Effective Approximations of Stochastic Partial Differential Equations based on Wiener Chaos expansions and the Malliavin Calculus

by

Chia Ying Lee

B. S., University of Michigan, Ann Arbor, 2005

Sc. M., Brown University, 2007

A Dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Division of Applied Mathematics at Brown University

Providence, Rhode Island

May 2011

This dissertation by Chia Ying Lee is accepted in its present form
by the Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date ______________ ____

Boris L. Rozovsky, Director

Recommended to the Graduate Council

Date ______________ ____

George Karniadakis, Reader

Date ______________ ____

Kavita Ramanan, Reader

Approved by the Graduate Council

Date ______________ ____

Peter M. Weber, Dean of the Graduate School

Curriculum Vitæ

Chia Ying Lee was born in Penang, Malaysia on October 16, 1983, and lived and received her basic education in Singapore. She received a Bachelor of Science with High Distinction in Mathematics and Music from the University of Michigan at Ann Arbor in August 2005. From August 2005-2006, she worked at the Bioinformatics Institute, part of the Agency for Science, Technology and Research, in Singapore. She began her graduate studies in the Division of Applied Mathematics at Brown University in September 2006, where she worked under the supervision of Professor Boris Rozovsky. While at Brown, she received a Masters of Science in Applied Mathematics in May 2007. She was also awarded the Stella Dafermos Award in May 2011.

Dedicated to My Parents, Family and Percy

Acknowledgements

I would like to express my deepest gratitude to my thesis advisor Professor Boris Rozovsky, who has provided invaluable guidance throughout my graduate career, and has helped to point me in the right directions of my research. I am thankful for the freedom he has allowed me to discover my own interests and for the wide opportunities which, under his tutelage, was opened to me. Thus, as my academic father, I am indebted to him.

I would also like to thank all the professors and fellow students in the Division and at Brown who have helped make my graduate experience at Brown an immensely fruitful and enjoyable period of professional and personal growth. Special thanks goes to my committee members Professors George Karniadakis and Kavita Ramanan, who have kindly devoted their time to referee my thesis. I must also include to thank Professors Karniadakis, Chi-Wang Shu, David Gottlieb, Hao-Min Zhou and Bjorn Sandstede, among many others, who all played instrumental roles, both indirectly and direcly, in shaping the various aspects of my research and learning.

Last but not least, I am eternally grateful to my parents, family and my fiance, Percy, who have all given me incredible support, and though far away have been a constant source of motivation and strength to pursue my dreams.

Contents

List of Tables

List of Figures

Abstract of "Effective Approximations of Stochastic Partial Differential Equations based on Wiener Chaos expansions and the Malliavin Calculus"

by Chia Ying Lee, Ph.D., Brown University, May 2011

This thesis studies the application of the Wiener chaos expansion in the analysis of stochastic partial differential equations (SPDEs). Specifically, linear parabolic SPDEs and the quantized stochastic Navier-Stokes equations are considered, under the framework of the Malliavin calculus. Especially for these highly singular SPDEs, the Wiener chaos expansion is a useful tool for our study of the basic questions of solvability, regularity and dynamical behaviour, and it enables us to study approximations of the solutions of SPDEs and to quantify the errors of approximation. For the quantized stochastic Navier-Stokes equations, we use the Malliavin calculus to formulate a random perturbation of the Navier-Stokes equations that is unbiased, and we will show the existence and uniqueness of steady and time-dependent solutions, as well as the convergence to steady solution, in a stochastic weighted space. We also study a stochastic finite element method for numerical simulation of the solution of linear parabolic SPDEs and derive error estimates for the numerical solution. Finally, we show how one basis of the Wiener chaos expansion can be more efficient than another for approximating the energy of the solution, so that computational efficiency can be increased when applied to some physical applications.

CHAPTER 1

Introduction

In this thesis, we present analyses and numerical analyses of stochastic partial differential equations using the Malliavin calculus and Wiener chaos expansions, for two classes of SPDEs, linear parabolic SPDEs and the quantized stochastic Navier-Stokes equation.

The motivation for choosing the Wiener chaos expansion and Malliavin calculus as a tool of analysis comes in large part from the type of SPDE considered. Since the discovery of the Itô integral spurred the development of stochastic analysis and the Itô calculus, the study of stochastic models in a myriad of physical, biological and economic applications has caught the wave of Itô calculus. However, many SPDE arising in physical and mathematical models, such as those in Uncertainty Quantification, reveal several obvious limitations of the Itô calculus, not least the fact that it requires a notion of adaptedness. Uncertainty Quantification frequently deals with models of phenomena for which coefficients or parameters are not known to full certainty. Rather than neglecting the uncertainty in the model parameters, one can turn to stochastic models as a way to incorporate the uncertainty into the equations. In the simplest case, the uncertainty in a parameter may take the form of being a single random variable of a known distribution—a uniform distribution being a common choice. More complex real world examples of stochastic modelling include the stochastic pressure equations or models of flow in heterogeneous porous media, where the permeability of the medium is difficult to measure at all locations, and is instead modelled as a random field. This is a case of a noncausal system without a natural notion of a filtration, and it would be unwise to confine it into the Itô calculus framework.

We are thus prompted to appeal to a more general stochastic calculus in order to formulate models outside of the Itô calculus framework. This is where the Malliavin calculus comes into the picture. In fact, the Malliavin calculus is not such a far flung idea, because it is an extension of the Itô calculus. The Skorokhod integral, one of the important constructs of the Malliavin calculus, extends the Itô integral to non-adapted integrands, and coincides with the Itô integral for adapted integrands. For this reason, a modeler, when considering to introduce stochasticity into a model, may find the Skorokhod integral a viable modelling choice.

Further reasons to work in a more general framework come from the need for more general solution concepts. A desired model for a random perturbation can, and often does, lead to an immediate difficulty of non-square integrable solutions. Early works, such as Walsh's [**64**], had already shown that certain equations commonly encountered do not possess solutions with finite variance. The models with uncertain coefficients, involving multiplicative noise that acts on the highest order partial differential operator, are prime examples of equations without square integrable solutions. Lacking a "usual" solution, we are forced to broaden our notion of a solution to include solutions with infinite variance in a larger space of random elements. These spaces are the so-called weighted stochastic spaces, which include the Hida spaces and Kondratiev spaces among others, and whose elements are characterized by their Wiener chaos expansion.

The Wiener chaos expansion is a classical orthogonal expansion theory for random functions that was first introduced by Cameron and Martin [**9**]. It representations a square integrable random element in an orthogonal expansion of the stochastic variable with respect to a basis derived from the Hermite polynomials. In our case of non-square integrable solutions or random elements, the Wiener chaos expansion is especially pertinent for representing the random elements.

The application of the Wiener chaos expansion here is two-fold: to analyze solutions of SPDE through approximate solutions derived from the Wiener chaos expansion; and to quantify the efficiency of approximations of the solutions.

The major theme underlying all the analysis in this thesis is the transformation of the single SPDE, via the Wiener chaos expansion, into the related *propagator system* of PDE. The utility of this transformation is no different from that of the Fourier transformation—the solution is understood as a collection of its expansion coefficients, and the analysis of the SPDE is achieved through the analysis of the propagator system of equations. As noted above, the Wiener chaos expansion is also key to obtaining finite approximations of solutions of SPDEs. The finite approximations, here specifically Galerkin approximations, render the analysis of the SPDE more tractable to analysis. In analogy to deterministic theory, creating approximate solutions is the first step in formulating energy estimates which are then used to deduce the existence of a solution. In fact, because the conversion to the propagator system separates out the stochastic variable and leaves behind a deterministic system of equations, a large part of the stochastic analysis is founded on deterministic theory. Thus, the strength of the stochastic results depend on the strength of deterministic PDE theory.

Through the use of the Wiener chaos expansion, the difficulty in the stochastic analysis is greatly *reduced* to understanding results from deterministic theory. Seen in a different light, the stochastic theory is in fact a *generalization* of the deterministic theory to the stochastic setting. One supporting argument for this is that the Malliavin calculus, historically developed to build a stochastic theory based on the integration by parts formula, accords us with an arsenal of conceptual tools familiar from deterministic theory—an integration by parts formula, adjoint operators and the possibility of defining weak or variational solutions by action on test functions. To give an example of such stochastic analogues, we will subsequently encounter the use two of the main constructs of the Malliavin calculus, the Malliavin derivative and the Malliavin divergence operator, which are adjoints of each other under the Gaussian measure. The latter is, in fact, a stochastic convolution and is equivalent to the Skorokhod integral. Related to the Malliavin divergence operator is the Wick product (see e.g. [**27**,**31**,**35**,**42**]). The Wick product is generally considered a suitable replacement of the usual product, especially when the product is between two generalized random elements for which the usual product is not well defined. Interestingly, although the Wick product was introduced independently in the seemingly unrelated field of quantum field theory, it turns out that the Wick product and the Malliavin divergence operator are closely related [**46**]. Both are stochastic convolutions between two random elements, and moreover, the two concepts coincide in some cases, so that it is possible in these cases to formulate stochastic equations using one or the other framework. Thus, we will see the use of both the Malliavin divergence operator and the Wick product.

The thesis is organized as follows. Chapter 2 introduces the mathematical framework of the Wiener chaos expansion, Malliavin calculus and Wick product. With these tools, we then present a basic technique of applying the propagator system to deduce the solvability of a stochastic parabolic equation under the Malliavin calculus approach. In Chapter 3, we discuss a numerical algorithm based on the finite element method for solving the stochastic parabolic equation, and quantify the error incurred by the numerical solutions by deriving a priori error estimates. We will consider a newly proposed, *unbiased*, random perturbation of the deterministic Navier-Stokes equations, called the quantized stochastic Navier-Stokes equation, in Chapter 4. We will analyze the solvability of the stationary equations, as well as the long time convergence of a time dependent solution to the steady solution. As an application of the Wiener chaos expansion, we will study in Chapter 5 how certain physical models of incoherent forcing sources can exploit a simple change of Wiener chaos expansion basis to drastically reduce computational cost. As the latter three topics are considerably distinct in their nature, we will leave further introduction to each topic to the start of their respective chapters.

CHAPTER 2

The Wiener Chaos Expansion and the Malliavin Calculus Framework

In this chapter, we present the tools and techniques which form the framework for the analysis of solutions of SPDEs, as studied in this thesis. We will first discuss the general construction of Gaussian noise, which is the main source of stochasticity in the random perturbations of stochastic equations. This will lead to the definition of the Wiener chaos expansion and the weighted stochastic spaces. The weighted stochastic spaces are introduced as spaces that our solutions live in, and the Wiener chaos expansion is the way our solutions are represented. We then define the main operators in Malliavin calculus and the Wick product, to be used as the stochastic models for the actual incorporation of the stochasticity into an equation. Finally, with the Wiener chaos expansion and the Malliavin calculus approach, we elucidate a technique for analyzing the solvability of a parabolic SPDE by considering the equivalent propagator system for the chaos modes of the solution.

1. Gaussian white noise and the Wiener Chaos expansion

Let $\xi = \{\xi_k\}_{k\geq 1}$ be a collection of i.i.d. $N(0,1)$ random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the σ-algebra generated by $\{\xi_k\}$. Let $\mathcal{U}$ be a real separable Hilbert space with complete orthonormal basis $\{\mathfrak{u}_k\}_{k\geq 1}$.

Definition 1.1. *The Gaussian white noise on $\mathcal{U}$ is the formal series*

$$\dot{W} := \sum_{k=1}^{\infty} \xi_k \mathfrak{u}_k. \tag{2.1}$$

Note that the white noise is not an element of $L_2(\Omega; \mathcal{U})$, because

$$\mathbb{E}\|\dot{W}\|_{\mathcal{U}}^2 = \sum_{k=1}^{\infty} \|\mathfrak{u}_k\|_{\mathcal{U}}^2 = \infty$$

We list some special examples of Gaussian noise commonly encountered:

1. For the standard 1-dimensional white noise $\dot{W}(t)$, we take $\mathcal{U} = L^2(0,T)$. The basis $\{\mathfrak{u}_k\}$ may, for example, be taken to be the cosine basis in $\mathcal{U}$, but this is not a unique choice of basis. If t represents time, then $\dot{W}(t)$ may be understood as the formal derivative of a 1-dimensional Brownian motion $W(t)$.
2. For stationary or spatial Gaussian white noise on a domain $D \subset \mathbb{R}^d$, we take $\mathcal{U} = L^2(D)$. Then the definition yields that $\dot{W}(x)$ is spatially uncorrelated; that is, for $x, y \in D$,

   $$\mathbb{E}[\dot{W}(x)\dot{W}(y)] = \delta_x(y)$$

   where δ_x is the Dirac delta function at x. Indeed, for smooth ϕ,

   $$\langle \mathbb{E}[\dot{W}(x)\dot{W}(y)], \phi \rangle = \left\langle \sum_{k=1}^{\infty} \mathfrak{u}_k(x)\mathfrak{u}_k(\cdot), \sum_{k=1}^{\infty} \phi_k \mathfrak{u}_k(\cdot) \right\rangle = \sum_{k=1}^{\infty} \phi_k \mathfrak{u}_k(x) = \phi(x)$$

3. If $\mathcal{U} = L_2(0,T;H)$ for some separable Hilbert space H, then the *cylindrical Wiener process* with values in H is

   $$W(t) := \sum_{k=1}^{\infty} \mathfrak{u}_k W_k(t), \tag{2.2}$$

   where $W_k(t), k = 1, 2, \ldots,$ are independent standard Wiener processes. If $H = L_2(D)$, the formal derivative $\dot{W}(t,x)$ is called *space-time white noise.*
4. *Correlated (or weighted) Gaussian noise $\dot{W}_Q(x)$ with covariance operator Q^2.* Let Q be an operator on $\mathcal{U}$ defined by

   $$Q\mathfrak{u}_k = \sigma_k \mathfrak{u}_k, \quad \text{for } k = 1, 2, \ldots.$$

   where $\{\sigma_k,\, k \geq 1\}$ are non-negative real numbers. The Gaussian noise with covariance operator Q^2 is defined by

   $$\dot{W}_Q(x) := \sum_{k=1}^{\infty} \sigma_k \mathfrak{u}_k \xi_k. \tag{2.3}$$

   If Q^2 is nuclear or trace class, i.e., $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$, then it is defined by the covariance function

   $$q(x,y) := \mathbb{E}[\dot{W}_Q(x)\dot{W}_Q(y)] = \sum_{k=1}^{\infty} \sigma_k^2 \mathfrak{u}_k(x)\mathfrak{u}_k(y)$$

as the operator $Q^2 f(x) = \int_D q(x,y)f(y)\,dy$, for $f \in L^2(D)$. The expansion (2.3) is the *Karhunen–Loève expansion* for $\dot{W}_Q$.

5. Let $\dot{W}_1, \dot{W}_2$ be two white noises on $\mathcal{U}_1, \mathcal{U}_2$ respectively,

$$\dot{W}_i := \sum_{k=1}^{\infty} \xi_k^{(i)} \mathfrak{u}_k^{(i)}, \quad \text{for } i = 1, 2.$$

We may assume $\dot{W}_i$ to be independent, or correlated in some way. We can define an abstract white noise $\dot{W}$ to accommodate both noises $\dot{W}_i$ into a single term, by defining $\mathfrak{u}_{2k-1} = \mathfrak{u}_k^{(1)}$, $\mathfrak{u}_{2k} = \mathfrak{u}_k^{(2)}$, and $\xi_{2k-1} = \xi_k^{(1)}$, $\xi_{2k} = \xi_k^{(2)}$. Then

$$\dot{W} := \sum_{k=1}^{\infty} \xi_k \mathfrak{u}_k.$$

The formulation of the abstract noise is useful for studying equations that are driven by two or more distinct white noises, or equations whose input data (initial or boundary conditions, or forcing terms) are measurable with respect to a different Gaussian noise than the noise driving the equation.

In order to develop the L_2 theory of $\mathcal{F}$-measurable random elements and the Wiener chaos expansions, we first introduce some housekeeping tools. Let $\mathcal{J} = \{\alpha = (\alpha_1, \alpha_2, \dots) : \alpha_k \in \mathbb{N}_0\}$ be the set of multi-indices of finite length, $|\alpha| := \sum_{k\geq 1} \alpha_k < \infty$. Denote $\dim(\alpha) = \min\{k : \alpha_\kappa = 0 \text{ for } \kappa > k\}$ and $d(\alpha) = \sum_{k=1}^{\infty} \mathbf{1}_{\alpha_k > 0}$. We denote the zero multi-index by $(0) = (0, 0, \dots)$, and the unit multi-index with 1 in the kth entry by ϵ_k. For $\alpha, \beta \in \mathcal{J}$,

$$\alpha + \beta = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots), \qquad \alpha! := \prod_{k\geq 1} \alpha_k!$$

$$\binom{\alpha+\beta}{\alpha} = \frac{(\alpha+\beta)!}{\alpha!\beta!}, \qquad \binom{|\alpha|}{\alpha} = \frac{|\alpha|!}{\alpha!}$$

For a sequence of nonnegative real numbers $q = (q_1, q_2, \dots)$, define $q^\alpha = \prod_{k\geq 1} q_k^{\alpha_k}$.

A multi-index α can be uniquely characterized by its *characteristic set* K_α. Let $n = |\alpha|$, and denote the ordered n-tuple $K_\alpha = (k_1, \dots, k_n)$ where $k_1 \leq k_2 \leq \dots \leq k_n$, with k_i defined as follows. Let $\kappa_1 < \dots < \kappa_{d(\alpha)}$ be the indices of α for which $\alpha_{\kappa_i} \neq 0$. Then

$$k_l = \kappa_i \quad \text{if} \quad \sum_{j=1}^{i-1} \alpha_{\kappa_j} < l \leq \sum_{j=1}^{i} \alpha_{\kappa_j}.$$

In words, the first α_{κ_1} entries of K_α are $k_1 = \cdots = k_{\alpha_{\kappa_1}} = \kappa_1$, followed by the next α_{κ_2} entries of K_α being $k_{\alpha_{\kappa_1}+1} = \cdots = k_{\alpha_{\kappa_1}+\alpha_{\kappa_1}} = \kappa_2$, etc.

The definitions of the multi-indices and their characteristic sets give rise to some useful combinatorial results which will come in handy later. We state two of these results here.

LEMMA 1.2. *(A multinomial sum in infinite dimensions) Let $\vec{\rho} = (\rho_1, \rho_2, \dots)$ with $\rho_k > 0$, and let $\bar{\rho} = \sum_{k\geq 1} \rho_k$. Then for any $n \in \mathbb{N}_0$,*

$$\sum_{|\alpha|=n} \frac{\rho^\alpha}{\alpha!} = \frac{\bar{\rho}^n}{n!}.$$

PROOF. Fix n and $|\alpha| = n$. We identify α with its characteristic set $K_\alpha = (k_1, \dots, k_n)$. Since there are $n!/\alpha!$ distinct permutations of $\{k_1, \dots, k_n\}$,

$$\sum_{|\alpha|=n} \frac{\rho^\alpha}{\alpha!} = \sum_{k_1 \leq \cdots \leq k_n} \frac{\prod_{j=1}^n \rho_{k_j}}{\alpha!} \cdot \frac{(n!/\alpha!)}{(n!/\alpha!)} = \sum_{k_1,\dots,k_n} \frac{\prod_{j=1}^n \rho_{k_j}}{\alpha!} \cdot \frac{1}{(n!/\alpha!)}$$

where we have multiplied by 1 and rearranged the sum over non-decreasing indices into a sum over all unordered indices. Finally, from the formula for the multinomial expansion

$$\sum_{|\alpha|=n} \frac{\rho^\alpha}{\alpha!} = \frac{1}{n!} \sum_{k_1,\dots,k_n} \prod_{j=1}^n \rho_{k_j} = \frac{1}{n!} \Big(\sum_k \rho_k \Big)^n$$

□

LEMMA 1.3. *For all $\alpha, \beta \in \mathcal{J}$,*

$$\frac{|\beta|!}{\beta!} \frac{|\alpha - \beta|!}{(\alpha - \beta)!} \leq \frac{|\alpha|!}{\alpha!}.$$

PROOF. Let $K_\alpha = (k_1, \dots, k_{|\alpha|})$ be the characteristic set of α. On the RHS, $\frac{|\alpha|!}{\alpha!}$ is the number of distinct permutations of K_α. On the LHS, we partition K_α into the two subsets corresponding to K_β and $K_{(\alpha-\beta)}$. Then, the number of distinct permutations of K_β times that of $K_{(\alpha-\beta)}$ cannot exceed the number of distinct permutations of K_α. □

1.1. The Wiener chaos expansion and weighted Wiener chaos spaces. The Wiener chaos expansion is an orthogonal expansion for random elements that are measurable with respect to the Gaussian white noise $\dot{W}$. Due to the Gaussian assumption, the Wiener chaos expansion is necessarily an expansion in the Hermite polynomials. Recall the Hermite polynomial $H_n(x)$ of degree n,

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}.$$

For each $\alpha \in \mathcal{J}$, define the random variables

$$\xi_\alpha = \prod_{k \geq 1} \frac{H_{\alpha_k}(\xi_k)}{\sqrt{\alpha_k!}}.$$

THEOREM 1.4. *(Cameron and Martin [**9**]) The collection $\Xi = \{\xi_\alpha, \alpha \in \mathcal{J}\}$ is an orthonormal basis of $L_2(\Omega)$. Ξ is referred to as the Cameron-Martin basis.*

Given a real separable Hilbert space X with norm $|\cdot|_X$, let $L_2(\Omega; X)$ be the Hilbert space of square integrable $\mathcal{F}$-measurable random elements with values in X. Then the Cameron-Martin theorem provides that any square integrable random element $\zeta \in L_2(\Omega; X)$ has the *Wiener Chaos expansion* with respect to the Cameron-Martin basis,

$$\zeta = \sum_{\alpha \in \mathcal{J}} \zeta_\alpha \xi_\alpha$$

where $\zeta_\alpha = \mathbb{E}[\zeta \xi_\alpha]$, and Parseval's identity holds,

$$\|\zeta\|_{L_2(\Omega;X)} \equiv \mathbb{E}|\zeta|_X^2 = \sum_{\alpha \in \mathcal{J}} |\zeta_\alpha|_X^2.$$

We will frequently encounter random elements that are not square integrable, and thus we describe a construction analogous to the construction of Sobolev scales. Define the test function space

$$\mathcal{D} = \{\zeta = \sum_\alpha \zeta_\alpha \xi_\alpha : \zeta_\alpha \in \mathbb{R} \text{ and only finite number of } \zeta_\alpha \text{ are non-zero}\}.$$

DEFINITION 1.5. *A* generalized random element *f with values in X is a formal series*

$$f = \sum_{\alpha \in \mathcal{J}} f_\alpha \xi_\alpha, \tag{2.4}$$

where $f_\alpha \in X$. f is identified with the sequence $\{f_\alpha, \alpha \in \mathcal{J}\}$. The expansion (2.4) *is also called the Wiener chaos expansion of f.*

The space $\mathcal{D}'(X)$ of generalized random elements in X is the dual space of $\mathcal{D}$ with respect to $L_2(\Omega)$, with duality pairing

$$\langle\!\langle f, \zeta \rangle\!\rangle = \sum_{\alpha} \zeta_\alpha f_\alpha$$

The space $\mathcal{D}'$ is a very large space. Its elements have Wiener chaos expansions that may exhibit severe blow-up. Next, we introduce the weighted Wiener Chaos spaces that quantify the asymptotic behaviour of the Wiener chaos modes. Let $\mathcal{R}$ be a bounded linear operator on $L_2(\Omega)$ defined by $\mathcal{R}\xi_\alpha = r_\alpha \xi_\alpha$ for every $\alpha \in \mathcal{J}$, where the weights $\{r_\alpha, \alpha \in \mathcal{J}\}$ are positive real numbers. Note that $\mathcal{R}$ is bounded if and only if the weights r_α are uniformly bounded from above, that is, $r_\alpha < C$ for all $\alpha \in \mathcal{J}$, for some constant C. Define the norm

$$\|f\|^2_{\mathcal{R}L_2(\Omega;X)} := \sum_{\alpha \in \mathcal{J}} |f_\alpha|^2_X r^2_\alpha$$

for $f = \sum_{\alpha \in \mathcal{J}} f_\alpha \xi_\alpha$. The space $\mathcal{R}L_2(\Omega; X)$ of random elements in X, is defined as the closure of $L^2(\Omega; X)$ under the norm $\| \cdot \|_{\mathcal{R}L_2(\Omega;X)}$; in other words, the elements of $\mathcal{R}L_2(\Omega; X)$ are identified with a formal series $\sum_{\alpha \in \mathcal{J}} f_\alpha \xi_\alpha$, where $\|f\|^2_{\mathcal{R}L_2(\Omega;X)} < \infty$. Clearly, $\mathcal{R}L_2(\Omega; X)$ is a Hilbert space with respect to $\| \cdot \|_{\mathcal{R}L_2(\Omega;X)}$.

The operator $\mathcal{R}^{-1}$ that is inverse to $\mathcal{R}$ is defined by $\mathcal{R}^{-1}\xi_\alpha = r_\alpha^{-1}\xi_\alpha$. Let $X \hookrightarrow Y \hookrightarrow X'$ be a normal triple of Hilbert spaces with duality pairing $\langle \cdot, \cdot \rangle_{X',X}$. We define the space $\mathcal{R}^{-1}L_2(\Omega; X)$ as the dual of $\mathcal{R}L_2(\Omega; X')$ relative to the inner product in the space $L_2(\Omega; Y)$. The duality pairing is given by

$$\langle\!\langle f, g \rangle\!\rangle_{\mathcal{R}L_2(\Omega;X'),\mathcal{R}^{-1}L_2(\Omega;X)} := \mathbb{E}\langle \mathcal{R}f_\alpha, \mathcal{R}^{-1}g_\alpha \rangle_{X',X} = \sum_{\alpha \in \mathcal{J}} \langle f_\alpha, g_\alpha \rangle_{X',X}$$

for $f \in \mathcal{R}L_2(\Omega; X')$ and $g \in \mathcal{R}^{-1}L_2(\Omega; X)$. Similarly, $\mathcal{R}^{-1}L_2(\Omega; X')$ is defined as the dual of $\mathcal{R}L_2(\Omega; X)$ relative to the inner product in $L_2(\Omega; Y)$. We may leave out notating the dual spaces in $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ where it is either obvious or inconsequential.

There are several classes of weights in the literature [**28**, **31**, **37**, **47**]. We list a few.

(1) In Section 3 and Chapter 3, we consider only admissible weights of the form

$$r^2_\alpha = \frac{q^\alpha}{|\alpha|!},$$

where $q = (q_1, q_2, \dots)$ is a decreasing sequence of positive real numbers. This class of weights arises naturally, for example, when studying equations where the driving white noise acts on the term of the second order partial differential operator [**44**, **48**, **49**].

(2) *Kondratiev spaces.* Denote the sequence $(2\mathbb{N})^{-q} := \big((2k)^{-q}\big)_{k=1,2,\dots}$. The Kondratiev space $\mathcal{S}_{-1,-q}(X)$ is a weighted space with weights

$$r_\alpha^2 = \frac{\big((2\mathbb{N})^{-q}\big)^\alpha}{\alpha!}$$

The Kondratiev spaces have been widely used to study various classes of SPDE (see e.g., [**10**, **29**, **31**]).

2. Generalized Malliavin calculus and the Wick product

In this section, we discuss a generalized form of the Malliavin calculus, and also its relation to the Wick product. Roughly speaking, the traditional development of the subject of the Malliavin calculus begins with defining the Malliavin derivative $\boldsymbol{D}_{\dot{W}}$ with respect to Gaussian white noise,

$$\boldsymbol{D}_{\dot{W}} F(W(h_1), \dots, W(h_N)) = \sum_{i=1}^{N} \frac{\partial F}{\partial x_i}(W(h_1), \dots, W(h_N)) h_i,$$

for a smooth function F and $h_i \in \mathcal{U}$, $i = 1, \dots, N$, and where $W(h_i) = \int_0^T h_i dW(t)$. The Malliavin derivative maps elements in $L_2(\Omega)$ into $L_2(\Omega; \mathcal{U})$. The Skorokhod integral $\boldsymbol{\delta}_{\dot{W}}$ is then a map $L_2(\Omega; \mathcal{U})$ into $L_2(\Omega)$, defined via the adjoint property,

$$\mathbb{E}[\boldsymbol{\delta}_{\dot{W}}(f)\phi] = \mathbb{E}[(f, \boldsymbol{D}_{\dot{W}}\phi)_{\mathcal{U}}].$$

(See [**50**, **56**] for details.)

