymbol{m}\times\hat{\mathcal{C}}(W(t),\boldsymbol{m})\Big)$.

The main result of [GLT16] is:

Theorem 1.27. *If* $\boldsymbol{m}$ *is a weak solution of* (1.76) *in the sense of Definition 1.26, then* $\boldsymbol{M} = e^{-WG}\boldsymbol{m}$ *is a weak martingale solution of* (1.68) *in the sense of Definition 1.25.*

Proof outline. Let us summarize the main steps of the proof:

1. From Definition 1.26, $\boldsymbol{m} \in \mathbb{H}^1(D_T)$, $|\boldsymbol{m}| = 1$ a.e. and it solves (1.76). Then, $\boldsymbol{m}$ also solves

$$\begin{aligned}\int_0^t \langle \partial_t \boldsymbol{m}, \boldsymbol{\psi}\rangle + \lambda_1 \langle \boldsymbol{m}\times\nabla\boldsymbol{m}, \nabla\boldsymbol{\psi}\rangle + \lambda_2 \langle \boldsymbol{m}\times\nabla\boldsymbol{m}, \nabla(\boldsymbol{m}\times\boldsymbol{\psi})\rangle &= \int_0^t \langle F(t,\boldsymbol{m}), \boldsymbol{\psi}\rangle \\ \forall \boldsymbol{\psi} \in L^2(0,T,\mathbb{W}^{1,\infty}(D)),\ \text{a.e. in } \Omega.&\end{aligned} \tag{1.80}$$

   To obtain this, note that for each $\boldsymbol{\psi} \in L^2(0,T,\mathbb{W}^{1,\infty}(D))$ there exists $\boldsymbol{\phi} \in L^2(0,T,\mathbb{H}^1(D))$ such that $\lambda_1\boldsymbol{\phi} + \lambda_2\boldsymbol{\phi}\times\boldsymbol{m} = \boldsymbol{\psi}$.

2. The Itô chain rule (see Lemma C.2) applied to $\boldsymbol{M} = e^{WG}\boldsymbol{m}$ (understood as a function of t and W) gives

$$\boldsymbol{M}(t) = \boldsymbol{M}(0) + \int_0^t \left[e^{WG}\partial_t\boldsymbol{m} + \frac{1}{2}G^2 e^{WG}\boldsymbol{m}\right]\mathrm{d}t + \int_0^t \left[Ge^{WG}\boldsymbol{m}\right]\mathrm{d}W.$$

   From the relation between Itô and Stratonovich differentials (see Lemma C.4) and the fact that $G^2\boldsymbol{u} = G'(\boldsymbol{u})[G\boldsymbol{u}]$, we can write

$$\boldsymbol{M}(t) = \boldsymbol{M}(0) + \int_0^t \left[e^{WG}\partial_t\boldsymbol{m}\right]\mathrm{d}t + \int_0^t \left[G\boldsymbol{M}\right]\circ\mathrm{d}W.$$

3. Multiplying by $\boldsymbol{\phi} \in C_0^\infty(D)^3$ and using (1.72), we get

$$\langle \boldsymbol{M}(t), \boldsymbol{\phi} \rangle = \langle \boldsymbol{M}(0), \boldsymbol{\phi} \rangle + \int_0^t \left\langle \partial_t \boldsymbol{m}, e^{-WG} \boldsymbol{\phi} \right\rangle \mathrm{d}t + \int_0^t \langle G\boldsymbol{M}, \boldsymbol{\phi} \rangle \circ \mathrm{d}W.$$

We may now use Equation (1.80) with $\boldsymbol{\psi} = e^{-WG}\boldsymbol{\phi}$ to substitute the second term in the right-hand side.

4. To prove that $\boldsymbol{m}$ satisfies (1.54), it is left to show that

$$\begin{aligned} \left\langle \boldsymbol{m} \times \nabla \boldsymbol{m}, \nabla e^{-WG} \boldsymbol{\psi} \right\rangle - \left\langle \boldsymbol{m} \times \hat{\mathcal{C}}(W, \boldsymbol{m}), e^{-WG} \boldsymbol{\psi} \right\rangle &= \langle \boldsymbol{M} \times \nabla \boldsymbol{M}, \nabla \boldsymbol{\psi} \rangle, \\ \left\langle \boldsymbol{m} \times \nabla \boldsymbol{m}, \nabla \left(\boldsymbol{m} \times e^{-WG} \boldsymbol{\psi} \right) \right\rangle - \left\langle \boldsymbol{m} \times \left(\boldsymbol{m} \times \hat{\mathcal{C}}(W, \boldsymbol{m}) \right), e^{-WG} \boldsymbol{\psi} \right\rangle & \\ = \langle \boldsymbol{M} \times \nabla \boldsymbol{M}, \nabla \left(\boldsymbol{M} \times \boldsymbol{\psi} \right) \rangle. & \end{aligned}$$

The first equality is proved applying the triple product (B.1), the relation between $\boldsymbol{m}$ and $\boldsymbol{M}$ (1.70), the property of $\hat{\mathcal{C}}$ (1.77). The second is proved analogously.

5. The pointwise unit modulus of $\boldsymbol{M}$ follow from properties (1.72) and (1.71):

$$|\boldsymbol{M}|^2 = e^{WG}\boldsymbol{m} \cdot e^{WG}\boldsymbol{m} = \boldsymbol{m} \cdot e^{-WG} e^{WG} \boldsymbol{m} = |\boldsymbol{m}|^2 = 1.$$

6. Boundary conditions for $\boldsymbol{M}$ follow from: $\partial_{\boldsymbol{n}} \boldsymbol{M} = \partial_{\boldsymbol{n}} \left(e^{WG} \boldsymbol{m} \right) = e^{WG} \left(\partial_{\boldsymbol{n}} \boldsymbol{m} \right) = 0$ (we assumed in (1.67) that $\partial_n \boldsymbol{g} = 0$ on ∂D) and continuity properties of e^{WG}.

7. The fact that $\boldsymbol{M}(0) = \boldsymbol{M}^0$ follows from: $e^{W(0)G} = \mathrm{id}$.

□

1.3.7 Space and time approximation of the random coefficient LLG equation

The random coefficient LLG equation (1.76) obtained in the previous section was discretized in [GLT16]. The authors employ a suitably adapted tangent plane scheme (see also Sections 1.3.3 and 1.3.5.

The algorithm reads:

Algorithm 6 $\left(\boldsymbol{M}_h^j \right)_{j=0}^J \leftarrow$ `Random tangent plane scheme (alt.)`$(\mathcal{T}_h, \boldsymbol{M}^0, J, \tau\ \lambda_1, \lambda_2, \theta)$

1: Compute $\boldsymbol{m}_h^0 = \Pi_h \boldsymbol{M}^0$
2: **for** $j = 0, 1, ..., J-1$ **do**
3: Find $\boldsymbol{v}_h^j \in \mathcal{K}_h(\boldsymbol{m}_h^j)$ such that

$$\left\langle \lambda_2 \boldsymbol{v}_h^j - \lambda_1 \boldsymbol{m}_h^j \times \boldsymbol{v}_h^j, \boldsymbol{\phi}_h \right\rangle = -\mu \left\langle \nabla \left(\boldsymbol{m}_h^j + \theta\tau \boldsymbol{v}_h^j \right), \nabla \boldsymbol{\phi}_h^j \right\rangle - \left\langle R_{h,\tau}(t_j, \boldsymbol{m}_h^j), \boldsymbol{\phi} \right\rangle \quad \forall \boldsymbol{\phi}_n \in \mathcal{K}_h(\boldsymbol{m}_h^j)$$

4: $\boldsymbol{m}_h^{j+1} = \mathfrak{I}_h \left[\frac{\boldsymbol{m}_h^j + \tau \boldsymbol{v}_h^j}{|\boldsymbol{m}_h^j + \tau \boldsymbol{v}_h^j|} \right]$
5: **end for**

In the previous algorithm we denoted $\mu = \lambda_1^2 + \lambda_2^2$, and $R_{h,\tau}(t, \boldsymbol{u})$ a suitably chosen discrete version of $R(t, \boldsymbol{u}) := \lambda_2^2 \boldsymbol{u} \times \left(\boldsymbol{u} \times \hat{\mathcal{C}}(t, \boldsymbol{u})\right) - \lambda_1^2 \hat{\mathcal{C}}(t, \boldsymbol{u})$. More precisely,

$$
\begin{aligned}
G_h \boldsymbol{u} &= \boldsymbol{u} \times \mathfrak{I}_h \boldsymbol{g} \\
\mathcal{C}_h \boldsymbol{u} &= \boldsymbol{u} \times \mathfrak{I}_h[\Delta \boldsymbol{g}] + 2\nabla \boldsymbol{u} \times \mathfrak{I}_h(\nabla \boldsymbol{g}) \\
D_{h,\tau}(t, \boldsymbol{u}) &= \left(\sin(W_\tau(t))\mathcal{C}_h + (1 - \cos(W_\tau(t))(G_h \mathcal{C}_h + \mathcal{C}_h G_h)\right) \boldsymbol{u} \\
\hat{\mathcal{C}}_{h,\tau}(t, \boldsymbol{u}) &= \left(\mathrm{id} - \sin(W_\tau(t)) G_h + (1 - \cos(W_\tau(t)) G_h^{(2)}\right) D_{h,\tau}(t, \boldsymbol{u}) \\
R_{h,\tau}(t, \boldsymbol{u}) &= \lambda_2^2 \boldsymbol{u} \times \left(\boldsymbol{u} \times \hat{\mathcal{C}}_{h,\tau}(t, \boldsymbol{u}))\right) - \lambda_1^2 \hat{\mathcal{C}}_{h,\tau}(t, \boldsymbol{u}).
\end{aligned}
$$

where $W_\tau(t) = W(t_j)$ if $t_j \leq t < t_{j+1}$.

The authors assume that $\boldsymbol{M}^0 \in \mathbb{H}^2(D)$, $\boldsymbol{g} \in \mathbb{W}^{2,\infty}(D)$, and $|\boldsymbol{g}| = 1$ a.e. They show convergence of a subsequence of the sequence of discrete solutions obtained as the discretization parameters h, τ vanish. In particular, they prove weak convergence in $\mathbb{H}^1(D_T)$ and a.e. in Ω. Moreover, they show that the random field $e^{WG} \boldsymbol{m}_{h,\tau}$, obtained applying the *inverse Doss-Sussman transform*, gives an approximation to the weak martingale solution of SLLG in the sense of Definition 1.25.

In [GGL20], the authors generalize the approach based on the Doss-Sussmann transformation to multi-dimensional (but with finitely many terms) noise (cf. Remark 1.23). Observe that this requires a more general strategy to apply the Doss-Sussmann transform. Nevertheless, many results generalize analogously to the 1-dimensional case.

1.3.8 Applications of the numerical approximation of the SLLG equation

In this short section, we highlight some engineering applications for which an efficient and reliable approximation of SLLG could be useful. The fact that the model is valid for the sub-micrometer length-scale made it very popular in the simulation of magnetic data storage devices such as hard disk drives.

- *Simulation of data corruption due heat fluctuations and stability of magnetic storage*: Increasing the data storage density requires reducing the bit size. However, the resistance of a magnetic body to perturbations is proportional to its size. Therefore, a point is reached when the heat perturbation present at room temperature is sufficient to introduce non-negligible errors. Unless different materials become available or the intensity of the writing magnetic field can be increased (see HAMR below), this poses a physical limit to magnetic storage density. It was estimated in [ECN+12] that the current density limit is 15 to 20 $Tbit/in^2$. An efficient simulation method for SLLG could be used to estimate the probability of an unwanted switching for a fixed bit size. However, in the regime of *rare event simulation*, i.e. for small target probabilities, further developments, both in terms of methods and of numerical analysis guarantees, would be needed to tackle this challenging problem;

- *Heat Assisted Magnetic Recording* (HAMR): An emerging magnetic storage technology which promises both higher storage density and resistance to noise-induced storage errors. A magnetically hard material is selected because of its resistance to perturbations. At room temperature, such material would need a strong magnetic field to cause a switch, possibly incompatible with the small size required of writing heads. A possible solution is to heat up the magnetic cell (e.g. with a laser) in order to reduce its coercivity and allow a weaker magnetic field to write the bits [KGM+08, RBB+06, IH06, AMM+12]. At the

time of writing, some of the first products based on this technology are already available on the market[1];

- *Spin-transfer torque magnetoresistive random access memory (STT-MRAM)*: Another emerging technology that leverages the *spin-transfer torque* effect: A spin-polarized current (i.e. one where spins have a nonzero net orientation in a specific direction) passes through a *magnetic tunnel junctions* (the equivalent of a bit) and a transfer of spin angular momentum causes the magnetization to re-align. The electric control of the magnetization guarantees high speeds, precision, scalability and, at the same time, lower power consumption. See [AKW+13, LW21] for more details. At the time of writing, some of the first products based on this technology are already available on the market[2];

- *Racetrack memory* [PY15, BKF+20]: Another novel magnetic data storage technology. It promises high reliability, capacity, and speed by storing data in the domain walls of nanowires, induced by spin-polarized currents. Advantages are the fact that it is a solid state drive (no moving parts) and it can be implemented as a 3D device.

[1] https://www.seagate.com/gb/en/news/news-archive/seagates-breakthrough-30tb-plus-hard-drives-ramp-volume-marking-an-inflection-point-in-the-storage-industry-pr/

[2] https://www.everspin.com/spin-transfer-torque-ddr-products#:~:text=Everspin's%20Spin%2Dtransfer%20Torque%20MRAM,use%20of%20supercapacitors%20or%20batteries.

1.4 Contributions of this thesis

In Chapter 2, we consider the adaptive sparse grid algorithm for the random affine diffusion Poisson problem from [GN18] and achieve the following results:

- Proof of plain convergence of the adaptive sparse grid algorithm (where the finite element error is assumed to be negligible, Theorem 2.15);
- Definition of an *adaptive sparse grid–finite element* (ASGFE) algorithm which adaptively refines both parametric and space approximations (Algorithm 7 and following listings). We use adaptive finite elements with h-refinement for the space approximation;
- Proof of plain convergence of the ASGFE algorithm (Theorems 2.26, 2.29), as well as an$h$-refinement optimality result (Theorem 2.24);
- Implementation and a number of numerical tests for the ASGFE algorithm.

Here we have to tackle a number of difficulties. Of particular relevance are:

- The fact that sparse grid is a collocation method and not a projection method (such as e.g. stochastic Galerkin or polynomial chaos expansion). This makes the task of deriving error estimates, both a priori and a-posteriori, more difficult;
- The interplay between parametric and finite element error and their effect of the respective a-posteriori error estimators.

These results were also published in [FS21]

In Chapter 3, we are interested in approximating physical quantities described by nonlinear SPDEs where the random noise, which models heat fluctuations, is a Gaussian process. We also consider, as an example, the *stochastic Landau–Lifshitz–Gilbert equation* (SLLG) from micromagnetics, which describes the magnetization of sub-micrometer magnetic bodies immersed in a heath bath. Our main achievements here are:

- Outline of a general strategy to study the parametric regularity of SPDEs driven by Gaussian noise. We reduce the SPDE to a parametric coefficient PDE with infinite-dimensional unbounded parameters using the Doss-Sussmann transform and the Lévy-Ciesielski expansion of the Wiener process. The theory is based on four "abstract" assumptions that have to be verified for each concrete problem;
- Application to the SLLG equation (in particular, verification of Assumptions 1-4 in Section 3.1 below). This is the first result on uniform holomorphic regularity of the parameter-to-solution map for the SLLG 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;
- A Hölder regularity result for the sample paths of solutions of SLLG assuming regular initial conditions which are sufficiently close to constant;
- Definition of a sparse grid scheme tailored to SLLG regularity;
- Convergence analysis of the sparse grid approximation of SLLG. This is the first rigorous convergence result for an approximation of a nonlinear and time-dependent parametric coefficient PDE with unbounded parameter space. In particular, we show convergence

of piecewise quadratic sparse grids for the stochastic LLG equation with order 1/2 and dimension dependent constant (Theorem 3.25). Under some reasonable assumptions and simplifications of the stochastic input, we show dimension independent convergence with order 1/2 (Theorem 3.29);

- Definition of a multilevel approximation applied to SLLG with proof of improved convergence rate under natural assumptions on the underlying finite element method;
- Numerical verification of the convergence of the above-mentioned numerical schemes.

To achieve these 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 [CCS15], 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 [CF01b, CDG98, Cim07, LLW15, Mel05, Mel12, Mos05] 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 3.7).

These results were also published in [ADF+24]

Other results we document in this dissertation are the following:

- In Chapter 4, we give remarks on the implementation and use of `SGMethods`, an object oriented Python implementation of sparse grid interpolation used to carry out numerical experiment of Chapter 3;
- In Chapter 5, additional and partial results. In Section 5.1, we report on partial results regarding optimality and convergence with rate of adaptive sparse grid interpolation, different marking strategies, dimension-adaptivity, cost reduction of the computation of the a-posteriori estimator, and additional numerical results. In Section 5.2, a convergence proof of the tangent plane scheme applied to the random coefficient LLG equation is proposed. In the proof, we had to assumes the existence of a test functions with specific properties.

Chapter 2

Convergence of an adaptive sparse grid–finite element scheme

In this chapter, we describe and analyses an *adaptive sparse grid–finite element algorithm* (ASGFE) for the approximation of the affine diffusion Poisson problem. The algorithm is based on the reliable a-posteriori estimator from [GN18], consisting of a sum of parametric and finite-element error estimators. We exploit the non-intrusive nature of sparse grid interpolation (a collocation algorithm described in Section 1.2.3) to effectively decouple the two adaptive procedures. We then go on to suggest several variations of the algorithm. As for the analysis, we prove plain convergence of the algorithm under the assumption that at each iteration of the adaptive loop the finite element estimator is smaller than (a function of) the current parametric estimator, which is proved to vanish. Results are confirmed by numerical experiments.

The chapter is structured as follows: In Section 2.1, we recall the definition of the parametric diffusion Poisson problem (already mentioned in Example 1.8), sparse grid interpolant (see Section 1.2.3) and introduce the adaptive sparse grid–finite elements algorithm (ASGFE). In Section 2.2, we prove convergence of the adaptive algorithm proposed in [GN18] for the pure parameter enrichment problem. Section 2.3 proves the convergence of the full adaptive algorithm including spatial adaptivity (with one adaptive mesh per collocation point or one global adaptive mesh). Section 2.4 presents some numerical experiments. Finally, in the final Section 2.5 we draw some conclusions.

The following is, up to small stylistic changes and new cross-references, verbatim to [FS21]. To achieve a coherent presentation, we made the following changes to the notation:

- Tensor product parameter space Γ to $\mathbf{\Gamma}$;
- Multi-index sets I to Λ;
- Sparse grid interpolant S to $\mathcal{I}$;
- Multi-index $\boldsymbol{i}$ to $\boldsymbol{\nu}$;
- Hierarchical surplus operator $\Delta^{m(\boldsymbol{i})}$ to $\Delta_{\boldsymbol{\nu}}$;
- Tensor product interpolant $I^{m(\boldsymbol{i})}$ to $I_{\boldsymbol{\nu}}$;
- Tensor product interpolation nodes $\mathcal{Y}^{m(\boldsymbol{i})}$ to $\mathcal{Y}_{\boldsymbol{\nu}}$;
- Denote by $\mathbb{P}_i$ the space of degree i polynomials and by $V_\nu = \mathbb{P}_{m(\nu)-1}$ the linear space corresponding to the multi-index ν, where $m : \mathbb{N}_0 \to \mathbb{N}$ is the level-to-knot function of the nodes family in use;

- Banach space of solutions of deterministic PDE V to $\mathbb{U}$;
- Adaptive sparse grid–finite element algorithm name SCFE (Stochastic Collocation Finite Element) to ASGFE (adaptive sparse grid–finite element);
- Adaptive sparse grid algorithm name SC (Stochastic Collocation) to ASG (Adaptive Sparse Grid);
- Space variable x to $\boldsymbol{x}$;
- Multi-indices in $\mathbb{N}^N$ to $\mathbb{N}_0^N$.

2.1 Problem and algorithm

2.1.1 Problem statement

Let us begin by recalling the problem setting, even if all concepts were already defined in Chapter 1. Consider an integer $d \geq 2$ and an open bounded domain $D \subset \mathbb{R}^d$ with Lipschitz continuous boundary ∂D. Let $(\Omega, \mathcal{E}, \mathcal{P})$ be a complete probability spaces. Let $Y_n : \Omega \to \mathbb{R}$ be independent random variables with ranges $\Gamma_n \coloneqq Y_n(\Omega)$ and densities $\rho_n : \Gamma_n \to \mathbb{R}$ for all $n \in 1, \ldots, N$. We assume that the ranges Γ_n are *bounded* subsets of $\mathbb{R}$. Let $\boldsymbol{\Gamma} \coloneqq \bigotimes_{n=1}^N \Gamma_n \subset \mathbb{R}^N$ and $\rho \coloneqq \prod_{n=1}^N \rho_n$. This means that, in this chapter, we are working under the finite-dimensional noise assumption (see Remark 1.7). The triple $(\boldsymbol{\Gamma}, \mathcal{B}(\boldsymbol{\Gamma}), \rho(\boldsymbol{y})\mathrm{d}\boldsymbol{y})$ ($\mathcal{B}(\boldsymbol{\Gamma})$ the Borel σ-algebra on $\boldsymbol{\Gamma}$) is a probability space. Consider $f \in L^2(D)$ and $a : \boldsymbol{\Gamma} \times D \to \mathbb{R}$ with the following properties:

- *Uniform boundedness*: There exist $0 < a_{\min} \leq a_{\max} < \infty$ such that
$$a_{\min} \leq a(\boldsymbol{y}, \boldsymbol{x}) \leq a_{\max} \qquad \rho\text{-a.e. } \boldsymbol{y} \in \boldsymbol{\Gamma}, \forall \boldsymbol{x} \in D;$$
- *Affine dependence on* $\boldsymbol{y} \in \boldsymbol{\Gamma}$:
$$\forall n = 1, \ldots, N \; \exists a_n : D \to \mathbb{R} : a(\boldsymbol{y}, \boldsymbol{x}) = a_0(x) + \sum_{n=1}^{N} a_n(\boldsymbol{x}) y_n;$$
- *Regularity in space*: $\nabla a(\boldsymbol{y}, \cdot)|_T \in L^\infty(T)$ for all elements T of a coarse *initial* mesh $\mathcal{T}_{\text{init}}$ of D.

We consider the energy space $\mathbb{U} \coloneqq H_0^1(D)$ with norm $\|v\|_{\mathbb{U}} \coloneqq \|\nabla v\|_{L^2(D)}$ and the parametric weak formulation of the Poisson problem:
Find $u : \boldsymbol{\Gamma} \to \mathbb{U}$ such that

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \nabla u(\boldsymbol{x}, \boldsymbol{y}) \cdot \nabla v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_D f(\boldsymbol{x}) v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \qquad \forall v \in \mathbb{U}, \; \rho\text{-a.e. } \boldsymbol{y} \in \boldsymbol{\Gamma}. \tag{2.1}$$

Due to uniform ellipticity of the problem, the exact solution is unique and (see also, e.g., [BNT10, Lemma 3.1]) there exists $\boldsymbol{\tau} \subset \mathbb{R}^N_{>0}$ such that $u : \boldsymbol{\Gamma} \to \mathbb{U}$ can be extended to a bounded holomorphic function on the set

$$\Sigma(\boldsymbol{\Gamma}, \boldsymbol{\tau}) \coloneqq \left\{ \boldsymbol{z} \in \mathbb{C}^N : \operatorname{dist}(z_n, \Gamma_n) \leq \tau_n \; \forall n = 1, \ldots, N \right\}. \tag{2.2}$$

2.1.2 Sparse grid (stochastic collocation) interpolation

We aim at building an approximation of the parameter-to-solution map $u : \mathbf{\Gamma} \to U$ of (2.1) in the space

$$\mathbb{P}(\mathbf{\Gamma}, \mathbb{U}) \cong \mathbb{P}(\mathbf{\Gamma}) \otimes \mathbb{U}, \tag{2.3}$$

where $\mathbb{P}(\mathbf{\Gamma})$ is a finite-dimensional polynomial space on $\mathbf{\Gamma}$ and $\mathbb{U}$ may be substituted by a finite-dimensional subspace of $\mathbb{U}$. In order to do so, we fix a set of collocation nodes $\mathcal{H} \subset \mathbf{\Gamma}$ and denote by $\{L_{\boldsymbol{y}}\}_{\boldsymbol{y}\in\mathcal{H}}$ the related Lagrange basis (i.e. the unique set of polynomials of minimal degree over $\mathbf{\Gamma}$ such that $L_{\boldsymbol{z}}(\boldsymbol{y}) = \delta_{\boldsymbol{y},\boldsymbol{z}}$ for any $\boldsymbol{y}, \boldsymbol{z} \in \mathcal{H}$). By $\mathbb{P}(\mathbf{\Gamma})$, we denote the polynomial space spanned by $\{L_{\boldsymbol{y}}\}_{\boldsymbol{y}\in\mathcal{H}}$. For any $\boldsymbol{y} \in \mathcal{H}$, we consider $\mathcal{T}_{\boldsymbol{y}}$, a shape-regular triangulation on D depending on $\boldsymbol{y}$, and $\mathbb{U}_{\boldsymbol{y}} := \mathcal{S}^1_0(\mathcal{T}_{\boldsymbol{y}})$, the classical finite elements space of piecewise-linear functions over $\mathcal{T}_{\boldsymbol{y}}$ with zero boundary conditions. We denote by $U_{\boldsymbol{y}} \in \mathbb{U}_{\boldsymbol{y}}$ the finite element approximation of $u(\boldsymbol{y})$, i.e. $U_{\boldsymbol{y}}$ solves the variational problem

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \nabla U_{\boldsymbol{y}}(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_D f(\boldsymbol{x}) v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \qquad \forall v \in \mathbb{U}_{\boldsymbol{y}}. \tag{2.4a}$$

Finally, the fully discrete approximation of u takes the form

$$\sum_{\boldsymbol{y}\in\mathcal{H}} U_{\boldsymbol{y}}(\boldsymbol{x}) L_{\boldsymbol{y}}(\boldsymbol{z}). \tag{2.4b}$$

The number of its degrees of freedom is $\sum_{\boldsymbol{y}\in\mathcal{H}} \dim(\mathbb{U}_{\boldsymbol{y}})$.

The set of collocation nodes $\mathcal{H}$, the Lagrange basis $\{L_{\boldsymbol{y}}\}_{\boldsymbol{y}\in\mathcal{H}}$, and the polynomial space $\mathbb{P}(\mathbf{\Gamma})$ are defined following the sparse grid construction , which we described in Section 1.2.3. Recall that a sparse grid interpolation method is defined given a 1D interpolation method, which in turn is defined by a 1D nodes family and a level-to-knot function. Here we consider *Clenshaw-Curtis (CC)* nodes (as in Example 1.15):

$$\begin{aligned} \mathcal{Y}_0 &= \{1\} \\ \mathcal{Y}_\nu &= \left\{ -\cos\left(\frac{\pi(i-1)}{m(\nu)-1}\right) : i = 1, \dots, m(\nu) \right\} \qquad \text{if } \nu \in \mathbb{N}. \end{aligned}$$

As level-to-knot function, we employ the *doubling rule*:

$$m(\nu) = \begin{cases} 1 & \text{if } \nu = 0, \\ 2^\nu + 1 & \text{if } \nu \in \mathbb{N}. \end{cases} \tag{2.5}$$

We additionally make the formal assumption

$$m(-1) = 1.$$

It is easy to prove that the resulting nodes family $\{\mathcal{Y}_\nu\}_{\nu\in\mathbb{N}_0}$ is nested. We will stick with this particular choice for the remainder of this work for simplicity, however that other choices are possible (see, e.g., [GN18]). The essential properties of $\mathcal{Y}_n$ used in the proofs below are nestedness $\mathcal{Y}_\nu \subseteq \mathcal{Y}_{\nu+1}$ and the fact that the Lebesgue constants of the associated interpolation operators grow sub-exponentially.

As mentioned in the beginning of the section, we want a *polynomial* interpolation. Therefore, we define

$$V_\nu = \mathbb{P}_{m(\nu)-1}(\Gamma) \qquad \forall \nu \in \mathbb{N}_0,$$

i.e. the linear space of polynomials on $\Gamma = [-1, 1]$ of degree $m(\nu) - 1$. The spaces $V_{\boldsymbol{\nu}}$ (1.28) for $\boldsymbol{\nu} \in \mathbb{N}_0^N$ and V_Λ for $\Lambda \subset \mathbb{N}_0^N$ remain defined as a function of V_ν as in in Section 1.2.3.
A sparse grid interpolant $\mathcal{I}_\Lambda$ 1.32 is defined once a downward-closed multi-index set $\Lambda \subset \mathbb{N}_0^N$ is given (see Section 1.2.3 for more details).

Since the exact solution u is analytic in $\boldsymbol{y}$, we may represent it with a sparse grid-type sum over all multi-indices (see again [GN18]):

$$u(\boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^N} \Delta_{\boldsymbol{\nu}} u(\boldsymbol{y}) \qquad \text{for a.e. } \boldsymbol{y} \in \boldsymbol{\Gamma}. \tag{2.6}$$

The sum converges absolutely in $\mathbb{U}$.

2.1.3 The adaptive sparse grid–finite element algorithm (ASGFE)

In the present section, we define an adaptive algorithm that iteratively refines both sparse grid and finite element approximation. An important ingredient is the reliable a-posteriori estimator from [GN18], which we already described in Section 1.2.4 (in particular, the estimator was defined in (1.44), (1.45) and the a-posteriori estimate in (1.46)). Here, we aim at controlling the $L^\infty(\boldsymbol{\Gamma}, \mathbb{U})$ error of the fully discrete approximation, but the same can be done for the L^p_ρ norm for any $1 \leq p \leq \infty$ and measure ρ on $\boldsymbol{\Gamma}$.
In [GN18, p. 3130] an algorithm is given to adaptively refine the parameter space only. We extend it here to the following *Adaptive Sparse Grid–Finite Element* algorithm (ASGFE), which also adaptively refines the finite element space.

Algorithm 7 $u_\varepsilon \leftarrow \texttt{ASGFE}(\varepsilon, \theta, \alpha, \mathcal{T}_{\text{init}})$

1: $\Lambda_0 := \{\mathbf{0}\}$
2: Compute $U_{0,\boldsymbol{y}}$ on $\mathcal{T}_{\text{init}}$ for all $\boldsymbol{y} \in \mathcal{H}_{\Lambda_0}$
3: **for** $\ell = 0, 1, 2, \ldots$ **do**
4: $\quad U_\ell \leftarrow \texttt{Refine\_FE\_spaces}\ (\Lambda_\ell, U_\ell, \alpha, \theta)$
5: $\quad$ Compute $(\zeta_{\boldsymbol{\nu},\Lambda_\ell})_{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}}, \zeta_{SC,\Lambda_\ell}, \eta_{\text{FE},\Lambda_\ell}$
6: $\quad$ **if** $a_{\min}^{-1}(\zeta_{SC,\Lambda_\ell} + C\eta_{\text{FE},\Lambda_\ell}) < \varepsilon$ **then**
7: $\quad\quad$ **return** $u_\varepsilon \leftarrow \mathcal{I}_{\Lambda_\ell}[U_\ell]$
8: $\quad$ **end if**
9: $\quad (U_{\ell+1}, \Lambda_{\ell+1}) \leftarrow \texttt{Refine\_parameter\_space}(\Lambda_\ell, U_\ell, (\zeta_{\boldsymbol{\nu},\Lambda_\ell})_{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}}, \mathcal{T}_{\text{init}})$
10: **end for**

Algorithm 7 consists of alternating between enriching the polynomial space V_Λ in Line 9 (Alg. 7) and refining the finite element spaces corresponding to each collocation point independently from each other in Line 4 (Alg. 7). The intuitive idea behind this choice is the following: In order for the parameter enrichment routine to make a meaningful choice, the finite element solution should be a relatively good approximation of the exact solution for parameters coinciding with the collocation nodes. The algorithm terminates when the a-posteriori estimator falls below a given tolerance $\varepsilon > 0$ in Line 6 (Alg. 7). The reliable upper bound (1.46) guarantees that the error of the discrete solution is also bounded by ε.

The sub-routine `Refine_FE_spaces` reads:

Algorithm 8 $U \leftarrow$`Refine_FE_spaces` $(\Lambda, U, \alpha, \theta)$

1: Compute $\left(\eta_{\boldsymbol{y},T_{\boldsymbol{y}}}\right)_{T_{\boldsymbol{y}}\in\mathcal{T}_{\boldsymbol{y}},\boldsymbol{y}\in\mathcal{H}_\Lambda}$, $\eta_{\mathrm{FE},\Lambda}$, $\zeta_{SC,\Lambda}$
2: Tol $\leftarrow \alpha \left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}\prod_{n=1}^{N}\nu_n\right)^{-1}\zeta_{SC,\Lambda}$
3: **while** $\eta_{\mathrm{FE},\Lambda} >$ Tol **do**
4: Find minimal $\mathcal{K}\subset\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\boldsymbol{y}}$ such that $\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\mathcal{K}}\eta^2_{\boldsymbol{y},T_{\boldsymbol{y}}}\|L_{\boldsymbol{y}}\|_{L^\infty(\boldsymbol{\Gamma})}\geq\theta\eta^2_{\mathrm{FE},\Lambda}$
5: **for** $\boldsymbol{y}\in\mathcal{H}$ **do**
6: Refine $\mathcal{T}_{\boldsymbol{y}}$ with $\mathcal{K}_{\boldsymbol{y}}:=\{T\in\mathcal{T}_{\boldsymbol{y}}\,:\,(\boldsymbol{y},T)\in\mathcal{K}\}$ as marked elements
7: Compute $U_{\boldsymbol{y}}$ over $\mathcal{T}_{\boldsymbol{y}}$
8: **end for**
9: Compute $\left(\eta_{\boldsymbol{y},T_{\boldsymbol{y}}}\right)_{T_{\boldsymbol{y}}\in\mathcal{T}_{\boldsymbol{y}},\boldsymbol{y}\in\mathcal{H}_\Lambda}$, $\eta_{\mathrm{FE},\Lambda}$, $\zeta_{SC,\Lambda}$
10: Tol $\leftarrow \alpha \left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}\prod_{n=1}^{N}\nu_n\right)^{-1}\zeta_{SC,\Lambda}$
11: **end while**

The aim of this sub-routine is to refine the finite element solutions in the collocation points until the finite element estimator falls below the tolerance defined in Line 2 (Alg. 8). We use newest-vertex-bisection with mesh closure for mesh refinement (see, e.g., [Ste08] for further details). Observe that, since the tolerance depends on the parametric estimator $\zeta_{SC,\Lambda}$, which in turn depends on the discrete solution, the tolerance needs to be re-computed after every finite-element refinement. This is necessary as one parametric refinement might uncover new features of the solution which need to be resolved in the finite-element refinement step. In practical computations, this rarely happens after the first few iterations of the algorithm. Note also that the linear convergence result in Proposition 2.22 shows that starting from the initial mesh does not significantly increase the computational cost (at worst case, it contributes logarithmically). In line 4 (Alg. 8), we considered the *disjoint union*:

$$\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\boldsymbol{y}}:=\bigcup_{\boldsymbol{y}\in\mathcal{H}}\bigcup_{T_{\boldsymbol{y}}\in\mathcal{T}_{\boldsymbol{y}}}(\boldsymbol{y},T_{\boldsymbol{y}}). \tag{2.7}$$

In Section 2.3 we will prove that the sub-routine terminates (i.e. that the finite element estimator eventually falls below the tolerance) and that the choice of tolerance made in Line 2 (Alg. 8) is a sufficient condition for convergence. It might be possible to prove convergence omitting the scaling $\left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}\prod_{n=1}^{N}\nu_n\right)^{-1}$ in the tolerance in Line 2 (Alg. 8). The experiments in Section 2.4 indicate that it is not necessary, however with the present techniques a proof is still out of reach.

Finally, the sub-routine `Refine_parameter_space` reads:

Algorithm 9 $(U',\Lambda')\leftarrow$`Refine_parameter_space` $(\Lambda,\, U,\, (\zeta_{\boldsymbol{\nu},\Lambda})_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda},\, \mathcal{T}_{\mathrm{init}})$

1: Compute $\mathcal{P}_{\boldsymbol{\nu},\Lambda}$ for all $\boldsymbol{\nu}\in\mathcal{M}_\Lambda$
2: $\boldsymbol{\nu}\leftarrow\arg\max_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}\mathcal{P}_{\boldsymbol{\nu},\Lambda}$
3: $\Lambda'\leftarrow\Lambda\cup A_{\boldsymbol{\nu},\Lambda}$
4: $U'\leftarrow$ Update U by computing finite element solution $U_{\boldsymbol{y}}$ on $\mathcal{T}_{\mathrm{init}}$ for all $\boldsymbol{y}\in\mathcal{H}_{\Lambda'}\setminus\mathcal{H}_\Lambda$

The aim here is to enrich the polynomial space V_Λ as done in [GN18, Algorithm 1]. At each iteration, the algorithm enlarges the multi-index set Λ by adding multi-indices from the margin of Λ depending on the values of the pointwise error estimators $(\zeta_{\boldsymbol{\nu},\Lambda})_{\boldsymbol{\nu}\in\Lambda}$. More precisely, in Line 2 (Alg. 9) we select a *profit maximizer*, i.e. a multi index in the margin that maximizes a

given *profit function* $\mathcal{P}_{\boldsymbol{\nu},\Lambda}$ (see below for some examples):

$$\boldsymbol{\nu} = \underset{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}{\arg\max}\, \mathcal{P}_{\boldsymbol{\nu},\Lambda} \tag{2.8}$$

(in case more than one multi-index maximizes the profit, we pick the one that comes first in the lexicographic ordering).
Then, in Line 3 (Alg. 9), Λ is enlarged by adding $A_{\boldsymbol{\nu},\Lambda}$, the smallest subset of $\mathcal{M}_\Lambda$ containing $\boldsymbol{\nu}$ such that $\Lambda \cup A_{\boldsymbol{\nu},\Lambda}$ is downward-closed. Finally, in Line 4(Alg. 9), we compute the finite element solution over the default mesh $\mathcal{T}_{\text{init}}$ corresponding to each new collocation point, while preserving the old ones.

We analyze two possible choices of profit:

- *Workless profit*:

$$\mathcal{P}_{\boldsymbol{\nu},\Lambda} \coloneqq \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu},\Lambda}} \zeta_{\boldsymbol{j},\Lambda} \tag{2.9}$$

- *Profit with work*:

$$\mathcal{P}_{\boldsymbol{\nu},\Lambda} \coloneqq \frac{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu},\Lambda}} \zeta_{\boldsymbol{j},\Lambda}}{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu},\Lambda}} W_{\boldsymbol{j}}} \tag{2.10}$$

  where the *work* is defined as $W_{\boldsymbol{j}} \coloneqq \prod_{n=1}^{N} (m(j_n) - m(j_{n-1}))$. This choice follows the profit maximization strategy outlined in Section 1.2.3. In this case, the "value", i.e. the numerator of the profit (2.10), is determined a-posteriori as a function of the pointwise error estimator $\zeta_{\boldsymbol{j},\Lambda}$ for $\boldsymbol{j} \in A_{\boldsymbol{\nu},\Lambda}$.

2.2 Convergence of the adaptive sparse grid algorithm (ASG)

We examine the convergence properties of a simplified version of Algorithm 7, also discussed in [GN18]. In the present case, we suppose to be able to sample the function $u : \boldsymbol{\Gamma} \to \mathbb{U}$ for any fixed parameter $\boldsymbol{y} \in \boldsymbol{\Gamma}$. Thus, a discrete solution is given by the sparse grid interpolant $\mathcal{I}_\Lambda[u] \in V_\Lambda(\boldsymbol{\Gamma}, \mathbb{U})$, for a downward-close multi-index set $\Lambda \subset \mathbb{N}_0^N$. Moreover, the a-posteriori estimator, which in the fully discrete setting was the sum of parametric estimator (1.44) and finite element estimator (1.45), simplifies to $\zeta_{SC,\Lambda} \coloneqq \sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \zeta_{\boldsymbol{\nu},\Lambda}$ (no additional term accounting for the finite element discretization), where the pointwise estimator now reads

$$\zeta_{\boldsymbol{\nu},\Lambda} \coloneqq \left\| \Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u] \right) \right\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))} .$$

In this setting, the reliable upper bound (1.46) simplifies to: $\|u - \mathcal{I}_\Lambda[u]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} \lesssim \zeta_{SC,\Lambda}$. Workless-profit and profit with work are defined analogously to (2.9) and (2.10) respectively. This simplified version of the algorithm reads:

Algorithm 10 $u_\varepsilon \leftarrow ASG(\varepsilon, u)$

```
Λ_0 ← {0}
Compute I_{Λ_ℓ}[u]
for ℓ = 0, 1, 2, ... do
    Compute ζ_{SC,Λ_ℓ}
    if a_min^{-1} ζ_{SC,Λ_ℓ} < ε then
        Return u_ε ← I_{Λ_ℓ}[u]
    end if
    Compute P_{ν,Λ_ℓ} for all ν ∈ M_{Λ_ℓ}
    ν_ℓ ← arg max_{ν∈M_{Λ_ℓ}} P_{ν,Λ_ℓ}
    Λ_ℓ ← Λ_ℓ ∪ A_{ν_ℓ,Λ_ℓ}
    Compute I_{Λ_ℓ}[u]
end for
```

The algorithm is given as input a tolerance $\varepsilon > 0$ and the permission to access point evaluations $u(\boldsymbol{y})$ for arbitrary $\boldsymbol{y} \in \boldsymbol{\Gamma}$. The output u_ε is an approximation of the exact parameter-to-solution map such that $\|u - u_\varepsilon\|_{L^\infty(\boldsymbol{\Gamma}, H_0^1(D))} < \varepsilon$, which is guaranteed by the reliability of the error estimator.

2.2.1 Preliminary results

Stability and convergence of the hierarchical surplus $\Delta_{\boldsymbol{\nu}}$

In this section, we recall basic results on the hierarchical surplus operator $\Delta_{\boldsymbol{\nu}}$ (see for instance [NTW08b]). The analysis is carried out in the $L^\infty(\boldsymbol{\Gamma}, \mathbb{U})$ norm as it is the most "stringent" among the $L^p(\boldsymbol{\Gamma}, \mathbb{U})$ norms for $p \in [1, \infty]$. We note that all the arguments below work for other choices of p.

We first state 1D results (corresponding to the case $N = 1$). For $\nu \in \mathbb{N}$, the Lebesgue constant λ_ν of the interpolant I_ν (1.26) satisfies the relation

$$\|I_\nu v\|_{L^\infty(\Gamma, \mathbb{U})} \leq \lambda_\nu \|v\|_{L^\infty(\boldsymbol{\Gamma}, \mathbb{U})} \qquad \forall v \in C^0(\Gamma, \mathbb{U}). \tag{2.11}$$

For CC nodes, it was proved in [DI83] that $\lambda_\nu \leq \frac{2}{\pi} \log(m(\nu) - 1) + 1$ for $m(\nu) \geq 2$ (obviously, $\lambda_1 \leq 1$). Recalling the definition of $m(\cdot)$ with the *doubling rule* (1.35), we obtain

$$\lambda_\nu \leq \nu + 1 \qquad \forall \nu \in \mathbb{N}_0. \tag{2.12}$$

The estimate can be used to derive a stability estimate for the detail operator

$$\|(I_\nu - I_{\nu-1})\, v\|_{L^\infty(\Gamma, \mathbb{U})} \lesssim \lambda_\nu \|v\|_{L^\infty(\Gamma, \mathbb{U})}.$$

Moving to the general case $N \in \mathbb{N}$, we can now exploit the tensor product structure of $\boldsymbol{\Gamma} = [-1, 1]^N$ to bound the $L^\infty(\boldsymbol{\Gamma}) \to L^\infty(\boldsymbol{\Gamma})$ operator norm of the hierarchical surplus operator:

$$\|\Delta_{\boldsymbol{\nu}} v\|_{L^\infty(\boldsymbol{\Gamma}, \mathbb{U})} \lesssim \lambda_{\boldsymbol{\nu}} \|v\|_{L^\infty(\boldsymbol{\Gamma}, \mathbb{U})} \qquad \forall v \in C^0(\boldsymbol{\Gamma}, \mathbb{U}), \tag{2.13}$$

where

$$\lambda_{\boldsymbol{\nu}} \leq \prod_{n=1}^{N} \lambda_{\nu_n} \lesssim \prod_{n=1}^{N} (\nu_n + 1). \tag{2.14}$$

Observe that the same holds for the tensor product interpolant $I_{\boldsymbol{\nu}}$.

Let us now assume that $u : \mathbf{\Gamma} \to \mathbb{U}$ is analytic with respect to $\boldsymbol{y}$ and derive another estimate of $\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})}$. The tensor product structure of $\mathbf{\Gamma}$ allows us again to start from a 1D results and then generalize to N dimensions. We state a result that relates the best approximation error in $\mathbb{P}_m(\Gamma)$ for $m \in \mathbb{N}$ to the size of the domain of the holomorphic extension of u (2.2).

Lemma 2.1 ([BNT10, Lemma 4.4])**.** *If $v \in C^0(\Gamma)$ and it exists $\tau > 0$ such that v admits a holomorphic extension to $\Sigma(\Gamma, \tau)$ (cf. (2.2)), then, for $m \in \mathbb{N}$,*

$$E_m(v) \coloneqq \min_{w \in \mathbb{P}_m(\Gamma)} \|v - w\|_{L^\infty(\Gamma)} \leq \frac{2}{e^\sigma - 1} e^{-\sigma m} \max_{z \in \Sigma(\Gamma,\tau)} |v(z)| \tag{2.15}$$

where $\sigma \coloneqq \log\left(\frac{2\tau}{|\Gamma|} + \sqrt{1 + \frac{4\tau^2}{|\Gamma|^2}}\right) > 0$. □

Since I_ν is exact on $\mathbb{P}_{m(\nu)-1}(\Gamma)$, its error can be expressed as (see [BNR00])

$$\|u - I_\nu u\|_{L^\infty(\Gamma)} \leq (1 + \lambda_\nu)\, E_{m(\nu)-1}(u).$$

Remembering (2.12) and the previous lemma, the error estimate for I_ν can be simplified to

$$\|u - I_\nu u\|_{L^\infty(\Gamma)} \lesssim \lambda_\nu e^{-\sigma m(\nu)} \max_{z \in \Sigma(\Gamma,\tau)} |u(z)|\,.$$

This estimate can be applied to the detail operator after a triangle inequality to obtain

$$\|\Delta_\nu u\|_{L^\infty(\Gamma)} \lesssim \lambda_\nu e^{-\sigma m(\nu-1)} \max_{z \in \Sigma(\Gamma,\tau)} |u(z)| \quad \forall \nu \in \mathbb{N}_0. \tag{2.16}$$

Applying (2.16) to the multidimensional case (by considering one component at a time) leads to the following estimate:

Lemma 2.2. *Consider a function $u : \mathbf{\Gamma} \to \mathbb{U}$ admitting a holomorphic extension to $\Sigma(\Gamma, \tau) \subset \mathbb{C}^{\mathbb{N}}$ (2.2). Its hierarchical surplus with multi-index $\boldsymbol{\nu} \in \mathbb{N}_0^N$ satisfies:*

$$\|\Delta_{\boldsymbol{\nu}} u\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \lesssim \lambda_{\boldsymbol{\nu}} e^{-\sigma |m(\boldsymbol{\nu}-\mathbf{1})|_1},$$

where the hidden constant depends on u and N, $\lambda_{\boldsymbol{\nu}}$ is estimated in (2.14), m is the level-to-knot function and

$$\sigma \coloneqq \min_{n \in 1,\dots,N} \sigma_n, \qquad \sigma_n \coloneqq \log\left(\frac{2\tau_n}{|\Gamma_n|} + \sqrt{1 + \frac{4\tau_n^2}{|\Gamma_n|^2}}\right).$$

A simplified formula for $\zeta_{\boldsymbol{\nu},\Lambda}$

In this section, we highlight elementary facts on the zeros of $\Delta_{\boldsymbol{\nu}} u$ and the kernel of $\Delta_{\boldsymbol{\nu}}$ for some $\boldsymbol{\nu} \in \mathbb{N}_0^N$. These facts are combined to show that the operator $\Delta_{\boldsymbol{\nu}}\,(a \nabla \Delta_{\boldsymbol{j}})$ is identically zero unless the multi-indices $\boldsymbol{\nu}$, $\boldsymbol{j} \in \mathbb{N}_0^N$ are "close to each other" (see also [GN18, Proposition 4.3] for partial results in this direction).

For two multi-indices $\boldsymbol{i}, \boldsymbol{j} \in \mathbb{N}_0^N$, $\boldsymbol{i} < \boldsymbol{j}$ means that for all $n \in 1, \dots, N$ $i_n < j_n$. Analogously, $\boldsymbol{i} \leq \boldsymbol{j}$ means that for all $n = 1, \dots, N$ $i_n \leq j_n$.

We denote by $\mathrm{Rect}_{\boldsymbol{\nu}} \subset \mathbb{N}_0^N$ the axis-aligned rectangle with opposite vertices $\mathbf{0}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^N$:

$$\mathrm{Rect}_{\boldsymbol{\nu}} \coloneqq \left\{\boldsymbol{j} \in \mathbb{N}_0^N : \boldsymbol{j} \leq \boldsymbol{\nu}\right\}. \tag{2.17}$$

It clearly holds that the corresponding sparse grid interpolant is a tensor product interpolant (1.29), i.e.

$$\mathcal{I}_{\mathrm{Rect}_{\boldsymbol{\nu}}} = I_{\boldsymbol{\nu}} \qquad \forall \boldsymbol{\nu} \in \mathbb{N}_0^N.$$

Theorem 2.3. *Let $\boldsymbol{\nu}$, $\boldsymbol{j} \in \mathbb{N}_0^N$ and assume that*

$$\exists n \in 1, \dots, N : \nu_n < j_n, \tag{2.18}$$

or that

$$\forall n \in 1, \dots, N : \boldsymbol{j} + \boldsymbol{e}_n < \boldsymbol{\nu}. \tag{2.19}$$

Then,

$$\Delta_{\boldsymbol{\nu}} \left(a \nabla \Delta_{\boldsymbol{j}} u\right) \equiv 0 \qquad \forall u \in C^0(\boldsymbol{\Gamma}, \mathbb{U}). \tag{2.20}$$

Proof. Fix $\boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{\nu}}$. By (2.18) and the nestedness of CC nodes, it exists $n \in 1, \dots, N$ such that $y_n \in \mathcal{Y}_{j_n - 1}$. This implies that $\Delta_{\boldsymbol{j}} u(\boldsymbol{y}) = 0$, as y_n is an interpolation point for both I_{j_n} and $I_{j_n - 1}$. Thus, recalling that ∇ acts on the space variable x only,

$$a(\boldsymbol{y}) \nabla \Delta_{\boldsymbol{j}} u(\boldsymbol{y}) = 0 \qquad \forall \boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{\nu}}. \tag{2.21}$$

Next, observe that a hierarchical surplus can be written as a linear combination of Lagrange interpolants

$$\Delta_{\boldsymbol{\nu}} = \sum_{\boldsymbol{\alpha} \in \{0,1\}^N} (-1)^{|\boldsymbol{\alpha}|} I_{\boldsymbol{\nu} - \boldsymbol{\alpha}}.$$

Again by the nestedness of CC nodes, (2.21) implies that $a \nabla \Delta_{\boldsymbol{j}} u$ is in the kernel of each of the interpolants $\left\{I_{\boldsymbol{\nu} - \boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0,1\}^N\right\}$, which in turn implies (2.20).

To show that (2.19) also implies (2.20), first observe that

$$a \nabla \Delta_{\boldsymbol{j}} u \in \sum_{n=1}^{N} \mathbb{P}_{m(\boldsymbol{j}) - \mathbf{1} + \boldsymbol{e}_n} \subseteq V_{\mathrm{Rect}_{\boldsymbol{\nu}} \setminus \{\boldsymbol{\nu}\}}, \tag{2.22}$$

where the last inclusion is due to assumption (2.19). Next, observe that a hierarchical surplus can be written as a difference of sparse grid interpolants:

$$\Delta_{\boldsymbol{\nu}} = \mathcal{I}_{\mathrm{Rect}_{\boldsymbol{\nu}}} - \mathcal{I}_{\mathrm{Rect}_{\boldsymbol{\nu}} \setminus \{\boldsymbol{\nu}\}}.$$

This implies that $V_{\mathrm{Rect}_{\boldsymbol{\nu}} \setminus \{\boldsymbol{\nu}\}}$ is a subset of the kernel of $\Delta_{\boldsymbol{\nu}}$, as both $\mathcal{I}_{\mathrm{Rect}_{\boldsymbol{\nu}}}$ and $\mathcal{I}_{\mathrm{Rect}_{\boldsymbol{\nu}} \setminus \{\boldsymbol{\nu}\}}$ are exact on this space. Together with (2.22), this concludes the proof. □

Remark 2.4. *Consider a multi-index set $\Lambda \subset \mathbb{N}_0^N$ and $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$. The previous theorem can be used to simplify the computation of $\zeta_{\boldsymbol{\nu},\Lambda}$. Define*

$$J_{\boldsymbol{\nu},\Lambda} \coloneqq \left\{\boldsymbol{j} \in \Lambda : \exists n \in 1, \dots, N : \boldsymbol{j} = \boldsymbol{\nu} - \boldsymbol{e}_n\right\}.$$

Then, thanks to the previous theorem:

$$\Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda [u]\right) = \Delta_{\boldsymbol{\nu}} \left(a \nabla \sum_{\boldsymbol{j} \in \Lambda} \Delta_{\boldsymbol{j}} u\right) = \Delta_{\boldsymbol{\nu}} \left(a \nabla \sum_{\boldsymbol{j} \in J_{\boldsymbol{\nu},\Lambda}} \Delta_{\boldsymbol{j}} u\right),$$

so

$$\zeta_{\boldsymbol{\nu},\Lambda} = \left\| \Delta_{\boldsymbol{\nu}} \left(a \nabla \sum_{\boldsymbol{j} \in J_{\boldsymbol{\nu},\Lambda}} \Delta_{\boldsymbol{j}} u\right) \right\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))}. \tag{2.23}$$

See Figure 2.1 for a graphical representation.

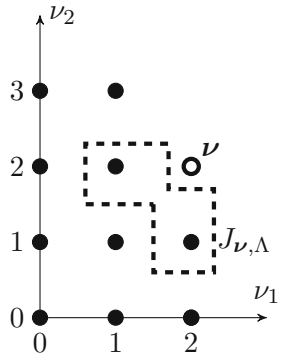

Figure 2.1: Graphical representation of the simplified computation of $\zeta_{\boldsymbol{\nu},\Lambda}$ from (2.23). We consider $N = 2$ parameters, each point in the plot corresponds to an element $\boldsymbol{\nu} = (\nu_1, \nu_2) \in \mathbb{N}_0^2$, where on the x-axis we represent ν_1 and on the y-axis ν_2. Filled dots represent Λ, the hollow one is $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$. The dashed line encircles the multi-indices in $J_{\boldsymbol{\nu},\Lambda}$, i.e. the only relevant ones in Λ for the computation of $\zeta_{\boldsymbol{\nu},\Lambda}$, as explained in Remark 2.4.

A-priori estimates for estimators and index sets

Proposition 2.5. *Given $u : \boldsymbol{\Gamma} \to \mathbb{U}$ a function that admits a holomorphic extension to $\Sigma(\boldsymbol{\Gamma}, \boldsymbol{\tau})$ (cf. (2.2)), a multi-index set $\Lambda \subset \mathbb{N}_0^N$, and $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$, the pointwise error estimator can be bounded as*

$$\zeta_{\boldsymbol{\nu},\Lambda} \lesssim N \lambda_{\boldsymbol{\nu}}^2 e^{-\sigma |m(\boldsymbol{\nu}-\mathbf{1})|_1},$$

where $\lambda_{\boldsymbol{\nu}}$ is as in (2.13) and m is the level-to-knot function (for example (2.5)).

Proof. Observe that $\mathcal{I}_\Lambda[u] : \boldsymbol{\Gamma} \to \mathbb{U}$ is analytic but not uniformly bounded with respect to Λ, so one cannot apply directly the convergence result for the hierarchical surplus. Recalling Remark 2.4, we can simplify the expression of $\zeta_{\boldsymbol{\nu},\Lambda}$ as

$$\zeta_{\boldsymbol{\nu},\Lambda} = \left\| \Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u] \right) \right\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))} = \left\| \Delta_{\boldsymbol{\nu}} \left(a \nabla \sum_{\substack{n \in 1,\dots,N \\ \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda}} \Delta_{\boldsymbol{\nu}-\boldsymbol{e}_n} u \right) \right\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))} .$$

Applying the stability estimate (2.13) of $\Delta_{\boldsymbol{\nu}}$, the uniform boundedness assumption (2.1.1) on a, and the triangle inequality, we obtain

$$\zeta_{\boldsymbol{\nu},\Lambda} \lesssim \lambda_{\boldsymbol{\nu}} \sum_{\substack{n \in 1,\dots,N \\ \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda}} \left\| \Delta_{\boldsymbol{\nu}-\boldsymbol{e}_n} \nabla u \right\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))} .$$

Observe finally that, since u is analytic, we can apply Lemma 2.2 to obtain

$$\zeta_{\boldsymbol{\nu},\Lambda} \leq \lambda_{\boldsymbol{\nu}} \sum_{\substack{n \in 1,\dots,N \\ \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda}} \lambda_{\boldsymbol{\nu}-\boldsymbol{e}_n} e^{-\sigma |m(\boldsymbol{\nu}-\boldsymbol{e}_n-\mathbf{1})|_1} \lesssim N \lambda_{\boldsymbol{\nu}}^2 e^{-\sigma |m(\boldsymbol{\nu}-\mathbf{1})|_1}.$$

□

Remark 2.6. *A direct consequence of the previous proposition is the uniform boundedness of the sequence of a-posteriori estimators $(\zeta_{SC,\Lambda_\ell})_\ell$ produced by the adaptive Algorithm 10 driven*

by either workless profit (2.9) *or profit with work* (2.10)*. Indeed, we have the following bound independently of the iteration number* ℓ*:*

$$\zeta_{SC,\Lambda_\ell} = \sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_\ell}} \zeta_{\boldsymbol{\nu},\Lambda_\ell} \lesssim N \sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_\ell}} \lambda_{\boldsymbol{\nu}} e^{-\sigma|m(\boldsymbol{\nu}-\mathbf{1})|_1} \leq N \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^N} \lambda_{\boldsymbol{\nu}} e^{-\sigma|m(\boldsymbol{\nu}-\mathbf{1})|_1} < \infty.$$

Lemma 2.7. *The profit maximizer* $\boldsymbol{\nu}_\ell \in \mathbb{N}_0^N$ *at iteration* ℓ *of Algorithm 10 satisfies*

$$\#Rect_{\boldsymbol{\nu}_\ell} = \prod_{n=1}^{N} (1 + \langle \boldsymbol{\nu}_\ell, \boldsymbol{e}_n \rangle) \leq \left(1 + \frac{\ell+1}{N}\right)^N. \tag{2.24}$$

In particular, there holds

$$\#A_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \leq \left(1 + \frac{\ell}{N}\right)^N \tag{2.25a}$$

and

$$\#\mathcal{M}_{\Lambda_\ell} \leq N \left(1 + \sum_{\ell'=0}^{\ell-1} \left(1 + \frac{\ell'+1}{N}\right)^N\right) \leq N \left(1 + \ell \left(1 + \frac{\ell}{N}\right)^N\right). \tag{2.25b}$$

Proof. The identity (2.24) is trivial. The following inequality is a consequence of the arithmetic-geometric inequality:

$$\prod_{n=1}^{N} (1 + j_n) \leq \left(\frac{\sum_{n=1}^{N}(1+j_n)}{N}\right)^N = \left(1 + \frac{|\boldsymbol{j}|_1}{N}\right)^N \qquad \forall \boldsymbol{j} \in \mathbb{N}_0^N,$$

and the fact that $|\boldsymbol{\nu}_\ell|_1 \leq \ell + 1$.

To prove (2.25a), simply observe that $A_{\boldsymbol{\nu},\Lambda} = \mathrm{Rect}_{\boldsymbol{\nu}} \setminus \Lambda$.

To prove (2.25b), first note that $\#\mathcal{M}_{\Lambda_\ell} \leq N \#\Lambda_\ell$. Then, an estimate of $\#\Lambda_\ell$ comes from partitioning $\Lambda_\ell = \{\mathbf{0}\} \cup \bigcup_{m=1}^{\ell-1} A_{\boldsymbol{\nu}_m,\Lambda_m}$ and the estimate of $\#A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}$ (2.25a) obtained above. □

Remarks on the algorithm driven by workless profit

In this section, we point out some elementary facts on the behavior of the algorithm when the workless profit defined in (2.9) is used. Inspired by [CCS14], we give the following definition:

Definition 2.8. *Given a downward closed multi-index set* $\Lambda \subset \mathbb{N}_0^N$*,* $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$ *is* maximal in $\mathcal{M}_\Lambda$ *if and only if*

$$\forall \boldsymbol{j} \in \mathcal{M}_\Lambda \setminus \{\boldsymbol{\nu}\}, \ \exists n \in 1, \ldots, N : \nu_n > j_n.$$

The set of maximal points in $\mathcal{M}_\Lambda$ *is denoted by* μ_Λ*.*

Example 2.9. *If* $\boldsymbol{\nu} \in \mathbb{N}_0^N$ *and* $\Lambda = Rect_{\boldsymbol{\nu}}$ *is an axis-aligned rectangle as defined in* (2.17)*, then*

$$\mu_\Lambda = \{\boldsymbol{\nu} + \boldsymbol{e}_n, n \in 1, \ldots, N\}.$$

Lemma 2.10. *Consider a downward-closed multi-index set* $\Lambda \subset \mathbb{N}_0^N$*. The multi-index corresponding to the largest workless profit* (2.9) *in* $\mathcal{M}_\Lambda$ *is maximal:*

$$\arg\max_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \mathcal{P}_{\boldsymbol{\nu},\Lambda} \in \mu_\Lambda. \tag{2.26}$$

In particular, the adaptive Algorithm 10 driven by workless profit always selects in Line 9 a maximal multi-index:

$$\boldsymbol{\nu}_\ell \in \mu_{\Lambda_\ell} \qquad \forall \ell \in \mathbb{N}_0. \tag{2.27}$$

Therefore, Λ_ℓ is always an axis-aligned rectangle in $\mathbb{N}_0^N$, i.e.

$$\Lambda_\ell = \mathit{Rect}_{\boldsymbol{\nu}_{\ell-1}} \qquad \forall \ell \in \mathbb{N}. \tag{2.28}$$

Proof. We prove (2.26) by contradiction. If $\boldsymbol{\nu}^* \coloneqq \arg\max_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \mathcal{P}_{\boldsymbol{\nu},\Lambda}$ is not maximal, there exists $\boldsymbol{j} \in \mathcal{M}_\Lambda \setminus \{\boldsymbol{\nu}^*\}$ such that for all $n \in 1,\dots,N$ $\langle \boldsymbol{\nu}^*, \boldsymbol{e}_n\rangle \leq j_n$. Thus, $\boldsymbol{\nu}^* \in A_{\boldsymbol{j},\Lambda_\ell}$ and by definition of workless profit (2.9), we have the contradiction $\mathcal{P}_{\boldsymbol{\nu}^*,\Lambda_\ell} < \mathcal{P}_{\boldsymbol{j},\Lambda_\ell}$.