The generalized form of the Malliavin calculus retains the properties of the traditional Malliavin calculus, including the adjoint property, but accords more flexibility when it comes to defining the Malliavin derivative and Malliavin divergence operator with respect to other random elements besides Gaussian white noise. For any ξ_k, define the Malliavin derivative

$\boldsymbol{D}_{\xi_k}$ and Malliavin divergence operator $\boldsymbol{\delta}_{\xi_k}$ by[1]

$$\boldsymbol{D}_{\xi_k}(\xi_\alpha) := \sqrt{\alpha_k}\xi_{\alpha-\epsilon_k}, \qquad \text{and} \qquad \boldsymbol{\delta}_{\xi_k}(\xi_\alpha) := \sqrt{\alpha_k+1}\xi_{\alpha+\epsilon_k}.$$

The Malliavin operators $\boldsymbol{D}_{\xi_k}$, $\boldsymbol{\delta}_{\xi_k}$ can be extended to any Cameron-Martin basis element ξ_β by

$$\boldsymbol{D}_{\xi_\beta}(\xi_\alpha) := \sqrt{\binom{\alpha}{\beta}}\xi_{\alpha-\beta}, \qquad \text{and} \qquad \boldsymbol{\delta}_{\xi_\beta}(\xi_\alpha) := \sqrt{\binom{\alpha+\beta}{\beta}}\xi_{\alpha+\beta}.$$

An important relationship between the Malliavin derivative and Malliavin divergence operator is the *adjoint property*: for $\alpha, \alpha', \beta \in \mathcal{J}$,

$$\langle\!\langle \boldsymbol{\delta}_{\xi_\beta}(\xi_\alpha), \xi_{\alpha'} \rangle\!\rangle = \langle\!\langle \xi_\alpha, \boldsymbol{D}_{\xi_\beta}(\xi_{\alpha'}) \rangle\!\rangle \tag{2.5}$$

By bilinearity, $\boldsymbol{D}_u(v)$ and $\boldsymbol{\delta}_u(f)$ can be defined for random elements u on $\mathcal{U}$, v on X, and f on $X \otimes \mathcal{U}$ [**46**]. Elementary computations with the Wiener chaos expansion yields explicit formulas for $\boldsymbol{D}_u(v)$ and $\boldsymbol{\delta}_u(f)$, as follows. Let $u = \sum_\alpha u_\alpha \xi_\alpha$ with $u_\alpha \in \mathcal{U}$, and $v = \sum_\alpha v_\alpha \xi_\alpha$ with $v_\alpha \in X$, and $f = \sum_\alpha f_\alpha \xi_\alpha$ with $f_\alpha \in X \otimes \mathcal{U}$. Then

$$\boldsymbol{D}_u(v) = \sum_\alpha \left(\sum_\beta \sqrt{\binom{\alpha+\beta}{\beta}} v_{\alpha+\beta} \otimes u_\beta \right) \xi_\alpha,$$

$$\boldsymbol{\delta}_u(f) = \sum_\alpha \left(\sum_{\beta \le \alpha} \sqrt{\binom{\alpha}{\beta}} (f_\beta, u_{\alpha-\beta})_{\mathcal{U}} \right) \xi_\alpha.$$

The Malliavin divergence operator is closely related to the Wick product. The Wick product is defined as

$$\xi_\alpha \diamond \xi_\beta = \sqrt{\binom{\alpha+\beta}{\beta}} \xi_{\alpha+\beta}$$

and is extended by linearity to $f \diamond \eta$, where f is a generalized X-valued random element and η is a generalized real-valued random element. Clearly, $f \diamond \eta = \boldsymbol{\delta}_\eta(f)$ in this case (with $\mathcal{U} = \mathbb{R}$). The difference between the Malliavin divergence operator and the Wick product lies in the fact that the Wick product is a point-wise product between X-valued and $\mathbb{R}$-valued generalized random elements, whereas the Malliavin divergence operator, being a stochastic integral, is a convolution between $\mathcal{U}$-valued and $\mathcal{U} \otimes X$-valued generalized random elements.

[1] Recall the multi-index notation on page 7.

Thus, the Wick product is a symmetric operator, whereas the Malliavin divergence operator is not symmetric (see [**46**]).

Next, we describe some situations in which we will encounter the Malliavin derivative, Malliavin divergence operator and the Wick product:

(1) We can define the Malliavin divergence operator with respect to white noise $\dot{W}$, which is a random element on $\mathcal{U}$. For $f \in \mathcal{R}L_2(\Omega; X \otimes \mathcal{U})$, $\boldsymbol{\delta}_{\dot{W}}(f)$ is the unique element of $\mathcal{R}L_2(\Omega; X)$ with the property that

$$\langle\!\langle \boldsymbol{\delta}_{\dot{W}}(f), \varphi \rangle\!\rangle_{\mathcal{R}L_2(\Omega;X),\mathcal{R}^{-1}L_2(\Omega;X')} = \langle\!\langle f, \boldsymbol{D}_{\dot{W}}(\varphi) \rangle\!\rangle_{\mathcal{R}L_2(\Omega;X\otimes\mathcal{U}),\mathcal{R}^{-1}L_2(\Omega;X'\otimes\mathcal{U})}$$

for every $\varphi \in \mathcal{R}^{-1}L_2(\Omega; X')$ such that $\boldsymbol{D}_{\dot{W}}(\varphi) \in \mathcal{R}^{-1}L_2(\Omega; X' \otimes \mathcal{U})$.

In the case of time white noise on a finite time interval, i.e., $\mathcal{U} = L_2(0, T)$, it can be shown that the Malliavin divergence operator coincides with the Itô integral under the assumption of adaptedness [**56**]. That is,

$$\boldsymbol{\delta}_{\dot{W}}(u) = \int_0^T u(t) dW(t),$$

provided $u(t)$ is a suitable random element that is adapted to the filtration generated by the Brownian motion $W(t)$.

(2) For a random element $g \in \mathcal{R}L_2(\Omega; X)$, it will often arise in modelling problems to consider the multiplication, or convolution, of g with white noise $\dot{W}$. Strictly speaking, the term $\boldsymbol{\delta}_{\dot{W}}(g)$ is not well defined. But, by abuse of notation, we interpret

$$[\boldsymbol{\delta}_{\dot{W}}(g)]_\alpha = \sum_{k\geq 1} \sqrt{\alpha_k} g_{k,\alpha-\epsilon_k}$$

where $g_{k,\alpha} = \mathfrak{u}_k \otimes g_\alpha$.

When $\dot{W}(x)$ is a white noise on $L_2(D)$ with orthonormal basis $\{\mathfrak{u}_k(x)\}$, and when $g(x)$ is a generalized random function in x, this interpretation coincides with the Wick product model $g(x) \diamond \dot{W}(x)$ by taking the Wick product pointwise in x. To see this, choose $g_{k,\alpha}(x) = \mathfrak{u}_k(x) g_\alpha(x)$, then

$$\Big(g(x) \diamond \dot{W}(x)\Big)_\alpha = \sum_l \sqrt{\alpha_l}\, g_{\alpha-\epsilon_l}(x) \mathfrak{u}_l(x) = \boldsymbol{\delta}_{\dot{W}}(g)$$

If $\{\mathfrak{u}_k\}$ is chosen such that $D^\gamma \mathfrak{u}_k \in L^\infty$ for all k and all $\gamma = (\gamma_1, \dots, \gamma_d)$ with $|\gamma| \leq l$, then $g \diamond \dot{W} \in \mathcal{D}'(W^{l,p})$ provided $g \in \mathcal{D}'(W^{l,p})$.

(3) In Chapter 4, we will consider a nonlinearity of the form $u^i \diamond \partial_{x_i} u$. Direct computation gives that

$$(u^i \diamond \partial_{x_i} u)_\alpha = \sum_{0 \leq \gamma \leq \alpha} \sqrt{\binom{\alpha}{\gamma}}\, (u_\gamma, \nabla) u_{\alpha-\gamma}. \tag{2.6}$$

Each chaos mode of the Wick product is determined by a convolution among between the lower order chaos modes. This observation is important, as it suggests a connection to the Catalan numbers which are themselves characterized recursively as convolutions.

3. A basic application of the Wiener chaos expansion to solving SPDEs via the propagator system

The Wiener chaos expansion is used as a separation of variables technique for stochastic ordinary or partial differential equations. Just like how the Fourier expansion is used to solve for the Fourier modes of a solution of a deterministic PDE, the Wiener chaos expansion is used in an analogous way to find the Wiener chaos modes of the solution of an SPDE. The separation of the independent variable of randomness leaves behind a deterministic system of equations for the Wiener chaos modes of the solution, termed the *propagator system* of equations. The analysis of the SPDE is thus reduced to the analysis of a deterministic PDE system for which deterministic theory can be applied.

In this section, we show a basic application of the Wiener chaos expansion to study the parabolic SPDE on a bounded domain $D \subset \mathbb{R}^d$,

$$\begin{aligned} &\tfrac{du}{dt} + \mathcal{A}u + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}u + g) = f \quad \text{on } D \times (0, T] \\ &u|_{\partial D} = 0 \\ &u|_{t=0} = v \end{aligned} \tag{2.7}$$

where $\dot{W}$ is a Gaussian white noise which may depend on space or time, $\mathcal{A}$ is a second order elliptic operator from $H_0^1(D)$ onto $H^{-1}(D)$, and $\mathcal{M}u := \sum_k \mathcal{M}_k u \otimes \mathfrak{u}_k$ where $\mathcal{M}_k$, $k = 1, 2, \dots$ are bounded operators from $H_0^1(D)$ into $H^{-1}(D)$. We will assume that the boundary ∂D and the coefficients of $\mathcal{A}, \mathcal{M}_k$ are sufficiently smooth in the space and time variables. The input data v, f, g are allowed to be generalized random elements, and we recall that they can be measurable with respect to a white noise different from $\dot{W}$.

3.1. Some notation and constants. Before we proceed, we state our notation for various constants that will show up often in the rest of this chapter and the next chapter.

We assume throughout that $\mathcal{A}(t)$ is uniformly elliptic on $(0, T]$ and is coercive and bounded,

$$\begin{aligned} \langle \mathcal{A}(t)u, u\rangle &\geq C_A^{coerc} \|u\|_{H_0^1}^2, \\ \langle \mathcal{A}(t)u, v\rangle &\leq C_A^b \|u\|_{H_0^1} \|v\|_{H_0^1}. \end{aligned} \tag{2.8}$$

Denote $C_A^{ellip} = (C_A^{coerc})^{-1}$ to be the constant in

$$\|w\|_{H_0^1} \leq C_A^{ellip} \|f\|_{H^{-1}} \tag{2.9}$$

for the solution of the zero Dirichlet problem $\mathcal{A}(t)w = f$, for any $t \in (0, T]$. Also denote C_A to be the constant in

$$\|w\|_{L^2(0,T;H_0^1(D))} \leq C_A(\|w_0\|_{L^2(D)} + \|f\|_{L^2(0,T;H^{-1}(D))}) \tag{2.10}$$

for the weak solution w of the zero Dirichlet problem $\frac{dw}{dt} + \mathcal{A}(t)w = f$ with $w(0) = w_0$.

For the Sobolev spaces $H^r(D)$, $r \geq 1$, let $\lambda_k^{(r)}$ be the constants in

$$\|\mathcal{M}_k(t)w\|_{H^{r-2}(D)} \leq \lambda_k^{(r)} \|w\|_{H^r(D)}, \qquad \forall w \in H^r(D), t \in (0, T] \tag{2.11}$$

For brevity, we write $\lambda_k = \lambda_k^{(1)}$. Observe that $C_k := \lambda_k C_A^{ellip}$ are the constants defined by $\|\mathcal{A}^{-1}\mathcal{M}_k v\|_{H_{0X}^1} \leq C_k \|v\|_{H_{0X}^1}$ for all $v \in H_{0X}^1$.

Finally, let $\mu_k^{(r)}$, $r = -1, 0, 1, \ldots$ be the constant arising in $\|g_k\|_{H_X^r} \leq \mu_k^{(r)} \|g\|_{H_X^r}$. We will write $\mu_k = \mu_k^{(-1)}$.

We will use shorthand to denote the spaces: for example, we will write $\mathcal{R}_\Omega L_T^2 H_X^{-1}$ to denote $\mathcal{R}L_2(\Omega; L^2((0, T); H^{-1}(D)))$. Also, H_{0X}^1 denotes $H_0^1(D)$.

3.2. Existence and uniqueness of solutions. We begin by defining the notion of a weak solution of (2.7).

Definition 3.1. *A weak solution of* (2.7)*, with* $f, g \in \mathcal{R}_\Omega L^2_T H^{-1}_X$ *and* $v \in \mathcal{R}_\Omega L^2_X$*, is a process* $u \in \mathcal{R}_\Omega L^2_T H^1_{0X}$ *such that for every* $\phi \in \mathcal{R}^{-1}_\Omega$ *with* $\boldsymbol{D}_{\dot W}\phi \in \mathcal{R}^{-1}_\Omega \mathcal{U}$*,*

$$\langle\langle u(t), \phi\rangle\rangle = \langle\langle v, \phi\rangle\rangle - \int_0^t \langle\langle \mathcal{A}u + \boldsymbol{\delta}_{\dot W}(\mathcal{M}u + g),\, \phi\rangle\rangle ds + \int_0^t \langle\langle f, \phi\rangle\rangle ds \tag{2.12}$$

with equality in $L^2_T H^{-1}_X$*.*

The existence and uniqueness result for a weak solution of (2.7) for v, f deterministic and $g \equiv 0$ has been shown in [**49**]. The proof relies on the Equivalence Theorem 3.2 that relates the weak solution to the *propagator system* (2.13).

Theorem 3.2. *The process* $u = \sum_\alpha u_\alpha \xi_\alpha \in \mathcal{R}_\Omega L^2_T H^1_{0X}$ *is a solution of* (2.7)*, if and only if, for each* $\alpha \in \mathcal{J}$*,*

$$u_\alpha(t) = v_\alpha - \int_0^t \mathcal{A}u_\alpha(s) + \sum_{k\geq 1} \sqrt{\alpha_k}(\mathcal{M}_k u_{\alpha-\epsilon_k} + g_{k,\alpha-\epsilon_k})\, ds + \int_0^t f_\alpha(s) ds \tag{2.13}$$

holds in H^{-1}_X *for a.e.* $t \in [0, T]$*.*

Proof. See [**49**]. □

Using the techniques from [**47**] (Theorem 9.4) or [**45**] (Proposition 4.2), we can extend the existence and uniqueness result to the case when v, f, g are random, and determine the conditions for the weighted spaces that u may belong to, in terms of the spaces that the input data belong to.

Theorem 3.3. *Let the weights* $\mathcal{R}$*, with* $r^2_\alpha = \frac{q^\alpha}{|\alpha|!}$*, satisfy*

$$\sum_{k\geq 1} q_k C^2_A \lambda^2_k < 1. \tag{2.14}$$

(1) *If the input data* $v \in L^2_X$ *and* $f, g \in L^2_T H^{-1}_X$ *are deterministic, then there exists a unique weak solution* $u \in \mathcal{R}_\Omega L^2_T H^1_{0X}$*, and*

$$\|u\|_{\mathcal{R}_\Omega L^2_T H^1_{0X}} \leq C\left(\|v\|_{L^2_X} + \|f\|_{L^2_T H^{-1}_X} + \|g\|_{L^2_T H^{-1}_X}\right)$$

where C *depends only on* $\mathcal{R}, \mathcal{A}, \mathcal{M}$ *and* T*.*

(2) Assume $v \in \bar{\mathcal{R}}_\Omega L^2_X$ and $f, g \in \bar{\mathcal{R}}_\Omega L^2_T H^{-1}_X$ for some $\bar{r}^2_\alpha = \frac{\rho^\alpha}{|\alpha|!}$. Also assume, in addition to (2.14), that q_k are chosen to satisfy

$$\sum_{k\geq 1} \frac{q_k}{\rho_k} < 1 \tag{2.15}$$

Then there exists a unique solution $u \in \mathcal{R}_\Omega L^2_T H^1_{0X}$, and

$$\|u\|_{\mathcal{R}_\Omega L^2_T H^1_{0X}} \leq C \left(\|v\|_{\bar{\mathcal{R}}_\Omega L^2_X} + \|f\|_{\bar{\mathcal{R}}_\Omega L^2_T H^{-1}_X} + \|g\|_{\bar{\mathcal{R}}_\Omega L^2_T H^{-1}_X} \right)$$

where C depends only on $\mathcal{R}, \bar{\mathcal{R}}, \mathcal{A}, \mathcal{M}$ and T.

Proof. Step 1.

Assume v, f, g are non-random. This case has been studied in [**49**] for $g = 0$. The proof here is essentially the same. The propagator system is

$$\begin{aligned} u_{(0)}(t) &= v + \int_0^t \mathcal{A}u_{(0)}(s) + f(s)ds \\ u_{\epsilon_k}(t) &= \int_0^t \mathcal{A}u_{\epsilon_k}(s) + \big(\mathcal{M}_k u_{(0)}(s) + g_k(s)\big)ds \\ u_\alpha(t) &= \int_0^t \mathcal{A}u_\alpha(s) + \sum_k \sqrt{\alpha_k}\mathcal{M}_k u_{\alpha-\epsilon_k}(s)ds, \quad |\alpha| \geq 2 \end{aligned}$$

Let $\Phi_t = e^{\mathcal{A}}t$ be the semigroup generated by $\mathcal{A}$. Then for $\alpha = (0)$,

$$u_{(0)}(t) = \Phi_t v + \int_0^t \Phi_{t-s} f(s)ds$$

and from the deterministic parabolic estimates,

$$\|u_{(0)}\|_{L^2_T H^1_{0X}} \leq C_A(\|v\|_{L^2_X} + \|f\|_{L^2_T H^{-1}_X})$$

For $\alpha = \epsilon_k$,

$$u_{\epsilon_k}(t) = \int_0^t \Phi_{t-s_1}\big(\mathcal{M}_k u_{(0)}(s_1) + g_k\big)ds_1$$

and again applying the deterministic parabolic estimates,

$$\begin{aligned}\|u_{\epsilon_k}\|_{L^2_T H^1_{0X}} &\le C_A\left(\|\mathcal{M}_k u_{(0)}\|_{L^2_T H^{-1}_X} + \|g_k\|_{L^2_T H^{-1}_X}\right)\\ &\le C_A\left(\lambda_k C_A \|u_{(0)}\|_{L^2_T H^1_{0X}} + \mu_k \|g\|_{L^2_T H^{-1}_X}\right)\\ &\le C_A M\big(\vec{\lambda} C_A\big)(\|v\|_{L^2_X} + \|f\|_{L^2_T H^{-1}_X} + \|g\|_{L^2_T H^{-1}_X})\end{aligned}$$

where $M = \sup_k(1 \vee \frac{\mu_k}{\lambda_k C_A})$.

For $|\alpha| = n \ge 2$, with characteristic set $K_\alpha = (k_1, \dots, k_n)$, it can be shown by induction that

$$u_\alpha(t) = \frac{1}{\sqrt{\alpha!}} \sum_{\sigma \in \mathcal{P}_n} \int_0^t \int_0^{s_n} \dots \int_0^{s_2} \Phi_{t-s_n} \mathcal{M}_{k_{\sigma(n)}} \dots \Phi_{s_2 - s_1}\big(\mathcal{M}_{k_{\sigma(1)}} u_{(0)}(s_1) + g_{k_{\sigma(1)}}\big) ds_1 \dots ds_n$$

where $\mathcal{P}_n$ is the group of permutations of $\{1, \dots, n\}$. Then, we obtain

$$\|u_\alpha\|_{L^2_T H^1_{0X}} \le C_A M \frac{\big(\vec{C}\big)^\alpha |\alpha|!}{\sqrt{\alpha!}} (\|v\|_{L^2_X} + \|f\|_{L^2_T H^{-1}_X} + \|g\|_{L^2_T H^{-1}_X})$$

where $\vec{C} = (C_1, C_2, \dots)$, $C_k = \lambda_k C_A$.

Taking the weights to satisfy (2.14), it follows from Lemma 1.2 that

$$\|u\|_{\mathcal{R}\Omega L^2_T H^1_{0X}} \le C(\|v\|_{L^2_X} + \|f\|_{L^2_T H^{-1}_X} + \|g\|_{L^2_T H^{-1}_X})$$

where C depends only on $\mathcal{R}, \mathcal{A}, \mathcal{M}$ and T.

Step 2.

Fix an arbitrary $\alpha^* \in \mathcal{J}$. Assume $v = V\xi_{\alpha^*}, f = F\xi_{\alpha^*}, g = G\xi_{\alpha^*}$; in other words, the randomness of the data is localized to a single mode. Let $u[\alpha^*; V, F, G](t, x)$ be the solution. By linearity, the chaos expansion coefficients with indices of the form $\alpha^* + \alpha$ satisfy

$$\frac{u_{\alpha^*+\alpha}[\alpha^*; V, F, G]}{\sqrt{(\alpha^* + \alpha)!}} = \frac{u_\alpha[(0); \frac{V}{\sqrt{\alpha^*!}}, \frac{F}{\sqrt{\alpha^*!}}, \frac{G}{\sqrt{\alpha^*!}}]}{\sqrt{\alpha!}}$$

and are zero otherwise. Then

$$\begin{aligned}&\int_0^T \|u[\alpha^*; V, F, G](t)\|^2_{\mathcal{R}\Omega H^1_{0X}} dt\\ &= \sum_\alpha \frac{q^{\alpha^*+\alpha}}{|\alpha^* + \alpha|!} \frac{(\alpha^* + \alpha)!}{\alpha!} \left\| u_\alpha \left[(0); \frac{V}{\sqrt{\alpha^*!}}, \frac{F}{\sqrt{\alpha^*!}}, \frac{G}{\sqrt{\alpha^*!}}\right]\right\|^2_{L^2_T H^1_{0X}}\end{aligned}$$

$$
= \sum_{\alpha} \frac{q^{\alpha^*+\alpha}}{|\alpha|!|\alpha^*|!} \frac{|\alpha|!|\alpha^*|!}{|\alpha^*+\alpha|!} \frac{(\alpha^*+\alpha)!}{\alpha!\alpha^*!} \|u_\alpha[(0); V, F, G]\|^2_{L^2_T H^1_{0X}}
$$

$$
\leq \frac{q^{\alpha^*}}{|\alpha^*|!} \|u[(0); V, F, G]\|^2_{\mathcal{R}_\Omega L^2_T H^1_{0X}}
$$

where the last inequality follows by Lemma 1.3.

Step 3.

For the general case with random data, assume $v \in \bar{\mathcal{R}}_\Omega L^2_X$ and $f, g \in \bar{\mathcal{R}}_\Omega L^2_T H^{-1}_X$. The solution can be written as

$$
u = \sum_{\alpha^*} u[\alpha^*; v_{\alpha^*}, f_{\alpha^*}, g_{\alpha^*}]
$$

Using the estimates from Step 2,

$$
\|u\|_{\mathcal{R}_\Omega L^2_T H^1_{0X}} \leq \sum_{\alpha^*} \|u[\alpha^*; v_{\alpha^*}, f_{\alpha^*}, g_{\alpha^*}]\|_{\mathcal{R}_\Omega L^2_T H^1_{0X}}
$$

$$
\leq C \left(\sum_{\alpha^*} \frac{q^{\alpha^*}}{|\alpha^*|!} \frac{|\alpha^*|!}{\rho^{\alpha^*}} \right)^{1/2} \left(\sum_{\alpha^*} \frac{\rho^{\alpha^*}}{|\alpha^*|!} \left(\|v_{\alpha^*}\|_{L^2_X} + \|f_{\alpha^*}\|_{L^2_T H^{-1}_X} + \|g_{\alpha^*}\|_{L^2_T H^{-1}_X} \right)^2 \right)^{1/2}
$$

$$
\leq C \left(\|v\|_{\bar{\mathcal{R}}_\Omega L^2_X} + \|f\|_{\bar{\mathcal{R}}_\Omega L^2_T H^{-1}_X} + \|g\|_{\bar{\mathcal{R}}_\Omega L^2_T H^{-1}_X} \right)
$$

where we have applied Cauchy-Schwartz inequality in the second inequality. The convergence of $\left(\sum_{\alpha^*} \frac{q^{\alpha^*}}{|\alpha^*|!} \frac{|\alpha^*|!}{\rho^{\alpha^*}} \right)$ follows from a sufficient condition such as (2.15).

Clearly, $\mathcal{R} \supseteq \bar{\mathcal{R}}$, so u is a weak solution of (2.7) in the sense of Definition 3.1. Uniqueness follows from the uniqueness of each equation in the propagator system. □

Remark. The validity of the assumption that $M := \sup_k (1 \vee \frac{\mu_k}{\lambda_k}) < \infty$ arises in some common examples. For example, taking $\mathcal{M}_k \phi = \mathfrak{u}_k \Delta \phi$ and $g_k = \mathfrak{u}_k g$, we have that μ_k, λ_k are both $\sim \mathcal{O}(k)$. If $M = \infty$, then in the estimate for $\|u_\alpha\|_{L^2_T H^1_{0X}}$ in Step 1, we should replace the factor $M \vec{\lambda}^\alpha$ by $(\vec{\lambda} C_A \vee \vec{\mu})^\alpha$, and use the criterion $\sum_k q_k (\lambda_k C_A \vee \mu_k)^2 < 1$ in place of (2.14).

Remark. If the input data is non-random, then it belongs to any weighted space $\bar{\mathcal{R}}$ for any ρ. In this case, condition (2.15) is automatically satisfied, and the condition for optimal solution weights $\mathcal{R}$ reduces to (2.14) alone.

3.3. Higher regularity of solutions. The weak solution of (2.7) is a generalized process in H^1_{0X}, and we now investigate when it possesses better smoothness in the spatial variable. Such a result will become useful in the analysis of the stochastic finite element method in Chapter 3, because the spatial regularity of the solution is closely related to the convergence order of finite element schemes. In fact, within the limitations of our analysis in Chapter 3, we require at the least that the minimum spatial regularity of the solution u be $H^2(D)$-smooth, and its time derivatives u_t, u_{tt} be $L_2(D)$ and $H^{-1}(D)$ functions respectively. However, obtaining higher spatial regularity comes at the expense of worsening the weights $\mathcal{R}$.

In line with the strategy of the previous section, we derive higher regularity results from the analogous results in the deterministic theory. Thus, we will see, for example, that certain *compatibility conditions* at time $t = 0$ are necessary conditions for higher regularity to hold, except that the compatibility conditions in the stochastic case are more extensive than those in the deterministic case.

We now recall a result from deterministic PDE.

Theorem 3.4. *(Evans [**14**], Theorems 5 and 6 in §7.1.3[2]). Let $\mathcal{A}$ be a uniformly elliptic second order operator whose coefficients belong to $H^1_T W^{k,\infty}_X$. Suppose $u \in L^2_T H^1_{0X}$ with $u_t \in L^2_T H^{-1}_X$ is the weak solution of*

$$\begin{aligned} &u_t + \mathcal{A}u = f \quad \textit{in } D \times (0,T] \\ &u|_{\partial D} = 0 \\ &u(0) = v \end{aligned}$$

(i) Assume

$$v \in H^1_{0X}, \qquad f \in L^2_T L^2_X.$$

Then in fact $u \in L^2_T H^2_X \cap L^\infty_T H^1_{0X}$ and $u_t \in L^2_T L^2_X$, and

$$\operatorname*{ess\,sup}_{0 \le t \le T} \|u(t)\|_{H^1_{0X}} + \|u\|_{L^2_T H^2_X} + \|u_t\|_{L^2_T L^2_X} \le C_0^{reg} \left(\|v\|_{H^1_{0X}} + \|f\|_{L^2_T L^2_X} \right)$$

where the constant C_0^{reg} depends only on D, T and $\mathcal{A}$.

(ii) Fix $m \ge 1$. Assume

$$v \in H^{2m+1}_X, \qquad \frac{d^k f}{dt^k} \in L^2_T H^{2m-2k}_X \quad \textit{for } k = 0, \dots, m$$

[2] The statement of the results assumes that the operator $\mathcal{A}$ does not depend on time. A careful analysis of the proof shows that a similar result holds for time dependent operators under the assumptions on the coefficients described in this section.

and suppose the m-th order compatibility conditions hold:

$$\begin{cases} v_0 := V_0 \in H^1_{0X}, \quad V_1 := f(0) - \mathcal{A}V_0 \in H^1_{0X}, \ldots, \\ V_m := \frac{d^{m-1}f}{dt^{m-1}} - \mathcal{A}V_{m-1} \in H^1_{0X}. \end{cases}$$

Then $\frac{d^k u}{dt^k} \in L^2_T H^{2m+2-2k}_X$ *for* $k = 0, \ldots m+1$, *and*

$$\sum_{k=0}^{m} \left\| \frac{d^k u}{dt^k} \right\|_{L^2_T H^{2m+2-2k}_X} \leq C^{reg}_m \left(\|v\|_{H^{2m+1}_X} + \sum_{k=0}^{m} \left\| \frac{d^k f}{dt^k} \right\|_{L^2_T; H^{2m-2k}_X} \right)$$

where the constant C^{reg}_m *depends only on* m, D, T *and* $\mathcal{A}$.

From Theorem 3.4(i), we can obtain the following higher regularity result for the stochastic equation (2.7), with deterministic input data. The case of random data can be shown in the same way as Steps 2 and 3 in the proof of Theorem 3.3.

Corollary 3.5. *Suppose* $u \in \mathcal{R}_\Omega L^2_T H^1_{0X}$ *is the weak solution of the SPDE* (2.7). *Also assume that* v, f, g *are deterministic with*

$$v \in H^1_{0X}, \quad \text{and} \quad f, g \in L^2_T L^2_X.$$

Then for the weights $\tilde{\mathcal{R}}$ *satisfying*

$$\sum_k \tilde{\rho}_k (\lambda^{(2)}_k C^{reg}_0)^2 < 1,$$

the weak solution $u \in \tilde{\mathcal{R}}_\Omega L^2_T H^2_X$ *and*

$$\|u\|_{\tilde{\mathcal{R}}_\Omega L^2_T H^2_X} \leq C \left(\|v\|_{H^1_{0X}} + \|f\|_{L^2_T L^2_X} + \|g\|_{L^2_T L^2_X} \right)$$

Proof. The proof is similar to the proof of Theorem 3.3. The estimates for each u_α are obtained by applying Theorem 3.4(i) to the propagator system. □

No special compatibility conditions were necessary for Corollary 3.5, but it is unable to ensure boundedness of u_{tt}. Thus, we next show how to obtain a smoother solution and the boundedness of u_{tt} using the 1st order compatibility conditions.

Corollary 3.6. *Suppose $u \in \mathcal{R}_\Omega L^2_T H^1_{0X}$ is the weak solution of the SPDE* (2.7)*. Also assume that v, f, g are deterministic with*

$$v \in H^3_X, \quad \textit{and} \quad f, g \in L^2_T H^2_X, \quad \textit{and} \quad \frac{df}{dt}, \frac{dg}{dt} \in L^2_T L^2_X,$$

and that the 1st order compatibility conditions hold for $\{v, f, g_k\}$:

$$\begin{cases} v \in H^1_{0X}, \quad f(0) - \mathcal{A}v \in H^1_{0X}, \\ \mathcal{M}_k v + g_k(0) \in H^1_{0X} \quad \forall k = 1, 2, \dots \end{cases} \tag{2.16}$$

Then for the weights $\mathcal{R}'$ satisfying

$$\sum_k \rho'_k \Big((\lambda_k^{(4)} \vee \lambda_k^{(2)}) C_1^{reg} \Big)^2 < 1, \tag{2.17}$$

the weak solution $u \in \mathcal{R}'_\Omega L^2_T H^4_X$, $u_t \in \mathcal{R}'_\Omega L^2_T H^2_X$ and $u_{tt} \in \mathcal{R}'_\Omega L^2_T L^2_X$ and

$$\begin{aligned} &\|u\|_{\mathcal{R}'_\Omega L^2_T H^4_X} + \|u_t\|_{\mathcal{R}'_\Omega L^2_T H^2_X} + \|u_{tt}\|_{\mathcal{R}'_\Omega L^2_T L^2_X} \\ &\quad \le C\big(\|v\|_{H^3_X} + \|f\|_{L^2_T H^2_X} + \|g\|_{L^2_T H^2_X} + \|f_t\|_{L^2_T L^2_X} + \|g_t\|_{L^2_T L^2_X} \big) \end{aligned}$$

Proof. For $\alpha = (0)$, the (deterministic) compatibility conditions hold, and from Theorem 3.4(ii),

$$\begin{aligned} &\|u_{(0)}\|_{L^2_T H^4_X} + \|u_{(0),t}\|_{L^2_T H^2_X} + \|u_{(0),tt}\|_{L^2_T L^2_X} \\ &\le C_1^{reg} \left(\|v\|_{H^3_X} + \|f\|_{L^2_T H^2_X} + \|f_t\|_{L^2_T L^2_X} \right). \end{aligned}$$

For $\alpha = \varepsilon_k$, since we have assumed the coefficients of $\mathcal{M}_k$ to be sufficiently smooth (e.g., at least $W^{3,\infty}_X$), so $u_{(0)} \in L^2_T H^4_X$ implies that $\mathcal{M}_k u_{(0)} + g_k \in L^2_T H^2_X$, and $u_{(0),t} \in L^2_T H^2_X$ implies that $(\mathcal{M}_k u_{(0)} + g_k)_t \in L^2_T L^2_X$. The compatibility conditions for $(\mathcal{M}_k u_{(0)} + g_k)\big|_{t=0} = \mathcal{M}_k v + g_k(0)$ are also satisfied. Again applying Theorem 3.4(ii),

$$\begin{aligned} &\|u_{\varepsilon_k}\|_{L^2_T H^4_X} + \|(u_{\varepsilon_k})_t\|_{L^2_T H^2_X} + \|(u_{\varepsilon_k})_{tt}\|_{L^2_T L^2_X} \\ &\le C_1^{reg} \left(\lambda_k^{(4)} \|u_{(0)}\|_{L^2_T H^4_X} + \theta_k^{(2)} \|g\|_{L^2_T H^2_X} + \lambda_k^{(2)} \|(u_{(0)})_t\|_{L^2_T H^2_X} + \theta_k^{(0)} \|g_t\|_{L^2_T L^2_X} \right) \\ &\le (C_1^{reg})^2 (\lambda_k^{(4)} \vee \lambda_k^{(2)}) \tilde{\tilde{M}} \left(\|v\|_{H^3_X} + \|f\|_{L^2_T H^2_X} + \|f_t\|_{L^2_T L^2_X} + \|g\|_{L^2_T H^2_X} + \|g_t\|_{L^2_T L^2_X} \right) \end{aligned}$$

where $\tilde{\tilde{M}} = \sup_k \big\{ 1 \vee \frac{(\mu_k^{(2)} \vee \mu_k^{(0)})}{(\lambda_k^{(4)} \vee \lambda_k^{(2)}) C_1^{reg}} \big\}$. (The remark following Theorem 3.3 applies.)

For $|\alpha| \geq 2$, we have $\mathcal{M}_k u_{\alpha-\varepsilon_k} \in L^2_T H^2_X$ and $(\mathcal{M}_k u_{\alpha-\varepsilon_k})_t \in L^2_T L^2_X$. The compatibility conditions hold trivially, since $u_{\alpha-\varepsilon_k}\big|_{t=0} \equiv 0$ whenever $|\alpha| \geq 2$. The usual computations give the estimates,

$$
\begin{aligned}
&\|u_\alpha\|_{L^2_T H^4_X} + \|u_{\alpha,t}\|_{L^2_T H^2_X} + \|u_{\varepsilon_k,tt}\|_{L^2_T L^2_X} \\
&\leq C^{reg}_1 \tilde{\tilde{M}} \frac{\left(C^{reg}_1(\lambda^{\vec{(4)}} \vee \lambda^{\vec{(2)}})\right)^\alpha |\alpha|!}{\sqrt{\alpha!}} \\
&\quad \times \left(\|v\|_{H^3_X} + \|f\|_{L^2_T H^2_X} + \|f_t\|_{L^2_T L^2_X} + \|g\|_{L^2_T H^2_X} + \|g_t\|_{L^2_T L^2_X} \right).
\end{aligned}
$$

The weighted norm $\|u\|_{\mathcal{R}'_\Omega L^2_T H^4_X} < \infty$ provided (2.17) holds. □

Due to the lower triangular property of the propagator system, the first order compatibility conditions for the stochastic parabolic equation call for additional conditions on the input data compared to the deterministic case. If the input data is smoother than what is assumed in Corollary 3.6, additional compatibility conditions are required on the derivatives $\{D^\gamma v, D^\gamma f, D^\gamma g\}$ in order to further increase the spatial regularity of u, u_t and u_{tt}, even if the boundedness of time derivatives beyond u_{tt} are not needed. Additionally, if the input data is random, similar arguments as Steps 2 and 3 in Theorem 3.3 extend Corollary 3.6 to the random input data case, this time with additional compatibility conditions on the modes $\{v_\alpha, f_\alpha, g_\alpha\}$. These results are summarized in the following theorem.

THEOREM 3.7. *Suppose $u \in \mathcal{R}_\Omega L^2_T H^1_{0X}$ is the weak solution of the SPDE* (2.7). *For fixed $m \geq 2$, also assume that*

$$
v \in \bar{\mathcal{R}}_\Omega H^{m+1}_X, \quad \textit{and} \quad f, g \in \bar{\mathcal{R}}_\Omega L^2_T H^m_X, \quad \textit{and} \quad \frac{df}{dt}, \frac{dg}{dt} \in \bar{\mathcal{R}}_\Omega L^2_T H^{m-2}_X,
$$

and that the compatibility conditions (2.16) *hold for $\{D^\gamma v_\alpha, D^\gamma f_\alpha, D^\gamma g_{k,\alpha}\}$, for all $\alpha \in \mathcal{J}$, and all indices $\gamma = (\gamma_1, \ldots, \gamma_d)$ with $|\gamma| \leq m-2$.*

Then for the weights $\mathcal{R}'$ satisfying

$$
\sum_k q'_k \Big(\big(\lambda^{(m+2)}_k \vee \lambda^{(m)}_k\big) C^{reg}_m \Big)^2 < 1 \quad \textit{and} \quad \sum_k \frac{q'_k}{\rho_k} < 1, \tag{2.18}
$$

we have for the weak solution

$$
u \in \mathcal{R}'_\Omega L^2_T H^{m+2}_X, \quad u_t \in \mathcal{R}'_\Omega L^2_T H^m_X, \quad u_{tt} \in \mathcal{R}'_\Omega L^2_T H^{m-2}_X, \tag{2.19}
$$

and

$$\begin{aligned}&\|u\|_{\mathcal{R}'_\Omega L^2_T H^{m+2}_X} + \|u_t\|_{\mathcal{R}'_\Omega L^2_T H^m_X} + \|u_{tt}\|_{\mathcal{R}'_\Omega L^2_T H^{m-2}_X} \\ &\le C\big(\|v\|_{\bar{\mathcal{R}}_\Omega H^{m+1}_X} + \|f\|_{\bar{\mathcal{R}}_\Omega L^2_T H^m_X} + \|g\|_{\bar{\mathcal{R}}_\Omega L^2_T H^m_X} + \|f_t\|_{\bar{\mathcal{R}}_\Omega L^2_T H^{m-2}_X} + \|g_t\|_{\bar{\mathcal{R}}_\Omega L^2_T H^{m-2}_X}\big).\end{aligned}$$

CHAPTER 3

Error Analysis for the Stochastic Finite Element Method

In this chapter, we study numerical solutions obtained with the stochastic finite element method applied to a parabolic SPDE driven by a multiplicative abstract noise $\dot{W}$,

$$
\begin{aligned}
&\tfrac{du}{dt} + \mathcal{A}u + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}u) = f \quad \text{on } D \times (0,T] \\
&u|_{\partial D} = 0, \\
&u|_{t=0} = v
\end{aligned}
\tag{3.1}
$$

and derive *a priori* error estimates for the numerical solution. Here, $\mathcal{A}$ is a uniformly elliptic operator and $\mathcal{A}, \mathcal{M}$ take the form

$$
\begin{aligned}
\mathcal{A}u &= -\textstyle\sum_{i,j} D_i(a^{ij}(x,t)D_j u) \\
\mathcal{M}_k u &= \textstyle\sum_{i,j} D_i(\sigma_k^{ij}(x,t)D_j u)
\end{aligned}
\tag{3.2}
$$

with a^{ij}, σ_k^{ij} measurable and uniformly bounded on $\bar{D}$. So, equation 3.1 is a special case of equation 2.7.

The stochastic finite element method combines discretization procedures from the classical finite element theory in numerical analysis with stochastic analysis in order to obtain computable solutions of SPDEs. The variable of randomness is discretized by a Galerkin approximation of the Wiener chaos expansion. This reduces the propagator system to a finite system of deterministic PDE that is then solved using the finite element discretization. Additionally, thanks to the lower triangular property of the propagator system, the stochastic finite element method becomes an iterative procedure of applying the finite element method to each equation in the propagator system recursively. This formulation of the stochastic finite element method for the corresponding stochastic elliptic equation has been described in [**65**], while the formulation for the parabolic case is essentially the same [**44**].

An important question to address upon formulating the numerical algorithm is to quantify, a priori, the error of the numerical solution. As a consequence of the discretization procedures, the numerical error estimates for the elliptic and parabolic problems are comprised of two terms. One term represents the error from the stochastic discretization, while the other term represents the numerical error from the application of the deterministic finite element method to each equation in the truncated propagator system. A feature of the error estimates that carries over from the deterministic theory to the stochastic case is the connection between the spatial regularity of the solution and the order of convergence of the finite element schemes—a smoother solution yields a higher order of convergence for the part of the numerical error coming from the finite element discretization. Moreover, the Malliavin calculus approach turns out to be an indispensable framework for the error analysis of the parabolic equations, because it avails us of the two *stochastic adjoint operators*, the Malliavin derivative and the Malliavin divergence operators, satisfying the adjoint property (2.5). This provides a tool to investigate the stochastic finite element method in a completely analogous way to the deterministic theory. The main idea brought over from the deterministic theory is the definition of the so-called Ritz projection that comes from the finite element method applied to the corresponding elliptic problem; additionally, where the notion of invoking an *adjoint problem* is required, the Malliavin calculus provides exactly this tool of a *stochastic adjoint problem*. In this sense, one may construe this error analysis to be a direct generalization of the deterministic theory to the stochastic case.

However, it should be noted that extensive research on variants of finite element methods, such as *hp*-element methods, has produced highly efficient deterministic solvers. As such, it is frequently the case that the errors incurred by the stochastic finite element method are largely dominated by the error due to the stochastic discretization, rather than the spatial discretization. Nevertheless, it is our hope that the techniques described in this chapter will elucidate a way of using the Malliavin calculus as a framework for direct generalization of the numerical analysis.

Before proceeding, we remark on the wealth of techniques in the literature that has been developed both for the stochastic analysis and numerical analysis of SPDEs. Though the basic conception of the discretization procedure is based on the basic protocols already familiar in the algorithms for deterministic PDE—finite differences, Galerkin approximations, collocation methods, finite element methods, etc.—these methods differ essentially depending on the way stochasticity is modelled in the equations. SPDEs that are essentially infinite dimensional Itô equations (for example, equations driven by cylindrical Brownian motion) are often treated by transforming the SPDE into an infinite system of SDE and applying the techniques of the usual Itô calculus. Numerical simulation of this type of SPDEs often involves discretizing the SDE system by finite differences in the time component of the Brownian motion increments [**13**, **25**, **26**, **33**]. Stochastic Taylor expansions are used in [**32**, **36**] for developing and analyzing high order methods. In problems of Uncertainty Quantification, equations that depend on finite dimensional noise (i.e. perturbation by finite number of random variables) have enjoyed the development of polynomial chaos, generalized polynomial chaos and stochastic collocation methods over the the past decade [**22**, **23**, **67**–**69**]. These methods make use of the Karhunen–Loève expansion, an orthogonal stochastic expansion akin to the Wiener chaos expansion, likewise treating the stochasticity in the equation as independent variables.

Within the realm of finite element methods for stochastic PDE, there has been much literature on the algorithms and analysis for both elliptic and parabolic SPDE. We describe a few studies that bear some connection to our present analysis. Convergence rates of the Wiener-Itô expansions of white noise and the errors from the Galerkin approximations, sans spatial discretization, have been studied in [**7**, **10**, **63**]. In [**38**, **39**], the finite element discretization for semilinear parabolic SPDE and linear stochastic wave equation, both with additive noise in the framework of Ito calculus, was studied. For general elliptic SPDEs, the stochastic finite element or stochastic collocation methods have been studied by [**1**–**3**, **21**]. The white noise functional approach using primarily the Wick product model has been studied by [**8**, **29**, **30**, **52**], appealing to a similar technique of transforming the SPDE into a deterministic system of PDEs. An analysis in which the existing deterministic finite element theory is extended to the stochastic setting has been studied by [**61**, **70**]. This idea of extending the deterministic theory turns out to be similar in spirit to our present work.

1. Review of the error estimates for the finite element method for deterministic PDE

We first describe the usual finite element set up for solving deterministic PDEs, and briefly review how the error estimates for elliptic and parabolic PDE are derived. The finite element set up will be used directly as the protocol for spatial discretization in the stochastic finite element method, but beyond that, we will elucidate the principles governing how the deterministic theory is developed, that will become the conception for the subsequent error analysis.

1.1. The finite element approximation. Let D be a domain in $\mathbb{R}^d$ with smooth boundary and let $\mathcal{T}_h$ be a family of quasi-uniform triangulations on D. Let $(K_{ref}, \mathcal{P}, \mathcal{N})$ be a reference finite element. For $K \in \mathcal{T}_h$, let $S_h^K = \{z \,:\, z \circ F_K^{-1} \in \mathcal{P}(K_{ref})\}$ where $F_K : K_{ref} \to K$ is affine. The finite element space is

$$S_h = \{z \in H_0^1(D) \,:\, z|_K \in S_h^K, K \in \mathcal{T}_h\}$$

A property of S_h we assume is that there exists $r \geq 2$ such that for h small,

$$\inf_{z_h \in S_h} \left\{ \|v - z_h\|_{L_2} + h\|\nabla(v - z_h)\|_{L_2} \right\} \leq Ch^s \|v\|_{H^s}, \quad \text{for } 1 \leq s \leq r \tag{3.3}$$

whenever $v \in H^s \cap H_0^1$ [**62**]. We also assume that, in particular, S_h consists of piecewise polynomials of degree at most $r-1$, so that the inverse inequality holds,

$$\|\nabla z_h\|_{L^2} \leq Ch^{-1}\|z_h\|_{L^2}, \quad \forall z_h \in S_h.$$

We denote the finite element basis of S_h by $\{\Phi_l\}_{l=1,\ldots,\dim S_h}$.

1.2. The finite element method for deterministic PDE. For illustration's sake, we consider a simple parabolic equation, the heat equation on D

$$\begin{aligned} &u_t - \Delta u = f, \quad \text{on } D \\ &u|_{\partial D} = 0 \\ &u(0, \cdot) = w \end{aligned}$$

and the corresponding elliptic problem

$$\begin{aligned} &-\Delta U = F, \quad \text{on } D \\ &U|_{\partial D} = 0 \end{aligned}$$

We will give an overview of the derivation of the error estimates for the parabolic equation, à la Thomée, which utilizes error estimates for the corresponding elliptic problem as well as utilizes properties of the adjoint problem (which in this case coincides with the elliptic problem, since Δ is self-adjoint.)

The finite element formulation for the elliptic problem is to find $U_h \in S_h$ such that

$$(\nabla U_h, \nabla \chi) = (F, \chi), \quad \forall \chi \in S_h$$

Then the following error estimate obtains, the proof of which is well documented in the literature and is not needed for our purposes.

Theorem 1.1. *Assume the solution U of the elliptic problem belongs to H^s for some $1 \leq s \leq r$. Then*

$$|U_h - U| \leq Ch^s \|U\|_s \quad \textit{and} \quad |\nabla U_h - \nabla U| \leq Ch^{s-1} \|U\|_s.$$

Similarly, the finite element formulation for the parabolic equation is to find $u_h(t) \in S_h$, $t \geq 0$, such that

$$(u_{h,t}, \chi) + (\nabla u_h, \nabla \chi) = (f, \chi), \quad \forall \chi \in S_h, t > 0,$$

with $u_h(0) = w_h$, where $w_h \in S_h$ is some approximation of w. This is a semi-discrete formulation where the time variable has not been discretized. We have the error estimate as follows.

Theorem 1.2. *Assume the initial condition $w \in H^r$, and for simplicity take $w_h = R_h w$. For the solution u of the parabolic problem, assume that $u, u_t \in H^r$. Then*

$$|u_h(t) - u(t)| \leq Ch^r \left(\|u\|_r + \left(\int_0^t \|u_t\|_r^2 dt' \right)^{1/2} \right), \quad \forall t \geq 0$$

We will highlight the key ideas of the proof to illustrate how the elliptic error estimates are being used to show the parabolic error estimates. We define the *elliptic* or *Ritz projection* $R_h : H_0^1 \to S_h$ by

$$(\nabla R_h v, \nabla \chi) = (\nabla v, \nabla \chi), \quad \forall \chi \in S_h$$

In order words, R_h is the finite element approximation operator for the corresponding elliptic problem, which maps an exact solution v of the elliptic problem to the finite element approximation $v_h = R_h v$. Then, the approximation error can be decomposed into the sum of two terms,

$$u_h(t) - u(t) = \big(u_h(t) - R_h u(t)\big) + \big(R_h u(t) - u(t)\big) = \theta(t) + \pi(t)$$

In an obvious way, $\pi(t)$ can be directly estimated using the elliptic error estimates, and in a less direct way, so too can $\theta(t)$ be estimated. Using the definitions of the weak and FE formulations, we can compute

$$(\theta_t, \chi) + (\nabla\theta, \nabla\chi) = -(\pi_t, \chi), \quad \forall\chi \in S_h,\, t > 0,$$

and choosing $\chi = \theta$

$$\frac{1}{2}\frac{d}{dx}|\theta|^2 + |\nabla\theta|^2 \leq |\pi_t||\theta| \leq C|\pi_t|^2 + C'|\theta|^2$$

and the error estimates follow from applying Gronwall's inequality. □

The last equation in the sketch of the proof uses the L_2 duality estimates, $|(\pi_t, \theta)| \leq |\pi_t||\theta|$, which is a sensible choice since elliptic error estimates provide knowledge of $|\pi_t|$, and which leads to an order of convergence h^r matching the norms of both $\|u_t\|_r$ and $\|u_t\|_r$. However, this manner of estimates does not exploit the structure of the regularity properties of $u, u_t, \dots$ that solutions of parabolic problems possess. An alternative estimate is to use instead the duality pairing between H^{-1} and H^1, $|(\pi_t, \theta)| \leq \|\pi_t\|_{-1}\|\theta\|_1$. This requires a different set of estimates in the negative norm, but also turns out to yield a higher order of convergence.

1.3. Using the adjoint problem for negative norm estimates. The adjoint problem of the elliptic problem, which in the simple model problem happens to coincide with the elliptic problem itself, is used in a duality argument to yield error estimates in negative order norms. The feature of these estimates is the give-and-take between spatial regularity and order of convergence—one can estimate the error in a lower order Sobolev space in exchange for a higher order of convergence—though such trade-off is quite typical in finite element theory. Subsequently, we will use this to improve the order of convergence of the error estimates for the parabolic problem.

For a nonnegative integer q, we define the spaces $H^{-q}(D)$ to be the dual of $H^q(D)$ with respect to the inner product in $L^2(D)$, with duality pairing $\langle \cdot, \cdot \rangle$. The norm is

$$\|v\|_{-q} = \sup_{\phi \in H^q} \frac{\langle v, \phi \rangle}{\|\phi\|_q}$$

We have the following analogue of the error estimates for the elliptic problem.

Theorem 1.3. *Let $U \in H^s$ for some $1 \leq s \leq r$. Then*

$$\|U_h - U\|_{-q} \leq Ch^{q+s}\|U\|_s, \quad \textit{for } 0 \leq q \leq r-2.$$

To illustrate the duality argument, we give the highlights of the proof. The negative norm in the sense of the sup norm is to be estimated; to this end, for any $\phi \in H^q$, consider

$$\langle U_h - U, \phi \rangle = (U_h - U, -\Delta\psi) = (\nabla(U_h - U), \nabla\psi)$$

The existence of ψ is granted by the solution of the adjoint problem $-\Delta\psi = \phi$ with $\psi|_{\partial D} = 0$, and moreover has the property that $\|\psi\|_{q+2} \leq C\|\phi\|_q$ for any $q \geq 0$. Consequently, by orthogonality of the error to S_h, the approximation property in S_h, and the elliptic error estimates,

$$\begin{aligned} |\langle U_h - U, \phi \rangle| &= |(\nabla(U_h - U), \nabla(\psi - \chi))| \qquad \text{for any } \chi \in S_h \\ &\leq C\|U_h - U\|_1 \inf_{\chi \in S_h} \|\psi - \chi\|_1 \\ &\leq Ch^{s-1}\|U\|_s \cdot h^{q+1}\|\psi\|_{q+2} \leq Ch^{q+s}\|U\|_s\|\phi\|_q \end{aligned}$$

The result follows. □

As noted above, the application of the negative norm estimates is to raise the order of convergence for the parabolic problem.

Theorem 1.4. *Let $r \geq 3$. Assume $w \in H^r$ and for simplicity, take $w_h = R_h w$. Also assume the* compatibility conditions *that yield $u(t) \in H^{r+1}$, $u_t(t) \in H^{r-1}$ for a.e. $t \geq 0$. Then*

$$|u_h(t) - u(t)| \leq Ch^r \left(\|u(t)\|_r + \left(\int_0^t \|u_t\|_{r-1}^2 dt' \right)^{1/2} \right)$$

To prove the theorem, we decompose the error into two terms similar to the previous proof of the parabolic estimates. The difference comes in estimating the term $\theta(t)$:

$$\frac{1}{2}\frac{d}{dt}|\theta|^2 + (\nabla\theta, \nabla\theta) \leq \|\pi_t\|_{-1}\|\theta\|_1 \leq C\|\pi_t\|_{-1}^2 + |\nabla\theta|^2$$

Integrating yields the desired estimate for θ. Together with the estimate for $|\pi(t)| \leq Ch^r\|u\|_r$, the result follows. □

2. The stochastic finite element method formulation

We will formulate the stochastic finite element method for the linear parabolic SPDE (3.1). The stochastic finite element method adopts the same idea as in the deterministic case, by elucidating a finite dimensional *stochastic finite element space* and casting the weak formulation of the problem into that finite dimensional setting. As with many numerical schemes for SPDE, the stochastic finite element method considered here forms the stochastic finite element space as a tensor product space of the spatial and stochastic variables, to which well-developed discretization techniques for each variable are applied separately: a finite element approximation in the spatial variable and the Galerkin approximation in the stochastic variable. In our analysis, we consider only the semi-discrete case, in that the time variable is kept continuous, thus yielding a system of ODE. The fully discrete scheme can be created by applying a suitable time stepping algorithm to the system of ODE.

Finite element approximation in space. We use the usual finite element set up described in Section 1.1; that is, S_h is a finite element space on a family $\mathcal{T}_h$ of quasi-uniform triangulations, with the assumption that S_h is spanned by the FE basis $\{\Phi_l\}_{l=1,\ldots,\dim S_h}$ consisting of piecewise polynomials of degree at most $r-1$.

Galerkin approximation in randomness. Letting

$$\mathcal{J}_{M,n} := \{\gamma \in \mathcal{J} \,:\, |\gamma| \leq n,\, \dim(\gamma) \leq M\},$$

we define the truncated Wiener chaos space

$$S^{M,n} = \Big\{ f = \sum_{\gamma \in \mathcal{J}_{M,n}} f_\gamma \xi_\gamma \,:\, f_\gamma \in \mathbb{R} \Big\}.$$

M represents the truncation of the white noise to a finite dimension, while n represents the highest polynomial degree of the Hermite polynomials that make up the Cameron-Martin basis.

SFEM formulation. The stochastic finite element method for the parabolic problem is

Find $u_h^{M,n} \in S_h \otimes S^{M,n}$ such that

$$\begin{aligned}&\langle\!\langle \tfrac{du_h^{M,n}}{dt}, z_h \rangle\!\rangle_{\mathcal{R}_\Omega^{\mp 1} L_X^2} + \langle\!\langle \mathcal{A} u_h^{M,n} + \textstyle\sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k u_h^{M,n}), z_h \rangle\!\rangle_{\mathcal{R}_\Omega^{\mp 1} H_X^{\mp 1}} \\ &= \langle\!\langle f, z_h \rangle\!\rangle_{\mathcal{R}_\Omega^{\mp 1} H_X^{\mp 1}}\end{aligned} \tag{3.4}$$

for all $z_h \in S^{M,n} \otimes S_h$, and for every $t \in [0, T]$.

Denote the numerical solution

$$u_h^{M,n}(x,t) = \sum_{\gamma \in \mathcal{J}_{M,n}} \hat{u}_\gamma(x,t)\xi_\gamma = \sum_{\gamma \in J_{M,n}} \sum_{l=1}^{\dim S_h} \hat{u}_{\gamma,l}(t)\Phi_l(x)\xi_\gamma$$

Due to 3.2, solving (3.1) via the SFEM is equivalent to solving each equation in the truncated propagator system via FEM: for $\alpha \in \mathcal{J}_{M,n}$,

$$\Big(\frac{d\hat{u}_{(0)}}{dt}, z_h\Big) + \boldsymbol{A}[\hat{u}_{(0)}, z_h] = \langle f_{(0)}, z_h \rangle, \tag{3.5}$$

$$\big(\frac{d\hat{u}_\alpha}{dt}, z_h\big) + \boldsymbol{A}[\hat{u}_\alpha, z_h] + \sum_{k=1}^{M} \sqrt{\alpha_k}\big(\boldsymbol{M}_k[\hat{u}_{\alpha-\varepsilon_k}, z_h]\big) = \langle f_\alpha, z_h \rangle, \tag{3.6}$$

for all $z_h \in S_h$, with initial conditions $\hat{u}_\alpha|_{t=0} = (v_h^{M,n})_\alpha$. The bilinear forms $\boldsymbol{A}, \boldsymbol{M}_k$ are the bilinear forms associated with $\mathcal{A}, \mathcal{M}_k$. Note that by our assumptions, $\boldsymbol{A}$ is coercive and $\boldsymbol{M}_k$ is bounded.

The algorithm. Next, we write out the SFEM algorithm explicitly to show the resulting system of ODE. We define the mass and stiffness matrices identically to the usual FEM case, and also a noise matrix arising from the stochastic term:

$$\mathbb{M}^{mass}_{l'l} = (\Phi_l, \Phi_{l'}), \qquad \mathbb{M}^{stiff}_{l'l} = \boldsymbol{A}[\Phi_l, \Phi_{l'}], \qquad \mathbb{M}^{noise}_{k;l'l} = \boldsymbol{M}_k[\Phi_l, \Phi_{l'}].$$

The lower triangular discrete propagator system is solved iteratively. Let the vector of coefficients of the solution vector be $\vec{\hat{u}}_\gamma = (\hat{u}_{\gamma,1}, \ldots, \hat{u}_{\gamma,\dim S_h})^T$. Then, for $\gamma = (0)$,

$$\mathbb{M}^{mass}(\vec{\hat{u}}_{(0)})_t + \mathbb{M}^{stiff}\vec{\hat{u}}_{(0)} = \vec{f}_{(0)}$$

and for $|\gamma| \geq 1$,

$$\mathbb{M}^{mass}(\vec{\hat{u}}_\gamma)_t + \mathbb{M}^{stiff}\vec{\hat{u}}_\gamma + \sum_k \sqrt{\gamma_k}\left(\mathbb{M}^{noise}\vec{\hat{u}}_{\gamma-\varepsilon_k} + \vec{g}_{k,\gamma-\varepsilon_k}\right) = \vec{f}_\gamma$$

where

$$\vec{f}_\gamma = (\langle f_\gamma, \Phi_1\rangle, \ldots, \langle f_\gamma, \Phi_{\dim S_h}\rangle)^T, \quad \text{and}$$
$$\vec{g}_{k,\gamma} = (\langle g_{k,\gamma}, \Phi_1\rangle, \ldots, \langle g_{k,\gamma}, \Phi_{\dim S_h}\rangle)^T.$$

Remark. We remark that the stochastic finite element formulation for the parabolic problem is identical to the formulation for the corresponding elliptic problem (3.8). For the elliptic problem,

Find $U_h^{M,n} \in S_h \otimes S^{M,n}$ such that

$$\langle\!\langle \mathcal{A}U_h^{M,n} + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k U_h^{M,n}), z_h \rangle\!\rangle_{\mathcal{R}_\Omega^{\mp1} H_X^{\mp1}} = \langle\!\langle f, z_h \rangle\!\rangle_{\mathcal{R}_\Omega^{\mp1} H_X^{\mp1}} \tag{3.7}$$

for all $z_h \in S^{M,n} \otimes S_h$, and for every $t \in [0, T]$.

In this case, the implementation of the algorithm involves defining the stiffness and noise matrices, but not the mass matrix.