The fact (2.28) can be proved by induction. For $\ell = 1$, $\Lambda_1 = \mathrm{Rect}_{\mathbf{0}} = \{\mathbf{0}\}$. Assume that for fixed $\ell \in \mathbb{N}$, $\Lambda_\ell = \mathrm{Rect}_{\boldsymbol{\nu}_{\ell-1}}$. Example 2.9 and (2.27) imply that

$$\boldsymbol{\nu}_\ell \in \mu_{\Lambda_\ell} = \mu_{\mathrm{Rect}_{\boldsymbol{\nu}_{\ell-1}}} = \{\boldsymbol{\nu}_{\ell-1} + \boldsymbol{e}_n, n \in 1,\dots,N\}.$$

Thus $\Lambda_{\ell+1} = \Lambda_\ell \cup A_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = \mathrm{Rect}_{\boldsymbol{\nu}_\ell}$. □

To summarize, the use of the workless profit (2.9) implies that, for all $\ell \in \mathbb{N}_0$,

- It exists *a unique* number $n(\ell) \in 1,\dots,N$ such that

$$\boldsymbol{\nu}_{\ell+1} = \boldsymbol{\nu}_\ell + \boldsymbol{e}_{n(\ell)}. \tag{2.29}$$

- As a consequence, the norm of $\boldsymbol{\nu}_\ell$ is given by:

$$|\boldsymbol{\nu}_{\ell+1}|_1 = |\boldsymbol{\nu}_\ell|_1 + 1 = \ell. \tag{2.30}$$

- Λ_ℓ is a rectangle:

$$\Lambda_{\ell+1} = \mathrm{Rect}_{\boldsymbol{\nu}_\ell}. \tag{2.31}$$

  Therefore, the sparse grid interpolant is actually a tensor product Lagrange interpolant:

$$\mathcal{I}_{\Lambda_{\ell+1}} = I_{\boldsymbol{\nu}_\ell} = \bigotimes_{n=1}^{N} I_{\langle \boldsymbol{\nu}_\ell, \boldsymbol{e}_n\rangle}.$$

- The multi-indices added at iteration ℓ are

$$A_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = \Lambda_{\ell+1} \setminus \Lambda_\ell = \left\{\boldsymbol{j} \in \mathrm{Rect}_{\boldsymbol{\nu}_\ell} : j_{n(\ell)} = \langle \boldsymbol{\nu}_\ell, \boldsymbol{e}_{n(\ell)}\rangle\right\}. \tag{2.32}$$

In other words, the evolution of the approximation space is determined by the sequence of integers $(n(\ell))_\ell$. This allows us to simplify the notation as follows

$$A_{n,\Lambda_\ell} \coloneqq A_{\boldsymbol{\nu}_{\ell-1}+\boldsymbol{e}_n,\Lambda_\ell} \tag{2.33}$$

$$\mathcal{P}_{n,\Lambda_\ell} \coloneqq \sum_{\boldsymbol{j}\in A_{n,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell} \tag{2.34}$$

Let us moreover denote the maximal n-th dimension of Λ_ℓ as

$$r_{n,\ell} \coloneqq \max_{\boldsymbol{j}\in\Lambda_\ell} j_n. \tag{2.35}$$

See Figure 2.2 for a graphical representation.

The estimate for the pointwise error estimator from Proposition 2.5 can be improved as follows: First observe that, due to (2.29) and (2.31), for any $\boldsymbol{\nu} \in A_{\nu_\ell,\Lambda_\ell}$,

$$J_{\boldsymbol{\nu},\Lambda_\ell} = \{\boldsymbol{j} \in \Lambda_\ell : \exists n \in 1,\dots,N : \boldsymbol{j} = \boldsymbol{\nu} - \boldsymbol{e}_n\} = \left\{\boldsymbol{\nu} - \boldsymbol{e}_{n(\ell)}\right\}.$$

Thus, $\#J_{\boldsymbol{\nu},\Lambda_\ell} = 1$ and we may reduce the dependence on N to

$$\zeta_{\boldsymbol{\nu},\Lambda_\ell} \lesssim \Lambda_{\boldsymbol{\nu}}^2 e^{-\sigma|m(\boldsymbol{\nu}-\mathbf{1})|}. \tag{2.36}$$

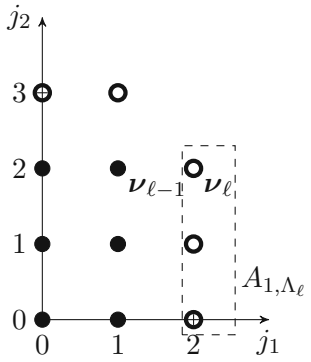

Figure 2.2: Example of approximation parameters at a generic step ℓ of the algorithm when the workless profit (2.9) is used. We consider $N = 2$ parameters, each point in the plot corresponds to a multi-index $\boldsymbol{j} = (j_1, j_2) \in \mathbb{N}_0^2$, where on the x-axis we represent j_1 and on the y-axis j_2. Filled dots represent Λ_ℓ, hollow ones $\mathcal{M}_{\Lambda_\ell}$. The multi-index selected by the algorithm at current step, $\boldsymbol{\nu}_\ell$ corresponds here to $n(\ell) = 1$). The dashed rectangle encircles multi-indices in $A_{n(\ell),\Lambda_\ell}$.

2.2.2 Convergence of the parametric estimator

In the following two lemmata, we prove that Algorithm 10 driven by workless profit and profit with work respectively forces the maximum profit over the margin to zero.

Proposition 2.11. *If the workless profit (2.9) is used, then*

$$\lim_{\ell\to\infty} \mathcal{P}_{n(\ell),\Lambda_\ell} = 0,$$

where we used the notation (2.34).

Proof. For fixed $n \in 1, \dots, N$, we estimate each pointwise error estimator appearing in $\mathcal{P}_{n,\Lambda_\ell}$ by (2.36) and the fact that for any $\boldsymbol{\nu}$ in A_{n,Λ_ℓ}, $\nu_n = r_{n,\ell} + 1$.

$$\begin{aligned}
\mathcal{P}_{n,\Lambda_\ell} &= \sum_{\boldsymbol{\nu}\in A_{n,\Lambda_\ell}} \zeta_{\boldsymbol{\nu},\Lambda_\ell} \lesssim \sum_{\boldsymbol{\nu}\in A_{n,\Lambda_\ell}} \lambda_{\boldsymbol{\nu}}^2 e^{-\sigma|m(\boldsymbol{\nu}-\mathbf{1})|_1} \\
&\lesssim \sum_{\boldsymbol{\nu}\in A_{n,\Lambda_\ell}} \prod_{k=1}^{N} \left((\nu_k+1)^2 e^{-\sigma|m(\nu_k-1)|_1} \right) \\
&\leq (r_{n,\ell}+2)^2\, e^{-\frac{\sigma}{2}m(r_{n,\ell})} \sum_{\boldsymbol{\nu}\in A_{n,\Lambda_\ell}} \left((\nu_n+1)^2 e^{-\frac{\sigma}{2}m(\nu_n+1)} \prod_{k=1,k\neq n}^{N} \left((\nu_k+1)^2 e^{-\sigma m(\nu_k-1)} \right) \right) \\
&\leq (r_{n,\ell}+2)^2\, e^{-\frac{\sigma}{2}m(r_{n,\ell})} \sum_{\boldsymbol{\nu}\in A_{n,\Lambda_\ell}} \left(\prod_{n=1}^{N} (\lambda_n+1) \right)^2 e^{-\frac{\sigma}{2}|m(\boldsymbol{\nu}-\mathbf{1})|_1}.
\end{aligned}$$

The last factor is uniformly bounded with respect to ℓ (but this bound depends on the number of dimensions N)

$$\sum_{\boldsymbol{\nu}\in A_{n,\Lambda_\ell}} \left(\prod_{n=1}^{N} (\lambda_n+1) \right)^2 e^{-\frac{\sigma}{2}|m(\boldsymbol{\nu}-\mathbf{1})|_1} \leq \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^N} \Lambda_{\boldsymbol{\nu}}^2 e^{-\frac{\sigma}{2}|m(\boldsymbol{\nu}-\mathbf{1})|_1} < \infty.$$

All in all, we have

$$\mathcal{P}_{n,\Lambda_\ell} \lesssim (r_{n,\ell}+2)^2\, e^{-\frac{\sigma}{2}m(r_{n,\ell})}$$

and the right-hand side vanishes as $\ell \to \infty$ since $\lim_{\ell\to\infty} r_{n(\ell),\ell} = \infty$. □

For the profit with work, we can even show convergence to zero of the profit without using the analyticity assumption on u. This is not relevant for the problem at hand, as the analyticity follows immediately, but may be relevant for more complicated and less regular random coefficients.

Proposition 2.12. *Consider the profit with work* (2.10). *There holds* $\lim_{\ell\to\infty} \mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = 0$.

Proof. As in the proof of Proposition 2.5, but without using any analyticity of u, we obtain with (2.13) that

$$\zeta_{\boldsymbol{\nu},\Lambda} \lesssim \lambda_{\boldsymbol{\nu}}^2 N \left\|\nabla u\right\|_{L^\infty(\boldsymbol{\Gamma},L^2(D))}. \tag{2.37}$$

We observe that the use of the doubling rule (1.35) implies that

$$m(\nu) - m(\nu-1) = \begin{cases} 2^\nu & \text{if } \nu \le 1 \\ 2^{\nu-1} & \text{otherwise.} \end{cases}$$

Therefore,

$$2^{|\boldsymbol{\nu}|_1 - N} \le W_{\boldsymbol{\nu}} \le 2^{|\boldsymbol{\nu}|_1} \qquad \forall \boldsymbol{\nu} \in \mathbb{N}_0^N. \tag{2.38}$$

Thus, using estimates 2.25a, 2.37 and 2.38, the profit can be estimated as:

$$\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = \frac{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell}}{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} W_{\boldsymbol{j}}} \lesssim \frac{\#A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}\mathbb{L}^2_{\boldsymbol{\nu}_\ell} N}{W_{\boldsymbol{\nu}_\ell}} \le N\left(1+\frac{\ell}{N}\right)^N \mathbb{L}^2_{\boldsymbol{\nu}_\ell} 2^{N-|\boldsymbol{\nu}_\ell|_1}.$$

Since $\lambda^2_{\boldsymbol{\nu}_\ell} \le \left(\prod_{n=1}^N (\langle \boldsymbol{\nu}_\ell, \boldsymbol{e}_n\rangle + 1)\right)^2$ is negligible with respect to $2^{|\boldsymbol{\nu}_\ell|_1}$, we conclude the proof. □

The following result shows that, if a multi-index $\boldsymbol{\nu} \in \mathbb{N}_0^N$ stays in the margin indefinitely, then it's pointwise estimator vanishes. This result is valid for both workless profit and profit with work.

Proposition 2.13. *Let* $\hat{\boldsymbol{\nu}} \in \mathbb{N}_0^N$ *and suppose it remains in the margin indefinitely, i.e.,*

$$\exists \ell_0 \in \mathbb{N} : \forall \ell \ge \ell_0,\ \hat{\boldsymbol{\nu}} \in \mathcal{M}_{\Lambda_\ell}.$$

Then, the pointwise error estimator corresponding to $\hat{\boldsymbol{\nu}}$ *vanishes:*

$$\lim_{\ell\to\infty} \zeta_{\hat{\boldsymbol{\nu}},\Lambda_\ell} = 0.$$

Proof. Let $\hat{\boldsymbol{\nu}} \in \mathbb{N}_0^N$ such that $\hat{\boldsymbol{\nu}} \in \mathcal{M}_{\Lambda_\ell}$ for all $\ell > \ell_0$. Thus, $\hat{\boldsymbol{\nu}} \neq \boldsymbol{\nu}_\ell$ for any $\ell > \ell_0$, which means that $\mathcal{P}_{\hat{\boldsymbol{\nu}},\Lambda_\ell} \le \mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \qquad \forall \ell > \ell_0$. In case the profit with work (2.10) is used, since $\lim_{\ell\to\infty} \mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = 0$ as proved in Proposition 2.12, we have that $\lim_{\ell\to\infty} \mathcal{P}_{\hat{\boldsymbol{\nu}},\Lambda_\ell} = 0$ (otherwise $\hat{\boldsymbol{\nu}}$ would be selected at some point). Moreover, since $\sum_{\boldsymbol{j}\in A_{\hat{\boldsymbol{\nu}},\Lambda_\ell}} W_{\boldsymbol{j}}$ (i.e. the denominator in the profit $\mathcal{P}_{\hat{\boldsymbol{\nu}},\Lambda_\ell}$) is eventually constant with respect to ℓ, we have that $\lim_{\ell\to\infty} \sum_{\boldsymbol{j}\in A_{\hat{\boldsymbol{\nu}},\Lambda_\ell}} \zeta_{\hat{\boldsymbol{\nu}},\Lambda_\ell} = 0$, and in particular we obtain the statement. The same holds if the profit without work (2.9) is employed, as in Proposition 2.11 we have proved that also in this case $\lim_{\ell\to\infty} \mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = 0$. □

Remark 2.14. *Recall the simplified formula* (2.23) *for* $\zeta_{\hat{\boldsymbol{\nu}},\Lambda_\ell}$*, where the (non downward-closed) multi-index is defined as* $J_{\hat{\boldsymbol{\nu}},\Lambda_\ell} \subset \{\hat{\boldsymbol{\nu}} - \boldsymbol{e}_n \ : n \in 1,\dots,N\}$*. Observe that* $\left(J_{\hat{\boldsymbol{\nu}},\Lambda_\ell}\right)_\ell$ *is eventually constant, i.e. it exists* $\ell_2 > \ell_0$ *(as defined in the previous proposition) such that for all* $\ell > \ell_2$ $J_{\hat{\boldsymbol{\nu}},\Lambda_\ell} = J_{\hat{\boldsymbol{\nu}},\Lambda_{\ell_2}}$*. Thus,* $\left(\zeta_{\hat{\boldsymbol{\nu}},\Lambda_\ell}\right)_\ell$ *is also eventually constant. Therefore,* $\left(\zeta_{\hat{\boldsymbol{\nu}},\Lambda_\ell}\right)_\ell$ *does not only vanish in the limit, but is actually eventually zero:*

$$\forall \ell > \ell_2, \zeta_{\hat{\boldsymbol{\nu}},\Lambda_\ell} = 0.$$

We can finally prove the convergence of the parameter-enrichment algorithm with a technique inspired by [BSX22b, Proposition 10].

Theorem 2.15 (Convergence of the parameter-enrichment algorithm)**.** *The adaptive sparse grid interpolation Algorithm 10, driven by either workless profit* (2.9) *or profit with work* (2.10)*, leads to a vanishing sequence of a-posteriori error estimators. Due to the reliability of the error estimator, it also leads to a convergent sequence of discrete solutions to the exact solution*

$$\lim_{\ell\to\infty} \zeta_{SC,\Lambda_\ell} = 0 = \lim_{\ell\to\infty} \|u - \mathcal{I}_{\Lambda_\ell}[u]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})}$$

Proof. The a-posteriori error estimator at step $\ell \in \mathbb{N}_0$ can be written as

$$\zeta_{SC,\Lambda_\ell} = \sum_{\boldsymbol{\nu}\in\mathbb{N}^N} \zeta_{\boldsymbol{\nu},\Lambda_\ell} \mathbb{1}_{\mathcal{M}_{\Lambda_\ell}}(\boldsymbol{\nu}),$$

where $\mathbb{1}_{\mathcal{M}_{\Lambda_\ell}}$ is the indicator function of the margin $\mathcal{M}_{\Lambda_\ell}$. In order to prove that the sequence vanishes by dominated convergence, it is sufficient to prove that (i) for any $\boldsymbol{\nu} \in \mathbb{N}_0^N$, $\lim_{\ell\to\infty} \zeta_{\boldsymbol{\nu},\Lambda_\ell} \mathbb{1}_{\mathcal{M}_{\Lambda_\ell}} = 0$ and (ii) that the sequence $(\zeta_{SC,\Lambda_\ell})_\ell$ is bounded. The uniform boundedness (ii) was proved in Remark 2.6. As for (i), observe that at least one of the following cases applies:

- $\boldsymbol{\nu}$ is eventually added to Λ_ℓ, thus $\mathbb{1}_{\mathcal{M}_{\Lambda_\ell}}(\boldsymbol{\nu})$ is eventually zero;
- $\boldsymbol{\nu}$ is never added to the margin (for all $\ell \in \mathbb{N}_0$, $\boldsymbol{\nu} \in \mathbb{N}_0^N \setminus \mathcal{M}_{\Lambda_\ell}$), thus $\zeta_{\boldsymbol{\nu},\Lambda_\ell}$ is constantly zero;
- it exists $\bar{\ell} \in \mathbb{N}$ such that for any $\ell \geq \bar{\ell}$, $\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}$. In this case, due to Proposition 2.13, $\lim_{\ell\to\infty} \zeta_{\boldsymbol{\nu},\Lambda_\ell} = 0$.

This concludes the proof. □

2.2.3 Convergence of the parametric error

In this section, we denote by $\mathcal{L}(L^\infty(\boldsymbol{\Gamma},\mathbb{U}))$ the space of linear bounded operators $T : L^\infty(\boldsymbol{\Gamma},\mathbb{U}) \to L^\infty(\boldsymbol{\Gamma},\mathbb{U})$. It is well known that this is a Banach space when equipped with the usual operator norm

$$\|T\|_{\mathcal{L}(L^\infty(\boldsymbol{\Gamma},\mathbb{U}))} \coloneqq \sup_{\substack{u\in L^\infty(\boldsymbol{\Gamma},\mathbb{U})\\ u\neq 0}} \frac{\|Tu\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})}}{\|u\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})}}.$$

We have the following monotonicity property of the approximation error of $\mathcal{I}_\Lambda[\cdot]$ with respect to Λ:

Lemma 2.16. *Let* $u \in C^0(\boldsymbol{\Gamma},\mathbb{U})$ *and* $\Lambda, J \subset \mathbb{N}_0^N$ *downward-closed multi-index sets such that* $J \subset \Lambda$*. Then*

$$\|u - \mathcal{I}_\Lambda[u]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} \leq \left(1 + \|\mathcal{I}_\Lambda\|_{\mathcal{L}(L^\infty(\boldsymbol{\Gamma},\mathbb{U}))}\right) \|u - \mathcal{I}_J[u]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})}.$$

Proof. The assumption $J \subset \Lambda$ and the use of nested nodes imply $\mathcal{I}_\Lambda[\mathcal{I}_J[u]] = \mathcal{I}_J[u]$. As a consequence, denoting by $\mathbf{1}$ the identity operator on $C^0(\mathbf{\Gamma}, \mathbb{U})$,

$$u - \mathcal{I}_\Lambda[u] = (\mathbf{1} - \mathcal{I}_\Lambda)\, u = (\mathbf{1} - \mathcal{I}_\Lambda)\,(\mathbf{1} - \mathcal{I}_J)\, u.$$

The triangle inequality concludes the proof. □

In this section, we provide error estimates for $\mathcal{I}_{\Lambda_\ell}$, i.e. the sparse grid interpolant adaptively generated after $\ell \in \mathbb{N}_0$ steps, with respect to the number of iterations ℓ. We consider both the possible definitions of profit (2.9) and (2.10).

Remark 2.17. *The quantity $\|\mathcal{I}_{\Lambda_\ell}\|_{\mathcal{L}(L^\infty(\mathbf{\Gamma},\mathbb{U}))}$ from Lemma 2.16 satisfies*

- Workless profit*: $\Lambda_\ell = Rect_{\boldsymbol{\nu}_{\ell-1}}$, i.e. $\mathcal{I}_{\Lambda_\ell}$ i.e. a tensor product Lagrange interpolant (recall Section 2.2.1). Therefore,*

$$\|\mathcal{I}_{\Lambda_\ell}\|_{\mathcal{L}(L^\infty(\mathbf{\Gamma},\mathbb{U}))} = \left\| \bigotimes_{n=1}^{N} I_{\langle \boldsymbol{\nu}_{\ell-1}, \boldsymbol{e}_n \rangle} \right\|_{\mathcal{L}(L^\infty(\mathbf{\Gamma},\mathbb{U}))} \leq \prod_{n=1}^{N} (1 + \langle \boldsymbol{\nu}_{\ell-1}, \boldsymbol{e}_n \rangle) \leq \left(1 + \frac{\ell}{N}\right)^N, \tag{2.39}$$

  where in the first inequality we used the stability bound for the Lagrange interpolant (2.11) and in the second Lemma 2.7;

- Profit with work*: Partitioning Λ_ℓ with Λ_0 and the sequence $(A_{\boldsymbol{\nu}_m,\Lambda_m})_{m=0}^{\ell-1}$ and using Lemma 2.7 we obtain*

$$\begin{aligned} \|\mathcal{I}_{\Lambda_\ell}\|_{\mathcal{L}(L^\infty(\mathbf{\Gamma},\mathbb{U}))} &\leq \sum_{\boldsymbol{\nu} \in \Lambda_\ell} \|\Delta_{\boldsymbol{\nu}}\|_{\mathcal{L}(L^\infty(\mathbf{\Gamma},\mathbb{U}))} \leq \lambda_{\mathbf{0}} + \sum_{m=0}^{\ell-1} \#A_{\boldsymbol{\nu}_m,\Lambda_m} \lambda_{\boldsymbol{\nu}_m} \\ &\leq 1 + \ell \left(1 + \frac{\ell}{N}\right)^{2N}. \end{aligned} \tag{2.40}$$

We finally prove the parametric error estimates, first with workless profit, then with profit with work with respect to the number of iterations.

Theorem 2.18. *Consider Algorithm 10 with workless profit defined in (2.9). Denote by Λ_ℓ the downward-closed multi-index sets chosen by the algorithm at step $\ell \in \mathbb{N}$ and by $\mathcal{I}_{\Lambda_\ell}[u]$ the corresponding sparse grid approximation of the analytic function $u : \mathbf{\Gamma} \to \mathbb{U}$. Then,*

$$\|u - \mathcal{I}_{\Lambda_\ell}[u]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \lesssim \left(1 + \left(1 + \frac{\ell - 1}{N}\right)^N\right) N \ell^2 e^{-\frac{\sigma}{2} m (1 + \frac{\ell}{N})} \qquad \forall \ell > 0. \tag{2.41}$$

Proof. Fix $\ell > 0$. Recall the definition of $r_{n,\ell}$ from (2.35) and consider the direction

$$\overline{n} = \arg\max_{n \in 1,\dots,N} r_{n,\ell}.$$

With $n(\ell)$ from (2.29), define

$$\ell' \coloneqq \max\left\{\ell' \in 1, \dots, \ell : n(\ell') = \bar{n}\right\},$$

and observe that with each iteration at least one side of the axis aligned rectangle Λ_ℓ is increased by one. Thus, from the definition of $|oln$, we have

$$r_{n(\ell'),\ell'} = r_{\bar{n},\ell} \geq 1 + \frac{\ell}{N}. \tag{2.42}$$

Applying Lemma 2.16 and estimate (2.39) form the previous remark, we bound

$$\|u - \mathcal{I}_{\Lambda_\ell}[u]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \leq \left(1 + \left(1 + \frac{\ell}{N}\right)^N\right) \left\|u - \mathcal{I}_{\Lambda_{\ell'}}[u]\right\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})}.$$

Te reliability of the error estimator proved in [GN18, Proposition 4.3] implies

$$\left\|u - \mathcal{I}_{\Lambda_{\ell'}}[u]\right\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \lesssim \sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} \zeta_{\boldsymbol{\nu},\Lambda_{\ell'}}.$$

Recalling the definition of $A_{n,\Lambda_{\ell'}}$ and $\mathcal{P}_{n,\Lambda_{\ell'}}$ for $n \in 1,\dots,N$ given in Section 2.2.1, we have

$$\sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} \zeta_{\boldsymbol{\nu},\Lambda_{\ell'}} = \sum_{n=1}^{N} \sum_{\boldsymbol{\nu}\in A_{n,\Lambda_{\ell'}}} \zeta_{\boldsymbol{\nu},\Lambda_{\ell'}} = \sum_{n=1}^{N} \mathcal{P}_{n,\Lambda_{\ell'}} \leq N\mathcal{P}_{n(\ell'),\Lambda_{\ell'}}.$$

The profit $\mathcal{P}_{n(\ell'),\Lambda_{\ell'}}$ can now be bounded as a function of $r_{n(\ell'),\ell'}$ as we did in Proposition 2.11:

$$\mathcal{P}_{n(\ell'),\Lambda_{\ell'}} = \sum_{\boldsymbol{j}\in A_{n(\ell'),\Lambda_{\ell'}}} \zeta_{\boldsymbol{j},\Lambda_{\ell'}} \leq \sum_{\boldsymbol{j}\in A_{n(\ell'),\Lambda_{\ell'}}} \left(\prod_{k=1}^{N} j_k + 1\right)^2 e^{-\frac{\sigma}{2}|m(\boldsymbol{j}-\mathbf{1})|_1} \lesssim r_{n(\ell'),\ell'}^2 e^{-\frac{\sigma}{2}m(r_{n(\ell'),\ell'})},$$

where in the first inequality we have applied the estimate (2.36) on $\zeta_{\boldsymbol{j},\Lambda_{\ell'}}$ and in the second we have exploited the fact that, for $\boldsymbol{j} \in A_{n(\ell'),\Lambda_{\ell'}}$, $j_{n(\ell')} = r_{n(\ell'),\ell'} + 1$. Recalling that $1 + \frac{\ell}{N} \leq r_{n(\ell'),\ell'} \leq \ell+1$, we obtain

$$\mathcal{P}_{n(\ell'),\Lambda_{\ell'}} \lesssim \ell^2 e^{-\frac{\sigma}{2}m(1+\frac{\ell}{N})}.$$

□

Let us now prove the analogous result for the algorithm driven by profit with work.

Theorem 2.19. *Consider Algorithm 10 with profit with work defined in (2.10). Denote by Λ_ℓ the downward-closed multi-index sets chosen by the algorithm at step $\ell \in \mathbb{N}_0$ and by $\mathcal{I}_{\Lambda_\ell}[u]$ the corresponding sparse grid approximation of the analytic function $u : \mathbf{\Gamma} \to \mathbb{U}$. Then,*

$$\|u - \mathcal{I}_{\Lambda_\ell}[u]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \lesssim \ell^5 \left(\frac{\ell}{N}\right)^{4N} 2^{\ell\left(1-\frac{1}{N}\right)} e^{-\frac{\sigma}{2}m\left(\ell^{\frac{1}{N}}\right)} \qquad \forall \ell > 0. \tag{2.43}$$

Proof. For brevity, we write $\zeta_{\boldsymbol{\nu}}$, $A_{\boldsymbol{\nu}}$ and $\mathcal{P}_{\boldsymbol{\nu}}$ instead of $\zeta_{\boldsymbol{\nu},\Lambda}$, $A_{\boldsymbol{\nu},\Lambda}$ and $\mathcal{P}_{\boldsymbol{\nu},\Lambda}$ respectively. Fix $\ell \in \mathbb{N}_0$ and consider $\bar{r} \coloneqq \max_{\boldsymbol{\nu}\in\Lambda_\ell} |\boldsymbol{\nu}|_\infty$ and $\bar{n} \in 1,\dots,N$ such that, for some $\boldsymbol{\nu} \in \Lambda_\ell$, $\nu_{\bar{n}} = \bar{r}$. Observe that $\#\Lambda_\ell \geq \ell$, hence

$$\bar{r} \geq \ell^{\frac{1}{N}}.$$

Consider the last step ℓ' in which Λ_ℓ has been extended in direction $\bar{n}$, i.e.,

$$\ell' \coloneqq \max\left\{\ell' \in 1,\dots,\ell : \langle \boldsymbol{\nu}_{\ell'}, \boldsymbol{e}_{\bar{n}}\rangle = \bar{r} \text{ and } \boldsymbol{\nu}_{\ell'} - \boldsymbol{e}_{\bar{n}} \in \Lambda_{\ell'}\right\}. \tag{2.44}$$

Applying estimate (2.40) from Remark 2.17, we can bound

$$\|u - \mathcal{I}_{\Lambda_\ell}[u]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \leq \left((2 + \ell\left(1 + \frac{\ell}{N}\right)^{2N}\right) \left\|u - \mathcal{I}_{\Lambda_{\ell'}}[u]\right\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})}. \tag{2.45}$$

In [GN18, Proposition 4.3], the reliability of the error estimator is proved:

$$\left\|u - \mathcal{I}_{\Lambda_{\ell'}}[u]\right\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} \lesssim \sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} \zeta_{\boldsymbol{\nu}}.$$

Recalling the definition of $\mu_{\Lambda_{\ell'}}$, the set of maximal elements in $\mathcal{M}_{\Lambda_{\ell'}}$ (Definition 2.8), the margin can be represented (but in general not partitioned) as

$$\mathcal{M}_{\Lambda_{\ell'}} = \bigcup_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} A_{\boldsymbol{j}}. \tag{2.46}$$

Thus, we can estimate

$$\begin{aligned}\sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} \zeta_{\boldsymbol{\nu}} \leq \sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} \zeta_{\boldsymbol{\nu}} = \sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \frac{\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} \zeta_{\boldsymbol{\nu}}}{\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}} = \sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \mathcal{P}_{\boldsymbol{j}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}} \\ \leq \mathcal{P}_{\boldsymbol{\nu}_{\ell'}} \sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}} = \left(\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{\nu}_{\ell'}}} \zeta_{\boldsymbol{\nu}}\right) \frac{1}{\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{\nu}_{\ell'}}} W_{\boldsymbol{\nu}}} \left(\sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}}\right),\end{aligned}$$

where in the second inequality we have used the fact that $\mathcal{P}_{\boldsymbol{\nu}_{\ell'}} \geq \mathcal{P}_{\boldsymbol{j}}$ for any $\boldsymbol{j} \in \mathcal{M}_{\Lambda_{\ell'}}$. Let us now estimate each of the three factors separately.

- $\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{\nu}_{\ell'}}} \zeta_{\boldsymbol{\nu}}$: As in the proof of Theorem 2.18 (using the estimate from Proposition 2.5 instead of (2.36)) we obtain with $\ell^{\frac{1}{N}} \leq \bar{r} \leq \ell + 1$ that

$$\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{\nu}_{\ell'}}} \zeta_{\boldsymbol{\nu}} \lesssim N\ell^2 e^{-\frac{\sigma}{2} m\left(\ell^{\frac{1}{N}}\right)}. \tag{2.47}$$

- $\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{\nu}_{\ell'}}} W_{\boldsymbol{\nu}}$: There holds

$$\sum_{\boldsymbol{\nu}\in A_{\boldsymbol{\nu}_{\ell'}}} W_{\boldsymbol{\nu}} \geq W_{\boldsymbol{\nu}_{\ell'}} \geq m(\langle \boldsymbol{\nu}_{\ell'}, \boldsymbol{e}_{\bar{n}}\rangle) - m(\langle \boldsymbol{\nu}_{\ell'}, \boldsymbol{e}_{\bar{n}}\rangle - 1) \geq 2^{\bar{r}-2} \geq 2^{\ell^{\frac{1}{N}}-2} \tag{2.48}$$

- $\sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}}$: We observe

$$\sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}} = \sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} \#\left\{\boldsymbol{j} \in \mu_{\Lambda_{\ell'}} : \boldsymbol{\nu} \in A_{\boldsymbol{j}}\right\} W_{\boldsymbol{\nu}}.$$

  Thus, being $\#\left\{\boldsymbol{j} \in \mu_{\Lambda_{\ell'}} : \boldsymbol{\nu} \in A_{\boldsymbol{j}}\right\} \leq \#\mathcal{M}_{\Lambda_{\ell'}}$, we can estimate

$$\sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}} \leq \#\mathcal{M}_{\Lambda_{\ell'}} \sum_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} W_{\boldsymbol{\nu}} \leq \left(\#\mathcal{M}_{\Lambda_{\ell'}}\right)^2 \max_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_{\ell'}}} W_{\boldsymbol{\nu}}. \tag{2.49}$$

  An estimate for $\#\mathcal{M}_{\Lambda_{\ell'}}$ is given in (2.25b). For the second factor, use the bound on $W_{\boldsymbol{\nu}}$ from (2.38) and the fact that for any $\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}, |\boldsymbol{\nu}|_1 \leq \ell + 1$ to obtain:

$$\sum_{\boldsymbol{j}\in\mu_{\Lambda_{\ell'}}} \sum_{\boldsymbol{\nu}\in A_{\boldsymbol{j}}} W_{\boldsymbol{\nu}} \leq \left(N + N\ell\left(1 + \frac{\ell}{N}\right)^N\right)^2 2^\ell. \tag{2.50}$$

Finally, the statement of the theorem is obtained combining (2.47), (2.48) and (2.50). □

Remark 2.20. *We note that the convergence rates in Theorems 2.18–2.19 above compare the error to the number of adaptive steps ℓ. This is hard to compare to classical a priori results which bound the error in terms of the number of collocation points (see, e.g., [NTW08a, BNT10]). Due to the adaptive nature of the algorithm we have no knowledge about the shape of Λ_ℓ and hence the number of collocation points $\#\mathcal{H}_{\Lambda_\ell}$. Additionally, we do not assume any a priori information about the anisotropy of the solution. Hence, the term $\ell^{1/N}$ is the* worst-case *for a fully isotropic solution. We point out that the observed rate of convergence is much better (see Section 2.4) and further research is required to explain the performance of the adaptive algorithm.*

2.3 Convergence of the adaptive sparse grid–finite element algorithm (ASGFE)

In order to prove the convergence of Algorithm 7, it is sufficient to prove that:

- In Algorithm 8 (the finite element refinement sub-routine) the finite element error eventually falls below the tolerance prescribed in Line 2 (Alg. 8) and iteratively updated in Line 10 (Alg. 8) (proved in Section 2.3.1);
- The parametric estimator ζ_{SC,Λ_ℓ} in Algorithm 7 vanishes (proved in Section 2.3.2).

Indeed, if this is the case, $\eta_{\mathrm{FE},\Lambda_\ell}$ will vanish with ζ_{SC,Λ_ℓ} because of the definition of the finite element refinement tolerance and the reliability of the estimator will ensure the convergence of the discrete solution to the exact one.

In the present section, we write $\zeta_{SC,\Lambda}(\cdot), \zeta_{\boldsymbol{\nu},\Lambda}(\cdot)$ to denote the dependence on the function explicitly. The same will be done for the finite element estimator $\eta_{\mathrm{FE},\Lambda}(\cdot)$. For instance, the parametric estimator (1.44) is written as $\zeta_{SC,\Lambda}(U)$, if we denote by U the current discrete finite element solution. In the previous section, in which we assumed to be able to sample the parameter-to-solution map exactly, we were dealing with $\zeta_{SC,\Lambda}(u)$.

The following lemma is used in the next sections.

Lemma 2.21. *Given a downward-closed multi-index set $\Lambda \subset \mathbb{N}_0^N$, there holds*

$$|\zeta_{SC,\Lambda}(u) - \zeta_{SC,\Lambda}(U)| \lesssim \left(\sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \lambda_{\boldsymbol{\nu}} \right) \eta_{FE,\Lambda}(U).$$

Proof. The definition of parametric estimator and the triangle inequality give

$$\begin{aligned} |\zeta_{SC,\Lambda}(u) - \zeta_{SC,\Lambda}(U)| &\leq \sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} |\zeta_{\boldsymbol{\nu},\Lambda}(u) - \zeta_{\boldsymbol{\nu},\Lambda}(U)| \\ &\leq \sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \|\Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u - U]\right)\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))}. \end{aligned}$$

Recall that $\lambda_{\boldsymbol{\nu}}$ as in 2.13 denotes the Lebesgue constant of $\Delta_{\boldsymbol{\nu}}$ and that the diffusion coefficient a is uniformly bounded by $0 < a_{\max} < \infty$ to estimate

$$|\zeta_{SC,\Lambda}(u) - \zeta_{SC,\Lambda}(U)| \lesssim \left(\sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \lambda_{\boldsymbol{\nu}} \right) \|\nabla \mathcal{I}_\Lambda[u - U]\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))}.$$

Observe now that

$$\|\nabla \mathcal{I}_\Lambda[u-U]\|_{L^\infty(\mathbf{\Gamma},L^2(D))} \leq \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} \|\nabla\left(u(\boldsymbol{y})-U_{\boldsymbol{y}}\right)\|_{L^2(D)} \|L_{\boldsymbol{y}}\|_{L^\infty(\mathbf{\Gamma})}\,.$$

Finally, the reliability of the residual-based error estimator in each collocation point $\boldsymbol{y}$ concludes the proof. □

2.3.1 Convergence under *h*-refinement

The ASGFE algorithm (Algorithm 7) delegates to Algorithm 8 the task of refining the finite element solutions corresponding to the collocation points until the finite element a-posteriori estimator (1.45) falls below a given tolerance. Recall that Algorithm 8 is given a multi-index set Λ, or equivalently a sparse grid $\mathcal{H}_\Lambda$ that will *not* change during its execution. Hence, we will drop the index Λ in the following. In this section, the index $\ell \in \mathbb{N}_0$ will denote the current iteration of the adaptive loop starting at Line 3 (Alg. 8) (so $U_{\ell,\boldsymbol{y}}$ and $\eta_{\ell,\boldsymbol{y}}$ will denote respectively the finite element solution and finite element estimator on the collocation point $\boldsymbol{y}\in\mathcal{H}$ at iteration ℓ and with respect to the mesh $\mathcal{T}_{\ell,\boldsymbol{y}}$).

From the theory of the classical h-adaptive finite element algorithm, we have the following contraction property (see, e.g., [CKNS08, Ste07, CFPP14]) for all $\boldsymbol{y}\in\mathcal{H}$:

$$\sum_{T\in\mathcal{T}_{\ell+1,\boldsymbol{y}}\setminus\mathcal{T}_{\ell,\boldsymbol{y}}} \eta^2_{\ell+1,\boldsymbol{y},T} \leq q \sum_{T\in\mathcal{T}_{\ell,\boldsymbol{y}}\setminus\mathcal{T}_{\ell+1,\boldsymbol{y}}} \eta^2_{\ell,\boldsymbol{y},T} + C\,\|U_{\ell+1,\boldsymbol{y}}-U_{\ell,\boldsymbol{y}}\|^2_{\mathbb{U}}\,, \tag{2.51}$$

as well as

$$\Big(\sum_{T\in\mathcal{T}_{\ell+1,\boldsymbol{y}}\cap\mathcal{T}_{\ell,\boldsymbol{y}}} \eta^2_{\ell+1,\boldsymbol{y},T}\Big)^{1/2} \leq \Big(\sum_{T\in\mathcal{T}_{\ell,\boldsymbol{y}}\cap\mathcal{T}_{\ell+1,\boldsymbol{y}}} \eta^2_{\ell,\boldsymbol{y},T}\Big)^{1/2} + C^{1/2}\,\|U_{\ell+1,\boldsymbol{y}}-U_{\ell,\boldsymbol{y}}\|_{\mathbb{U}} \tag{2.52}$$

for $0<q<1$ and $C>0$ independent of ℓ but depending on the shape-regularity of the mesh and the regularity assumptions on the coefficient $a(\boldsymbol{y},\cdot)$ on $\mathcal{T}_{\text{init}}$. Since we use newest-vertex-bisection for mesh refinement, the shape regularity depends only on $\mathcal{T}_{\text{init}}$.

As in the deterministic setting, Dörfler marking together with (2.51)–(2.52) can be used to prove a contraction property of the estimator (see also [BPRR19] for a similar argument with a slightly different marking strategy).

Proposition 2.22. *Given an arbitrary downward closed index set $\Lambda\subseteq\mathbb{N}_0^N$, Algorithm 8 satisfies*

$$\sum_{\boldsymbol{y}\in\mathcal{H}} \eta^2_{\ell+k,\boldsymbol{y}} \leq C_{lin} q^k_{lin} \sum_{\boldsymbol{y}\in\mathcal{H}} \eta^2_{\ell,\boldsymbol{y}} \tag{2.53}$$

for all $\ell,k\in\mathbb{N}$ and some uniform constants $0<q_{lin}<1$, $C_{lin}>0$. In particular, we have:

$$\lim_{\ell\to\infty} \|\mathcal{I}_\Lambda[u]-\mathcal{I}_\Lambda[U_\ell]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} = 0 = \lim_{\ell\to\infty} \eta_{FE,\Lambda}(U_\ell).$$

Proof. We show with (2.51)–(2.52) that all $\delta>0$ satisfy (recall the definition of disjoint union

$\bigsqcup$ in (2.7))

$$\begin{aligned}
\sum_{\boldsymbol{y}\in\mathcal{H}}\eta_{\ell+1,\boldsymbol{y}}^2 &= \sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell+1,\boldsymbol{y}}\setminus\mathcal{T}_{\ell,\boldsymbol{y}}}\eta_{\ell+1,\boldsymbol{y},T_{\boldsymbol{y}}}^2 + \sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell+1,\boldsymbol{y}}\cap\mathcal{T}_{\ell,\boldsymbol{y}}}\eta_{\ell+1,\boldsymbol{y},T_{\boldsymbol{y}}}^2\\
&\le q\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}\setminus\mathcal{T}_{\ll+1,\boldsymbol{y}}}\eta_{\ell,\boldsymbol{y},T_{\boldsymbol{y}}}^2 + (1+\delta)\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}\cap\mathcal{T}_{\ell+1,\boldsymbol{y}}}\eta_{\ell,\boldsymbol{y},T_{\boldsymbol{y}}}^2\\
&\quad + C(2+\delta^{-1})\sum_{\boldsymbol{y}\in\mathcal{H}}\|U_{\ell+1,\boldsymbol{y}}-U_{\ell,\boldsymbol{y}}\|_{\mathbb{U}}^2\\
&\le (q-1)\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}\setminus\mathcal{T}_{\ell+1,\boldsymbol{y}}}\eta_{\ell,\boldsymbol{y},T_{\boldsymbol{y}}}^2 + (1+\delta)\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}}\eta_{\ell+1,\boldsymbol{y},T_{\boldsymbol{y}}}^2\\
&\quad + C(2+\delta^{-1})\sum_{\boldsymbol{y}\in\mathcal{H}}\|U_{\ell+1,\boldsymbol{y}}-U_{\ell,\boldsymbol{y}}\|_{\mathbb{U}}^2.
\end{aligned}$$

The Dörfler marking from Algorithm 8 ensures $\mathcal{K}\subseteq\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}\setminus\mathcal{T}_{\ell+1,\boldsymbol{y}}$ and hence

$$(q-1)\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}\setminus\mathcal{T}_{\ell+1,\boldsymbol{y}}}\eta_{\ell,\boldsymbol{y},T_{\boldsymbol{y}}}^2 \le \theta(q-1)\sum_{(\boldsymbol{y},T_{\boldsymbol{y}})\in\bigsqcup_{\boldsymbol{y}\in\mathcal{H}}\mathcal{T}_{\ell,\boldsymbol{y}}}\eta_{\ell,\boldsymbol{y},T_{\boldsymbol{y}}}^2.$$

Altogether, we obtain for $\kappa := 1+\delta-\theta(1-q)$ and $\widetilde{C} := C(2+\delta^{-1})$ that

$$\sum_{\boldsymbol{y}\in\mathcal{H}}\eta_{\ell+1,\boldsymbol{y}}^2 \le \kappa\sum_{\boldsymbol{y}\in\mathcal{H}}\eta_{\ell,\boldsymbol{y}}^2 + \widetilde{C}\sum_{\boldsymbol{y}\in\mathcal{H}}\|U_{\ell+1,\boldsymbol{y}}-U_{\ell,\boldsymbol{y}}\|_{\mathbb{U}}^2.$$

With the Galerkin orthogonality

$$\begin{aligned}
&\sum_{\boldsymbol{y}\in\mathcal{H}}\left\|a(\boldsymbol{y})^{1/2}\nabla(U_{\ell+1,\boldsymbol{y}}-U_{\ell,\boldsymbol{y}})\right\|_{L^2(D)}^2\\
&\quad=\sum_{\boldsymbol{y}\in\mathcal{H}}\Big(\left\|a(\boldsymbol{y})^{1/2}\nabla(u(\boldsymbol{y})-U_{\ell,\boldsymbol{y}})\right\|_{L^2(D)}^2-\left\|a(\boldsymbol{y})^{1/2}\nabla(u(\boldsymbol{y})-U_{\ell+1,\boldsymbol{y}})\right\|_{L^2(D)}^2\Big)
\end{aligned}$$

we may follow [CFPP14, Section 4] verbatim in order to prove (2.53). Since $\#\mathcal{H}$ is fixed, we have $\sum_{\boldsymbol{y}\in\mathcal{H}}\eta_{\ell,\boldsymbol{y}}^2\simeq\eta_{\mathrm{FE},\Lambda}(U_\ell)^2$ and reliability proves

$$\lim_{\ell\to\infty}\eta_{\mathrm{FE},\Lambda}(U_\ell)=\lim_{\ell\to\infty}\|\mathcal{I}_\Lambda[u]-\mathcal{I}_\Lambda[U_\ell]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})}=0.$$

This concludes the statement. □

Remark 2.23. *The previous proposition implies that Algorithm 8 terminates. In particular, the algorithm will eventually satisfy the condition $\eta_{FE,\Lambda}(U_\ell) < Tol_\ell$, where $Tol_\ell := \alpha\left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}\Lambda_{\boldsymbol{\nu}}\right)^{-1}\zeta_{SC,\Lambda}(U_\ell)$. Indeed, due to Lemma 2.21, we have that, as $(\eta_{FE,\Lambda_\ell}(U_\ell))_\ell$ vanishes, $\zeta_{SC,\Lambda}(U_\ell)$ converges to $\zeta_{SC,\Lambda}(u)\ge\varepsilon>0$. Therefore,*

$$\lim_{\ell\to\infty}Tol_\ell=\alpha\left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda}\Lambda_{\boldsymbol{\nu}}\right)^{-1}\zeta_{SC,\Lambda}(u)>0.$$

Note that the convergence proof uses an ℓ_2-type estimator instead of an ℓ_1-type as in $\eta_{FE,\Lambda}$. In this regard, the ℓ_2-type might seem more natural and we refer to Section 2.3.3 for further discussion.

Theorem 2.24. *Given an arbitrary downward closed index set $\Lambda \subseteq \mathbb{N}_0^N$, Algorithm 8 converges with the optimal rate in the following sense: Let $\mathbb{T}$ denote the set of all meshes which can be obtained from $\mathcal{T}_{init}$ by iterated newest-vertex-bisection with mesh closure. Let $s > 0$ such that*

$$\sup_{N\in\mathbb{N}} \inf_{\substack{\mathcal{T}_{\boldsymbol{y}}\in\mathbb{T}\\ \sum_{\boldsymbol{y}\in\mathcal{H}} \#\mathcal{T}_{\boldsymbol{y}}\le N}} \Big(\sum_{\boldsymbol{y}\in\mathcal{H}} \big\|u(\boldsymbol{y}) - U_{\mathcal{T}_{\boldsymbol{y}}}\big\|_{\mathbb{U}}^2 + \big\|h_{\mathcal{T}_{\boldsymbol{y}}}(1-\Pi_{\mathcal{T}_{\boldsymbol{y}}})f\big\|_{L^2(D)}^2 \Big)^{1/2} N^s < \infty, \tag{2.54}$$

where $h_{\mathcal{T}}$ denotes the local mesh-size function and $\Pi_{\mathcal{T}}$ is the $L^2(D)$-orthogonal projection onto $\mathcal{T}$-elementwise constant functions. Then, there holds

$$\sup_{\ell\in\mathbb{N}} \|\mathcal{I}_\Lambda[u] - \mathcal{I}_\Lambda[U_\ell]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} \Big(\sum_{\boldsymbol{y}\in\mathcal{H}} \#\mathcal{T}_{\ell,\boldsymbol{y}} \Big)^s < \infty.$$

Proof. First note that standard upper/lower bounds for the residual error estimator together with the regularity assumptions on $a(\boldsymbol{y},\cdot)$ show $\eta_{\boldsymbol{y}} \simeq \sqrt{\big\|u(\boldsymbol{y}) - U_{\mathcal{T}_{\boldsymbol{y}}}\big\|_{\mathbb{U}}^2 + \big\|h_{\mathcal{T}_{\boldsymbol{y}}}(1-\Pi_{\mathcal{T}_{\boldsymbol{y}}})f\big\|_{L^2(D)}^2}$ and hence (2.54) is equivalent to

$$\sup_{N\in\mathbb{N}} \inf_{\substack{\mathcal{T}_{\boldsymbol{y}}\in\mathbb{T}\\ \sum_{\boldsymbol{y}\in\mathcal{H}} \#\mathcal{T}_{\boldsymbol{y}}\le N}} \Big(\sum_{\boldsymbol{y}\in\mathcal{H}} \eta_{\boldsymbol{y}}^2 \Big)^{1/2} N^s < \infty.$$

With the error norm $|||u - U_\ell||| := \sqrt{\sum_{\boldsymbol{y}\in\mathcal{H}} \big\|u(\boldsymbol{y}) - U_{\mathcal{T}_{\ell,\boldsymbol{y}}}\big\|_{\mathbb{U}}^2}$ and (2.51)–(2.52), the estimator $\eta_{\mathrm{FE},\Lambda}$ satisfies (A1) and (A2) from [CFPP14, Section 3]. From the classical theory of h-adaptivity [CKNS08], we immediately obtain discrete reliability (A3) in the sense

$$|||U_{\ell+k} - U_\ell|||^2 = \sum_{\boldsymbol{y}\in\mathcal{H}} \|U_{\ell+k,\boldsymbol{y}} - U_{\ell,\boldsymbol{y}}\|_{\mathbb{U}}^2 \le C_{\mathrm{drel}} \sum_{\boldsymbol{y}\in\mathcal{H}} \sum_{T\in\omega(\mathcal{T}_{\ell,\boldsymbol{y}}\setminus\mathcal{T}_{\ell+k,\boldsymbol{y}})} \eta_{\ell,\boldsymbol{y},T}^2,$$

where $\omega(\cdot)$ denotes the set of elements with non-empty intersection with $(\cdot)$. With these ingredients and the linear convergence from Proposition 2.22, [CFPP14, Proposition 4.12 & Proposition 4.15] show optimal convergence of the error estimator

$$\sup_{\ell\in\mathbb{N}} \sqrt{\sum_{\boldsymbol{y}\in\mathcal{H}} \eta_{\ell,\boldsymbol{y}}^2} \Big(\sum_{\boldsymbol{y}\in\mathcal{H}} \#\mathcal{T}_{\ell,\boldsymbol{y}} \Big)^s < \infty.$$

With constants depending only on the size of Λ, the quantity $\sqrt{\sum_{\boldsymbol{y}\in\mathcal{H}} \eta_{\ell,\boldsymbol{y}}^2}$ is equivalent to $\eta_{\mathrm{FE},\Lambda}$ and hence reliability concludes the proof. □

2.3.2 Proof of convergence of the ASGFE algorithm

The tolerance for finite element refinement was defined in Algorithm 8 as:

$$\mathrm{Tol} = \mathrm{Tol}(\Lambda, \zeta_{SC,\Lambda}(U); \alpha) := \alpha \frac{1}{\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \lambda_{\boldsymbol{\nu}}} \zeta_{SC,\Lambda}(U). \tag{2.55}$$

where $\alpha \in (0,1)$, $\Lambda_{\boldsymbol{\nu}}$ was defined in (2.14) and $\zeta_{SC,\Lambda}(U)$ is the parametric a-posteriori error estimator. This choice is motivated by the following estimate: For fixed downward closed $\Lambda \subset \mathbb{N}_0^N$, Lemma 2.21 shows

$$\zeta_{SC,\Lambda}(U) \le \zeta_{SC,\Lambda}(u) + \left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \lambda_{\boldsymbol{\nu}} \right) \eta_{\mathrm{FE},\Lambda}(U) \le \zeta_{SC,\Lambda}(u) + \alpha\zeta_{SC,\Lambda}(U),$$

and hence

$$\zeta_{SC,\Lambda}(U) \leq \frac{1}{1-\alpha}\zeta_{SC,\Lambda}(u). \tag{2.56}$$

In the context of the adaptive algorithm, this implies that $(\zeta_{SC,\Lambda_\ell}(U_\ell))_\ell$ is uniformly bounded since $(\zeta_{SC,\Lambda_\ell}(u))_\ell$ is. This last fact was proved in Remark 2.6 using the estimate on the pointwise error estimator from Proposition 2.5.

Lemma 2.25. *Consider Algorithm 7 with either workless profit* (2.9) *and* $0 < \alpha < 1$ *sufficiently small, or profit with work* (2.10) *and arbitrary* $0 < \alpha < 1$*. Define the finite element refinement tolerance as* (2.55)*. The sequence of adaptive refinements is such that, denoting* $\boldsymbol{\nu}_\ell$ *the profit maximizer at step* $\ell \in \mathbb{N}_0$*,* $\lim_{\ell\to\infty} \mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} = 0$*.*

Proof. We consider the two definitions of profit separately:

Profit with work: $\mathcal{P}_{\boldsymbol{\nu},\Lambda} \coloneqq \frac{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu},\Lambda}} \zeta_{\boldsymbol{j},\Lambda}(U)}{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu},\Lambda}} W_{\boldsymbol{j}}}$: The uniform boundedness of the parametric a-posteriori error estimator, together with the fact that the work over $A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}$ diverges, gives

$$\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \leq \frac{\zeta_{SC,\Lambda_\ell}(U_\ell)}{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} W_{\boldsymbol{j}}} \lesssim \frac{1}{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} W_{\boldsymbol{j}}} \to 0.$$

Workless profit: $\mathcal{P}_{\boldsymbol{\nu},\Lambda} \coloneqq \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu},\Lambda}} \zeta_{\boldsymbol{j},\Lambda}(U)$. We recall that, for the profit-maximizer $\boldsymbol{\nu}_\ell \in \mathcal{M}_{\Lambda_\ell}$, $\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \geq \frac{1}{N}\zeta_{SC,\Lambda_\ell}(U)$ (2.46). Thus, Lemma 2.21 shows

$$\begin{aligned}
\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} &\leq \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell}(u) + \alpha\frac{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \lambda_{\boldsymbol{j}}}{\sum_{\boldsymbol{j}\in \mathcal{M}_{\Lambda_\ell}} \lambda_{\boldsymbol{j}}}\zeta_{SC,\Lambda_\ell}(U_\ell) \\
&\leq \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell}(u) + \alpha\frac{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \lambda_{\boldsymbol{j}}}{\sum_{\boldsymbol{j}\in \mathcal{M}_{\Lambda_\ell}} \lambda_{\boldsymbol{j}}}N\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \\
&\leq \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell}(u) + \alpha N\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda_\ell},
\end{aligned}$$

so

$$\mathcal{P}_{\boldsymbol{\nu}_\ell,\Lambda} \leq \frac{1}{1-\alpha N} \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell}(u) \to 0 \qquad \text{as } \ell \to \infty.$$

Observe that this introduces the constraint on α with respect to the number of dimensions: $\alpha < N^{-1}$. This constraint can be improved by replacing the crude estimate

$$\frac{\sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \Lambda_{\boldsymbol{j}}}{\sum_{\boldsymbol{j}\in \mathcal{M}_{\Lambda_\ell}} \Lambda_{\boldsymbol{j}}} \leq 1,$$

with the better bound

$$\alpha \leq \left(\frac{\max_{n\in 1,\dots,N} \sum_{\boldsymbol{j}\in A_{\boldsymbol{\nu}_{\ell-1}+\boldsymbol{e}_n,\Lambda_\ell}} \Lambda_{\boldsymbol{j}}}{\sum_{\boldsymbol{j}\in \mathcal{M}_{\Lambda_\ell}} \Lambda_{\boldsymbol{j}}} N\right)^{-1}.$$

This concludes the proof. □

We can finally prove that the error estimator vanishes with a technique similar to that used in Theorem 2.15 for the parametric algorithm.

Theorem 2.26. *Algorithm 7 with either workless profit (and $0 < \alpha < 1$ sufficiently small) or profit with work (and arbitrary $0 < \alpha < 1$) and the tolerance (2.55) satisfies the following: The sequence of parametric a-posteriori error estimators $(\zeta_{SC,\Lambda_\ell}(U_\ell))_\ell$ vanishes*

$$\lim_{\ell\to\infty} \zeta_{SC,\Lambda_\ell}(U_\ell) = 0.$$

Thus, also the finite element error estimator vanishes

$$\lim_{\ell\to\infty} \eta_{FE,\Lambda_\ell}(U_\ell) = 0,$$

and the reliability of the a-posteriori error estimator implies error convergence

$$\lim_{\ell\to\infty} \|u - \mathcal{I}_{\Lambda_\ell}[U_\ell]\|_{L^\infty(\mathbf{\Gamma},\mathbb{U})} = 0.$$

Proof. The a-posteriori error estimator can be expressed as

$$\zeta_{SC,\Lambda_\ell}(U_\ell) = \sum_{\boldsymbol{\nu}\in\mathbb{N}_0^N} \zeta_{\boldsymbol{\nu},\Lambda_\ell}(U_\ell)\mathbb{1}_{\mathcal{M}_{\Lambda_\ell}}(\boldsymbol{\nu}).$$

Since the sequence $(\zeta_{SC,\Lambda_\ell}(U_\ell))_\ell$ is uniformly bounded (2.56), it is sufficient to prove that $\left(\zeta_{\boldsymbol{\nu},\Lambda_\ell}(U_\ell)\mathbb{1}_{\mathcal{M}_{\Lambda_\ell}}\right)_\ell$ vanishes for any fixed $\boldsymbol{\nu}\in\mathbb{N}_0^N$. We can distinguish three cases:

- If $\boldsymbol{\nu}$ is eventually added to Λ_ℓ, then $\mathbb{1}_{\mathcal{M}_{\Lambda_\ell}}(\boldsymbol{\nu})$ is eventually zero;
- If $\boldsymbol{\nu}$ is never added to the margin $\mathcal{M}_{\Lambda_\ell}$, then $\zeta_{\boldsymbol{\nu},\Lambda_\ell}(U_\ell)$ is constantly zero;
- Finally, if it exists $\bar{\ell}\in\mathbb{N}_0$ such that for all $\ell > \bar{\ell}$, $\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_\ell}$, then $\lim_{\ell\to\infty}\zeta_{\boldsymbol{\nu},\Lambda_\ell}(U_\ell) = 0$. Indeed, because of Lemma 2.25, $\lim_{\ell\to\infty}\mathcal{P}_{\boldsymbol{\nu},\Lambda_\ell} = 0$ (for both workless profit and profit with work), thus $(\zeta_{\boldsymbol{\nu},\Lambda_\ell}(U_\ell))_\ell$ vanishes as in Proposition 2.13.

This concludes the proof. □

2.3.3 Alternative finite element estimators

In the previous section we followed [GN18] to derive the estimator via

$$\begin{aligned}\mathcal{I}_\Lambda\left[\int_D fv - a\nabla\mathcal{I}_\Lambda[U]\cdot\nabla v\right] &= \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}\left[\int_D fv - a(\boldsymbol{y})\nabla\mathcal{I}_\Lambda[U](\boldsymbol{y})\cdot\nabla v\right]L_{\boldsymbol{y}}\\ &\le C\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}\eta_{\boldsymbol{y}}|L_{\boldsymbol{y}}|\,\|\nabla v\|_{L^2(D)}.\end{aligned}$$

Choosing $v = u - \mathcal{I}_\Lambda[U]$ and taking the $L^\infty(\mathbf{\Gamma})$ norm leads to the estimator we used above, i.e.,

$$\left\|\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}\eta_{\boldsymbol{y}}|L_{\boldsymbol{y}}|\right\|_{L^\infty(\mathbf{\Gamma})} \le \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}\eta_{\boldsymbol{y}}\,\|L_{\boldsymbol{y}}\|_{L^\infty(\mathbf{\Gamma})} = \eta_{\mathrm{FE},\Lambda}(U).$$

Using the Hölder estimates with other combinations of $(p,q)\in\{(2,2),(\infty,1)\}$, we obtain

$$\left\|\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}\eta_{\boldsymbol{y}}|L_{\boldsymbol{y}}|\right\|_{L^\infty(\mathbf{\Gamma})} \le \left\|\left(\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}\eta_{\boldsymbol{y}}^2\right)^{\frac{1}{2}}\left(\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda}|L_{\boldsymbol{y}}|^2\right)^{\frac{1}{2}}\right\|_{L^\infty(\mathbf{\Gamma})} = \eta_{p,\Lambda}\lambda_{q,\Lambda},$$

where

$$\eta_{p,\Lambda} \coloneqq \begin{cases} \left(\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} \eta_{\boldsymbol{y}}^2\right)^{\frac{1}{2}} & p=2, \\ \max_{\boldsymbol{y}\in\mathcal{H}_\Lambda} \eta_{\boldsymbol{y}} & p=\infty, \end{cases} \quad \text{and} \quad \lambda_{q,\Lambda} \coloneqq \begin{cases} \left\| \left(\sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} |L_{\boldsymbol{y}}|^2\right) \right\|_{L^\infty(\boldsymbol{\Gamma})}^{\frac{1}{2}} & q=2, \\ \left\| \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} |L_{\boldsymbol{y}}| \right\|_{L^\infty(\boldsymbol{\Gamma})} & q=1. \end{cases}$$

The perturbation result from Lemma 2.21 can be analogously modified to obtain:

$$|\zeta_{SC,\Lambda}(u) - \zeta_{SC,\Lambda}(U)| \lesssim \left(\sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \lambda_{\boldsymbol{\nu}}\right) \eta_{p,\Lambda}\lambda_{q,\Lambda}$$

From these results, the sufficient condition in (2.55) for convergence becomes respectively

$$\eta_{p,\Lambda}(U) \le \alpha \left(\lambda_{q,\Lambda} \sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \lambda_{\boldsymbol{\nu}}\right)^{-1} \zeta_{SC,\Lambda}(U).$$

With these ingredients, all the other results of the previous sections hold for the variants of the finite element estimator discussed above.