3. Error analysis for SPDE with time independent operators

We first study (3.1) under the assumption that $\mathcal{A}, \mathcal{M}_k$ do not depend on time. In particular, the white noise $\dot{W}(x)$ is restricted to become a purely spatial noise. The main goal of this section is to show the main error estimates for the parabolic problem (3.1) (see Theorem 3.9). This will be achieved by an analogous analysis as the deterministic theory, of going through the elliptic error estimates and adjoint problem. In view of this, we will begin by studying the formal stochastic adjoint problem as well as the negative norm error estimates for the elliptic SPDE, and finally stating and deriving the parabolic estimates.

3.1. The corresponding elliptic SPDE and the formal stochastic adjoint problem. The corresponding stochastic elliptic problem is

$$\begin{aligned} &\mathcal{A}U + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}U) = F \quad \text{in } D \\ &U|_{\partial D} = 0 \end{aligned} \tag{3.8}$$

where $F \in \bar{\mathcal{R}}_\Omega H_X^{-1}$. For non-random F, [**49**] has shown the unique existence of the weak solution U in some $\mathcal{R}_\Omega H_{0X}^1$. For arbitrary random F, an argument by induction yields the following result.

Theorem 3.1. *Let $F \in \bar{\mathcal{R}}_\Omega H_X^{-1}$. Then there exists a unique weak solution U of* (3.8) *belonging to $\mathcal{R}_\Omega H_{0X}^1$, provided the weights $r_\alpha^2 = \frac{q^\alpha}{|\alpha|!}$ satisfy*

$$\sum_k q_k (\lambda_k C_A^{ellip})^2 < 1, \quad \textit{and} \quad \sum_k \frac{q_k}{\bar{\rho}_k} < 1, \tag{3.9}$$

Moreover, we have the bounds

$$\|U_\alpha\|_{H_{0X}^1} \le C_A^{ellip} \sqrt{|\alpha|!} \sum_{\beta \le \alpha} \|F_{\alpha-\beta}\|_{H_X^{-1}} \prod_{k=1}^{\infty} (\lambda_k C_A^{ellip})^{\beta_k} \sqrt{\frac{|\beta|!}{\beta!(\alpha-\beta)!}}. \tag{3.10}$$

We first state a result on the boundedness of the stochastic operator in the LHS of equation (3.8) that will come in handy subsequently.

Lemma 3.2. *Let $\chi \in \mathcal{R}_\Omega H_X^r \cap \mathcal{R}_\Omega H_{0X}^1$, $r \ge 1$, where the weights satisfy $\sum_k q_k (\lambda_k^r)^2 < \infty$. Then there exists C depending only on $\mathcal{R}, \mathcal{A}, \mathcal{M}$ such that*

$$\|\mathcal{A}\chi + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}\chi)\|_{\mathcal{R}_\Omega H_X^{r-2}} \le C \|\chi\|_{\mathcal{R}_\Omega H_X^r}.$$

Proof. We show the lemma for $r = 1$, for ease of notation; the proof for $r > 1$ is identical. By direct computation,

$$\begin{aligned}
\|\mathcal{A}\chi + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}\chi)\|^2_{\mathcal{R}_\Omega H_X^{-1}} &= \sum_\alpha r_\alpha^2 \|\mathcal{A}\chi_\alpha + \sum_{k=1}^\infty \sqrt{\alpha_k} \mathcal{M}_k \chi_{\alpha-\varepsilon_k}\|^2_{H_X^{-1}} \\
&\le \sum_\alpha r_\alpha^2 \left(C_A^b \|\chi_\alpha\|_{H_0^1} + \sum_{k=1}^\infty \sqrt{\alpha_k} \lambda_k \|\chi_{\alpha-\varepsilon_k}\|_{H_0^1} \right)^2 \\
&\le 2 (C_A^b)^2 \|\chi\|^2_{\mathcal{R}_\Omega H_{0X}^1} + 2 \underbrace{\sum_\alpha r_\alpha^2 \left(\sum_{k=1}^\infty \sqrt{\alpha_k} \lambda_k \|\chi_{\alpha-\varepsilon_k}\|_{H_0^1} \right)^2}_{(*)}
\end{aligned}$$

where C_A^b is the constant in $\|\mathcal{A}\phi\|_{H_X^{-1}} \leq C_A^b \|\phi\|_{H^1}$, for all $\phi \in H_{0X}^1$. To estimate $(*)$, we apply Jensen's inequality to obtain

$$\begin{aligned}
(*) &= \sum_\alpha r_\alpha^2 \left(\sum_{\substack{k=1 \\ \alpha_k \neq 0}}^\infty \frac{\alpha_k}{|\alpha|} \frac{|\alpha|}{\sqrt{\alpha_k}} \lambda_k \|\chi_{\alpha-\varepsilon_k}\|_{H_0^1} \right)^2 \\
&\leq \sum_\alpha \frac{q^\alpha}{|\alpha|!} \sum_{\substack{k=1 \\ \alpha_k \neq 0}}^\infty \frac{\alpha_k}{|\alpha|} \frac{|\alpha|^2}{\alpha_k} \lambda_k^2 \|\chi_{\alpha-\varepsilon_k}\|_{H_0^1}^2 \\
&= \sum_\alpha \sum_k \mathbf{1}_{\{\alpha_k \neq 0\}} q_k \lambda_k^2 \frac{q^{\alpha-\varepsilon_k}}{(|\alpha|-1)!} \|\chi_{\alpha-\varepsilon_k}\|_{H_0^1}^2 \\
&= \sum_k q_k \lambda_k^2 \sum_{\substack{\alpha \\ \alpha_k \neq 0}} r_{\alpha-\varepsilon_k} \|\chi_{\alpha-\varepsilon_k}\|_{H_0^1}^2 = \left(\sum_k q_k \lambda_k^2 \right) \|\chi\|_{\mathcal{R}_\Omega H_{0X}^1}^2
\end{aligned}$$

Hence,

$$\|\mathcal{A}\chi + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}\chi)\|_{\mathcal{R}_\Omega H_X^{-1}}^2 \leq 2 \left((C_A^b)^2 + \sum_k q_k \lambda_k^2 \right) \|\chi\|_{\mathcal{R}_\Omega H_{0X}^1}^2.$$

□

Let the operators $\mathcal{A}^*, \mathcal{M}_k^*$ be the formal adjoints of $\mathcal{A}, \mathcal{M}_k$, respectively. The formal *stochastic adjoint problem* of (3.8) is

$$\begin{aligned}
\mathcal{A}^*\psi + \mathcal{M}^* \cdot \boldsymbol{D}_{\dot{W}}\psi &= \phi \qquad \text{on } D \\
\psi|_{\partial D} &= 0
\end{aligned} \tag{3.11}$$

for $\phi \in \mathcal{R}_\Omega^{-1} H_X^{-1}$. Although for our error estimates, we consider only self-adjoint operators $\mathcal{A}, \mathcal{M}_k$ of the form (3.2), the results in this section apply to nonself-adjoint operators as well.

By definition, the term $\mathcal{M}^* \cdot \boldsymbol{D}_{\dot{W}}\psi$ can be formally written as

$$\left(\mathcal{M}^* \cdot \boldsymbol{D}_{\dot{W}}\psi\right)_\alpha = \sum_{k=1}^\infty \sqrt{\alpha_k + 1} \mathcal{M}_k^* \psi_{\alpha+\varepsilon_k}, \quad \text{for } \alpha \in \mathcal{J}$$

where the infinite sum is interpreted as convergent in an appropriate space. Due to the adjoint property (2.5) between $\boldsymbol{D}_{\dot{W}}$ and $\boldsymbol{\delta}_{\dot{W}}$, we have the adjoint property between the operators

$$\langle\!\langle \chi, \mathcal{A}^*\psi + \mathcal{M}^* \cdot \boldsymbol{D}_{\dot{W}}\psi \rangle\!\rangle_{\mathcal{R}_\Omega H_{0X}^1, \mathcal{R}_\Omega^{-1} H_X^{-1}} = \langle\!\langle \mathcal{A}\chi + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}\chi), \psi \rangle\!\rangle_{\mathcal{R}_\Omega H_X^{-1}, \mathcal{R}_\Omega^{-1} H_{0X}^1}.$$

Definition 3.3. *A weak solution of (3.11), with $\phi \in \mathcal{R}_\Omega^{-1} H_X^{-1}$, is a process $\psi \in \mathcal{R}_\Omega^{-1} H_{0X}^1$ such that*

$$\langle\!\langle \chi, \mathcal{A}^* \psi + \mathcal{M}^* \cdot \boldsymbol{D}_{\dot{W}} \psi \rangle\!\rangle_{\mathcal{R}_\Omega H_{0X}^1, \mathcal{R}_\Omega^{-1} H_X^{-1}} = \langle\!\langle \chi, \phi \rangle\!\rangle_{\mathcal{R}_\Omega H_{0X}^1, \mathcal{R}_\Omega^{-1} H_X^{-1}}$$

for all $\chi \in \mathcal{R}_\Omega H_{0X}^1$.

Note that $\|U^*\|_{H_0^1} \le C_A^{ellip} \|F\|_{H^{-1}}$ for the solution of $\mathcal{A}^* U^* = F$, and $\|\mathcal{M}_k^* \phi\|_{H^{-1}} \le \lambda_k \|\phi\|_{H_0^1}$.

Proposition 3.4. *Suppose there exists $\{\psi_\alpha, \, \alpha \in \mathcal{J}\}$ belonging to H_0^1 such that for all α,*

(i) $\displaystyle\sum_{k=1}^\infty \sqrt{\alpha_k + 1} \mathcal{M}_k^* \psi_{\alpha + \varepsilon_k} \in H_X^{-1}$*;*

(ii) $\displaystyle\mathcal{A}^* \psi_\alpha + \sum_{k=1}^\infty \sqrt{\alpha_k + 1} \mathcal{M}_k^* \psi_{\alpha + \varepsilon_k} = \phi_\alpha$ *in the weak sense.*

Let the weights $\mathcal{R}$ satisfy

$$\sum_k q_k (\lambda_k C_A^{ellip})^2 < \frac{1}{2}. \tag{3.12}$$

Then there exists C depending on $\mathcal{R}, \mathcal{A}^*, \mathcal{M}^*$, such that

$$\|\psi\|_{\mathcal{R}_\Omega^{-1} H_{0X}^1} \le C \|\phi\|_{\mathcal{R}_\Omega^{-1} H_X^{-1}}. \tag{3.13}$$

Proof. From the deterministic elliptic estimates,

$$\|\psi_\alpha\|_{H_{0X}^1} \le C_A^{ellip} \left(\|\phi_\alpha\|_{H^{-1}} + \sum_k \sqrt{\alpha_k + 1} \|\mathcal{M}_k^* \psi_{\alpha + \varepsilon_k}\|_{H^{-1}} \right)$$

So

$$\sum_\alpha r_\alpha^{-2} \|\psi_\alpha\|_{H_{0X}^1}^2 \le 2 (C_A^{ellip})^2 \sum_\alpha r_\alpha^{-2} \|\phi_\alpha\|_{H^{-1}}^2$$
$$+ 2 \sum_\alpha (C_A^{ellip})^2 \left(\sum_k r_\alpha^{-1} \sqrt{\alpha_k + 1} \lambda_k \|\psi_{\alpha + \varepsilon_k}\|_{H_{0X}^1} \right)^2$$

In the second term,

$$
\begin{aligned}
&(C_A^{ellip})^2 \left(\sum_k r_\alpha^{-1} \sqrt{\alpha_k + 1} \lambda_k \|\psi_{\alpha+\varepsilon_k}\|_{H^1_{0X}} \right)^2 \\
&= \left(\sum_k \frac{\sqrt{|\alpha|!}}{q^{\alpha/2}} \frac{\sqrt{|\alpha|+1}}{q_k^{1/2}} \sqrt{\frac{\alpha_k+1}{|\alpha|+1}} q_k^{1/2} \lambda_k C_A^{ellip} \|\psi_{\alpha+\varepsilon_k}\|_{H^1_0} \right)^2 \\
&\leq \left(\sum_k r^{-2}_{\alpha+\varepsilon_k} \|\psi_{\alpha+\varepsilon_k}\|^2_{H^1_0} \frac{\alpha_k+1}{|\alpha|+1} \right) \left(\sum_k q_k (\lambda_k C_A^{ellip})^2 \right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_\alpha \sum_k r^{-2}_{\alpha+\varepsilon_k} \|\psi_{\alpha+\varepsilon_k}\|^2_{H^1_0} \frac{\alpha_k+1}{|\alpha|+1} \\
&= \sum_k \sum_{\beta:\beta_k \neq 0} r_\beta^{-2} \|\psi_\beta\|^2_{H^1_0} \frac{\beta_k}{|\beta|} = \sum_\beta \sum_k \frac{\beta_k}{|\beta|} r_\beta^{-2} \|\psi_\beta\|^2_{H^1_0} = \|\psi\|^2_{\mathcal{R}_\Omega^{-1} H^1_{0X}}
\end{aligned}
$$

Hence,

$$
\left(1 - 2 \Big(\sum_k q_k (\lambda_k C_A^{ellip})^2 \Big) \right) \|\psi\|^2_{\mathcal{R}_\Omega^{-1} H^1_{0X}} \leq 2 (C_A^{ellip})^2 \|\phi\|^2_{\mathcal{R}_\Omega^{-1} H_X^{-1}}.
$$

The estimate follows from the condition (3.12). □

Theorem 3.5. *There exists a weak solution $\psi \in \mathcal{R}_\Omega^{-1} H^1_{0X}$ to the adjoint problem* (3.11) *satisfying* (3.13), *provided* (3.12) *holds.*

Proof. The weak solution is constructed via the usual Galerkin approach. Fix an integer p, and let $\phi^p := \sum_{|\alpha| \leq p} \phi_\alpha \xi_\alpha$. We will first construct the weak solution ψ^p of

$$
\mathcal{A}^* \psi^p + \mathcal{M}^* \cdot \boldsymbol{D}_{\dot{W}} \psi^p = \phi^p. \tag{3.14}
$$

Let $\psi^p_\alpha = 0$ if $|\alpha| > p$. For $|\alpha| = p$, define ψ^p_α by the solution of $\mathcal{A}^* \psi^p_\alpha = \phi_\alpha$. For $|\alpha| < p$,

$$
\mathcal{A}^* \psi^p_\alpha = \phi_\alpha - \sum_{k=1}^\infty \sqrt{\alpha_k + 1} \mathcal{M}_k^* \psi^p_{\alpha+\varepsilon_k}.
$$

The solvability of the equation for $|\alpha| = p$ follows from the usual deterministic theory, and

$$
\|\psi^p_\alpha\|_{H^1_0} \leq C_A \|\phi_\alpha\|_{H^{-1}}.
$$

The solvability of the equation for $|\alpha| < p$ requires that $\sum_k \sqrt{\alpha_k + 1} \mathcal{M}_k^* \psi^p_{\alpha+\varepsilon_k}$ belongs to H_X^{-1}, which we now verify.

Denote by $\Phi_\alpha^{(i)}$ the quantity

$$(\Phi_\alpha^{(i)})^2 = \sum_{k_1,\dots,k_i=1}^{\infty} r^{-2}_{\alpha+\varepsilon_{k_1}+\dots+\varepsilon_{k_i}} \|\phi_{\alpha+\varepsilon_{k_1}+\dots+\varepsilon_{k_i}}\|^2_{H^{-1}} \prod_{j=1}^{i} \frac{(\alpha+\varepsilon_{k_1}+\dots+\varepsilon_{k_{j-1}})_{k_j}+1}{|\alpha|+j}$$

Clearly, $\Phi_\alpha^{(i)} < \infty$. If $|\alpha| = p-l$, for $l = 1,\dots,p$, it is easy to show by induction on l that

$$\|\psi_\alpha^p\|_{H_0^1} \le C_A^{ellip} \left(\|\phi_\alpha\|_{H^{-1}} + r_\alpha^{-1}\sqrt{(l-1)!}\sum_{i=1}^{l} 2^{i/2}\hat{q}^{i/2}\Phi_\alpha^{(i)} \right)$$

where $\hat{q} = \sum_k q_k(\lambda_k C_A^{ellip})^2$, and hence

$$r_\alpha^{-2}\|\sum_k \sqrt{\alpha_k+1}\mathcal{M}_k^*\psi^p_{\alpha+\varepsilon_k}\|^2_{H^{-1}} \le (l-1)!\sum_{i=1}^{l} 2^i\hat{q}^i\Phi_\alpha^{(i)} < \infty.$$

This verifies that $\sum_k \sqrt{\alpha_k+1}\mathcal{M}_k^*\psi^p_{\alpha+\varepsilon_k} \in H^{-1}$, and hence $\psi^p := \sum_\alpha \psi_\alpha^p \xi_\alpha$ is well-defined.

By construction, ψ^p solves equation (3.14). Moreover, by similar calculations as Proposition 3.4,

$$\begin{aligned}\left(1 - 2\Big(\sum_k q_k(\lambda_k C_A^{ellip})^2\Big)\right)\|\psi^p\|^2_{\mathcal{R}_\Omega^{-1}H^1_{0X}} &\le 2(C_A^{ellip})^2\|\phi^p\|^2_{\mathcal{R}_\Omega^{-1}H_X^{-1}} \\ &\le 2(C_A^{ellip})^2\|\phi\|^2_{\mathcal{R}_\Omega^{-1}H_X^{-1}}\end{aligned}$$

and by (3.12), the sequence ψ^p is uniformly bounded in $\mathcal{R}_\Omega^{-1}H^1_{0X}$. Thus, there exists a weakly converging subsequence, say, with abuse of notation, $\psi^p \rightharpoonup \psi$ weakly in $\mathcal{R}_\Omega^{-1}H^1_{0X}$.

Fix an arbitrary $\chi \in \mathcal{R}_\Omega H^1_{0X}$. From Lemma 3.2, $F := \mathcal{A}\chi + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}\chi)$ belongs to $\mathcal{R}_\Omega H_X^{-1}$. Then

$$\begin{aligned}\langle\!\langle \mathcal{A}^*\psi + \mathcal{M}^*\cdot \boldsymbol{D}_{\dot{W}}\psi,\, \chi\rangle\!\rangle &= \langle\!\langle \psi,\, \mathcal{A}\chi + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}\chi)\rangle\!\rangle = \lim_{p\to\infty}\langle\!\langle \psi^p, F\rangle\!\rangle \\ &= \lim_{p\to\infty}\langle\!\langle \mathcal{A}^*\psi^p + \mathcal{M}^*\cdot \boldsymbol{D}_{\dot{W}}\psi^p,\, \chi\rangle\!\rangle = \lim_{p\to\infty}\langle\!\langle \phi^p, \chi\rangle\!\rangle = \langle\!\langle \phi, \chi\rangle\!\rangle.\end{aligned}$$

By definition, the solution ψ satisfies the hypothesis of Proposition 3.4, hence the estimate (3.13) holds. □

Remark. Higher spatial regularity results follow as usual from the corresponding deterministic results for each equation in the propagator. In a similar fashion to the proof of

Theorem 3.5, one can obtain higher regularity estimates such as

$$\|\psi\|_{\mathcal{R}_\Omega^{-1} H_X^r} \leq C\|\phi\|_{\mathcal{R}_\Omega^{-1} H_X^{r-2}}$$

for $r \geq 1$, if $\phi \in \mathcal{R}_\Omega^{-1} H_X^{r-2}$, and if the boundary ∂D and the coefficients of $\mathcal{A}, \mathcal{M}_k$ are sufficiently smooth.

3.2. Error estimates for the corresponding elliptic SPDE. An extension of [**65**] to random forcing terms yields the following result for the approximation error of the SFEM approximation $U_h^{M,n}$ of equation (3.8).

Theorem 3.6. *Suppose $U \in \mathcal{R}_\Omega H_{0X}^1 \cap \mathcal{R}_\Omega H_X^{m+1}$, where the weights satisfy*

$$\sum_k q_k (\lambda_k C_A^{ellip})^2 < \frac{1}{2}, \quad \text{and} \quad \sum_k \frac{q_k}{\bar{\rho}_k} < \frac{1}{2}. \tag{3.15}$$

Then the error of approximation of the stochastic finite element method is given by

$$\|U - U_h^{M,n}\|_{\mathcal{R}_\Omega H_{0X}^1} \leq C_{M,n} h^m \|U\|_{\mathcal{R}_\Omega H_X^{m+1}} + C\|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} Q_{M,n}(\mathcal{R}, \bar{\mathcal{R}}) \tag{3.16}$$

Here, $C_{M,n}$ can be taken as

$$C_{M,n} = C' \binom{M+n}{M}$$

and the constants C, C' are independent of h, M, n. The term

$$Q_{M,n}(\mathcal{R}, \bar{\mathcal{R}}) = \sqrt{\frac{\hat{Q}_W}{(1-\hat{Q})^2} + \frac{\hat{Q}^{n+1}}{1-\hat{Q}}}$$

where

$$\hat{Q} = \sum_{k \geq 1} q_k (\lambda_k C_A^{ellip})^2 + \frac{q_k}{\bar{\rho}_k} < 1, \quad \text{and} \quad \hat{Q}_W = \sum_{k > M} q_k (\lambda_k C_A^{ellip})^2 + \frac{q_k}{\bar{\rho}_k}.$$

Proof. The first part of this proof closely follows the proof in [**65**]. Denote the numerical solution by $U_h^{M,n} = \sum_{\alpha \in \mathcal{J}_{M,p}} \hat{U}_\alpha \xi_\alpha$. We decompose the approximation error into two components,

$$\begin{aligned} \|U - U_h^{M,n}\|^2_{\mathcal{R}_\Omega H_{0X}^1} &= \sum_{\alpha \in \mathcal{J}_{M,p}} \|U_\alpha - \hat{U}_\alpha\|^2_{H_X^1} r_\alpha^2 + \sum_{\alpha \in \mathcal{J} \setminus \mathcal{J}_{M,p}} \|U_\alpha\|^2_{H_X^1} r_\alpha^2 \\ &=: I_1 + I_2 \end{aligned}$$

For Term I_1, we use the definitions of the weak and numerical solution for each equation in the propagator system,

$$\left\langle \mathcal{A}\hat{U}_\alpha + \sum_{k=1}^{M} \sqrt{\alpha_k}\mathcal{M}_k \hat{U}_{\alpha-\epsilon_k}\, v_h \right\rangle = \langle f_\alpha, v_h \rangle = \left\langle \mathcal{A}U_\alpha + \sum_{k=1}^{M} \sqrt{\alpha_k}\mathcal{M}_k U_{\alpha-\epsilon_k}\, v_h \right\rangle$$

for all $v_h \in S_h$. Note that we are assuming complete knowledge of the forcing term F. An application of the approximation techniques in the classical finite element theory yields, (see the Online Supplementary Material of [**65**] for details),

$$\|U_\alpha - \hat{U}_\alpha\|_{H^1_X} \le \hat{C}_A \inf_{v_h \in S_h} \|U_\alpha - v_h\|_{H^1_X} + \sum_{k=1}^{M} \sqrt{\alpha_k} C_k \|U_{\alpha-\varepsilon_k} - \hat{U}_{\alpha-\varepsilon_k}\|_{H^1_X} \tag{3.17}$$

where $\hat{C}_A = (1 + C^b_A C^{ellip}_A)$ and $C_k := \lambda_k C^{ellip}_A$. By induction,

$$\|U_\alpha - \hat{U}_\alpha\|_{H^1_X} \le \hat{C}_A \sum_{\beta \le \alpha} c_{\alpha,\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X} \tag{3.18}$$

where $c_{\alpha,\beta}$ are constants depending on α, β. The following Lemma gives a possible choice for $c_{\alpha,\beta}$.

Lemma 3.7. *Denote $\vec{C} = (C_1, C_2, \dots)$. Then the constants $c_{\alpha,\beta}$ in* (3.18) *may be taken as*

$$c_{\alpha,\beta} = \frac{|\alpha - \beta|!}{\sqrt{(\alpha-\beta)!}} \sqrt{\binom{\alpha}{\beta}}\, \vec{C}^{\alpha-\beta}.$$

Proof. This is done by induction. Suppose

$$\|U_\gamma - \hat{U}_\gamma\|_{H^1_X} \le \hat{C}_A \sum_{\beta \le \gamma} c_{\gamma,\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}$$

for all $|\gamma| \le n-1$, dim $\gamma \le M$. Let $|\alpha| = n$. Then the second term on the RHS of (3.17) is

$$\begin{aligned}
&\sum_{k=1}^{M} \sqrt{\alpha_k} C_k \|U_{\alpha-\varepsilon_k} - \hat{U}_{\alpha-\varepsilon_k}\|_{H^1_X} \\
&= \hat{C}_A \sum_{k=1}^{M} \sqrt{\alpha_k} C_k \sum_{\beta \le \alpha-\varepsilon_k} c_{\alpha-\varepsilon_k,\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X} \\
&= \hat{C}_A \sum_{k=1}^{M} \sqrt{\alpha_k} \sum_{\beta \le \alpha-\varepsilon_k} \frac{|\alpha - 1 - \beta|!}{\sqrt{(\alpha-\varepsilon_k-\beta)!}} \sqrt{\binom{\alpha-\varepsilon_k}{\beta}} \vec{C}^{\alpha-\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}
\end{aligned}$$

$$= \hat{C}_A \sum_{\substack{k=1 \\ \alpha_k \neq 0}}^{M} \sum_{\beta \leq \alpha - \varepsilon_k} \frac{|\alpha - 1 - \beta|!}{\sqrt{(\alpha-\beta)!}} \sqrt{\binom{\alpha}{\beta}} (\alpha_k - \beta_k) \vec{C}^{\alpha-\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}$$

$$\leq \hat{C}_A \sum_{\substack{k=1 \\ \alpha_k \neq 0}}^{M} \sum_{\beta < \alpha} \frac{|\alpha - 1 - \beta|!}{\sqrt{(\alpha-\beta)!}} \sqrt{\binom{\alpha}{\beta}} (\alpha_k - \beta_k) \vec{C}^{\alpha-\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}$$

$$= \hat{C}_A \sum_{\beta < \alpha} \sum_{\substack{k=1 \\ \alpha_k \neq 0}}^{M} (\alpha_k - \beta_k) \frac{|\alpha - 1 - \beta|!}{\sqrt{(\alpha-\beta)!}} \sqrt{\binom{\alpha}{\beta}} \vec{C}^{\alpha-\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}$$

$$= \hat{C}_A \sum_{\beta < \alpha} \frac{|\alpha - \beta|!}{\sqrt{(\alpha-\beta)!}} \sqrt{\binom{\alpha}{\beta}} \vec{C}^{\alpha-\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}$$

$$= \hat{C}_A \sum_{\beta < \alpha} c_{\alpha,\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}$$

Hence,

$$\begin{aligned} \|U_\alpha - \hat{U}_\alpha\|_{H^1_X} &\leq \hat{C}_A \inf_{v_h \in S_h} \|U_\alpha - v_h\|_{H^1_X} + \sum_{k=1}^{M} \sqrt{\alpha_k} C_k \|U_{\alpha-\varepsilon_k} - \hat{U}_{\alpha-\varepsilon_k}\|_{H^1_X} \\ &\leq \hat{C}_A \sum_{\beta \leq \alpha} c_{\alpha,\beta} \inf_{v_h \in S_h} \|U_\beta - v_h\|_{H^1_X}. \end{aligned}$$

□

From Lemma 3.7 and denoting the constant in (3.3) by C_{FE}, we obtain

$$\begin{aligned} \|U_\alpha - \hat{U}_\alpha\|^2_{H^1_X} r^2_\alpha &\leq h^{2m} C^2_{FE} \hat{C}^2_A \left(\sum_{\beta \leq \alpha} c_{\alpha,\beta} \|U_\beta\|_{H^{m+1}} r_\alpha \right)^2 \\ &\leq h^{2m} C^2_{FE} \hat{C}^2_A \left(\sum_{\beta \leq \alpha} \frac{r^2_\alpha}{r^2_\beta} c^2_{\alpha,\beta} \right) \left(\sum_{\beta \leq \alpha} r^2_\beta \|U_\beta\|^2_{H^{m+1}_X} \right) \\ &\leq h^{2m} C^2_{FE} \hat{C}^2_A \left(\sum_{\beta \leq \alpha} \binom{|\alpha|}{|\beta|}^{-1} r^2_{\alpha-\beta} c^2_{\alpha,\beta} \right) \|U\|^2_{\mathcal{R}\Omega H^{m+1}_X} \end{aligned}$$

So

$$\sum_{\alpha \in \mathcal{J}_{M,p}} \|U_\alpha - \hat{U}_\alpha\|^2_{H^1_X} r^2_\alpha \leq h^{2m} C^2_{FE} \hat{C}^2_A \|U\|^2_{\mathcal{R}\Omega H^{m+1}_X} \underbrace{\left(\sum_{\alpha \in \mathcal{J}_{M,p}} \sum_{\beta \leq \alpha} \binom{|\alpha|}{|\beta|}^{-1} r^2_{\alpha-\beta} c^2_{\alpha,\beta} \right)}_{(*)}$$

To estimate $(*)$, since $\binom{|\alpha|}{|\beta|}^{-1}\binom{\alpha}{\beta} < 1$ due to Lemma 1.3,

$$
\begin{aligned}
(*) &= \sum_{\alpha\in\mathcal{J}_{M,p}} \sum_{\beta\le\alpha} \binom{|\alpha|}{|\beta|}^{-1} r^2_{\alpha-\beta}\frac{|\alpha-\beta|!^2}{(\alpha-\beta)!}\binom{\alpha}{\beta}\vec{C}^{2(\alpha-\beta)} \\
&\le \sum_{\alpha\in\mathcal{J}_{M,p}} \sum_{\beta\le\alpha} (q^2\vec{C}^2)^{\alpha-\beta}\frac{|\alpha-\beta|!}{(\alpha-\beta)!} \\
&= \sum_{\beta\in\mathcal{J}_{M,p}} \sum_{\substack{\alpha\ge\beta \\ \alpha\in\mathcal{J}_{M,p}}} (q^2\vec{C}^2)^{\beta}\frac{|\beta|!}{\beta!} \\
&= \sum_{\beta\in\mathcal{J}_{M,p}} (q^2\vec{C}^2)^{\beta}\frac{|\beta|!}{\beta!} \times (\#\{\alpha\in\mathcal{J}_{M,p} : \alpha\ge\beta\}) \\
&= \sum_{n=0}^{p} \sum_{\substack{|\beta|=n \\ \dim\,\beta\le M}} (q^2\vec{C}^2)^{\beta}\frac{n!}{\beta!} \times \left(\binom{M+p}{M} - \frac{2^n}{\beta!}\right) \\
&\le \binom{M+p}{M}\sum_{n=0}^{p}[q]^n_{\le M} \le \binom{M+p}{M}\frac{1}{1-\hat{q}}
\end{aligned}
$$

where $[q]_{\le M} := \sum_{k=1}^{M} q_k^2 C_k^2 = \hat{q} - \hat{q}_W$. This gives the first term in the RHS of (3.16).

For term I_2, we recall the estimates (3.10). We decompose the sum in Term I_2 into

$$
\sum_{\alpha\in\mathcal{J}\setminus\mathcal{J}_{M,p}} = \sum_{n=0}^{p}\sum_{i=0}^{n-1}\sum_{\left\{\alpha:\,\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}\right\}} + \sum_{n=p+1}^{\infty}\sum_{i=0}^{n}\sum_{\left\{\alpha:\,\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}\right\}} .
$$

Consider the innermost sum

$$
\begin{aligned}
&\sum_{\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}} \|U_\alpha\|^2_{H^1_X} r^2_\alpha \le \sum_{\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}} (C_A^{ellip})^2 q^\alpha \left(\sum_{\beta\le\alpha}\|F_{\alpha-\beta}\|_{H_X^{-1}}\vec{C}^\beta\sqrt{\frac{|\beta|!}{\beta!(\alpha-\beta)!}}\right)^2 \\
&\le (C_A^{ellip})^2 \sum_{\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}} q^\alpha \left(\sum_{\beta\le\alpha}\bar{r}^2_{\alpha-\beta}\|F_{\alpha-\beta}\|^2_{H_X^{-1}}\right)\left(\sum_{\beta\le\alpha}\bar{r}^{-2}_{\alpha-\beta}\vec{C}^{2\beta}\frac{|\beta|!}{\beta!(\alpha-\beta)!}\right) \\
&\le (C_A^{ellip})^2\|F\|^2_{\bar{\mathcal{R}}_\Omega H_X^{-1}} \sum_{\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}} \sum_{\beta\le\alpha} (q\vec{C}^2)^\beta\left(\frac{q}{\bar{\rho}}\right)^{\alpha-\beta}\frac{|\beta|!|\alpha-\beta|!}{\beta!(\alpha-\beta)!} \\
&= (C_A^{ellip})^2\|F\|^2_{\bar{\mathcal{R}}_\Omega H_X^{-1}} \sum_{k=0}^{i}\sum_{l=0}^{n-i}\sum_{\substack{|\beta^{(1)}|=k \\ |\beta^{(2)}|=l}}\sum_{\substack{|\gamma^{(1)}|=i-k \\ |\gamma^{(2)}|=n-i-l}} (q\vec{C}^2)^\beta\left(\frac{q}{\bar{\rho}}\right)^{\alpha-\beta}\frac{|\beta|!|\alpha-\beta|!}{\beta!(\alpha-\beta)!}
\end{aligned}
$$

We introduce the notation, for $\rho = (\rho_1, \rho_2, \dots)$,

$$[\rho]_{\le M} = \sum_{k=1}^{M} \rho_k, \qquad [\rho]_{>M} = \sum_{k=M+1}^{\infty} \rho_k.$$

Then

$$\begin{aligned}
&\sum_{\substack{|\alpha^{(1)}|=i \\ |\alpha^{(2)}|=n-i}} \|U_\alpha\|^2_{H^1_X} r^2_\alpha \\
&\le (C_A^{ellip})^2 \|F\|^2_{\bar{\mathcal{R}}_\Omega H_X^{-1}} \sum_{k=0}^{i} \sum_{l=0}^{n-i} [q\vec{C}^2]^k_{\le M} [q\vec{C}^2]^l_{>M} \binom{k+l}{k} \left[\frac{q}{\bar{\rho}}\right]^{i-k}_{\le M} \left[\frac{q}{\bar{\rho}}\right]^{n-i-l}_{>M} \binom{n-k-l}{i-k} \\
&\le (C_A^{ellip})^2 \|F\|^2_{\bar{\mathcal{R}}_\Omega H_X^{-1}} \binom{n}{i} \left([q\vec{C}^2]_{\le M} + \left[\frac{q}{\bar{\rho}}\right]_{\le M} \right)^i \left([q\vec{C}^2]_{>M} + \left[\frac{q}{\bar{\rho}}\right]_{>M} \right)^{n-i}
\end{aligned}$$

The rest of the proof proceeds identically to the proof in [**65**], and we obtain the second term in the RHS of (3.16).

□

Remark. Define the (Ritz) projection $\Pi_h^{M,n} : \bar{\mathcal{R}}_\Omega H^1_{0X} \to S_h \otimes S^{M,n}$ as the *stochastic finite element approximation operator* for the stochastic elliptic problem (3.8). More precisely, for $U \in \bar{\mathcal{R}}_\Omega H^1_{0X}$, the projection $\Pi_h^{M,n} U$ is the stochastic finite element method's solution of the elliptic SPDE (3.8), satisfying

$$\langle\!\langle \mathcal{A}U + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k U), z \rangle\!\rangle = \langle\!\langle \mathcal{A}(\Pi_h^{M,n} U) + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k(\Pi_h^{M,n} U)), z \rangle\!\rangle \tag{3.19}$$

for all $z \in S_h \otimes S^{M,n}$. Due to Lemma 3.2, $F := \mathcal{A}U + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}U) \in \bar{\mathcal{R}}_\Omega H^1_{0X}$, and in view of 3.6, the estimates (3.16) hold with $U_h^{M,n} = \Pi_h^{M,n} U$. This also implies that $\Pi_h^{M,n}$ is a continuous linear map from $\mathcal{R}_\Omega H^1_{0X}$ into itself.

We will also need error estimates in the $L_2(D)$ and $H^{-1}(D)$ norms.

Proposition 3.8. *Under the same assumptions as Theorem 3.6, the error of approximation of the SFEM has the bounds*

$$\|U - U_h^{M,n}\|_{\mathcal{R}_\Omega H_X^{1-k}} \le C_{M,n} h^{m+k} \|U\|_{\mathcal{R}_\Omega H_X^{m+1}} + C\|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} Q_{M,n}(\mathcal{R}, \bar{\mathcal{R}}) \tag{3.20}$$

for $k = 1, 2$.

Proof. As in the proof of Theorem 3.6,

$$U - U_h^{M,p} = \sum_{\alpha \in \mathcal{J}_{M,p}} (U_\alpha - \hat{U}_\alpha)\xi_\alpha + \sum_{\alpha \in \mathcal{J} \setminus \mathcal{J}_{M,p}} U_\alpha \xi_\alpha =: e_1 + e_2,$$

with

$$\|e_1\|_{\mathcal{R}_\Omega H^1_{0X}} \le C_{M,n} h^m \|U\|_{\mathcal{R}_\Omega H^{m+1}_X}, \quad \text{and}$$
$$\|e_2\|_{\mathcal{R}_\Omega H^1_{0X}} \le C\|F\|_{\bar{\mathcal{R}}_\Omega H^{-1}_X} Q_{M,n}(\mathcal{R}, \bar{\mathcal{R}})$$

We leave the estimate for e_2 untouched. For e_1, we consider the two cases.

<u>Case: $k = 1$.</u> Let $\psi \in \mathcal{R}_\Omega^{-1} H^2_X$ be the solution of $\mathcal{A}\psi + \mathcal{M} \cdot \boldsymbol{D}_{\dot{W}}\psi = \mathcal{R}^2 e_1$, with $\|\psi\|_{\mathcal{R}_\Omega^{-1} H^2_X} \le C\|\mathcal{R}^2 e_1\|_{\mathcal{R}_\Omega^{-1} L^2_X} = \|e_1\|_{\mathcal{R}_\Omega L^2_X}$. Note that, in fact, $\psi \in S^{M,n} \otimes H^3_X$ also. Then,

$$\begin{aligned}\|e_1\|^2_{\mathcal{R}_\Omega L^2_X} &= \langle\!\langle e_1, \mathcal{R}^2 e_1 \rangle\!\rangle_{\mathcal{R}_\Omega L^2_X, \mathcal{R}_\Omega^{-1} L^2_X} = \langle\!\langle e_1, \mathcal{R}^2 e_1 \rangle\!\rangle_{\mathcal{R}_\Omega H^{-1}_X, \mathcal{R}_\Omega^{-1} H^1_X} \\ &= \langle\!\langle e_1, \mathcal{A}\psi + \mathcal{M} \cdot \boldsymbol{D}_{\dot{W}}\psi \rangle\!\rangle_{\mathcal{R}_\Omega H^{-1}_X, \mathcal{R}_\Omega^{-1} H^1_X} \\ &= \langle\!\langle \mathcal{A} e_1 + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M} e_1),\ \psi - \chi \rangle\!\rangle_{\mathcal{R}_\Omega H^{-1}_X, \mathcal{R}_\Omega^{-1} H^1_X}\end{aligned}$$

for all $\chi \in S^{M,n} \otimes S_h$. So

$$\|e_1\|^2_{\mathcal{R}_\Omega L^2_X} \le \|\mathcal{A} e_1 + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M} e_1)\|_{\mathcal{R}_\Omega H^{-1}_X} \inf_{\chi \in S^{M,n} \otimes S_h} \|\psi - \chi\|_{\mathcal{R}_\Omega^{-1} H^1_X}$$

To estimate the first term, Lemma 3.2 implies that

$$\|\mathcal{A} e_1 + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M} e_1)\|_{\mathcal{R}_\Omega H^{-1}_X} \le C\|e_1\|_{\mathcal{R}_\Omega H^1_{0X}}$$

To estimate the second term, we make use of the FE estimate (3.3), in particular

$$\inf_{\chi_h \in S_h} \|\Phi - \chi_h\|_{H^1_{0X}} \le Ch\|\Phi\|_{H^2_X}, \quad \forall \Phi \in H^2_X \cap H^1_{0X}.$$

This FE estimate is usually obtained by finding a projection operator I_h for which $\|\Phi - I_h\Phi\|_{H^1_{0X}} \le Ch^2\|\Phi\|_{H^3_X}$, from which the desired estimate follows immediately. But here, we will show the estimate by constructing a near-infimizing χ. Fix $\epsilon > 0$. For each $\alpha \in \mathcal{J}_{M,n}$, there exists $\chi_\alpha \in S_h$ such that

$$\|\psi_\alpha - \chi_\alpha\|_{H^1_{0X}} \leq \inf_{\chi_h \in S_h} \|\psi_\alpha - \chi_h\|_{H^1_{0X}} + \kappa_\alpha(\epsilon) \leq Ch\|\psi_\alpha\|_{H^2_X} + \kappa_\alpha(\epsilon)$$

where we choose $\kappa_\alpha(\epsilon) = \epsilon^{1/2} r_\alpha \bar{\kappa}_\alpha$, with $\sum_\alpha \bar{\kappa}_\alpha^2 = \frac{1}{2}$. Set $\chi = \sum_{\alpha \in \mathcal{J}_{M,n}} \chi_\alpha \xi_\alpha \in S^{M,n} \otimes S_h$. Then

$$\|\psi - \chi\|^2_{\mathcal{R}_\Omega^{-1} H^1_{0X}} \leq \sum_{\alpha \in \mathcal{J}_{M,n}} r_\alpha^{-2} \left(Ch\|\psi_\alpha\|_{H^2_X} + \kappa_\alpha(\epsilon) \right)^2 \leq Ch^2 \|\psi\|^2_{\mathcal{R}_\Omega^{-1} H^2_X} + \epsilon$$

and

$$\inf_{\chi \in S^{M,n} \otimes S_h} \|\psi - \chi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}} \leq Ch\|\psi\|_{\mathcal{R}_\Omega^{-1} H^2_X} \leq Ch\|e_1\|_{\mathcal{R}_\Omega H^1_{0X}}$$

Hence,

$$\begin{aligned} \|e_1\|^2_{\mathcal{R}_\Omega L^2_X} &\leq \|e_1\|_{\mathcal{R}_\Omega H^1_{0X}} Ch \|\psi\|_{\mathcal{R}_\Omega^{-1} H^2_X} \\ &\leq C_{M,n} h^{m+1} \|U\|_{\mathcal{R}_\Omega H^{m+1}_X} \|e_1\|_{\mathcal{R}_\Omega L^2_X}. \end{aligned}$$

Case: $k = 2$. Since $e_1 \in S^{M,n} \otimes H^1_{0X}$, we compute the norm

$$\|e_1\|_{\mathcal{R}_\Omega H^{-1}_X} = \sup_{\phi \in \mathcal{R}_\Omega^{-1} H^1_{0X}} \frac{|\langle\langle e_1, \phi \rangle\rangle|}{\|\phi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}}} = \sup_{\phi \in S^{M,n} \otimes H^1_{0X}} \frac{|\langle\langle e_1, \phi \rangle\rangle|}{\|\phi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}}}$$

For any $\phi \in S^{M,n} \otimes H^1_{0X}$, let $\psi \in \mathcal{R}_\Omega^{-1} H^3_X$ be the solution of $\mathcal{A}\psi + \mathcal{M} \cdot \boldsymbol{D}_{\dot{W}} \psi = \phi$, with $\|\psi\|_{\mathcal{R}_\Omega^{-1} H^3_X} \leq C\|\phi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}}$. Note that, in fact, $\psi \in S^{M,n} \otimes H^3_X$ also. Then,

$$\langle\langle e_1, \phi \rangle\rangle = \langle\langle e_1, \mathcal{A}\psi + \mathcal{M} \cdot \boldsymbol{D}_{\dot{W}} \psi \rangle\rangle = \langle\langle \mathcal{A} e_1 + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M} e_1), \, \psi - \chi \rangle\rangle$$

for all $\chi \in S^{M,n} \otimes S_h$, and by a similar argument in the previous case, we have that

$$\begin{aligned} |\langle\langle e_1, \phi \rangle\rangle| &\leq \|\mathcal{A} e_1 + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M} e_1)\|_{\mathcal{R}_\Omega H^{-1}_X} \inf_{\chi \in S^{M,n} \otimes S_h} \|\psi - \chi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}} \\ &\leq Ch^2 \|e_1\|_{\mathcal{R}_\Omega H^1_{0X}} \|\phi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}} \\ &\leq C_{M,n} h^{m+2} \|U\|_{\mathcal{R}_\Omega H^{m+1}} \|\phi\|_{\mathcal{R}_\Omega^{-1} H^1_{0X}}. \end{aligned}$$

The result follows. □

3.3. The parabolic error estimates.

Theorem 3.9. *Let $m \geq 2$ be an even integer. Assume for the input data*

$$v \in \bar{\mathcal{R}}_\Omega H_X^{m+1}, \quad f \in \bar{\mathcal{R}}_\Omega L_T^2 H_X^m, \quad f_t \in \bar{\mathcal{R}}_\Omega L_T^2 H_X^{m-2},$$

with weights $\bar{r}_\alpha^2 = \frac{\bar{\rho}^\alpha}{|\alpha|!}$, and assume that the appropriate compatibility conditions hold, so that

$$u \in \mathcal{R}'_\Omega L_T^2 H_{0X}^1 \cap \mathcal{R}'_\Omega L_T^2 H_X^{m+2}, \quad u_t \in \mathcal{R}'_\Omega L_T^2 H_X^{-1} \cap \mathcal{R}'_\Omega L_T^2 H_X^m,$$
$$u_{tt} \in \mathcal{R}'_\Omega L_T^2 H_X^{m-2},$$

where the weights ${\rho'_\alpha}^2 = \frac{\rho'^\alpha}{|\alpha|!}$ are chosen using the conditions (2.15) *and* (2.17)*. Also assume, for simplicity, that the discretized initial condition is $v_h = \Pi_h^{M,n} v$. Then, for every $t \in (0, T]$, we have the error estimate for the stochastic finite element solution $u_h^{M,n}(t)$,*

$$\begin{aligned} \|e_h(t)\|_{\mathcal{R}_\Omega L_X^2} \leq C_{M,n} h^{m+1} \Big(\|u_t\|_{\mathcal{R}_\Omega L_T^2 H_X^m} + \|u(t)\|_{\mathcal{R}_\Omega H_X^{m+1}}\Big) \\ + C Q_{M,n}(\mathcal{R}, \mathcal{R}') \Big(\|f_t - u_{tt}\|_{\mathcal{R}'_\Omega L_T^2 H_X^{-1}} + \|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H_X^{-1}}\Big) \end{aligned} \tag{3.21}$$

where the weights $\mathcal{R}$, ${r_\alpha}^2 = \frac{q^\alpha}{|\alpha|!}$, satisfy

$$\sum_k q_k \lambda_k^2 \left(C_A^{ellip}\right)^2 < \frac{1}{2}, \quad \textit{and} \quad \sum_k \frac{q_k}{\rho'_k} < \frac{1}{2}. \tag{3.22}$$

Proof. Let $\Pi_h^{M,n}$ denote the stochastic finite element approximation operator for the stochastic *elliptic* problem (3.8). In particular,

$$\left\langle\!\left\langle \mathcal{A}U + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k U), z \right\rangle\!\right\rangle = \left\langle\!\left\langle \mathcal{A}(\Pi_h^{M,n} U) + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k(\Pi_h^{M,n} U)), z \right\rangle\!\right\rangle$$

for all $z \in S^{M,n} \otimes S_h$. The error estimates (3.16) also imply that $\Pi_h^{M,n}$ is a continuous linear map from $\mathcal{R}_\Omega H_{0X}^1$ into itself.

Decompose the error into

$$\begin{aligned} e_h(t) := u_h^{M,n}(t) - u(t) &= \Big(u_h^{M,n}(t) - \Pi_h^{M,n} u(t)\Big) + \Big(\Pi_h^{M,n} u(t) - u(t)\Big) \\ &= \theta(t) + \pi(t). \end{aligned}$$

Analysis for π. For every $t \in (0,T]$, we have that $\mathcal{A}u(t) + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}u(t)) = f(t) - u_t(t) \in \mathcal{R}'_\Omega H_X^{m-1}$. Hence the elliptic estimates (3.16) and lower norm estimates (3.20) imply

$$\begin{aligned}
&\|\pi(t)\|_{\mathcal{R}_\Omega L^2_X} = \|\Pi_h^{M,n} u(t) - u(t)\|_{\mathcal{R}_\Omega L^2_X} \\
&\quad \le C_{M,n} h^{m+1} \|u(t)\|_{\mathcal{R}_\Omega H_X^{m+1}} + C\|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H_X^{-1}} Q_{M,n}(\mathcal{R}, \mathcal{R}')
\end{aligned}$$

provided (3.22) holds.

Analysis for θ. From the definitions of the numerical and weak solutions,

$$\begin{aligned}
&\langle\!\langle \theta_t, z \rangle\!\rangle + \Big\langle\!\!\Big\langle \mathcal{A}\theta + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k \theta), z \Big\rangle\!\!\Big\rangle \\
&= \langle\langle f, z \rangle\rangle - \Big\langle\!\!\Big\langle (\Pi_h^{M,n} u)_t, z \Big\rangle\!\!\Big\rangle - \Big\langle\!\!\Big\langle \mathcal{A}\Pi_h^{M,n} u + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k(\Pi_h^{M,n} u)), z \Big\rangle\!\!\Big\rangle \\
&= \langle\!\langle f, z \rangle\!\rangle - \Big\langle\!\!\Big\langle (\Pi_h^{M,n} u)_t, z \Big\rangle\!\!\Big\rangle - \Big\langle\!\!\Big\langle \mathcal{A}u + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k u), z \Big\rangle\!\!\Big\rangle \pm \langle\!\langle u_t, z \rangle\!\rangle \\
&= -\Big\langle\!\!\Big\langle (\Pi_h^{M,n} u - u)_t, z \Big\rangle\!\!\Big\rangle
\end{aligned}$$

for all $z \in S^{M,n} \otimes S_h$. Choosing $z = \mathcal{R}^2\theta$,

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\theta\|^2_{\mathcal{R}_\Omega L^2_X} + \sum_{\alpha \in \mathcal{J}_{M,n}} r_\alpha^2 \boldsymbol{A}[\theta_\alpha, \theta_\alpha] \\
&\le \|(\Pi_h^{M,n} u - u)_t\|_{\mathcal{R}_\Omega H_X^{-1}} \|\theta\|_{\mathcal{R}_\Omega H^1_{0X}} + \sum_{\alpha \in \mathcal{J}_{M,n}} \sum_{k=1}^{M} \sqrt{\alpha_k}\lambda_k r_\alpha^2 \|\theta_{\alpha - \epsilon_k}\|_{H^1_{0X}} \|\theta_\alpha\|_{H^1_{0X}} \\
&= (I) + (II)
\end{aligned}$$

where λ_k are the constants in (2.11).

For (II),

$$\begin{aligned}
(II) &= \sum_{\alpha \in \mathcal{J}_{M,n}} \sum_{k=1}^{M} \sqrt{\alpha_k}\lambda_k r_\alpha \|\theta_{\alpha - \varepsilon_k}\|_{H^1_X} r_\alpha \|\theta_\alpha\|_{H^1_X} \\
&\le \left(\sum_{\alpha \in \mathcal{J}_{M,n}} \left(\sum_{k=1}^{M} \sqrt{\alpha_k}\lambda_k r_\alpha \|\theta_{\alpha - \varepsilon_k}\|_{H^1_X} \right)^2 \right)^{1/2} \left(\sum_{\alpha \in \mathcal{J}_{M,n}} r_\alpha^2 \|\theta_\alpha\|^2_{H^1_X} \right)^{1/2}
\end{aligned}$$

$$\le \left(\sum_{\alpha \in \mathcal{J}_{M,n}} \left(\sum_{\substack{k=1 \\ \alpha_k \neq 0}}^{M} \frac{\alpha_k}{|\alpha|} \sqrt{\frac{|\alpha|}{\alpha_k}} \lambda_k q_k^{1/2} r_{\alpha-\varepsilon_k} \|\theta_{\alpha-\varepsilon_k}\|_{H^1_X} \right)^2 \right)^{1/2} \|\theta\|_{\mathcal{R}_\Omega H^1_X}$$

$$\le \left(\sum_{\alpha \in \mathcal{J}_{M,n}} \sum_{\substack{k=1 \\ \alpha_k \neq 0}}^{M} \frac{\alpha_k}{|\alpha|} \left(\sqrt{\frac{|\alpha|}{\alpha_k}} \lambda_k q_k^{1/2} r_{\alpha-\varepsilon_k} \|\theta_{\alpha-\varepsilon_k}\|_{H^1_X} \right)^2 \right)^{1/2} \|\theta\|_{\mathcal{R}_\Omega H^1_X}$$

where we applied Jensen's inequality in the last inequality. Continuing,

$$(II) \le \left(\sum_{k=1}^{M} \sum_{\substack{\alpha \in \mathcal{J}_{M,n} \\ \alpha_k \neq 0}} \lambda_k^2 q_k r^2_{\alpha-\varepsilon_k} \|\theta_{\alpha-\varepsilon_k}\|^2_{H^1} \right)^{1/2} \|\theta\|_{\mathcal{R}_\Omega H^1_X}$$

$$\le \left(\sum_{k=1}^{M} \lambda_k^2 q_k \right)^{1/2} \|\theta\|^2_{\mathcal{R}_\Omega H^1_X} := [q\vec{\lambda}^2]^{1/2}_{\le M} \|\theta\|^2_{\mathcal{R}_\Omega H^1_X}$$

where $[q\vec{\lambda}^2]_{\le M} = \sum_{k=1}^{M} \lambda_k^2 q_k$. Then

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\theta\|^2_{\mathcal{R}_\Omega L^2_X} + C_A^{coerc}\|\theta\|^2_{\mathcal{R}_\Omega H^1_{0X}}$$
$$\le \epsilon_0 \|(\Pi_h^{M,n} u - u)_t\|^2_{\mathcal{R}_\Omega H^{-1}_X} + \left(\frac{1}{4\epsilon_0} + [q\vec{\lambda}^2]^{1/2}_{\le M} \right) \|\theta\|^2_{\mathcal{R}_\Omega H^1_{0X}}$$

where C_A^{coerc} is the coercivity constant in (2.8). By the first condition in (3.22), we can find ϵ_0 such that $\frac{1}{4\epsilon_0} + [q\vec{\lambda}^2]^{1/2}_{\le M} = C_A^{coerc}$. So

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\theta\|^2_{\mathcal{R}_\Omega L^2_X} \le 2\epsilon_0 \|(\Pi_h^{M,n} u - u)_t\|^2_{\mathcal{R}_\Omega H^{-1}_X}$$

and

$$\|\theta(t)\|^2_{\mathcal{R}_\Omega L^2_X} \le \|\theta(0)\|^2_{\mathcal{R}_\Omega L^2_X} + 2\epsilon_0 \int_0^t \|(\Pi_h^{M,n} u - u)_t(s)\|^2_{\mathcal{R}_\Omega H^{-1}_X} ds.$$

Due to our assumption on the initial condition, $v_h = \Pi_h^{M,n} v$, the term $\theta(0)$ vanishes. The estimate for the second term in the last inequality is similar to the analysis for $\pi(t)$, but since the norm appears inside a time integral, it suffices to show a bound for a.e. t. Since $\Pi_h^{M,n}$ is a continuous linear map from $\mathcal{R}_\Omega H^1_{0X}$ into itself, it follows that $(\Pi_h^{M,n} u)_t = \Pi_h^{M,n} u_t$. For a.e. $s \in (0, T]$, we have that

$$\mathcal{A} u_t(s) + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M} u_t(s)) = f_t(s) - u_{tt}(s) \in \mathcal{R}'_\Omega H^{m-2}_X.$$

Then

$$\begin{aligned}(3.23)\qquad \|(\Pi_h^{M,n}u-u)_t(s)\|_{\mathcal{R}\Omega H_X^{-1}} &= \|\Pi_h^{M,n}u_t-u_t(s)\|_{\mathcal{R}\Omega H_X^{-1}}\\ &\le C_{M,n}h^{m+1}\|u_t(s)\|_{\mathcal{R}\Omega H_X^m}+C\|f_t(s)-u_{tt}(s)\|_{\mathcal{R}'_\Omega H_X^{-1}}Q_{M,n}(\mathcal{R},\mathcal{R}')\end{aligned}$$

for a.e. s, and hence

$$\|\theta(t)\|^2_{\mathcal{R}\Omega L_X^2}\le C^2_{M,n}h^{2(m+1)}\|u_t\|^2_{\mathcal{R}\Omega L_T^2H_X^m}+C\|f_t-u_{tt}\|^2_{\mathcal{R}'_\Omega L_T^2H_X^{-1}}Q_{M,n}(\mathcal{R},\mathcal{R}')^2$$

for all $t\in(0,T]$.

Putting together the estimates for $\theta(t)$ and $\pi(t)$, we obtain

$$\begin{aligned}\|e_h(t)\|^2_{\mathcal{R}\Omega L_X^2}&\le C^2_{M,n}h^{2(m+1)}\left(\|u_t\|^2_{\mathcal{R}\Omega L_T^2H_X^m}+\|u(t)\|^2_{\mathcal{R}\Omega H_X^{m+2}}\right)\\ &\quad+CQ_{M,n}(\mathcal{R},\mathcal{R}')^2\left(\|f_t-u_{tt}\|^2_{\mathcal{R}'_\Omega L_T^2H_X^{-1}}+\|f(t)-u_t(t)\|^2_{\mathcal{R}'_\Omega H_X^{-1}}\right)\end{aligned}$$

The constant C depends only on $\mathcal{R}$, $\mathcal{A}$, $\mathcal{M}$ and the elliptic estimate constant in (3.16). □

Remarks. If the discrete initial condition v_h is not $\Pi_h^{M,n}v$, additional terms will arise from approximating the initial error, but those can be subsumed into the two main terms of the error estimate.

If the boundary is not smooth enough, the use of regularity estimates for the stochastic adjoint problem in the proof of Proposition 3.8 will no longer hold. In this case, the application of the lower norm estimate to the term $\|(\Pi_h^{M,n}u-u)_t(s)\|_{\mathcal{R}\Omega H_X^{-1}}$ is no longer valid, but we can nonetheless obtain a convergence rate of $\mathcal{O}(h^{m-1})$ in the first term of (3.21).

In analogy to the deterministic equation case, the finite element convergence rate of h^{m+1} for the solution $u\in\mathcal{R}_\Omega H_T^1H_X^m$ is optimal. Without invoking the stochastic adjoint problem, it is easy to obtain a convergence rate of h^{m-1} for the solution $u\in\mathcal{R}_\Omega H_T^1H_X^m$, which is two orders worse than optimal. The gain of two orders is achieved by extracting some crucial information from the estimates of lower norms, through the application of the stochastic adjoint problem in the duality technique.

The term $Q_{M,n}(\mathcal{R},\mathcal{R}')$ in the estimate (3.21) is, as usual, the error from truncating the Wiener chaos expansion up to $\mathcal{J}_{M,n}$. It arises from invoking the error estimates for the corresponding elliptic problem, and depends on the *choice* of the weighted space $\mathcal{R}$ in which to bound the error, as well as on the weights $\mathcal{R}'$ of the forcing term in the sense of the *elliptic problem.* It also implicitly assumes that $\mathcal{R}, \mathcal{R}'$ are related by the condition (3.22). However, the second inequality in (3.22) is a somewhat strict condition. If we consider the optimal weights $\mathcal{R}'$ to behave like $\rho_k' \sim k^{-(1+\epsilon)}\lambda_k^{-2}$ for any $\epsilon > 0$, then the optimal weights $\mathcal{R}$ can behave like $q_k \sim k^{-(2+\epsilon)}\lambda_k^{-2}$ for any $\epsilon > 0$. Thus, the error estimate holds in a weighted space that is generally worse than the optimal space that the solution u belongs to. Additionally, the validity of the first and third term in the RHS of (3.21) requires the boundedness of u_{tt} in the H_X^{-1} norm. This marks the departure of the SFEM from the deterministic FEM.