2.3.4 Convergence of a single mesh version of the ASGFE algorithm

We also consider the ASGFE algorithm with the same *adaptively* refined mesh in all collocation points. The idea is that, if the set of singularities of the solution u depends weakly on the value of the parameter $\boldsymbol{y} \in \boldsymbol{\Gamma}$, one single adaptive mesh can resolve all of them simultaneously and thus substantially reduce the computational effort. We employ the following estimator from [GN18, Remark 4.4] for the finite element part:

$$\eta_{\mathrm{FE},\Lambda}(U) \coloneqq \left(\sum_{T\in\mathcal{T}} \eta_T^2(U)\right)^{1/2}, \qquad \eta_T(U) \coloneqq \|\eta_T(\,\cdot\,;U)\|_{L^\infty(\boldsymbol{\Gamma})},$$
$$\eta_T^2(\boldsymbol{y};U) \coloneqq h_T^2 \left\|\mathcal{I}_\Lambda\left[f + \nabla\cdot(a\nabla U)\right](\boldsymbol{y})\right\|_{L^2(T)}^2 + \sum_{e\subset\partial T} h_T \left\|\mathcal{I}_\Lambda\left[[a\nabla U\cdot\boldsymbol{n}_e]_{\boldsymbol{n}_e}\right](\boldsymbol{y})\right\|_{L^2(e)}^2.$$

Since we use a single mesh for all collocation points $\boldsymbol{y} \in \mathcal{H}_\Lambda$, we replace $\mathcal{T}_{\boldsymbol{y}}$ in Algorithm 8 by $\mathcal{T}$. We change the Dörfler marking in Line 4 (Alg. 8) to: Find minimal $\mathcal{K} \subseteq \mathcal{T}$ such that

$$\sum_{T\in\mathcal{K}} \eta_T(U)^2 \ge \theta \eta_{\mathrm{FE},\Lambda}^2.$$

Moreover, we replace the refinement loop in Line 5 (Alg. 8) by a single refinement of the mesh $\mathcal{T}$ with marked elements $\mathcal{K}$.

Due to the fact that $U\colon \boldsymbol{\Gamma} \to \mathcal{S}_0^1(\mathcal{T})$ admits a holomorphic extension to $\Sigma(\boldsymbol{\Gamma},\boldsymbol{\tau})$ just like u does (the same arguments work also for the discrete approximation), the convergence analysis of the parametric enrichment algorithm remains unchanged (Section 2.2), we now have to show convergence of the adaptive finite element subroutine. With this, we may analogously employ the results of Section 2.3.2 to obtain convergence of the full algorithm. Note that we cannot directly transfer the proof of Proposition 2.22 as the definition of $\eta_{\mathrm{FE},\Lambda}$ in this section mixes L^2-norms and L^∞-norms.

In this setting, the multi-index set $\Lambda \subset \mathbb{N}_0^N$ is fixed. We denote by $\mathcal{T}_\ell$ the finite element mesh at step $\ell > 0$ (the same for every collocation point). U_ℓ represents the discrete solution at step ℓ and $U_{\ell,\boldsymbol{y}} \in \mathcal{S}^1_0(\mathcal{T}_\ell)$ its value on a collocation point $\boldsymbol{y} \in \mathcal{H}_\Lambda$. We simplify the notation for the estimator as $\eta_\ell \coloneqq \eta_{\mathrm{FE},\Lambda}(U_\ell)$, $\eta_\ell(U) \coloneqq \left(\sum_{T\in\mathcal{T}_\ell} \eta_T(U)^2\right)^{1/2}$.
We first give a perturbation estimate localized on one element T of a mesh $\mathcal{T}$, analogously to [CKNS08, Proposition 3.3].

Lemma 2.27. *Consider a shape-regular mesh $\mathcal{T}$ obtained by NVB from a mesh $\mathcal{T}_{init}$. There holds for $U, W \in C^0(\boldsymbol{\Gamma}, \mathcal{S}^1_0(\mathcal{T}))$ that*

$$\eta_T(U) \le \eta_T(W) + C\,\|\mathcal{I}_\Lambda\|_{\mathcal{L}(L^\infty(\boldsymbol{\Gamma}))} \max_{\boldsymbol{y}\in\mathcal{H}} \|\nabla(U(\boldsymbol{y}) - W(\boldsymbol{y}))\|_{L^2(\omega(T))} \qquad \forall T \in \mathcal{T}, \tag{2.57}$$

where $\omega(T)$ is the union of the elements sharing an edge with T, $C > 0$ depends only on a and $\mathcal{T}_{init}$.

Proof. For any fixed $\boldsymbol{y} \in \boldsymbol{\Gamma}$, the linearity of $\mathcal{I}_\Lambda$ and the triangle inequality yield

$$\begin{aligned}\eta_T(\boldsymbol{y}; U) \le \eta_T(\boldsymbol{y}; W) &+ h_T\,\|\mathcal{I}_\Lambda\left[\nabla\cdot(a\nabla(U-W))\right](\boldsymbol{y})\|_{L^2(T)} \\ &+ h_T^{1/2} \sum_{e\subset\partial T} \|\mathcal{I}_\Lambda\left[[a\nabla(U-W)\cdot\boldsymbol{n}_e]_{\boldsymbol{n}_e}\right](\boldsymbol{y})\|_{L^2(e)}\,.\end{aligned}$$

With the operator norm of $\mathcal{I}_\Lambda$, we obtain

$$\|\mathcal{I}_\Lambda\left[\nabla\cdot(a\nabla(U-W))\right](\boldsymbol{y})\|_{L^2(T)} \le \|\mathcal{I}_\Lambda\|_{\mathcal{L}(L^\infty(\boldsymbol{\Gamma}))} \max_{\boldsymbol{y}\in\mathcal{H}} \|\nabla a(\boldsymbol{y})\|_{L^\infty(T))}\,\|\nabla(U-W)(\boldsymbol{y})\|_{L^2(T)}\,.$$

Analogously, for the jump terms $[a\nabla(U-W)\cdot\boldsymbol{n}_e]_{\boldsymbol{n}_e}$ with $e \subset \partial T$, we obtain, following the same steps as in [CKNS08, Proposition 3.3],

$$\begin{aligned}&\sum_{e\subset\partial T} \|\mathcal{I}_\Lambda\left[[a\nabla(U-W)\cdot\boldsymbol{n}_e]_{\boldsymbol{n}_e}\right](\boldsymbol{y})\|_{L^2(e)} \\ &\qquad\lesssim \|\mathcal{I}_\Lambda\|_{\mathcal{L}(L^\infty(\boldsymbol{\Gamma}))} \max_{\boldsymbol{y}\in\mathcal{H}} \|a(\boldsymbol{y})\|_{L^\infty(\omega(T)))}\,\|\nabla(U-W)(\boldsymbol{y})\|_{L^2(\omega(T))}\,.\end{aligned}$$

This concludes the proof. □

Proposition 2.28. *The sequence of finite element estimators η_ℓ obtained from the single mesh adaptive algorithm satisfies*

$$\lim_{\ell\to\infty} \eta_\ell = 0.$$

Proof. With the perturbation estimate from Lemma 2.27, we may follow [CFPP14, Section 4.3] to show estimator reduction: There exist $0 < q < 1$, $C > 0$ such that

$$\eta_{\ell+1}^2 \le q\eta_\ell^2 + C^2\,\|\mathcal{I}_\Lambda\|^2_{\mathcal{L}(L^\infty(\boldsymbol{\Gamma}))} \sum_{T\in\mathcal{T}_\ell} \max_{\boldsymbol{y}\in\mathcal{H}} \|\nabla(U_{\ell+1}(\boldsymbol{y}) - U_\ell(\boldsymbol{y}))\|^2_{L^2(\omega(T))} \qquad \forall \ell \in \mathbb{N}. \tag{2.58}$$

To show that the second term in (2.58) vanishes, we first observe that $(U_\ell(\boldsymbol{y}))_{\ell\in\mathbb{N}}$ converges in $\mathbb{U}$ for all $\boldsymbol{y} \in \mathcal{H}$. Indeed, for any fixed $\boldsymbol{y} \in \mathcal{H}$, the nestedness of the finite element spaces $\mathbb{U}_\ell$ guarantees the existence of $U_\infty(\boldsymbol{y}) \in \mathbb{U}$ such that $\lim_{\ell\to\infty} \|U_\infty(\boldsymbol{y}) - U_\ell(\boldsymbol{y})\|_{\mathbb{U}} = 0$ by Céa's lemma (see, e.g., [CFPP14, Section 3.6]). This implies that

$$\lim_{\ell\to\infty} \|\nabla(U_{\ell+1} - U_\ell)(\boldsymbol{y})\|_{L^2(D)} = 0 \tag{2.59}$$

for all $\boldsymbol{y} \in \mathcal{H}$. Since $\#\mathcal{H}$ is fixed in the finite element refinement loop of the adaptive algorithm, we have

$$\begin{aligned}\sum_{T\in\mathcal{T}_\ell} \max_{\boldsymbol{y}\in\mathcal{H}} \|\nabla(U_{\ell+1}(\boldsymbol{y}) - U_\ell(\boldsymbol{y}))\|^2_{L^2(\omega(T))} &\leq \sum_{T\in\mathcal{T}_\ell}\sum_{\boldsymbol{y}\in\mathcal{H}} \|\nabla(U_{\ell+1}(\boldsymbol{y}) - U_\ell(\boldsymbol{y}))\|^2_{L^2(\omega(T))} \\ &\lesssim \sum_{\boldsymbol{y}\in\mathcal{H}} \|\nabla(U_{\ell+1}(\boldsymbol{y}) - U_\ell(\boldsymbol{y}))\|^2_{L^2(D)} \to 0\end{aligned}$$

as $\ell \to \infty$. Passing to the limit superior in (2.58) shows $0 \leq \limsup_{\ell\to\infty} \eta_{\ell+1} \leq q \limsup_{\ell\to\infty} \eta_\ell$ and thus concludes $\lim_{\ell\to\infty} \eta_\ell = 0$. □

Altogether, we obtain the convergence result analogously to Theorem 2.26.

Theorem 2.29. *The single mesh ASGFE algorithm discussed in this section satisfies the following: The sequence of parametric a-posteriori error estimators $(\zeta_{SC,\Lambda_\ell}(U_\ell))_\ell$ vanishes*

$$\lim_{\ell\to\infty} \zeta_{SC,\Lambda_\ell}(U_\ell) = 0.$$

Thus, also the finite element error estimator vanishes

$$\lim_{\ell\to\infty} \eta_{FE,\Lambda_\ell}(U_\ell) = 0,$$

and the reliability of the a-posteriori error estimator implies error convergence

$$\lim_{\ell\to\infty} \|u - \mathcal{I}_{\Lambda_\ell}[U_\ell]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} = 0.$$

2.3.5 Cost of the ASGFE algorithms

Under the assumption that the pointwise estimators $\zeta_{\boldsymbol{\nu},\Lambda}(U)$ and $\eta_{\boldsymbol{y},T}(U)$ can be computed from the discrete solution in $\mathcal{O}(1)$, each step of the adaptive loop (all algorithms) is linear with respect to the number of degrees of freedom of the current sparse grid and spatial meshes. Indeed, a properly preconditioned iterative solver computes U in linear cost (depending on $a_{\min}$ and $a_{\max}$). The Dörfler marking in Algorithm 8 requires sorting when done in a naive way, but can be improved to linear cost by binning [Ste07] or by a clever variation of the quick-select algorithm [PP20]. Finally, the refinement of the finite element meshes $\mathcal{T}_{\boldsymbol{y}}$ via newest-vertex-bisection can be done in linear cost [Ste08].

As discussed in Section 2.4 below, the computation of the $L^\infty(\boldsymbol{\Gamma})$ and $L^2(D)$-norms for $\zeta_{\boldsymbol{\nu},\Lambda}$ is done via a random sample/Monte-Carlo procedure. This results in constant cost $\mathcal{O}(1)$ and the numerical experiments below show that the approximation error is negligible. A precise convergence analysis of this procedure would be interesting but is beyond the scope of this work. Each random sample requires the evaluation of the sparse grid interpolant. Theoretically, the cost of the evaluation of the sparse grid interpolation operator is linear in terms of collocation points, after a quadratic set-up cost. Practically, however, the cost of computing the discrete solutions is expected to dominate significantly.

2.4 Numerical experiments

The Matlab implementation of Algorithm 7 used to produce the numerical results presented in this section is based on two Matlab libraries. For sparse grid interpolation algorithms, the *Sparse Grids Kit* [BNTT11b] was used. The implementation of the adaptive P1 finite element methods is from the *p1afem* Matlab package [FPW11], which uses Matlab's direct solver for

sparse matrices. For further details about parameters and algorithm used within these libraries, the reader is referred to the respective documentations. The parts of the algorithm that deal with parameter enrichment (e.g. Algorithm 9) were implemented following the guidelines from [GN18].

In order to compute the $L^\infty(\boldsymbol{\Gamma})$ norm approximately, we consider a set Θ of 500 uniformly distributed random points in $\boldsymbol{\Gamma}$ and approximate, for any $g \in C^0(\boldsymbol{\Gamma})$, $\|g\|_{L^\infty(\boldsymbol{\Gamma})} \approx \max_{\boldsymbol{y}\in\Theta} |g(\boldsymbol{y})|$. The computation of the $L^2(D)$ norm is carried out with Monte Carlo integration: Given $f \in L^2(D)$, we denote by Π a set of 500 uniformly distributed random points in D and approximate $\|f\|^2_{L^2(D)} \approx \frac{1}{\#\Pi} \sum_{\boldsymbol{x}\in\Pi} f(\boldsymbol{x})^2$. The reason Monte Carlo integration is used is that for a generic $\boldsymbol{y} \in \boldsymbol{\Gamma}$ the discrete solution $\mathcal{I}_\Lambda[U](\boldsymbol{y})$ belongs to the finite element space $\mathcal{S}^1_0(\mathcal{T})$, where $\mathcal{T}$ is the coarsest common refinement of the meshes $\{\mathcal{T}_{\boldsymbol{y}} : \boldsymbol{y} \in \mathcal{H}_\Lambda\}$. Therefore, in order to compute the exact $L^2(D)$ norm of the function, it would be necessary to compute $\mathcal{T}$, which would lead to a significant computational overhead. In numerical experiments, we have observed that increasing $\#\Pi$ does not lead to a significant improvement in the approximation of the $L^2(D)$ norm, thus suggesting that the approximation error can be neglected. In the numerical examples presented in the next sections, we approximate the error between the exact solution u and a discrete solution $\mathcal{I}_\Lambda[U]$ by $\|u - \mathcal{I}_\Lambda[U]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})} \approx \|u_{\text{approx}} - \mathcal{I}_\Lambda[U]\|_{L^\infty(\boldsymbol{\Gamma},\mathbb{U})}$, where u_{approx} is a discrete solution obtained as the last iteration of the single mesh version of ASGFE. To approximate the $L^\infty(\boldsymbol{\Gamma},\mathbb{U})$-norm appearing in the error, we use the same methods detailed above, but with $\#\Theta = \#\Pi = 5000$.

To drive parametric refinement, we employ only profits with work as defined in (2.10). As observed in Section 2.2.1, workless profits lead to a tensor product interpolant and thus less interesting results. In all examples, we consider the finite element estimator $\eta_{FE,\Lambda} = \eta_{2,\Lambda}\Lambda_{2,\Lambda}$ as defined in Section 2.3.3. The Dörfler parameter for refinement is chosen as $\theta = 0.7$ and, as default initial mesh $\mathcal{T}_{\text{init}}$, a quasi-uniform mesh with 512 triangles and 289 vertices.

In order to decrease the memory requirements of the program, the finite element refinement tolerance from Section 2.3.3 is modified as follows:

$$\text{Tol} \coloneqq \alpha\lambda^{-1}_{2,\Lambda}\zeta_{SC,\Lambda}, \tag{2.60}$$

i.e. we neglect the term depending on the margin of Λ. In the experiments below, we observe that this choice does not compromise convergence. Further investigations are needed to understand whether or not the theoretical sufficient condition for convergence can be weakened. The constant α appearing in (2.60) is chosen as $\alpha = 0.9$. A value of α close to one shifts the balance between finite element refinement and parameter enrichment towards the latter one. In order to improve the computational efficiency, we use the following shortcut in the implementation of Algorithm 8: Instead of re-computing the tolerance Tol at each iteration of the loop, we update it only at the end and, if needed, keep refining the finite element solutions. We alternate these two steps until the finite element estimator falls below the tolerance.

In the following two sections, we consider a physical domain $D = (0,1)^2$ and denote $\boldsymbol{x} = (x_1, x_2) \in D$. The parametric domain is $\boldsymbol{\Gamma} = [-1,1]^N$ for an integer $N \in \mathbb{N}$ representing the number of parametric dimensions of the problem.

We recall that for a numerical solution $\mathcal{I}_\Lambda[U]$ obtained with ASGFE, its number of degrees of freedom is proportional to $M \coloneqq \sum_{\boldsymbol{y}\in\mathcal{H}_\Lambda} \#\mathcal{T}_{\boldsymbol{y}}$, where $\#\mathcal{T}_{\boldsymbol{y}}$ is the number of vertices of the mesh corresponding to the collocation point $\boldsymbol{y}$ (or equivalently the dimension of the finite element space $\mathbb{U}_{\boldsymbol{y}}$ up to boundary conditions).

2.4.1 First example: Karhunen–Loève expansion with $N = 5, 11$ parameters

We consider a constant forcing term $f \equiv 1$ and the following diffusion coefficient with affine dependence on the parameter $\boldsymbol{y} \in \boldsymbol{\Gamma}$:

$$a(\boldsymbol{x}, \boldsymbol{y}) = a_0 + \frac{1}{3}\left(a_1(\boldsymbol{x})y_1 + \sum_{n=2}^{N} a_n(\boldsymbol{x})y_n\right), \tag{2.61}$$

where $a_0 = 1$, $a_1 = \left(\frac{\sqrt{\pi}L}{2}\right)^{1/2}$ and, for $n > 1$,

$$\lambda_n = \left(\sqrt{\pi}L\right)^{1/2} \exp\left(-\frac{\left(\left\lfloor \frac{n}{2} \right\rfloor \pi L\right)^2}{8}\right),$$

$$a_n(\boldsymbol{x}) = \begin{cases} \sqrt{\lambda_n}\sin\left(n\pi x_1\right) & \text{if } n \text{ even} \\ \sqrt{\lambda_n}\cos\left(n\pi x_1\right) & \text{if } n \text{ odd,} \end{cases}$$

where $L \in (0, 1)$ is a constant. Such a diffusion coefficient is the result of the Karhunen–Loève expansion [ST06] of the random field $a(\boldsymbol{x}, \omega)$ with mean a_0 and covariance $Cov(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{3^2}\exp\left(-\frac{(x_1 - x_1')^2}{L^2}\right)$, for $x, x' \in D$. The constant L denotes the "correlation length" of the stochastic parameter. We choose $L = 0.5$, which implies

$$a_1 \approx 6.7 \cdot 10^{-1},$$
$$\lambda_1 \approx 6.9 \cdot 10^{-1},\ \lambda_2 \approx 2.7 \cdot 10^{-1},\ \lambda_3 \approx 5.8 \cdot 10^{-2},\ \lambda_4 \approx 6.8 \cdot 10^{-3},\ \lambda_5 \approx 4.2 \cdot 10^{-4}.$$

In the rest of this section, we truncate the expansion to $N = 5$ and $N = 11$ terms. The aim is to study how the algorithm performs for different numbers of parametric dimensions N on an anisotropic problem, where the first parameters are more relevant than the last ones.

In Figure 2.3, we use the problem with $N = 5$ parameters to provide the reader with a concrete example of the steps of the algorithm. On the left, we plot the evolution of the estimators with respect to the number of degrees of freedom. We plot the values of the estimators any time they are computed (not only once per iteration). The algorithm alternates between steps of parameter enrichment and mesh refinement. The spikes in the value of the finite element estimator correspond to the parametric enrichment steps, when new collocation points are added to the sparse grid with the initial (coarse) mesh $\mathcal{T}_{\text{init}}$. When finite element refinement is carried out, the finite element estimator eventually decreases with order $M^{-1/2}$, as has to be expected for lowest order adaptive FEM (see also Theorem 2.24). On the right-hand side of Figure 2.3, we plot the estimator only once per iteration. As prescribed in (2.60), the finite element estimator is bounded from above by the parametric estimator after each finite element refinement loop.

In Figure 2.4 we compare the results for $N = 5$ and $N = 11$. On the left, the value of total estimator and reference error are plotted as a function of the number of degrees of freedom. The problem with $N = 11$ gives larger estimator and reference error. However, the difference is marginal, suggesting that the algorithm successfully detects the anisotropy of the problem. On the right, we plot the effectivity index (ratio between estimator and error). As observed in [GN18], the number of problem dimensions affects the efficiency of the estimator. In view of these facts, the algorithm may benefit from an adaptive dimension selection step as the one proposed in [GN18, Section 7].

In Figure 2.5, we consider the problem with $N = 11$ and plot projections of the final multi-index set Λ. The projections are obtained selecting pairs of parametric dimensions $n_1, n_2 \in 1, ..., N$ and plotting the 2D set $\{(\nu_{n_1}, \nu_{n_2}), \boldsymbol{\nu} \in \Lambda\}$. Observe how larger values are achieved by the first parametric dimensions, confirming that the algorithm manages to detect the anisotropy of the problem.

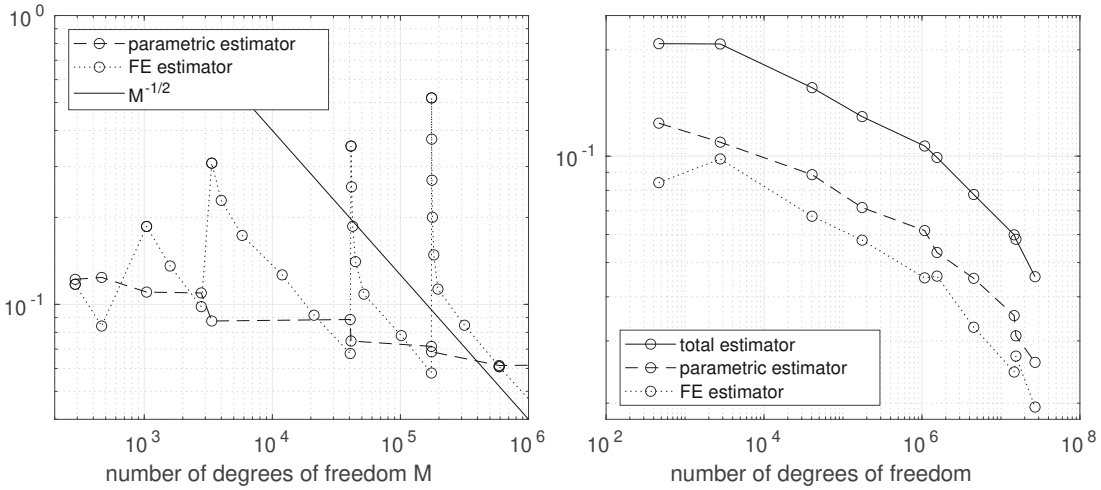

Figure 2.3: First results for ASGFE applied to the problem with Karhunen–Loève expansion $N = 5$. Left: "detailed" evolution of the estimators, i.e. reporting their values any time they are computed during the execution. Right: Total, parametric and finite element estimators at every iteration.

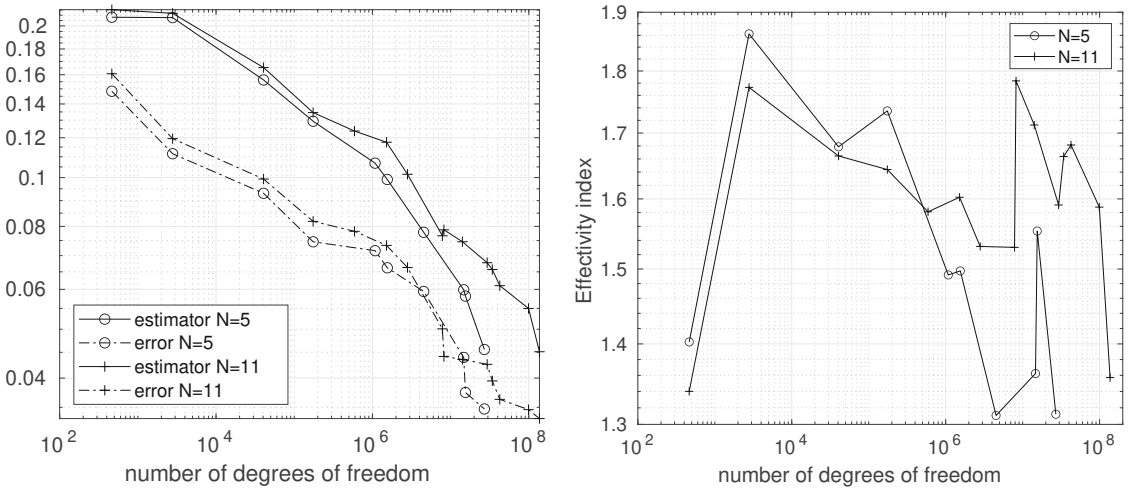

Figure 2.4: Comparing ASGFE applied to the problem with Karhunen–Loève expansion for $N = 5$ and $N = 11$. Left: Total estimator and error. Right: Effectivity index.

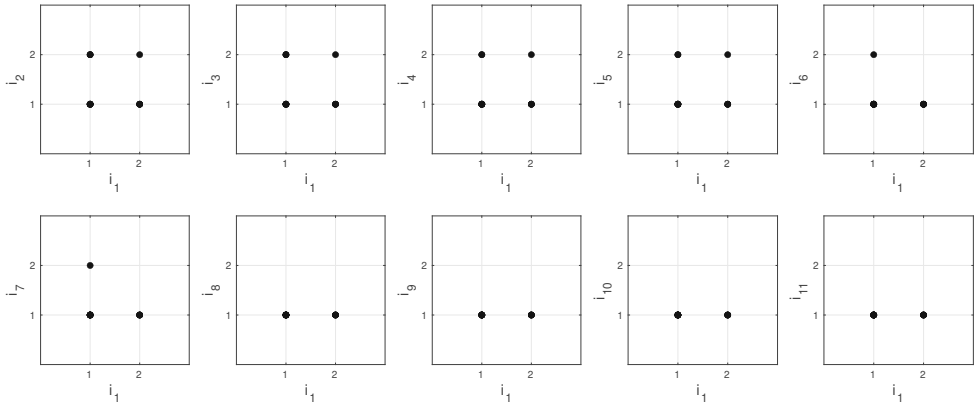

Figure 2.5: Projections of the final multi-index set from ASGFE applied to the the problem with Karhunen–Loève expansion for $N = 11$.

2.4.2 Second example: Inclusion problem with $N = 8$ parameters

We consider an inclusion problem with $N = 8$ parameters similar to that in [GN18]. Within D, we identify nine disjoint subdomains F and $\{C_n\}_{n=1}^8$ depicted in Figure 2.6. The diffusion coefficient reads

$$a(\boldsymbol{x}, \boldsymbol{y}) = a_0 + \sum_{n=1}^{8} \gamma_n \chi_n y_n \qquad \text{with } a_0 = 1.1, \tag{2.62}$$

where $(\gamma_n)_{n=1}^8 = (1, 0.8, 0.4, 0.2, 0.1, 0.05, 0.02, 0.01)$ are constants used to introduce anisotropy in the problem and χ_n is the characteristic function of C_n, for all $n \in 1, ..., 8$. The forcing term reads $f(\boldsymbol{x}) \coloneqq 100\chi_F(\boldsymbol{x})$, where χ_F is the characteristic function of F.

In order to highlight the importance of adaptive finite element refinement in space, we present a comparison between the single mesh version of ASGFE from Section 2.3.4, where the unique mesh is adaptively refined with Dörfler marking, and an analogous version where only uniform refinement on the whole mesh is allowed. In Figure 2.7 (top left) we report for both algorithms the value of the estimator and reference error. The adaptive version clearly outperforms the one with uniform refinement. In Figure 2.6 (right) we show a density plot of a mesh produced by the single mesh adaptive algorithm with $\approx 2 \cdot 10^7$ degrees of freedom. We see that mesh refinement occurs along the boundary of the inclusions and is more pronounced for the inclusions corresponding to larger anisotropy parameter γ_n, confirming that the algorithm detects the parametric structure of the problem.

In Figure 2.7 (top right) we study the fully adaptive ASGFE algorithm and observe that this is also markedly slower than the single mesh algorithm. Additional insight is given in the plot on the bottom left of Figure 2.7. Here we show only the value of the parametric estimator with respect to the number of collocation point for both the fully adaptive and single mesh versions of ASGFE. This implies that the fully adaptive ASGFE algorithm seems to overrefine the finite-element meshes. We suspect that this is due to the fact that in the derivation of $\eta_{\mathrm{FE},\Lambda}$ in Section 2.1.3, one is required to use the triangle inequality and thus sacrifices local information of the sparse grid interpolant. This is not necessary in the single-mesh estimator from Section 2.3.4.

Finally, we ran Algorithm 7 with fixed tolerance Tol (not depending on the parametric estimator) and plot the results in Figure 2.7 (bottom right). We observe that the results are

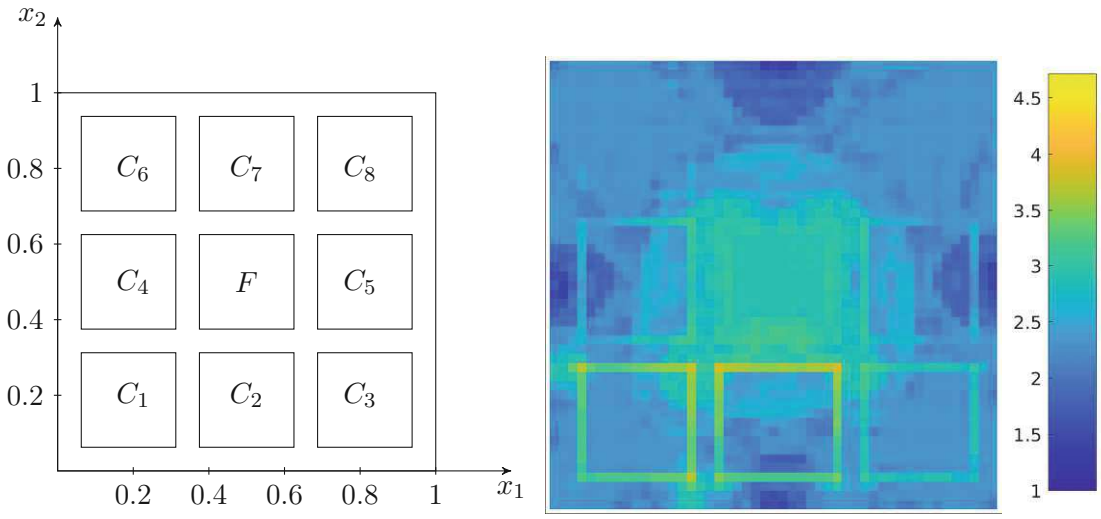

Figure 2.6: Left: Domain for the inclusion problem. Right: Logarithmic density plot of a mesh of ASGFE with single adaptive mesh. The colors refer to the number of mesh elements within one pixel of the plot.

very much comparable to the standard fully adaptive ASGFE algorithm, except for significant over refinement in the early stages of the computation (see the flat line of the finite element error estimator satisfying the tolerance). In terms of computational effort, the algorithms are nearly identical, as the same spatial refinements are performed, only at different stages of the algorithm.

We also tested the algorithms with respect to the $L^2(\mathbf{\Gamma})$-norm instead of the $L^\infty(\mathbf{\Gamma})$-norm. The necessary changes in the estimators are straightforward, essentially we replace the search for the maximum by a Monte Carlo quadrature. The theoretical results of this manuscript all hold verbatim for the $L^2(\mathbf{\Gamma})$-norm. Figure 2.8 shows the results. Again the single mesh algorithm outperforms the fully adaptive algorithm.

Distribution of computational cost

In Figure 2.8 (right-hand side) we compare the total number of degrees of freedom and collocation points achieved by the three methods, i.e. the single adaptive mesh algorithm from Section 2.3.4, adaptivity in the parameter space but uniform refinement in the spatial domain, and the fully adaptive ASGFE algorithm (Algorithm 7). The adaptive strategy with a single adaptive mesh performs parametric refinement more often than the other two, leading to a higher number of collocation points and lower average number of degrees of freedom per collocation point. In Figure 2.9 (compare also Figure 2.6) we provide logarithmic density plots of the meshes produced by the multiple adaptive mesh algorithm (with $\approx 2 \cdot 10^7$ degrees of freedom). We observe that the mesh corresponding to a collocation point is locally refined along the edges of the corresponding inclusion. Furthermore, the intensity of the refinement around a certain inclusion is related to the constant γ_n of the diffusion coefficient, confirming that the numerical methods detects the anisotropy of the problem.

2.5 Conclusion

We analyses the adaptive sparse grid interpolations algorithm from [GN18] and prove convergence of several different versions of the algorithm:

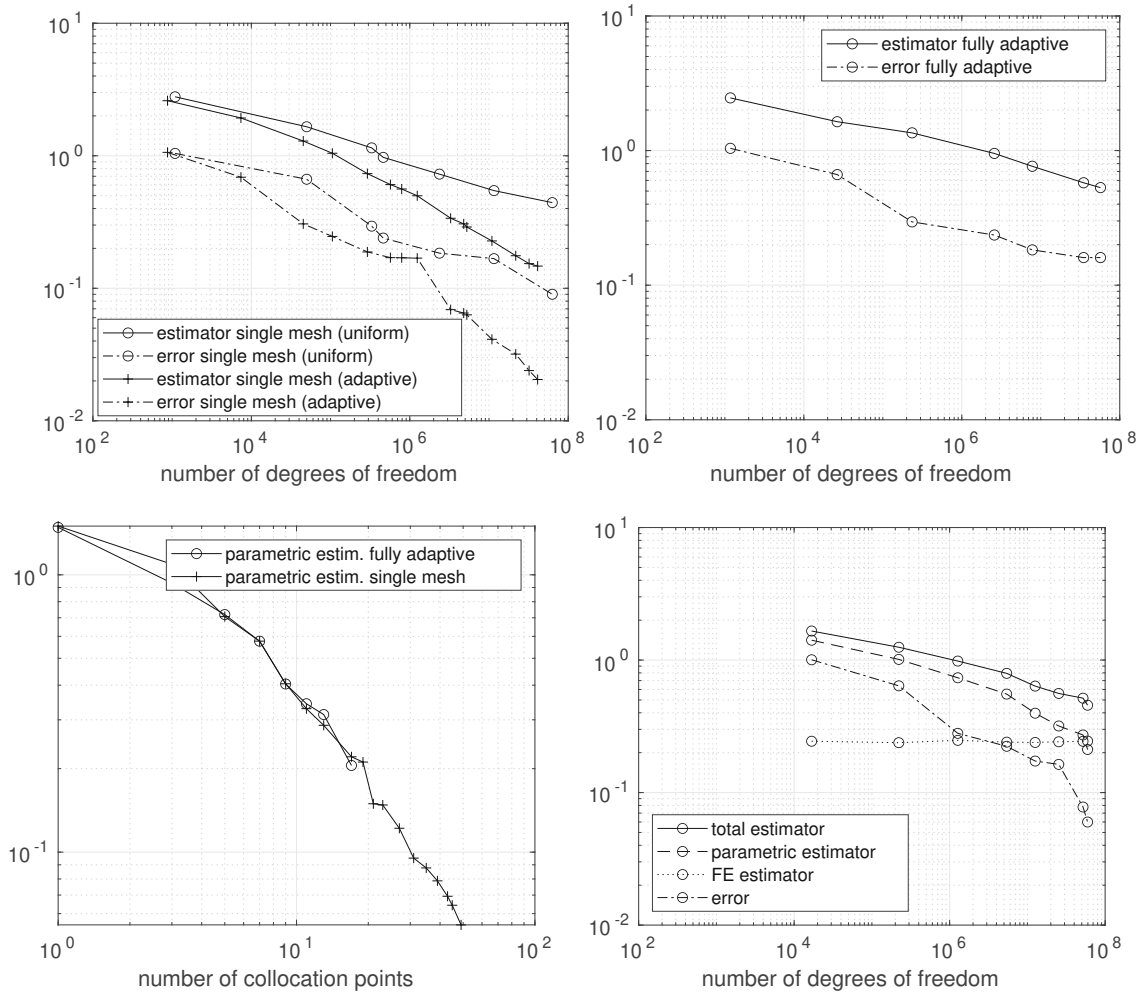

Figure 2.7: Results for ASGFE on the 8D inclusion problem. Top left: Comparison between adaptive and uniform space refinement with the single mesh algorithm from Section 2.3.4. Top right: Total estimator and error for the fully adaptive algorithm. Bottom left: Parametric estimator as a function of the number of collocation points, for both the fully adaptive and the single mesh ASGFE. Bottom right: the fully adaptive ASGFE with fixed value for finite element tolerance Tol.

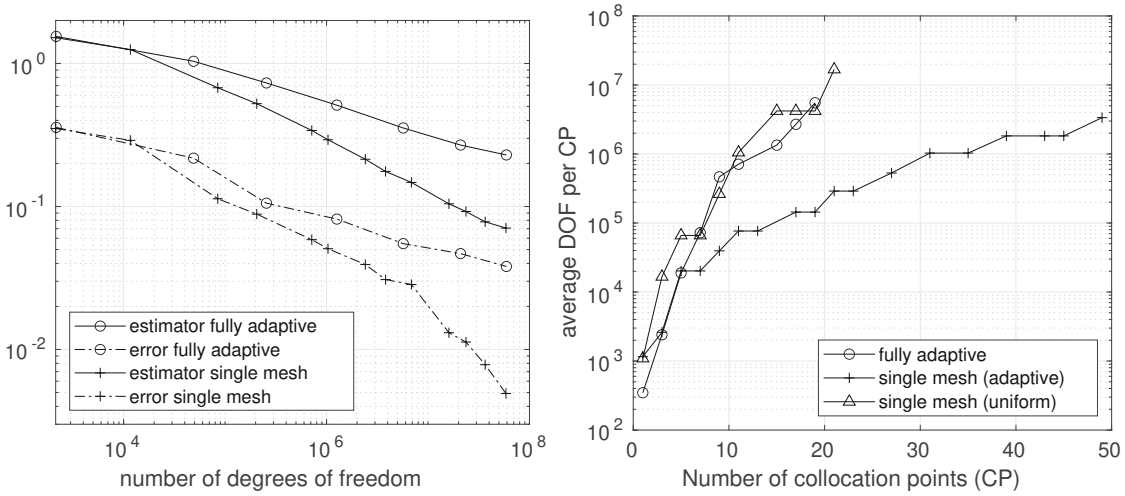

Figure 2.8: Left: Numerical results on the 8D inclusion problem in the $L^2(\mathbf{\Gamma})$-norm. Right: Total estimator and error for ASGFE (fully adaptive). Right: Average number of degrees of freedom (DOF) per collocation point (CP) plotted versus the number of collocation points. Each line corresponds to one of the three proposed algorithms. Each marker corresponds to one step of the adaptively refined discrete solution.

- Convergence of the parametric enrichment algorithm without finite element refinement (Section 2.2);
- Convergence of the fully adaptive algorithm (Algorithm 7) even with optimal convergence of the finite element loop (Theorem 2.24);
- Convergence of a single-mesh variant of Algorithm 7 (Section 2.3.4) proposed in [GN18].

The numerical examples clearly show the superiority of spatial adaptive refinement combined with parametric enrichment over pure parametric enrichment algorithms. While the theoretical results are strongest for the fully adaptive algorithm (linear convergence in Proposition 2.22 for Algorithm 7) the single mesh algorithm from Section 2.3.4 seems to be more efficient. This is underlined by the numerical experiments in the previous section, which clearly show an advantage of the single mesh version over the fully adaptive version. Based on the theoretical results from Theorem 2.24 and the experiments, we come to the conclusion that the finite element error estimator of Algorithm 7 severely over-estimates the total error and hence leads to over-refinement of the finite element meshes. This does not seem to happen for the single-mesh error estimator. We suspect that the application of the triangle inequality in the derivation in Section 2.3.3 is mainly responsible for this over-estimation and further research is required to see whether this can be avoided.

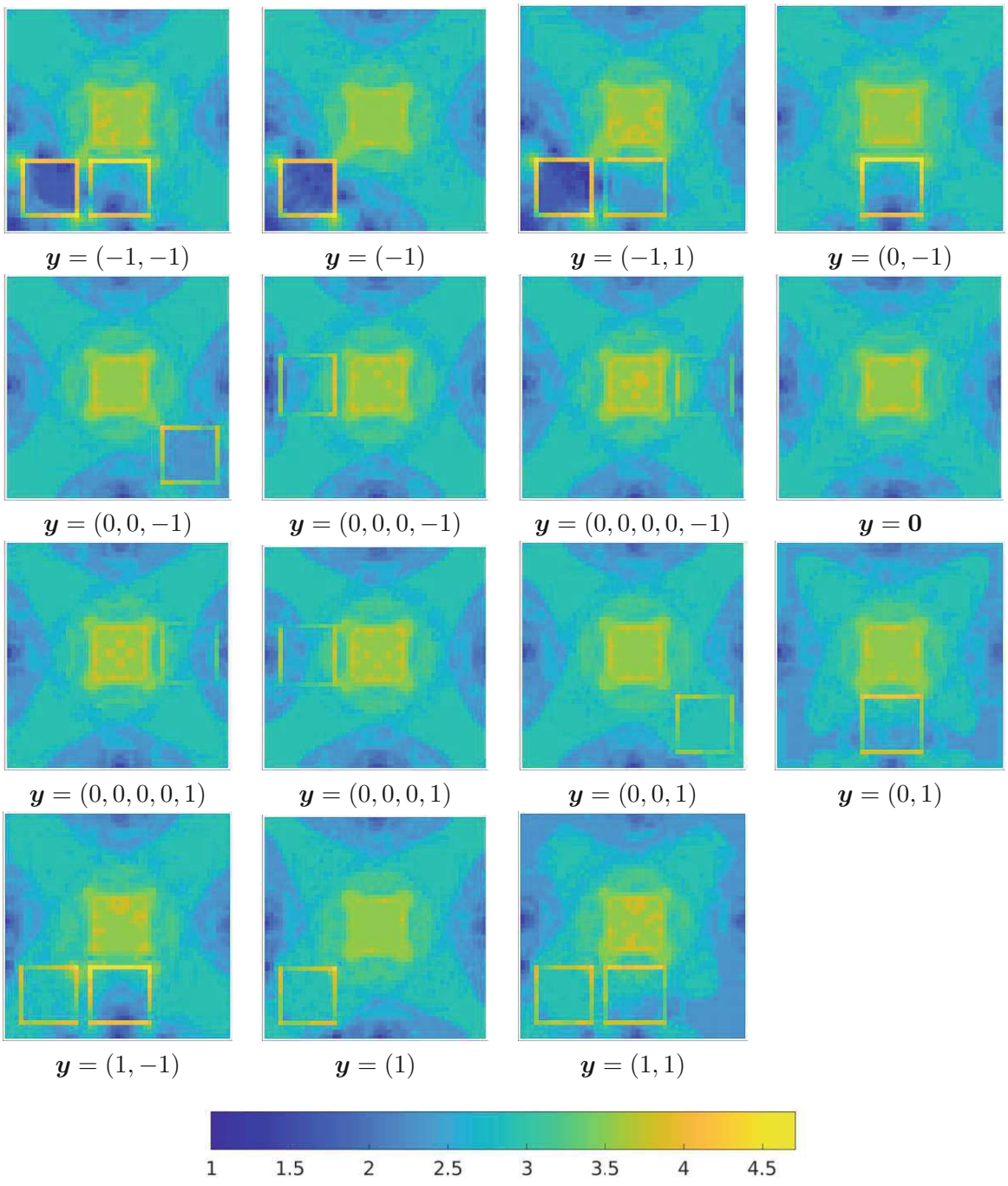

Figure 2.9: Density plot of the meshes produced by the fully adaptive ASGFE algorithm. The corresponding collocation point is indicated below, ignoring the trailing components equal to zero. The color-bar at the bottom indicates the base-10 logarithm of the density of elements.

Chapter 3

Sparse grid approximation of nonlinear SPDEs: The SLLG equation

In this chapter, we present a methodology to approximate the solution to (possibly) nonlinear stochastic PDEs driven by Gaussian noise. To that end, we employ the Doss-Sussmann transform (Section 1.1.5) and discretion the resulting Wiener process with the Lévy-Ciesielski expansion (Section 1.1.2). This leads to a parameterized nonlinear time-dependent PDE with infinite dimensional and unbounded parameter space, which can be approximated by using sparse grid techniques (see Sections 1.2.3 for an 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 (SLLG) as it contains many of the difficulties one encounters in nonlinear and stochastic partial differential equations. The present work gives a first efficient approximation of the probability distribution of the solution of the SLLG equation.

This chapter is structured as follows: In Section 3.1 we introduce a general framework for the study of the parametric regularity of solutions of SPDEs. We first explain in Section 3.1.1 how to reduce a SPDE to a parametric coefficients PDE. Then, in Sections 3.1.2 and 3.1.3 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. In Section 3.2, we consider the SLLG equation (already discussed in Section 1.3) as a concrete example of nonlinear SPDE with noise given by a function of the Wiener process. We recall how it can be reduced to a random coefficient PDE in Section 3.2.1 and further to a parametric coefficient PDE in Section 3.2.2. In Section 3.2.3, we prove that, under regularity assumption on the problem data, the solutions' sample paths are Hölder-continuous and uniformly bounded with respect to the Wiener process sample paths. In Section 3.3, we apply the regularity analysis from Sections 3.1.2 and 3.1.3 to the parametric 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 3.4, we do the same for a simplified version of the parametric LLG problem obtained with additional modeling 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 3.5. The results are confirmed by numerical experiments. The final Section 3.6 derives the multilevel version of the stochastic collocation method and provides numerical tests.

This chapter is verbatim to [ADF+24] up to stylistic changes and the following changes of notation:

- Sigma-algebra in the probability triple $\mathcal{F}$ to $\mathcal{E}$;
- Real parameter space $\mathcal{X}_{\mathbb{R}}$ to $\mathbf{\Gamma}$;
- Complex parameter space $\mathcal{X}$ to $\mathbf{\Sigma}$;
- SPDE drift and noise integrands μ, σ to $\mathfrak{D}$, $\mathfrak{N}$;
- Banach space of residual functions R to $\mathbb{F}$;
- Imaginary parts Im to $\Im$;
- Space of real solution functions $\mathbb{U}_{\mathbb{R}}$ to $\mathcal{U}(\mathbf{\Gamma})$;
- Space of real coefficient functions $\mathbb{W}_{\mathbb{R}}$ to $W(\mathbf{\Gamma})$;
- Constant introduced in Assumption 1 C_r to $C_{\mathbf{\Gamma}}$;
- "LLG problem" and "SLLG problem" substituted respectively with (problem related to the) "LLG equation" and (problem related to the)"SLLG problem".

3.1 General derivation of parametric regularity of SPDEs

In the present section, we outline a 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 problem formulation and arguments presented in this section are formal and need respectively to be rigorously defined and verified for each concrete problem. The most important assumptions are listed explicitly below.

3.1.1 Reduction to a parametric problem

Let us recall the Doss-Sussmann transform, which allows to reduced a stochastic PDE (SPDE) and random coefficient PDE (see also Section (1.1.5)). We then also recall the Lévy-Ciesielski expansion (see also Section 1.1.2) to obtain (from the random coefficient PDE) a parametric coefficient PDE.

Consider a bounded, connected, Lipschitz domain $D \subset \mathbb{R}^d$ of dimension $d \in \mathbb{N}$ and a final time $T > 0$. Recall the normal derivative ∂_n and the space-time cylinder $D_T := [0,T] \times D$. Given the initial condition $U^0 : D \to \mathbb{R}^m$ for $m \in \mathbb{N}$, a drift coefficient $\mathfrak{D} : \mathbb{R}^m \times D_T \to \mathbb{R}^m$, a noise coefficient $\mathfrak{N} : \mathbb{R}^m \times D \to \mathbb{R}^m$, and the probability triple $(\Omega, \mathcal{E}, \mathcal{P})$, we consider the SPDE problem:
Find a random field $U : \Omega \times D_T \to \mathbb{R}^m$ such that, $\mathcal{P}$-a.s.

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

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

Consider the *Doss-Sussmann transform* (see Section 1.1.5) of U, i.e. the random field $u : \Omega \times D_T \to \mathbb{R}^m$, $u = e^{-W\mathfrak{N}}U$. It can be proved to be solution to a *random coefficient partial differential equation* (PDE)

$$\mathcal{R}(W(\omega), u(\omega)) = 0 \qquad \text{in } \mathbb{F},\ \mathcal{P}\text{-a.e. } \omega \in \Omega. \tag{3.1}$$

The *residual operator* $\mathcal{R} : \mathbb{W} \times \mathbb{U} \to \mathbb{F}$ is defined for Banach spaces $\mathbb{W}, \mathbb{U}$ and $\mathbb{F}$, representing respectively coefficients, solutions, and residuals sample paths. Observe that this is an instance of the framework described in Section 1.1.1, but we denote coefficients with $\mathbb{W}$ instead of $\mathbb{A}$ because we work only with the Wiener process. In general, the residual map $\mathcal{R}$ is a differential operator in time and space with respect to $u \in \mathbb{U}$ while it does not contain Itô or Stratonovich differentials of W.

With the aim of making the distribution of u amenable to *approximation*, we parameterize 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. Recall that the Lévy-Ciesielski expansion (LCE) (see Section 1.1.2) 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 the coefficients $Y_{\ell,j}$ are independent standard normal random variables and $\{\eta_{\ell,j} : \ell \in \mathbb{N}_0, j = 1, \dots, \lceil 2^{\ell-1} \rceil\}$ denotes the *Faber-Schauder* hat-function basis.

We consider a parametric version of the random field W in the form W as $W : \mathbb{R}^{\mathbb{N}} \times [0,1] \to \mathbb{R}$ so that

$$W(\boldsymbol{y}, t) = \sum_{\ell=0}^{\infty} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} y_{\ell,j} \eta_{\ell,j}(t), \tag{3.2}$$

where $y_{\ell,j} \in \mathbb{R}$ for all $\ell \in \mathbb{N}_0, j = 1, \dots, \lceil 2^{\ell-1} \rceil$. 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 $W_N(\boldsymbol{y}, t) = \sum_{n=0}^{N} y_n \eta_n(t)$. The two indexing systems, hierarchical and linear, are related via (1.11).

The fact that the parameter domain is unbounded requires the use of appropriate collocation nodes, a topic we treat in Section 3.5, below.

We denote by $\boldsymbol{\Gamma}$ an appropriate separable Banach space of real sequences such that if $\boldsymbol{y} \in \boldsymbol{\Gamma}$, then $W(\boldsymbol{y}, \cdot) \in \mathbb{W}$, the desired Banach space of coefficient sample paths. The Banach space $\boldsymbol{\Gamma}$ is assumed to be separable in order for $u : \boldsymbol{\Gamma} \to \mathbb{U}$ to be separably valued (i.e. its image $u(\boldsymbol{\Gamma})$ be separable) under the mild regularity assumption that u is continuous. As a consequence of Pettis measurably theorem, the parameter-to-solution map is also measurable. Measurability is a naturally important property because it is necessary for the well posedness of integral quantities such as the moments of the random field.

Example 3.1. *Consider the Banach space of sequences:*

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

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

Assume without loss of generality that $T = 1$. As explained in Section 1.1.3, by substituting the random field $W(\omega, t)$ in the random coefficient PDE (3.1) with the parametric expansion (3.2), we obtain a *parametric coefficient PDE*:
Find $u : \boldsymbol{\Gamma} \to \mathbb{U}$ such that

$$\mathcal{R}(W(\boldsymbol{y}), u(\boldsymbol{y})) = 0 \qquad \text{in } \mathbb{F}, \ \boldsymbol{\mu}\text{-a.e. } \boldsymbol{y} \in \boldsymbol{\Gamma}, \tag{3.3}$$

where $\boldsymbol{\mu}$ denotes the standard Gaussian measure on $\mathbb{R}^{\mathbb{N}}$, i.e. the product measure $\boldsymbol{\mu} := \bigotimes_{n \in \mathbb{N}} \mu_n$, where $(\mu_n)_{n \in \mathbb{N}}$ is a sequence of standard Gaussian probability measures on $\mathbb{R}$ (see, e.g., [Kak48] and [DNSZ23b, Section 2.4] for details on infinite product measures).

3.1.2 Holomorphic regularity of the solution operator

While holomorphic parameter regularity of random elliptic equations is well-known by now, the literature is much sparser for nonlinear and time-dependent problems. In this section, we follow an approach from [CCS15] which uses the implicit function theorem to obtain analyticity. While the authors in [CCS15] can rely on a compact parameter domain to ensure a non-trivial domain of extension and a uniformly bounded 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 [HSS23] could also be used to achieve similar results in a less explicit fashion.

We require some assumptions to work in a general setting.

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

Assumption 2. *The residual operator $\mathcal{R} : \mathbb{W} \times \mathbb{U} \to \mathbb{F}$ admits an extension to* complex *Banach spaces, which we again denote $\mathbb{W}$, $\mathbb{U}$, and $\mathbb{F}$ such that the extended residual (again denoted by $\mathcal{R}$) satisfies the following properties:*

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

(ii) $\partial_u \mathcal{R}(W(\boldsymbol{y}), u(\boldsymbol{y})) : \mathbb{U} \to \mathbb{F}$ is a homeomorphism for all $\boldsymbol{y} \in \boldsymbol{\Gamma}$.

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

$$\begin{aligned} B_\varrho(W_0) &:= \big\{W \in \mathbb{W} : \|W - W_0\|_{\mathbb{W}} < \varrho\big\}, \\ B_\varrho(u_0) &:= \big\{u \in \mathbb{U} : \|u - u_0\|_{\mathbb{U}} < \varrho\big\}. \end{aligned} \tag{3.4}$$

Let us recall the implicit function theorem for maps between Banach spaces (see, e.g., [Die69, Theorem 10.2.1]).

Theorem 3.2 (Implicit function)**.** *Let E, F, G be Banach spaces, $A \subset E \times F$ and $f : A \to G$ be a Fréchet 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*

$$\mathcal{U}'(x) = -\left(D_2 f(x, \mathcal{U}(x))\right)^{-1} \circ \left(D_1 f(x, \mathcal{U}(x))\right) \quad \forall x \in U. \tag{3.5}$$

Invoking Theorem 3.2 for the operator $\mathcal{R} : \mathbb{W} \times \mathbb{U} \to \mathbb{F}$, with $\boldsymbol{y} \in \boldsymbol{\Gamma}$ and $u(\boldsymbol{y}) \in \mathbb{U}$ 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 $W \in B_{\varepsilon(\boldsymbol{y})}(W(\boldsymbol{y}))$ (recall (3.4)).

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 (3.4), 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 $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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*

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

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 (3.5) from Theorem 3.2, this assumption implies the existence of a continuous increasing function $\mathcal{G} = \mathcal{G}(\|\mathcal{U}(W)\|_{\mathbb{U}}) > 0$ such that

$$\left\|\mathcal{U}'(W)\right\|_{\mathcal{L}(\mathbb{W},\mathbb{U})} \leq \mathcal{G}(\|\mathcal{U}(W)\|_{\mathbb{U}}) \qquad \forall W \in B_{\min(\varepsilon(\boldsymbol{y}),\varepsilon_W)}(W(\boldsymbol{y})). \tag{3.6}$$

3.1.3 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 $\boldsymbol{y} \in \boldsymbol{\Gamma}$. We can assume, without loss of generality, that $\varepsilon(\boldsymbol{y}) \leq \varepsilon_W$ introduced in Assumtpion 3.

Definition 3.3. *Fixed $\boldsymbol{y} \in \boldsymbol{\Gamma}$, 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 $\mathcal{H}(\boldsymbol{y})$,*
- *$\mathcal{H}(\boldsymbol{y})$ is star-shaped around $W(\boldsymbol{y})$, i.e. for all $W \in \mathcal{H}(\boldsymbol{y})$, $0 \leq \sigma \leq 1$, we have $\sigma W + (1 - \sigma)W(\boldsymbol{y}) \in \mathcal{H}(\boldsymbol{y})$.*

Recall that $\varepsilon(\boldsymbol{y}) > 0$ is the radius of the holomorphic extension given by the implicit function theorem (Theorem 3.2) and $\varepsilon_u > 0$ is defined in Assumption 3.

In contrast to [CCS15], the domain of real parameters $W(\boldsymbol{\Gamma})$ may not be compact for relevant choices of $\boldsymbol{\Gamma}$. Therefore, $\varepsilon(\boldsymbol{y})$ can be arbitrarily small and in turn $\mathcal{H}(\boldsymbol{y})$ might become very small for certain parameters $\boldsymbol{y}$. The goal of the next arguments is to show that there exists $\varepsilon > 0$ such that for all $\boldsymbol{y} \in \boldsymbol{\Gamma}$ $\mathcal{H}(\boldsymbol{y}) = B_\varepsilon(W(\boldsymbol{y}))$ is a valid choice. Instead of relying on compactness, we exploit estimate (3.6) through the following nonlinear generalization of Gronwall's lemma:

Lemma 3.4 ([Dra03], Theorem 27)**.** *Let $0 \leq c \leq d < \infty$, $\varphi : [c, d] \to \mathbb{R}$ 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*

$$\varphi(t) \leq a + b \int_c^t k(s)\mathcal{G}(\varphi(s))\,ds \quad \forall t \in [c, d],$$

then

$$\varphi(t) \leq G^{-1}\left(G(a) + b \int_c^t k(s)\,ds\right) \quad \forall c \leq t \leq d_1 \leq d$$

where G is defined, for some fixed $\xi > 0$, as

$$G(\lambda) \coloneqq \int_\xi^\lambda \frac{ds}{\mathcal{G}(s)} \qquad for all \lambda > \xi \tag{3.7}$$

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

Theorem 3.5. *Assume the validity of Assumptions 1, 2, and 3. With $C_{\mathbf{\Gamma}} > 0$ given in Assumption 1, choose $0 < \varepsilon < \varepsilon_W$ such that $G(C_{\mathbf{\Gamma}}) + \varepsilon$ belongs to the domain of G^{-1} (where G is defined in (3.7) with the corresponding $\mathcal{G}$ given in (3.6)). Then, ε is independent of $\boldsymbol{y}$ and $\mathcal{H}(\boldsymbol{y})$ from Definition 3.3 can be chosen as $\mathcal{H}(\boldsymbol{y}) = B_\varepsilon(W(\boldsymbol{y}))$ for all $\boldsymbol{y} \in \mathbf{\Gamma}$. 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 $\boldsymbol{y} \in \mathbf{\Gamma}$.

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

$$\varphi(t) - \varphi(s) = \int_s^t \varphi'(\sigma)\mathrm{d}\sigma = \int_s^t \mathcal{U}'(W_\sigma)[W - W(\boldsymbol{y})]\mathrm{d}\sigma \quad \text{for all } s,t \in [0,1]. \tag{3.8}$$

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

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

Assumption 1 and estimate (3.6) (consequence of Assumption 3) imply the estimate

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

Apply Lemma 3.4 to conclude (note that, in the notation of Lemma 3.4, we have $d_1 = d = 1$ because of the definition of ε as well as $k(s) = 1$)

$$\|\varphi(t)\|_{\mathbb{U}} \le G^{-1}(G(C_{\mathbf{\Gamma}}) + \varepsilon t) \le G^{-1}(G(C_{\mathbf{\Gamma}}) + \varepsilon) \quad \text{for all } 0 \le t \le 1. \tag{3.9}$$

Since $\|\mathcal{U}(W)\|_{\mathbb{U}} = \|\varphi(1)\|_{\mathbb{U}} \le C_\varepsilon$, where $C_\varepsilon := G^{-1}(G(C_{\mathbf{\Gamma}}) + \varepsilon)$, 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 (3.8) implies, for $0 \le s < t \le 1$,

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

The desired results then follows from (3.9).

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 $B_\varepsilon(W(\boldsymbol{y}))$. 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 3.3) 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 3.5 provides all the tools to estimate parametric regularity through Cauchy's integral theorem. The Lévy-Ciesielski expansion (3.2) can be (formally) extended to the complex parameters $\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}}$. Thus, in view of Theorem 3.5, $\boldsymbol{z} \mapsto \mathcal{U}(W(\boldsymbol{z}))$ 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 3.5 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 *complex* polydisk

$$\boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y}) \coloneqq \{\boldsymbol{z} \in \boldsymbol{\Sigma} : |z_n - y_n| < \rho_n \ \forall n \in \mathbb{N}\}, \tag{3.10}$$

where $\boldsymbol{\Gamma} \subset \boldsymbol{\Sigma} \subset \mathbb{C}^{\mathbb{N}}$ is the set of complex parameters.

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

$$\boldsymbol{z} \in \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y}) \Rightarrow W(\boldsymbol{z}) \in B_\varepsilon(W(\boldsymbol{y})).$$

In conclusion, for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $\mathcal{U} \circ W : \boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y}) \to \mathbb{U}$ is holomorphic because it is a composition of holomorphic functions. Moreover, $\mathcal{U} \circ W$ is uniformly bounded by C_ε (see Theorem 3.5) independently of $\boldsymbol{y}$.

Consider a multi-index $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n) \in \mathbb{N}_0^n$ and denote by $\partial^{\boldsymbol{\nu}}$ the mixed derivative $\partial_1^{\nu_1} \dots \partial_n^{\nu_n}$ 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 3.6. *Consider $u : \boldsymbol{\Gamma} \to \mathbb{U}$, the parameter-to-solution map that solves the parametric coefficient PDE (3.3). Let Assumptions 1, 2, 3 hold and fix $\varepsilon > 0$ as in Theorem 3.5. 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*

$$\|\partial^{\boldsymbol{\nu}} u(\boldsymbol{y})\|_{\mathbb{U}} \leq \prod_{j=1}^{n} \nu_j! \rho_j^{-\nu_j} C_\varepsilon \qquad \forall \boldsymbol{y} \in \boldsymbol{\Gamma}, \tag{3.11}$$

where $C_\varepsilon > 0$ from Theorem 3.5 is independent of $\boldsymbol{\nu}$ or $\boldsymbol{y}$. The same bound holds for the norm $\|\partial^{\boldsymbol{\nu}} u\|_{L^2_{\boldsymbol{\mu}}(\boldsymbol{\Gamma};\mathbb{U})}$, where $\boldsymbol{\mu}$ denotes a probability measure on $\boldsymbol{\Gamma}$.

Proof. Apply Cauchy's formula [Her89, Theorem 2.1.2] to each of the n variables $y_1, \dots, y_n$ recursively and then differentiate. □

Note that Theorem 3.6 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.

3.2 Stochastic, random and parametric Landau–Lifshitz–Gilbert equation

We begin by recalling the stochastic Landau–Lifshitz–Gilbert (SLLG) equation, already discussed in Section 1.3. We then also recall some facts from Section 1.3.6, in which the problem is

reduced to solving a random coefficient PDE by means of the Doss-Sussmann transform. Finally, we apply the Lévy-Ciesielski expansion of the Wiener process (already discussed in Section 1.1.2) to obtain yet another equivalent problem, this time a parametric PDE. This two steps are an instance of the general theory outlined in Section 3.1.1.

Consider a bounded connected Lipschitz domain $D \subset \mathbb{R}^3$ representing a ferromagnetic body in the time interval $[0,T]$. Recall that by $D_T \coloneqq [0,T] \times D$ we denote the space-time cylinder and by ∂_n the outward pointing normal derivative on ∂D. Consider $\boldsymbol{M}_0 : D \to \mathbb{S}^2 \coloneqq \left\{\boldsymbol{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\right\}$ (the magnetization of the magnetic body at initial time), $\lambda > 0$ (called the *Gilbert damping parameter*) and $\lambda_1 = \frac{1}{1+\lambda^2}$, $\lambda_2 = \frac{\lambda}{1+\lambda^2}$.

The effect of heat fluctuations on the systems is described with a random model. Denote by $(\Omega, \mathcal{E}, \mathcal{P})$ a probability triple and let $\mathrm{d}\boldsymbol{W} : \Omega \times D_T \to \mathbb{R}^3$ be a suitable space-time noise. For simplicity, we assume a 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 problem associated to the SLLG equation reads: Find $\boldsymbol{M} : \Omega \times D_T \to \mathbb{S}^2$ such that

$$\begin{cases} \mathrm{d}\boldsymbol{M} &= (\lambda_1 \boldsymbol{M} \times \Delta \boldsymbol{M} - \lambda_2 \boldsymbol{M} \times (\boldsymbol{M} \times \Delta \boldsymbol{M}))\,\mathrm{d}t + (\lambda_1 \boldsymbol{M} \times \boldsymbol{g}) \circ \mathrm{d}W \quad \text{in } D_T,\ \mathcal{P}\text{-a.s.} \\ \partial_n \boldsymbol{M} &= \boldsymbol{0} \qquad \text{on } \partial D \times [0,T], \\ \boldsymbol{M}(0) &= \boldsymbol{M}_0 \qquad \text{on } D. \end{cases}$$

By $\circ \mathrm{d}W$ we denote the Stratonovich differential. Recall that solution has constant magnitude in space and time. A meaningful notion of weak solution is given in Definition 1.25.

3.2.1 Random LLG equation by Doss-Sussmann transform

Let us recall some facts from Section 1.3.6, in which we explained how the SLLG equation can be reduced to a random coefficient PDE through the Doss-Sussmann transform.

While in [GLT16] the random coefficient problem is considered 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, we recall the definition of $\hat{\mathcal{C}}(s, \boldsymbol{v})$ (1.75):

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

We can equivalently express it as a sum

$$\hat{\mathcal{C}}(s, \boldsymbol{v}) = \sum_{i=1}^{6} b_i(s) F_i(\boldsymbol{v}), \tag{3.12}$$

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 $\boldsymbol{g}$.

The *Doss-Sussmann transform* (1.19) of the SLLG solution $\boldsymbol{M}$ reads $\boldsymbol{m} = e^{-WG}\boldsymbol{M}$ and solves the following *random coefficients LLG equation*: Given $\boldsymbol{M}^0 : D \to \mathbb{S}^2$, find $\boldsymbol{m} : \Omega \times D_T \to$

$\mathbb{S}^2$ such that for $\mathcal{P}$-a.e. $\omega \in \Omega$

$$\begin{cases} \partial_t \boldsymbol{m}(\omega) & = \lambda_1 \boldsymbol{m}(\omega) \times \Big(\Delta \boldsymbol{m}(\omega) + \hat{\mathcal{C}}(W(\omega), \boldsymbol{m}(\omega))\Big) \\ & \quad -\lambda_2 \boldsymbol{m}(\omega) \times \Big(\boldsymbol{m}(\omega) \times \Big(\Delta \boldsymbol{m}(\omega) + \hat{\mathcal{C}}(W(\omega), \boldsymbol{m}(\omega))\Big)\Big) \qquad \text{in } D_T, \\ \partial_n \boldsymbol{m}(\omega) & = 0 \qquad \text{on } [0,T] \times \partial D, \\ \boldsymbol{m}(\omega, 0, \cdot) & = \boldsymbol{M}_0 \qquad \text{on } D. \end{cases} \tag{3.13}$$

It is shown in [GLT16, Lemma 4.6] that any weak solution $\boldsymbol{m}$ of (3.13) corresponds to a weak martingale solution $\boldsymbol{M} = e^{WG}\boldsymbol{m}$ of (1.68) through the *inverse Doss-Sussmann transform*. Existence of solutions to (3.13) is shown in [GLT16] but uniqueness is open.

3.2.2 Parametric LLG equation by Lévy-Ciesielski expansion

Following Section 3.1.1, we derive a parametric coefficient PDE problem using the Lévy-Ciesielski expansion of the Wiener process (see Section 1.1.2). The problem related to the *parametric LLG equation* reads: Given $\boldsymbol{M}_0 : D \to \mathbb{S}^2$, find $\boldsymbol{m} : \boldsymbol{\Gamma} \times D_T \to \mathbb{S}^2$ such that for a.e. $\boldsymbol{y} \in \boldsymbol{\Gamma}$

$$\begin{cases} \partial_t \boldsymbol{m}(\boldsymbol{y}) & = \boldsymbol{m}(\boldsymbol{y}) \times \Big(\Delta \boldsymbol{m}(\boldsymbol{y}) + \hat{\mathcal{C}}(W(\boldsymbol{y}), \boldsymbol{m}(\boldsymbol{y}))\Big) \\ & \quad -\boldsymbol{m}(\boldsymbol{y}) \times \Big(\boldsymbol{m}(\boldsymbol{y}) \times \Big(\Delta \boldsymbol{m}(\boldsymbol{y}) + \hat{\mathcal{C}}(W(\boldsymbol{y}), \boldsymbol{m}(\boldsymbol{y}))\Big)\Big) \qquad \text{in } D_T, \\ \partial_n \boldsymbol{m}(\boldsymbol{y}) & = 0 \qquad \text{on } [0,T] \times \partial D, \\ \boldsymbol{m}(\boldsymbol{y}, 0, \cdot) & = \boldsymbol{M}_0 \qquad \text{on } D, \end{cases} \tag{3.14}$$

where we set $\lambda_1 = \lambda_2 = 1$ for simplicity. The precise definition of the Banach space of sequences $\boldsymbol{\Gamma}$ will be given below in (3.30).

Applying the triple cross-product formula (B.2) on $\boldsymbol{m}(\boldsymbol{y}) \times (\boldsymbol{m}(\boldsymbol{y}) \times (\Delta \boldsymbol{m}(\boldsymbol{y})))$, together with $|\boldsymbol{m}| \equiv 1$, gives an equivalent equation valid again for a.e. $\boldsymbol{y} \in \boldsymbol{\Gamma}$:

$$\partial_t \boldsymbol{m}(\boldsymbol{y}) = \Delta \boldsymbol{m}(\boldsymbol{y}) + \boldsymbol{m}(\boldsymbol{y}) \times \Delta \boldsymbol{m}(\boldsymbol{y}) - (\nabla \boldsymbol{m}(\boldsymbol{y}) : \nabla \boldsymbol{m}(\boldsymbol{y}))\, \boldsymbol{m}(\boldsymbol{y}) \tag{3.15}$$

$$+ \boldsymbol{m}(\boldsymbol{y}) \times \hat{\mathcal{C}}(W, \boldsymbol{m}(\boldsymbol{y})) - \boldsymbol{m}(\boldsymbol{y}) \times \Big(\boldsymbol{m}(\boldsymbol{y}) \times \hat{\mathcal{C}}(W, \boldsymbol{m}(\boldsymbol{y}))\Big) \quad \text{in } D_T. \tag{3.16}$$

3.2.3 Space and time Hölder regularity of solutions of the random LLG equation

In the present section, we prove that the sample paths of solutions of the random LLG equation (3.13) 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 $C^{k+\alpha}(D) := \{v : D \to \mathbb{C} : D^j v \in C^\alpha(D)\ \forall j = 0, \dots, k\}$ 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

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

For a function $v : D_T \to \mathbb{C}$, define the seminorm $|v|_{C^{\alpha/2,\alpha}(D_T)} \coloneqq \sup_{\substack{P,Q\in D_T \\ P\neq Q}} \frac{|v(P)-v(Q)|}{d(P,Q)^\alpha}$. Define the Banach spaces $C^{\alpha/2,\alpha}(D_T) \coloneqq \left\{v : D_T \to \mathbb{C} : v \in C^0(D_T) \text{ and } |v|_{C^{\alpha/2,\alpha}(D)} < \infty\right\}$ with the norm (see [WYW06, Section 1.2.3] for details) $\|v\|_{C^{\alpha/2,\alpha}(D_T)} \coloneqq \|v\|_{C^0(D_T)} + |v|_{C^{\alpha/2,\alpha}(D_T)}$. Finally,

$$C^{1+\alpha/2,2+\alpha}(D_T) \coloneqq \left\{v : D_T \to \mathbb{C} : \partial_t v \text{ and } D^j v \in C^{\alpha/2,\alpha}(D_T), j = 0,1,2\right\} \tag{3.17}$$

is a Banach space when endowed with the norm

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

In what follows, we work with the Hölder seminorm

$$|v|_{C^{1+\alpha/2,2+\alpha}(D_T)} \coloneqq |v|_{C^{\alpha/2,\alpha}(D_T)} + \sum_{j=1}^{2} \left\|D^j v\right\|_{C^{\alpha/2,\alpha}(D_T)} + \|\partial_t v\|_{C^{\alpha/2,\alpha}(D_T)}. \tag{3.18}$$

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 $\|uv\|_{C^{\alpha/2,\alpha}(D_T)} \leq \|u\|_{C^{\alpha/2,\alpha}(D_T)} \|v\|_{C^{\alpha/2,\alpha}(D_T)}$.

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 (3.13), obtained analogously to (LLA) (see Section 1.3.2 for more on equivalent forms of the LLG equation):

$$\lambda\partial_t \boldsymbol{m} + \boldsymbol{m} \times \partial_t \boldsymbol{m} = \Delta \boldsymbol{m} + |\nabla \boldsymbol{m}|^2 \boldsymbol{m} - \boldsymbol{m} \times \left(\boldsymbol{m} \times \hat{\mathcal{C}}(W, \boldsymbol{m})\right), \tag{3.19}$$

where we recall that $\lambda > 0$ is the Gilbert damping parameter, $\hat{\mathcal{C}}$ was defined in (1.75), and we wrote $\Delta \boldsymbol{m} \cdot \boldsymbol{m} = -|\nabla \boldsymbol{m}|^2$ (a result of the a.e. unit modulus condition on $\boldsymbol{m}$).

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

Theorem 3.7. *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 $\|\boldsymbol{M}_0\|_{2+\alpha} \leq \varepsilon$, $\|\Delta \boldsymbol{g}\|_\alpha \leq \varepsilon$, and $\|\nabla \boldsymbol{g}\|_\alpha \leq \varepsilon$, then the solution $\boldsymbol{m}$ of equation (3.19) 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,*

$$\|\boldsymbol{m}\|_{1+\alpha/2,2+\alpha} \leq C_{\boldsymbol{\Gamma}}, \tag{3.20}$$

where $C_{\boldsymbol{\Gamma}} > 0$ 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 [FT17b]. The proofs in the mentioned work require higher temporal regularity than is available for SLLG, which we circumvent by the use of Hölder spaces instead of Sobolev spaces. In the following, we require some notation:

$$H(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \coloneqq \boldsymbol{u} \times (\boldsymbol{v} \times \hat{\mathcal{C}}(W, \boldsymbol{w})) \quad \forall \boldsymbol{u}, \boldsymbol{v} \in C^{\alpha/2,\alpha}(D_T), \boldsymbol{w} \in C^{\alpha/2,1+\alpha}(D_T), \tag{3.21}$$

$$\mathcal{R}_a(\boldsymbol{v}) \coloneqq \lambda\partial_t \boldsymbol{v} + \boldsymbol{v} \times \partial_t \boldsymbol{v} - |\boldsymbol{v}|^2 \Delta \boldsymbol{v} - |\nabla \boldsymbol{v}|^2 \boldsymbol{v} + H(\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{v}) \quad \forall \boldsymbol{v} \in C^{1+\alpha/2,2+\alpha}(D_T), \tag{3.22}$$

$$L\boldsymbol{v} \coloneqq L_{\boldsymbol{x}_0}\boldsymbol{v} \coloneqq \lambda \boldsymbol{v} + \boldsymbol{M}_0(\boldsymbol{x}_0) \times \boldsymbol{v} \quad \forall \boldsymbol{x}_0 \in D, \boldsymbol{v} \in C^{\alpha/2,\alpha}(D_T). \tag{3.23}$$

We note that $\mathcal{R}_a$ is the residual defined from the alternative form (3.19) of the LLG equation; confer (3.3).

We will require a couple of technical results.

Lemma 3.8 (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*

$$\|H(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\|_{\alpha/2,\alpha} \le C_{\boldsymbol{g}} \|\boldsymbol{u}\|_{\alpha/2,\alpha} \|\boldsymbol{v}\|_{\alpha/2,\alpha} \|\boldsymbol{w}\|_{\alpha/2,1+\alpha}, \tag{3.24}$$

where $C_{\boldsymbol{g}} \coloneqq \left(1+\|\boldsymbol{g}\|_{1+\alpha}\right)^3 \left(\|\nabla \boldsymbol{g}\|_\alpha + \|\Delta \boldsymbol{g}\|_\alpha\right)$. 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

$$\|\mathcal{R}_a(\boldsymbol{v})\|_{\alpha/2,\alpha} \le \left(|\boldsymbol{v}|_{1+\alpha/2,2+\alpha} + |\boldsymbol{v}|^2_{1+\alpha/2,2+\alpha}\right)\left(\lambda + \|\boldsymbol{v}\|_{1+\alpha/2,2+\alpha}\right)^2 + C_{\boldsymbol{g}} \|\boldsymbol{v}\|^2_{\alpha/2,\alpha} \|\boldsymbol{v}\|_{\alpha/2,1+\alpha}. \tag{3.25}$$

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 (3.24), note the following elementary estimates

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

Putting these facts together, one obtains (3.24). To get the second inequality (3.25), estimate

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

Using (3.24) to estimate the last term yields (3.25).

□

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 3.9 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 3.9. *For $\boldsymbol{v},\boldsymbol{w} \in C^{1+\alpha/2,2+\alpha}(D_T)$ and $\boldsymbol{x}_0 \in D$, there holds*

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

where $C_{\boldsymbol{g}} > 0$ is defined in Lemma 3.8.

Proof. All but the last term in the definition of $\mathcal{R}_a$ are estimated as in [FT17b]. As for the last term, observe that

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

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 3.7, we use a fixed point iteration.

Proof of Theorem 3.7. 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 $\ell = 0, 1, \dots$

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 3.8 implies that $\boldsymbol{r}_\ell \in C^{\alpha/2,\alpha}(D_T)$. The parabolic regularity result [LSU68, 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

$$\|\boldsymbol{R}_{j+1}\|_{1+\alpha/2,2+\alpha} \le C_s \|\boldsymbol{r}_{j+1}\|_{\alpha/2,\alpha} = C_s \|\mathcal{R}_a(\boldsymbol{m}_{j+1})\|_{\alpha/2,\alpha} = C_s \|\mathcal{R}_a(\boldsymbol{m}_j - \boldsymbol{R}_j)\|_{\alpha/2,\alpha}, \quad (3.26)$$

where $C_s > 0$ is the stability constant from [LSU68, Theorem 10.4,§10, VII], which only depends on D_T and L (particularly, it is independent of ℓ). We invoke Lemma 3.9 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

$$\begin{aligned} \|\mathcal{R}_a(\boldsymbol{m}_j - \boldsymbol{R}_j)\|_{\alpha/2,\alpha} \le& \left\|\boldsymbol{m}_j - \boldsymbol{M}^0(\boldsymbol{x}_0)\right\|_{\alpha/2,\alpha} \|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha} + \left\|\left(1 - |\boldsymbol{m}_j|^2\right)\Delta \boldsymbol{R}_j\right\|_{\alpha/2,\alpha} \\ &+ \|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha} \left(|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha} + C_{\boldsymbol{g}}\right) \left(1 + \|\boldsymbol{m}_j\|_{1+\alpha/2,2+\alpha}\right)^2 \\ &+ \|\boldsymbol{R}_j\|^2_{1+\alpha/2,2+\alpha} \left(1 + (1 + C_{\boldsymbol{g}})\,|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}\right) \\ &+ \|\boldsymbol{R}_j\|^3_{1+\alpha/2,2+\alpha} (1 + C_{\boldsymbol{g}}). \end{aligned} \quad (3.27)$$

Let us estimate the first term in (3.27). For any $(t,\boldsymbol{x}) \in D_T$, the fundamental theorem of calculus yields $|\boldsymbol{m}_j(\boldsymbol{x},t) - \boldsymbol{M}^0(\boldsymbol{x}_0)| \lesssim \|(\partial_t,\nabla)\boldsymbol{m}_j\|_{C^0(D_T)} \leq |\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}$. Analogously, we get

$$\big\|\boldsymbol{m}_j - \boldsymbol{M}^0(\boldsymbol{x}_0)\big\|_{\alpha/2,\alpha} \leq 2\,|\boldsymbol{m}_j|_{1+\alpha/2,2+\alpha}\,.$$

Let us estimate the second term in (3.27). Since $\boldsymbol{m}_j = \boldsymbol{m}_0 + \sum_{i=0}^{j-1}\boldsymbol{R}_i$ and $|\boldsymbol{m}_0| = 1$ a.e., we have $|\boldsymbol{m}_j|^2 = 1 + 2\boldsymbol{m}_0 \cdot \sum_{i=0}^{j-1}\boldsymbol{R}_i + \Big(\sum_{i=0}^{j-1}\boldsymbol{R}_i\Big)^2$. Thus, the fact that Hölder spaces are closed under multiplication and the triangle inequality imply

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

All in all, we obtain

$$\|\boldsymbol{R}_{j+1}\|_{1+\alpha/2,2+\alpha} \leq \tilde{C}Q_j\,\|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}\,, \tag{3.28}$$

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

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

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 [FT17b], using additionally the assumption on the smallness of $\nabla\boldsymbol{g}$ and $\Delta\boldsymbol{g}$. Therefore, $\|\boldsymbol{R}_{j+1}\|_{1+\alpha/2,2+\alpha} \leq q\,\|\boldsymbol{R}_j\|_{1+\alpha/2,2+\alpha}$, 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 3.7.

Step 3. $\boldsymbol{m}$ solves (3.19)*:* $\boldsymbol{m}$ fulfills 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 (3.28) 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 [FT17b, Lemma 4.8] show that $\mathcal{R}_a(\boldsymbol{m}) = 0$ implies that $\boldsymbol{m}$ solves (3.19) and hence concludes the proof. □

3.3 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 3.1 for the case of the SLLG equation. Fix $0 < \alpha < 1$ and consider the complex parameter set

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\alpha) \coloneqq \Big\{\boldsymbol{z} \in \mathbb{C}^{\mathbb{N}} : \|\boldsymbol{z}\|_{\boldsymbol{\Sigma},\alpha} < \infty\Big\}, \quad \text{where } \|\boldsymbol{z}\|_{\boldsymbol{\Sigma},\alpha} \coloneqq \sum_{\ell\in\mathbb{N}_0}\max_{j=1,\dots,\lceil 2^{\ell-1}\rceil}|z_{\ell,j}|\,2^{-(1-\alpha)\ell/2}. \tag{3.29}$$

where we used the hierarchical indexing (1.11). For real parameters consider the subspace

$$\boldsymbol{\Gamma} \coloneqq \boldsymbol{\Sigma} \cap \mathbb{R}^{\mathbb{N}} \tag{3.30}$$

with the same norm. The definition of the Banach spaces for real and complex coefficients sample paths follows from the Lévy-Ciesielski expansion (3.2):

$$\mathbb{W} \coloneqq \{W : [0,T] \to \mathbb{C}\}\, \exists \boldsymbol{z} \in \mathcal{X} \text{ such that } W(t) = \sum_{n\in\mathbb{N}} z_n \eta_n(t) \forall t \in [0,T],$$
$$\mathbb{W}_{\mathbb{R}} \coloneqq \{W : [0,T] \to \mathbb{R}\}\, \exists \boldsymbol{y} \in \mathcal{X}_{\mathbb{R}} \text{ such that } W(t) = \sum_{n\in\mathbb{N}} y_n \eta_n(t) \forall t \in [0,T].$$

It is however interesting to identify classical spaces to which they belong.

Remark 3.10. *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 3.1.1) we have*

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

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

$$W(\boldsymbol{\Gamma}) \subset C^{\alpha/2}([0,T];\mathbb{R}), \tag{3.31}$$
$$\mathbb{W} = W(\boldsymbol{\Sigma}) \subset C^{\alpha/2}([0,T]), \tag{3.32}$$

with the same techniques used in the proof of Lemma 3.14 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

$$\boldsymbol{m}(\omega,t,\boldsymbol{x}) = \boldsymbol{M}^0(\boldsymbol{x}) + u(\omega,t,\boldsymbol{x}) \qquad \text{for a.e. } \omega \in \Omega, (t,\boldsymbol{x}) \in D_T,$$

where we recall $\boldsymbol{M}^0$ is the given unit-modulus initial condition, assume to belong to $C^{2+\alpha}(D)$. Consider then

$$u \in \mathbb{U} = C_0^{1+\alpha/2,2+\alpha}(D_T) \coloneqq \left\{ \boldsymbol{v} \in C^{1+\alpha/2,2+\alpha}(D_T) : \boldsymbol{v}(0) = \boldsymbol{0} \text{ in } D,\ \partial_n \boldsymbol{v} = \boldsymbol{0} \text{ on } \partial D \right\}. \tag{3.33}$$

See Section 3.2.3 for the definition of the relevant Hölder spaces. Given a noise coefficient $\boldsymbol{g} \in C^{2+\alpha}(D)$, we define the residual as:

$$\begin{aligned} \mathcal{R}(W,u) \coloneqq& \widetilde{\mathcal{R}}(W, \boldsymbol{M}^0 + u), \qquad \text{where} \\ \widetilde{\mathcal{R}}(W,\boldsymbol{m}) \coloneqq& \partial_t \boldsymbol{m} - \Delta \boldsymbol{m} - \boldsymbol{m} \times \Delta \boldsymbol{m} + (\nabla \boldsymbol{m} : \nabla \boldsymbol{m})\, \boldsymbol{m} - \boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{m}) + \\ &+ \boldsymbol{m} \times \left(\boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{m}) \right). \end{aligned} \tag{3.34}$$

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

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

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

$$\mathbb{F} = C^{\alpha/2,\alpha}(D_T), \tag{3.35}$$

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

$$\mathcal{R} : \mathbb{W} \times \mathbb{U} \to \mathbb{F}, \qquad (W,\boldsymbol{m}) \mapsto \mathcal{R}(W,\boldsymbol{m}). \tag{3.36}$$

Observe that we already proved Assumption 1 in Theorem 3.7.

3.3.1 Proof of Assumptions 2 and 3

In order to apply the general strategy outlined in Section 3.1, we need to prove Assumption 2 and 3 for the problem defined by (3.34).

To this end, we apply the following lemma found in much more general form, e.g., in [LSU68, Chapter VII, § 10, Theorem 10.3].

Lemma 3.11 (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 let $D_T := [0,T] \times D$. Denote by a_{ij}, a_i, a for $i,j = 1,\dots,d$ 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 sense that the principal part $\mathcal{L}_0$ of the elliptic operator satisfies: There exists $\delta > 0$ such that for a.e. $(t, \boldsymbol{x}) \in D_T$,*

$$\Re\langle \mathcal{L}_0(t,\boldsymbol{x})\boldsymbol{z}, \boldsymbol{z}\rangle \geq |\boldsymbol{z}|^2 \qquad \forall \boldsymbol{z} \in \mathbb{C}^3,$$

where $\langle\cdot,\cdot\rangle$ and $|\cdot|$ denote the standard scalar product and norm on $\mathbb{C}^3$ (see also , e.g., [LSU68, 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 $\|u\|_{1+\alpha/2,2+\alpha} \leq C_{stab} \|f\|_{\alpha/2,\alpha}$. 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 3.12. *Note that the compatibility conditions in [LSU68, 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 [LSU68, 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, analyzing the proof of [LSU68, Chapter VII, § 10, Theorem 10.3] gives the dependence of C_{stab} on the coefficients of the problem.*

Lemma 3.13. *Let $\alpha \in (0,1)$, $\boldsymbol{g} \in C^{2+\alpha}(D)$ and $\boldsymbol{M}^0 \in C^{2+\alpha}(D)$. Consider the spaces $\mathbb{W}, \mathbb{U}, \mathbb{F}$ defined at the beginning of the present section. Then, the residual $\mathcal{R}$ (cf. (3.34), (3.36)) 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 $\mathbb{F}$ if*

$$W \in \mathbb{W} \quad \text{and} \quad u \in \mathbb{U} \text{ satisfies } \|\Im u\|_{L^\infty(D_T)} \leq \frac{1}{4}. \tag{3.37}$$

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

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

and $C_{stab} = C_{stab}(\|u\|_{\mathbb{U}}) > 0$ is as in Lemma 3.11, i.e. it guarantees that for any $f \in \mathbb{F}$, $W \in \mathbb{W}$, $u \in \mathbb{U}$ $\left\|(\partial_u \mathcal{R}(W,u))^{-1} f\right\|_{\mathbb{U}} \leq C_{stab}(\|u\|_{\mathbb{U}}) \|f\|_{\mathbb{F}}$.

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 of (3.34) 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 $\omega \in C^{\alpha/2}([0,T])$,

$$\partial_W \mathcal{E}(W,\boldsymbol{m})[\omega] = (\cos(W)\mathcal{C}\boldsymbol{m} + \sin(W)(G\mathcal{C}+\mathcal{C}G)\boldsymbol{m})\,\omega, \tag{3.38}$$

$$\partial_W \hat{\mathcal{C}}(W,\boldsymbol{m})[\omega] = e^{WG}\partial_W \mathcal{E}(W,\boldsymbol{m})[\omega] + \big(\cos(W)G\mathcal{E}(W,\boldsymbol{m}) + \sin(W)G^2\mathcal{E}(W,\boldsymbol{m})\big)\,\omega, \tag{3.39}$$

$$\partial_W \widetilde{\mathcal{R}}(W,\boldsymbol{m})[\omega] = -\boldsymbol{m} \times \partial\hat{\mathcal{C}}(W,\boldsymbol{m})[\omega] + \boldsymbol{m} \times \Big(\boldsymbol{m} \times \partial\hat{\mathcal{C}}(W,\boldsymbol{m})[\omega]\Big). \tag{3.40}$$

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

$$\begin{aligned}\|\partial_W \mathcal{R}(W,u)[\omega]\|_{C^{\alpha/2,\alpha}(D_T)} \le \Big(1 + \big\|e^{\Im W}\big\|_{C^{\alpha/2}([0,T])}\Big)^2 \Big(1 + \|\boldsymbol{g}\|_{C^{2+\alpha}(D)}\Big)^4 \\ \Big(1 + \big\|\boldsymbol{M}^0 + u\big\|_{C^{\alpha/2,1+\alpha}(D_T)}\Big)^3 \|\omega\|_{C^{\alpha/2}([0,T])}.\end{aligned} \tag{3.41}$$

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

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

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

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

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 $\big|e^{\Im W}\big|_{C^{\alpha/2}([0,T])} \lesssim e^{\varepsilon}\varepsilon$ and

$$\big\|e^{\Im W}\big\|_{C^{\alpha/2}([0,T])} \lesssim e^{\varepsilon}(1+\varepsilon). \tag{3.42}$$

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

$$\begin{aligned}\partial_{\boldsymbol{m}}\widetilde{\mathcal{R}}(W,\boldsymbol{m})[\boldsymbol{v}] =& \partial_t \boldsymbol{v} - \Delta \boldsymbol{v} - \boldsymbol{v} \times \Delta \boldsymbol{m} - \boldsymbol{m} \times \Delta \boldsymbol{v} + 2(\nabla \boldsymbol{v} : \nabla \boldsymbol{m})\boldsymbol{m} + (\nabla \boldsymbol{m} : \nabla \boldsymbol{m})\, \boldsymbol{v} \\ &- \left(\boldsymbol{v} \times \hat{\mathcal{C}}(W,\boldsymbol{m}) + \boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{v})\right) \\ &- \left(\boldsymbol{v} \times \left(\boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{m})\right) + \boldsymbol{m} \times \left(\boldsymbol{v} \times \hat{\mathcal{C}}(W,\boldsymbol{m}) + \boldsymbol{m} \times \hat{\mathcal{C}}(W,\boldsymbol{v})\right)\right),\end{aligned} \tag{3.43}$$

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 of real coefficients $W \in W(\boldsymbol{\Gamma})$, $u \in u(\boldsymbol{\Gamma})$ such that $\mathcal{R}(W,u) = 0$, let us consider the more general case (3.37) for future use. Consider $\boldsymbol{f} \in \mathbb{F}$ (the residuals space defined in (3.35)) and the problem

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

With the aim of applying Lemma 3.11, we note that the principal part of $\partial_u \mathcal{R}(W,u)[\boldsymbol{v}]$ is $-\Delta \boldsymbol{v} - u \times \Delta \boldsymbol{v}$. We now show that for any $(t,\boldsymbol{x}) \in D_T$ and $\boldsymbol{w} \in \mathbb{C}^3$,

$$\Re\langle \boldsymbol{w} + u(t,\boldsymbol{x}) \times \boldsymbol{w}, \boldsymbol{w}\rangle \geq \frac{1}{2}\|\boldsymbol{w}\|^2, \tag{3.44}$$

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

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

and algebraic manipulations lead to the identity

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

which implies the estimate

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

Thus, by virtue of Assumption (3.37), we obtain (3.44). This shows that $\partial_u \mathcal{R}(W,u)$ is parabolic in the sense of Lemma 3.11 and hence, we obtain that $\partial_u \mathcal{R}(W,u)$ admits a continuous inverse. Together with its continuity, this implies that it is a homeomorphism. The norm of the inverse can be estimated as

$$\left\|\partial_u \mathcal{R}(W,u)^{-1}[\boldsymbol{f}]\right\|_{C^{1+\alpha/2,2+\alpha}(D_T)} \leq C_{\text{stab}}(W,u)\,\|\boldsymbol{f}\|_{C^{\alpha/2,\alpha}(D_T)}, \tag{3.45}$$

where $C_{\text{stab}}(W,u) > 0$ 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 (3.41) and (3.42) with

$$\mathcal{G}_1(s) = (1 + e^{\varepsilon_W}(1+\varepsilon_W))^2 \left(1 + \|\boldsymbol{g}\|_{C^{2+\alpha}(D)}\right)^4 \left(1 + \left\|\boldsymbol{M}^0\right\|_{C^{2+\alpha}(D)} + s\right)^3,$$

where $\varepsilon_W > 0$. The bound on $(\partial_u \mathcal{R}(W,u))^{-1}$ is proved in (3.45) with $\varepsilon_u = \frac{1}{4}$ and $\mathcal{G}_2 = C_{\text{stab}}$. 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 (3.37). □

We recall that, as shown in Section 3.1.2, the implicit function theorem (Theorem 3.2) and Theorem 3.5 prove the existence of $\varepsilon > 0$ such that for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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 (3.6) on the differential $\mathcal{U}'(W)$ as a function of $\mathcal{U}(W)$ through $\|\mathcal{U}(W)\|_{\mathbb{U}}$ under the assumption that $W \in B_{\min(\varepsilon(\boldsymbol{y}),\varepsilon_W)}(W(\boldsymbol{y}))$ in $\mathbb{W}$.

3.3.2 Proof of Assumption 4 and estimates of derivatives

Let us now estimate the derivatives of the parameter-to-solution map. While this is a standard technique established already in [CDS11], it turns out this will not be quite sharp enough to obtain dimension independent convergence of the sparse grid approximation. In Section 3.4, 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 (3.29) and (3.30).

Lemma 3.14. *Assumption 4 holds in the present setting. In particular, it is sufficient to choose* $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ *such that*

$$\|\boldsymbol{\rho}\|_{\boldsymbol{\Sigma}} \leq \frac{\varepsilon}{2}. \tag{3.46}$$

Proof. Fix $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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 (1.11) 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.,

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

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 3.10), we find

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

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

$$\rho_n = \varepsilon 2^{\frac{(1-\alpha)\lceil \log_2(n)\rceil}{2}} \qquad \forall n \in \mathbb{N}. \tag{3.47}$$

Having so concluded that for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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 3.6.

Proposition 3.15. *Consider $\boldsymbol{m} = \boldsymbol{M}^0 + u : \boldsymbol{\Gamma} \to C^{1+\alpha/2,2+\alpha}(D_T)$, the parameter-to-solution map solution of the parametric LLG equation with Hölder spaces ($\boldsymbol{\Gamma}$ and $C^{1+\alpha/2,2+\alpha}(D_T)$ defined in (3.30) and (3.17) respectively). Fix $\varepsilon > 0$ as in Theorem 3.5 and let $\boldsymbol{\rho} = (\rho_n)_{n\in\mathbb{N}}$ a positive sequence that satisfies (3.46). Then, for any $n \in \mathbb{N}_0$, $\boldsymbol{\nu} = (\nu_i)_{i=1}^n \subset \mathbb{N}_0^n$, it holds that*

$$\|\partial^{\boldsymbol{\nu}} \boldsymbol{m}(\boldsymbol{y})\|_{C^{1+\alpha/2,2+\alpha}(D_T)} \leq \prod_{j=1}^{n} \nu_j! \rho_j^{-\nu_j} C_\varepsilon \qquad \forall \boldsymbol{y} \in \boldsymbol{\Gamma}, \tag{3.48}$$

where $C_\varepsilon > 0$ from Theorem 3.5 is independent of $\boldsymbol{\nu}$ or $\boldsymbol{y}$.

Remark 3.16. *Note that we essentially proved "$(\boldsymbol{b}, \xi, \delta, X)$-holomorphy" [DNSZ23b, Definition 4.1] for the SLLG equation in the case of a Hölder-valued parameter-to-solution map. However, this regularity is not sufficient to apply the theory in [DNSZ23b], as the summability coefficient is $p = 2$, which lies out of the range $(0, \frac{2}{3})$ considered in [DNSZ23b]. This fact is analogous to what happens in our analysis.*

3.4 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 3.1. 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 SLLG equation, the results hold only for a simplified version of the stochastic input. We make the following modeling 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 (3.34)):

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

Consequently, we approximate $\hat{\mathcal{C}}(W, \boldsymbol{m})$ defined in (1.75) with the first order expansion

$$\tilde{\mathcal{C}}(W, \boldsymbol{m}) \coloneqq W \boldsymbol{m} \times \Delta \boldsymbol{g}, \tag{3.49}$$

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

$$\begin{aligned} \mathcal{R}_s(W, u) \coloneqq& \widetilde{\mathcal{R}}_s(W, \boldsymbol{M}^0 + u), \qquad \text{where} \\ \widetilde{\mathcal{R}}_s(W, \boldsymbol{m}) \coloneqq& \partial_t \boldsymbol{m} - \Delta \boldsymbol{m} - \boldsymbol{m} \times \Delta \boldsymbol{m} + (\nabla \boldsymbol{m} : \nabla \boldsymbol{m})\, \boldsymbol{m} - \boldsymbol{m} \times \tilde{\mathcal{C}}(W, \boldsymbol{m}) \\ &+ \boldsymbol{m} \times \Big(\boldsymbol{m} \times \tilde{\mathcal{C}}(W, \boldsymbol{m}) \Big). \end{aligned} \tag{3.50}$$

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 complex coefficients, we again start from the parameters: Define, for $1 < q < \infty$,

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

and we denoted $\boldsymbol{z}_\ell = (z_{\ell,1}, \dots, z_{\ell,\lceil 2^{\ell-1} \rceil})$. We then define the space of complex coefficients through the Lévy-Ciesielski expansion (3.2): $\mathbb{W} = W(\boldsymbol{\Sigma}) = \{W(\boldsymbol{z}, \cdot) : [0,T] \to \mathbb{C} : \boldsymbol{z} \in \boldsymbol{\Sigma}\}$.

For real parameters, we fix $\theta > 0$ and let

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\alpha, \theta) := \left\{ \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} : \|\boldsymbol{y}\|_{\boldsymbol{\Sigma},\alpha} < \theta \right\}, \tag{3.51}$$

where $\|\cdot\|_{\boldsymbol{\Sigma},\alpha}$ was defined in (3.29).

Lemma 3.17. *For fixed $1 < q < \infty$ and $\theta > 0$, there holds,*

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

Proof. To prove the first inclusion, fix $\boldsymbol{z} \in \boldsymbol{\Sigma}$ 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, recall the definition of Faber-Schauder basis functions (cf. Section 1.1.2) to compute $\|\eta_{\ell,j}\|_{L^q(0,T)} = 2^{-\ell(1/2+1/q)} 2^{-1/2} \left(\frac{2}{q+1}\right)^{1/q}$, so we get $\|W(\boldsymbol{z})\|_{L^q(0,T)} \le \|\boldsymbol{z}\|_{\boldsymbol{\Sigma}^q}$, which implies the first inclusion. The second inclusion follows with the methods of the proof of Lemma 3.14. □

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

$$\begin{aligned} \mathbb{U} = \big\{ u : D_T \to \mathbb{C}^3 : u \in L^q(0,T,C^{2+\alpha}(D)),\ \partial_t u \in L^q(0,T,C^\alpha(D)), \\ u(0,\cdot) = \mathbf{0} \text{ on } D,\ \partial_n u = \mathbf{0} \text{ on } [0,T] \times \partial D \big\}, \end{aligned} \tag{3.52}$$

and for solutions corresponding to real parameters, we have that

$$u(\boldsymbol{\Gamma}) \subset \left\{ u : D_T \to \mathbb{S}^2 : u \in \mathbb{U} \right\}. \tag{3.53}$$

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

$$\mathcal{R}_s : \mathbb{W} \times \mathbb{U} \to \mathbb{F}. \tag{3.54}$$

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 $U \in \mathbb{U}$, $\|u\|_{L^\infty(D_T)} + \|u\|_{L^2(0,T,C^1(D))} \leq \|u\|_{\mathbb{U}}$. 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

$$\left\|\tilde{\mathcal{C}}(W,\boldsymbol{m})\right\|_{L^q([0,T],C^\alpha(D))} \leq \|W\|_{L^q([0,T])} \|\boldsymbol{m}\|_{C^0([0,T],C^\alpha(D))} \|\boldsymbol{g}\|_{C^{2+\alpha}(D)}. \tag{3.55}$$

The proof of Theorem 3.7 can be transferred to this simplified version of LLG and hence we have that there exists $\overline{C}_{\boldsymbol{\Gamma}} = \overline{C}_{\boldsymbol{\Gamma}}(\theta) > 0$ such that

$$\|\mathcal{U}(W)\|_{\mathbb{U}} \leq \overline{C}_r \qquad \forall W \in \mathbb{W}_{\mathbb{R}}. \tag{3.56}$$

This gives the validity of Assumption 1 with $C_{\boldsymbol{\Gamma}} = \overline{C}_{\boldsymbol{\Gamma}}$ for the present problem.

3.4.1 Proof of Assumptions 2 and 3

In order to apply the general strategy outlined in Section 3.1.2, we need to prove Assumptions 2 and 3 for the spaces and residual chosen at the beginning of this section.

Remark 3.18. *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 $\partial_u \mathcal{R}_s(W,u) : \mathbb{U} \to \mathbb{F}$ which coincides with* (3.43) *but $\hat{\mathcal{C}}$ replaced by $\tilde{\mathcal{C}}$. For scalar problems, this can be found in [PS01, Section 4]. Strictly speaking, however, Lemma 3.19 only holds under the assumption that [PS01] can be generalized to the vector valued case.*

We can prove, analogously to Lemma 3.13, the following result:

Lemma 3.19. *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 $\mathbb{W}, \mathbb{U}, \mathbb{F}$ defined at the beginning of the present section. Then, the residual $\mathcal{R}_s$* (3.50), (3.54) *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 $\mathbb{F}$ if*

$$W \in \mathbb{W}, u \in \mathbb{U} : \|\Im u\|_{L^\infty(D_T)} \leq \frac{1}{4}. \tag{3.57}$$

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

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

where $C_{stab}(\|W\|_{\mathbb{W}}, \|u\|_{\mathbb{U}}) > 0$ is as c_p in [PS01, Theorem 2.5], i.e. it guarantees that $\left\|\left(\partial_u \mathcal{R}(W,u)\right)^{-1} f\right\|_{\mathbb{U}} \leq C_{stab}(\|W\|_{\mathbb{W}}, \|u\|_{\mathbb{U}}) \|f\|_{\mathbb{F}}$ for any $f \in \mathbb{F}$, $W \in \mathbb{W}$, $u \in \mathbb{U}$.

We recall that, as shown in Section 3.1.2, the implicit function theorem and Theorem 3.5 prove the existence of $\varepsilon > 0$ such that for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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 (3.6) on the differential $\mathcal{U}'(W)$ as a function of $\mathcal{U}(W)$ through $\|\mathcal{U}(W)\|_{\mathbb{U}}$ for all $W \in B_\varepsilon(W(\boldsymbol{y}))$ in $\mathbb{W}$.

3.4.2 Proof of Assumption 4 and estimates of derivatives

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 3.3.2, here $\boldsymbol{\rho}$ depends on which mixed derivative $\partial^{\boldsymbol{\nu}}$ is considered: Given a multi-index $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n) \in \mathbb{N}_0^n$,

$0 < \delta < \frac{1}{2}$ and $0 < \gamma < 1$ consider a sequence of positive numbers $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\nu}, \delta, \gamma)$ defined as follows:

$$\rho_{\ell,j} \coloneqq \gamma \begin{cases} 1 & \text{if } \nu_{\ell,j} = 0 \\ 2^{\left(\frac{3}{2}-\delta\right)\ell} \frac{1}{r_\ell(\boldsymbol{\nu})} & \text{if } \nu_{\ell,j} = 1 \\ 2^{\left(\frac{1}{2}-\delta\right)\ell} & \text{otherwise} \end{cases} \qquad \forall \ell \in \mathbb{N}_0, j = 1, \dots, \lceil 2^{\ell-1} \rceil, \tag{3.58}$$

where we used the hierarchical indexing (1.11) and $r_\ell(\boldsymbol{\nu}) \coloneqq \#\left\{j \in 1, \dots, \lceil 2^{\ell-1} \rceil : \nu_{\ell,j} = 1\right\}$.

Lemma 3.20. *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 (3.58) verifies Assumption 4.*

Proof. Let $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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 (cf Section 1.1.2), $\|\eta_{\ell,j}\|_{L^q(0,T)} \leq 2^{-(1/q+1/2)\ell}$ 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

$$\|W(\boldsymbol{z}) - W(\boldsymbol{y})\|_{L^q(0,T)} \leq \sum_{\ell \in \mathbb{N}_0} 2^{-(1/q+1/2)\ell} \sum_{j=1}^{\lceil 2^{\ell-1} \rceil} \rho_{\ell,j}. \tag{3.59}$$

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

$$\sum_{j=1}^{\lceil 2^{\ell-1} \rceil} \rho_{\ell,j} = \gamma \left(\#\left\{i : \nu_{\ell,i} = 0\right\} + 2^{\left(\frac{3}{2}-\delta\right)\ell} \frac{1}{r_\ell(\boldsymbol{\nu})} r_\ell(\boldsymbol{\nu}) + 2^{\left(\frac{1}{2}-\delta\right)\ell} \#\left\{i : \nu_{\ell,i} > 1\right\} \right). \tag{3.60}$$

Trivially, $\#\left\{i : \nu_{\ell,i} = 0\right\} \leq 2^\ell$ and $\#\left\{i : \nu_{\ell,i} > 1\right\} \leq 2^\ell$. This, together with (3.59) and (3.60) yields

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

The last series 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 $\boldsymbol{y} \in \boldsymbol{\Gamma}$ 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 3.6.

Proposition 3.21. *Consider $\boldsymbol{m} = \boldsymbol{M}^0 + u : \boldsymbol{\Gamma} \to \boldsymbol{M}^0 + u(\boldsymbol{\Gamma})$, the parameter-to-solution map of the parametric LLG equation defined in the beginning of this section, where $\boldsymbol{\Gamma}$ is defined in (3.51). Fix $\varepsilon > 0$ as in Theorem 3.5, 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 (3.58) and choose $0 < \gamma < 1$ such that Assumption 4 holds. Then, it holds that*

$$\|\partial^{\boldsymbol{\nu}} \boldsymbol{m}(\boldsymbol{y})\|_{\mathbb{U}} \leq \prod_{j=1}^n \nu_j! \rho_j^{-\nu_j} C_\varepsilon \qquad \forall \boldsymbol{y} \in \boldsymbol{\Gamma}, \tag{3.61}$$

where $C_\varepsilon > 0$ from Theorem 3.5 is independent of $\boldsymbol{\nu}$ or $\boldsymbol{y}$.

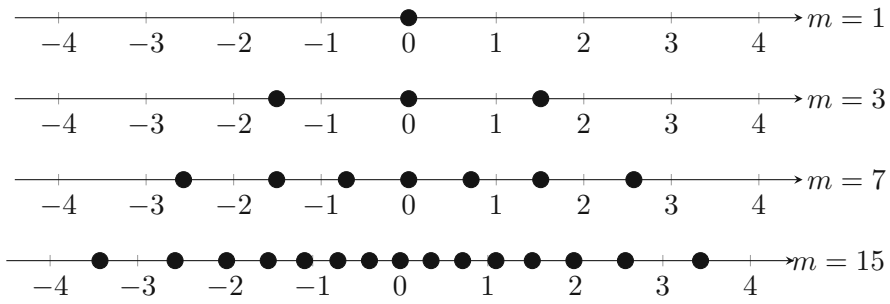

Figure 3.1: Examples of knots (3.62) for $p = 2$ on $\mathbb{R}$. It can be seen that the knots span a wider and wider portion of the real line and, at the same time, become denser. If the number of knots is suitably increased (for example using (3.69)), the resulting knots family is nested.

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

In the present section, we apply the sparse grid method (see Section 1.2.3) to approximate the parameter-to-solution map of the parametric LLG equation (3.14) derived above from the SLLG equation.

3.5.1 1D piecewise polynomial interpolation on $\mathbb{R}$

In this section, we define appropriate interpolation nodes on $\mathbb{R}$ and the corresponding piecewise-polynomial interpolant. We then prove algebraic convergence in the $L^2_\mu(\mathbb{R})$ norm, where μ denotes the Gaussian measure (in 1D).

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 the *knot sequence* $\mathcal{Y}^m = \{y_1^m, \dots, y_m^m\} \subset \mathbb{R}$ by

$$y_i^m := \phi\left(-1 + \frac{i}{m+1}\right) \qquad i = 1, \dots, m, \tag{3.62}$$

where

$$\phi(x) := \alpha \operatorname{erf}^{-1}(x) \qquad \forall x \in (-1,1), \tag{3.63}$$

$$\alpha = \alpha(p,\sigma^2) := \sqrt{\frac{4p}{1 - \frac{1}{\sigma^2}}}, \qquad \text{where } p \in \mathbb{N},\ \sigma^2 > 0 \tag{3.64}$$

The m knots define $m+1$ intervals (the first and last are unbounded). See Figure 3.1.

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

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

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 $u \in C^0(\mathbb{R})$,

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

We assume that for each $i = 1, \dots, m-1$, the interval (y_i, y_{i+1}) contains additional $p-2$ distinct interpolation *nodes* so that $I_m^p[u]$ is uniquely defined. Observe that in our case each interpolation *knot* (3.62) is also an interpolation *node* because we require a globally continuous piecewise polynomial interpolant.

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

$$C_\phi = \sqrt{\sigma^2} \left(\frac{\alpha\sqrt{\pi}}{2} \right)^{2p}, \tag{3.65}$$

where α was defined in (3.64).

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

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

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

$$\|u - I_m^p[u]\|_{L^2_{\tilde{\mu}}(\mathbb{R})} \le \tilde{C}_2 (m+1)^{-p} \frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}(\mathbb{R})}}{p!} \qquad \forall m \ge 3 \text{ odd},$$

where $\tilde{C}_2 = \sqrt{C_\phi \frac{p}{2} \left(m - 3 + 2^{2p+1}\right)}$ and C_ϕ was defined in (3.65).

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

For the second estimate, let $i \in \left\{ \frac{m(\nu)+1}{2}, \dots, m(\nu) - 2 \right\}$. 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:

$$\left(u - I_m^p[u]\right)(y) = \int_{y_i}^{y} \int_{\xi_1}^{z_1} \cdots \int_{\xi_{p-1}}^{z_{p-1}} \partial^p u \qquad \forall y \in [y_i, y_{i+1}] \tag{3.66}$$

where $\xi_j \in [y_i, y_{i+1}]$ is such that $\partial^j (u - I_m^p[u])(\xi_j) = 0$ for all $j = 1, \dots, p-1$. The Cauchy-Schwarz inequality applied to the last integral gives

$$\int_{\xi_{p-1}}^{z_{p-1}} \partial^p u \le \|\partial^p u\|_{L^2_{\tilde{\mu}}(\xi_{p-1}, z_{p-1})} \sqrt{\int_{\xi_{p-1}}^{z_{p-1}} \tilde{\mu}^{-1}}.$$

The monotonicity of the integral with respect to the integration domain and the fact that $\tilde{\mu}^{-1}$ is monotonically increasing on the positive semi-axis give

$$\int_{\xi_{p-1}}^{z_{p-1}} \partial^p u \le \|\partial^p u\|_{L^2_{\tilde{\mu}}(y_i, y_{i+1})}\, \tilde{\mu}^{-1}(y) \sqrt{z_{p-1} - y_i}.$$

Applying this to (3.66) and after integration, we obtain

$$\left(u - I_m^p[u]\right)(y) \leq \|\partial^p u\|_{L^2_{\tilde{\mu}}(y_i,y_{i+1})} \tilde{\mu}^{-1/2}(y) \frac{|y-y_i|^{p-1+\frac{1}{2}}}{(p-1)!}.$$

Then,

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

In order to estimate the last integral, change variables using ϕ defined in (3.63). 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

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

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

$$\int_{y_i}^{y_{i+1}} |u - I_m^p[u]|^2 (y)\mathrm{d}\mu(y) \leq C_\phi \frac{p^2}{2p} (m+1)^{-2p} \left(\frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])}}{p!} \right)^2. \tag{3.67}$$

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

$$\int_{y_{m-1}}^{+\infty} |u - I_m^p[u]|^2 (y)\mathrm{d}\mu(y) \leq C_\phi \frac{p^2}{2p} 2^{2p} (m+1)^{-2p} \left(\frac{\|\partial^p u\|_{L^2_{\tilde{\mu}}([y_i,y_{i+1}])}}{p!} \right)^2. \tag{3.68}$$

Combine (3.67) and (3.68) with analogous estimates for $i = 1, \ldots, \frac{m-1}{2}$ to finally obtain the second error estimate in the statement. □

Let us relate this to the notation introduced in Section 1.2.3: The *level-to-knot function* of the nodes family introduced above is:

$$m(\nu) \coloneqq 2^{\nu+1} - 1 \qquad \forall \nu \in \mathbb{N}_0. \tag{3.69}$$

Therefore, 1D nodes sequences and interpolants are expressed with respect to their index $\nu \in \mathbb{N}_0$ as:

$$\begin{aligned} \mathcal{Y}_\nu &= \mathcal{Y}_{m(\nu)}, \\ I_\nu &= I^p_{m(\nu)}. \end{aligned}$$

As mentioned above, since we use polynomial interpolation of degree $p-1$ on each interval, the 1D interpolant I_ν, $\nu \in \mathbb{N}_0$, has $(m(\nu)-1)(p-1)+1$ interpolation nodes. This includes the interpolation knots in order to give a globally continuous interpolant.

Observe that $(\mathcal{Y}_\nu)_{\nu\in\mathbb{N}_0}$ is nested, i.e. $\mathcal{Y}_\nu \subset \mathcal{Y}_{\nu+1}$ for all $\nu \in \mathbb{N}_0$. The 1D interpolant defines the detail operators $\Delta_\nu = I_\nu - I_{\nu-1}$ for $\nu \in \mathbb{N}_0$ and hierarchical surpluses $\Delta_{\boldsymbol{\nu}} = \bigotimes_n \Delta_{\nu_n}$ for $\boldsymbol{\nu} \in \mathcal{F}$ as explained in Section 1.2.3. We now apply the previous results to estimate 1D detail operators.

Lemma 3.23. *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 $C_1 = 2^{3/2}\tilde{C}_1 \sqrt{\int_0^\infty \sum_{j=1}^p \left|l_j'\right|^2 d\tilde{\mu}} \sqrt[4]{\int_0^{y_3} \tilde{\mu}^{-1}}$, $\tilde{C}_1 > 0$ was defined in the previous lemma, y_1, y_2, y_3 delimit the intervals of definition of the piecewise polynomial $I_3^p[u]$ 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

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

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

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

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

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

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

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

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. □

Recall the definition of finite-support multi-indices $\mathcal{F}$ (1.22). We can finally estimate hierarchical surpluses as follows.

Proposition 3.24. *Let $u : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, $p \geq 2$ and $\boldsymbol{\nu} \in \mathcal{F}$. Then*

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

where $\tilde{\boldsymbol{\mu}}$ denotes the infinite product measure $\tilde{\boldsymbol{\mu}} := \bigotimes_{n\in\mathbb{N}} \tilde{\mu}_n$ and $\tilde{\mu}_n = \tilde{\mu}$ for all $N \in \mathbb{N}$, 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, denoting by μ^N the N-dimensional standard Gaussian measure, by $\hat{\boldsymbol{\nu}}_1 = (\nu_2, \ldots, \nu_N)$ and $\hat{\mu}_1$ the $(N-1)$-dimensional standard Gaussian measure,

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

We apply the previous lemma (assume that $\nu_1 = 1$, the other case is analogous) to get

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

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

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

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

3.5.2 Basic profits and dimension dependent convergence

In the present section, we discuss the convergence of sparse grid approximation built using the profit-maximization technique explain in Section 1.2.3 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 3.3.

We estimate the derivatives appearing in the estimate from Proposition 3.24 with Proposition 3.15. We find

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

and $\rho_i = \varepsilon 2^{\frac{(1-\alpha)\lceil \log_2(i)\rceil}{2}}$. Recall now the framework presented in Section 1.2.3. Given a multi-index $\boldsymbol{\nu} \in \mathcal{F}$, we define as its value and work respectively

$$\tilde{v}_{\boldsymbol{\nu}} = \prod_{i\in\mathrm{supp}(\boldsymbol{\nu})} \tilde{v}_{\nu_i}, \tag{3.70}$$

$$w_{\boldsymbol{\nu}} = \prod_{i\in\mathrm{supp}(\boldsymbol{\nu})} p 2^{\nu_i}. \tag{3.71}$$

The definition of work is justified by the fact that the 1D interpolant is a degree $p-1$ piecewise-polynomial interpolant (see above). So in the bounded intervals defined by the $m(\nu) = 2^{\nu+1} - 1$

(for $\nu \in \mathbb{N}_0$) knots, there are $p-1$ additional interpolation nodes, which together with the knots give enough collocation nodes to uniquely define the piecewise polynomial.

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

$$\tilde{\mathcal{P}}_{\boldsymbol{\nu}} = \frac{\tilde{v}_{\boldsymbol{\nu}}}{w_{\boldsymbol{\nu}}}. \tag{3.72}$$

We apply Theorem 1.18 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 3.25. *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 equation under the assumption that $W(\boldsymbol{y}, t) = \sum_{n=1}^{N} y_n \eta_n(t)$ for all $t \in [0,T]$ and all $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$. Let $\Lambda_n \subset \mathbb{N}_0^N$ denote the optimal multi-index set (1.43) with $\#\Lambda_n = n$ defined using $\tilde{\mathcal{P}}_{\boldsymbol{\nu}}$ (3.72). Let $\mathcal{I}_{\Lambda_n}$ denote the corresponding piecewise polynomial sparse grids interpolant of degree $p-1$ with knots (3.62) and $p \geq 2$. Denote $\mathcal{H}_{\Lambda_n} \subset \mathbb{R}^N$ the corresponding sparse grid. Under the assumptions of Theorem 3.7, for any $\frac{2}{(1+\alpha)p} < \tau < 1$,*

$$\|\boldsymbol{m}_N - \mathcal{I}_{\Lambda_n} \boldsymbol{m}_N\|_{L^2_{\mu}(\mathbb{R}^N, C^{1+\alpha/2,2+\alpha}(D_T))} \leq C_{\tau,p}(N) \left(\#\mathcal{H}_{\Lambda_n}\right)^{1-1/\tau}, \tag{3.73}$$

where $C_{\tau,p}(N)$ is a function of τ, p, N defined as

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

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 3.23. In particular, the bound grows root-exponentially in the number of dimensions.

Proof. With the aim of applying the convergence Theorem 1.18, we estimate:

$$\begin{aligned} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^N} \mathcal{P}_{\boldsymbol{\nu}}^{\tau} w_{\boldsymbol{\nu}} &= \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^N} v_{\boldsymbol{\nu}}^{\tau} w_{\boldsymbol{\nu}}^{1-\tau} \leq \prod_{i=1}^{N} \sum_{\nu_i \geq 0} v_{\nu_i}^{\tau} w_{\nu_i}^{1-\tau} \\ &= \prod_{i=1}^{N} \left(1 + \left(C_1 \rho_i^{-1}\right)^{\tau} (2p)^{1-\tau} + \sum_{\nu_i \geq 2} \left(C_2 (2^{\nu_i} \rho_i)^{-p}\right)^{\tau} (p 2^{\nu_i})^{1-\tau} \right). \end{aligned}$$

The remainder of the proof consists of estimating the product under the condition on τ. □

3.5.3 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 3.4, 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 (3.58). This will be sufficient to obtain dimension-*independent* convergence.

Again, we work within the "profit maximization" framework described in Section 1.2.3.