4. The SFEM for SPDE with time-dependent operators

In this section, we extend the results in [**44**] to allow the noise term, as well as the operator $\mathcal{A}$, to depend on time. As discussed in the previous chapter, such time-dependent noise encompasses a variety of equations driven by an abstract noise, such as equations with space-time white noise, or equations having two independent noise terms, one purely spatial and the other purely temporal.

Most steps of the analysis for the time-independent noise case carry over to the time-dependent case, except for the step estimating the norm $\|(\Pi_h^{M,n}u - u)_t\|_{\mathcal{R}_\Omega H_X^{-1}}$, where $\Pi_h^{M,n}$ is the *elliptic* SFEM approximation operator. When the noise is purely spatial (and the operators $\mathcal{A}, \mathcal{M}_k$ are independent of time), the equality $(\Pi_h^{M,n}u - u)_t = \Pi_h^{M,n}u_t - u_t$ allows an immediate application of the elliptic error estimates in (3.23). On the other hand, if the noise depends on time, the elliptic SFEM approximation operator $\Pi_h^{M,n}(t)$ also depends on the time parameter, and by the product rule,

$$(\Pi_h^{M,n}(t)u(t))_t = \dot{\Pi}_h^{M,n}(t)u(t) + \Pi_h^{M,n}(t)u_t(t).$$

Thus, to complete the error estimates, we need to derive estimates for the time derivative of the SFEM approximation operator, $\dot{\Pi}_h^{M,n}(t)$.

We assume the operators $\mathcal{A}(t), \mathcal{M}_k(t)$ are of the form (3.2), with $a^{ij}, \sigma_k^{ij} \in H_T^1 W_X^{m+2,\infty}$ for some $m \geq 2$. Then a^{ij}, σ_k^{ij} are Lipschitz continuous in time and their time derivatives $a_t^{ij}, (\sigma_k^{ij})_t$ exist a.e. Because much of the analysis hinges on studying the corresponding elliptic problem, which obviously has no time evolution, we emphasize that the operators

$\mathcal{A}(t), \mathcal{M}_k(t)$ will also be understood as being *parameterized* by time t. We define the time-parameterized operators $\dot{\mathcal{A}}(t), \dot{\mathcal{M}}_k(t)$ by

$$\dot{\mathcal{A}}(t)u = -\sum_{i,j} D_i(a_t^{ij}(x,t)D_j u),$$
$$\dot{\mathcal{M}}_k(t)u = \sum_{i,j} D_i((\sigma_k^{ij})_t(x,t)D_j u).$$

We recall the constants C_A and $\lambda_k^{(r)}$ in (2.10), (2.11), and also define $\dot{C}_A$ and $\dot{\lambda}_k^{(r)}$ to be the constants in

$$\|w\|_{L^2(0,T;H_0^1(D))} \leq \dot{C}_A(\|w_0\|_{L^2(D)} + \|f\|_{L^2(0,T;H^{-1}(D))})$$

for the weak solution w of the zero Dirichlet problem $\frac{dw}{dt} + \dot{\mathcal{A}}(t)w = f$ with $w(0) = w_0$; and in

$$\|\dot{\mathcal{M}}_k(t)w\|_{H^{r-2}(D)} \leq \dot{\lambda}_k^{(r)}\|w\|_{H^r(D)}, \qquad \forall w \in H^r(D), t \in (0,T]$$

For brevity, we write $\dot{\lambda}_k = \dot{\lambda}_k^{(1)}$.

4.1. The stochastic elliptic problem with time parameterized operators. For fixed $t \in (0,T]$, define the operator $\mathcal{L}(t)U := \mathcal{A}(t)U + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}(t)U)$. We will study the time-parameterized elliptic problem

$$\begin{aligned} &\mathcal{L}(t)U = F \quad \text{in } D \\ &U|_{\partial D} = 0. \end{aligned} \tag{3.24}$$

The elliptic SFEM approximation (Ritz) operator $\Pi_h^{M,n}(t)$ satisfies

$$\left\langle\!\left\langle \mathcal{A}(t)U + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k(t)U), z \right\rangle\!\right\rangle = \left\langle\!\left\langle \mathcal{A}(t)(\Pi_h^{M,n}(t)U) + \sum_{k=1}^{M} \boldsymbol{\delta}_{\xi_k}(\mathcal{M}_k(t)(\Pi_h^{M,n}(t)U)), z \right\rangle\!\right\rangle$$

for all $z \in S^{M,n} \otimes S_h$, and all $t \in (0,T]$. We have termed $\Pi_h^{M,n}(t)$ the "approximation" operator because $\Pi_h^{M,n}(t)U$ produces a finite approximation of U. An alternative take on the SFEM approximation operator $\Pi_h^{M,n}(t)U$ is to consider instead the SFEM *solution* operator $T_h^{M,n}(t)F$, which creates from F a finite solution of the elliptic problem. We now define $T_h^{M,n}(t)$.

For $F \in \bar{\mathcal{R}}_\Omega H_X^{-1}$, and for the weights $\widetilde{\mathcal{R}}$ satisfying

$$\sum_k \tilde{q}_k \lambda_k^2 C_A^2 < 1, \quad \text{and} \quad \sum_k \frac{\tilde{q}_k}{\bar{\rho}_k} < 1, \tag{3.25}$$

define $T(t) : \bar{\mathcal{R}}_\Omega H_X^{-1} \to \widetilde{\mathcal{R}}_\Omega H_{0X}^1$ to be the solution operator for the equation (3.24). That is, for each $t \in (0, T]$, $U = U(t) := T(t)F$ is the weak solution of (3.24).

Define the SFEM solution operator $T_h^{M,n}(t) : \bar{\mathcal{R}}_\Omega H_X^{-1} \to S^{M,n} \otimes S_h$ by $T_h^{M,n}(t)F \equiv \Pi_h^{M,n}(t)U$. Then,

$$e(t) := \Pi_h^{M,n}(t)U - U = T_h^{M,n}(t)F - T(t)F,$$

and also

$$e_t(t) = (\dot{T}_h^{M,n}(t) - \dot{T}(t))F.$$

The dot $\dot{}$ stands for time differentiation, and the derivatives $\dot{T}_h^{M,n}(t), \dot{T}(t)$ are understood in the weak sense,

$$\int_0^T \dot{T}(t)\varphi(t)dt = -\int_0^T T(t)\dot{\varphi}(t)dt$$

for all smooth functions φ.

From the usual elliptic error estimates (3.16), (3.20), given $F \in \bar{\mathcal{R}}_\Omega H_X^{-1}$ and $U \in \widetilde{\mathcal{R}}_\Omega H_X^{r+1}$, with $\widetilde{\mathcal{R}}, \bar{\mathcal{R}}$ satisfying (3.25) with $\frac{1}{2}$ on the RHS of both inequalities, we have that

$$\|(T_h^{M,n}(t) - T(t))F\|_{\widetilde{\mathcal{R}}_\Omega H_X^{1-k}} \le C_{M,n} h^{r+k} \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} + C\|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} Q_{M,n}(\widetilde{\mathcal{R}}, \bar{\mathcal{R}})$$

for all $t \in (0, T]$, and for $k = 0, 1, 2$. The following two propositions show that similar estimates hold for $(\dot{T}_h^{M,n}(t) - \dot{T}(t))F$, with $k = 0, 2$.

Proposition 4.1. *Assume $U(t) \in \widetilde{\mathcal{R}}_\Omega H_{0X}^1 \cap \widetilde{\mathcal{R}}_\Omega H_X^{r+1}$, with the weights $\tilde{r}_\alpha^2 = \tilde{q}^\alpha/|\alpha|!$ satisfying* (3.25), (3.28), *and*

$$\sum_k \tilde{q}_k \dot{\lambda}_k^2 < \infty \quad \textit{and} \quad \sum_k \tilde{q}_k (\dot{\lambda}_k^{(r+1)})^2 < \infty, \tag{3.26}$$

Let the weights $\mathcal{R} : r_\alpha^2 = q^\alpha/|\alpha|!$ satisfy

$$\sum_k q_k \lambda_k^2 C_A^2 < \frac{1}{2} \quad \textit{and} \quad \sum_k \frac{q_k}{\tilde{q}_k} < \frac{1}{2}, \tag{3.27}$$

and

$$\sum_k q_k \big(\lambda_k^{(r+1)} C_A^{(r+1)}\big)^2 < 1, \tag{3.28}$$

where $C_A^{(r+1)}$ *is the constant in* $\|w\|_{H^{r+1}} \le C_A^{(r+1)} \|\mathcal{A}^{-1} w\|_{H^{r-1}}$.

Then, we have the estimate

$$\|(\dot{T}_h^{M,n}(t) - \dot{T}(t))F\|_{\mathcal{R}_\Omega H_X^1} \le C_{M,n} h^r \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} + C\|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}).$$

Proof. From the definitions of $T_h^{M,n}(t)$ and $T(t)$,

$$\langle\!\langle \mathcal{L}(t)e, \chi \rangle\!\rangle = 0, \quad \forall \chi \in S^{M,n} \otimes S_h$$

Differentiating both sides w.r.t. t,

$$\langle\!\langle \dot{\mathcal{L}}(t)e + \mathcal{L}(t)e_t, \chi \rangle\!\rangle = 0, \quad \forall \chi \in S^{M,n} \otimes S_h$$

where $\dot{\mathcal{L}}(t) = \dot{\mathcal{A}}(t) + \boldsymbol{\delta}_{\dot{W}}(\mathcal{M}(t))$ is the elliptic operator obtained by differentiating the coefficients of $\mathcal{L}$ w.r.t. t. Consider

$$\begin{aligned}
\langle\!\langle \mathcal{L}e_t, \mathcal{R}^2 e_t \rangle\!\rangle_{\mathcal{R}_\Omega^{\pm 1} H_X^{\mp 1}} &= \sum_\alpha r_\alpha^2 \big\langle \mathcal{A} e_{t,\alpha} + \sum_k \sqrt{\alpha_k} \mathcal{M}_k e_{t,\alpha-\varepsilon_k},\, e_{t,\alpha} \big\rangle_{H_X^{\mp 1}} \\
&\ge \sum_\alpha r_\alpha^2 \left(C_A^{coerc} \|e_{t,\alpha}\|_{H_{0X}^1}^2 - \sum_k \sqrt{\alpha_k} \lambda_k \|e_{t,\alpha-\varepsilon_k}\|_{H_{0X}^1} \|e_{t,\alpha}\|_{H_{0X}^1} \right)
\end{aligned}$$

so

$$\begin{aligned}
C_A^{coerc} \|e_t\|_{\mathcal{R}_\Omega H_{0X}^1}^2 &\le \langle\!\langle \mathcal{L}e_t, \mathcal{R}^2 e_t \rangle\!\rangle_{\mathcal{R}_\Omega^{\pm 1} H_X^{\mp 1}} + \sum_\alpha \sum_k r_\alpha^2 \sqrt{\alpha_k} \lambda_k \|e_{t,\alpha-\epsilon_k}\|_{H_{0X}^1} \|e_{t,\alpha}\|_{H_{0X}^1} \\
&= (I) + (II)
\end{aligned}$$

For Term (I), for any $\chi \in S^{M,n} \otimes S_h$,

$$\begin{aligned}
(I) &= \langle\!\langle \mathcal{L}e_t, \mathcal{R}^2 e_t \rangle\!\rangle + \langle\!\langle \dot{\mathcal{L}}e + \mathcal{L}e_t, \chi \rangle\!\rangle \pm \langle\!\langle \dot{\mathcal{L}}e, \mathcal{R}^2 e_t \rangle\!\rangle \\
&= \langle\!\langle \mathcal{L}e_t, \mathcal{R}^2 e_t + \chi \rangle\!\rangle + \langle\!\langle \dot{\mathcal{L}}e, \mathcal{R}^2 e_t + \chi \rangle\!\rangle - \langle\!\langle \dot{\mathcal{L}}e, \mathcal{R}^2 e_t \rangle\!\rangle \\
&\le \left(\|\mathcal{L}e_t\|_{\mathcal{R}_\Omega H_X^{-1}} + \|\dot{\mathcal{L}}e\|_{\mathcal{R}_\Omega H_X^{-1}} \right) \inf_{\chi \in S^{M,n} \otimes S_h} \|\mathcal{R}^2 e_t + \chi\|_{\mathcal{R}_\Omega^{-1} H_X^1} \\
&\quad + \|\dot{\mathcal{L}}e\|_{\mathcal{R}_\Omega H_X^{-1}} \|\mathcal{R}^2 e_t\|_{\mathcal{R}_\Omega^{-1} H_X^1}.
\end{aligned}$$

For Term (II), by Cauchy-Schwartz and Jensen's inequalities,

$$\begin{aligned}(II) &\leq \left(\sum_\alpha \left(\sum_k r_\alpha \sqrt{\alpha_k}\lambda_k \|e_{t,\alpha-\varepsilon_k}\|_{H^1_{0X}}\right)^2\right)^{1/2} \left(\sum_\alpha r^2_\alpha \|e_{t,\alpha}\|^2_{H^1_{0X}}\right)^{1/2} \\ &\leq \left(\sum_\alpha \left(\sum_{k:\alpha_k\neq 0} \frac{\alpha_k}{|\alpha|}\frac{|\alpha|^2}{\alpha_k}\frac{q^{\alpha-\varepsilon_k}q_k}{(|\alpha|-1)!|\alpha|}\lambda^2_k \|e_{t,\alpha-\varepsilon_k}\|^2_{H^1_{0X}}\right)\right)^{1/2} \|e_t\|_{\mathcal{R}_\Omega H^1_{0X}} \\ &= \left(\sum_\alpha \sum_k \mathbf{1}_{\{\alpha_k\neq 0\}} q_k \lambda^2_k r^2_{\alpha-\varepsilon_k} \|e_{t,\alpha-\varepsilon_k}\|^2_{H^1_{0X}}\right)^{1/2} \|e_t\|_{\mathcal{R}_\Omega H^1_{0X}} \\ &= \left(\sum_k q_k\lambda^2_k\right)^{1/2} \|e_t\|^2_{\mathcal{R}_\Omega H^1_{0X}}.\end{aligned}$$

Combining,

$$\begin{aligned}&\left(C^{coerc}_A - \left(\sum_k q_k\lambda^2_k\right)^{1/2}\right)\|e_t\|^2_{\mathcal{R}_\Omega H^1_{0X}} \\ &\leq \left(\|\mathcal{L}e_t\|_{\mathcal{R}_\Omega H^{-1}_X} + \|\dot{\mathcal{L}}e\|_{\mathcal{R}_\Omega H^{-1}_X}\right) \inf_{\chi\in S^{M,n}\otimes S_h} \|\mathcal{R}^2 e_t + \chi\|_{\mathcal{R}^{-1}_\Omega H^1_X} \\ &\quad + \|\dot{\mathcal{L}}e\|_{\mathcal{R}_\Omega H^{-1}_X}\|\mathcal{R}^2 e_t\|_{\mathcal{R}^{-1}_\Omega H^1_X}.\end{aligned}$$

Since $C^{ellip}_A = (C^{coerc}_A)^{-1}$, and from (3.27), the LHS of the last equation is strictly positive, and we obtain a valid bound for $\|e_t\|^2_{\mathcal{R}_\Omega H^1_{0X}}$. From Lemma 3.2, (3.25) and (3.26), we have $\|\mathcal{L}\chi\|_{\mathcal{R}_\Omega H^{-1}_X} \leq C\|\chi\|_{\mathcal{R}_\Omega H^1_{0X}}$ and $\|\dot{\mathcal{L}}\chi\|_{\mathcal{R}_\Omega H^{-1}_X} \leq C\|\chi\|_{\mathcal{R}_\Omega H^1_{0X}}$, for any $\chi \in \mathcal{R}_\Omega H^1_{0X}$. Then

$$\begin{aligned}\|e_t\|^2_{\mathcal{R}_\Omega H^1_{0X}} \leq C &\left(\|e_t\|_{\mathcal{R}_\Omega H^1_{0X}} + \|e\|_{\mathcal{R}_\Omega H^1_{0X}}\right) \inf_{\chi\in S^{M,n}\otimes S_h} \|e_t - \mathcal{R}^{-2}\chi\|_{\mathcal{R}_\Omega H^1_X} \\ &+ C\|e\|_{\mathcal{R}_\Omega H^1_{0X}}\|e_t\|_{\mathcal{R}_\Omega H^1_X}\end{aligned}$$

In the "inf" term, because any $\chi \in S^{M,n}\otimes S_h$ has only finite non-zero Wiener chaos modes, infimizing $\|e_t - \mathcal{R}^{-2}\chi\|_{\mathcal{R}_\Omega H^1_X}$ over $\chi \in S^{M,n}\otimes S_h$ is equivalent, by a simple rescaling, to infimizing $\|e_t - \tilde{\chi}\|_{\mathcal{R}_\Omega H^1_X}$ over $\tilde{\chi} \in S^{M,n}\otimes S_h$. Thus, upon dividing through by $\|e_t\|_{\mathcal{R}_\Omega H^1_{0X}}$,

$$\begin{aligned}\|e_t\|_{\mathcal{R}_\Omega H^1_{0X}} &\leq C \inf_{\chi\in S^{M,n}\otimes S_h} \|e_t - \chi\|_{\mathcal{R}_\Omega H^1_X} \\ &\quad + C\|e\|_{\mathcal{R}_\Omega H^1_{0X}}\left(1 + \inf_{\chi\in S^{M,n}\otimes S_h} \frac{\|e_t - \chi\|_{\mathcal{R}_\Omega H^1_X}}{\|e_t\|_{\mathcal{R}_\Omega H^1_{0X}}}\right) \\ &\leq C \inf_{\chi\in S^{M,n}\otimes S_h} \|e_t - \chi\|_{\mathcal{R}_\Omega H^1_X} + C\|e\|_{\mathcal{R}_\Omega H^1_{0X}}\end{aligned}$$

since $\inf_{\chi \in S^{M,n} \otimes S_h} \|e_t - \chi\|_{\mathcal{R}_\Omega H_X^1} \le \|e_t\|_{\mathcal{R}_\Omega H_{0X}^1}$. By translation, we have that

$$\|e_t\|_{\mathcal{R}_\Omega H_{0X}^1} \le C \left(\inf_{\chi \in S^{M,n} \otimes S_h} \|\dot{T}(t)F - \chi\|_{\mathcal{R}_\Omega H_X^1} + \|e\|_{\mathcal{R}_\Omega H_{0X}^1} \right).$$

Continuing, the estimation of the "inf" term is as follows. An element $\phi = \sum_\alpha \phi_\alpha \xi_\alpha$ will be decomposed into $\phi = \phi^\sharp + \phi^\perp$, where $\phi^\sharp = \sum_{\alpha \in \mathcal{J}_{M,n}} \phi_\alpha \xi_\alpha$. Then,

$$\begin{aligned}
&\inf_{\chi \in S^{M,n} \otimes S_h} \|\dot{T}(t)F - \chi\|_{\mathcal{R}_\Omega H_X^1} \\
&\qquad\qquad \le \inf_{\chi \in S^{M,n} \otimes S_h} \|(\dot{T}(t)F)^\sharp - \chi\|_{\mathcal{R}_\Omega H_X^1} + \|(\dot{T}(t)F)^\perp\|_{\mathcal{R}_\Omega H_X^1} \\
&\qquad\qquad = (III) + (IV)
\end{aligned}$$

Note that $\dot{U}(t) = \dot{T}(t)F$ solves the equation $\mathcal{L}(t)\dot{U}(t) = -\dot{\mathcal{L}}(t)U(t)$. Since $U(t) \in \widetilde{\mathcal{R}}_\Omega H_{0X}^1$, it follows from Lemma 3.2 that $\dot{\mathcal{L}}(t)U(t) \in \widetilde{\mathcal{R}}_\Omega H_X^{-1}$. So (IV) is the error from Wiener chaos truncation of $\dot{U}$, and by the same proof for Term I_2 in Theorem 3.6, we have that

$$\begin{aligned}
(IV) = \|\dot{U}(t)^\perp\|_{\mathcal{R}_\Omega H_X^1} &\le C Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}) \|\dot{\mathcal{L}}(t)U(t)\|_{\widetilde{\mathcal{R}}_\Omega H_X^{-1}} \\
&\le C Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}) \|U(t)\|_{\widetilde{\mathcal{R}}_\Omega H_{0X}^1} \\
&\le C Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}) \|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}}.
\end{aligned}$$

For (III), we use the same arguments as the proof of Proposition 3.8 to obtain

$$\begin{aligned}
\inf_{\chi \in S^{M,n} \otimes S_h} \|\dot{U}(t)^\sharp - \chi\|_{\mathcal{R}_\Omega H_X^1} &\le C h^r \|\dot{U}(t)^\sharp\|_{\mathcal{R}_\Omega H_X^{r+1}} \\
&\le C h^r \|\dot{\mathcal{L}}(t)U(t)\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r-1}} \\
&\le C h^r \|U(t)\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}}.
\end{aligned}$$

The 2nd inequality follows from the boundedness of $\mathcal{L}^{-1}$ from $\widetilde{\mathcal{R}}_\Omega H_X^{r-1}$ into $\mathcal{R}_\Omega H_X^{r+1}$ ensured by (3.28); the 3rd inequality follows from the boundedness of $\dot{\mathcal{L}}$ from $\widetilde{\mathcal{R}}_\Omega H_X^{r+1}$ into $\widetilde{\mathcal{R}}_\Omega H_X^{r-1}$ ensured by (3.26).

Combining $(III), (IV)$ with the known estimates for $\|e\|_{\mathcal{R}_\Omega H_{0X}^1}$,

$$\begin{aligned}
\|e_t\|_{\mathcal{R}_\Omega H_{0X}^1} \le &\Big(C h^r \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} + C Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}) \|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} \Big) \\
&\quad + \Big(C_{M,n} h^r \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} + C Q_{M,n}(\mathcal{R}, \bar{\mathcal{R}}) \|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} \Big)
\end{aligned}$$

We abuse notation again to write $C_{M,n}$ in place of $C(1 + C_{M,n})$. The result follows by noting that $Q_{M,n}(\mathcal{R}, \bar{\mathcal{R}}) \leq Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}})$. □

Proposition 4.2. *Assume, in addition to the conditions in Proposition 4.1, that*

$$\sum_k q_k(\lambda_k^{(3)} C_A^{(3)})^2 < \frac{1}{2} \quad \text{and} \quad \sum_k q_k(\dot{\lambda}_k^{*,(3)})^2 < \infty. \tag{3.29}$$

Then

$$\|(\dot{T}_h^{M,n}(t) - \dot{T}(t))F\|_{\mathcal{R}_\Omega H_X^{-1}} \leq C_{M,n} h^{r+2} \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} + C\|F\|_{\bar{\mathcal{R}}_\Omega H_X^{-1}} Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}).$$

Proof. From the proof of the $\|e_t\|_{\mathcal{R}_\Omega H_{0X}^1}$ estimate, it is clear that the first term with h^r is due to the norm from $\|e_t^\sharp\|_{S^{M,n} \otimes H_{0X}^1}$, whereas the second term is due to the norm from $\|e_t^\perp\|_{(S^{M,n})^\perp \otimes H_{0X}^1}$. As usual, we leave the second term untouched, and consider only the first term.

We want to estimate $\|e_t^\sharp\|_{\mathcal{R}_\Omega H_X^{-1}}$. For any $\phi \in S^{M,n} \otimes H_{0X}^1$, let $\psi = \psi(t) \in S^{M,n} \otimes H_X^3$ be the solution of $\mathcal{L}^*(t)\psi = \phi$.

Since $\langle\langle \mathcal{L}e^\sharp, \chi \rangle\rangle = 0$ for all $\chi \in S^{M,n} \otimes S_h$, differentiating w.r.t. t gives that $\langle\langle \dot{\mathcal{L}}e^\sharp, \chi \rangle\rangle + \langle\langle \mathcal{L}e_t^\sharp, \chi \rangle\rangle = 0$. Hence

$$\begin{aligned}
\langle\langle e_t^\sharp, \phi \rangle\rangle &= \langle\langle e_t^\sharp, \mathcal{L}^*\psi \rangle\rangle = \langle\langle \mathcal{L}e_t^\sharp, \psi \rangle\rangle \\
&= \langle\langle \mathcal{L}e_t^\sharp + \dot{\mathcal{L}}e^\sharp, \psi + \chi \rangle\rangle - \langle\langle \dot{\mathcal{L}}e^\sharp, \psi \rangle\rangle =: (V) - (VI)
\end{aligned}$$

for all $\chi \in S^{M,n} \otimes S_h$.

For Term (V),

$$\begin{aligned}
|(V)| &\leq \|\mathcal{L}e_t^\sharp + \dot{\mathcal{L}}e^\sharp\|_{\mathcal{R}_\Omega H_X^{-1}} \inf_{\chi \in S^{M,n} \otimes S_h} \|\psi - \chi\|_{\mathcal{R}_\Omega^{-1} H_X^1} \\
&\leq C \left(\|e_t^\sharp\|_{\mathcal{R}_\Omega H_{0X}^1} + \|e^\sharp\|_{\mathcal{R}_\Omega H_{0X}^1} \right) h^2 \|\phi\|_{\mathcal{R}_\Omega^{-1} H_X^1} \\
&\leq C_{M,n} h^{r+2} \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} \|\phi\|_{\mathcal{R}_\Omega^{-1} H_X^1}.
\end{aligned}$$

For Term (VI), notice that $\dot{\mathcal{L}}^*\psi \in \mathcal{R}_\Omega^{-1} H_X^1$, so

$$|(VI)| = |\langle\langle e^\sharp, \dot{\mathcal{L}}^*\psi \rangle\rangle_{\mathcal{R}_\Omega^{\pm 1} H_X^{\pm 1}}| = |\langle\langle e^\sharp, \dot{\mathcal{L}}^*\psi \rangle\rangle_{\mathcal{R}_\Omega^{\pm 1} H_X^{\mp 1}}|$$

$$
\begin{aligned}
&\leq \|e^\sharp\|_{\mathcal{R}_\Omega H_X^{-1}} \|\dot{\mathcal{L}}^* \psi\|_{\mathcal{R}_\Omega^{-1} H_X^1} \leq C \|e^\sharp\|_{\mathcal{R}_\Omega H_X^{-1}} \|\psi\|_{\mathcal{R}_\Omega^{-1} H_X^3} \\
&\leq C_{M,n} h^{r+2} \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}} \|\phi\|_{\mathcal{R}_\Omega^{-1} H_X^1}.
\end{aligned}
$$

The penultimate inequality holds due to the second inequality in (3.29).

Combining,

$$
\|e_t^\sharp\|_{\mathcal{R} H_X^{-1}} = \sup_{\phi \in S^{M,n} \otimes H_{0X}^1} \frac{|\langle\langle e_t^\sharp, \phi \rangle\rangle|}{\|\phi\|_{\mathcal{R}_\Omega^{-1} H_X^1}} \leq C_{M,n} h^{r+2} \|U\|_{\widetilde{\mathcal{R}}_\Omega H_X^{r+1}}.
$$

□

4.2. The parabolic problem with time-dependent operators. We are now in the position to prove the parabolic estimates.

THEOREM 4.3. *Let $u \in \mathcal{R}'_\Omega L_T^2 H_{0X}^1 \cap \mathcal{R}'_\Omega L_T^2 H_X^{m+2}$ be the solution to the stochastic parabolic equation, and assume that the conditions in Theorem 3.9 hold. For the weights $\mathcal{R}$, assume that* (3.27) *and* (3.29) *hold, where $\widetilde{\mathcal{R}} : \tilde{r}_\alpha^2 = \tilde{q}^\alpha / |\alpha|!$ satisfies* (3.26) *and*

$$
\sum_k \frac{\tilde{q}_k}{q'_k} < \frac{1}{2}.
$$

Then we have the estimates

$$
\begin{aligned}
&\|u_h^{M,n} - u\|_{\mathcal{R}_\Omega L_X^2} \\
&\leq C_{M,n} h^{m+1} \left(\|f\|_{\bar{\mathcal{R}}_\Omega L_T^2 H_X^m} + \|f_t\|_{\bar{\mathcal{R}}_\Omega L_T^2 H_X^{m-2}} + \|v\|_{\bar{\mathcal{R}}_\Omega H_X^{m+1}} \right) \\
&\quad + C Q_{M,n}(\mathcal{R}, \mathcal{R}') \Big(\|f - u_t\|_{\mathcal{R}'_\Omega L_T^2 H_X^{-1}} + \|f_t - u_{tt}\|_{\mathcal{R}'_\Omega L_T^2 H_X^{-1}} + \|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H_X^{-1}} \Big).
\end{aligned}
$$

PROOF. We set

$$
\begin{aligned}
e_h(t) &:= u_h^{M,n}(t) - u(t) \\
&= \Big(u_h^{M,n}(t) - T_h^{M,n}(t)(f(t) - u_t(t)) \Big) + \Big((T_h^{M,n}(t) - T(t))(f(t) - u_t(t)) \Big) \\
&= \theta(t) + \pi(t)
\end{aligned}
$$

Up to the point of equation (3.23), the proof of Theorem 3.9 is followed identically to yield estimates for θ, π. Thus,

$$\begin{aligned}\|\pi(t)\|_{\mathcal{R}_\Omega L^2_X} &= \|(T_h^{M,n}(t) - T(t))(f - u_t(t))\|_{\mathcal{R}_\Omega L^2_X} \\ &\leq C_{M,n} h^{m+1} \|u(t)\|_{\mathcal{R}_\Omega H^{m+1}_X} + C\|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H^{-1}_X} Q_{M,n}(\mathcal{R}, \mathcal{R}')\end{aligned}$$

and

$$\|\theta(t)\|^2_{\mathcal{R}_\Omega L^2_X} \leq \|\theta(0)\|^2_{\mathcal{R}_\Omega L^2_X} + 2\epsilon_0 \int_0^t \|(\Pi_h^{M,n} u - u)_t(s)\|^2_{\mathcal{R}_\Omega H^{-1}_X} ds.$$

To estimate $\|(\Pi_h^{M,n} u - u)_t(s)\|_{\mathcal{R}_\Omega H^{-1}_X}$,

$$\begin{aligned}(\Pi_h^{M,n} u - u)_t(t) &= \frac{\mathrm{d}}{\mathrm{d}t}(T_h^{M,n}(t) - T(t))(f(t) - u_t(t)) \\ &= (\dot{T}_h^{M,n}(t) - \dot{T}(t))(f(t) - u_t(t)) + (T_h^{M,n}(t) - T(t))(f_t(t) - u_{tt}(t))\end{aligned}$$

So from Proposition 4.2,

$$\begin{aligned}&\|(\Pi_h^{M,n} u - u)_t(t)\|_{\mathcal{R}_\Omega H^{-1}_X} \\ &\leq C_{M,n} h^{m+3} \|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H^m_X} + C Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}) \|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H^{-1}_X} \\ &\quad + C_{M,n} h^{m+1} \|f_t(t) - u_{tt}(t)\|_{\mathcal{R}'_\Omega H^{m-2}_X} + C Q_{M,n}(\mathcal{R}, \widetilde{\mathcal{R}}) \|f_t(t) - u_{tt}(t)\|_{\mathcal{R}'_\Omega H^{-1}_X}\end{aligned}$$

Putting together the estimates for $\theta(t) = u_h^{M,n}(t) - \Pi_h^{M,n} u(t)$ and $\pi(t) = \Pi_h^{M,n} u(t) - u(t)$,

$$\begin{aligned}&\|u_h^{M,n} - u\|_{\mathcal{R}_\Omega L^2_X} \\ &\leq C_{M,n} h^{m+1} \left(h^2 \|u\|_{\mathcal{R}'_\Omega L^2_T H^{m+2}_X} + \|u_t\|_{\mathcal{R}'_\Omega L^2_T H^m_X} + \|u(t)\|_{\mathcal{R}'_\Omega H^{m+1}_X} \right) \\ &+ C Q_{M,n}(\mathcal{R}, \mathcal{R}') \Big(\|f - u_t\|_{\mathcal{R}'_\Omega L^2_T H^{-1}_X} + \|f_t - u_{tt}\|_{\mathcal{R}'_\Omega L^2_T H^{-1}_X} + \|f(t) - u_t(t)\|_{\mathcal{R}'_\Omega H^{-1}_X} \Big)\end{aligned}$$

as desired. □

5. Numerical Simulations

We perform numerical simulations in order to test the error estimates (3.21). There are several aspects of the error estimates that are worth investigation.