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. Recall that here $\mathbb{U}$, the Banach space of solutions sample paths, is defined as (3.52). In particular, it consists

of Lebesgue integrable functions in time with index $q \approx 1$. We estimate the derivatives appearing in Proposition 3.24 with Proposition 3.21. This 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

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

Here, i and (ℓ, j) are related through the hierarchical indexing (1.11), $\delta > 0$ 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 (3.71), the profits now read

$$\mathcal{P}_{\boldsymbol{\nu}} = \frac{v_{\boldsymbol{\nu}}}{w_{\boldsymbol{\nu}}}. \tag{3.74}$$

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

$$\mathcal{F}^* \coloneqq \{\boldsymbol{\nu} \in \mathcal{F} : \nu_n \neq 1 \ \forall n \in \mathbb{N}\}$$

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,

$$\begin{aligned} \sum_{\boldsymbol{\nu} \in \mathcal{F}} v_{\boldsymbol{\nu}}^\tau w_{\boldsymbol{\nu}}^{1-\tau} &= \sum_{\boldsymbol{\nu} \in \mathcal{F}^*} \sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} v_{\hat{\boldsymbol{\nu}}}^\tau w_{\hat{\boldsymbol{\nu}}}^{1-\tau} \\ &= \sum_{\boldsymbol{\nu} \in \mathcal{F}^*} \sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} \prod_{n:\hat{\nu}_n \leq 1} \left(v_{\hat{\nu}_n}^\tau w_{\hat{\nu}_n}^{1-\tau}\right) \prod_{n:\hat{\nu}_n > 1} \left(v_{\hat{\nu}_n}^\tau w_{\hat{\nu}_n}^{1-\tau}\right) \\ &= \sum_{\boldsymbol{\nu} \in \mathcal{F}^*} \prod_{n:\nu_i > 1} \left(v_{\nu_n}^\tau w_{\nu_n}^{1-\tau}\right) \sum_{\hat{\boldsymbol{\nu}} \in K_{\boldsymbol{\nu}}} \prod_{n:\hat{\nu}_n \leq 1} \left(v_{\hat{\nu}_n}^\tau w_{\hat{\nu}_n}^{1-\tau}\right). \end{aligned} \tag{3.75}$$

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

$$\mathcal{F}\{0,1\} \coloneqq K_{\mathbf{0}} = \{\boldsymbol{\nu} \in \mathcal{F} : \nu_n \in \{0,1\} \, \forall n \in \mathbb{N}\}.$$

Lemma 3.26. *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 $\frac{1}{1+\varepsilon} \geq p$ 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 $\alpha^{|\boldsymbol{\nu}|_1} \leq C_\varepsilon \left(|\boldsymbol{\nu}|_1!\right)^\varepsilon$ 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

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

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

$$\sum_{\boldsymbol{\nu}\in\mathcal{F}\{0,1\}} (|\boldsymbol{\nu}|_1!\,\boldsymbol{a}^{\boldsymbol{\nu}})^q \lesssim \sum_{\boldsymbol{\nu}\in\mathcal{F}\{0,1\}} \left(\frac{|\boldsymbol{\nu}|_1!}{\boldsymbol{\nu}!}\left(\frac{\boldsymbol{a}}{\alpha}\right)^{\frac{1}{1+\varepsilon}\boldsymbol{\nu}}\right)^{(1+\varepsilon)q}. \tag{3.76}$$

Observe that $\sum_j (\frac{a_j}{\alpha})^{\frac{1}{1+\varepsilon}} < 1$ 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 $r \geq p(1+\varepsilon)$. Then, [CDS11, Theorem 1] implies that the second sum in (3.76) is finite, thus proving the statement. □

Lemma 3.27. *If $\tau > \frac{1}{\frac{3}{2}-\delta}$, there exists $C>0$ such that for any $\boldsymbol{\nu}\in\mathcal{F}^*$,*

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

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

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

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

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

and substitute

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

From the definition of $r_{\ell(n)}(\boldsymbol{d})$, we estimate $\prod_{n:d_n=1} r_{\ell(n)}(\boldsymbol{d}) \leq \prod_{\ell:\exists j:d_{\ell,j}=1} r_\ell(\boldsymbol{d})^{r_\ell(\boldsymbol{d})}$. Stirling's formula (see (D.1)) gives $r_\ell(\boldsymbol{d})^{r_\ell(\boldsymbol{d})} \leq r_\ell(\boldsymbol{d})! e^{r_\ell(\boldsymbol{d})}$. Denote $\boldsymbol{d}_\ell = \left(d_{\ell,1},\dots,d_{\ell,\lceil 2^{\ell-1}\rceil}\right)$ 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

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

Let $c_j \coloneqq C_1 2^{-\left(\frac{3}{2}-\delta\right)\ell(j)} (2p)^{(1-\tau)/\tau} e$ for all $j\in\mathbb{N}$ to get

$$\sum_{\hat{\boldsymbol{\nu}}\in K_{\boldsymbol{\nu}}} \prod_{n:\hat{\nu}_n\leq 1} \left(v_{\hat{\nu}_n}^{\tau} w_{\hat{\nu}_n}^{1-\tau}\right) \leq \sum_{\boldsymbol{d}\in D_{\boldsymbol{\nu}}} \left(|\boldsymbol{d}|_1! \boldsymbol{c}^{\boldsymbol{d}}\right)^{\tau}.$$

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

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

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

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

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 $\sum_{\boldsymbol{\nu}\in\mathcal{F}^*}\prod_{n:\nu_n>1}\left(v_{\nu_n}^{\tau}w_{\nu_n}^{1-\tau}\right) \leq \prod_{n\in\mathbb{N}}\left(1+F_{\ell(n)}\right)$. We can further estimate, recalling the hierarchical indexing (1.11),

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

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 3.28. *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 1.18 implies the following convergence.

Theorem 3.29. *Consider the parameter-to-solution map $\boldsymbol{m} = \boldsymbol{M}^0 + u$ solution of the parametric LLG equation as in Section 3.4. Recall that $\boldsymbol{M}^0 \in C^{2+\alpha}(D)$ is the initial condition such that $|\boldsymbol{M}| = 1$ a.e. on D and $u : \boldsymbol{\Gamma} \to \mathbb{U}$, with $\boldsymbol{\Gamma}$ and $\mathbb{U}$ defined in (3.51) and (3.52) respectively. Let $\Lambda_n \subset \mathcal{F}$ denote the multi-index set (1.43) with $\#\Lambda_n = n \in \mathbb{N}$ defined using the profits $\mathcal{P}_\nu$ (3.74). Let $\mathcal{I}_{\Lambda_n}$ denote the corresponding piecewise polynomial sparse grids interpolant of degree $p-1$ for $p \geq 3$ with knots (3.62). Assume that the corresponding sparse grid satisfies $\mathcal{H}_{\Lambda_n} \subset \boldsymbol{\Gamma}$. Under the assumptions of Theorem 3.7, for any $\frac{2}{3} < \tau < 1$,*

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

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

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

where in turn C is defined in Lemma 3.27 and C_2 is defined in Lemma 3.23.

Remark 3.30 (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*

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

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 3.31. *Given an approximation* $\boldsymbol{m}_\Lambda(\boldsymbol{y})$ *of the solution to the parametric LLG equation* (3.14)*, such as the sparse grid interpolant*

$$\mathcal{I}_\Lambda[\boldsymbol{m}](\boldsymbol{y})$$

defined and studied in this document, it is easy to sample an approximate random solution of the random *LLG equation* (3.13) *too: Sample i.i.d. standard normal random variables* $\boldsymbol{Y}_{N_\Lambda} = (Y_i)_{i=1}^{N_\Lambda}$ *and evaluate* $\boldsymbol{m}_\Lambda(\boldsymbol{Y}_{N_\Lambda})$*. Here* $N_\Lambda \in \mathbb{N}$ *denotes the number of active parameters in the sparse grid interpolant* $\mathcal{I}_\Lambda$*, i.e.* $N_\Lambda \coloneqq \min\{n \in \mathbb{N} : \forall \boldsymbol{\nu} \in \Lambda,\ supp\,(\boldsymbol{\nu}) \subset \{1, \dots, n\}\}$*. The root-mean-square error is naturally the same as the one we estimated in the previous theorem*

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

We can also draw approximate samples from the random solution of the stochastic *PDE* (1.68)*:*

1. *Sample a Wiener process* $W(\omega, \cdot)$*;*
2. *Compute the first* N_Λ *coordinates* $\boldsymbol{Y}_{N_\Lambda} = (Y_1, \dots, Y_{N_\Lambda}) \in \mathbb{R}^{N_\Lambda}$ *of its Lévy-Ciesielski expansion* $W(\omega, \cdot) = \sum_{i=1}^{\infty} Y_i \eta_i(\cdot)$ *(see Section 1.1.2);*
3. *Compute* $\boldsymbol{m}_\Lambda(\boldsymbol{Y}_{N_\Lambda})$*, the approximate solution to the random LLG equation* (3.13)*;*
4. *Finally compute the inverse Doss-Sussmann transform to obtain an approximation of the SLLG solution:* $\boldsymbol{M}_\Lambda \coloneqq e^{WG} \boldsymbol{m}_\Lambda$ *(recall the convenient expression* (1.74) *for* e^{WG}*).*

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 the identity* $\sqrt{\mathbb{E}_W \|\boldsymbol{M} - \boldsymbol{M}_\Lambda\|_{\mathbb{U}}} = \left\|e^{WG}(\boldsymbol{m} - \boldsymbol{m}_\Lambda)\right\|$*. Equation* (1.74) *followed by a triangle inequality then give*

$$\|\boldsymbol{M} - \boldsymbol{M}_\Lambda\| \leq \left(1 + \|\boldsymbol{g}\|_{L^\infty(D)} + \|\boldsymbol{g}\|^2_{L^\infty(D)}\right) \|\boldsymbol{m} - \boldsymbol{m}_\Lambda\|. \tag{3.77}$$

Remark 3.32 (Comparison with Monte Carlo quadrature)**.** *While the sparse grid interpolation does not offer any rate advantages over Monte Carlo to compute statistical quantities, it provides much more information on the solution. For example, it can be used to approximate minima and maxima of scalar quantities of interest of the solution.*

3.5.4 Numerical tests

We numerically test the convergence of the sparse grid interpolation defined above. Since no exact sample path of the solution is available, we approximate them with the high-order linearly implicit BDF scheme we discussed in Section 1.3.3. We recall that this method is based on the *tangent plane scheme* and has the advantage of solving one *linear* elliptic problem per timestep with finite elements. The method is high-order for both the finite elements and BDF discretization, but here we use for simplicity only order 1.

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

$$\boldsymbol{g}(\boldsymbol{x}) = \left(-\frac{1}{2}\cos(\pi x_1), -\frac{1}{2}\cos(\pi x_2), \sqrt{1 - \left(\frac{1}{2}\cos(\pi x_1)\right)^2 - \left(\frac{1}{2}\cos(\pi x_2)\right)^2} \right). \tag{3.78}$$

Observe that $\partial_n \boldsymbol{g} = 0$ on ∂D and $|\boldsymbol{g}| = 1$ 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 timesteps 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 3.5.2, namely

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

  where $\ell(i) = \lceil \log_2(i) \rceil$. Compared to (3.72), we set $C_1 = C_2 = \varepsilon = 1$ and $\alpha = 0$ for simplicity.

- The improved profit from Section 3.5.3, namely

$$\mathcal{P}_{\boldsymbol{\nu}} = \prod_{i:\nu_i=1} 2^{-\frac{3}{2}\ell(i)} \prod_{i:\nu_i>1} \left(2^{\nu_i+\frac{1}{2}\ell(i)}\right)^{-p} \left(\prod_{i:\nu_i\geq 1} p 2^{\nu_i} \right)^{-1} \qquad \forall \boldsymbol{\nu} \in \mathcal{F}, \tag{3.79}$$

  where again $\ell(i) = \lceil \log_2(i) \rceil$. Compared to the profits in Section 3.5.3, we set $C_1 = C_2 = \gamma = 1$ and neglected the factor $r_\ell(\boldsymbol{\nu})$.

We estimate the sparse grid approximation error with the Monte Carlo sum $\frac{1}{N}\sum_{i=1}^N \|\boldsymbol{m}_{\tau h}(\boldsymbol{y}_i) - \mathcal{I}_\Lambda[\boldsymbol{m}_{\tau h}](\boldsymbol{y}_i)\|_{L^2(0,T,\mathbb{H}^1(D))}$, 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 timestep 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 3.2. In the top left 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 3.5.3 and it

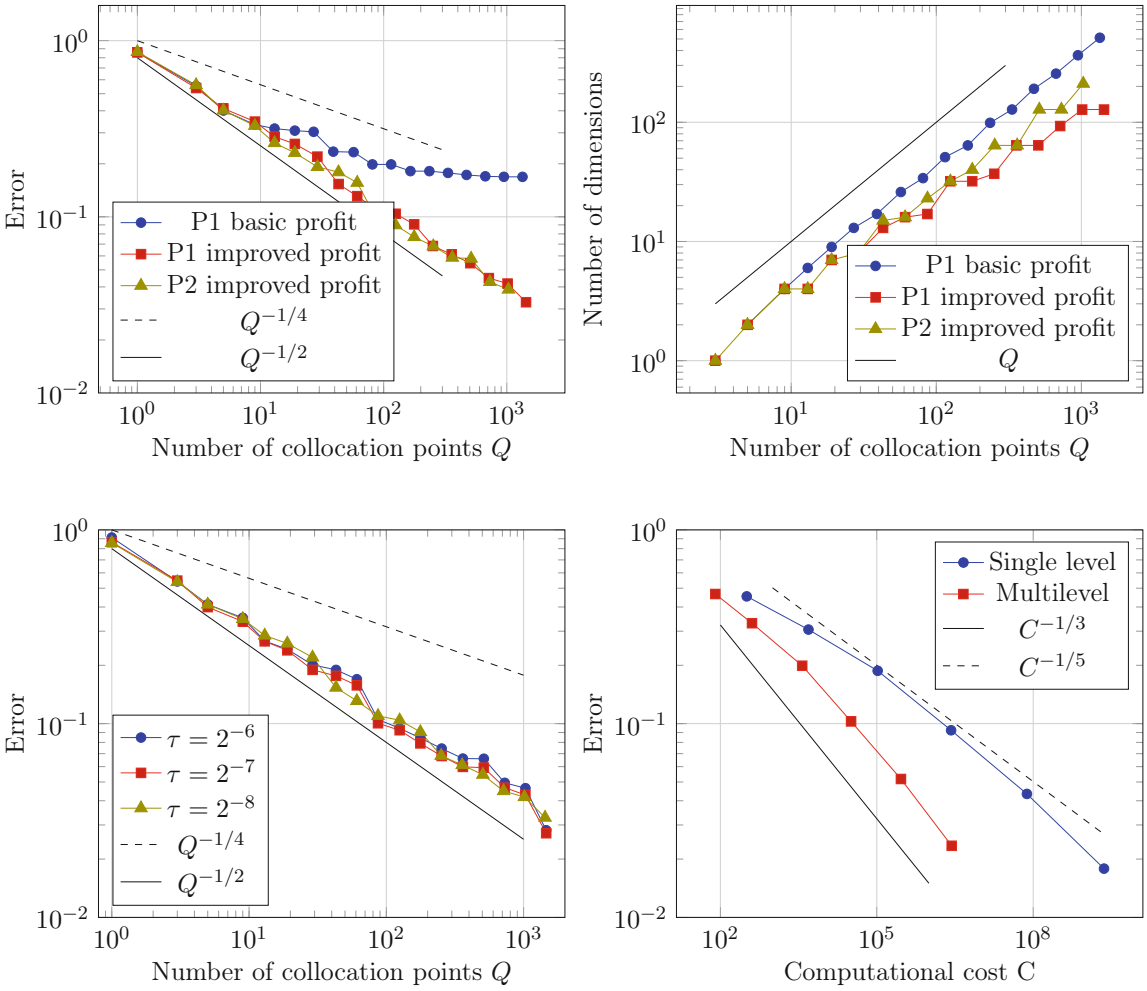

Figure 3.2: Approximation of parameter-to-solution map and multilevel approximation of full unknown. Top left: Error vs. number of collocation nodes; Top right: Number of effective dimensions vs. number of collocation nodes; Bottom left: Comparison of convergence of the sparse grid approximation ($p = 3$, i.e. piecewise quadratic) for different space and time discretization parameters. In all cases timestep τ and mesh size h are related by $h = 8\tau$. Bottom right: 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 [AFKL21] for time and space approximation.

delivers the same convergence rate as piecewise linear interpolation. Hence, the restriction in Theorem 3.29 is possibly an artifact of the proof. In view of Remark 3.30, it seems unnecessary to test higher polynomial degrees. In the top right 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, in the bottom left plot, we verify numerically that the approximation power of the method does not degrade when space and time approximations are refined.

3.6 Multilevel sparse grid–finite element approximation

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 high-order linearly implicit BDF scheme discussed in Section 1.3.3.

Given $\tau > 0$, consider $N_\tau = \frac{T}{\tau}$ equispaced timesteps 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 $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $\boldsymbol{m}_{\tau h}(\boldsymbol{y})$ 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

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

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

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

Indeed, the numerical scheme requires, at each timestep, 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 [KPP+19], for a mathematically rigorous preconditioning strategies for LLG.

Theorem 3.29 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,

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

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

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

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 $C^{\mathrm{SL}}_{\Lambda,\tau,h}$:

$$\left\|\boldsymbol{m} - \boldsymbol{m}^{\mathrm{SL}}_{\Lambda,\tau,h}\right\|_{L^2_\mu(\boldsymbol{\Gamma};\mathbb{U})} \lesssim \left(C^{\mathrm{SL}}_{\Lambda,\tau,h}\right)^{-\frac{1}{\frac{1}{r}+(d+1)}}. \tag{3.80}$$

A *Multilevel* approximation of $\boldsymbol{m}$ can be defined following [TJWG15]: Let $K \in \mathbb{N}_0$ 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 $k = 0,\dots,K$ and $\boldsymbol{m}_{-1} \equiv 0$. Define the multilevel approximation as

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

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

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

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

$$\left(\#\mathcal{H}_{\Lambda_{K-k}}\right)^{-r} \le C_{\mathrm{FE}} \left(C_{\mathrm{SG}}(K+1)\right)^{-1} \frac{\tau_K + h_K}{\tau_k + h_k}. \tag{3.81}$$

As a consequence, the multilevel error with optimal sparse grids sizes (3.81) can be estimated as $\left\|\boldsymbol{m} - \boldsymbol{m}_K^{\mathrm{ML}}\right\|_{L^2_\mu(\boldsymbol{\Gamma};\mathbb{U})} \le 2C_{\mathrm{FE}}(\tau_K + h_K)$. The error can be related to the computational cost as done in [TJWG15]. We obtain the improved error-to-cost relation

$$\left\|\boldsymbol{m} - \boldsymbol{m}_K^{\mathrm{ML}}\right\|_{L^2_\mu(\boldsymbol{\Gamma};\mathbb{U})} \lesssim \left(C_K^{\mathrm{ML}}\right)^{-\frac{1}{d+1}}. \tag{3.82}$$

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 (3.78). The initial condition $\boldsymbol{M}^0$ coincides with (3.78). 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 (3.79) 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 $K = 0,\dots,5$ and define $\tau_k = 2^{-k-2}$, $h_k = 2^{-k}$, and Λ_k using the same profit-maximization as in the previous section.

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 3.2, bottom right plot.

For the multilevel approximation, we follow formula (3.81). 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 cardinality of $\mathcal{H}_{\Lambda_5}$ for $K = 5$ should actually be at least 4082 if formula (3.81) is used. Here, we reduce it to 1500 in order to guarantee reasonable computational times.

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 timestep $\tau_{\text{ref}} = 2^{-9}$ and mesh size $h_{\text{ref}} = 2^{-7}$. 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.

Chapter 4

`SGMethods`: A Python implementation of sparse grid interpolation

In Chapter 3, we considered a sparse grid interpolation scheme for functions on $\mathbb{R}^{\mathbb{N}}$. We based the scheme on piecewise-polynomial interpolation (of fixed but arbitrary degree) with knots defined according to a normal distribution. This is a somewhat non-canonical choice (it is more common to work with interpolation in the zeros of orthogonal polynomials). In fact, we could not find an implementation of sparse grid interpolation that allowed for the easy implementation of this precise method.

This necessity sparked the idea of producing an in-house implementation of sparse grid interpolation that allows the user to input their 1D interpolation method of choice. Together with additional input, this is used by the library to produce a sparse grid interpolation operator, with a sampling and interpolation method and other functionalities. The name `SGMethods` stands for *Sparse Grid Methods* (plural), to highlight that the user is free to input any 1D interpolation method (together with nodes and a multi-index set), and the library automatically generates a corresponding sparse grid interpolant.

While `SGMethods` can be used to approximate any high-dimensional function, our main application was in the context of approximation of parametric coefficient PDEs in Chapter 3. To this end, we also needed to evaluate each collocation sample for fixed random parameters with an implementation of the chosen time approximation and/or finite element method. To this end, we chose finite elements as implemented in the library `Dolfin` [LW10], which offers a user-friendly Python interface and efficient computations.

The code of `SGMethods` and the implementation from Chapter 3 (used to produce the numerical experiments reported there) is available on the GitHub profile of the author[1]. The project is growing and additional documentation, tests and functionalities, such as the adaptive method from Chapter 2, will be added in the future.

The chapter is structured as follows: In Section 4.1, we list the main features of the library. In Section 4.2, we clarify the structure of the project by giving an overview of the content of each module. In Section 4.3, we present a commented example on the approximation of the affine diffusion Poisson problem.

4.1 Main Features

Some of the main features of the library are:

[1] `https://github.com/andreascaglioni/SGMethods`

- Python object oriented implementation, with documentation and tests;
- The user can define a tensor product interpolant class used for the sparse grid computation. This is possible because any sparse grid interpolant can be reduced to a linear combination of tensor product interpolants through the *inclusion-exclusion formula* (1.33). The user-defined tensor product interpolation class is given to the sparse grid interpolant class upon initialization and must fit the interface defined by `TPInterpolantWrapper`. In practice, it must be a class with the same visible attributes and methods of the class `TPLagrangeInterpolator` and `TPPwQuadraticInterpolator`;
- A rich choice of implemented nodes, multi-index sets, and tensor product interpolation methods is included (see respective modules description below);
- Efficient algorithms to enlarge the multi-index set by adding multi-indices to the margin. In particular, the ability to increase its dimensionality by adding appropriated multi-indices;
- When computing a sparse grid interpolant, some operations can be implemented in such a way that their complexity scales unfavorably with the number of parametric dimensions. Here, we implemented these operations in a way that, where possible, avoids this problem. An example is the computation of the combination coefficients $\{\alpha_{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda\}$ (1.33) from the inclusion-exclusion formula. The definition reads

$$\alpha_{\boldsymbol{\nu}} \coloneqq \sum_{\boldsymbol{i} \in \{0,1\}^N : \boldsymbol{\nu} + \boldsymbol{i} \in \Lambda} (-1)^{|\boldsymbol{i}|_1} . \tag{4.1}$$

  It is tempting to compute this quantity using the definition, i.e. through a loop over $\boldsymbol{i} \in \{0,1\}^N$. However, $\#\{0,1\}^N = 2^N$, so for example $\#\{0,1\}^N = 1024$ for $N = 10$ and $\#\{0,1\}^N > 10^{30}$ for $N = 100$. The complexity of this computation is affected by the curse of dimensionality. This problem can be circumvented by computing $\alpha_{\boldsymbol{\nu}}$ with a loop over the multi-indices in Λ instead:

Algorithm 11 $\alpha_{\boldsymbol{\nu}} \leftarrow$`compute_combination_coefficent`$(\boldsymbol{\nu}, \Lambda)$

```
1: α_ν = 0
2: for ν̂ ∈ Λ do
3:     i = ν̂ − ν
4:     if i ∈ {0,1}^N then
5:         α_ν ← α_ν + (−1)^|i|_1
6:     end if
7: end for
```

  This algorithm has complexity $\mathcal{O}(\#\Lambda N)$ (because of the loop and of the property to check in line 4), which is much smaller than 2^N if the problem has a relatively weak sparsity. Conversely, if $\#\Lambda N \gg 2^N$, the definition in (4.1) should be used. In our experience, this strategy brings a marked reduction in computational time compared to the use of the naive algorithm. All in all, the result of this and other optimizations is that, when approximating parametric PDEs, most of the computational time is spent sampling the parameter-to-solution map;
- We implement recycling when sampling the function to interpolate on the sparse grid. This means that the user is allowed to provide any saved values obtained e.g. from a previous interpolation on a coarser sparse grid (see method `SGInterpolant.sampleOnSG`);

- A similar recycling strategy is implemented in the computation of the a-posteriori parametric error (see `compute_aposteriori_estimator.py`), whose definition and theoretical aspects are discussed in Chapter 2. Here, we recycle values of the pointwise estimator when they are unchanged from a previous computation;

- Finally, we implement a multilevel sparse grid interpolation class `MLInterpolant` with the structure analysed in [TJWG15]. This is a class initialized with a sequence of interpolants with increasing resolution and a function to compute a finite element approximation at any desired resolution (indexed by a scalar parameter). Therefore, it may be used for other multilevel algorithms as well, possibly with small modifications.

4.2 Project structure

The project is structured in several directories or modules:

- `SGMethods`: Core functions for the defining the sparse grid interpolant and ancillary functions;

- `SLLG`: Functions used to sample the SLLG parameter-to-solution map. It includes `Fenics` code for the `HOLIBDF` method (see Section 1.3.3), functions to compute Karhunen-Loève and Lévy-Ciesielski expansions of the Wiener process, functions to compute several error metrics in parameter and physical space, and some example with the (parametric version of the) SLLG equation;

- `tests`: Unit-test files for `SGMethods`;

- `tutorials`: Simple examples with detailed comments aimed at giving a practical introduction to `SGMethods`;

- `utils`: Helper functions used in several scripts;

The functions within the `SGMethods` directory are contained in the following modules:

- `ScalarNodes`: Functions to generate 1D nodes families. It includes bounded nodes (equispaced on $[-1, 1]$, equispaced on $(-1, 1)$, Clenshaw-Curtis, Leja) and unbounded nodes families on $\mathbb{R}$ (including Hermite interpolation nodes and Leja nodes with respect to the Gaussian measure). Nodes families are either nested or non-nested;

- `TPKnots`: Simple function to generate tensor product nodes grids given 1D nodes;

- `TPLagrangeInterpolator`: Class for tensor product Lagrange interpolation on given tensor product grid;

- `TPPwCubicInterpolator`: Class for tensor product piecewise cubic interpolation on given tensor product grid;

- `TPPwQuadraticInterpolator`: Class for tensor product piecewise quadratic interpolation on given tensor product grid;

- `TPInterpolantWrapper`: Wrapper class to unify the interface of several tensor product interpolant classes (in particular the `scipy.interpolate.RegularGridInterpolator` piecewise linear interpolation, `TPLagrangeInterpolator`, `TPPwCubicInterpolator`, `TPPwQuadraticInterpolator`);

- `MidSets`: Functions to generate multi-index sets (tensor product, isotropic and anisotropic simplex); a class `MidSert` to grow a multi-index set starting from $\{0\}$ and adding multi-indices in the (reduced) margin;
- `SGInterpolant`: Class for the sparse grid interpolant. The user gives the tensor product interpolant and the class interpolates using the inclusion-exclusion formula;
- `compute_aposteriori_estimator`: Contain a function to compute the pointwise estimator $\zeta_{\boldsymbol{\nu},\Lambda}$ (1.44) for a downward-closed multi-index set Λ and a multi-index $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$;
- `MLInterpolant`: Class for the multilevel sparse grid interpolant as in Section 3.6.

4.3 Exemplary use

We conclude the section with a small example of use of the package and specific comments about the most important commands.

Example 4.1 (Use of `SGMethods`)**.** *The aim of this example is to interpolate the parameter-to-solution map associated to the (finite element discretization of) the affine diffusion Poisson problem: Fix $N \in \mathbb{N}$, consider $\boldsymbol{y} \in \boldsymbol{\Gamma} \coloneqq [-1,1]^N$, i.e. the parametric domain. Consider the physical domain $D = (0,1)^2 \subset \mathbb{R}^2$. The parametric affine diffusion reads $a(\boldsymbol{y},\boldsymbol{x}) = 1 + \frac{1}{C}\sum_{n=1}^{N} y_n \frac{\sin(2\pi n x_1)}{n^2}$, defined for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $\boldsymbol{x} \in D$, where $C > 0$ is chosen to make a uniformly bounded above 0. On D, we have the quasi-uniform triangular mesh $\mathcal{T}_h$ with element size $h > 0$. Denote $V_h \coloneqq \mathcal{S}_0^1(\mathcal{T}_h)$ the finite element space of piecewise-affine functions on $\mathcal{T}_h$ with 0 boundary-condition. For any $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $U(\boldsymbol{y}) \in V_h$ is the unique solution of the problem*

$$\int_D a(\boldsymbol{y},\cdot)\nabla U(\boldsymbol{y},\cdot)\cdot\nabla\phi = \int_D \phi \qquad \forall \phi \in V_h,$$

whose unique solvability is a consequence of the coercivity of $a(\boldsymbol{y},\cdot)$ and regularity of the domain D and forcing term (identically 1). The $H_0^1(D)$ norm is the natural candidate to compute the error.

In the module `sample_affine_diffusion_Poisson`, we define the following ancillary functions to sample the parameter-to-finite-element-solution map and compute the error of a parametric sample respectively. We make large use of the finite element method implemented in the Python package `Dolfin`.

```
1  from math import pi
2  import numpy as np
3  from fenics import set_log_level, DOLFIN_EPS, Constant, Expression,UnitSquareMesh, \
4  FunctionSpace, DirichletBC, TrialFunction, TestFunction, inner, grad, dx, solve, Function, norm
5
6
7  def sampleParametricPoisson(yy, nH, orderDecayCoefficients):
8      """Sample parameter-to-finite element solution map of parametric affine diffusion Poisson:
9          -\nabla \cdot (a(\by) \nabla u(\by)) = 1 on D
10                                          u = 0 on \partial D
11     Domain D is 2D unit square;
12     Diffusion a(\by) is affine in y, coercive on D
13     Parameters are in [-1,1], any length of yy is allowed
14     Mesh on D is structured with nH edges on each side
```

```
    Args:
        yy (double array): Vector of scalar parameters in [-1,1]
        nH (int): Number of mesh edges on each edge of square domain (structured mesh)
        orderDecayCoefficients (positive int): Order of decay terms affine diffusion
    Returns:
        np.array: Corrdinates of the solution (w.r.t. selected hat functions basis)"""

    assert(np.min(yy) >= -1 and np.max(yy) <= 1)
    set_log_level(31)  # reduce fenics logging
    def boundary(x):
        return x[0]<DOLFIN_EPS or x[0]>1.-DOLFIN_EPS or x[1]<DOLFIN_EPS or x[1]>1.-DOLFIN_EPS
    u0 = Constant(0.)  # boundary condition
    f = Constant(1.)  # right-hand-side
    # Diffusion coefficient
    C = pi**2/6 * 1.1  # make the diffusion uniformly positive
    strA = "1."
    for n in range(len(yy)):
        strA = strA + "+" + "sin(x[0]*2.*pi*" + str(n+1) + ")/("+str(C)+"*pow(" + str(n+1) \
         + ","+ str(orderDecayCoefficients) + "))" + "*" + str(yy[n])
    a = Expression(strA, degree=2)
    # Computational objects
    mesh = UnitSquareMesh(nH, nH)
    V = FunctionSpace(mesh, "Lagrange", 1)
    bc = DirichletBC(V, u0, boundary)
    u = TrialFunction(V)
    v = TestFunction(V)
    L = inner(a*grad(u), grad(v))*dx
    rhs = f*v*dx
    u = Function(V)
    solve(L == rhs, u, bc)  # solve linear system
    return u.vector()[:]

def computeErrorSamplesPoisson(u, uExa, nH):
    """Compute H^1_0(D) norm (see previous function) of difference of two functions
            given through their coordinates in the hat functions basis
    Args:
        u (double array): Coordinates of fucntion 1
        uExa (double array): Coordinates of Function 2
        nH (int): Number of mesh edges on each edge of square domain (structured mesh)
    Returns:
        double: error in the H^1_0(D) norm """

    assert(len(u) == len(uExa))
    mesh = UnitSquareMesh(nH, nH)
    V = FunctionSpace(mesh, "Lagrange", 1)
    errSamples = np.zeros(len(u))
    errFun = Function(V)
    for n in range(len(u)):
        errFun.vector()[:] = u[n] - uExa[n]
        errSamples[n] = norm(errFun, 'H1')
```

```
65      return errSamples
```

Then, in the main script, we define the sparse grid interpolant, interpolate the parameter-to-finite-element-solution map defined above, and measure the error in parametric and physical space. The number of parametric dimensions $N = 10$ is chosen. We consider the sparse grid interpolation method with 1D interpolation given by Lagrange interpolation over Clenshaw-Curtis nodes (1.38) *and the doubling rule* (1.35)*) as level-to-knot function. The multi-index set is chosen as $\Lambda_{\boldsymbol{\alpha}}(w)$ for $w \geq 0$ and $\boldsymbol{\alpha} \in \mathbb{R}^N_{>0}$ as in Definition 5.6, i.e. an anisotropic simplex that gives optimal dimensions-independent convergence for the present problem. The error is estimated in the $L^2(\boldsymbol{\Gamma}, H^1(D))$-norm through a Monte Carlo estimate with 128 samples (without truncation).*

Let us now read the code and give some comments:

```
1   import numpy as np
2   import matplotlib.pyplot as plt
3   from math import pi, sqrt
4   import sys, os
5   sys.path.insert(1, os.path.join(os.path.expanduser("~"), 'workspace/SGMethods'))
6   from SGMethods.ScalarNodes import CCNodes
7   from SGMethods.TPLagrangeInterpolator import TPLagrangeInterpolator
8   from SGMethods.MidSets import anisoSmolyakMidSet
9   from SGMethods.SGInterpolant import SGInterpolant
10  from tutorials.sample_affine_diffusion_Poisson import sampleParametricPoisson, \
11          computeErrorSamplesPoisson
```

Lines 6-9: Importing from `SGMethods` *respectively scalar nodes, 1D interpolant, multi-index set, and sparse grid interpolant class;*

```
12  """Tutorial on sparse grid interpolation with SGMethods.
13  We see how to use SGMethods to approximate the parametric affine diffusion Poisson problem."""
14
15  # Problem parameters
16  np.random.seed(36157)
17  N = 10  # number of parametric dimensions
18  C = pi**2/6 * 1.1  # normalization diffusion coefficient
19  a_min = 1-pi**2/(6*C)  # minimum diffusion coefficient
20  nH = 16 # mesh resolution i.e. # edges per side of square domain mesh
21  def F(y):  # target function sampler
22      return sampleParametricPoisson(y, nH=nH, orderDecayCoefficients=2)
```

Lines 20-21: Definition of the target function (i.e. the function to interpolate). `SGMethods` *accepts functions in the form used here, i.e. both input and input parameter are instance of* `1d array`*;*

```
23  # Error computation
24  yyRnd = np.random.uniform(-1, 1, (128, 1000))  # Random parameters for MC error estimation
25  uExa = np.squeeze(np.array(list(map(F, yyRnd))))  # Random sample of exact function
```

Line 24: Random sample of target function for Monte Carlo estimation of the sparse grid

error (consisting of truncation error to N dimensions and by approximation error in the first N parameters). The value 1000 gives a practically exact (non-truncated) KLE;

```
26 # Sparse grid parameters
27 TPInterpolant = lambda nodesTuple, fOnNodes : TPLagrangeInterpolator(nodesTuple, fOnNodes)
```

Line 26: Definition of the 1D interpolant. It must be given through its tensor product version on an arbitrary number of dimensions through a function with input: i. *A Python tuple with the "active" nodes (i.e. when a parameter/direction has only 1 collocation node, it can be left out and is assumed to be 0);* ii. *Array of values of function to interpolate on the nodes;*

```
28 lev2knots = lambda n: np.where(n==0, 1, 2**n+1)  # doubling rule: 1 if n=0, 2^n+1 otherwise
29 knots = lambda n : CCNodes(n)  # Clenshaw-Curtis nodes
30 # Anisotropy vector for parametric Poisson
31 gamma = lambda N : 1/(C*a_min*np.linspace(1, N, N)**2)
32 tau = lambda N : 1/(2*gamma(N))
33 anisoVec = lambda N : 0.5* np.log(1+tau(N))
34 midSet = anisoSmolyakMidSet(w=2, N=N, a=anisoVec(N))  # Anisotropic Smolyak multi-index set
```

Line 33: Definition of the multi-index set as an anisotropic simplex. We limit the dimension to $N = 10$ (line 16). However, the dimension N can be left out maintaining the rest of the signature and the function determines it automatically based on the anisotropy;

```
35 interpolant = SGInterpolant(midSet, knots, lev2knots, TPInterpolant, NParallel=8)
```

Line 34: Definition of the sparse grid interpolant instance. As explained in Section 1.2.3, a sparse grid interpolant is defined by a multi-index set, and a 1D interpolation method (in particular, nodes and level-to-knot function). Upon initialization, the class computes all relevant parameters e.g. the sparse grid $\mathcal{H} \subset \mathbf{\Gamma}$ and the inclusion-exclusion coefficients;

```
36 uOnSG = interpolant.sampleOnSG(F)  # Sample the function on the sparse grid
```

Line 35: Sampling of the target function on the sparse grid giving a handle to the parameter-to-solution map.

```
37 uInterp = interpolant.interpolate(yyRnd, uOnSG)  # Compute the interpolated value
```

Line 36: Interpolation the target function on new parameter values. Here we choose to Monte Carlo sample used to compute the error. Also the values of F on the sparse grid need to be given as an input. This increases the flexibility of the method and allows to interpolate different a quantity of interest *using the same interpolant.*

```
38 errSamples = computeErrorSamplesPoisson(uInterp, uExa, nH)  # Compute error samples in H^1_0(D)
39 errorL2 = sqrt(np.mean(np.square(errSamples)))  # L^2 error in parameter space
40 print("Error:", errorL2)
41 print("Sparse grid:\n", interpolant.SG)
42 # refine SG and compute refinement
```

```
43 midSet2 = anisoSmolyakMidSet(w=4, N=N, a=anisoVec(N))
44 interpolant2 = SGInterpolant(midSet2, knots, lev2knots, TPInterpolant=TPInterpolant, \
45         NParallel=8)
46 print("Number of collocation nodes:", interpolant.numNodes)
47 # Sample on refined sparse grid and recycle old values
48 uOnSG = interpolant.sampleOnSG(F, oldX=interpolant.SG, oldSamples=uOnSG)
49 uInterp = interpolant.interpolate(yyRnd, uOnSG,)
50 errSamples = computeErrorSamplesPoisson(uInterp, uExa, nH)
51 errorL2 = sqrt(np.mean(np.square(errSamples)))  # L^2 error in parameter space
52 print("Error:", errorL2)
53 print("Sparse grid:\n", interpolant.SG)
```

Lines 42-51: Interpolation with a refined interpolant. The multi-index set is refined by increasing w (line 42). Since the method is nested, we can recycle the previously computed interpolation samples (line 46). This is done again with the method ***interpolate*** *of the class* ***SGInterpolant****, but additionally providing as input:* i. *a list of nodes; and* ii. *a list of the corresponding parameter-to-solution map values.*

Chapter 5

Additional and partial results

In this chapter, we gather additional and partial results related to the topics discussed in the previous two chapters.

In Section 5.1, we discuss optimality and convergence with rate of the adaptive sparse grid algorithm discussed in Chapter 2. We additionally suggest and analyze possible variations of the algorithm. In Section 5.2, we present additional results regarding the space and time approximation of the stochastic Landau–Lifshitz–Gilbert equation, which we also considered in Chapter 3.

5.1 Additional results on adaptive sparse grid interpolation of random diffusion Poisson

In this section, we again adopt the notation introduced in Chapters 1 and 2 for the random diffusion Poisson problem and the adaptive sparse grid (ASG) algorithm (Algorithm 10).

The section in structured as follows: In Section 5.1.1, we discuss the quasi-optimality of the ASG algorithm and variations thereof. In Section 5.1.2, we derive a convergence with rate result for the same adaptive algorithm driven by the *simplified workless profit* (introduced in Section 5.1.1) and a sparse grid interpolant using Leja sequences. In Section 5.1.3, we present a version of the ASG algorithm with a threshold-type marking that adds multi-indices to the multi-index set more quickly to accelerate the execution. In Section 5.1.4, we define a dimension-adaptive version of the ASG algorithm that adaptively increases the number of parameters actively approximated by sparse grid interpolation. All results are supported by a number of numerical experiments, documented in Section 5.1.5. In the same section, we also numerically investigate some conjectures: The reliability of a simplified a-posteriori estimator, which is easier to compute because it is a sum over the *reduced* margin (Definition 1.19) only; An alternative estimator with improved efficiency properties; The special case of the $L^2(\mathbf{\Gamma}, H^1(D))$ error; and an alternative dimension-refinement strategy.

To prove this results, we need to study the cardinality of margins of nested multi-index sets. Observe that, for downward-closed multi-index sets $\Lambda, \Lambda' \in \mathbb{N}_0^N$, *nestedness does not imply increasing margin cardinality*, i.e.

$$\Lambda \subset \Lambda' \not\Rightarrow \#\mathcal{M}_\Lambda \leq \#\mathcal{M}_{\Lambda'}.$$

A simple counterexample in 2D is given by $\Lambda = \{(0,0),(0,1),(0,2),(1,0),(2,0)\}$ and $\Lambda' = \Lambda \cup \{(1,1)\}$. This example can be generalized to higher dimensions to show that the constant $C > 0$ such that makes the following statement true:

For any $N \in \mathbb{N}$, and any downward-closed multi-index set $\Lambda \subset \Lambda' \subset \mathbb{N}_0^N$, $\#\mathcal{M}_\Lambda \leq C \#\mathcal{M}_{\Lambda'}$

actually depends on N. The following is a positive bound of this type.

Lemma 5.1. *Let $\Lambda \subset \Lambda' \subset \mathbb{N}_0^N$ be downward-closed multi-index sets. Then, $\#\mathcal{M}_\Lambda \leq N \#\mathcal{M}_{\Lambda'}$.*

Proof. Denote $K_0 \coloneqq \mathcal{M}_\Lambda$. Define for any $n \in 1, \dots, N$

$$R_n \coloneqq \{\boldsymbol{\nu} \in K_{n-1} : \forall r \in \mathbb{N},\ \boldsymbol{\nu} + r\boldsymbol{e}_n \notin K_{n-1}\}, \qquad K_n \coloneqq K_{n-1} \setminus R_n.$$

The statement follows from : *1.* $\mathcal{M}_\Lambda = \bigcup_{n=1}^N R_n$; and *2.* $\#R_n \leq \#\mathcal{M}_{\Lambda'}$ for all $n = 1, \dots, N$. Let us prove them.

Proof of 1. While $\bigcup_{n=1}^N R_n \subset \mathcal{M}_\Lambda$ is trivial, the converse requires more work. Suppose by contradiction that $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$ but $\boldsymbol{\nu} \notin R_n$ for any $n = 1, \dots, N$. Then $\boldsymbol{\nu} \in \bigcap_{n=0}^N K_n$. Thus, $\boldsymbol{\nu} \notin R_N$ implies that there exists $r_N \in \mathbb{N}$ such that $\boldsymbol{\nu}^{(N)} \coloneqq \boldsymbol{\nu} + r\boldsymbol{e}_N \in K_{N-1}$. Since $\boldsymbol{\nu}^{(N)} \notin R_{N-1}$ by definition, we may repeat this argument and obtain: There exists $r_n \in \mathbb{N}$ such that $\boldsymbol{\nu}^{(n)} \coloneqq \boldsymbol{\nu}^{(n+1)} + r_n \boldsymbol{e}_n \in K_n$ for all $n = N, \dots, 1$. All in all, we have proved $\boldsymbol{\nu} < \boldsymbol{\nu}^{(1)} = \boldsymbol{\nu} + \sum_{n=1}^N r_n \boldsymbol{e}_n \in K_0 = \mathcal{M}_\Lambda$. This is a contradiction to $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$.

Proof of 2. Fix $n = 1, \dots, N$. The statement is equivalent to the injectivity of the function $f : R_n \to \mathcal{M}_{\Lambda'}$ defined for any $\boldsymbol{\nu} \in R_n$ as $f(\boldsymbol{\nu}) \coloneqq \boldsymbol{\nu} + \min\{t \in \mathbb{N} : \boldsymbol{\nu} + t\boldsymbol{e}_n \in \mathcal{M}_{\Lambda'}\}\boldsymbol{e}_n$. Suppose that for $\boldsymbol{\nu}, \boldsymbol{\nu}' \in R_n$, $f(\boldsymbol{\nu}) = f(\boldsymbol{\nu}')$. Thus, for $t, t' > 0$, $\boldsymbol{\nu} + t\boldsymbol{e}_n = \boldsymbol{\nu}' + t'\boldsymbol{e}_n$ or, rearranging the terms, $\boldsymbol{\nu} = \boldsymbol{\nu}' + (t' - t)\boldsymbol{e}_n$. Because of the definition of R_n, $t - t' = 0$ and so $\boldsymbol{\nu} = \boldsymbol{\nu}'$.

□

5.1.1 Quasi-optimality

We define notions of (quasi)-optimality for the approximation of the parametric Poisson Problem with sparse grid and prove them for the adaptive sparse grid algorithm (Algorithm 10). Recall that the optimality of the finite element adaptive approximation was proved in Theorem 2.24.

Setting 1: Workless profit

We show that the adaptive parametric refinement Algorithm 10 with workless profit (2.9) generates a sparse grid approximation that is, in a sense, quasi-optimal.

Proposition 5.2. *Consider for $N \in \mathbb{N}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^N$ a rectangular multi-index set $\Lambda = \mathrm{Rect}_{\boldsymbol{\nu}} \subset \mathbb{N}_0^N$ (2.17). Let $\varepsilon > 0$ such that $\zeta_{SC,\Lambda} = \frac{\varepsilon}{N}$. Denote by Λ_ε the multi-index set obtained running Algorithm 10 with workless profit* (2.9) *and tolerance ε. It holds that $\Lambda_\varepsilon \subset \Lambda$.*

Proof. By contradiction, consider $\ell > 0$ such that $\Lambda_\ell \subset \Lambda$, the profit-maximizing multi-index $\boldsymbol{\nu}_\ell = \arg\max_{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}} \mathcal{P}_{\boldsymbol{\nu},\Lambda_\ell}$ belongs to the margin of Λ, and $\zeta_{SC,\Lambda_\ell} \geq \varepsilon$. Since both Λ and Λ_ℓ are rectangles (for Λ_ℓ this fact was proved in Section 2.2.1), $A_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \subset \mathcal{M}_\Lambda$, where we recall that $\Lambda_{\ell+1} = \Lambda_\ell \cup A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}$. The contradiction is then found by estimating

$$\varepsilon \leq \zeta_{SC,\Lambda_\ell} \leq N \sum_{\boldsymbol{j} \in A_{\boldsymbol{\nu}_\ell,\Lambda_\ell}} \zeta_{\boldsymbol{j},\Lambda_\ell} = N \sum_{\boldsymbol{j} \in A_{\boldsymbol{\nu}_\ell,\Lambda}} \zeta_{\boldsymbol{j},\Lambda} \leq N \zeta_{SC,\Lambda},$$

where the first estimate is from the contradiction assumption, the second is a consequence of the definition of workless profit (2.9), and the equality comes from the formula for the simplified computation of the pointwise error estimator (see Remark 2.4), since $J_{\boldsymbol{j},\Lambda_\ell} = J_{\boldsymbol{j},\Lambda}$. □

Remark 5.3. *We make some observations regarding the previous result.*

1. *In the previous proposition, quasi-optimality is understood with respect to* rectangular *multi-index sets ($\Lambda = Rect_{\boldsymbol{\nu}}$ for $\boldsymbol{\nu} \in \mathbb{N}_0^N$) only, and not with respect to general downward-closed multi-index sets. The reason is that Algorithm 13 driven by workless profit can only generate rectangular multi-index sets itself (Lemma 2.10). It is therefore natural to have an optimality result with respect to this family of approximations only;*

2. *The result is a useful statement about the optimality of the a-posteriori estimator only is $\varepsilon \ll N^{-1}$. In other words, it can be understood as an* asymptotic *optimality statement under the finite-dimensional noise assumption (Remark 1.7). Conversely, the result is of little use when dimension-adaptivity is considered and the function has weak sparsity, i.e. the number of dimensions $N \in \mathbb{N}$ is also determined adaptively (see Section 5.1.4 below) and it grows rapidly. In this case, ε may not be smaller than N^{-1}.*

Setting 2: A simplified workless profit

Consider the following *simplified workless profit*:

$$\mathcal{P}_{\boldsymbol{\nu},\Lambda} \coloneqq \zeta_{\boldsymbol{\nu},\Lambda} \qquad \forall \boldsymbol{\nu} \in \mathcal{M}_\Lambda, \tag{5.1}$$

and use it to drive Algorithm 10. To state the next proposition, we need the following

Definition 5.4. *Given $N \in \mathbb{N}$, a multi-index $\boldsymbol{\nu} \in \mathbb{N}_0^N$, and a downward-closed multi-index set $\Lambda \subset \mathbb{N}_0^N$, we call* strengthened pointwise estimator *and* strengthened sparse grid a-posteriori estimator *respectively:*

$$\tilde{\zeta}_{\boldsymbol{\nu}} \coloneqq \max\left\{\zeta_{\boldsymbol{\nu},J} : J \subset \mathbb{N}_0^N \text{ downward such that } \boldsymbol{\nu} \in \mathcal{M}_J\right\}, \qquad \tilde{\zeta}_{SC,\Lambda} \coloneqq \sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \tilde{\zeta}_{\boldsymbol{\nu}}.$$

Proposition 5.5. *Consider an arbitrary downward-closed multi-index set $\Lambda \subset \mathbb{N}_0^N$. Let $\ell \in \mathbb{N}_0$ be the unique iteration of Algorithm 10 such that $\Lambda_\ell \subset \Lambda$ and $\boldsymbol{\nu}_\ell \notin \Lambda$. Then,*

$$\zeta_{SC,\Lambda_\ell} \leq N \# \mathcal{M}_\Lambda \tilde{\zeta}_{SC,\Lambda},$$

where $\tilde{\zeta}_{SC,\Lambda}$ is the strengthened a-posteriori estimator (Definition 5.4).

Proof. Since $\boldsymbol{\nu}_\ell \notin \Lambda$ and $\Lambda_\ell \subset \Lambda$, $\boldsymbol{\nu}_\ell \in \mathcal{M}_\Lambda$. Then, we estimate

$$\zeta_{SC,\Lambda_\ell} \leq \#\mathcal{M}_{\Lambda_\ell} \zeta_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \leq \#\mathcal{M}_{\Lambda_\ell} \tilde{\zeta}_{\boldsymbol{\nu}_\ell} \leq \#\mathcal{M}_{\Lambda_\ell} \tilde{\zeta}_{SC,\Lambda} \leq N \#\mathcal{M}_\Lambda \tilde{\zeta}_{SC,\Lambda}.$$

In the first inequality we used the fact that $\boldsymbol{\nu}_\ell$ is the profit maximizer on $\mathcal{M}_{\Lambda_\ell}$, in the second the definition of $\tilde{\zeta}_{\boldsymbol{\nu},\Lambda}$, in the third the definition of $\tilde{\zeta}_{SC,\Lambda}$, and in the last one we used Lemma 5.1. □

Let us finally recall that the the *cost* of adding a multi-index can be included in the profit definition through a *simplified profit with work* inspired by the equivalence of the multi-index set selection and the Knapsack problem (cf. Section 1.2.3):

$$\mathcal{P}_{\boldsymbol{\nu},\Lambda} = \frac{\zeta_{\boldsymbol{\nu},\Lambda}}{W_{\boldsymbol{\nu}}} \qquad \forall \boldsymbol{\nu} \in \mathcal{M}_\Lambda, \tag{5.2}$$

where the *work* $W_{\boldsymbol{\nu}}$ denotes the number of collocation points added to the sparse grid $\mathcal{H}_\Lambda$ when $\boldsymbol{\nu}$ is added to the multi-index set Λ. Here, we do not analyze the case of a general work definition. We only observe that when $W_{\boldsymbol{\nu}} = 1$ for all $\boldsymbol{\nu} \in \mathbb{N}_0^N$, this profit coincides with the simplified workless profit (5.1).

5.1.2 Convergence with rate

We now want to estimate ζ_{SC,Λ_ℓ} with respect to $\#\mathcal{H}_{\Lambda_\ell}$, i.e. the number of collocation points at step $\ell \in \mathbb{N}_0$. This is a better quantitative measure of the convergence speed of the ASG algorithm compared to the estimates from Theorem 2.18, 2.19. We work with Leja sequences (see also Example 1.17), which we recall are defined as

$$\mathcal{Y}^1 = \{y_1\}\,, \text{where } y_1 \in [-1,1] \text{ arbitrary}, \tag{5.3}$$

$$\mathcal{Y}^m = \mathcal{Y}^{m-1} \cup \left\{ \underset{y\in[-1,1]}{\arg\max} \prod_{z\in\mathcal{Y}^{m-1}} |y-z| \right\} \quad \forall m \in \mathbb{N}, \tag{5.4}$$

and the level-to-knot function:

$$m(\nu) = \nu + 1 \qquad \forall \nu \in \mathbb{N}_0. \tag{5.5}$$

Leja sequences have properties that make them popular interpolation nodes:

- The sparse grid and multi-index set have the same cardinality, i.e. $\#\mathcal{H}_\Lambda = \#\Lambda$. As a consequence, the simplified workless profit (5.1) coincides with the simplified profit with work (5.2);
- Their Lebesgue constant (1.34) grows sub-exponentially [TT08]: $\lim_{\nu\to\infty} (\lambda_\nu)^{1/(\nu+1)} = 0$.

Proposition 5.5 is not directly applicable here. However, it serves as a guideline for the proof of Proposition 5.8. Consider the following family of multi-index sets, which in the next proposition play the same role of $\Lambda \subset \mathbb{N}_0^N$ in Proposition 5.5.

Definition 5.6 (Anisotropic simplex multi-index set)**.** *Let $N \in \mathbb{N}$, $w \geq 0$ and a non-decreasing sequence of positive real numbers $\boldsymbol{\alpha} = (\alpha_n)_{n=1}^N$. The* anisotropic simplex multi-index set *is by definition*

$$\Lambda_{\boldsymbol{\alpha}}(w) = \Lambda_{\boldsymbol{\alpha},N}(w) \coloneqq \left\{ \boldsymbol{\nu} \in \mathbb{N}_0^N : \sum_{n=1}^N \nu_n \alpha_n \leq w \right\}.$$

In order to choose $\boldsymbol{\alpha}$ appropriately for the affine random diffusion Poisson problem, recall the following regularity fact:

Lemma 5.7 (From [BNT10])**.** *If for all $n \in 1,\dots,N$ it exists $\gamma_n \in \mathbb{R}_{>0}$ such that for all $\boldsymbol{y} \in \boldsymbol{\Gamma} = [-1,1]^N$, $\left\| \frac{\partial_{y_n} a(\boldsymbol{y})}{a(\boldsymbol{y})} \right\|_{L^\infty(D)} \leq \gamma_n$, then $y_n \mapsto u(\boldsymbol{y})$, the parameter-to-solution map of the parametric affine diffusion Poisson problem (cf. Example 1.8), admits a bounded holomorphic extension to the complex neighbourhood $\Sigma(\Gamma_n, \tau_n) \coloneqq \{z \in \mathbb{C} : \operatorname{dist}(\Gamma_n, z) < \tau_n\}$, where $\tau_n \leq \frac{1}{2\gamma_n}$.*

We use the result to define the vector $\boldsymbol{g} = (g_n)_{n=1}^N$ as

$$g_n \coloneqq \frac{1}{2}\log(\tau_n + \sqrt{1+\tau_n^2}) \qquad \forall n = 1,\dots,N. \tag{5.6}$$

The strengthened pointwise estimator (Definition 5.4) can be estimated a priori as done in Proposition 2.5. We get

$$\tilde{\zeta}_{\boldsymbol{\nu}} \leq C(u,a) N \lambda_{\boldsymbol{\nu}}^2 e^{-\boldsymbol{g}\cdot\boldsymbol{\nu}} \qquad \forall \boldsymbol{\nu} \in \mathbb{N}_0^N, \tag{5.7}$$

where

$$C(u,a) \coloneqq a_{max} \prod_{n=1}^{N} \frac{4}{e^{g_n}-1} \max_{\boldsymbol{z}\in\Sigma(\boldsymbol{\Gamma},\boldsymbol{\tau})} \|u(\boldsymbol{z})\|_V\,, \tag{5.8}$$

$\lambda_{\boldsymbol{\nu}} \coloneqq \prod_{n=1}^{N} 2\lambda_{\nu_n}$, and λ_{ν_n} denotes the Lebesgue constant of the 1D interpolant of index $\nu \in \mathbb{N}_0$. The sub-exponential growth of the Lebesgue constant of Leja nodes (see above) implies:

$$\forall \varepsilon \in (0,1)\ \exists C(\varepsilon) > 0 : \forall \boldsymbol{\nu} \in \mathbb{N}_0^N\ \tilde{\zeta}_{\boldsymbol{\nu}} \leq C(u,a,\varepsilon) N e^{-\varepsilon \boldsymbol{g}\cdot\boldsymbol{\nu}}, \tag{5.9}$$

where $C(u,a,\varepsilon) = C(u,a)C(\varepsilon)$ and $C(\varepsilon) > 0$ is independent of N. Finally, choose the anisotropy vector $\boldsymbol{\alpha}$ as

$$\alpha_n \coloneqq \varepsilon g_n \qquad \forall n = 1,\dots,N. \tag{5.10}$$

Let us now use Lemma 5.1 to obtain a useful a priori bound for the a-posteriori estimator of Λ_ℓ for $\ell \in \mathbb{N}_0$.

Proposition 5.8. *Let $w \geq 0$, $\ell \in \mathbb{N}_0$ such that the multi-index set produced by Algorithm 10 at step ℓ, $\Lambda_\ell \subset \Lambda_{\boldsymbol{\alpha}}(w)$ and the simplified workless profit* (5.1) *maximizer $\boldsymbol{\nu}_\ell \in \mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)}$. Then,*

$$\zeta_{SC,\Lambda_\ell} \leq C(u,a,\varepsilon) N^2 \sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}},$$

where $C(u,a,\varepsilon) = a_{max} C(\varepsilon) \prod_{n=1}^{N} \frac{4}{e^{\sigma}-1} \max_{\boldsymbol{z}\in\boldsymbol{\Sigma}(\boldsymbol{\Gamma},\boldsymbol{\tau})} \|u(\boldsymbol{z})\|_V$, and $C(\varepsilon)$ was defined in (5.9).

Proof. Fix $w \geq 0, \ell \in \mathbb{N}_0$ as in the statement. Estimate

$$\zeta_{SC,\Lambda_\ell} \leq \#\mathcal{M}_{\Lambda_\ell} \tilde{\zeta}_{\boldsymbol{\nu}_\ell} \leq\ N \#\mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)} C(u,a,\varepsilon) N e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}_\ell} \leq C(u,a,\varepsilon) N^2 e^{\alpha_N} \sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}}.$$

The first estimate is based on the fact that $\boldsymbol{\nu}_\ell$ is the simplified workless profit maximizer. In the second estimate, we use Lemma 5.1 to estimate the margin and (5.9) to estimate the pointwise estimator recalling that $\boldsymbol{\alpha}$ is as in (5.10). To justify the last estimate, first observe that for any $\boldsymbol{j} \in \mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)}$, either $\boldsymbol{j} - e_N \in \Lambda_{\boldsymbol{\alpha}}(w)$ or $j_N = 0$. Thus, by definition of $\Lambda_{\boldsymbol{\alpha}}(w)$ (Definition 5.6), $\boldsymbol{\alpha}\cdot\boldsymbol{j} - \alpha_N \leq w \leq \boldsymbol{\alpha}\cdot\boldsymbol{j}$. Being in particular $\boldsymbol{\nu}_\ell \in \mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)}$, we have that for all $\boldsymbol{j} \in \mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)}$, $e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}_\ell} \leq e^{-w} \leq e^{-\boldsymbol{\alpha}\cdot\boldsymbol{j}+\alpha_N}$, which proves the last estimate in the chain above. □

The series $\sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}}$ appearing in the right-hand side of the estimate was studied in [GO16, Corollary 2.7]. If $w > N$, it holds that

$$\sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}} \leq N \prod_{j=1}^{N} \frac{e^{\alpha_j}}{e^{\alpha_j}-1} e^{-w} \frac{w^{N-1}}{(N-1)!} \leq D(N,\boldsymbol{\alpha}) e^{-w} w^{N-1},$$

where $D(\boldsymbol{\alpha},N)$ is a constant depending only on N, $\boldsymbol{\alpha}$. The analogous estimate with respect to $\#\Lambda_{\boldsymbol{\alpha}}(w)$ is also proved in the same reference:

$$\sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}} \leq E(N,\boldsymbol{\alpha}) \exp\left(-\frac{N}{e}\mathrm{gm}(\boldsymbol{\alpha})\left(\#\Lambda_{\boldsymbol{\alpha}}(w)\right)^{\frac{1}{N}}\right), \tag{5.11}$$

where $\mathrm{gm}(\boldsymbol{\alpha}) \coloneqq \left(\prod_{n=1}^{N} \alpha_n\right)^{\frac{1}{N}}$ and $E(\boldsymbol{\alpha},N)$ is again a constant depending only on N, $\boldsymbol{\alpha}$.

This finally leads to a convergence with rate result for our adaptive algorithm:

Corollary 5.9. *Consider sparse grid interpolation with Leja sequences* (5.3) *and the affine level-to-knot function* (5.5)*, so that* $\#\Lambda = \#\mathcal{H}_\Lambda$ *for any downward-closed multi-index set* $\Lambda \subset \mathbb{N}_0^N$*. Denote* $(\Lambda_\ell)_{\ell\in\mathbb{N}_0}$ *the nested sequence of downward-closed multi-index sets produced by Algorithm 10 with simplified workless profit* (5.1)*. The a-posteriori estimator at step* $\ell \in \mathbb{N}_0$ *admits the following estimate:*

$$\zeta_{SC,\Lambda_\ell} \leq C'(u,a,\varepsilon,N)N^2 \exp\left(-\frac{N}{e} gm(\boldsymbol{\alpha})\,(\#\Lambda_\ell)^{\frac{1}{N}}\right),$$

where $C'(u,a,\epsilon,N) = C(u,a,\varepsilon)E(N,\boldsymbol{\alpha})$*,* $C(u,a,\varepsilon,\boldsymbol{\alpha})$ *was defined in Proposition 5.8, and* $E(N,\boldsymbol{\alpha})$ *was defined just above.*

5.1.3 Threshold-type marking of the margin

While the sparse grid interpolant produced by the adaptive algorithms described above may have quasi-optimality properties (see the discussion in Section 5.1.1), the adaptive algorithm as a whole may take a lot of time due to the small difference between subsequent multi-index sets. A possible solution is to introduce a threshold-type marking on the margin, tuned by a parameter $\theta \in (0,1)$. We adopt again Leja sequences (5.3) and the level-to-knot function (5.5), as well as the simplified workless estimator (5.1) (in this case, it coincides with the simplified profit with work (5.2)).

Algorithm 12 $u_\varepsilon \leftarrow TASG(\varepsilon,\theta)$

$\Lambda_0 := \{\mathbf{0}\}$
Compute $\mathcal{I}_{\Lambda_\ell}[u]$
for $\ell = 0,1,\dots$ **do**
  Compute ζ_{SC,Λ_ℓ}
  if $a_{\min}^{-1}\zeta_{SC,\Lambda_\ell} < \varepsilon$ **then**
    Return $u_\varepsilon \leftarrow \mathcal{I}_{\Lambda_\ell}[u]$
  end if
  $\boldsymbol{\nu}_\ell \leftarrow \arg\max_{\boldsymbol{\nu}\in\mathcal{M}_{\Lambda_\ell}} \zeta_{\boldsymbol{\nu},\Lambda_\ell}$
  Find $K_\ell := \{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell} : \zeta_{\boldsymbol{\nu},\Lambda_\ell} \geq \theta\zeta_{\boldsymbol{\nu}_\ell,\Lambda_\ell}\}$
  $\Lambda_{\ell+1} \leftarrow \Lambda_\ell \cup K_\ell$
  Make $\Lambda_{\ell+1}$ downward-closed
end for

A quasi-optimality analogous to Proposition 5.5 holds:

Proposition 5.10. *Consider an arbitrary downward-closed multi-index set* $\Lambda \subset \mathbb{N}_0^N$*. Let* $\ell \in \mathbb{N}_0$ *be the unique iteration of Algorithm 12 with marking parameter* $\theta \in (0,1)$ *such that* $\Lambda_\ell \subset \Lambda$ *and* $\Lambda_{\ell+1} \not\subset \Lambda$*. Then,*

$$\zeta_{SC,\Lambda_\ell} \leq \theta^{-1}N\#\mathcal{M}_\Lambda\tilde{\zeta}_{SC,\Lambda},$$

where $\tilde{\zeta}_{SC,\Lambda}$ *denotes the strengthened sparse grid estimator (Definition 5.4).*

Proof. As $\Lambda_{\ell+1} \not\subset \Lambda$, it exists $\boldsymbol{j} \in K_\ell$ such that $\boldsymbol{j} \in \mathcal{M}_\Lambda$. We then proceed as in Proposition 5.5:

$$\zeta_{SC,\Lambda_\ell} \leq \#\mathcal{M}_{\Lambda_\ell}\zeta_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \leq \#\mathcal{M}_{\Lambda_\ell}\theta^{-1}\zeta_{\boldsymbol{j},\Lambda_\ell} \leq \theta^{-1}N\#\mathcal{M}_\Lambda\tilde{\zeta}_{SC,\Lambda},$$

where in the first estimate we use the fact that $\boldsymbol{\nu}_\ell$ is the simplified workless profit maximizer, in the second we use the definition of K_ℓ, in the third the definition of strengthened a-posteriori estimator and Lemma 5.1. □

As seen in Proposition 5.8, Algorithm 10 produces a sparse grid interpolant with exponential accuracy with respect to its number of collocation points. We now prove the same of Algorithm 12.

Proposition 5.11. *For all $w \geq 0$, there exists $\ell \in \mathbb{N}_0$ such that Λ_ℓ, the multi-index set produced by Algorithm 12 with marking parameter $\theta \in (0,1)$ at step ℓ, satisfies*

$$\zeta_{SC,\Lambda_\ell} \leq \theta^{-1} C(u,a,\varepsilon) N^2 \sum_{\boldsymbol{\nu} \notin \Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}},$$

where $C(u,a,\varepsilon)$ was defined in Proposition 5.8.

Proof. Fix $w \geq 0$. It exists $\ell \in \mathbb{N}_0$ such that $\Lambda_\ell \subset \Lambda_{\boldsymbol{\alpha}}(w)$ and $K_\ell \cap \mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)} \neq \emptyset$. Observe that, for $\boldsymbol{j} \in K_\ell \cap \mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)}$,

$$\zeta_{SC,\Lambda_\ell} \leq \#\mathcal{M}_{\Lambda_\ell} \zeta_{\boldsymbol{\nu}_\ell,\Lambda_\ell} \leq \theta^{-1} \#\mathcal{M}_{\Lambda_\ell} \zeta_{\boldsymbol{j},\Lambda_\ell} \leq \theta^{-1} \#\mathcal{M}_{\Lambda_\ell} \tilde{\zeta}_{\boldsymbol{j}}.$$

It is now sufficient to proceed as in Proposition 5.8, using $\boldsymbol{j}$ in place of $\boldsymbol{\nu}_\ell$. □

As a consequence, also a statement analogous to Corollary 5.9 is valid for the current algorithm:

Corollary 5.12. *The downward-closed multi-index set $\Lambda_\ell \subset \mathbb{N}_0^N$ produced by Algorithm 12 with marking parameter $\theta \in (0,1)$ at step $\ell \in \mathbb{N}_0$ is such that*

$$\zeta_{SC,\Lambda_\ell} \leq \theta^{-1} C'(u,a,\varepsilon,N) \exp\left(-\frac{N}{e} gm(\boldsymbol{\alpha}) \left(\#\Lambda_\ell\right)^{\frac{1}{N}}\right),$$

where $C(u,a,\varepsilon,N)$ was defined in Corollary 5.9.

5.1.4 Dimension-adaptive parametric approximation

Rather than working under the finite dimensional noise assumption (Remark 1.7), we can adaptively determine the number of parameters approximated by the sparse grid interpolant.

Recall the definition of the finite-support multi-index set $\mathcal{F}$ (1.22) and of support of a multi-index $\operatorname{supp}(\boldsymbol{\nu})$ (1.23). Also recall that $\mathcal{F}$ is countable.

For a multi-index with finite support $\Lambda \subset \mathcal{F}$, define its *support size* by

$$N_\Lambda \coloneqq \max\{N \in \mathbb{N} : \exists \boldsymbol{\nu} \in \Lambda : \nu_N \neq 0\}.$$

As proved in [GN18, Proposition 7.1], if $\dim(\boldsymbol{\Gamma}) > N_\Lambda$, the a-posteriori estimate must be modified as

$$\|u - \mathcal{I}_\Lambda[u]\|_{L^\infty(\boldsymbol{\Gamma}, H_0^1(D))} \lesssim \zeta_{SC,\Lambda} + \zeta_{TR,\Lambda},$$

where the *dimension truncation error estimator* reads:

$$\zeta_{TR,\Lambda} \coloneqq \|(a - a_{N_\Lambda}) \nabla \mathcal{I}_\Lambda[u]\|_{L^\infty(\boldsymbol{\Gamma}, L^2(D))}. \tag{5.12}$$

To define a dimension-adaptive algorithm, we add one parametric dimension at the end of each iteration of Algorithm 12 if a marked multi-index $\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}$ has nonzero last component $\nu_{N_\ell} > 0$. This way, we obtain an increasing sequence $(N_\ell)_{\ell \in \mathbb{N}_0} \subset \mathbb{N}_0$ of the number of approximated scalar parameters.

The definition of margin (Definition 1.19) does not generalize in a a useful way to the present setting. Consider instead:

$$\begin{aligned} \mathcal{M}_\Lambda^{(Q)} &\coloneqq \{\boldsymbol{\nu} \in \mathcal{F} \setminus \Lambda : \exists n \in 1, \dots, Q : \text{ either } \boldsymbol{\nu} - \boldsymbol{e}_n \in \Lambda \text{ or } \nu_n = 0\} \qquad \forall Q \in \mathbb{N}, \\ \mathcal{M}_\Lambda &\coloneqq \mathcal{M}_\Lambda^{(N_\Lambda)}. \end{aligned} \tag{5.13}$$

The *Threshold Dimension Adaptive Sparse Grid* algorithm (TDASG) with Leja sequences, simplified pointwise estimator, threshold-type marking (as in in the previous section), and the new dimension-adaptivity reads:

Algorithm 13 $u_\varepsilon \leftarrow TDASG(\varepsilon, \theta)$

$N_0 \leftarrow 1$
$\Lambda_0 \leftarrow \{0\}$
Compute $\mathcal{I}_{\Lambda_\ell}[u]$
for $\ell = 0, 1, \dots$ **do**
  Compute $(\zeta_{\boldsymbol{\nu},\Lambda_\ell})_{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}}, \zeta_{SC,\Lambda_\ell}, \zeta_{TR,\Lambda_\ell}$
  if $a_{\min}^{-1}(\zeta_{SC,\Lambda_\ell} + \zeta_{TR,\Lambda_\ell})$ **then**
    Return $\mathcal{I}_{\Lambda_\ell}[u]$
  end if
  $\boldsymbol{\nu}_\ell \coloneqq \arg\max_{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell}} \zeta_{\boldsymbol{\nu},\Lambda_\ell}$
  Find $K_\ell \coloneqq \{\boldsymbol{\nu} \in \mathcal{M}_{\Lambda_\ell} : \zeta_{\boldsymbol{\nu},\Lambda_\ell} \geq \theta \zeta_{\boldsymbol{\nu}_\ell,\Lambda_\ell}\}$
  $\Lambda_{\ell+1} \leftarrow \Lambda_\ell \cup K_\ell$
  Make $\Lambda_{\ell+1}$ downward-closed
  if $\exists \boldsymbol{\nu} \in K : \nu_{N_\ell} > 0$ **then**
    $N_{\ell+1} \leftarrow N_\ell + 1$
    Update Λ_ℓ and $\mathcal{M}_{\Lambda_\ell}$
  else
    $N_{\ell+1} \leftarrow N_\ell$
  end if
end for

We consider $\boldsymbol{g} = (g_n)_{n \in \mathbb{N}} \in \mathbb{R}_{>0}^{\mathbb{N}}$ with g_n defined as in (5.6). In particular, $\lim_{n\to\infty} g_n = \infty$. For $\Lambda \subset \mathcal{F}$ downward-closed and $\boldsymbol{\nu} \in \mathcal{M}_\Lambda$, we can repeat verbatim the proof of Proposition 2.5 to obtain, now in a general number of parametric dimensions,

$$\zeta_{\boldsymbol{\nu},\Lambda} \lesssim \#\mathrm{supp}\,(\boldsymbol{\nu})\, \lambda_{\boldsymbol{\nu}}^2 e^{-\boldsymbol{g}\cdot\boldsymbol{\nu}}.$$

Additionally, the sub-exponential growth of the Lebesgue constant for Leja nodes implies that for all $\varepsilon \in (0,1)$, it exists $C(\varepsilon) > 0$ such that

$$\zeta_{\boldsymbol{\nu},\Lambda} \lesssim \#\mathrm{supp}\,(\boldsymbol{\nu})\, C(\varepsilon) e^{-\varepsilon \boldsymbol{g}\cdot\boldsymbol{\nu}}. \tag{5.14}$$

Since $g_n \to \infty$, $C(\varepsilon)$ is independent of $\boldsymbol{\nu}$ or Λ. For a fixed $\varepsilon \in (0,1)$, define $\boldsymbol{\alpha} = \varepsilon \boldsymbol{g} \in \mathbb{R}_{>0}^{\mathbb{N}}$ as in (5.10).

The following lemma is a direct consequence of Lemma 5.1 (first estimate) and of the definition of margin (5.13, second estimate).

Lemma 5.13. *Consider $\Lambda, J \subset \mathcal{F}$ downward-closed such that $\Lambda \subset J$. Then,*

$$\#\mathcal{M}_\Lambda \leq N_\Lambda \#\mathcal{M}_J^{(N_\Lambda)} \leq N_\Lambda \#\mathcal{M}_J^{(N_J)}.$$

We define the a-posteriori estimator in the infinite-dimensional parameters setting using the definition of $\mathcal{M}_\Lambda$ (5.13):

$$\zeta_{SC,\Lambda} \coloneqq \sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \zeta_{\boldsymbol{\nu},\Lambda}. \tag{5.15}$$

Proposition 5.14. *Fix $w \geq 0$. There exists $\ell \in \mathbb{N}_0$ such that the multi-index set Λ_ℓ produced by Algorithm 13 at step $\ell \in \mathbb{N}_0$ satisfies*

$$\Lambda_\ell \subset \Lambda_{\boldsymbol{\alpha}}(w), \qquad \zeta_{SC,\Lambda_\ell} \leq C(\varepsilon) N_\ell^2 \, e^{\alpha_{N_\ell}} \sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}},$$

where $C(\varepsilon) > 0$ depends only on ε, used to define $\boldsymbol{\alpha}$ as in (5.10).

Proof. Fix $w \geq 0$. Then, it exists $\ell \in \mathbb{N}_0$ such that $\Lambda_\ell \subset \Lambda_{\boldsymbol{\alpha}}(w)$ and $K_\ell \not\subset \Lambda_{\boldsymbol{\alpha}}(w)$. The parametric estimator ζ_{SC,Λ_ℓ} (5.15) is estimated using the definition of profit-maximizer $\boldsymbol{\nu}_\ell$ and θ-marking. There exists $\boldsymbol{j} \in K_\ell \cap \mathcal{M}^{(N_\ell)}_{\Lambda_{\boldsymbol{\alpha}}(w)}$ such that $\zeta_{SC,\Lambda_\ell} \leq \theta^{-1} \# \mathcal{M}_{\Lambda_\ell} \zeta_{\boldsymbol{j},\Lambda_\ell}$. Estimate the pointwise estimator $\zeta_{\boldsymbol{j},\Lambda_\ell}$ with (5.14) and the margin cardinality with Lemma 5.13. We get

$$\zeta_{SC,\Lambda_\ell} \leq N_\ell \# \mathcal{M}^{(N_\ell)}_{\Lambda_{\boldsymbol{\alpha}}(w)} \theta^{-1} C(\varepsilon) \# \operatorname{supp}(\boldsymbol{\nu}_\ell) \, e^{-\boldsymbol{\alpha}\cdot\boldsymbol{j}}. \tag{5.16}$$

Observe that $\#\operatorname{supp}(\boldsymbol{\nu}_\ell) \leq N_\ell$, and that, as in Proposition 5.8, $e^{-\boldsymbol{j}\cdot\boldsymbol{\alpha}} < e^{-w} \leq e^{-\boldsymbol{\nu}\cdot\boldsymbol{\alpha}} e^{\alpha_N}$ for any $\boldsymbol{\nu} \in \mathcal{M}^{(N_\ell)}_{\Lambda_{\boldsymbol{\alpha}}(w)}$. This implies

$$\# \mathcal{M}^{(N_\ell)}_{\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{j}} \leq e^{\alpha_{N_\ell}} \sum_{\boldsymbol{\nu}\in\mathcal{M}^{(N_\ell)}_{\Lambda_{\boldsymbol{\alpha}}(w)}} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}}.$$

Trivially, we can substitute $\mathcal{M}_{\Lambda_{\boldsymbol{\alpha}}(w)}$ with the complementary of $\Lambda_{\boldsymbol{\alpha}}(w)$ in the sum. Together with (5.16), these two facts prove the statement. □

To find a convergence result with respect to the number of collocation points $\#\mathcal{H}_{\Lambda_\ell}$, we apply the estimate from [GO16, Theorem 3.2] to the sum in the right-hand side. We obtain

Corollary 5.15. *If there exists $\beta > 0$ such that*

$$M(\boldsymbol{\alpha}, \beta) \coloneqq \sum_{n\in\mathbb{N}} \frac{1}{e^{\alpha_n/\beta} - 1} < \infty,$$

then the TDASG algorithm (Algorithm 13) satisfies

$$\sum_{\boldsymbol{\nu}\notin\Lambda_{\boldsymbol{\alpha}}(w)} e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\nu}} \leq \beta^{-1} e^{\beta M(\boldsymbol{\alpha},\beta)} \left(\#\mathcal{H}_{\Lambda_\ell}\right)^{-(\beta-1)}.$$

However, in this case the factor $N_\ell^2 \, e^{\alpha_{N_\ell}}$ in the estimate obtain in Proposition (5.14) depends on ℓ. In order to have optimal convergence (at least in the limit) this term has to grow at most logarithmically in $\#\mathcal{H}_{\Lambda_\ell}$. However, numerical experiments (see Figure 5.3) suggest that $N_\ell = \mathcal{O}\left(\#\mathcal{H}_{\Lambda_\ell}\right)$.

5.1.5 Numerical experiments

We carry out a number of numerical experiments to confirm the statements made above. In the following experiments, we always consider a parametric affine diffusion coefficient

$$a(\boldsymbol{y},\boldsymbol{x}) = a_0 + \frac{1}{\nu}\sum_{n=1}^{\infty} y_n \lambda_n \psi_n(\boldsymbol{x}) \qquad \forall \boldsymbol{y} \in [-1,1]^{\mathbb{N}},\ \boldsymbol{x} \in D, \tag{5.17}$$

where $a_0 > 0$ is a constant (a function of $\boldsymbol{x} \in D$ is also possible), $\psi_n(\boldsymbol{x}) = \frac{1}{2}(\cos(\pi x_1)+\cos(2\pi x_2))$ for all $\boldsymbol{x} \in D$, $(\lambda_n)_{n\in\mathbb{N}}$ is a given absolutely summable sequence, $\nu \coloneqq (a_0 - a_{\min})^{-1}\sum_{n=1}^{\infty}|\lambda_n|$, $a_{\min} = \min\left\{a(\boldsymbol{y},\boldsymbol{x}) : \boldsymbol{y} \in [-1,1]^{\mathbb{N}},\ \boldsymbol{x} \in D\right\} \in (0, a_0)$. Indeed, for any $\boldsymbol{y} \in \boldsymbol{\Gamma}$, $\boldsymbol{x} \in D$, $a(\boldsymbol{y},\boldsymbol{x}) \geq a_0 - \frac{1}{\nu}\sum_{n=1}^{\infty}\lambda_n \|\psi_n\|_{L^\infty(D)} = a_0 - (a_0 - a_{\min}) = a_{\min}$. The decay of $(|\lambda_n|)_{n\in\mathbb{N}}$ determines how "smooth" the random field is (cf. Section 1.1.2). For example consider, for $\gamma > 1$,

$$\lambda_n = n^{-\gamma} \qquad \forall n \in \mathbb{N}, \tag{5.18}$$

so that the random field becomes smoother and smoother as γ increases. In order to predict the convergence rate of sparse grid interpolation methods, we have to estimate the regularity vector $\boldsymbol{g}$ (5.6). Due to Lemma 5.7, for all $n \in \mathbb{N}$,

$$\left\|\frac{\partial_{y_n} a(\boldsymbol{y})}{a(\boldsymbol{y})}\right\|_{L^\infty(D)} \leq \left\|\frac{\lambda_n \psi_n(\boldsymbol{x})}{\nu a(\boldsymbol{y})}\right\|_{L^\infty(D)} \leq \frac{\lambda_n}{\nu a_{\min}} =: \gamma_n,$$

$$\tau_n = \frac{1}{2\gamma_n} = \frac{\nu a_{\min}}{2} n^{\gamma},\ g_n = \log(\tau_n + \sqrt{1+\tau_n^2}) \approx \log(2\tau_n) = \log(\nu a_{\min} n^{\gamma}).$$

Supposed that $\varepsilon \approx 1$, so that the anisotropy vector in (5.10) can be chosen $\boldsymbol{\alpha} \approx \boldsymbol{g}$. Then, the following series is summable for parameter $\beta > 0$:

$$\sum_{n=1}^{\infty} \frac{1}{e^{\alpha_n/\beta} - 1} = \sum_{n=1}^{\infty} \frac{1}{C n^{\varepsilon\gamma/\beta} - 1} < \infty \Leftrightarrow \beta < \varepsilon\gamma,$$

where $C = (\nu a_{\min})^{\varepsilon/\beta}$. Thus, the results in the previous sections predict e.g. that the dimension-adaptive Algorithm 13 converges at least algebraically with order $\gamma - 1$ (with respect to the number of collocation points). Observe that for $\gamma \leq 1$ the sequence defined by $\lambda_n = n^{-\gamma}$ is not summable, so not even the random diffusion would be well-defined.

In all the following examples, we use Leja sequences (5.3) and the corresponding level-to-know function (5.5). Every adaptive algorithm is steered by the simplified workless profit (5.1).

Quasi-optimality (under the finite dimensional noise assumption)

We verify numerically the results from Sections 5.1.1 and 5.1.2 by comparing a priori and adaptive sparse grid interpolation applied to problems with a fixed finite number of dimensions.

The a priori method consists of a sequence of interpolations with multi-index set $\Lambda_{\boldsymbol{\alpha}}(w)$ for increasing values of $w \geq 0$. For each a priori interpolant, we additionally compute the a-posteriori estimator $\zeta_{SC,\Lambda_{\boldsymbol{\alpha}}(w)}$. We compare this to the fixed dimension, threshold marking adaptive (Algorithm 12). We run both algorithms until we reach 200 collocation nodes. In Figure 5.1, we plot results for $N = 5, 10, 15$. A qualitative comparison shows clearly that the adaptive algorithm has smaller estimator and error that the sequence of a priori approximations. This can be understood as a quasi-optimality property as suggested in Proposition 5.10. The difference in performance seems to depend only mildly on the number of dimensions N. In all cases, convergence rates seem at first algebraic, and then exponential. The transition happens later and later as N increases and convergence gets slower and slower as N increases.

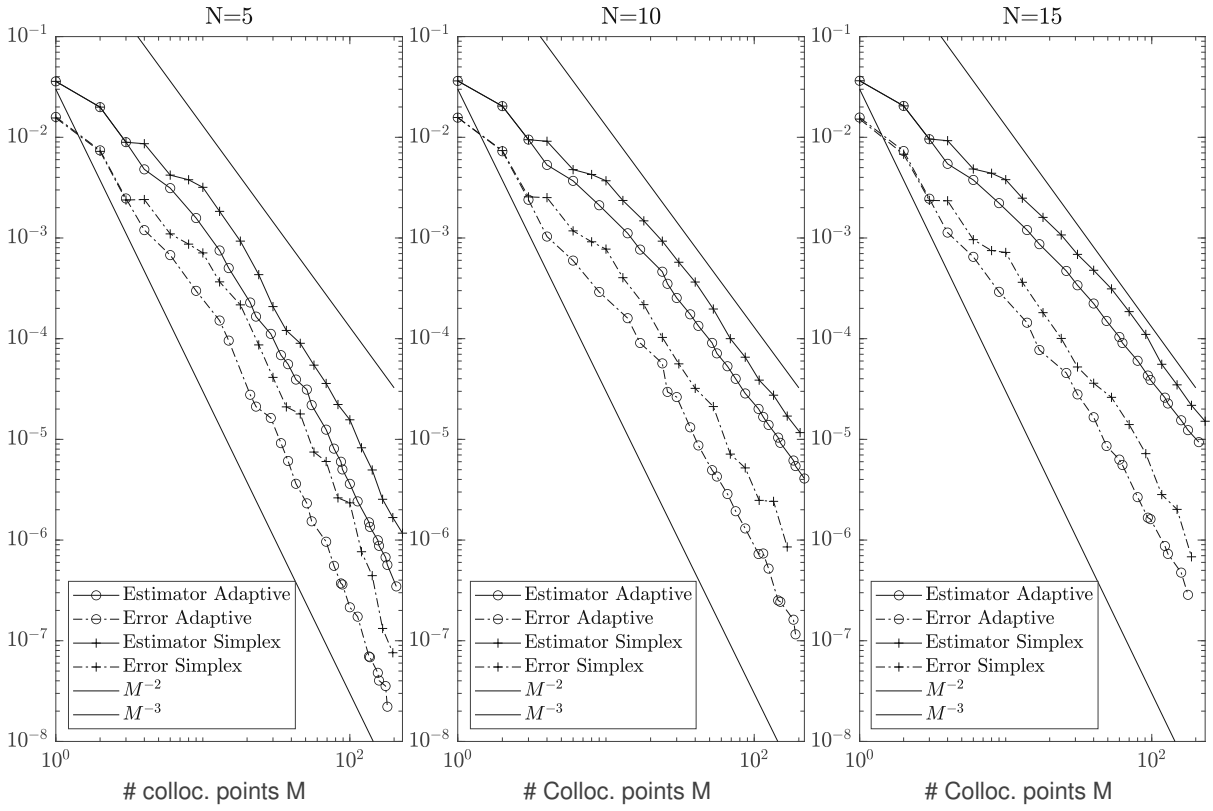

Figure 5.1: Comparison of adaptive and a priori (in the legend "Simplex") sparse grid constructions under the finite dimensional noise assumption. Problem with diffusion (5.17), $\lambda_n = n^{-3}$, $a_{\min}/a_0 = 0.5$. Left to right: $N = 5, 10, 15$.

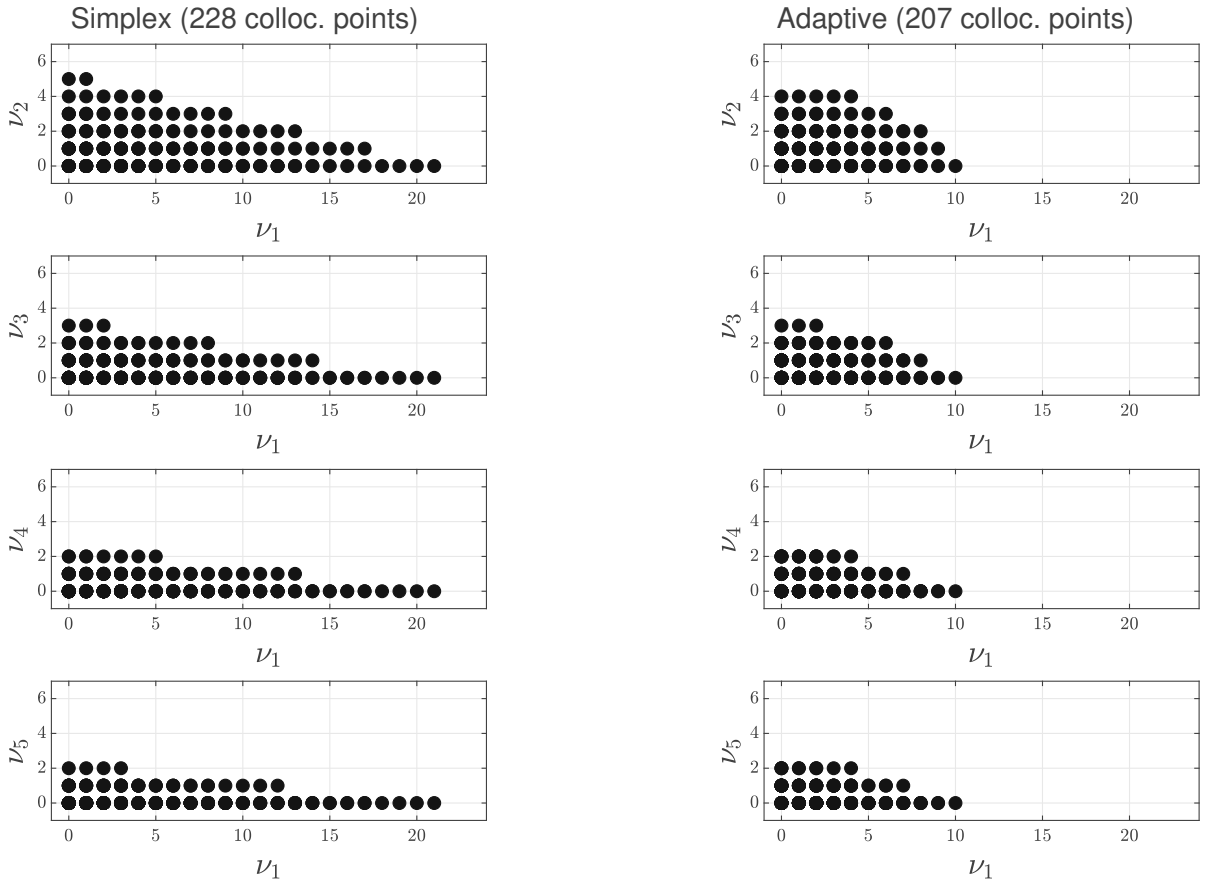

Figure 5.2: Comparison of multi-index sets correspond to about 200 collocation nodes from adaptive and a priori sparse grid in $N = 5$ dimensions. Problem with diffusion (5.17), $\lambda_n = n^{-3}$, $a_{\min}/a_0 = 0.5$. Left: Final multi-index set from a priori constructions (anisotropic simplex). Right: Final multi-index set from ASG algorithm.

In figure 5.2, we compare the adaptively-determined and a priori multi-index sets for $N = 5$ and the maximum number of collocation nodes (about 200). The adaptive algorithm produces multi-index sets that are fairly different than simplexes, with more multi-indices in the "center".

Dimension-adaptive sparse grind interpolation

Let us now test the TDASG algorithm (Algorithm 13) on the same problem as above. We compare it to the finite-dimensional version and check whether the algebraic convergence predicted in Corollary 5.15 is achieved. See Figure 5.3. We observe that the estimator from the dimension-adaptive algorithm coincides with the one of the fixed-dimension algorithm with dimension $N \in \mathbb{N}$ for the iteration $\ell \in \mathbb{N}_0$ such that $N_\ell = N$.

It can also be seen that the TDASG algorithm adds a dimension at almost every iteration. This can be understood observing that the pointwise estimator $\zeta_{\boldsymbol{\nu},\Lambda}$ for $\boldsymbol{\nu} = \boldsymbol{e}_n$ decreases only *algebraically* in n (Figure 5.3 bottom right). More precisely, one proves (recall that for the current numerical experiment $\|a_n\|_{L^\infty(D)} \cong n^{-3}$)

$$\zeta_{\mathbf{0}+\boldsymbol{e}_n,\{\mathbf{1}\}} \leq \|a_n\|_{L^\infty(D)} \|\nabla u(\mathbf{0})\|_{L^2(D)} .$$

On the other hand, estimate (5.14) shows that adding multi-indices in the "bulk" of the margin makes the pointwise estimators decrease exponentially.

The number of dimensions N_ℓ seems to grow linearly with respect to both the number of iterations ℓ and the number of collocation nodes $\#\mathcal{H}_{\Lambda_\ell}$ (this is theoretically explained in the previous paragraph). This suggest that the bound in Proposition 5.14 is not tight.

In Figure 5.4, we study how the convergence of the dimension-adaptive algorithm depends on the problem regularity, by changing the decay parameter $\gamma > 0$ in (5.17). We also compare it with a sequence of a priori solutions defined using the anisotropic simplex (Definition 5.6) with maximum nonzero component $N_\Lambda = \max\{n \in \mathbb{N} : \alpha_n \leq w\}$ (increasing for $w \geq 0$ increasing). For both a priori and adaptive discretizations, we compute the (total) a-posteriori estimator and the reference error. We first observe that, as expected, in all cases the decay rate of estimators and errors increase as γ increases and the difference is more pronounced if the regularity, proportional to γ, is low. However, the estimator consistently converges at a slower rate than the error. This can be attributed to the fact that the error estimator is a sum of error *indicators* (i.e. pointwise estimators) obtained through a triangle inequality (see [GN18, end of proof of Proposition 4.3]). This causes an over-estimation in the uniform norm because of a lack of orthogonality (more on this below). The adaptive method always has a marginal advantage over the a priori one. This confirms the idea that the a priori approximation is quasi-optimal for a whole class of parameter-to-solution maps sharing the same regularity, while the adaptively determined approximation is always tailored to the function at hand.

A simplified a-posteriori parametric estimator on the reduced margin

Consider, for a downward-closed multi-index set $\Lambda \subset \mathcal{F}$, the following heuristic a-posteriori estimator defined as a sum over the *reduced* margin (Definition 1.19): $\overline{\zeta}_{SC,\Lambda} \coloneqq \sum_{\boldsymbol{\nu}\in\mathcal{RM}_\Lambda} \zeta_{\boldsymbol{\nu},\Lambda}$. Since the reduced margin is a subset of the margin, clearly $\overline{\zeta}_{SC,\Lambda} \leq \zeta_{SC,\Lambda} = \sum_{\boldsymbol{\nu}\in\mathbb{M}_\Lambda} \zeta_{\boldsymbol{\nu},\Lambda}$.

We conjecture that, if the terms of the sparse grid expansions $u = \sum_{\boldsymbol{\nu}\in\mathcal{F}} \Delta_{\boldsymbol{\nu}} u$ decay sufficiently quickly (in a specific sense), then there exists $C > 0$ independent of Λ such that the opposite estimate holds: $\zeta_{SC,\Lambda} \leq C\overline{\zeta}_{SC,\Lambda}$.

In Figure 5.5, we compare the reliable estimator $\zeta_{SC,\Lambda}$ the simplified estimator $\overline{\zeta}_{SC,\Lambda}$. In the second case, we also alter the refinement routine: We look for a profit maximizer over the

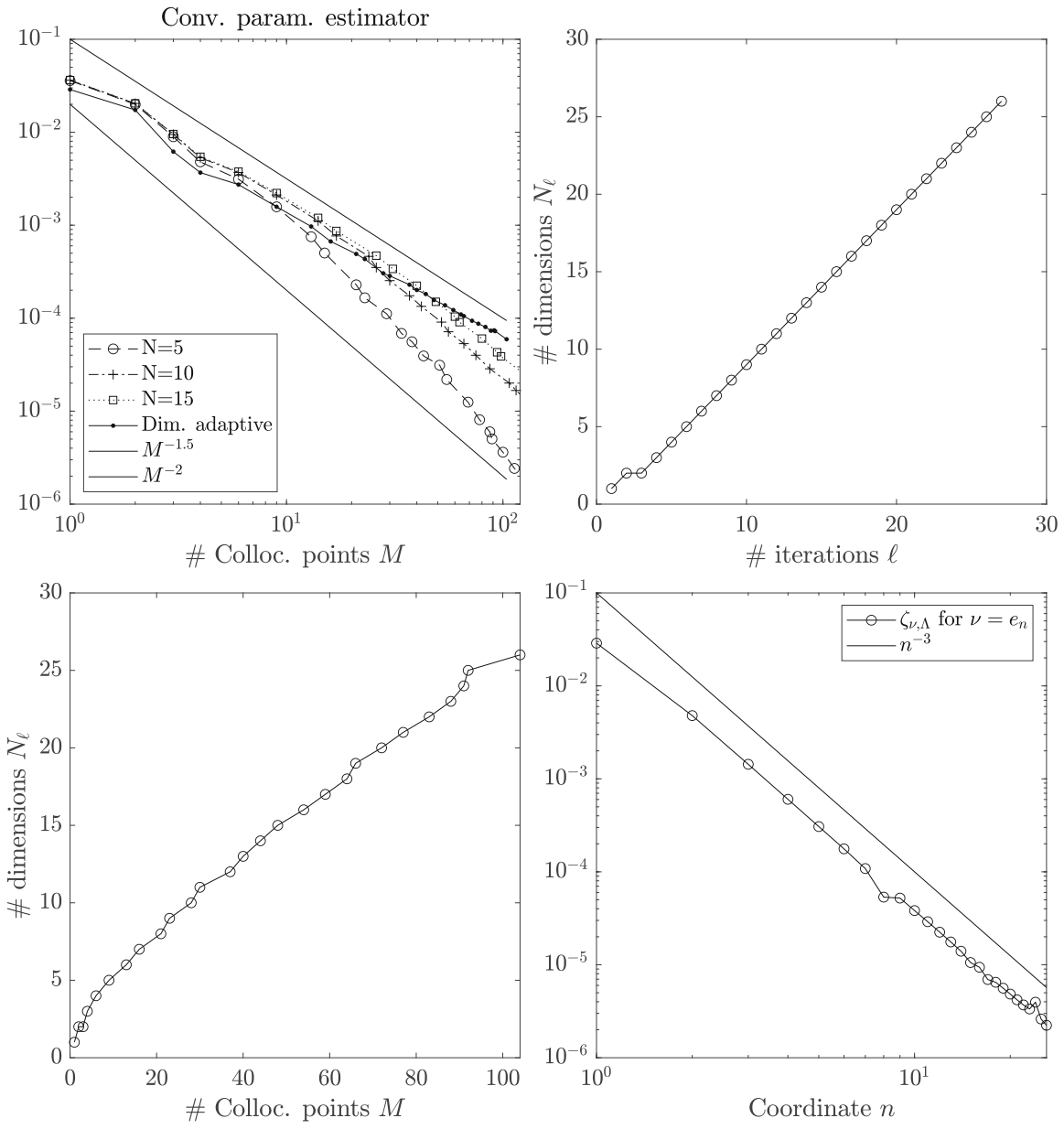

Figure 5.3: Tests for the dimension-adaptive sparse grid algorithm. Problem with diffusion (5.17), $\lambda_n = n^{-3}$, $a_{\min}/a_0 = 0.5$. Top left: Comparison of parametric estimator between adaptive-dimension algorithm and fixed dimensions $N = 5, 10, 15$. Top right: Growth of number of dimensions N_ℓ with respect to ℓ for the TDASG algorithm. Bottom left: Growth of number of dimensions N_ℓ with respect to number of collocation points M_ℓ Bottom right: Decay of $\zeta_{\boldsymbol{\nu},\Lambda}$ for $\boldsymbol{\nu} = \boldsymbol{e}_n$ as $n \to \infty$.

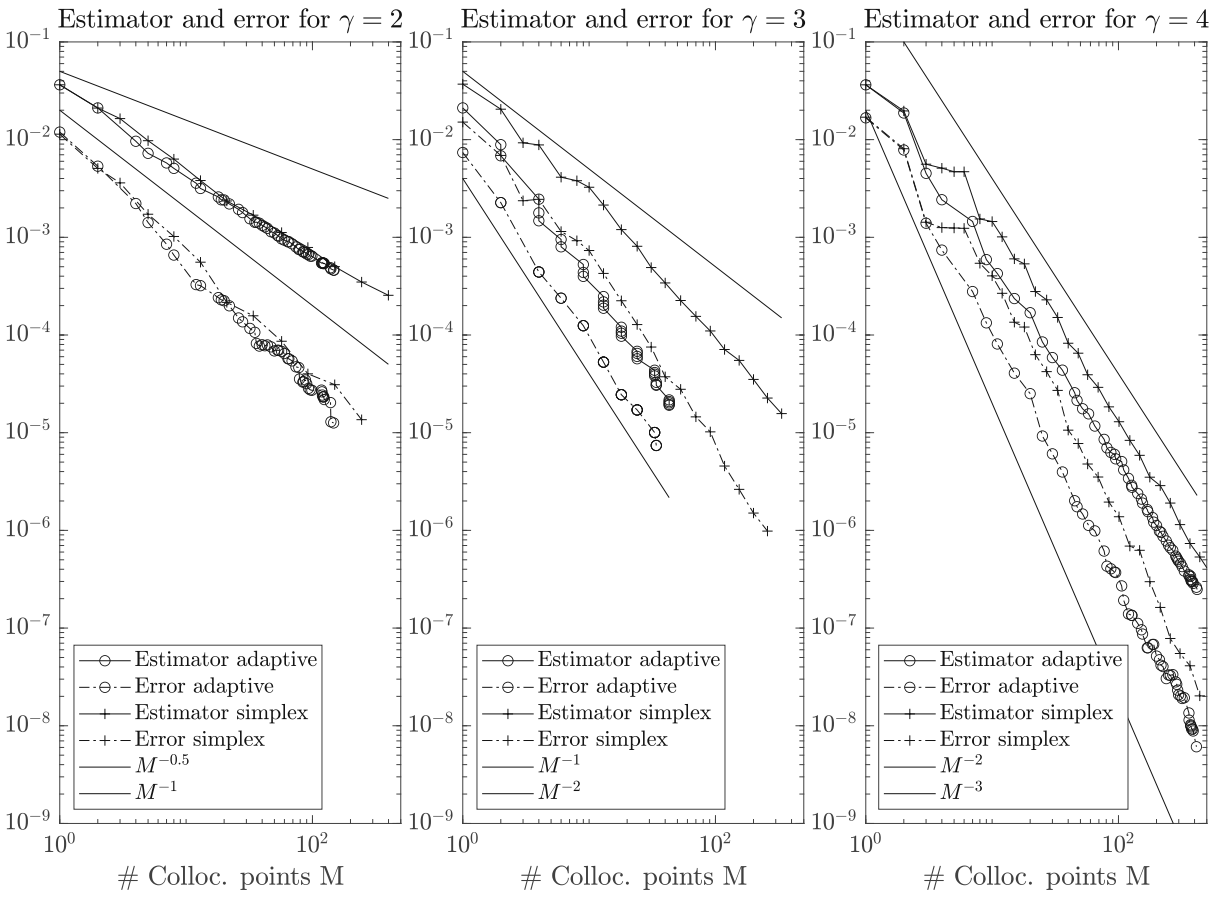

Figure 5.4: Comparison of dimension-adaptive (circles) and simplexes (crosses) discretization for smoother and smoother problems defined by $\gamma \in (2, 3, 4)$ and $a_{\min}/a_0 = 0.5$ in the diffusion (5.17). For each, we plot the total error estimator (continuous line) and the reference error (dashed line).

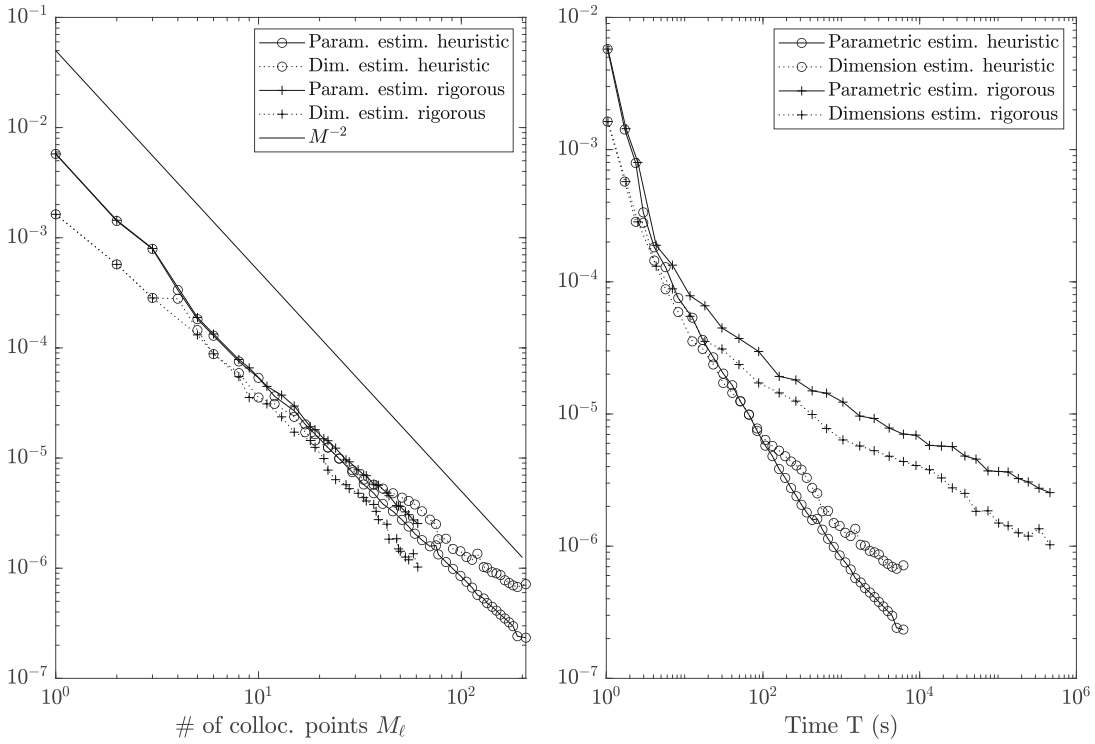

Figure 5.5: Comparison of adaptive algorithm steered by the a-posteriori estimator $\zeta_{SC,\Lambda}$ ("rigorous") and by the simplified counterpart $\overline{\zeta}_{SC,\Lambda}$ ("heuristic"). Problem with diffusion (5.17) and $\lambda_n = n^{-3}$, $a_{\min}/a_0 = 0.9$, and $a_0 = 1$. Left: Parametric and dimension estimators for the algorithm driven by the reliable and simplified parametric estimators; Right: The same information as in the previous plots, with computational time on the x axis (in seconds).

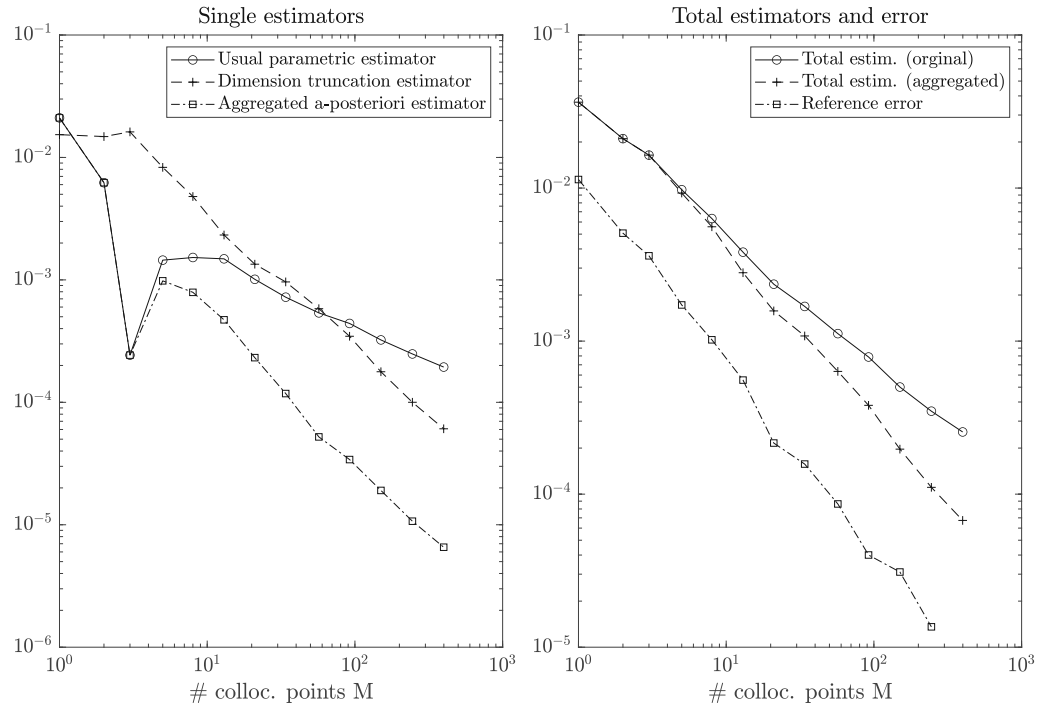

Figure 5.6: Comparison of original and aggregated error estimator for the $L^\infty(\mathbf{\Gamma}, H^1_0(D))$-error. All quantities are computed over the sequence of solutions generated by the dimension-adaptive Algorithm 13. Problem with diffusion (5.17) and $\lambda_n = n^{-2}$, $a_{\min}/a_0 = 0.5$. Left: Comparison between original and aggregated parametric estimator. We also plot the dimension truncation estimator ζ_{TR,Λ_ℓ} (5.12). Right: comparison of total a-posteriori estimator using either original or aggregated parametric estimator. We also plot the reference error.

reduced margin only: $\boldsymbol{\nu}_\ell \in \mathcal{RM}_{\Lambda_\ell}$. This has the added benefit of directly obtaining a downward-closed multi-index set as $\Lambda_\ell \cup \{\boldsymbol{\nu}_\ell\}$. The simplified estimator is very close to the reliable one and their distance is uniformly bounded with respect to the number of collocation points. The heuristic estimator seems to be a good approximation of the reliable one, but it is much faster to compute.

Efficiency of an "aggregated" estimator of the parametric $L^\infty(\mathbf{\Gamma})$ error

Numerical experiments (see e.g. Figure 5.4) clearly show that the parametric estimator $\zeta_{SC,\Lambda}$ is not always an efficient estimator of the $L^\infty(\mathbf{\Gamma}, H^1_0(D))$-error. This phenomenon becomes more and more pronounced as the number of dimensions increases or as the regularity of the problem decreases. Consider instead the following *aggregated a-posteriori error estimator*: $\tilde{\zeta}_{SC,\Lambda} \coloneqq \left\| \sum_{\boldsymbol{\nu}\in\mathcal{M}_\Lambda} \Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u]\right) \right\|_{L^\infty(\mathbf{\Gamma}, L^2(D))}$. We conjecture that it converges with the same rate as the error, i.e. its efficiency is robust with respect to problem parameters such as number of parametric dimensions. The aggregated estimator (up to addition of a truncation error estimator (5.12)) can be proved to bound the error following the proof of [GN18, Proposition 4.3], but skipping the triangle inequality at the end of the proof. The results in Figure 5.6 suggests that the conjecture indeed holds (at least for the multi-index sets generated by the adaptive algorithm).

Estimator of the parametric $L^2(\mathbf{\Gamma})$ error

Let us now consider the $L^2(\mathbf{\Gamma}, H_0^1(D))$-error. Recall that the a-posteriori estimate introduced in [GN18, Proposition 4.3] is valid for any $p \in [1, \infty]$. We conjecture that the following "ℓ^2-type" a-posteriori estimate (under the finite dimensional noise assumption):

$$\|u - \mathcal{I}_\Lambda[u]\|_{L^2(\mathbf{\Gamma}, H_0^1(D))} \lesssim \left(\sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \zeta_{\boldsymbol{\nu},\Lambda}^2 \right)^{1/2}. \tag{5.19}$$

In other words, we claim that in the case of the L^2 error, a valid estimator is given by an ℓ^2 sum of parametric error *indicators* (i.e. pointwise error estimators).

This is obtained from the estimator of [GN18, Proposition 4.3] by ignoring the scalar products $\langle \Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u]\right), \Delta_{\boldsymbol{j}} \left(a \nabla \mathcal{I}_\Lambda[u]\right) \rangle_{L^2(\mathbf{\Gamma}, L^2(D))}$ when $\boldsymbol{\nu} \neq \boldsymbol{j}$. The numerical results in Figure 5.7 suggest that (unlike the case of the $L^\infty(\mathbf{\Gamma}, H_0^1(D))$), this estimator is, up to constant, equivalent to the aggregated L^2 parametric estimator:

$$\left\| \sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u]\right) \right\|_{L^2(\mathbf{\Gamma}, L^2(D))} \cong \left(\sum_{\boldsymbol{\nu} \in \mathcal{M}_\Lambda} \| \Delta_{\boldsymbol{\nu}} \left(a \nabla \mathcal{I}_\Lambda[u]\right) \|_{L^2(\mathbf{\Gamma}, L^2(D))}^2 \right)^{1/2}, \tag{5.20}$$

and a reliable and efficient estimator of the L^2 parametric error.

Alternative dimension-adaptive strategy

When Algorithm 13 performs dimension refinement, the value of the dimension truncation estimator $\zeta_{TR,\Lambda}$ (5.12) is not taken into account. We can modify the adaptive loop in the TDASG algorithm (Algorithm 13) to perform *either* parametric *or* dimension refinement based on which estimator is larger. The numerical experiments in Figure 5.8 suggest that the two strategies are equivalent.

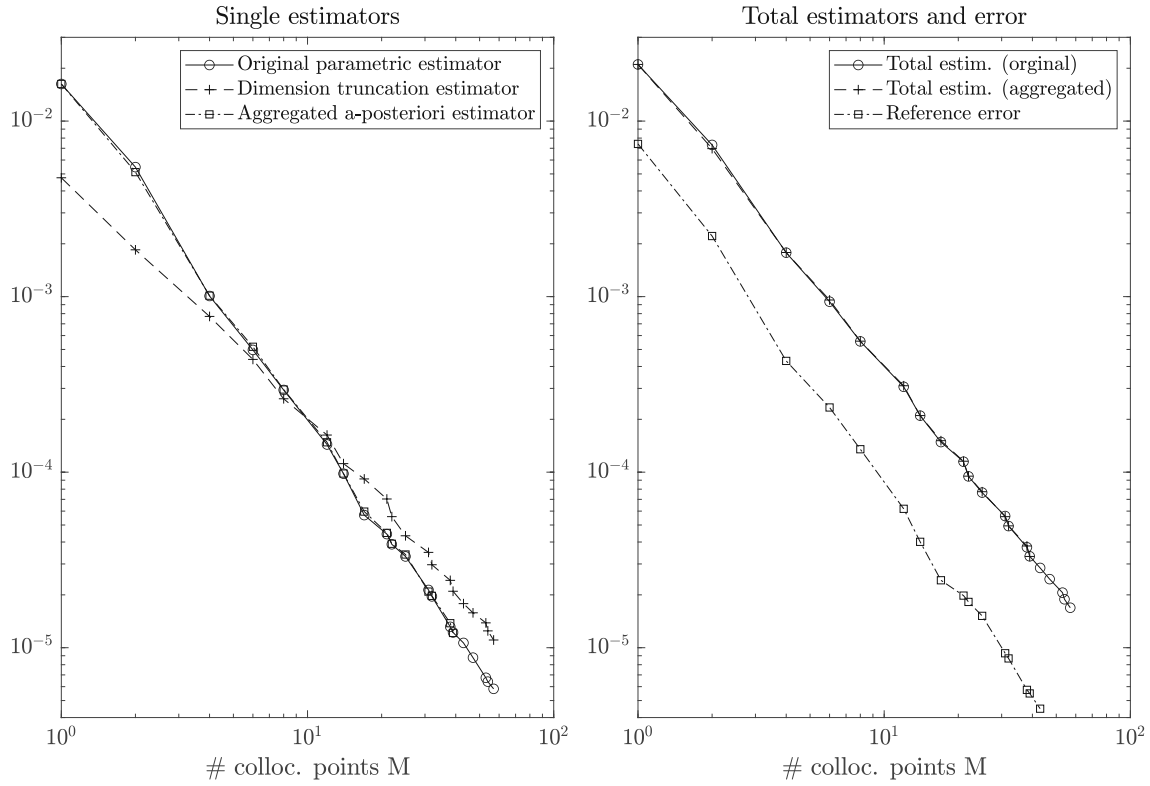

Figure 5.7: Comparison between original, "ℓ^2", and aggregated parametric estimators for the $L^2(\mathbf{\Gamma}, H_0^1(D))$-error. All quantities are computed over the sequence of solutions generated by the TDASG algorithm (Algorithm 13). Problem with diffusion (5.17) and $\lambda_n = n^{-3}$, $a_{\min}/a_0 = 0.5$. Left: Comparison between original and aggregated parametric estimator. We also plot the dimension truncation estimator ζ_{TR,Λ_ℓ}. Right: Comparison of total a-posteriori estimator using either original or aggregated parametric estimator. We also plot the reference error.

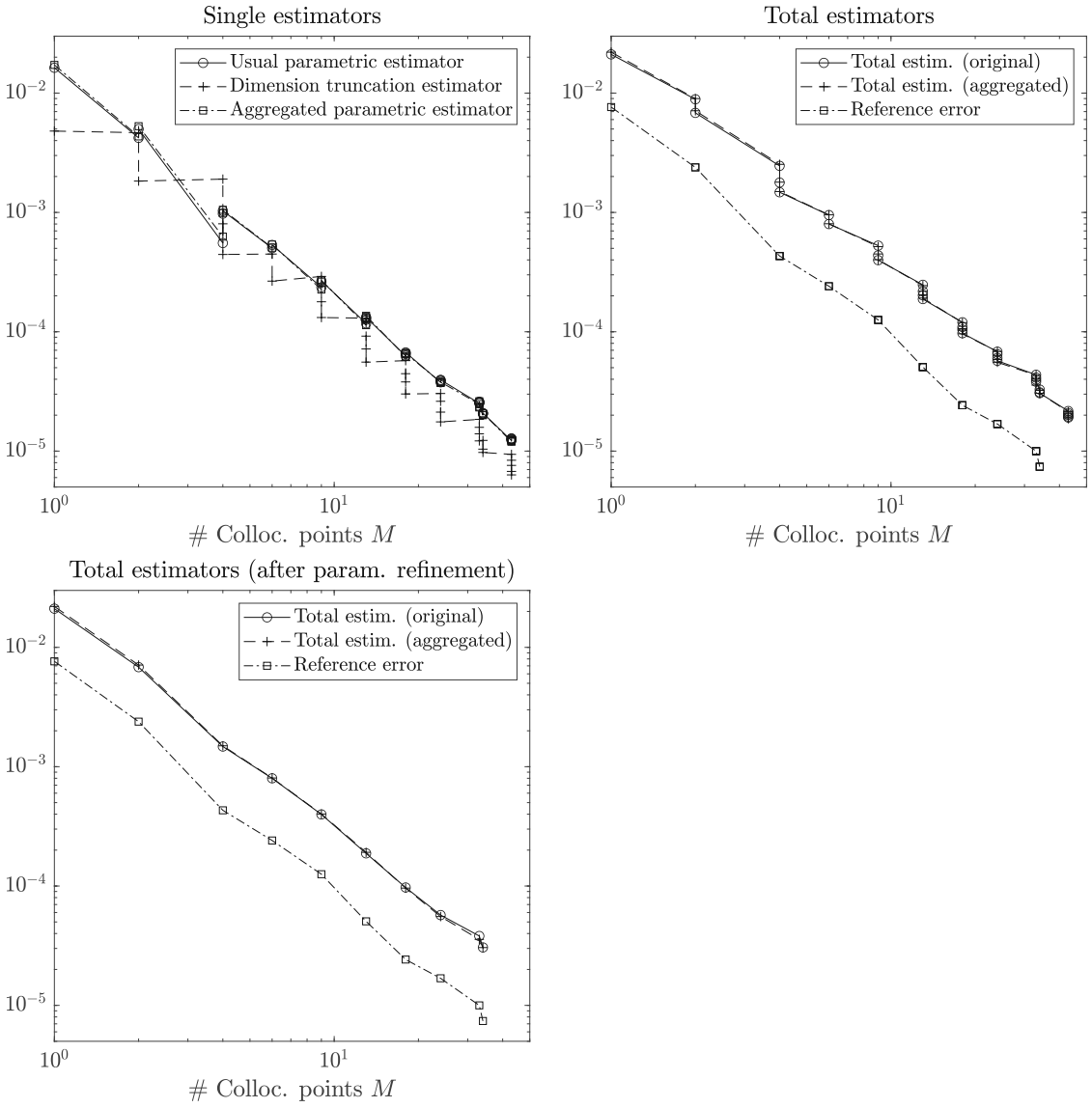

Figure 5.8: Comparison of original and aggregated error estimators for the dimension-refinement Problem with diffusion (5.17) and $\lambda_n = n^{-3}$, $a_{\min}/a_0 = 0.5$ in $L^2(\mathbf{\Gamma}, L^2(D))$ norm. Top left: Comparison between original and aggregated parametric estimator and dimensions truncation estimator. We also plot the dimension truncation estimator ζ_{TR,Λ_ℓ}. Top right: Comparison of total a-posteriori estimator, using either original or aggregated parametric estimator. We also plot the reference error. Bottom left: As in the previous plot, but plotting the estimator and reference error only after a parametric refinement.

5.2 Additional results on the space and time approximation of the SLLG equation

In this section, we give a partial proof of convergence of a tangent plane scheme applied to sample paths of the random coefficients LLG equation introduced in Section 1.3.6. Importantly, the Wiener process appears in the equation as a problem datum. One expects that the fact that its sample paths are only Hölder continuous with exponent less than $\frac{1}{2}$ reduces the convergence of the tangent plane scheme. Nevertheless, numerical experiments consistently show the same convergence rate of the classical tangent plane scheme applied to the deterministic LLG equation, i.e. order one in both space and time, when the error is measured in the $L^\infty(0,T,\mathbb{H}^1(D))$-norm. In Theorem 5.21, we motivate this by proving convergence in expectation ("strong convergence" in SDE terminology) and in the $L^\infty(0,T,\mathbb{H}^1(D))$-norm for space and time. The section ins structured as follows: In Section 5.2.1, we prove some technical lemmata needed later. In Section 5.2.2, we consider a random coefficient ODE, which can be interpreted as a simplification of the main problem and demonstrate how the proof strategy can be applied in this easier setting. Finally, in Section (5.2.3), we state and give a partial proof of the mean convergence of the tangent plane scheme for the random coefficient LLG equation. The validity of the proof depends on the existence of an appropriate test function, which we could not prove.

For the rest of the section, let $(\Omega,\mathcal{E},|PP)$ denote a probability triple, formed respectively by a set, a sigma-algebra on Ω, and a probability measure on the measurable set $\Omega,\mathcal{E})$. Denote by $W:\Omega\times[0,T]\to\mathbb{R}$ the Wiener process, where $T>0$. Since we work with time discretization, we also fix a notation for the time-stepping: Consider $J\in\mathbb{N}$, the time step size $k=\frac{T}{J}$ and time steps $t_j\coloneqq jk$ for all $j=0,\dots,J$.

5.2.1 Technical results

We begin by proving some technical results.

Lemma 5.16. *Denote by $B:\Omega\times[0,T]\to\mathbb{R}$ the Brownian bridge with $B(T)=B(0)=0$. Then*

$$\int_0^T B(t)\,dt\sim\mathcal{N}\left(0,\frac{T^3}{12}\right),\qquad \int_0^T tB(t)\,dt\sim\mathcal{N}\left(0,\frac{T^5}{45}\right).$$

Proof. We prove the first fact. The proof of the second is analogous. Since the integral is a linear operator and B is a Gaussian stochastic process, $\int_0^T B(s)\mathrm{d}s$ is a Gaussian random variable. Its expectation is $\mathbb{E}\int_0^T B(s)\mathrm{d}s=\int_0^T\mathbb{E}B(s)\mathrm{d}s=0$ since the mean of the Brownian Bridge is identically zero. As for the variance, first recall that the covariance of the Brownian Bridge is $\mathbb{E}\left[B(s)B(t)\right]=\min(s,t)-\frac{st}{T}$. Then,

$$\mathbb{E}\left(\int_0^T B(s)\mathrm{d}s\right)^2=\int_0^T\int_0^T\mathbb{E}\left(B(s)B(t)\right)\mathrm{d}s\mathrm{d}t=\int_0^T\int_0^T\min(s,t)-\frac{st}{T}\mathrm{d}s\mathrm{d}t=\frac{T^3}{3}-\frac{T^3}{4}=\frac{T^3}{12}.$$

□

Remark 5.17 (Discrete Itô isometry)**.** *Consider $(X_j)_{j=1}^J$ random variables such that X_i and $W(t_{j+1})-W(t_j)$ are independent if $j\geq i$. Then,*

$$\mathbb{E}\left[\left(\sum_{j=0}^J X_j\left(W(t_{j+1})-W(t_j)\right)\right)^2\right]=\sum_{j=0}^J\mathbb{E}\left[X_j^2\right](t_{j+1}-t_j).$$

Indeed,

$$\mathbb{E}\left[\left(\sum_{j=0}^{J} X_j \left(W(t_{j+1}) - W(t_j)\right)\right)^2\right] = \mathbb{E}\left[\sum_{i,j=0}^{N} X_i X_j \left(W(t_{i+1}) - W(t_j)\right)\left(W(t_{j+1}) - W(t_j)\right)\right].$$

Assume $i < j$. Then, because of the assumption on X_i, X_j and the independence of Wiener process increments, we have

$$\begin{aligned}&\mathbb{E}\left[X_i X_j \left(W(t_{i+1}) - W(t_i)\right)\left(W(t_{j+1}) - W(t_j)\right)\right]\\ &\quad = \mathbb{E}\left[X_i X_j \left(W(t_{i+1}) - W(t_i)\right)\right]\mathbb{E}\left[W(t_{j+1}) - W(t_j)\right] = 0.\end{aligned}$$

The same identity holds for $i > j$. Finally, if $i = j$,

$$\mathbb{E}\left[X_i^2 \left(W(t_{i+1}) - W(t_i)\right)^2\right] = \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[\left(W(t_{i+1}) - W(t_i)\right)^2\right] = \mathbb{E}\left[X_i^2\right](t_{i+1} - t_i).$$

The two previous facts imply the following lemma, which we apply to the next two problems.