(1) The spectral convergence in n, coming from the second term in $Q_{M,n}$. This spectral convergence has been shown in [**65**] for the corresponding elliptic equation.

(2) The factor $C_{M,n}$. Theorem 3.6 gives an upper bound for $C_{M,n} = C'\binom{M+n}{M}$. However, a tighter upper bound has been conjectured[1] that for $C_{M,n}$ that is independent of M, n.

(3) The optimality of the relationship between the convergence order h^{m+1} and the weighted space $\mathcal{R}_\Omega H_X^{m+1}$. How critical is the stochastic weights in the order of convergence?

5.1. A (Simple) Model Problem. To investigate the above questions, we simulate the 1D equation

$$\begin{aligned} &\tfrac{\partial u}{\partial t} = \Delta u + \boldsymbol{\delta}_{\dot{W}}(\Delta u) + f, \quad \text{for } x \in (0,\pi),\, t \in (0,T] \\ &u(0,t) = u(\pi,t) = 0, \quad u(x,0) = v(x) \end{aligned} \tag{3.30}$$

where $\dot{W}(x) = \sum_k \mathfrak{u}_k(x)\xi_k$ is a spatial Gaussian white noise on $L_2(0,\pi)$. For the CONS in $L_2(0,\pi)$, we take

$$\mathfrak{u}_1(x) = \sqrt{\frac{1}{\pi}},$$
$$\mathfrak{u}_k(x) = \sqrt{\frac{2}{\pi}}\cos(k-1)\pi.$$

In the notation of (3.1), $\mathcal{A} = -\Delta$ and $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \dots)$, where $\mathcal{M}_k u = (\mathfrak{u}_k(x)u_x(x))_x$. Then the propagator system is

$$\begin{aligned} &\frac{\partial u_\alpha}{\partial t} = \Delta u_\alpha + \sum_k \sqrt{\alpha_k}\mathcal{M}_k u_{\alpha-\varepsilon_k}, \quad \text{for } x \in (0,\pi),\, t \in (0,T] \\ &u_\alpha(0,t) = u_\alpha(\pi,t) = 0, \quad u_\alpha(x,0) = v_\alpha(x). \end{aligned}$$

The relevant constants are: $C_A^{ellip} = (1 + C^{poinc}) = (1+\pi)$; the constants $\lambda_k^{(s)}$ in $\|\mathcal{M}_k u\|_{H^{s-2}} \le \lambda_k^{(s)}\|u\|_{H^s}$ are

$$\begin{aligned} \lambda_1 &= \sqrt{\tfrac{1}{X}}, & \lambda_1^{(s)} &\sim k^0 \\ \lambda_k &= \sqrt{\tfrac{2}{X}}, & \lambda_k^{(s)} &\sim k^{s-1} \end{aligned}$$

[1] The author is grateful to Zhongqiang Zhang for bringing this conjecture to her attention.

For the numerical simulations, we solve the SPDE on the interval $(0, \pi)$ up to final time $T = 0.1$. We fabricate the solution of the SPDE, by fixing the solution to be

$$\begin{aligned} u_{(0)}(x,t) &= 0 \\ u_\alpha(x,t) &= \frac{|\alpha|!}{\sqrt{\alpha!}} \sum_{\substack{k\geq 1 \\ \alpha_k \neq 0}} \sin(kx)(e^{-k^2 t} - 1) \end{aligned}$$

This fabricated solution u is rather spatially rough, but this is all the better for demonstrating some important features of the error estimates. Based on the fabricated solution, we reverse engineer the input data of the equation; that is, the solution u is obtained with zero initial conditions and with forcing term

$$f_\alpha = -\frac{|\alpha|!}{\sqrt{\alpha!}} \sum_{\substack{k\geq 1 \\ \alpha_k \neq 0}} k^2 \sin(kx) - \frac{|\alpha|!}{\sqrt{\alpha!}} \sum_{\substack{k\geq 1 \\ \alpha_k \neq 0}} \sum_{\substack{l\geq 1 \\ \alpha_l - \delta_{kl} \neq 0}} \frac{\alpha_k}{|\alpha|}(e^{-k^2 t} - 1)(\mathfrak{u}_k(x)(\sin(lx))_x)_x$$

For the weighted space $\mathcal{R}L_2(\Omega; H^s)$ with weights $r_\alpha^2 = q^\alpha/|\alpha|!$, and for any $s \in \mathbb{N}_0$,

$$\|u\|_{\mathcal{R}_\Omega H_X^s} \qquad \text{provided} \quad q_k \sim k^{-(2s+1+\delta)}. \tag{3.31}$$

Indeed,

$$\begin{aligned} \sum_{\alpha\in\mathcal{J}} \frac{q^\alpha}{|\alpha|!}\|u_\alpha\|_{H^s}^2 &= \sum_{n=0}^{\infty} \sum_{|\alpha|=n} \sum_{k\geq 1} \mathbf{1}_{\{\alpha_k\neq 0\}} q^\alpha \frac{|\alpha|!}{\alpha!} \|(e^{-k^2 t} - 1)\sin(kx)\|_{H^s}^2 \\ &= \sum_{n=0}^{\infty} \sum_{k\geq 1} (e^{-k^2 t} - 1)^2 \|\sin(kx)\|_{H^s}^2 \left(\sum_{|\alpha|=n} q^\alpha \frac{|\alpha|!}{\alpha!} - \sum_{|\alpha|=n} \mathbf{1}_{\{\alpha_k=0\}} q^\alpha \frac{|\alpha|!}{\alpha!} \right) \\ &= \sum_{n=0}^{\infty} \sum_{k\geq 1} (e^{-k^2 t} - 1)^2 \|\sin(kx)\|_{H^s}^2 \left(\bar{q}^n - (\bar{q} - q_k)^n\right) \end{aligned}$$

where $\bar{q} = \sum_{k\geq 1} q_k$. Since $\bar{q}^n - (\bar{q} - q_k)^n \leq n\bar{q}^{n-1} q_k$ by the mean value theorem,

$$\|u\|_{\mathcal{R}_\Omega H_X^s}^2 \leq \sum_{n=0}^{\infty} n\bar{q}^{n-1} \sum_{k\geq 1} (e^{-k^2 t} - 1)^2 \frac{C}{\pi} k^{2s} q_k$$

Similarly, we find that $u_t \in \mathcal{R}_\Omega H_X^{s-2}$, and it follows by Lemma 3.2 that $f \in \mathcal{R}_\Omega H_X^{s-2}$.

The weights. We consider $u(t)$ belonging to $\mathcal{R}'_\Omega L_T^2 H_X^3$. For concreteness' sake, we take the weights $(r'_\alpha)^2 = \rho^\alpha/|\alpha|!$, with $\rho_k = \hat{\rho} k^{-8}$ and $\hat{\rho} = 0.5 < \sum_k k^{-8}$. Then according to Theorem 3.9, we measure the error in the norm $\mathcal{R}_\Omega L^2_X$ with weights $r^2_\alpha = q^\alpha/|\alpha|!$, and $q_k = \hat{q}k^{-10}$ satisfying the two conditions (3.22),

$$\sum_{k\geq 1} \hat{q}k^{-10}(\lambda_k C_A^{ellip})^2 < \frac{1}{2}, \quad \text{and} \quad \sum_{k\geq 1} \frac{\hat{q}k^{-10}}{\hat{\rho}k^{-9}} < \frac{1}{2}.$$

So, we take

$$\hat{q} = 0.01 < \min\{\frac{\hat{\rho}}{2}\big(\sum_k k^{-2}\big)^{-1}, \frac{\pi}{4(C_A^{ellip})^2}\big(\sum_k k^{-10}\big)^{-1}\}.$$

Code Specifications. For the finite element discretization, we implemented the cG(2) method. The domain $(0, \pi)$ is partitioned into $M + 1$ uniform subintervals of length $h := \pi/(M + 1)$. The subintervals are labelled $I_i = [x_{i-1}, x_i]$ for $i = 1, \dots M + 1$, where the grid points are $x_i = ih$ for $i = 0, \dots M + 1$. The nodal basis $\Phi_l(x) = \phi_{l/2}(x)$ is given by

$$\phi_{i-1/2}(x) = \mathbf{1}_{I_i}(x)4h^{-2}(x_i - x)(x - x_{i-1})$$
$$\phi_i(x) = \mathbf{1}_{I_i}(x)2h^{-2}(x - x_{i-1/2})(x - x_{i-1}) + \mathbf{1}_{I_{i+1}}(x)2h^{-2}(x - x_{i+1/2})(x - x_{i+1})$$

for $l = 1, \dots, 2M+1$. The mass, stiffness and noise matrices are symmetric 5-banded sparse matrices given by $\mathbb{M}^{mass}_{l,l'} = (\Phi_l, \Phi_{l'})$, and $\mathbb{M}^{stiff}_{l,l'} = (\Phi'_l, \Phi'_{l'})$, and $\mathbb{M}^{noise}_{k;l,l'} = (\mathfrak{u}_k\Phi'_l, \Phi'_{l'})$.

Time stepping is implemented via a 2nd order Runge Kutta method

$$u^{(1)} = u^n + \Delta t L(u^n)$$
$$u^{n+1} = \frac{1}{2}\big(u^n + u^{(1)} + \Delta t L(u^{(1)})\big).$$

We took $\Delta t/h^2 = 0.03$. Then the maximal order of convergence in each deterministic equation is h^3.

Test of Error Estimate. The basic error estimate has 3 terms[2]

$$C\binom{N+p}{p}h^{m+1}\|u\|_{\mathcal{R}_\Omega H^{m+1}_X} + C\hat{Q}_N^{1/2} + C\hat{Q}^{(p+1)/2}.$$

with $m = 2$. In the second term, $\hat{Q}_N = \sum_{k\geq N+1} q_k(\lambda_k C_A)^2 + \frac{q_k}{\tilde{q}_k} \sim N^{-1}$, by our choice of $q_k \sim k^{-10}$.

To investigate the relationship between N and h, in the first term we couple h and N by setting $h = N^{-p'}$ for p' chosen in the following way. We estimate $\binom{N+p}{p} = \frac{(N+p)\dots(N+1)}{p!} \sim$

[2] The notation of some parameters has been reshuffled. The finite stochastic subspace is now $\mathcal{J}_{N,p}$.

Error (and order of convergence) from finite element approximation

	$N=4$	8	12	16	20
$p=2$	1.1609e-11	6.0927e-13 (4.25)	8.5121e-14 (4.85)	2.6567e-14 (4.05)	9.9758e-15 (4.39)
$p=3$	7.0086e-13	9.4364e-15 (6.21)	6.715e-16 (6.52)	1.0689e-16 (6.39)	2.3028e-17 (6.88)
$p=4$	3.8629e-14	1.4847e-16 (8.02)	4.8386e-18 (8.44)	4.114e-19 (8.57)	5.7714e-20 (8.80)

TABLE 1. Absolute errors (and convergence orders) from the finite element part under the weights $q_k \sim k^{-10}$. (Values are squared of the error norm.)

Error from white noise truncation (fixed p)

	$N=4$	8	12	16	20
$p=2$	1.1903e-10	2.6821e-12	1.7075e-13	1.7931e-14	2.7177e-15
$p=3$	1.1906e-10	2.6828e-12	1.708e-13	1.7936e-14	2.7185e-15
$p=4$	1.1906e-10	2.6829e-12	1.708e-13	1.7936e-14	2.7185e-15
Order		(5.47)	(6.79)	(7.83)	(8.46)

TABLE 2. Truncation error (and convergence order) for each fixed p. (Values are squared of the error norm.)

N^p, and match the powers of N in the first two terms, $\binom{N+p}{p}h^m \sim N^{p-mp'}$ v.s. $N^{-1/2}$. So, we choose $p' = \frac{1+2p}{2(m+1)}$. Then, the error estimate reduces to 2 terms

$$CN^{-1/2} + C\hat{Q}^{(p+1)/2}.$$

In this way, we expect the convergence to be order $-1/2$ in N, and spectral in p.

ERROR RESULTS. We set $h = 2N^{-p'}$, where $p' = \frac{2p+1}{2(m+1)}$ with $m = 2$. We vary $p = 2, 3, 4$ and $N = 4, 8, 12, 16, 20$. Clearly, the mesh size h varies with both N and p.

For the fabricated solution, it is fairly straightforward to compute the exact truncation error from the stochastic truncation. Thus, we focus our results mostly on the finite element error from the $\mathcal{J}_{N,p}$ part.

The tables 1 and 2 show the errors from the finite element approximation of the modes in $\mathcal{J}_{N,p}$ and from the truncation of the white noise to N dimensions, up to polynomial order p. The truncation error for fixed p is the error from the terms $\sum_{\substack{\alpha\notin\mathcal{J}_{N,p}\\ |\alpha|\le p}} r_\alpha^2 \|u_\alpha\|^2_{L^2_X}$. The weights were taken to be $q_k \sim k^{-10}$. Notice that in some cases, both errors are comparable, while other times the finite element error is insignificant compared to the truncation error. For the error from the $J_{N,p}$ part, the orders of convergence is much higher than the $\mathcal{O}(N^{-1/2})$ that we tried to match. The order is approximately $5/2, 7/2$ and $9/2$ for $p = 2, 3, 4$ respectively,

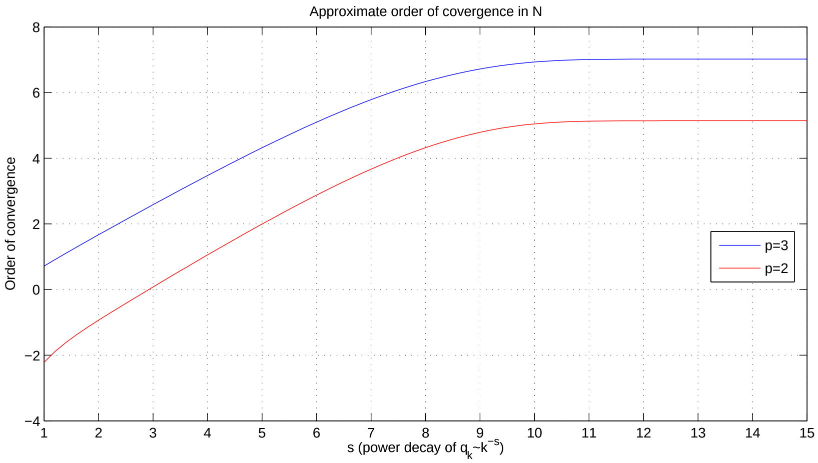

FIGURE 1. Order of convergence (in N) for the $\mathcal{J}_{N,p}$ part, as the power decay of the weights $q_k \sim k^{-s}$ varies. (Orders are computed for the square of the error norm.)

which is exactly N^p higher than expected. This over-correction calls into question whether the term $\binom{N+p}{p} \sim N^p$ is indeed present in the error estimate. Calculating backwards, we might conjecture the term $N^\tau h^{m+1}$ with $\tau = 0$ instead. This conjecture warrants further investigation. Additionally, we note that the order of convergence in the truncation of white noise is order N^{-9} instead of N^{-1}. This is due to the fact that we are estimating the error in the $\mathcal{R}_\Omega L^2_X$ norm, with the weights required for u to possess higher spatial regularity, $\mathcal{R}_\Omega H^3_X$, and which are worse than the weights required for u to be merely L^2_X in space.

Next, we investigate the relationship between the stochastic weights and the order of convergence in the $\mathcal{J}_{N,p}$ part. For the stochastic weights $\mathcal{R}^{(s)}$ characterized by the decay $q_k \sim k^{-s}$, (3.31) relates the power s to the spatial regularity in $H^{s'}$ of the solution in $\mathcal{R}^{(s)}_\Omega H^{s'}_X$, and hence also to the convergence order h^{m+1}. Figure 1 shows the approximate order of convergence in N as the power s varies, also with the coupling $h = 2N^{-p'}$ for p' found above. Two features are observed. First, the order of convergence is higher for $p = 3$ than for $p = 2$, possibly an artifact of the over-correction by the $\binom{N+p}{p}$ term. Second, for $s \lesssim 8$, the order of convergence is almost linear with slope ≈ 1, whereas for $s \gtrsim 10$, the order of convergence plateaus off. The plateau is due to the convergence order being limited by the h^3 from the cG(2) implementation. In view of (3.31), the linear portion of the graph corroborates qualitatively the trade-off between the stochastic weights on the one hand, and the spatial smoothness and order of convergence on the other.

In conclusion, we have seen qualitatively that the finite element order of convergence does depend on the spatial smoothness of the stochastically weighted solution, but further work remains to be done to investigate (the absence of) the $\binom{N+p}{p}$ factor.

CHAPTER 4

Unbiased perturbations of the Navier-Stokes equations

Stochastic perturbations of the Navier-Stokes equation have received much attention over the past few decades. Among the early studies of the stochastic Navier-Stokes equations are those by Bensoussan and Temam [**6**], Foias et al. [**17**–**19**], Flandoli [**15**, **16**], etc. Traditionally, the types of perturbations that were proposed includes stochastic forcing by a noise term such as a Gaussian random field or a cylindrical Wiener process, and are broadly accepted as a natural way to incorporate stochastic effects into the system. The stochastic Navier-Stokes equation is underpinned by a familiar physical basis, because it can be derived from Newton's Second Law via the the fluid flow map, using a particular assumption on the stochasticity of the governing SODE of the flow map, known as the Kraichnan turbulence. (See [**53**, **54**] and the references therein.) However, due to the nonlinearity in the equations, the stochastic Navier-Stokes equation leads to a biased perturbation; that is, the mean solution of the stochastic equation does not coincide with the solution of the unperturbed equation. In fact, the mean solution and unperturbed solution can differ quite drastically, an observation that is also true for other nonlinear equations such as the stochastic Burgers equation.

To derive a model for an unbiased perturbation of the Navier-Stokes equation, the nonlinear term must be modified. In [**55**], a quantized stochastic Navier-Stokes equation has been proposed as an unbiased perturbation. The quantized equation replaces the usual product in the nonlinear term with the Wick product, thereby turning the nonlinear term into a stochastic convolution. The replacement with the Wick product preserves the mean because of the identity

$$\mathbb{E}[u \diamond v] = \mathbb{E}u \, \mathbb{E}v.$$

Thus, this perturbation is unbiased in the sense that the mean $\mathbb{E}u$ of the solution satisfies the unperturbed Navier-Stokes equation.

Apart from the interpretation of being an unbiased perturbation, the quantized stochastic Navier-Stokes equation also has a physical derivation based on Newton's Second Law. This derivation likewise representation of the fluid flow map, but differs from the aforementioned derivation by the stochasticity assumption in the governing equation of the fluid flow map. In the quantized case, it can be shown that the flow map $\Phi(t,x)$ satisfies

$$\frac{d\Phi(t)}{dt} = u^{\diamond}(t, \Phi(t))$$

for smooth functions u, where the function $u^{\diamond}$ is the *Wick version* of u. However, we will not delve into the study of the fluid flow map here.

Thus, we will consider the quantized stochastic Navier-Stokes equation on an open bounded domain $D \in \mathbb{R}^d$, $d = 2, 3$, driven by purely spatial noise,

$$\begin{aligned}
&u_t + u^i \diamond u_{x_i} + \nabla P^f = \nu \Delta u + f(t,x) + \big(\sigma^i(t,x) u_{x_i} + \nabla P^g + g(t,x)\big) \diamond \dot{W}(x),\\
&\operatorname{div} u \equiv 0,\\
&u(0,x) = w(x), \quad u|_{\partial D} = 0.
\end{aligned}\tag{4.1}$$

where the diffusivity constant is $\nu > 0$, and the functions f, g, σ are given deterministic $\mathbb{R}^d$-valued functions. Here, the driving noise $\dot{W}(x) = \sum_k \mathfrak{u}_l(x)\xi_l$ is a stationary Gaussian white noise on $L_2(D)$, and we assume that $\sup_l \|\mathfrak{u}_l\|_{L^\infty} < \infty$. Since we restrict the forcing term to be a stationary noise, it is natural to study the related steady solution of the stationary Navier-Stokes equation,

$$\begin{aligned}
&\bar{u}^i \diamond \bar{u}_{x_i} + \nabla \bar{P}^f = \nu \Delta \bar{u} + \bar{f}(x) + \big(\bar{\sigma}^i(x) \bar{u}_{x_i} + \nabla \bar{P}^g(x) + \bar{g}(x)\big) \diamond \dot{W}(x)\\
&\operatorname{div} \bar{u} \equiv 0\\
&\bar{u}|_{\partial D} = 0.
\end{aligned}\tag{4.2}$$

where $\bar{f}(x), \bar{g}(x), \bar{\sigma}(x)$ are given deterministic $\mathbb{R}^d$-valued functions.

The analysis of the quantized Navier-Stokes equation relies on studying the propagator system. The propagator system is a lower triangular system, thus the analysis is amenable to the same induction procedure used in the earlier chapters. As a side note, we remark that in comparison, the usual stochastic Navier-Stokes equation has a propagator system that is a full system of equations which, comparatively, are a much tougher beast to tackle using the Wiener chaos expansion. Additionally, apart from the zero-th chaos mode which, being the mean, solves the deterministic Navier-Stokes equation, all higher modes in the propagator system solves a linearized Stokes equation. Thus, where a result is known for the deterministic Navier-Stokes equation, it is sometimes the case that an analogous result may be shown for the quantized equation. For instance, the existence of a unique stationary solution of (4.2) requires the same condition on the largeness of the viscosity ν as does the existence of a unique steady solution of the deterministic equation (4.8a).

There is substantial theory on the steady solutions of the deterministic Stokes and Navier-Stokes equations, the long time convergence of a time-dependent solution to the steady solution, as well as other dynamical behaviour of the solution. In the subsequent sections, we begin to study some of these same questions for the quantized Navier-Stokes equation, focusing on the large viscosity case where the uniqueness of steady solutions and long time convergence has been established in the deterministic setting. We will study the existence of a unique stationary solution of (4.2) as well as the existence of a unique time-dependent solution of (4.1) on a finite time interval. The Wiener chaos expansion and the propagator system will be the central tool in obtaining a generalized solution, but to place the solution in a Kondratiev space involves a useful result invoking the Catalan numbers. The Catalan numbers arises naturally from the convolution of the Wiener chaos modes in the nonlinear term. It was used to study the Wick version of the Burgers equation [**34**]. The long time convergence of a time-dependent solution to a steady state solution is also presented, though the theory here is not complete—the convergence in the generalized sense and in a Kondratiev space are shown separately with differing sets of assumptions. Continuing work is done to reconcile the disparity.

1. Functional analysis framework

To study equations (4.1) and (4.2), we adopt the variational/weak formulation in [**59**, **60**]. Denote the following spaces

$$\mathcal{V} := \{v \in C_0^\infty(D)^n : \operatorname{div} v = 0\}$$

$$V := \text{closure of } \mathcal{V} \text{ in the } H_0^1(D) \text{ norm } \equiv \{u \in H_0^1(D) : \operatorname{div} u = 0\}$$

$$H := \text{closure of } \mathcal{V} \text{ in the } L^2(D) \text{ norm } |\cdot|$$

$$V' := \text{dual space of } V \text{ w.r.t. inner product in } H$$

Also denote the norms in V and V' by $\|w\|_V = |\nabla w|$ and $\|f\|_{V'}$, respectively.

The operator[1] $-\Delta$ on H, defined on the domain $\mathrm{dom}(-\Delta)$, is symmetric positive definite and thus defines a norm via $|\Delta w|$, which is equivalent to the norm $\|w\|_{H^2}$. For $m \in \mathbb{R}$, define the norms $|w|_m = |(-\Delta)^{m/2} w|$ on the closed subspace

$$V_m := \mathrm{dom}((-\Delta)^{m/2}) = \{v \in H^m : \mathrm{div}\, v = 0\}$$

The norms $|w|_m$ and $\|w\|_{H^m}$ are equivalent. We have a constant c_1 so that

$$c_1 \| \cdot \|_{H^1}^2 \leq |w|_1^2 \leq \frac{1}{c_1} \| \cdot \|_{H^1}^2.$$

Note that $|w|_1 = \|w\|_V$. Denote $\lambda_1 > 0$ to be the smallest eigenvalue of $-\Delta$, then we have a Poincare inequality,

$$\lambda_1 |v|^2 \leq \|v\|_V^2, \qquad \text{for } v \in V. \tag{4.3}$$

Define the trilinear continuous form b on $V \times V \times V$ by

$$b(u, v, w) = \int_D u^k \partial_{x_k} v^j w^j \, dx,$$

and the mapping $B : V \times V \to V'$ by

$$\langle B(u, v), w \rangle = b(u, v, w).$$

It is easy to check that

$$b(u, v, w) = -b(u, w, v), \quad \text{and} \quad b(u, v, v) = 0$$

for all $u, v, w \in V$. B and b have many useful properties that follow from the following lemma.

Lemma 1.1 (Lemma 2.1 in [**59**]). *The form b is defined and is trilinear continuous on $H^{m_1} \times H^{m_2+1} \times H^{m_3}$, where $m_i \geq 0$ and*

$$\begin{aligned} &m_1 + m_2 + m_3 \geq \tfrac{d}{2} \quad \text{if } m_i \neq \tfrac{d}{2},\ i = 1, 2, 3, \\ &m_1 + m_2 + m_3 > \tfrac{d}{2} \quad \text{if } m_i = \tfrac{d}{2},\ \text{some } i. \end{aligned} \tag{4.4}$$

[1] Technically, the correct operator is $Au := -P\Delta u$, where P is the orthogonal projection onto H. We abuse notation here and continue writing $-\Delta$.

In view of Lemma 1.1, let c_b be the constant in

$$|b(u,v,w)| \le c_b |u|_{m_1} |v|_{m_2+1} |w|_{m_3}$$

where m_i satisfies (4.4). Also let c_d, $d = 2, 3$, be the constants in

$$\begin{aligned} &|b(u,v,w)| \le c_2 |u|^{1/2} \|u\|_V^{1/2} \|v\|_V^{1/2} |\Delta v|^{1/2} |w| && \text{if } d = 2 \\ &|b(u,v,w)| \le c_3 \|u\|_V \|v\|_V^{1/2} |\Delta v|^{1/2} |w| && \text{if } d = 3 \end{aligned}$$

for all $u \in V$, $v \in \operatorname{dom}(-\Delta)$, and $w \in H$ (equations (2.31-32) in [**59**]). Other useful consequences of Lemma 1.1 is that $B(\cdot,\cdot)$ is a bilinear continuous operator from $V \times H^2 \to L^2$, and also from $H^2 \times V \to L^2$.

In order to define the weak solution of (4.1), we recall that for a smooth function p, $(\nabla p, v) = 0$ for all $v \in V$. This leads us to define the weak solution by taking the test function space V, so that the pressure term drops out.

Definition 1.2. *A generalized weak solution of* (4.1) *is a generalized random element* $u \in \mathcal{D}'(L^2(0,T;H_0^1(D)))$ *such that*

$$\langle\!\langle u_t + u^i \diamond u_{x_i},\, \phi \rangle\!\rangle = \langle\!\langle \nu \Delta u + f(t,x) + \big(\sigma^i(t,x) u_{x_i} + \nabla P^g + g(t,x)\big) \diamond \dot{W}(x),\, \phi \rangle\!\rangle \tag{4.5}$$

for all test functions $\phi \in \mathcal{D}(V)$.