Lemma 5.18. *Let $(X_j)_{j=0}^J$ be a sequence of (possibly dependent) random variables such that X_j depends on $W(t)$ only through $W(t_0), \dots, W(t_j)$. Then,*

$$\mathbb{E}\left|\sum_{j\le J} X_j \int_{t_j}^{t_{j+1}} W(s) - W(t_j)\,ds\right| \le k\sqrt{\mathbb{E}\left[k\sum_{j\le J} X_j^2\right]}\left(\frac{1}{\sqrt{6\pi}} + \frac{1}{2}\right),$$

$$\mathbb{E}\left|\sum_{j\le J} X_j \int_{t_j}^{t_{j+1}} (s - t_j)\left(W(s) - W(t_j)\right) ds\right| \le k^2\sqrt{\mathbb{E}\left[k\sum_{j\le J} X_j^2\right]}\left(\sqrt{\frac{2}{45\pi}} + \frac{1}{2}\right).$$

Proof. We prove the first estimate. The proof of the second is analogous. By the law of iterated expectations,

$$\begin{aligned}&\mathbb{E}\left|\sum_{j\le J} X_j \int_{t_j}^{t_{j+1}} W(s) - W(t_j)\mathrm{d}s\right|\\ &\quad = \mathbb{E}_{\boldsymbol{\omega}}\left[\mathbb{E}\left[\left|\sum_{j\le J} X_j \int_{t_j}^{t_{j+1}} W(s) - W(t_j)\mathrm{d}s\right| \middle| W(t_j) = \omega_j \text{ for } j = 1, \dots, J\right]\right],\end{aligned}$$

where we conditioned the value of the Wiener process at J discrete times. Observe that, assuming the conditioning above,

$$\int_{t_j}^{t_{j+1}} W(s) - W(t_j)\mathrm{d}s = \int_{t_j}^{t_{j+1}} \mathcal{L}_j(s) + B_j(s)\mathrm{d}s = \frac{k}{2}(\omega_{j+1} - \omega_j) + \int_{t_j}^{t_{j+1}} B_j(s)\mathrm{d}s,$$

where $\mathcal{L}_j(s)$ is an appropriate affine function and $B_j(s)$ is the Brownian bridge with value 0 at t_j and t_{j+1}. The distribution of the last integral was determined in Lemma 5.16. Thus,

$$\sum_{j\le J} X_j \int_{t_j}^{t_{j+1}} W(s) - W(t_j)\mathrm{d}s \sim \mathcal{N}\left(\mu_k(\boldsymbol{\omega}), \sigma_k^2(\boldsymbol{\omega})\right),$$

where

$$\mu_k(\boldsymbol{\omega}) = \frac{k}{2}\sum_{j\leq J} X_j(\omega_{j+1}-\omega_j), \qquad \sigma_k^2(\boldsymbol{\omega}) = \frac{k^3}{12}\left(\sum_{j\leq J} X_j^2\right).$$

The absolute value of the previous random variable has a folded normal distribution. Therefore, we get:

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{j\leq J} X_j \int_{t_j}^{t_{j+1}} W(s)-W(t_j)\mathrm{d}s\right| \middle| W(t_j)=\omega_j \text{ for } j=1,\dots,J\right] \\ = \sigma_k(\boldsymbol{\omega})\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu_k^2(\boldsymbol{\omega})}{\sigma_k^2(\boldsymbol{\omega})}\right) + \mu_k(\boldsymbol{\omega})\left(1-2\Phi\left(-\frac{\mu_k(\boldsymbol{\omega})}{\sigma_k(\boldsymbol{\omega})}\right)\right).\end{aligned}$$

Let us finally estimate the expectation with respect to $\boldsymbol{\omega}$. Observe that $\left|\exp\left(-\frac{\mu_k^2(\boldsymbol{\omega})}{\sigma_k^2(\boldsymbol{\omega})}\right)\right| \leq 1$ and $\left|\left(1-2\Phi\left(-\frac{\mu_k(\boldsymbol{\omega})}{\sigma_k(\boldsymbol{\omega})}\right)\right)\right| \leq 1$. Also observe that

$$\mathbb{E}\left|\sigma_k(\boldsymbol{\omega})\right| \leq k\frac{1}{2\sqrt{3}}\sqrt{\mathbb{E}\left[k\sum_{j\leq J} X_j^2\right]}.$$

By Cauchy-Schwarz, we estimate

$$E\left|\mu_k(\boldsymbol{\omega})\right| = \frac{k}{2}\mathbb{E}\left|\sum_{j\leq J} X_j(\omega_{j+1}-\omega_j)\right| \leq \frac{k}{2}\sqrt{\mathbb{E}\left[\left|\sum_{j\leq J} X_j(\omega_{j+1}-\omega_j)\right|^2\right]}.$$

By Remark 5.17, we get

$$\mathbb{E}\left|\mu_k(\boldsymbol{\omega})\right| \leq \frac{k}{2}\sqrt{\sum_{j\leq J}\mathbb{E}\left[X_j^2\right](t_{j+1}-t_j)} = \frac{k}{2}\sqrt{\mathbb{E}\left[k\sum_{j\leq J} X_j^2\right]}.$$

All in all, we have found

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\omega}}\left[\sigma_k(\boldsymbol{\omega})\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu_k^2(\boldsymbol{\omega})}{\sigma_k^2(\boldsymbol{\omega})}\right) + \mu_k(\boldsymbol{\omega})\left(1-\Phi\left(-\frac{\mu_k(\boldsymbol{\omega})}{\sigma_k(\boldsymbol{\omega})}\right)\right)\right] \\ \leq k\sqrt{\mathbb{E}\left[k\sum_{j\leq J} X_j^2\right]}\left(\frac{1}{\sqrt{6\pi}}+\frac{1}{2}\right).\end{aligned}$$

□

5.2.2 Convergence of explicit Euler for a family of random ODEs

Consider the final time $T>0$ and the following function with "separation of variables structure":

$$f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}, \qquad f(x,z)=F(x)b(z). \tag{5.21}$$

We assume $F : \mathbb{R} \to \mathbb{R}$ to be globally Lipschitz with Lipschitz constant $L > 0$ and bounded with constant $1 + L$, i.e. for all $x, y \in \mathbb{R}$,

$$|F(x) - F(y)| \leq L|x - y|, \qquad |F(x)| \leq (1 + L)|x|.$$

We assume $b : \mathbb{R} \to \mathbb{R}$ to be differentiable and uniformly bounded up to second order:

$$\sup_{x \in \mathbb{R}} |b(x)| + \sup_{x \in \mathbb{R}} |b'(x)| + \sup_{x \in \mathbb{R}} |b''(x)| = \beta < \infty.$$

Given the initial datum $y_0 \in \mathbb{R}$, consider the following *Random Ordinary Differential Equation* (RODE): Find a stochastic process $y : \Omega \times [0, T] \to \mathbb{R}$ such that

$$y(\omega, t) = y_0 + \int_0^t f(y(\omega, s), W(\omega, s))\mathrm{d}s \qquad \forall t \in [0, T], \text{ for a.e. } \omega \in \Omega. \tag{5.22}$$

The specific form of the right-hand side is chosen so that Equation (5.22) encodes some of the basic properties of the random LLG equation (1.76). it has, up to a sum, an analogous separation of variables structure (3.12).

Gronwall's lemma and the properties of f yield a stability estimate for y:

$$|y(t)| \leq y_0 e^{\beta(1+L)t} \qquad \forall t \in [0, T], \text{ for a.e. } \omega \in \Omega. \tag{5.23}$$

The explicit Euler approximation of the RODE (5.22) is a sequence of random variables $(y_j)_{j=0}^J \subset \mathbb{R}^J$ such that

$$y_{j+1}(\omega) = y_j(\omega) + kf(y_j(\omega), W((\omega), t_j)) \qquad \forall j = 0, \ldots, J, \text{ for a.e. } \omega \in \Omega. \tag{5.24}$$

The aim is to prove that y_j approximates $y(t_j)$ in the appropriate sense. Denote by $\overline{y} : \Omega \times [0, T] \to \mathbb{R}$ the constant extension of $(y_j)_{j=0}^J$, i.e. for any $t \in [0, T]$,

$$\overline{y}(t) = y_j \qquad \text{if } t_j \leq t < t_{i+1}. \tag{5.25}$$

The assumptions on f implies

$$y_{j+1} \leq (1 + k(1 + L)\beta)\, |y_j| \qquad \forall j - 0, \ldots, J - 1,$$

and the discrete Gronwall's inequality gives the following discrete stability bound (analogous to (5.23)):

$$|y_j(\omega)| \leq e^{\beta(1+L)t_j} \qquad \forall j = 0, \ldots, J, \text{ for a.e. } \omega \in \Omega \tag{5.26}$$

The Euler approximation of the sample paths of the RODE is expected to converge with order smaller than $\frac{1}{2}$ since the right-hand side of Equation (5.22) is only Hölder regular with exponent smaller than $\frac{1}{2}$. Nevertheless, numerical tests such as the one reported in Figure 5.9 clearly show convergence with order 1.

Proposition 5.19. *Consider the final time $T > 0$, the probability $(\Omega, \mathcal{E}, \mathcal{P})$, the RODE (5.22) with right-hand side (5.21) depending on the Wiener process W and deterministic initial datum $y_0 \in \mathbb{R}$. Denote by $y : \Omega \times [0, T] \to \mathbb{R}$ its stochastic process solution. For $J \in \mathbb{N}$, define the timestep size $k = \frac{T}{J}$ and the timesteps $t_j = jk$ for $j = 0, \ldots, J$. Consider the Euler approximation (5.24) of Equation (5.22). There holds*

$$\mathbb{E}\, |y(t_j) - y_j| \lesssim k \qquad \forall j = 0, \ldots, J,$$

where the hidden constant depends only on T, L, β.

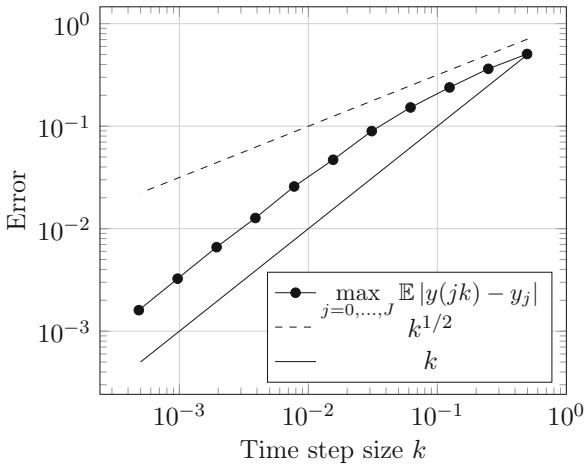

Figure 5.9: Convergence of the Euler scheme applied on $y'(t) = y(t)^2 \sin(W(t))$ for $t \in [0,1]$, $y(0) = 1$ and decreasing timestep size k. Error computed with 100 Monte Carlo samples. Reference solutions computed with reference timestep size $k_{\text{ref}} = 2^{-14}$.

Proof. Recall that $\overline{y}$ denotes the piecewise constant extension of the Euler approximation (cf. Equation 5.25). Analogously denote the piecewise constant extension of the Wiener process

$$\overline{W}(t) \coloneqq W(t_j) \qquad \text{if } t_j \le t < t_{i+j}.$$

This allows us to rewrite the Euler approximation (5.24) as

$$\overline{y}(t) = y_i = y_0 + \int_0^{t_j} f(\overline{y}(s), \overline{W}(s))\mathrm{d}s \qquad \text{if } t_j \le t < t_{i+j}$$

Taking the difference with the RODE (5.22) gives

$$y(t) - \overline{y}(t) = \int_0^{t_i} \left(f(y(s), W(s)) - f(\overline{y}(s), \overline{W}(s))\right) \mathrm{d}s + \int_{t_i}^{t} f(y(s), W(s))\mathrm{d}s.$$

Add and subtract the mixed term $f(\overline{y}, W) = F(\overline{y})b(W)$ to get

$$\begin{aligned} y(t) - \overline{y}(t) = &\int_0^{t_i} \left(F(y(s) - F(\overline{y}(s))\right) b(W(s))\mathrm{d}s + \int_0^{t_i} F(\overline{y}(s)) \left(b(W(s)) - b(\overline{W}(s))\right) \mathrm{d}s \\ &+ \int_{t_i}^{t} f(y(s), W(s))\mathrm{d}s. \end{aligned}$$

The expectation of the absolute value can be estimated with a triangle inequality as

$$\begin{aligned} \mathbb{E}\,|y(t) - \overline{y}(t)| \le &\mathbb{E}\left|\int_0^{t_i} \left(F(y(s) - F(\overline{y}(s))\right) b(W(s))\mathrm{d}s\right| \\ &+ \mathbb{E}\left|\int_0^{t_i} F(\overline{y}(s)) \left(b(W(s)) - b(\overline{W}(s))\right) \mathrm{d}s\right| + \mathbb{E}\left|\int_{t_i}^{t} f(y(s), W(s))\mathrm{d}s\right|. \end{aligned}$$

We now aim at showing the following three estimates, where the constants $C_1, C_2, C_3 > 0$ depend

only on problem data:

$$\mathbb{E}\left|\int_0^{t_i}\left(F(y(s)-F(\overline{y}(s))\,b(W(s))\mathrm{d}s\right|\leq\int_0^{t_i}C_1(s)\mathbb{E}\left|y(s)-\overline{y}(s)\right|\mathrm{d}s, \tag{5.27}$$

$$\mathbb{E}\left|\int_0^{t_i}F(\overline{y}(s))\left(b(W(s))-b(\overline{W}(s))\right)\mathrm{d}s\right|\leq C_2k, \tag{5.28}$$

$$\mathbb{E}\left|\int_{t_i}^{t}f(y(s),W(s))\mathrm{d}s\right|\leq C_3k. \tag{5.29}$$

This, together with a Gronwall's inequality, will give the statement in the form:

$$\mathbb{E}\left|y(t)-\overline{y}(t)\right|\leq e^{\int_0^t C_1(s)\mathrm{d}s}(C_2+C_3)k.$$

We prove the estimates one by one:

- Estimate (5.27): Since F is globally Lipschitz with constant L and b is bounded by $\beta>0$, we get

  $$\left(F(y(s)-F(\overline{y}(s))\,b(W(s))\leq L\beta\left|y(s)-\overline{y}(s)\right|\qquad\forall s\in[0,T],\text{for a.e. }\omega\in\Omega.$$

  Therefore, (5.27) follows with $C_1=L\beta$.

- Estimate (5.29): Recalling the definition of f, the fact that F is bounded by $1+L$ and that b is bounded by $\beta>0$, we get that

  $$\sup_{s\in[0,T]}|f(y,W)|\leq\beta(1+L)\sup_{s\in[0,T]}|y(s)|\qquad\text{for a.e. }\omega\in\Omega.$$

  Finally, recall the stability estimate on the exact solution (5.23). This implies the desired estimate with $C_3=\beta(1+L)y_0e^{\beta(1+L)T}$.

- Estimate (5.28). First, by definition of $\overline{y}$, we have

  $$\int_0^{t_i}F(\overline{y}(s))\left(b(W(s))-b(\overline{W}(s))\right)\mathrm{d}s=\sum_{j=0}^{i-1}F(y_j)\int_{t_j}^{t_{j+1}}b(W(s))-b(W(t_j))\mathrm{d}s.$$

  From Taylor's theorem, there exists $\xi\in[\min\left(W(s),W(t_j)\right),\max\left(W(s),W(t_j)\right)]$ such that

  $$b(W(s))=b(W(t_j))+b'(W(t_j))(W(s)-W(t_j))+b''(\xi)\frac{(W(s)-W(t_j))^2}{2}.$$

  This in turn gives

  $$\begin{aligned}\mathbb{E}\left|\int_0^{t_i}F(\overline{y}(s))\left(b(W(s))-b(\overline{W}(s))\right)\mathrm{d}s\right|\leq&\mathbb{E}\left|\sum_{j=0}^{i-1}F(y_j)b'(W(t_j))\int_{t_j}^{t_{j+1}}(W(s)-W(t_j))\mathrm{d}s\right|\\&+\mathbb{E}\left|\sum_{j=0}^{i-1}F(y_j)\int_{t_j}^{t_{j+1}}b''(\xi)\frac{(W(s)-W(t_j))^2}{2}\mathrm{d}s\right|.\end{aligned} \tag{5.30}$$

We apply Lemma 5.18 to the first right-hand side term with $X_j = F(y_j)b'(W(t_j))$ and get

$$\mathbb{E}\left|\sum_{j<i} F(y_j)b'(W(t_j))\int_{t_j}^{t_{j+1}}(W(s)-W(t_j))\mathrm{d}s\right| \\ \leq k\sqrt{\mathbb{E}\left[k\sum_{j\leq i}(F(y_j)b'(W(t_j))^2\right]}\left(\frac{1}{\sqrt{3}\pi}+\frac{1}{2}\right).$$

The uniform boundedness of b', boundedness of F, and definition of $\overline{y}(s)$ imply

$$\mathbb{E}\left|\sum_{j<i} F(y_j)b'(W(t_j))\int_{t_j}^{t_{j+1}}(W(s)-W(t_j))\mathrm{d}s\right| \\ \leq k\beta(1+L)\sqrt{\mathbb{E}\left\|\overline{y}\right\|^2_{L^2(0,t_i)}}\left(\frac{1}{\sqrt{3}\pi}+\frac{1}{2}\right). \tag{5.31}$$

As for the second right-hand side term, we first estimate the integral using the boundedness of b'' and Hölder continuity of W (with exponent $\frac{1}{2}$ and constant $C_W > 0$ almost surely finite),

$$\left|\int_{t_j}^{t_{j+1}} b''(\xi)\frac{(W(s)-W(t_j))^2}{2}\mathrm{d}s\right| \leq \frac{\beta}{2}\int_{t_j}^{t_{j+1}} C_W^2(s-t_j)\mathrm{d}s = \frac{\beta C_W^2}{4}k^2.$$

A triangle inequality and the previous estimate imply

$$\mathbb{E}\left|\sum_{j=0}^{i-1} F(y_j)\int_{t_j}^{t_{j+1}} b''(\xi)\frac{(W(s)-W(t_j))^2}{2}\mathrm{d}s\right| \leq \frac{\beta C_W^2}{4}k\mathbb{E}\left[\sum_{j=0}^{i-1} k\left|F(y_j)\right|\right].$$

We bound the expectation with Cauchy-Schwarz inequality to get

$$\mathbb{E}\left[\sum_j kF(y_j)\right] \leq \sqrt{\mathbb{E}\left[\sum_j (kF(y_j))^2\right]}\sqrt{J} = \sqrt{\sum_j k\mathbb{E}\left[F(y_j)^2\right]}\sqrt{k}\sqrt{J} \\ = \sqrt{\sum_j k\mathbb{E}\left[F(y_j)^2\right]}\sqrt{T}.$$

The boundedness of F gives

$$\mathbb{E}\left|\sum_{j=0}^{i-1} F(y_j)\int_{t_j}^{t_{j+1}} b''(\xi)\frac{(W(s)-W(t_j))^2}{2}\mathrm{d}s\right| \leq \frac{\beta C_W^2\sqrt{T}(1+L)}{4}k\sqrt{\mathbb{E}\left\|\overline{y}\right\|^2_{L^2(0,t_i)}}. \tag{5.32}$$

Finally, combining (5.30), (5.31), (5.32), and the stability estimate (5.26), we obtain inequality (5.28) with

$$C_2 = \left(\beta(1+L)\left(\frac{1}{\sqrt{3}\pi}+\frac{1}{2}\right)+\frac{\beta C_W^2\sqrt{T}(1+L)}{4}\right)Te^{\beta(1+L)T}.$$

□

5.2.3 Convergence of the tangent plane scheme for random LLG

In the previous section we showed that the Euler scheme applied to a random ODE with a specific structure has convergence order 1 when the error is measured in the expectation of the uniform norm. We did not use the consistency of the Euler method, which requires the boundedness of y'', the second time derivative of the exact solution. Rather, we considered the problem in integral form and relied on: *1.* The smallness of the timesteps (t_i, t_{i+1}), and *2.* The way the right-hand side depends on the Wiener process.

The previous section cannot be directly extended to the tangent plane scheme, but it might be possible for the midpoint rule (see Section 1.3.5).

Let us recall some facts about the random coefficient LLG equation (See Section 1.3.6) and its tangent plane approximation (see Section 1.3.5). We denote $\langle \cdot, \cdot \rangle$ the $\mathbb{L}^2(D)$ scalar product and by $\|\cdot\|$ the corresponding norm. The spaces $\mathbb{L}^2(\cdot)$ and $\mathbb{H}^1(\cdot)$ etc. denote Banach spaces of functions intro $\mathbb{R}^3$. We recall that the tangent plane to $\boldsymbol{m} : D_T \to \mathbb{S}^2$ is, by definition,

$$\mathcal{K}(\boldsymbol{m}) \coloneqq \left\{ \boldsymbol{v} \in L^2(0,T,\mathbb{H}^1(D)) : \boldsymbol{m} \cdot \boldsymbol{v} = 0 textrm a.e. in D_T \right\}.$$

Recall the definitions

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

The random LLG equation in weak form (1.79) may be written:

$$\int_0^T \langle \partial_t \boldsymbol{m} + \boldsymbol{m} \times \partial_t \boldsymbol{m}, \boldsymbol{\phi} \rangle + \langle \nabla \boldsymbol{m}, \nabla \boldsymbol{\phi} \rangle - \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}), \boldsymbol{\phi} \right\rangle = 0 \qquad \forall \boldsymbol{\phi} \in \mathcal{K}(\boldsymbol{m}),$$

Consider a quasi-uniform mesh $\mathcal{T}_h$ with maximal element diameter $h > 0$ and vertices $\mathcal{N}_h$. Denote $\boldsymbol{V}_h \coloneqq \mathcal{S}^1(\mathcal{T}_h)^3$, a space of piecewise-linear functions on the elements of $\mathcal{T}_h$. Denote

$$\mathcal{M}_h \coloneqq \{ \boldsymbol{m}_h \in \boldsymbol{V}_h : |\boldsymbol{m}_h(\boldsymbol{z})| = 1 \quad \forall \boldsymbol{z} \in \mathcal{N}_h \}.$$

The discrete tangent plane to $\boldsymbol{m}_h \in \mathcal{M}_h$ reads:

$$\mathcal{K}_h(\boldsymbol{m}_h) \coloneqq \{ \boldsymbol{v}_h \in \boldsymbol{V}_h : \boldsymbol{v}_h(\boldsymbol{z}) \cdot \boldsymbol{m}_h(\boldsymbol{z}) = 0 \quad \forall \boldsymbol{z} \in \mathcal{N}_h \}.$$

Recall that Π_h denotes the L^2-orthogonal projection onto $\boldsymbol{V}_h$ as defined in (1.57).

Let us consider here the following alternative form of the tangent plane scheme (see also Algorithm 5) based on the alternative formulation (LLA).

Algorithm 14 $\left(\boldsymbol{m}_h^j\right)_{j=0}^J \leftarrow$ `Random tangent plane scheme (alt. 2)`$(\mathcal{T}_h, \boldsymbol{M}^0, J, \tau, \lambda_1, \lambda_2, \theta)$

1: Compute $\boldsymbol{m}_h^0 = \Pi_h \boldsymbol{M}^0$
2: **for** $j = 0, 1, ..., J-1$ **do**
3: Find $\boldsymbol{v}_h^j \in \mathcal{K}_h(\boldsymbol{m}_h^j)$ such that

$$\left\langle \lambda_1 \boldsymbol{v}_h^j + \lambda_2 \boldsymbol{m}_h^j \times \boldsymbol{v}_h^j, \boldsymbol{\phi}_h \right\rangle + \left\langle \nabla \left(\boldsymbol{m}_h^j + k\theta \boldsymbol{v}_h^j \right), \nabla \boldsymbol{\phi}_h \right\rangle = \left\langle \hat{\mathcal{C}}(W(t_j), \boldsymbol{m}_h^j), \boldsymbol{\phi}_h \right\rangle \quad \forall \boldsymbol{\phi}_h \in \mathcal{K}_h(\boldsymbol{m}_h^j). \tag{5.33}$$

4: $\boldsymbol{m}_h^{j+1} = \boldsymbol{m}_h^j + k\boldsymbol{v}_h^j$
5: **end for**

Remark 5.20 (Stability of the tangent plane scheme)**.** *Stability properties of the scheme we proved in [GLT16, Lemma 5.4]:*

$$\left\|\nabla \boldsymbol{m}_h^J\right\|^2 + k\sum_{i=0}^{J-1}\left\|\boldsymbol{v}_h^j\right\|^2 + k^2(2\theta-1)\sum_{j=0}^{J-1}\left\|\nabla \boldsymbol{v}_h^j\right\|^2 \le C \qquad \mathcal{P}\text{-a.s.}$$

where the constant $C>0$ is independent of h and k. This estimate mimics the continuous energy estimate (1.55).

A time-continuous approximation is obtained by affine interpolation of the result of Algorithm 14:

$$\hat{\boldsymbol{m}}_{hk}(t) = \boldsymbol{m}_h^i + (t-t_i)\boldsymbol{v}_h^i \qquad \text{if } t_i \le t < t_{i+1}. \tag{5.34}$$

Let us additionally introduce the following functions

$$\bar{\boldsymbol{m}}_{hk}(t) = \boldsymbol{m}_h^i \qquad \text{if } t_i \le t < t_{i+1},$$
$$\bar{\boldsymbol{v}}_{hk}(t) = \boldsymbol{v}_h^i \qquad \text{if } t_i \le t < t_{i+1}.$$

Observe that, with this definitions, $\partial_t \hat{\boldsymbol{m}}_{hk}(t) = \bar{\boldsymbol{v}}_{hk}(t)$ for $t \in [0,T]\setminus\{t_i\}_{i=1}^{N-1}$.
We take inspiration from [AFKL21] to prove the following convergence result.

Theorem 5.21 (Mean convergence of the tangent plane scheme)**.** *Denote by $\boldsymbol{m}$ the random field solution of the random coefficient LLG equation under the same assumption of Theorem 3.7. Let $\hat{\boldsymbol{m}}_{hk}$ denote the piecewise-affine interpolation* (5.34) *of solution of the tangent plane scheme (Algorithm 14). Assume that there exists a function $\boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk})$ such that*

(A1) $\mathbb{E}\ \sup_{s\in[0,t]}\|\boldsymbol{\phi}_h\|^2 \lesssim \mathbb{E}\int_0^t \|\boldsymbol{e}\|^2_{\mathbb{H}^1(D)} + (h+k)^2 + \mathbb{E}\int_0^t\|\dot{\boldsymbol{e}}\|^2$;

(A2) The random variable $\boldsymbol{\phi}_h$ depends on $W(\cdot)$ exclusively through $(W(t_j))_{j=0}^J$;

(A3) $\boldsymbol{q}_h \coloneqq \dot{\boldsymbol{e}} - \boldsymbol{\phi}_h$ is such that $\mathbb{E}\int_0^t\|\boldsymbol{q}_h\|^2_{\mathbb{H}^1(D)} \lesssim \mathbb{E}\int_0^t\|\boldsymbol{e}\|^2_{\mathbb{H}^1(D)} + (h+k)^2$.

If additionally $\frac{1}{2} < \theta \le 1$, then there exists a constant $C>0$ depending only on the problem data such that

$$\mathbb{E}\,\|\boldsymbol{m}(t) - \hat{\boldsymbol{m}}_{hk}(t)\|_{\mathbb{H}^1(D)} \le C\left(h+k+\mathbb{E}\,\|\boldsymbol{m}(0)-\hat{\boldsymbol{m}}_{hk}(0)\|_{\mathbb{H}^1(D)}\right) \qquad \forall t\in[0,T].$$

Proof. We divide the proof into several steps.

1. *A time-continuous equation for $\hat{\boldsymbol{m}}_{hk}$.* Fix $i=0,\dots,N-1$ and $t_i \le t < t_{i+1}$. Multiply (5.33) by the timestep k for $j<i$, and by $t-t_i$ for $j=i$. Sum them up to obtain

$$\int_0^t \langle \bar{\boldsymbol{v}}_{hk} + \bar{\boldsymbol{m}}_{hk}\times\bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h\rangle + \langle \nabla(\bar{\boldsymbol{m}}_{hk} + k\bar{\boldsymbol{v}}_{hk}), \nabla\boldsymbol{\phi}_h\rangle\,\mathrm{d}s = \int_0^t \left\langle \hat{\mathcal{C}}(\bar{W}_k, \bar{\boldsymbol{m}}_{hk}), \boldsymbol{\phi}_h\right\rangle \mathrm{d}s$$
$$\forall \boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk}).$$

Then, add equal terms on both sides to get: For all $\boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk})$,

$$\begin{aligned}
&\int_0^t \langle \bar{\boldsymbol{v}}_{hk} + \hat{\boldsymbol{m}}_{hk}\times\bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h\rangle + \langle \nabla\hat{\boldsymbol{m}}_{hk}, \nabla\boldsymbol{\phi}_h\rangle\,\mathrm{d}s \\
&= \int_0^t \left\langle \hat{\mathcal{C}}(\bar{W}_k, \hat{\boldsymbol{m}}_{hk}), \boldsymbol{\phi}_h\right\rangle \mathrm{d}s + \int_0^t \langle (s-s.)\bar{\boldsymbol{v}}_{hk}\times\bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h\rangle - \langle (s..-s)\nabla\bar{\boldsymbol{v}}_{hk}, \nabla\boldsymbol{\phi}_h\rangle \\
&\quad - \left\langle (s-s.)\hat{\mathcal{C}}(\bar{W}_k, \bar{\boldsymbol{v}}_{hk}), \boldsymbol{\phi}_h\right\rangle \mathrm{d}s,
\end{aligned} \tag{5.35}$$

where we denoted $s. = t_j$ and $s.. = t_{j+1}$ if $t_j \le s < t_{j+1}$ and

$$\mathcal{K}_h(\bar{\boldsymbol{m}}_{hk}) \coloneqq \left\{ \boldsymbol{v}_h : D_T \to \mathbb{R}^3 : \boldsymbol{v}_h(t) = \boldsymbol{v}_h(t_i) \text{ if } t_i \le t < t_{i+1},\ \boldsymbol{v}_h(t_i) \in \mathcal{K}_h(\boldsymbol{m}_h^i) \right\}. \quad (5.36)$$

As for the terms in the right-hand side, observe that

$$\begin{aligned} \langle (s - s.)\bar{\boldsymbol{v}}_{hk} \times \bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h \rangle &= 0, \\ \int_0^t \langle (s.. - s)\nabla \bar{\boldsymbol{v}}_{hk}, \nabla \boldsymbol{\phi}_h \mathrm{d}s \rangle &\le k \left\| \nabla \bar{\boldsymbol{v}}_{hk} \right\|_{\mathbb{L}^2(D_T)} \left\| \nabla \boldsymbol{\phi}_h \right\|_{\mathbb{L}^2(D_T)}, \\ \int_0^t \left\langle (s - s.)\hat{\mathcal{C}}(\bar{W}_k, \bar{\boldsymbol{v}}_{hk}), \boldsymbol{\phi}_h \right\rangle \mathrm{d}s &\le k \left\| \hat{\mathcal{C}}(\bar{W}_k, \bar{\boldsymbol{v}}_{hk}) \right\|_{\mathbb{L}^2(D_T)} \left\| \boldsymbol{\phi}_h \right\|_{\mathbb{L}^2(D_T)}. \end{aligned}$$

Denote now

$$\int_0^t \langle R_h, \boldsymbol{\phi}_h \rangle \coloneqq \int_0^t \langle (s.. - s)\nabla \bar{\boldsymbol{v}}_{hk}, \nabla \boldsymbol{\phi}_h \rangle - \left\langle (s - s.)\hat{\mathcal{C}}(\bar{W}_k, \bar{\boldsymbol{v}}_{hk}), \boldsymbol{\phi}_h \right\rangle \mathrm{d}s.$$

Estimate the absolute value of the last quantity with Cauchy-Schwarz and Young inequalities with constant $\varepsilon > 0$ to get

$$\left| \int_0^t \langle R_h, \boldsymbol{\phi}_h \rangle \right| \le \frac{k^2}{4\varepsilon} C_{\text{stab}} + \varepsilon \int_0^t \left\| \boldsymbol{\phi}_h \right\|_{\mathbb{H}^1(D)}^2. \quad (5.37)$$

where $C_{\text{stab}} \coloneqq \int_0^t \left(\left\| \nabla \bar{\boldsymbol{v}}_{hk} \right\|^2 + \left\| \hat{\mathcal{C}}(\bar{W}_k, \bar{\boldsymbol{v}}_{hk}) \right\|^2 \right)$ is finite because of (1.78), the assumption $\boldsymbol{g} \in C^{2+\alpha}(D)^3$, and the fact that $\bar{\boldsymbol{v}}_{hk} \in L^2(0, T, \mathbb{H}^1(D))$ (Remark 5.20).

2. *"Consistency" of the Tangent plane scheme.* We test the exact random coefficient LLG equation (1.76) with a function in the discrete tangent plane:

$$\int_0^t \langle \partial_t \boldsymbol{m} + \boldsymbol{m} \times \partial_t \boldsymbol{m}, \boldsymbol{\phi}_h \rangle + \langle \nabla \boldsymbol{m}, \nabla \boldsymbol{\phi}_h \rangle = \int_0^t \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}), \boldsymbol{\phi}_h \right\rangle + \langle D(\boldsymbol{m})\boldsymbol{m}, \boldsymbol{\phi}_h \rangle \quad \forall \boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk}), \quad (5.38)$$

where $D(\boldsymbol{m}) \coloneqq |\nabla \boldsymbol{m}|^2 + \boldsymbol{m} \cdot \hat{\mathcal{C}}(W, \boldsymbol{m})$. By definition of discrete tangent plane (5.36),

$$\langle D(\boldsymbol{m})\boldsymbol{m}, \boldsymbol{\phi}_h \rangle = \langle D(\boldsymbol{m}) \left(\boldsymbol{m} - \bar{\boldsymbol{m}}_{hk} \right), \boldsymbol{\phi}_h \rangle = \langle D(\boldsymbol{m}) \left(\boldsymbol{m} - \hat{\boldsymbol{m}}_{hk} + (t - t.)\bar{\boldsymbol{v}}_{hk} \right), \boldsymbol{\phi}_h \rangle.$$

Young's inequality implies

$$|\langle D(\boldsymbol{m})\boldsymbol{m}, \boldsymbol{\phi}_h \rangle| \le \frac{1}{4\varepsilon} \int_0^t \left\| D(\boldsymbol{m}) \left(\boldsymbol{m} - \hat{\boldsymbol{m}}_{hk} + (t - t.)\bar{\boldsymbol{v}}_{hk} \right) \right\|^2 + \varepsilon \int_0^t \left\| \boldsymbol{\phi}_h \right\|^2.$$

Moreover, simple estimates give

$$\begin{aligned} &\int_0^t \left\| D(\boldsymbol{m}) \left(\boldsymbol{m} - \hat{\boldsymbol{m}}_{hk} + (t - t.)\bar{\boldsymbol{v}}_{hk} \right) \right\|^2 \\ &\qquad \le 2 \left\| D(\boldsymbol{m}) \right\|_{\mathbb{L}^\infty(D_T)}^2 \left(\int_0^t \left\| \boldsymbol{m} - \hat{\boldsymbol{m}}_{hk} \right\|^2 \mathrm{d}s + k^2 \left\| \bar{\boldsymbol{v}}_{hk} \right\|_{\mathbb{L}^2(D_T)}^2 \right) \end{aligned}$$

All in all, we obtained

$$\begin{aligned} \int_0^t |\langle D(\boldsymbol{m})\boldsymbol{m}, \boldsymbol{\phi}_h \rangle| \le &\frac{1}{2\varepsilon} \left\| D(\boldsymbol{m}) \right\|_{\mathbb{L}^\infty(D_T)}^2 \left(\int_0^t \left\| \boldsymbol{m} - \hat{\boldsymbol{m}}_{hk} \right\|^2 \mathrm{d}s + k^2 \left\| \bar{\boldsymbol{v}}_{hk} \right\|_{\mathbb{L}^2(D_T)}^2 \right) \\ &+ \varepsilon \int_0^t \left\| \boldsymbol{\phi}_h \right\|^2. \end{aligned} \quad (5.39)$$

Observe that $\|D(\boldsymbol{m})\|_{\mathbb{L}^\infty(D_T)}$ is finite since

$$\|D(\boldsymbol{m})\|_{\mathbb{L}^\infty(D_T)} \lesssim \|\nabla \boldsymbol{m}\|^2_{\mathbb{L}^\infty(D_T)} + \left(1 + \|\boldsymbol{g}\|_{[C^{2+\alpha}(D)]^3}\right)^4 \left(\|\boldsymbol{m}\|_{\mathbb{L}^\infty(D_T)} + \|\nabla \boldsymbol{m}\|_{\mathbb{L}^\infty(D_T)}\right),$$

where $\boldsymbol{g} \in \left[C^{2+\alpha}(D)\right]^3$ by assumption, $\boldsymbol{m} \in \left[C^{1+\alpha/2,2+\alpha}(D_T)\right]^3$ as proved in Theorem 3.7, and $\bar{\boldsymbol{v}}_{hk} \in \mathbb{L}^2(D_T)$ was proved above.

3. *Error equation.* We subtract (5.35) from (5.38) to obtain:

$$\begin{aligned}\int_0^t \langle \partial_t \boldsymbol{m} - \bar{\boldsymbol{v}}_{hk} + \boldsymbol{m} \times \partial_t \boldsymbol{m} - \hat{\boldsymbol{m}}_{hk} \times \bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h \rangle + \langle \nabla (\boldsymbol{m} - \hat{\boldsymbol{m}}_{hk}), \nabla \boldsymbol{\phi}_h \rangle \\ = \int_0^t \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}) - \hat{\mathcal{C}}(\bar{W}_k, \hat{\boldsymbol{m}}_{hk}) + D(\boldsymbol{m})\boldsymbol{m} - R_h, \boldsymbol{\phi}_h \right\rangle \qquad \forall \boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk}).\end{aligned} \tag{5.40}$$

Denote

$$\boldsymbol{e} \coloneqq \boldsymbol{m} - \hat{\boldsymbol{m}}_{hk}, \qquad \dot{\boldsymbol{e}} \coloneqq \partial_t \boldsymbol{e} = \partial_t \boldsymbol{m} - \bar{\boldsymbol{v}}_{hk}.$$

The previous equation can be rewritten as:

$$\begin{aligned}\int_0^t \langle \dot{\boldsymbol{e}} + \boldsymbol{m} \times \dot{\boldsymbol{e}} + \boldsymbol{e} \times \bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h \rangle + \langle \nabla \boldsymbol{e}, \nabla \boldsymbol{\phi}_h \rangle = \int_0^t \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}) - \hat{\mathcal{C}}(\bar{W}_k, \hat{\boldsymbol{m}}_{hk}), \boldsymbol{\phi}_h \right\rangle \\ + \int_0^t \langle D(\boldsymbol{m})\boldsymbol{m} - R_h, \boldsymbol{\phi}_h \rangle \qquad \forall \boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk}).\end{aligned} \tag{5.41}$$

4. *Energy estimate.* We would now like to test the error equation with $\dot{\boldsymbol{e}}$. However, this does not belong to the discrete tangent plane $\mathcal{K}_h(\bar{\boldsymbol{m}}_{hk})$. Instead, we test with $\boldsymbol{\phi}_h \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk})$, which satisfies assumptions (A1)-(A3) from the statement. We add $\int_0^t \langle \dot{\boldsymbol{e}}, \dot{\boldsymbol{e}} \rangle + \langle \nabla \boldsymbol{e}, \nabla \dot{\boldsymbol{e}} \rangle$ to both sides of (5.41). Recall that $\boldsymbol{q}_h \coloneqq \dot{\boldsymbol{e}} - \boldsymbol{\phi}_h$ (A3) to obtain

$$\begin{aligned}\int_0^t \|\dot{\boldsymbol{e}}\|^2 + \langle \nabla \boldsymbol{e}, \nabla \dot{\boldsymbol{e}} \rangle = \int_0^t \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}) - \hat{\mathcal{C}}(\bar{W}_k, \hat{\boldsymbol{m}}_{hk}) + D(\boldsymbol{m})\boldsymbol{m} - R_h - \boldsymbol{e} \times \bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h \right\rangle \\ + \int_0^t \langle \dot{\boldsymbol{e}} - \boldsymbol{m} \times \dot{\boldsymbol{e}}, \boldsymbol{q}_h \rangle + \langle \nabla \boldsymbol{e}, \nabla \boldsymbol{q}_h \rangle,\end{aligned}$$

where we observed that $\langle \boldsymbol{m} \times \dot{\boldsymbol{e}}, \dot{\boldsymbol{e}} \rangle = 0$. Since $\langle \nabla \boldsymbol{e}, \nabla \dot{\boldsymbol{e}} \rangle = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|\nabla \boldsymbol{e}(t)\|^2$, the fundamental theorem of calculus gives

$$\int_0^t \langle \nabla \boldsymbol{e}, \nabla \dot{\boldsymbol{e}} \rangle = \frac{1}{2} \|\nabla \boldsymbol{e}(t)\|^2 - \frac{1}{2} \|\nabla \boldsymbol{e}(0)\|^2.$$

We can finally take the the expectation both sides to obtain the energy equation

$$\begin{aligned}\mathbb{E} \int_0^t \|\dot{\boldsymbol{e}}\|^2 + \frac{1}{2}\mathbb{E} \|\nabla \boldsymbol{e}(t)\|^2 = \frac{1}{2}\mathbb{E} \|\nabla \boldsymbol{e}(0)\|^2 + \mathbb{E} \int_0^t \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}) - \hat{\mathcal{C}}(\bar{W}_k, \hat{\boldsymbol{m}}_{hk}), \boldsymbol{\phi}_h \right\rangle \\ + \mathbb{E} \int_0^t \langle D(\boldsymbol{m})\boldsymbol{m}, \boldsymbol{\phi}_h \rangle - \mathbb{E} \int_0^t \langle R_h, \boldsymbol{\phi}_h \rangle \\ - \mathbb{E} \int_0^t \langle \boldsymbol{e} \times \bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h \rangle + \mathbb{E} \int_0^t \left(\langle \dot{\boldsymbol{e}} - \boldsymbol{m} \times \dot{\boldsymbol{e}}, \boldsymbol{q}_h \rangle + \langle \nabla \boldsymbol{e}, \nabla \boldsymbol{q}_h \rangle \right).\end{aligned} \tag{5.42}$$

5. *Estimating the right-hand side of the energy equation* (5.42). Let us estimate one term at a time.

Second term (5.42) : Recall that $\hat{\mathcal{C}}(s, \boldsymbol{v}) = \sum_{i=1}^{6} F_i(\boldsymbol{v}) b_i(W)$ (3.12). The functions b_i are uniformly bounded with uniformly bounded derivatives up to second order, and the functions F_i are linear, globally Lipschitz with Lipschitz constants $L > 0$ and bounded with constant $1 + L$:

$$\|b_i\|_{L^\infty(\mathbb{R})} \leq \beta, \quad \left\|b_i'\right\|_{L^\infty(\mathbb{R})} \leq \beta, \quad \left\|b_i''\right\|_{L^\infty(\mathbb{R})} \leq \beta, \quad \|F_i(\boldsymbol{u}) - F_i(\boldsymbol{v})\| \leq L \|u - v\|_{\mathbb{H}^1(D)}.$$

We rewrite the second term of (5.42) as

$$\mathbb{E} \int_0^t \left\langle \hat{\mathcal{C}}(W, \boldsymbol{m}) - \hat{\mathcal{C}}(\bar{W}_k, \hat{\boldsymbol{m}}_{hk}), \boldsymbol{\phi}_h \right\rangle = \mathbb{E} \int_0^t \left\langle \sum_i F_i(\boldsymbol{m}) b_i(W) - F_i(\hat{\boldsymbol{m}}_{hk}) b_i(\bar{W}_k), \boldsymbol{\phi}_h \right\rangle.$$

By adding and subtracting appropriate quantities, we get

$$\begin{aligned} &\mathbb{E} \int_0^t \left\langle \sum_i F_i(\boldsymbol{m}) b_i(W) - F_i(\hat{\boldsymbol{m}}_{hk}) b_i(\bar{W}_k), \boldsymbol{\phi}_h \right\rangle \\ &= \sum_i \mathbb{E} \int_0^t \langle (F_i(\boldsymbol{m}) - F_i(\hat{\boldsymbol{m}}_{hk}))\, b_i(W), \boldsymbol{\phi}_h \rangle + \sum_i \mathbb{E} \int_0^t \left\langle F_i(\hat{\boldsymbol{m}}_{hk}) \left(b_i(W) - b_i(\bar{W}_k)\right), \boldsymbol{\phi}_h \right\rangle. \end{aligned} \tag{5.43}$$

As for the first expectation in the right-hand side, apply Young inequality with constant $\varepsilon > 0$ to get

$$\begin{aligned} \mathbb{E} \int_0^t \langle (F_i(\boldsymbol{m}) - F_i(\hat{\boldsymbol{m}}_{hk}))\, b_i(W), \boldsymbol{\phi}_h \rangle \leq &\varepsilon \mathbb{E} \int_0^t \|\boldsymbol{\phi}_h\|^2 \\ &+ \frac{1}{4\varepsilon} \mathbb{E} \int_0^t \|(F_i(\boldsymbol{m}) - F_i(\hat{\boldsymbol{m}}_{hk}))\, b_i(W)\|^2. \end{aligned}$$

Use the boundedness of b_i and Lipschitz continuity of F_i to bound the second term:

$$\mathbb{E} \left| \int_0^t \langle (F_i(\boldsymbol{m}) - F_i(\hat{\boldsymbol{m}}_{hk}))\, b_i(W), \boldsymbol{\phi}_h \rangle \right| \leq \varepsilon \mathbb{E} \int_0^t \|\boldsymbol{\phi}_h\|^2 + \frac{\beta^2 L^2}{4\varepsilon} \mathbb{E} \int_0^t \|\boldsymbol{e}\|^2_{\mathbb{H}^1(D)}. \tag{5.44}$$

As for the second expectation in the right-hand side of (5.43), a Taylor expansion gives: For any $s \in [0, T]$, $j \in \mathbb{N}_0$ such that $t_j \leq s \leq t_{j+1}$ there exists $\xi = \xi(s) \in [t_j, s]$ such that

$$b_i(W(s)) - b_i(W(t_j)) = b_i'(W(t_j))\,(W(s) - W(t_j)) + b_i''(\xi) \frac{(W(s) - W(t_j))^2}{2}.$$

Let $J \in \mathbb{N}$ such that $t_J \leq t < t_{J+1}$, recall that $\hat{\boldsymbol{m}}_{hk}(t) = \bar{\boldsymbol{m}}_{hk} + (t - t.)\bar{\boldsymbol{v}}_{hk}$ (5.34), and

that F_i is linear to obtain:

$$\begin{aligned}
&\mathbb{E}\int_0^t \left\langle F_i(\hat{\boldsymbol{m}}_{hk})\left(b_i(W)-b_i(\bar{W}_k)\right),\boldsymbol{\phi}_h\right\rangle\\
&\leq\mathbb{E}\left|\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle b_i'(W(t_j))\int_{t_j}^{t_{j+1}} W(s)-W(t_i)\mathrm{d}s\right|\\
&+\mathbb{E}\left|\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{v}_h^j),\boldsymbol{\phi}_h\right\rangle b_i'(W(t_j))\int_{t_j}^{t_{j+1}}(s-t_j)\left(W(s)-W(t_i)\right)\mathrm{d}s\right|\\
&+\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(W(s)-W(t_i))^2 b''(\xi(s))\mathrm{d}s\right|\\
&+\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{v}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(s-t_j)(W(s)-W(t_i))^2 b''(\xi(s))\mathrm{d}s\right|\\
&+\mathbb{E}\left|\int_{t_J}^t\langle\left(b_i(W(s))-b_i(W(t_{J-1}))\right)F_i(\hat{\boldsymbol{m}}_{hk}),\boldsymbol{\phi}_h\rangle\,\mathrm{d}s\right|.
\end{aligned}$$

We estimate one term of the right-hand side at a time.

• *First expectation*: Thanks to (A2), this falls in the setting of Lemma 5.18 with $X_j=\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle b_i'(W(t_j))$, which gives

$$\begin{aligned}
&\mathbb{E}\left|\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle b_i'(W(t_j))\int_{t_j}^{t_{j+1}} W(s)-W(t_i)\mathrm{d}s\right|\\
&\lesssim k\sqrt{\mathbb{E}\left[k\sum_{j=0}^{J-1}\left(\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h^j\right\rangle b_i'(W(t_j))\right)^2\right]}.
\end{aligned}$$

Observe now that there exists $M>0$ such that

$$\left\|\boldsymbol{m}_h^j\right\|_{\mathbb{H}^1(D)}\leq\left\|\boldsymbol{m}^0\right\|_{\mathbb{H}^1(D)}+\|\bar{\boldsymbol{v}}_{hk}\|_{L^1(0,T,\mathbb{H}^1(D))}=:M\qquad\forall j=0,\dots,J. \tag{5.45}$$

This follows from the following facts: $\boldsymbol{m}_h^j=\boldsymbol{m}_h^0+k\sum_{\ell=1}^j\boldsymbol{v}_h^\ell$, $\boldsymbol{m}_h^0=\Pi_h\boldsymbol{m}^0$ (the finite element projection of $\boldsymbol{m}^0$), and $\|\bar{\boldsymbol{v}}_{hk}\|_{L^1(0,T,\mathbb{H}^1(D))}$ is uniformly bounded with respect to h,k (a consequence of the fact that the same holds for the $L^2(0,T,\mathbb{H}^1(D))$-norm). The assumptions on b_i and F_i give

$$\sum_{j=0}^{J-1}\left(\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h^j\right\rangle b_i'(W(t_j))\right)^2\leq\beta^2(1+L)^2M^2\sum_{j=0}^{J-1}\left\|\boldsymbol{\phi}_h^j\right\|^2.$$

These facts followed by a Young inequality with constant $\varepsilon>0$ give

$$\begin{aligned}
&\mathbb{E}\left|\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle b_i'(W(t_j))\int_{t_j}^{t_{j+1}} W(s)-W(t_i)\mathrm{d}s\right|\\
&\lesssim k\beta(1+L)M\sqrt{\mathbb{E}\int_0^{t_J}\|\boldsymbol{\phi}_h\|^2}\leq\frac{(\beta(1+L)M)^2}{4\varepsilon}k^2+\varepsilon\mathbb{E}\int_0^{t_J}\|\boldsymbol{\phi}_h\|^2.
\end{aligned}$$

• *Second expectation*: Analogous computations, culminating in the use of the second estimate in Lemma 5.18, give

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{v}_h^j),\boldsymbol{\phi}_h^j\right\rangle b_i'(W(t_j))\int_{t_j}^{t_{j+1}}(s-t_j)\,(W(s)-W(t_i))\,\mathrm{d}s\right|$$
$$\lesssim k^2\sqrt{\mathbb{E}\left[k\sum_{j=0}^{J-1}\left(\left\langle F_i(\boldsymbol{v}_h^j),\boldsymbol{\phi}_h^j\right\rangle b_i'(W(t_j))\right)^2\right]}.$$

We apply again the uniform boundedness of b_i and continuity of F_i. The definition of $\boldsymbol{m}_h^{j+1}$ in Algorithm 14 implies that $\boldsymbol{v}_h^j=\frac{\boldsymbol{m}_h^{j+1}-\boldsymbol{m}_h^j}{k}$ for any $j=0,\dots,J-1$. Thus,

$$\max_{j=0,\dots,J-1}\left\|\boldsymbol{v}_h^j\right\|_{\mathbb{H}^1(D)}\leq\frac{2}{k}\max_{j=0,\dots,J}\left\|\boldsymbol{m}_{hk}^j\right\|_{\mathbb{H}^1(D)}\leq\frac{2}{k}M. \tag{5.46}$$

One of the two factors k can be simplified to obtain

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{v}_h^j),\boldsymbol{\phi}_h^j\right\rangle b_i'(W(t_j))\int_{t_j}^{t_{j+1}}(s-t_j)\,(W(s)-W(t_i))\,\mathrm{d}s\right|$$
$$\lesssim\frac{(\beta(1+L)M)^2}{\varepsilon}k^2+\varepsilon\mathbb{E}\int_0^{t_J}\|\boldsymbol{\phi}_h\|^2,$$

under the same regularity assumptions needed for the previous term.

• *Third expectation*:: We treat it using facts from the analogous term in the proof of Proposition 5.19 and from the First expectation above. The uniform boundedness of b'' and the Hölder continuity of the Wiener process sample paths imply

$$\left|\frac{1}{2}\int_{t_j}^{t_{j+1}}(W(s)-W(t_j))^2b''(\xi(s))\mathrm{d}s\right|\leq\beta C_W^2\frac{k^2}{4},$$

This implies

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(W(s)-W(t_j))^2b''(\xi(s))\mathrm{d}s\right|$$
$$\leq\frac{\beta C_W^2}{4}k\;\mathbb{E}\left[\sum_{j=0}^{J-1}k\left|\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle\right|\right]. \tag{5.47}$$

A Cauchy-Schwarz inequality gives

$$\sum_{j=0}^{J-1}k\left|\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle\right|\leq\sqrt{\sum_{j=0}^{J-1}\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle^2k^2\sqrt{J}}.$$

Observe now that $\sqrt{k}\sqrt{J}=\sqrt{T}$, recall the boundedness of F_i, and the bound on $\boldsymbol{m}_h^j$ (5.45). Another Cauchy-Schwarz inequality implies

$$\sum_{j=0}^{J-1}k\left|\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle\right|$$
$$\leq\sqrt{T}(1+L)M\sqrt{\int_0^{t_J}\|\phi_h\|^2}.$$

Returning to (5.47), we obtain,

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(W(s)-W(t_i))^2b''(\xi(s))\mathrm{d}s\right| \leq \frac{\beta C_W^2\sqrt{T}(1+L)M}{4}k\,\mathbb{E}\sqrt{\int_0^{t_J}\|\phi_h\|^2}.$$

A Young inequality with constant $\varepsilon>0$ implies

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(W(s)-W(t_i))^2b''(\xi(s))\mathrm{d}s\right| \leq \frac{\left(\beta C_W^2\sqrt{T}(1+L)M\right)^2}{64\varepsilon}k^2 + \varepsilon\mathbb{E}\int_0^{t_J}\|\phi_h\|^2.$$

- *Fourth expectation*:: We proceed analogously to the previous term. Let us highlight the differences. From the boundedness of b'' and Wiener process sample paths Hölder continuity, we get.

$$\left|\frac{1}{2}\int_{t_j}^{t_{j+1}}(s-t_j)(W(s)-W(t_j))^2b''(\xi(s))\mathrm{d}s\right| \leq \beta C_W^2\frac{k^3}{6}.$$

Recall the bound (5.46) of $\left\|\boldsymbol{v}_h^j\right\|$ for any $j=0,\ldots,J$ (we use it in place of the bound (5.45) of $\boldsymbol{m}_h^j$). Analogously to the previous term, we estimate

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{v}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(s-t_j)(W(s)-W(t_i))^2b''(\xi(s))\mathrm{d}s\right| \leq \frac{\beta C_W^2\sqrt{T}(1+L)M}{3}k\,\mathbb{E}\sqrt{\int_0^{t_J}\|\phi_h\|^2}.$$

Again, a Young inequality with constant $\varepsilon>0$ gives

$$\mathbb{E}\left|\sum_{j=0}^{J-1}\frac{\left\langle F_i(\boldsymbol{m}_h^j),\boldsymbol{\phi}_h\right\rangle}{2}\int_{t_j}^{t_{j+1}}(W(s)-W(t_i))^2b''(\xi(s))\mathrm{d}s\right| \leq \frac{\left(\beta C_W^2\sqrt{T}(1+L)M\right)^2}{36\varepsilon}k^2 + \varepsilon\mathbb{E}\int_0^{t_J}\|\phi_h\|^2.$$

- *Fifth expectation*: Apply Young inequality with constant $\varepsilon>0$ to obtain

$$\mathbb{E}\left|\int_{t_J}^{t}\langle(b_i(W(s))-b_i(W(t_{J+1})))\,F_i(\hat{\boldsymbol{m}}_{hk}),\boldsymbol{\phi}_h\rangle\,\mathrm{d}s\right| \leq \frac{1}{4\varepsilon}\mathbb{E}\int_{t_J}^{t}\left\|(b_i(W(s))-b_i(W(t_{J+1})))\,F_i(\hat{\boldsymbol{m}}_{hk}^j)\right\|^2+\varepsilon\mathbb{E}\int_{t_J}^{t}\|\boldsymbol{\phi}_h\|^2.$$

Estimate the first term using the Lipschitz continuity of b_i (with constant $\beta > 0$), the Hölder continuity of W (with exponent $\frac{1}{2}$ and constant $C_W > 0$ almost surely finite), the continuity of F_i (with Lipschitz constant $L > 0$), and the boundedness of $\hat{\boldsymbol{m}}_{hk}$:

$$\mathbb{E}\left|\int_{t_J}^{t}\left\langle \left(b_i(W(s)) - b_i(W(t_J))\right) F_i(\hat{\boldsymbol{m}}_{hk}^{j}), \boldsymbol{\phi}_h\right\rangle \mathrm{d}s\right| \leq \frac{(\beta C_W L M)^2}{8\varepsilon} k^2 + \varepsilon\mathbb{E}\int_{t_J}^{t}\|\boldsymbol{\phi}_h\|^2 .$$

All in all, we have obtained:

$$\begin{aligned}\mathbb{E}\int_0^t\left\langle \hat{\mathcal{C}}(W,\boldsymbol{m}) - \hat{\mathcal{C}}(\bar{W}_k,\hat{\boldsymbol{m}}_{hk}), \boldsymbol{\phi}_h\right\rangle \leq & \frac{2\left(\beta(1+L)(1+M)(1+C_W)(1+\sqrt{T})\right)^2}{\varepsilon} k^2 \\ & + \varepsilon\mathbb{E}\int_0^t\|\boldsymbol{\phi}_h\|^2 .\end{aligned}$$

Third term (5.42): Estimate as in (5.39) to obtain

$$\int_0^t |\langle D(\boldsymbol{m})\boldsymbol{m}, \boldsymbol{\phi}_h\rangle| \leq \frac{1}{2\varepsilon}\|D(\boldsymbol{m})\|_{L^\infty(D_T)}^2\left(\int_0^t\|\boldsymbol{e}\|^2 + \|\bar{\boldsymbol{v}}_{hk}\|_{\mathbb{L}^2(D_T)}^2 k\right) + \varepsilon\int_0^t\|\boldsymbol{\phi}_h\|^2 .$$

Fourth term (5.42): Estimate as in (5.37) to obtain

$$\int_0^t\langle R_h, \boldsymbol{\phi}_h\rangle \leq \frac{k^2}{4\varepsilon}C_{\mathrm{stab}} + \varepsilon\int_0^t\|\boldsymbol{\phi}_h\|^2 \leq \frac{k^2}{4\varepsilon}C_{\mathrm{stab}} + \varepsilon\int_0^t\|\boldsymbol{\phi}_h\|^2 .$$

Fifth term (5.42): Apply a three-terms Hölder inequality and the continuity of the embedding $\mathbb{H}^1(D) \hookrightarrow \mathbb{L}^4(D)$ to obtain

$$\left|\int_0^t\langle \boldsymbol{e}\times\bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h\rangle\right| \leq \sqrt{\int_0^t\|\boldsymbol{e}\|_{\mathbb{H}^1(D)}^2}\sqrt{\int_0^t\|\bar{\boldsymbol{v}}_{hk}\|_{\mathbb{H}^1(D)}^2}\sup_{[0,t]}\|\boldsymbol{\phi}_h\| .$$

A Young inequality with constant $\varepsilon > 0$ gives

$$\left|\int_0^t\langle \boldsymbol{e}\times\bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h\rangle\right| \leq \frac{1}{4\varepsilon}\|\bar{\boldsymbol{v}}_{hk}\|_{L^2(0,T,\mathbb{H}^1(D))}^2\int_0^t\|\boldsymbol{e}\|_{\mathbb{H}^1(D)}^2 + \varepsilon\sup_{[0,t]}\|\boldsymbol{\phi}_h\|^2 .$$

Recalling the stability of the tangent plane scheme (Remark 5.20), we have that $\|\bar{\boldsymbol{v}}_{hk}\|_{L^2(0,T,\mathbb{H}^1(D))}$ is finite, uniformly with respect to h and k.

Sixth term (5.42): We apply a Young inequality with constant $\varepsilon > 0$ to get

$$\begin{aligned}\mathbb{E}\int_0^t\langle \dot{\boldsymbol{e}} - \boldsymbol{m}\times\dot{\boldsymbol{e}}, \boldsymbol{q}_h\rangle + \langle\nabla\boldsymbol{e}, \nabla\boldsymbol{q}_h\rangle \leq & \varepsilon\mathbb{E}\int_0^t\left(\|\dot{\boldsymbol{e}} - \boldsymbol{m}\times\dot{\boldsymbol{e}}\|^2 + \|\nabla\boldsymbol{e}\|^2\right) \\ & + \frac{1}{4\varepsilon}\mathbb{E}\int_0^t\left(\|\boldsymbol{q}_h\|^2 + \|\nabla\boldsymbol{q}_h\|^2\right) .\end{aligned}$$

For the first term, observe that $\|\dot{\boldsymbol{e}} - \boldsymbol{m}\times\dot{\boldsymbol{e}}\|^2 \leq (1+|D|)^2\|\dot{\boldsymbol{e}}\|^2$ from the unit-modulus property of $\boldsymbol{m}$. We estimate the second summand with Assumption (A3). Thus, we get

$$\begin{aligned}\mathbb{E}\int_0^t\left(\langle \dot{\boldsymbol{e}} - \boldsymbol{m}\times\dot{\boldsymbol{e}}, \boldsymbol{q}_h\rangle + \langle\nabla\boldsymbol{e}, \nabla\boldsymbol{q}_h\rangle\right) \leq & \varepsilon(1+|D|)^2\,\mathbb{E}\int_0^t\|\dot{\boldsymbol{e}}\|^2 + \varepsilon\mathbb{E}\int_0^t\|\boldsymbol{e}\|_{\mathbb{H}^1(D)}^2 \\ & + \frac{1}{4\varepsilon}\mathbb{E}\int_0^t\left(\|\boldsymbol{q}_h\|^2 + \|\nabla\boldsymbol{q}_h\|^2\right) .\end{aligned}$$

Finally, we put together all previous estimates and use Assumptions (A1), (A3) on the test function $\boldsymbol{\phi}_h$ to obtain: There exists constants $C_1, C_2, C_3 > 0$ and $\varepsilon > 0$ such that $1 - C_1\varepsilon > 0$ and

$$\frac{1}{2}\mathbb{E}\left[\|\nabla \boldsymbol{e}(t)\|^2\right] + (1 - C_1\varepsilon)\mathbb{E}\int_0^t \|\dot{\boldsymbol{e}}\|^2 \leq \frac{1}{2}\mathbb{E}\left[\|\nabla \boldsymbol{e}(0)\|^2\right] + \mathbb{E}\int_0^t C_2 \|\boldsymbol{e}(s)\|^2_{\mathbb{H}^1(D)}\,\mathrm{d}s \tag{5.48}$$

$$+C_3(h+k)^2. \tag{5.49}$$

6. *Conclusion.* Observe now that

$$\|\boldsymbol{e}(t)\|^2 - \|\boldsymbol{e}(0)\|^2 \leq \|\boldsymbol{e}(t) - \boldsymbol{e}(0)\|^2 = \left\|\int_0^t \dot{\boldsymbol{e}}(s)\mathrm{d}s\right\|^2 \leq \int_0^t \|\dot{\boldsymbol{e}}(s)\|^2\,\mathrm{d}s.$$

So we can estimate the left-hand side of (5.48) from below to obtain

$$\frac{1}{2}\mathbb{E}\left[\|\nabla \boldsymbol{e}(t)\|^2\right] + (1 - C_1\varepsilon)\mathbb{E}\left[\|\boldsymbol{e}(t)\|^2\right] \leq \frac{1}{2}\mathbb{E}\left[\|\nabla \boldsymbol{e}(0)\|^2\right] + (1 - C_1\varepsilon)\mathbb{E}\left[\|\boldsymbol{e}(0)\|^2\right]$$
$$+ \mathbb{E}\int_0^t C_2 \|\boldsymbol{e}(s)\|^2_{\mathbb{H}^1(D)}\,\mathrm{d}s, +C_3(k+h)^2,$$

or

$$\min\left(\frac{1}{2}, 1 - C_1\varepsilon\right)\mathbb{E}\,\|\boldsymbol{e}(t)\|^2_{\mathbb{H}^1(D)} \leq \max\left(\frac{1}{2}, 1 - C_1\varepsilon\right)\mathbb{E}\,\|\boldsymbol{e}(0)\|^2_{\mathbb{H}^1(D)}$$
$$+ \mathbb{E}\int_0^t C_2 \|\boldsymbol{e}(s)\|^2_{\mathbb{H}^1(D)}\,\mathrm{d}s + C_3(k+h)^2.$$

Finally, Gronwall's lemma gives the statement.

□

Let us conclude by making some remarks on the previous theorem and its proof.

Remark 5.22.

- *In Assumption (A1), the supremum appearing in the left-hand side is needed to bound the trilinear form* $\int_0^t \langle \boldsymbol{e} \times \bar{\boldsymbol{v}}_{hk}, \boldsymbol{\phi}_h\rangle$ *in* Step 5 *of the proof above. For all occurrences of* $\boldsymbol{\phi}_n$*, the following weaker assumption would suffice:*

  (A1') $\mathbb{E}\int_0^t \|\boldsymbol{\phi}_h\|^2 \lesssim \mathbb{E}\int_0^t \|\boldsymbol{e}\|^2_{\mathbb{H}^1(D)} + (h+k)^2 + \varepsilon\mathbb{E}\int_0^t \|\dot{\boldsymbol{e}}\|^2$*;*

- *Proving the existence of the test function* $\boldsymbol{\phi}$ *satisfying (A1)-(A3) is not trivial (even if (A1) is replaced by (A1')). The following is a somewhat natural attempt that however fails. Denoting* $\mathcal{P}_{hk}$ *the* L^2*-orthogonal projection onto* $\mathcal{K}_h(\bar{\boldsymbol{m}}_{hk})$*, the choice* $\boldsymbol{\phi}_h := \mathcal{P}_{hk}[\partial_t \boldsymbol{m}] - \bar{\boldsymbol{v}}_{hk} \in \mathcal{K}_h(\bar{\boldsymbol{m}}_{hk})$ *does* not *work because (A2) is not satisfied. On the other hand, this choice would give*

$$\|\boldsymbol{\phi}_h\|_{\mathbb{L}^2(D_T)} \leq \|\dot{\boldsymbol{e}}\|_{\mathbb{L}^2(D_T)}, \tag{5.50}$$

  since $\boldsymbol{q}_h = \partial_t \boldsymbol{m} - \mathcal{P}_{hk}\partial_t \boldsymbol{m}$ *and* $\|\dot{\boldsymbol{e}}\|^2 = \|\boldsymbol{\phi}_h\|^2 + \|(1 - \mathcal{P}_{hk})\partial_t \boldsymbol{m}\|^2$*.*

- *Observe that the estimate in* (5.48) *motivates the choice of the specific form of Assumption (A3): Had we assumed the presence of* $\varepsilon\mathbb{E}\int_0^t \|\dot{\boldsymbol{e}}\|^2$ *in the right hand side, we would have ended up with a term* $\frac{1}{4\varepsilon}\mathbb{E}\int_0^t \|\dot{\boldsymbol{e}}\|^2$ *that cannot be made arbitrarily small when multiplied by* ε*.*

Bibliography

[ABW22] Ben Adcock, Simone Brugiapaglia, and Clayton G. Webster. *Sparse Polynomial Approximation of High-Dimensional Functions*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2022.

[AdBH14] 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.

[ADF+24] Xin An, Josef Dick, Michael Feischl, Andrea Scaglioni, and Thanh Tran. Sparse grid approximation of nolinear SPDEs: The Landau–Lifshitz–Gilbert equation, 2024.

[AFKL21] 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.

[AG07] Søren Asmussen and Peter W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[AJ06] François Alouges and Pascal Jaisson. Convergence of a finite element discretization for the Landau-Lifshitz equations in micromagnetism. *Math. Models Methods Appl. Sci.*, 16(2):299–316, 2006.

[AKW+13] Dmytro Apalkov, Alexey Khvalkovskiy, Steven Watts, Vladimir Nikitin, Xueti Tang, Daniel Lottis, Kiseok Moon, Xiao Luo, Eugene Chen, Adrian Ong, et al. Spin-transfer torque magnetic random access memory (stt-mram). *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 9(2):1–35, 2013.

[Alo08] François Alouges. A new finite element scheme for Landau–Lifchitz equations. *Discrete Contin. Dyn. Syst. Ser. S*, 1(2):187–196, 2008.

[AMM+12] Fumiko Akagi, Masaki Mukoh, Masafumi Mochizuki, Junko Ushiyama, Takuya Matsumoto, and Harukazu Miyamoto. Thermally assisted magnetic recording with bit-patterned media to achieve areal recording density beyond 5tb/in2. *Journal of Magnetism and Magnetic Materials*, 324(3):309–313, 2012. Selected papers of the 9th Perpendicular Magnetic Recording Conference(PMRC 2010).

[AS92] François Alouges and Alain Soyeur. On global weak solutions for Landau-Lifshitz equations: existence and nonuniqueness. *Nonlinear Anal.*, 18(11):1071–1084, 1992.

[Bab60] Ivan K. Babenko. Approximation of a certain class of periodical functions of many variables by trigonometric polynomials. *Dokl. Akad. Nauk SSSR*, 132:982–985, 1960.

[Bac23] Markus Bachmayr. Low-rank tensor methods for partial differential equations. *Acta Numerica*, 32:1–121, 2023.

[Bar05] Sören Bartels. Stability and convergence of finite-element approximation schemes for harmonic maps. *SIAM J. Numer. Anal.*, 43(1):220–238, 2005.

[BASL99] Ivo Babuška, Börje Andersson, Paul J. Smith, and Klas Levin. Damage analysis of fiber composites. I. Statistical analysis on fiber scale. *Comput. Methods Appl. Mech. Engrg.*, 172(1-4):27–77, 1999.

[BBNP14a] 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.

[BBNP14b] Ľubomír Baňas, Zdzisław Brzeźniak, Mikhail Neklyudov, and Andreas Prohl. *Stochastic Ferromagnetism.* De Gruyter, Berlin, Boston, 2014.

[BBP13] 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.

[BBRS15] Jean-Luc Bouchot, Benjamin Bykowski, Holger Rauhut, and Christoph Schwab. Compressed sensing petrov-galerkin approximations for parametric PDEs. In *2015 International Conference on Sampling Theory and Applications (SampTA)*, pages 528–532, 2015.

[BCD17] Markus Bachmayr, Albert Cohen, and Wolfgang Dahmen. Parametric PDEs: sparse or low-rank approximations? *IMA Journal of Numerical Analysis*, 38(4):1661–1708, 09 2017.

[BCDM17] 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.

[Ber07] 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.

[BG04] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[BGJ12] Zdzisław Brzeźniak, Beniamin Goldys, and Terence Jegaraj. Weak Solutions of a Stochastic Landau–Lifshitz–Gilbert Equation. *Applied Mathematics Research eXpress*, 2013(1):1–33, 04 2012.

[BGJ17] 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.

[BJ63] William Fuller Brown Jr. Thermal fluctuations of a single-domain particle. *Physical review*, 130(5):1677, 1963.

[BK59] R. Bellman and R. Kalaba. On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2):1–9, 1959.

[BKF+20] Robin Bläsing, Asif Ali Khan, Panagiotis Ch. Filippou, Chirag Garg, Fazal Hameed, Jeronimo Castrillon, and Stuart S. P. Parkin. Magnetic racetrack memory: From physics to the cusp of applications within a decade. *Proceedings of the IEEE*, 108(8):1303–1321, 2020.

[BKW23] Sören Bartels, Balázs Kovács, and Zhangxian Wang. Error analysis for the numerical approximation of the harmonic map heat flow with nodal constraints. *IMA Journal of Numerical Analysis*, 44(2):633–653, 06 2023.

[BL16] Zdzisław Brzeźniak and Liang Li. Weak solutions of the Stochastic Landau–Lifshitz–Gilbert Equations with nonzero anisotrophy energy. *Applied Mathematics Research eXpress*, 2016(2):334–375, 09 2016.

[BM05] Giorgio Bertotti and Isaak D. Mayergoyz. *The science of hystheresis - Volume 2: Physical modelling, micromagnetics, and magnetization dynamic*. Elsevier academic press, Berlin, 2005.

[BMM19] 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.

[BMP15] Simone Brugiapaglia, Stefano Micheletti, and Simona Perotto. Compressed solving: a numerical approximation technique for elliptic PDEs based on compressed sensing. *Comput. Math. Appl.*, 70(6):1306–1335, 2015.

[BNR00] Volker Barthelmann, Erich Novak, and Klaus Ritter. High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.*, 12(4):273–288, 2000. Multivariate polynomial interpolation.

[BNT10] 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.

[BNTT11a] Joakim Bäck, Fabio Nobile, Lorenzo Tamellini, and Raul Tempone. Implementation of optimal Galerkin and collocation approximations of PDEs with random coefficients. In *CANUM 2010, 40^e Congrès National d'Analyse Numérique*, volume 33 of *ESAIM Proc.*, pages 10–21. EDP Sci., Les Ulis, 2011.

[BNTT11b] Joakim Bäck, Fabio Nobile, Lorenzo Tamellini, and Raul Tempone. Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison. In J.S. Hesthaven and E.M. Ronquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, volume 76 of *Lecture Notes in Computational Science and Engineering*, pages 43–62. Springer, 2011. Selected papers from the ICOSAHOM '09 conference, June 22-26, Trondheim, Norway.

[BP06] 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.

[BPR21] Alex Bespalov, Dirk Praetorius, and Michele Ruggeri. Convergence and rate optimality of adaptive multilevel stochastic Galerkin FEM. *IMA Journal of Numerical Analysis*, 42(3):2190–2213, 05 2021.

[BPR22] Alex Bespalov, Dirk Praetorius, and Michele Ruggeri. Convergence and rate optimality of adaptive multilevel stochastic Galerkin FEM. *IMA J. Numer. Anal.*, 42(3):2190–2213, 2022.

[BPRR19] Alex Bespalov, Dirk Praetorius, Leonardo Rocchi, and Michele Ruggeri. Convergence of adaptive stochastic Galerkin FEM. *SIAM J. Numer. Anal.*, 57(5):2359–2382, 2019.

[BPRS24] Alex Bespalov, Dirk Praetorius, Thomas Round, and Andrey Savinov. Goal-oriented error estimation and adaptivity for stochastic collocation fem, 2024.

[Bre11] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[BS23] Alex Bespalov and David Silvester. Error estimation and adaptivity for stochastic collocation finite elements part ii: Multilevel approximation. *SIAM Journal on Scientific Computing*, 45(2):A781–A797, 2023.

[BS24] Alex Bespalov and Andrey Savinov. Convergence analysis of the adaptive stochastic collocation finite element method, 2024.

[BSX22a] Alex Bespalov, David J. Silvester, and Feng Xu. Error estimation and adaptivity for stochastic collocation finite elements part i: Single-level approximation. *SIAM Journal on Scientific Computing*, 44(5):A3393–A3412, 2022.

[BSX22b] Alex Bespalov, David J. Silvester, and Feng Xu. Error estimation and adaptivity for stochastic collocation finite elements Part I: Single-level approximation. *SIAM J. Sci. Comput.*, 44(5):A3393–A3412, 2022.

[BTNT12] Joakim Bäck, Raul Tempone, Fabio Nobile, and Lorenzo Tamellini. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Models Methods Appl. Sci.*, 22(9):1250023, 33, 2012.

[BY89] Alexei N Bogdanov and DA Yablonskii. Thermodynamically stable "vortices" in magnetically ordered crystals. the mixed state of magnets. *Zh. Eksp. Teor. Fiz*, 95(1):178, 1989.

[CCM+15] Abdellah Chkifa, Albert Cohen, Giovanni Migliorati, Fabio Nobile, and Raul Tempone. Discrete least squares polynomial approximation with random evaluations—application to parametric and stochastic elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 49(3):815–837, 2015.

[CCS14] Abdellah Chkifa, Albert Cohen, and Christoph Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Found. Comput. Math.*, 14(4):601–633, 2014.

[CCS15] 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.

[CD15] Albert Cohen and Ronald DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numer.*, 24:1–159, 2015.

[CDG98] 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.

[CDL13] Albert Cohen, Mark A. Davenport, and Dany Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013.

[CDS11] 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.

[CF01a] Gilles Carbou and Pierre Fabrie. Regular solutions for Landau-Lifschitz equation in a bounded domain. *Differential Integral Equations*, 14(2):213–229, 2001.

[CF01b] Gilles Carbou and Pierre Fabrie. Regular solutions for Landau-Lifschitz equation in $\mathbb{R}^3$. *Commun. Appl. Anal.*, 5(1):17–30, 2001.

[CFPP14] Carsten Carstensen, Michael Feischl, Marcus Page, and Dirk Praetorius. Axioms of adaptivity. *Comput. Math. Appl.*, 67(6):1195–1253, 2014.

[Cim07] 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.

[CKNS08] J Manuel Cascon, Christian Kreuzer, Ricardo H Nochetto, and Kunibert G Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM Journal on Numerical Analysis*, 46(5):2524–2550, 2008.

[CM17] Albert Cohen and Giovanni Migliorati. Optimal weighted least-squares methods. *The SMAI Journal of computational mathematics*, 3:181–203, 2017.

[CM23] Albert Cohen and Giovanni Migliorati. Near-optimal approximation methods for elliptic PDEs with lognormal coefficients. *Math. Comp.*, 92(342):1665–1691, 2023.

[CQR17] Peng Chen, Alfio Quarteroni, and Gianluigi Rozza. Reduced basis methods for uncertainty quantification. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):813–869, 2017.

[CSZ18] Albert Cohen, Christoph Schwab, and Jakob Zech. Shape holomorphy of the stationary Navier-Stokes equations. *SIAM J. Math. Anal.*, 50(2):1720–1752, 2018.

[DFIP20] Giovanni Di Fratta, Michael Innerberger, and Dirk Praetorius. Weak-strong uniqueness for the Landau-Lifshitz-Gilbert equation in micromagnetics. *Nonlinear Anal. Real World Appl.*, 55:103122, 13, 2020.

[DI83] VK Dzjadyk and VV Ivanov. On asymptotics and estimates for the uniform norms of the lagrange interpolation polynomials corresponding to the chebyshev nodal points. *Analysis Mathematica*, 9(2):85–97, 1983.

[Die69] Jean Dieudonné. *Foundations of modern analysis*. Pure and Applied Mathematics, Vol. 10-I. Academic Press, New York-London, 1969. Enlarged and corrected printing.

[DILP18] Josef Dick, Christian Irrgeher, Gunther Leobacher, and Friedrich Pillichshammer. On the optimal order of integration in hermite spaces with finite smoothness. *SIAM Journal on Numerical Analysis*, 56(2):684–707, 2018.

[DKLM15] Sergey Dolgov, Boris N. Khoromskij, Alexander Litvinenko, and Hermann G. Matthies. Polynomial chaos expansion of random coefficients and the solution of stochastic partial differential equations in the tensor train format. *SIAM/ASA J. Uncertain. Quantif.*, 3(1):1109–1135, 2015.

[DKP22] Josef Dick, Peter Kritzer, and Friedrich Pillichshammer. *Lattice rules—numerical integration, approximation, and discrepancy*, volume 58 of *Springer Series in Computational Mathematics*. Springer, Cham, [2022] ©2022. With an appendix by Adrian Ebert.

[DKS13] Josef Dick, Frances Y. Kuo, and Ian H. Sloan. High-dimensional integration: The quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013.

[DNSZ23a] Dinh Dũng, Van Kien Nguyen, Christoph Schwab, and Jakob Zech. *Multilevel Smolyak Sparse-Grid Interpolation and Quadrature*, pages 145–195. Springer International Publishing, Cham, 2023.

[DNSZ23b] Dinh Dũng, Van Kien Nguyen, Christoph Schwab, and Jakob Zech. Analyticity and sparsity in uncertainty quantification for PDEs with gaussian random field inputs, 2023.

[Dos77] Halim Doss. Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 13(no. 2,):99–125, 1977.

[DP10] Josef Dick and Friedrich Pillichshammer. *Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration*. Cambridge University Press, Cambridge, 2010.

[Dra03] Sever Silvestru Dragomir. *Some Gronwall type inequalities and applications*. Nova Science Publishers, Inc., Hauppauge, NY, 2003.

[DS14] Eric Dumas and Franck Sueur. On the weak solutions to the Maxwell-Landau-Lifshitz equations and to the Hall-magneto-hydrodynamic equations. *Comm. Math. Phys.*, 330(3):1179–1225, 2014.

[dSC+06] Massimiliano d'Aquino, Claudio Serpico, Gennaro Coppola, Isaak D Mayergoyz, and Giorgio Bertotti. Midpoint numerical technique for stochastic Landau-Lifshitz-Gilbert dynamics. *Journal of applied physics*, 99(8):08B905, 2006.

[Du23] Dinh Dũng. Numerical weighted integration of functions having mixed smoothness. *J. Complexity*, 78:Paper No. 101757, 20, 2023.

[ECN+12] Richard F. L. Evans, Roy W. Chantrell, Ulrich Nowak, Andreas Lyberatos, and Hannah J. Richter. Thermally induced error: Density limit for magnetic data storage. *Applied Physics Letters*, 100(10):102402, 03 2012.

[EEST22] Martin Eigel, Oliver G. Ernst, Björn Sprungk, and Lorenzo Tamellini. On the convergence of adaptive stochastic collocation for elliptic partial differential equations with affine diffusion. *SIAM J. Numer. Anal.*, 60(2):659–687, 2022.

[EGSZ14] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. Adaptive stochastic galerkin fem. *Computer Methods in Applied Mechanics and Engineering*, 270:247–269, 2014.

[EGSZ15] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *ESAIM Math. Model. Numer. Anal.*, 49(5):1367–1398, 2015.

[EHL+14] Mike Espig, Wolfgang Hackbusch, Alexander Litvinenko, Hermann G. Matthies, and Philipp Wähnert. Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats. *Comput. Math. Appl.*, 67(4):818–829, 2014.

[Eli99] I. Elishakoff. Are probabilistic and anti-optimization approaches compatible? In Isaac Elishakoff, editor, *Whys and Hows in Uncertainty Modelling*, pages 263–355, Vienna, 1999. Springer Vienna.

[EMSU12] Ernst, Oliver G., Mugler, Antje, Starkloff, Hans-Jörg, and Ullmann, Elisabeth. On the convergence of generalized polynomial chaos expansions. *ESAIM: M2AN*, 46(2):317–339, 2012.

[Ern14] Oliver Ernst. Mathematische methoden der unsicherheitsquantifizierung, Sommersemester 2014.

[EST18] 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.

[Eva13] Lawrence C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, Providence, RI, 2013.

[FKS18] Michael Feischl, Frances Y. Kuo, and Ian H. Sloan. Fast random field generation with H-matrices. *Numer. Math.*, 140(3):639–676, 2018.

[FPW11] Stefan Funken, Dirk Praetorius, and Philipp Wissgott. Efficient implementation of adaptive p1-fem in matlab. *Computational Methods in Applied Mathematics*, 11(4):460–490, 2011.

[FS21] Michael Feischl and Andrea Scaglioni. Convergence of adaptive stochastic collocation with finite elements. *Comput. Math. Appl.*, 98:139–156, 2021.

[FST05] Philipp Frauenfelder, Christoph Schwab, and Radu Alexandru Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):205–228, 2005.

[FT17a] 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.

[FT17b] 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.

[GG03] Thomas Gerstner and Michael Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, 2003.

[GGL20] 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.

[Gha99] Roger Ghanem. Ingredients for a general purpose stochastic finite elements implementation. *Comput. Methods Appl. Mech. Engrg.*, 168(1-4):19–34, 1999.

[GHJvW23] Philipp Grohs, Fabian Hornung, Arnulf Jentzen, and Philippe von Wurstemberger. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *Mem. Amer. Math. Soc.*, 284(1410):v+93, 2023.

[Gil55] Thomas L Gilbert. A lagrangian formulation of the gyromagnetic equation of the magnetization field. *Phys. Rev.*, 100:1243, 1955.

[Gil15] Michael B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

[GKN+11] Ivan G. Graham, Frances Y. Kuo, Dirk Nuyens, Robert Scheichl, and Ian H. Sloan. Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications. *J. Comput. Phys.*, 230(10):3668–3694, 2011.

[GKN+15] Ivan G. Graham, Frances Y. Kuo, James A. Nichols, R. Scheichl, Christoph Schwab, and Ian H. Sloan. Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. *Numer. Math.*, 131(2):329–368, 2015.

[GLT16] 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.

[GN18] Diane Guignard and Fabio Nobile. A posteriori error estimation for the stochastic collocation finite element method. *SIAM J. Numer. Anal.*, 56(5):3121–3143, 2018.

[GO16] Michael Griebel and Jens Oettershagen. On tensor product approximation of analytic functions. *J. Approx. Theory*, 207:348–379, 2016.

[GPL98] 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.

[GS91] Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[GS24] Michael Griebel and Uta Seidler. A dimension-adaptive combination technique for uncertainty quantification. *Int. J. Uncertain. Quantif.*, 14(2):21–43, 2024.

[GSZ90] Michael Griebel, Micha Schneider, and Christoph Zenger. A combination technique for the solution of sparse grid problems. *Forschungsberichte, TU Munich*, TUM I 9038:1–24, 1990.

[GV23] Philipp Grohs and Felix Voigtlaender. Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces. *Foundations of Computational Mathematics*, 23(2):329–392, 2023.

[GZ07] Baskar Ganapathysubramanian and Nicholas Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *Journal of Computational Physics*, 225(1):652–685, 2007.

[Her89] Michel Hervé. *Analyticity in Infinite Dimensional Spaces*. De Gruyter, Berlin, New York, 1989.

[Hic14] Fred J. Hickernell. *Koksma-Hlawka Inequality*. John Wiley & Sons, Ltd, 2014.

[HRW12] Martin Hairer, Marc Ryser, and Hendrik Weber. Triviality of the 2d stochastic allen-cahn equation. *Electron. J. Probab.*, 17:no. 39, 1–14, 2012.

[HS19] Lukas Herrmann and Christoph Schwab. Multilevel quasi-Monte Carlo integration with product weights for elliptic PDEs with lognormal coefficients. *ESAIM Math. Model. Numer. Anal.*, 53(5):1507–1552, 2019.

[HSS23] Helmut Harbrecht, Marc Schmidlin, and Christoph Schwab. The gevrey class implicit mapping theorem with application to uq odissertation.pdff semilinear elliptic PDEs, 2023.

[IH06] V Koji Matsumoto V Akihiro Inomata and V Shin-ya Hasegawa. Thermally assisted magnetic recording. *Fujitsu Sci. Tech. J*, 42(1):158–167, 2006.

[JWZ18] Peter Jantsch, Clayton G Webster, and Guannan Zhang. On the Lebesgue constant of weighted Leja points for Lagrange interpolation on unbounded domains. *IMA Journal of Numerical Analysis*, 39(2):1039–1057, 06 2018.

[Kak48] Shizuo Kakutani. On equivalence of infinite product measures. *Annals of Mathematics*, 49(1):214–224, 1948.

[KGM+08] Mark H Kryder, Edward C Gage, Terry W McDaniel, William A Challener, Robert E Rottmayer, Ganping Ju, Yiao-Tee Hsia, and M Fatih Erden. Heat assisted magnetic recording. *Proceedings of the IEEE*, 96(11):1810–1835, 2008.

[KH70] Ryogo Kubo and Natsuki Hashitsume. Brownian motion of spins. *Progress of Theoretical Physics Supplement*, 46:210–220, 1970.

[KN16] Frances Y. Kuo and Dirk Nuyens. Application of quasi-Monte Carlo methods to elliptic PDEs with random diffusion coefficients: a survey of analysis and implementation. *Found. Comput. Math.*, 16(6):1631–1696, 2016.

[Kor59] Anton N. Korobov. The approximate computation of multiple integrals. In *Dokl. Akad. Nauk SSSR*, volume 124, pages 1207–1210, 1959.

[KP06] Martin Kružík and Andreas Prohl. Recent developments in the modeling, analysis, and numerics of ferromagnetism. *SIAM Rev.*, 48(3):439–483, 2006.

[KPP+19] 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.

[KRVE05] 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.

[KSS12] Frances Y. Kuo, Christoph Schwab, and Ian H. Sloan. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM Journal on Numerical Analysis*, 50(6):3351–3374, 2012.

[KSS15] Frances Y. Kuo, Christoph Schwab, and Ian H. Sloan. Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients. *Found. Comput. Math.*, 15(2):411–449, 2015.

[KTB11] Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. Wiley, 2011.

[LL35] Evgeny Lifshitz and Lev D. Landau. On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. *Phys. Z. Sowjetunion*, 8(135), 1935.

[LLW15] 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.

[LP14] Gunther Leobacher and Friedrich Pillichshammer. *Introduction to quasi-Monte Carlo integration and applications*. Compact Textbooks in Mathematics. Birkhäuser/Springer, Cham, 2014.

[LPS14] Gabriel J. Lord, Catherine E. Powell, and Tony Shardlow. *An introduction to computational stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[LSS20] Jens Lang, Robert Scheichl, and David Silvester. A fully adaptive multilevel stochastic collocation strategy for solving elliptic PDEs with random data. *J. Comput. Phys.*, 419:109692, 17, 2020.

[LSU68] 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.

[LW10] Anders Logg and Garth N. Wells. Dolfin: Automated finite element computing. *ACM Trans. Math. Softw.*, 37(2), apr 2010.

[LW21] Wai Cheung Law and Shawn De Wei Wong. Spin transfer torque magnetoresistive random access memory. *Emerging Non-volatile Memory Technologies: Physics, Engineering, and Applications*, pages 45–102, 2021.

[MBJ+09] Sebastian Mühlbauer, Benedikt Binz, Florian Jonietz, Christian Pfleiderer, Achim Rosch, Anja Neubauer, Robert Georgii, and Peter Boöni. Skyrmion lattice in a chiral magnet. *Science*, 323(5916):915–919, 2009.

[MBS09] Isaak D Mayergoyz, Giorgio Bertotti, and Claudio Serpico. *Nonlinear magnetization dynamics in nanosystems*. Elsevier, 2009.

[Mel05] 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.

[Mel12] Christof 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.

[MG12] Lionel Mathelin and Kyle A. Gallivan. A compressed sensing approach for partial differential equations with random input data. *Communications in Computational Physics*, 12(4):919–954, 2012.

[MK05] Hermann G. Matthies and Andreas Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1295–1331, 2005.

[Mos05] Roger Moser. *Partial regularity for harmonic maps and related problems*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[Mus17] Eleonora Musharbash. Dynamical low rank approximation of PDEs with random parameters. Technical report, EPFL, 2017.

[NS23] Dirk Nuyens and Yuya Suzuki. Scaled lattice rules for integration on $\mathbb{R}^d$ achieving higher-order convergence with error analysis in terms of orthogonal projections onto periodic spaces. *Math. Comp.*, 92(339):307–347, 2023.

[NT09] 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.

[NTT16] 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.

[NTW08a] 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.

[NTW08b] Fabio Nobile, Raul Tempone, and Clayton G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[Øks03] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[OPS20] Joost A. A. Opschoor, Philipp C. Petersen, and Christoph Schwab. Deep relu networks and high-order finite element methods. *Analysis and Applications*, 18(05):715–770, 2020.

[OS24] Joost A.A. Opschoor and Christoph Schwab. Deep relu networks and high-order finite element methods ii: Chebyšev emulation. *Computers & Mathematics with Applications*, 169:142–162, 2024.

[Owe23] Art B. Owen. *Practical Quasi-Monte Carlo Integration*. `https://artowen.su.domains/mc/practicalqmc.pdf`, 2023.

[Pav14] Grigorios A Pavliotis. Stochastic processes and applications. *Texts in applied mathematics*, 60, 2014.

[PP20] Carl-Martin Pfeiler and Dirk Praetorius. Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comp.*, 89(326):2735–2752, 2020.

[PS01] 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.

[PY15] Stuart Parkin and See-Hun Yang. Memory on the racetrack. *Nature nanotechnology*, 10(3):195–198, 2015.

[QMN15] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*. UNITEXT. Springer International Publishing, 2015.

[RBB+06] Robert E Rottmayer, Sharat Batra, Dorothea Buechel, William A Challener, Julius Hohlfeld, Yukiko Kubota, Lei Li, Bin Lu, Christophe Mihalcea, Keith Mountfield, et al. Heat-assisted magnetic recording. *IEEE Transactions on Magnetics*, 42(10):2417–2421, 2006.

[RHM+13] Niklas Romming, Christian Hanneken, Matthias Menzel, Jessica E Bickel, Boris Wolter, Kirsten von Bergmann, André Kubetzka, and Roland Wiesendanger. Writing and deleting single magnetic skyrmions. *Science*, 341(6146):636–639, 2013.

[RNT12] Marc D. Ryser, Nilima Nigam, and Paul F. Tupper. On the well-posedness of the stochastic allen–cahn equation in two dimensions. *Journal of Computational Physics*, 231(6):2537–2550, 2012.

[Rug20] Michele Ruggeri. The mathematics of magnetic materials, 2020. Lecture notes of a course given at TU Wien.

[Smo63] Sergei Abramovich Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Doklady Akademii Nauk*, volume 148, pages 1042–1045. Russian Academy of Sciences, 1963.

[Sob67] Ilya M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.

[SSF01] 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.

[ST06] Christoph Schwab and Radu Alexandru Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.

[Ste01] J. Michael Steele. *Stochastic calculus and financial applications*, volume 45 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001.

[Ste07] Rob Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.

[Ste08] Rob Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241, 2008.

[Sus78] Héctor J. Sussmann. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probability*, 6(1):19–41, 1978.

[TJWG15] 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.

[Tre19] Lloyd N. Trefethen. *Approximation Theory and Approximation Practice, Extended Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019.

[TT08] Rodney Taylor and Vilmos Totik. Lebesgue constants for Leja points. *IMA Journal of Numerical Analysis*, 30(2):462–486, 12 2008.

[TT10] Rodney Taylor and Vilmos Totik. Lebesgue constants for Leja points. *IMA J. Numer. Anal.*, 30(2):462–486, 2010.

[WK06] Xiaoliang Wan and George Em Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM Journal on Scientific Computing*, 28(3):901–928, 2006.

[WYW06] Zhuoqun Wu, Jingxue Yin, and Chunpeng Wang. *Elliptic & parabolic equations*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006.

[WZ65] Eugene Wong and Moshe Zakai. On the Convergence of Ordinary Integrals to Stochastic Integrals. *The Annals of Mathematical Statistics*, 36(5):1560 – 1564, 1965.

[WZZ22] Juncheng Wei, Qidi Zhang, and Yifu Zhou. Finite-time singularity formations for the Landau-Lifshitz-Gilbert equation in dimension two, 2022.

[XE02] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 191(43):4927–4948, 2002.

[XH05] 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.

[XK03] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187(1):137–167, 2003.

[Zen91] Christoph Zenger. Sparse grids. In *Parallel algorithms for partial differential equations (Kiel, 1990)*, volume 31 of *Notes Numer. Fluid Mech.*, pages 241–251. Friedr. Vieweg, Braunschweig, 1991.

[ZG12] 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.

Appendix A

Physical units

Physical quantity	Unit	Note
Current	Ampere A	SI base unit
Distance	Meter m	SI base unit
Weight	Gram g	SI base unit
Time	Second s	SI base unit
Temperature	Kelvin K	SI base unit
Magnetic flux density	Tesla $\boldsymbol{T}$	$1\boldsymbol{T} = 1\frac{kg}{s^2 A}$
Force	Newton N	$1N = 10^3 gms^{-2}$
Work, energy	Joule $\boldsymbol{J}$	$1\boldsymbol{J} = 1Nm$

Appendix B

Algebraic identities and estimates

Consider $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{K}^3$, with $\mathbb{K}$ a field. As usual, we define the dot and cross products of vectors as

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 + a_3 b_3, \qquad \boldsymbol{a} \times \boldsymbol{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ -a_1 b_3 + a_3 b_1 \\ a_1 b_2 - a_2 b_1 \end{pmatrix};$$

We have the following elementary algebraic identities:

- Triple product for cross and dot product

$$(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c} = (\boldsymbol{b} \times \boldsymbol{c}) \cdot \boldsymbol{a} = (\boldsymbol{c} \times \boldsymbol{a}) \cdot \boldsymbol{b}. \tag{B.1}$$

- Triple product for the cross-product

$$\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b}(\boldsymbol{a} \cdot \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a} \cdot \boldsymbol{c}). \tag{B.2}$$

When working with the Landau–Lifshitz–Gilbert equation, it is useful to consider the following non-standard definitions: Given vector fields $\boldsymbol{u}, \boldsymbol{v} : D \to \mathbb{R}^3$

- $\boldsymbol{u} \times \nabla \boldsymbol{v} \coloneqq \left(\boldsymbol{u} \times \frac{\partial \boldsymbol{v}}{\partial x_1}, \boldsymbol{u} \times \frac{\partial \boldsymbol{v}}{\partial x_2}, \boldsymbol{u} \times \frac{\partial \boldsymbol{v}}{\partial x_3} \right)$;
- $\nabla \boldsymbol{u} \times \nabla \boldsymbol{v} \coloneqq \sum_{i=1}^{3} \frac{\partial \boldsymbol{u}}{\partial x_i} \times \frac{\partial \boldsymbol{v}}{\partial x_i}$;
- $\langle \boldsymbol{u} \times \nabla \boldsymbol{v}, \boldsymbol{w} \rangle = \sum_{i=1}^{3} \left\langle \boldsymbol{u} \times \frac{\partial \boldsymbol{v}}{\partial x_i}, \frac{\partial \boldsymbol{w}}{\partial x_i} \right\rangle$.

If a vector field $\boldsymbol{m} : D \to \mathbb{R}^3$ is such that $|\boldsymbol{m}| = 1$ a.e., then

$$\Delta \boldsymbol{m} \cdot \boldsymbol{m} = -\left|\nabla \boldsymbol{m}\right|^2 \boldsymbol{m}. \tag{B.3}$$

This is a consequence of the fact that $\Delta \left|\boldsymbol{m}\right| = 0$ through the chain rule.

Appendix C

Facts from probability theory

As above, let $(\Omega, \mathcal{E}, \mathcal{P})$ denote the classical probability triple made respectively of a set, a σ-algebra on Ω, and a probability measure on the measurable space $(\Omega, \mathcal{E})$.

Characteristic function of a random variable

Definition C.1. *Let $X : \Omega \to \mathbb{R}$ be a random variable. Its* characteristic function $\phi_X : \mathbb{R} \to \mathbb{C}$ *is $\phi_X(t) \coloneqq \mathbb{E}\left[e^{itX}\right]$.*

It is a simple exercise to prove that the characteristic function of the normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ of mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ is $\phi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$. If instead $X : \Omega \to [a, b]$ is uniform on the real interval $[a, b]$, then $\phi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$. Some important properties of the characteristic function of a random variable are:

- Two random variables X_1, X_2 have the same probability distribution if and only if they have the same characteristic function: $X_1 \sim X_2 \Leftrightarrow \phi_{X_1} = \phi_{X_2}$;
- The characteristic function of the linear combination $a_1X_1 + \cdots + a_nX_n$ of random variables $X_1, \ldots, X_n$ with real coefficients $a_1, \ldots, a_n$ is $\phi_{a_1X_1 + \cdots + a_nX_n}(t) = \phi_{X_1}(a_1t) \ldots \phi_{X_n}(a_nt)$;
- Two random variables X, Y are independent if and only if $\phi_{X,Y}(s, t) = \phi_X(s)\phi_Y(t)$, where $\phi_{X,Y}(s, t)$ is the characteristic function of the joint random variable $(X, Y) : \Omega \to \mathbb{R}^2$.

Relation between Stratonovich and Itô differentials

We do not precisely define Itô and Stratonovich stochastic integrals. The interested reader can find them e.g. in [Eva13, Sections 4.2 and 6.5 respectively].

The following are the Itô (respectively Stratonovich) chain rule, which, given a stochastic process $X = X(\omega, t)$ through an Itô (respectively Stratonovich) SDE and a (sufficiently regular) function $f = f(t, x)$, can be used to determine which Itô (respectively Stratonovich) SDE $f(t, X(\omega, t))$ obeys.

Lemma C.2 (Itô chain rule)**.** *Let $T > 0$, $X : [0, T] \to \mathbb{R}$ be solution of the following Itô SDE:*

$$dX(t) = \mathfrak{D}(t, X(t))dt + \mathfrak{N}(t, X(t))dW(t) \qquad \forall t \in [0, T], \quad a.e.\ in\ \Omega,$$

where $\mathfrak{D} : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $\mathfrak{N} : [0, T] \times \mathbb{R} \to \mathbb{R}$ are sufficiently regular for the problem to have solution. Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a twice-differentiable function. Then, $f(\cdot, X(\cdot)) : \Omega \times [0, T] \to \mathbb{R}$ is solution of the following Itô SDE:

$$df = \left(\frac{\partial f}{\partial t} + \mathfrak{D}\frac{\partial f}{\partial x} + \frac{\mathfrak{N}^2}{2}\frac{\partial^2 f}{\partial x^2}\right)dt + \mathfrak{N}\frac{\partial f}{\partial x}dW \qquad in\ [0, T], \quad a.e.\ in\ \Omega,$$

where for all functions we omitted the dependence on $(t, X(t))$.

See [Pav14, Lemma 3.2] for a proof.

The stochastic chain rule for Stratonovich SDEs is formally the same as for deterministic functions:

Lemma C.3 (Stratonovich chain rule)**.** *Let* $T > 0$, $X : [0,T] \to \mathbb{R}$ *be solution of the following Stratonovich SDE:*

$$dX(t) = \mathfrak{D}(t, X(t))\,dt + \mathfrak{N}(t, X(t)) \circ dW(t) \qquad \forall t \in [0,T], \quad a.e.\ in\ \Omega,$$

where $\mathfrak{D} : [0,T] \times \mathbb{R} \to \mathbb{R}$ *and* $\mathfrak{N} : [0,T] \times \mathbb{R} \to \mathbb{R}$ *are sufficiently regular for the problem to have solution. Let* $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *be a twice-differentiable function. Then,* $f(\cdot, X(\cdot)) : \Omega \times [0,T] \to \mathbb{R}$ *is solution of the following Stratonovich SDE:*

$$df = \left(\frac{\partial f}{\partial t} + \mathfrak{D}\frac{\partial f}{\partial x}\right) dt + \mathfrak{N}\frac{\partial f}{\partial x} \circ dW \qquad in\ [0,T], \quad a.e.\ in\ \Omega,$$

where for all functions we omitted the dependence on $(t, X(t))$.

See [Pav14, Proposition 3.4] for a proof.

Another important fact is the "conversion formula" between SDEs formulated with respect to one or the other stochastic integrals.

Lemma C.4. *There holds the following identity involving Stratonovich and Itô differentials of a sufficiently regular function* $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$*:*

$$f(t, W) \circ dW = \frac{1}{2}\partial_W f(t, W)\,dt + f(t, W)\,dW.$$

As a consequence, for $T > 0$, $X : [0,T] \to \mathbb{R}$ *solves the following Ito SDE:*

$$dX(t) = \mathfrak{D}(t, X(t))\,dt + \mathfrak{N}(t, X(t))\,dW(t) \qquad \forall t > 0, a.e.\ in\ \Omega$$

if and only if it solves the following Stratonovich SDE (omitting the dependence on t, X*):*

$$dX = \left(\mathfrak{D} - \frac{1}{2}\mathfrak{N}'\mathfrak{N}\right) dt + \mathfrak{N} \circ dW,$$

where $\mathfrak{D} : [0,T] \times \mathbb{R} \to \mathbb{R}$ *and* $\mathfrak{N} : [0,T] \times \mathbb{R} \to \mathbb{R}$ *are sufficiently regular for the problems to have solution.*

Appendix D

Facts from analysis

Stirling's formula

For $n \in \mathbb{N}$, there holds the asymptotic approximation

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \tag{D.1}$$

Exponential of an operator

Given a Banach space B, an linear operator $A : B \to B$, the *exponential of* A is by definition

$$e^A \coloneqq \sum_{n \in \mathbb{N}_0} \frac{A^{(n)}}{n!}, \tag{D.2}$$

where $A^{(n)}$ denotes the composition of A with itself n times, i.e., $A^{(0)} \coloneqq \mathrm{id}$, $A^{(1)} \coloneqq A$, and $A^{(n)} = A^{(n-1)} \circ A$ for any $n = 2, 3, \dots$ $e^A : U \to U$ is a linear bounded operator if $\left(\sum_{n=0}^{N} \frac{A^{(n)}}{n!}\right)_{N \in \mathbb{N}}$ converges to one in the norm of $\mathcal{L}(U)$, i.e. the Banach space of linear operators $U \to U$.

Sequence spaces and Stechkin's lemma

Recall that by $\mathcal{F}$ we denote the space of sequences with finite support. The set of $\ell^p(\mathcal{F})$-*summable sequences* is by definition

$$\ell^p(\mathcal{F}) \coloneqq \left\{ a_\cdot : \mathcal{F} \to \mathbb{R} : \sum_{\boldsymbol{\nu} \in \mathcal{F}} |a_{\boldsymbol{\nu}}|^p < \infty \right\} \qquad \text{if } 0 < p < \infty,$$
$$\ell^\infty(\mathcal{F}) \coloneqq \left\{ a_\cdot : \mathcal{F} \to \mathbb{R} : \sup_{\boldsymbol{\nu} \in \mathcal{F}} |a_{\boldsymbol{\nu}}| < \infty \right\}.$$

It is a Banach space for $1 \leq p$
$leq\infty$. The following is a useful lemma to estimate the tails of ℓ^p-summable sequences.

Lemma D.1 (Stechkin)**.** *Let $0 < p \leq q \leq \infty$, $(a_n)_{n \in \mathbb{N}}$ a non-increasing sequence such that $(a_n)_n \in \ell^p(\mathbb{N})$. Then,*

$$\left(\sum_{n > N} a_n^q\right)^{\frac{1}{q}} \leq \|(a_n)_n\|_{\ell^p} \, N^{\frac{1}{q} - \frac{1}{p}}.$$

Stechkin's lemma can be extended to $\ell^p(\mathcal{F})$-summable sequences since $\mathcal{F}$ is countable.

Corollary D.2 (Stechkin lemma on $\mathcal{F}$)**.** *Let $0 < p \leq q \leq \infty$ and consider a sequence $(a_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{F}}$ such that $(a_{\boldsymbol{\nu}})_{\boldsymbol{\nu}} \in \ell^p(\mathcal{F})$. Then, denoting $\Lambda_n \subset \mathcal{F}$ the set of multi-indices $\boldsymbol{\nu} \in \mathcal{F}$ with largest $a_{\boldsymbol{\nu}}$ we have*

$$\left(\sum_{\boldsymbol{\nu}\notin\Lambda_n} a_{\boldsymbol{\nu}}^q\right)^{\frac{1}{q}} \leq \|(a_{\boldsymbol{\nu}})_{\boldsymbol{\nu}}\|_{\ell^p(\mathbb{N})}\, n^{\frac{1}{q}-\frac{1}{p}}.$$

See [CDS11] for a proof.

Spectral theorem for compact self-adjoint operators

The following is based on [Ern14].
Let H, K be Hilbert spaces (on the real or complex field). Denote by $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{R}$ the scalar product of H and by $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle} : H \to \mathbb{R}_{\geq 0}$. Denote by $\mathcal{L}(H)$ the linear space of linear bounded operators on H and by $\mathcal{L}(H, K)$ the linear space of linear bounded operators from H to K.

Definition D.3 ((Sequentially) compact subset of a Hilbert space)**.** *A subset A of a Banach space B is* (sequentially) compact *if and only if every sequence $(u_n)_{n\in\mathbb{N}} \subset A$ admits a convergent subsequence $(u_{n_k})_{k\in\mathbb{N}}$.*

Definition D.4 (Compact, self-adjoint operator)**.** *$L \in \mathcal{L}(H, K)$ is called:*

- Compact *if and only if for any bounded subset B of H, $L(B)$ is pre-compact (i.e. it has a compact closure) in K;*
- Self-adjoint *if and only if $\langle Lu, v\rangle = \langle u, Lv\rangle \;\; \forall u, v \in H$.*

The following is an example of compact, self adjoint operator:

Proposition D.5. *Let $D \subset \mathbb{R}^n$ a smooth domain with $n \in \mathbb{N}$ and $k \in L^2(D \times D)$. The corresponding integral operator:*

$$K : L^2(D) \to L^2(D) \qquad Ku \coloneqq \int_D k(\boldsymbol{x}, \boldsymbol{y})u(\boldsymbol{y})\,d\boldsymbol{y} \quad \forall u \in L^2(D).$$

is a compact operator. Moreover, if k is symmetric (i.e. $k(\boldsymbol{x}, \boldsymbol{y}) = k(\boldsymbol{y}, \boldsymbol{x})$ for any $\boldsymbol{x}, \boldsymbol{y} \in D$), then K is self-adjoint with respect to the $L^2(D)$ scalar product.

Definition D.6. *Let H be a Hilbert space and $L \in \mathcal{L}(H)$. $\lambda \in \mathbb{C}$ is called* eigenvalue of L *if and only if there exists $\phi \in H \setminus \{\mathbf{0}\}$ such that $L\phi = \lambda\phi$. In this case, ϕ is called an* eigenvector *of L corresponding to λ.*

Theorem D.7 (Spectral theorem for compact self-adjoint operators, [Bre11, Theorem 6.11])**.** *Let H be a separable Hilbert space, $L \in \mathcal{L}(H)$ a compact self-adjoint operator. Denote by λ_n, $n \in \mathbb{N}$ the eigenvalues of L such that $|\lambda_n| \geq |\lambda_{n+1}|$ and denote $\phi_n \in H$, $n \in \mathbb{N}$ the corresponding eigenfunctions. Then,*

- *All eigenvalues are real and $\lim_{n\to\infty} \lambda_n = 0$;*
- *The set of eigenvectors $\{\phi_n\}_{n\in\mathbb{N}}$ can be chosen as a complete orthonormal basis of $L(H)$ (the range of L);*
- *In particular, for any $u \in H$, $Ku = \sum_{n\in\mathbb{N}} \lambda_n \langle u, \phi_n\rangle \phi_n$.*

Cauchy's integral formula

The following classical theorem is very useful when proving approximation properties of sparse polynomials. Let us begin by the basic case of a 1D complex domain.

Theorem D.8. *Consider $\mathfrak{N}$, an open, connected domain in $\mathbb{C}$, $\mathbb{U}$ a Banach space, and $u : \mathfrak{N} \to \mathbb{U}$. Then, for any $z_0 \in \mathfrak{N}$, $k \in \mathbb{N}_0$,*

$$\frac{d^k u}{dz^k}(z_0) = \frac{k!}{2\pi i} \int_\gamma \frac{u(z)}{(z - z_0)^{k+1}} dz$$

for any rectifiable non self-intersecting closed curve $\gamma \subset \mathbb{C}$ around z_0.

A direct consequence is the estimate obtained with γ such that $\operatorname{supp}(\gamma) = \mathbb{S}^2(z_0, \rho) \subset \Sigma$, where $\rho > 0$:

$$\left\| \frac{\mathrm{d}^k u}{\mathrm{d}z^k}(z_0) \right\|_{\mathbb{U}} \leq \frac{k!}{\rho^{k+1}} \max_{z \in B(z_0, \rho)} \|u(z)\|_{\mathbb{U}} .$$

When the function has a domain of dimension $N \in \mathbb{N}$, $u : \boldsymbol{\Sigma} \subset \mathbb{C}^N \to \mathbb{U}$, the Cauchy's integral formula generalizes easily to finitely-many partial derivatives:

Corollary D.9. *Let $u : \boldsymbol{\Sigma} \to \mathbb{U}$ with $\boldsymbol{\Sigma} = \bigotimes_{n=1}^N \mathfrak{N}_n \subset \mathbb{C}^N$, and $N \in \mathbb{N}$. Let $\mathbb{U}$ be a Banach space. For any $\boldsymbol{z}_0 \in \boldsymbol{\Sigma}$, $\boldsymbol{k} \in \mathbb{N}_0^N$,*

$$\partial^{k_1} \dots \partial^{k_N} u(\boldsymbol{z}) = \frac{\boldsymbol{k}!}{(2\pi i)^N} \int_{\gamma_1} \cdots \int_{\gamma_N} \frac{u(\boldsymbol{z})}{(z_1 - z_{0,1})^{k_1+1} \dots (z_N - z_{0,N})^{k_N+1}} d\boldsymbol{z},$$

where $\boldsymbol{k}! = \prod_{n=1}^N k_n!$, $\gamma_1 \subset \Sigma_1, \dots, \gamma_N \subset \Sigma_N$ are closed rectifiable non self-intersecting curves around $z_{0,1}, \dots, z_{0,N}$ respectively.

The corresponding estimate choosing γ_n such that $\operatorname{supp}(\gamma_n) = \mathbb{S}(z_{0,n}, \rho_n)$ for some $\rho_n > 0$ and taking the norm both sides gives

$$\left\| \partial^{k_1} \dots \partial^{k_N} u(\boldsymbol{z}) \right\|_{\mathbb{U}} \leq \boldsymbol{k}! \prod_{n=1}^N \rho_n^{-k_n} \max_{\boldsymbol{z} \in \boldsymbol{B}(\boldsymbol{z}_0, \boldsymbol{\rho})} \|u(\boldsymbol{z})\|_{\mathbb{U}} , \tag{D.3}$$

where $\boldsymbol{B}(\boldsymbol{z}_0, \boldsymbol{\rho}) = \bigotimes_{n=1}^N B(z_{0,n}, \rho_n)$ is called a *polydisk* of center $\boldsymbol{z}_0 \in \mathbb{C}^N$ and radius $\boldsymbol{\rho} \in \mathbb{R}^N_{>0}$.

Appendix E

Notation and index of symbols

In the following tables we list the most important symbols defined and used in this document. We do not include the most commonly accepted symbols (e.g. $L^2(D)$ for the Lebesgue space of square-integrable functions).

Symbol	Name	First occurrence
$\mathbb{A}$ (or $\mathbb{W}$)	Banach space of PDE coefficients	Section 1.1.1
$\mathbb{U}$	Banach space of PDE solutions	ibid.
R	Banach space of PDE residuals	ibid.
$\mathcal{R}$	PDE residuals	ibid.
$(\Omega, \mathcal{E}, \mathcal{P})$	Probability triple	ibid.
D	Space domain	ibid.
∂D	Boundary of domain D	ibid.
$c(\cdot, \cdot)$	Covariance of stochastic process	Section 1.1.2
$\mathcal{C}(\cdot)$	Covariance operator	(1.7)
Y_j	Scalar random variable	Section 1.1.2
$h_{\ell,j}$	Haar basis function	ibid.
$\mathbb{N}_0$	Non-negative integers	ibid.
$\eta_{\ell,j}$	Faber-Schauder basis function	ibid.
$\boldsymbol{Y}$	Sequence of i.i.d. random variables	Section 1.1.3
Γ_j	Codomain of Y_j	ibid.
y_j	Scalar parameter, running in Γ_j	ibid.
$\boldsymbol{\Gamma}$	Parametric domain	ibid.
y_N	N-dimensional parameter	Remark 1.7
$\boldsymbol{\Gamma}_N$	N-dimensional parameter domain	ibid.
$\mathbb{R}_{>0}$	Positive real numbers	Section 1.1.4
$\boldsymbol{B}_{\boldsymbol{\rho}}(\boldsymbol{y})$	Complex polydisk	ibid.
$\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_N$	Complex extension domain	ibid.
D_T	Space-time cylinder	Section 1.1.5
$\mathfrak{D}$	Drift coefficient in SPDE	ibid.
$\mathfrak{N}$	Noise coefficient in SPDE	ibid.
$\mathcal{L}(B)$	Banach space of linear bounded operators	ibid.
$\mathcal{U}$	Solution operator	Section 1.2.2
$\boldsymbol{\nu}$	Multi-index	ibid.
$\mathcal{F}$	Multi-indices with finite support	(1.22)
$\operatorname{supp}(\boldsymbol{\nu})$	Support of a multi-index $\boldsymbol{\nu} \in \mathcal{F}$	(1.23)
Λ, Λ_ℓ	Multi-index set	Section 1.2.2
$\mathcal{Y}_\nu, \mathcal{Y}_{\boldsymbol{\nu}}$	Nodes family	Section 1.2.3

Symbol	Name	First occurrence
$L_{\boldsymbol{y}}$	Lagrange basis function	ibid.
V_ν, $V_{\boldsymbol{\nu}}$	Linear space of collocation type	ibid.
I_ν, $I_{\boldsymbol{\nu}}$	Lagrange interpolant	(1.26)
Δ_ν (resp. $\Delta_{\boldsymbol{\nu}}$)	Detail operator (resp. hierarchical surplus)	Section 1.2.3
$\mathcal{H}_\Lambda$	Sparse grid with respect to Λ	ibid.
$\mathcal{I}_\Lambda$	Sparse grid interpolant	ibid.
$\mathbb{L}_\Lambda$	Lebesgue constant of $\mathcal{I}_\Lambda$	Equation 1.34
$\lambda_{\boldsymbol{\nu}}$	Lebesgue constant of $I_{\boldsymbol{\nu}}$	Section 1.2.3
$m(\cdot)$	Level-to-knot function	ibid.
$v_{\boldsymbol{\nu}}$	Value of the multi-index $\boldsymbol{\nu}$	Equation 1.40
$w_{\boldsymbol{\nu}}$	Value of the multi-index $\boldsymbol{\nu}$	Equation 1.41
$\mathcal{P}_{\boldsymbol{\nu}}$	Profit of the multi-index $\boldsymbol{\nu}$	Equation 1.42
$\boldsymbol{e}_n$	n-th coordinate unit vector	Section 1.2.3
$\mathcal{M}_\Lambda$	Margin of the multi-index set Λ	Definition 1.19
$\mathcal{RM}_\Lambda$	Reduced margin of multi-index set Λ	ibid.
$\mathcal{T}_{\boldsymbol{y}}$	Mesh in collocation point $\boldsymbol{y}$	Section 1.2.4
$\mathbb{U}_{\boldsymbol{y}}$	Finite dimensional space in $\boldsymbol{y}$	ibid.
$\zeta_{\boldsymbol{\nu},\Lambda}$	Pointwise estimator	(1.44)
$\zeta_{SC,\Lambda}$	A-posteriori error estimator	ibid.
$\eta_{\boldsymbol{y},T}$	Element-wise finite element estimator	(1.45)
$\eta_{\boldsymbol{y}}$	Parameter-wise finite element estimator	ibid.
$\eta_{FE,\Lambda}$	Finite element estimator	ibid.
M_s	Saturation magnetization	Section 1.3.2
T	Final time for parabolic problems	Also mesh element
$\boldsymbol{m}$	(Adimensional) magnetization	(1.49)
$\mathbb{S}^2$	Unit sphere in $\mathbb{R}^3$	Section 1.3.2
$\boldsymbol{H}_{\text{eff}}$	Effective field	(1.51)
λ	Gilbert damping	Section 1.3.2
$\mathbb{L}^2(D)$, $\mathbb{H}^1(D)$, etc.	Banach spaces of vector valued function	Section 1.3.2
$\langle\cdot,\cdot\rangle$	$\mathbb{L}^2(D)$-scalar product	ibid.
$\|\cdot\|$	$\mathbb{L}^2(D)$ norm	ibid.
h	Mesh-size	Section 1.3.3
$\mathcal{N}_h$	Mesh nodes	ibid.
$\boldsymbol{V}_h$	Vector-valued piecewise linear function	ibid.
$\mathfrak{I}_h$	Nodal interpolant onto $\boldsymbol{V}_h$	(1.56)
Π_h	L^2-projection onto $\boldsymbol{V}_h$	(1.57)
τ	Time-step	Section 1.3.3
$\langle\cdot,\cdot\rangle_h$	Mass-lumped scalar product	(1.58)
Δ_h	Discrete Laplacian	(1.59)
d_t	Backward finite difference	(1.60)
$\boldsymbol{m}^{j+\frac{1}{2}}$	Midpoint operator	(1.61)
$\mathcal{M}_h$	Discrete magnetizations space	Section 1.3.3
$\mathcal{K}(\boldsymbol{m}_h)$	Discrete tangent space	(1.62)
$\boldsymbol{m}_{\tau,h}$	Magnetization approximation	Section 1.3.3
$\boldsymbol{M}$	Random magnetization	Section 1.3.4
$\boldsymbol{H}_{\text{thermal}}$	Thermal noise	ibid.
$\boldsymbol{g}$	Noise coefficient	ibid.

Symbol	Name	First occurrence
G	Noise operator	Section 1.3.6
$\hat{\mathcal{C}}(s, \boldsymbol{u})$		(1.78)
$\mathcal{C}(s, \boldsymbol{u})$		ibid.
$\mathcal{E}(s, \boldsymbol{u})$		ibid.
ρ, ρ_N	Densities of random variables	Section 2.1.1
$U_{\boldsymbol{y}}$	Finite element solution for parameter $\boldsymbol{y}$	Section 2.1.2
$\bigsqcup$	Disjoint union	Algorithm 8
$A_{\boldsymbol{\nu},\Lambda}$	downward-closure	Section 2.1.3
θ	Dörfler parameter	Algorithm 8
λ_m	Lebesgue constant w m nodes	(2.11)
$\mathrm{Rect}_{\boldsymbol{\nu}}$	Multi-index rectangle in $\mathcal{F}$	Definition 2.17
$J_{\boldsymbol{\nu},\Lambda}$	Set of relevant multi-indices	Remark 2.4
μ_Λ	Set of maximal multi-indices in margin $\mathcal{M}_\Lambda$	Definition 2.8
$\|\cdot\|_{\mathcal{L}(U,V)}$	Norm of linear operator $U \to V$	Section 2.2.3
$\mathbb{T}$	Set of admissible meshes	Theorem 2.24
$\lambda_{q,\Lambda}$		Section 2.3.3
$\mathbb{W}_R$ (resp. $\mathbb{W}$)	Coefficients (resp. complex coefficients) space	Section 3.1.1
$\mathbb{U}_R$ (resp. $\mathbb{U}$)	Solutions (resp. complex solutions) space	ibid.
$\mathcal{X}_{\mathbb{R}}$, $\mathcal{X}$	Parameter space	Section 3.1.1
$\mathfrak{D}$, $\boldsymbol{\mu}$	Gaussian measure or density	Section 3.1.1
$\mathcal{R}_a$	Alternative SLLG residual	(3.21)
$\mathcal{R}_s$	Simplified SLLG residual	(3.50)
$H(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$	Trilinear form SLLG	(3.21)
$\tilde{\mathcal{C}}(s, \boldsymbol{u})$	Linearized term in simplified SLLG	(3.49)
$\tilde{\zeta}_{\boldsymbol{\nu}}$	Conservative pointwise parametric estimator	Definition 5.4
$X_{\boldsymbol{\alpha}}(w)$	Anisotropic simplex multi-index set	Definition 5.6