The pressure terms. The pressure terms P^f, P^g are determined from the velocity field u. The term P^g is defined so that the white noise multiplies a divergence-free term. Specifically, define P^g to be the solution of

$$-\Delta P^g = u^j_{x_k} \sigma^k_{x_j} + \operatorname{div} g, \tag{4.6}$$

and set $\beta := \nabla P^g + u_{x_k} \sigma^k + g$. Clearly, $\operatorname{div} \beta = 0$ as desired. Then, from the equation, P^f solves

$$-\Delta P^f = u^j_{x_k} \diamond u^k_{x_j} - \operatorname{div} f - \beta \diamond (\nabla \dot{W}(x)).$$

Using the Wiener chaos expansion, we will study equations (4.1) and (4.2) through the analysis of the propagator system of the QsNS equations. The technique is similar to

Chapter 3. We recall from (2.6) that

$$(u^i \diamond \partial_{x_i} u)_\alpha = \sum_{0 \le \gamma \le \alpha} \sqrt{\tbinom{\alpha}{\gamma}}\, (u_\gamma, \nabla) u_{\alpha-\gamma}.$$

Applying the above formula, we obtain the propagator system of (4.1),

$$\begin{aligned} &\partial_t u_0 + B(u_0, u_0) = \nu \Delta u_0 + f \\ &\operatorname{div} u_{(0)} = 0 \\ &u_{(0)}(0, x) = w(x), \quad u_\alpha|_{\partial D} = 0. \end{aligned} \qquad , \tag{4.7a}$$

$$\begin{aligned} &\partial_t u_\alpha + B(u_\alpha, u_{(0)}) + B(u_{(0)}, u_\alpha) + \textstyle\sum_{0<\gamma<\alpha} \sqrt{\tbinom{\alpha}{\gamma}}\, B(u_\gamma, u_{\alpha-\gamma}) \\ &\quad = \nu \Delta u_\alpha + \textstyle\sum_l \sqrt{\alpha_l} \mathfrak{u}_l(x) \big(\sigma^i \partial_{x_i} u_{\alpha-\epsilon_l} + \nabla P^g_{\alpha-\epsilon_l} + \mathbf{1}_{\alpha=\epsilon_l} g\big) \\ &\operatorname{div} u_\alpha = 0 \\ &u_\alpha(0, x) = 0, \quad u_\alpha|_{\partial D} = 0 \end{aligned} \tag{4.7b}$$

with equality holding in V'. Similarly, the propagator system of (4.2) is

$$\begin{aligned} &B(\bar{u}_0, \bar{u}_0) = \nu \Delta \bar{u}_0 + \bar{f} \\ &\operatorname{div} \bar{u}_{(0)} = 0, \quad \bar{u}_{(0)}|_{\partial D} = 0 \end{aligned} \quad , \tag{4.8a}$$

$$\begin{aligned} &B(\bar{u}_\alpha, \bar{u}_{(0)}) + B(\bar{u}_{(0)}, \bar{u}_\alpha) + \textstyle\sum_{0<\gamma<\alpha} \sqrt{\tbinom{\alpha}{\gamma}}\, B(\bar{u}_\gamma, \bar{u}_{\alpha-\gamma}) \\ &\quad = \nu \Delta u_\alpha + \textstyle\sum_l \sqrt{\alpha_l} \mathfrak{u}_l(x) \big(\bar{\sigma}^i \partial_{x_i} \bar{u}_{\alpha-\epsilon_l} + \nabla \bar{P}^g_{\alpha-\epsilon_l} + \bar{g}_{\alpha-\epsilon_l}\big) \\ &\operatorname{div} \bar{u}_\alpha = 0, \quad \bar{u}_\alpha|_{\partial D} = 0 \end{aligned} \tag{4.8b}$$

with equality holding in V'.

The zeroth mode $u_{(0)} = \mathbb{E} u$ is the mean of (4.1) and solves the the unperturbed Navier-Stokes equations (4.7a).

2. Stationary QSNS

Given deterministic functions $\bar{f}, \bar{g}, \bar{\sigma} \in L_2(D)$, we seek a weak/variational solution $\bar{u} \in \mathcal{D}'(V)$ (generalized random element with values in V) satisfying

$$-\nu \langle\!\langle \Delta \bar{u}, \varphi \rangle\!\rangle + \langle\!\langle \bar{u}^i \diamond \partial_{x_i} \bar{u},\, \varphi \rangle\!\rangle = \langle\!\langle \bar{f}, \varphi \rangle\!\rangle + \langle\!\langle \big(\bar{\sigma}^i \partial_{x_i} \bar{u} + \nabla \bar{P}^g + \bar{g}\big) \diamond \dot{W}(x),\, \varphi \rangle\!\rangle \tag{4.9}$$

for all test random elements $\varphi \in \mathcal{D}(V)$.

We will first show the existence and uniqueness of a generalized strong solution.

PROPOSITION 2.1. *Assume the dimension $d = 2, 3$. Assume $\bar{f}, \bar{g}, \bar{\sigma}$ are deterministic functions satisfying*

$$\bar{f}, \bar{g}, \bar{\sigma} \in H, \tag{A0}$$

$$\nu^2 > c_b \|\bar{f}\|_{V'}, \tag{A1}$$

$$\bar{g} \in H^1(D), \quad \bar{\sigma} \in W^{1,\infty}(D). \tag{A2}$$

Then there exists a unique generalized strong solution $u \in \mathcal{D}'(H^2(D) \cap V)$ of (4.2).

REMARK. It is interesting to note that condition (A1) in Proposition 2.1, that ensures the existence of a generalized strong solution, is the same condition that ensure the uniqueness of the strong solution of the deterministic Navier-Stokes equation. Thus, Proposition 2.1 generalizes the analogous result in the deterministic Navier-Stokes theory, which is the special subcase when $\bar{g} = \bar{\sigma} = 0$.

PROOF.

Solution for $\alpha = (0)$. The equation for $\bar{u}_0$ is the deterministic stationary Navier-Stokes equation, for which the existence and uniqueness of weak solutions is well-known [**59**, **60**]. From (A1), there exists a unique weak solution $\bar{u}_0 \in V$ of (4.8a) satisfying

$$\|\bar{u}_0\|_V \leq \frac{1}{\nu} \|\bar{f}\|_{V'} < \frac{\nu}{c_b}. \tag{4.10}$$

Moreover, since $\bar{f} \in L_2(D)$, then $\bar{u}_0 \in \mathrm{dom}(-\Delta)$, with

$$|\Delta \bar{u}_0| \leq \frac{2}{\nu} |\bar{f}| + \frac{c_d^2}{\nu^5 \lambda_1^{3/2}} |\bar{f}|^3.$$

THE BILINEAR FORM $\bar{a}_0(\cdot, \cdot)$. Define the bilinear continuous form $\bar{a}_0$ on $V \times V$ by

$$\bar{a}_0(u, v) = \nu(\nabla u, \nabla v) + b(u, \bar{u}_0, v) + b(\bar{u}_0, u, v) \tag{4.11}$$

where $\bar{u}_0(x)$ is the solution of the stationary (deterministic) Navier-Stokes equation (4.8a) just found. Also define the mapping $\bar{A}_0 : V \to V'$, by

$$\langle \bar{A}_0(u), v \rangle = \bar{a}_0(u, v), \quad \text{for all } v \in V.$$

Then (4.8b) can be written as

$$\bar{A}_0(\bar{u}_\alpha) = -\sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}}\, B(\bar{u}_\gamma, \bar{u}_{\alpha-\gamma}) + \sum_l \sqrt{\alpha_l}\mathfrak{u}_l(x)\big(\bar{\sigma}^i \partial_{x_i} \bar{u}_{\alpha-\epsilon_l} + \nabla \bar{P}^g_{\alpha-\epsilon_l} + \mathbf{1}_{\alpha=\epsilon_l}\bar{g}\big)$$

for $|\alpha| \geq 1$.

To obtain the existence and uniqueness of u_α, we intend to apply the Lax-Milgram lemma to the bilinear form $\bar{a}_0(\cdot,\cdot)$. To do this, we first check the coercivity of $\bar{a}_0(\cdot,\cdot)$ on V.

Lemma 2.2. *Assume (A1), and assume u_0 solves (4.8a) with $f \in V'$. Then $\bar{a}_0(\cdot,\cdot)$ defined in* (4.11) *is coercive and bounded on V.*

Proof. Indeed, for any $v \in V$,

$$\begin{aligned} \bar{a}_0(v,v) &= \nu|\nabla v|^2 + b(v,\bar{u}_0,v) + b(\bar{u}_0,v,v) \\ &\geq \nu|\nabla v|^2 - c_b\|\bar{u}_0\|_V\|v\|_V^2 \\ &= \big(\nu - c_b\|\bar{u}_0\|_V\big)\|v\|_V^2 = \bar{\beta}\|v\|_V^2, \end{aligned}$$

where $\bar{\beta} := \nu - c_b\|\bar{u}_0\|_V > 0$ by (4.10). Next, $\bar{a}_0(\cdot,\cdot)$ is bounded, because

$$\begin{aligned} |\bar{a}_0(v,w)| &\leq \nu\|v\|_V\|w\|_V + |b(v,\bar{u}_0,w)| + |b(\bar{u}_0,v,w)| \\ &\leq \big(\nu + c_b\|\bar{u}_0\|_V\big)\|v\|_V\|w\|_V \end{aligned}$$

for any $v, w \in V$. □

We continue with the proof of Proposition 2.1.

Solutions for $\alpha = \epsilon_l$. Equation (4.8b) in variational form reduces to finding $\bar{u}_{\epsilon_l} \in V$ such that

$$\bar{a}_0(\bar{u}_{\epsilon_l}, v) = \langle \mathfrak{u}_l\big(\bar{\sigma}^i\partial_{x_i}\bar{u}_0 + \nabla\bar{P}^g_0 + \bar{g}\big),\, v\rangle =: \langle G_{\epsilon_l}, v\rangle$$

for all $v \in V$. To apply the Lax-Milgram lemma to (4.8b), we check that the forcing term

$$G_{\epsilon_l} := \mathfrak{u}_l\big(\bar{\sigma}^i\partial_{x_i}\bar{u}_0 + \nabla\bar{P}^g_0 + \bar{g}\big)$$

belongs to V'. In fact, we have that G_{ϵ_l} belongs to $L^2(D)$. Indeed, due to assumption (A2), $|\bar{\sigma}^i\partial_{x_i}\bar{u}_0| \leq \|\bar{\sigma}\|_{L^\infty}\|\bar{u}_0\|_V$, and from (4.6) we have $\|\bar{P}^g_0\|_{H^2} \leq C(\|\bar{\sigma}\|_{W^{1,\infty}}\|\bar{u}_0\|_V + \|\bar{g}\|_{H^1})$.

So, from (4.10),

$$\begin{aligned}|G_{\epsilon_l}| &\leq C\|\mathfrak{u}_l\|_{L^\infty}\Big(\|\bar{\sigma}\|_{W^{1,\infty}}\|\bar{u}_0\|_V + \|\bar{g}\|_{H^1}\Big)\\ &\leq C\|\mathfrak{u}_l\|_{L^\infty}\Big(\frac{\nu}{c_b}\|\bar{\sigma}\|_{W^{1,\infty}} + \|\bar{g}\|_{H^1}\Big)\end{aligned}$$

By the Lax-Milgram lemma, there exists a unique variational solution $\bar{u}_{\epsilon_l} \in V$ with the estimate

$$\|\bar{u}_{\epsilon_l}\|_V \leq \frac{1}{\bar{\beta}}C\|\mathfrak{u}_l\|_{L^\infty}\Big(\frac{\nu}{c_b}\|\bar{\sigma}\|_{W^{1,\infty}} + \|\bar{g}\|_{H^1}\Big).$$

Additionally, by a standard technique in [**60**], there exists $\bar{P}^f_{\epsilon_l} \in L^2(D)$ such that (4.8b) holds in V'.

Next, observe that by the continuity property $B : V \times H^2 \to L^2$,

$$-\nu\Delta\bar{u}_{\epsilon_l} = G_{\epsilon_l} - B(\bar{u}_{\epsilon_l}, \bar{u}_0) - B(\bar{u}_0, \bar{u}_{\epsilon_l}) \in L^2(D)$$

Hence, $\bar{u}_{\epsilon_l} \in \mathrm{dom}(-\Delta)$, and we have the estimate

$$\begin{aligned}|\Delta\bar{u}_{\epsilon_l}| &\leq \frac{1}{\nu}\Big(|G_{\epsilon_l}| + |B(\bar{u}_{\epsilon_l}, \bar{u}_0)| + |B(\bar{u}_0, \bar{u}_{\epsilon_l})|\Big)\\ &\leq \frac{1}{\nu}\Big(|G_{\epsilon_l}| + 2c_b|\Delta\bar{u}_0|\;\|\bar{u}_{\epsilon_l}\|_V\Big)\\ &\leq \frac{C\sup_l\|\mathfrak{u}_l\|_{L^\infty}}{\nu\bar{\beta}}\Big(\frac{\nu}{c_b}\|\bar{\sigma}\|_{W^{1,\infty}} + \|\bar{g}\|_{H^1}\Big)\Big(1 + \frac{2c_b}{\bar{\beta}}|\Delta\bar{u}_0|\Big)\\ &= \bar{K}\end{aligned}$$

and $\bar{K} = \bar{K}(\nu, \bar{f}, \bar{g}, \bar{\sigma})$ does not depend on l.

Solutions for $|\alpha| \geq 2$. Denote

$$\begin{aligned}G_\alpha &:= \sum_l \sqrt{\alpha_l}\,\mathfrak{u}_l\big(\bar{\sigma}^i\partial_{x_i}\bar{u}_{\alpha-\epsilon_l} + \nabla\bar{P}^g_{\alpha-\epsilon_l}\big),\\ F_\alpha &:= -\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\;B(\bar{u}_\gamma, \bar{u}_{\alpha-\gamma})\end{aligned}$$

We first find $\bar{u}_\alpha \in V$ such that

$$\bar{a}_0(\bar{u}_{\epsilon_l}, v) = \langle F_\alpha + G_\alpha, v\rangle$$

for all $v \in V$.

We prove by induction. Assume we have shown the existence of a unique solution $\bar{u}_\gamma \in \mathrm{dom}(-\Delta)$ for all $|\gamma| \leq n-1$. By a similar argument as above, we have $G_\alpha \in L^2(D)$ with

$$|G_\alpha| \leq C \sum_l \sqrt{\alpha_l} \|\mathfrak{u}_l\|_{L^\infty} \|\bar{\sigma}\|_{W^{1,\infty}} \|\bar{u}_{\alpha-\epsilon_l}\|_V < \infty.$$

Also, since $B(\cdot,\cdot)$ is a bilinear continuous from $H^2 \times H^2 \to L^2$, we deduce that $F_\alpha \in L^2(D)$ with

$$|F_\alpha| \leq c_b \sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}}\ |\Delta \bar{u}_\gamma|\, |\Delta \bar{u}_{\alpha-\gamma}| < \infty$$

Applying the Lax-Milgram lemma, there exists a unique solution $\bar{u}_\alpha \in V$ with the estimates

$$\|\bar{u}_\alpha\|_V \leq \frac{1}{\bar{\beta}}(|G_\alpha| + |F_\alpha|).$$

Finally, since

$$-\nu \Delta \bar{u}_\alpha = F_\alpha + G_\alpha - B(\bar{u}_\alpha, \bar{u}_0) - B(\bar{u}_0, \bar{u}_\alpha) \in L^2(D),$$

we deduce that $u_\alpha \in \mathrm{dom}(-\Delta)$, with

$$\begin{aligned} |\Delta \bar{u}_\alpha| &\leq \frac{1}{\nu}\left(|F_\alpha| + |G_\alpha| + |B(\bar{u}_\alpha, \bar{u}_0)| + |B(\bar{u}_0, \bar{u}_\alpha)|\right) \\ &\leq \frac{1}{\nu}\left(|F_\alpha| + |G_\alpha| + 2c_b \|\bar{u}_\alpha\|_V |\Delta \bar{u}_0|\right) \\ &\leq \frac{1}{\nu}(|F_\alpha| + |G_\alpha|)\Big(1 + \frac{2c_b}{\bar{\beta}}|\Delta \bar{u}_0|\Big) < \infty \end{aligned}$$

Hence, we have found a solution $u \in \mathcal{D}'(H^2(D) \cap V)$. □

Next, we find the appropriate Kondratiev space to which the solution u belongs. As described previously, the estimation of the Kondratiev norm makes use of the recursion properties of the Catalan numbers. The details of how the Catalan number rescaling is used in our estimates is put off until Section 6, though in the proofs presented before that section, we will apply the results from there.

Proposition 2.3. *Assume (A0-2) hold. Then there exists $q_0 > 2$, depending on ν, $\bar{f}$, $\bar{g}$, $\bar{\sigma}$ such that $\bar{u}$ belongs to the Kondratiev space $S_{-1,-q}(H^2(D) \cap V)$, for $q > q_0$.*

Proof. For $|\alpha| \geq 1$, we have found in the proof of Proposition 2.1 estimates for $|\Delta \bar{u}_\alpha|$,

$$|\Delta \bar{u}_{\epsilon_l}| \leq \bar{K}$$

$$\frac{1}{\sqrt{\alpha!}}|\Delta\bar{u}_\alpha| \leq \bar{B}_0\left(\sum_{0<\gamma<\alpha}\frac{|\Delta\bar{u}_\gamma|}{\sqrt{\gamma!}}\frac{|\Delta\bar{u}_{\alpha-\gamma}|}{\sqrt{(\alpha-\gamma)!}} + \mathbf{1}_{\sigma\neq 0}\sum_l \mathbf{1}_{\alpha_l\neq 0}\frac{1}{\sqrt{(\alpha-\epsilon_l)!}}\|\bar{u}_{\alpha-\epsilon_l}\|_V\right).$$

where $\bar{B}_0$ depends on $\nu, \bar{f}, \bar{\sigma}$. Let $L_{\epsilon_l} = 1 + |\Delta\bar{u}_{\epsilon_l}|$, and $L_\alpha = \frac{1}{\sqrt{\alpha!}}|\Delta\bar{u}_\alpha|$ for $|\alpha| \geq 2$. Then the rest of the proof follows in a similar way to the proof of Lemma 3 in [**55**]:

$$L_\alpha \leq \bar{B}_0 \sum_{0<\gamma<\alpha} L_{\alpha-\gamma}L_\gamma$$

and by the Catalan numbers method in Appendix 6,

$$|\Delta\bar{u}_\alpha|^2 \leq \alpha!\mathcal{C}^2_{|\alpha|-1}\binom{|\alpha|}{\alpha}(2\mathbb{N})^\alpha \bar{B}_0^{2(|\alpha|-1)}\bar{K}^{2|\alpha|} \tag{4.12}$$

for $|\alpha| \geq 1$, and the result holds with q_0 satisfying

$$\bar{B}_0^2\bar{K}^2 2^{5-q_0}\sum_{i=1}^{\infty} i^{1-q_0} = 1. \tag{4.13}$$

□

3. The time-dependent QSNS (4.1)

In this section, we consider for simplicity equation (4.1) with $\sigma(t,x) = 0$ and, wlog, $P^g = 0$ and $\operatorname{div} g = 0$. We will consider the time-dependent solution $u(t)$ of (4.1) on a finite time interval $[0,T]$ if $d = 2, 3$, and also study its uniform boundedness on $[0,\infty)$ for $d = 2$. The former result allows an arbitrarily large time interval, thereby ensuring a global-in-time solution. On the other hand, the latter result will become useful for showing the long-time convergence of the solution to a steady state solution.

For any $T < \infty$, it is known that a strong solution $u_0(t)$ of the deterministic Navier-Stokes equation (4.7a) exists on the finite interval $[0,T]$ if $d = 2$, and exists on $[0,(T \wedge T_1)]$ for a specific $T_1 = T_1(u_0(0))$ depending on $u_0(0)$ if $d = 3$. Without further conditions, we have the following result for a generalized strong solution of the quantized Navier-Stokes equation.

Lemma 3.1. *For $d = 2, 3$, let $T < \infty$ if $d = 2$, or $T \leq T_1$ if $d = 3$. Assume the forcing terms $\bar{f}, \bar{g}$ and initial condition $u(0)$ are deterministic functions satisfying*

$$f, g \in L^2(0,T;H), \qquad u(0) \in V. \tag{A0$'$}$$

Then there exists a unique generalized strong solution $u(t) \in \mathcal{D}'(H^2(D) \cap V)$ for a.e. $t \in [0,T]$. Moreover, $u_\alpha \in C([0,T],V)$ for all α.

Proof. For $\alpha = (0)$, it is well-known that (4.7a) has a unique solution u_0, and

$$u_0 \in L^2([0,T];\mathrm{dom}(-\Delta)), \quad u_0 \in C([0,T];V).$$

The bilinear form $a_0(t)$. For $t \in [0,T]$, define the bilinear continuous form $a_0(t)$ on $V \times V$ by

$$a_0(u,v;t) = \nu(\nabla u, \nabla v) + b(u, u_0(t), v) + b(u_0(t), u, v)$$

where $u_0(t,x)$ is the solution of the time-dependent (deterministic) Navier-Stokes equations given in (4.7a) just found. Also define the mapping $A_0(t) : V \to V'$, for $t \in [0,T]$, by

$$\langle A_0(t)u, v\rangle = a_0(u,v;t), \quad \text{for all } v \in V.$$

Then (4.7b) can be written as

$$\partial_t u_\alpha + A_0(t)u_\alpha + \sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}}\, B(u_\gamma, u_{\alpha-\gamma}) = \sum_l \sqrt{\alpha_l}\mathfrak{u}_l(x)\big(\sigma^i\partial_{x_i}u_{\alpha-\epsilon_l} + \nabla P^g_{\alpha-\epsilon_l} + \mathbf{1}_{\alpha=\epsilon_l}g\big)$$

This is a linearized Stokes equation of the form

$$\begin{aligned} &\partial_t U + A_0(t)U = F \\ &U|_{\partial D} = 0, \quad U(0) = w \end{aligned}.$$

Since $u_0 \in L^2([0,T];\mathrm{dom}(-\Delta))$, it can be shown by standard compactness techniques that if $F \in L^2(0,T;H)$ and $w \in V$, then there exists a unique strong solution $U \in L^2(0,T;\mathrm{dom}(-\Delta))$ with

$$U \in L^2(0,T;\mathrm{dom}(-\Delta)), \quad U_t \in L^2(0,T;H), \quad \text{and} \quad U \in C(0,T;V)$$

We prove the lemma by induction. For $|\alpha| \geq 1$, assume that $u_\gamma \in L^2([0,T];\mathrm{dom}(-\Delta))$ for all $\gamma < \alpha$. We check for the RHS of (4.7b),

$$-\sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}}\, B(u_\gamma, u_{\alpha-\gamma}) + \mathbf{1}_{\alpha=\epsilon_l}\mathfrak{u}_l g \ \in L^2([0,T];H)$$

This follows from (A0$'$) and the fact that $|B(u_\gamma, u_{\alpha-\gamma})| \leq c_b|\Delta u_\gamma|\,|\Delta u_{\alpha-\gamma}|$. It follows from linear theory that there exists a unique solution u_α of (4.7b) with

$$u_\alpha \in L^2([0,T]; \mathrm{dom}(-\Delta)), \quad \partial_t u_\alpha \in L^2([0,T]; H), \quad \text{and } u_\alpha \in C([0,T]; V).$$

□

Remark. If $\sigma \neq 0$, then in addition to (A0$'$), we must require $g \in L^2(0,T;H^1(D))$ and $\sigma \in L^2(0,T;W^{1,\infty}(D))$. (Compare with (A2).)

Next, we study $\|u(t)\|_{-1,-q;H^2}$ on a finite interval $[0,T]$ as well as the uniform boundedness of $\|u(t)\|_{-1,-q;V}$ for all time $t \in [0,\infty)$. We recall the following established result on the uniform bounds of u_0 in the V and $H^2(D)$ norms.

Lemma 3.2. *(Lemma 11.1 in [**59**]; see also [**24**]) Assume for the initial condition that $u_0(0,\cdot) \in V$, and assume*

$$f \text{ is continuous and bounded from } [0,\infty) \text{ into } H$$

$$f' \text{ is continuous and bounded from } [0,\infty) \text{ into } V'$$

Let $u_0(t)$ be the strong solution of the deterministic Navier-Stokes equations (4.7a), defined on $[0,\infty)$ if $d = 2$, or on $[0,T_1]$ if $d = 3$. Then

$$\sup_{t\geq 0} \|u_0(t)\|_V \leq c'(\|u_{(0)}(0,\cdot)\|_V, \nu, \bar{f}, D). \tag{4.14a}$$

$$\sup_{t\geq \tau} |\Delta u_0(t)| \leq c''(\tau, \|u_{(0)}(0,\cdot)\|_V, \nu, \bar{f}, D). \tag{4.14b}$$

for any $\tau > 0$.

Proposition 3.3. *(i) For $d = 2,3$, let $[0,T]$ be a finite subinterval of $[0,\infty)$ if $d = 2$, or of $[0,T_1(u_0(0))]$ if $d = 3$. Assume (A0$'$) and assume*

$$\nu > 4c_b c'. \tag{A1$''$}$$

where $c' = c'(\|u(0,\cdot)\|_V, \nu, \bar{f}, D)$ in (4.14a).

Then there exists some $q_1 > 2$ depending on ν, c', c_b and T, such that for $q > q_1$,

$$u \in \mathcal{S}_{-1,-q}(\,L^2(0,T;dom(-\Delta))\,) \cap \mathcal{S}_{-1,-q}(L^\infty(0,T;V)).$$

(ii) For $d = 2$, assume the hypothesis of Lemma 3.2, and assume g is bounded from $[0, \infty)$ into H. Also assume

$$\text{(A1}'\text{)} \qquad \nu^4 > \frac{2^7 c_b^4 c'^4}{\lambda_1}$$

where $c' = c'(\|u(0, \cdot)\|_V, \nu, \bar{f}, D)$ in (4.14a).

Then there exists $q_2 > 2$ depending on ν, c' and c_b, such that for $q > q_2$,

$$\sup_{t \geq 0} \|u(t)\|_{-1,-q;V} < \infty.$$

Proof. (i) For $\alpha = (0)$, (4.14a) and the usual deterministic theory implies that $u_0 \in L^2(0, T; \operatorname{dom}(-\Delta)) \cap L^\infty(0, T; V)$.

For $|\alpha| = 1$, $\alpha = \epsilon_l$, choose in (4.7b) the test function $v = (-\Delta)u_\alpha$,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|u_{\epsilon_l}\|_V^2 + \nu|\Delta u_{\epsilon_l}|^2 &\leq |b(u_{\epsilon_l}, u_0, \Delta u_{\epsilon_l})| + |b(u_0, u_{\epsilon_l}, \Delta u_{\epsilon_l})| + |\langle \mathfrak{u}_l g, \Delta u_{\epsilon_l}\rangle| \\ &\leq 2c_b\|u_0\|_V|\Delta u_{\epsilon_l}|^2 + |\mathfrak{u}_l g|\,|\Delta u_{\epsilon_l}| \\ &\leq \big(2c_b c' + \frac{\nu}{2}\big)|\Delta u_{\epsilon_l}|^2 + \frac{1}{2\nu}|\mathfrak{u}_l g|^2\end{aligned}$$

So

$$\sup_{0\leq t\leq T}\|u_{\epsilon_l}(t)\|_V^2 + (\nu - 4c_b c')\int_0^T |\Delta u_{\epsilon_l}|^2 dt \leq \frac{1}{\nu}\int_0^T |\mathfrak{u}_l g|^2 dt$$

By (A1$''$),

$$\begin{aligned}L_{\epsilon_l} &:= \sup_{0\leq t\leq T}\|u_{\epsilon_l}(t)\|_V + \left(\int_0^T |\Delta u_{\epsilon_l}|^2 dt\right)^{1/2} \\ &\leq \frac{1}{\sqrt{\nu}}\left(1 + \frac{1}{\sqrt{\nu - 4c_b c'}}\right)\left(\int_0^T |\mathfrak{u}_l g|^2 dt\right)^{1/2} \leq K_1\end{aligned}$$

where K_1 does not depend on l.

For $|\alpha| \geq 2$,

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\|u_\alpha\|_V^2 + \nu|\Delta u_\alpha|^2 \\ &\leq |b(u_\alpha, u_0, \Delta u_\alpha)| + |b(u_0, u_\alpha, \Delta u_\alpha)| + \sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,|b(u_\gamma, u_{\alpha-\gamma}, \Delta u_\alpha)| \\ &\leq 2c_b\|u_0\|_V|\Delta u_\alpha|^2 + \sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,c_b\|u_\gamma\|_V|\Delta u_{\alpha-\gamma}|\,|\Delta u_\alpha|\end{aligned}$$

So

$$\begin{aligned}
&\frac{1}{2}\sup_{0\le t\le T}\|u_\alpha(t)\|_V^2+(\nu-2c_bc')\int_0^T|\Delta u_\alpha|^2dt\\
&\le c_b\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\Big(\int_0^T\|u_\gamma\|_V^2|\Delta u_{\alpha-\gamma}|^2dt\Big)^{1/2}\Big(\int_0^T|\Delta u_\alpha|^2dt\Big)^{1/2}\\
&\le\frac{c_b^2}{2\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\Big(\int_0^T\|u_\gamma\|_V^2|\Delta u_{\alpha-\gamma}|^2dt\Big)^{1/2}\right)^2+\frac{\nu}{2}\int_0^T|\Delta u_\alpha|^2dt
\end{aligned}$$

and for $\tilde{L}_\gamma:=\sup_{0\le t\le T}\|u_\gamma(t)\|_V+\big(\int_0^T|\Delta u_\gamma|^2dt\big)^{1/2}$,

$$\begin{aligned}
&\sup_{0\le t\le T}\|u_\alpha(t)\|_V^2+(\nu-4c_bc')\int_0^T|\Delta u_\alpha|^2dt\\
&\le\frac{c_b^2}{\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\Big(\sup_{0\le t\le T}\|u_\gamma(t)\|_V\Big)\Big(\int_0^T|\Delta u_{\alpha-\gamma}|^2dt\Big)^{1/2}\right)^2\\
&\le\frac{c_b^2}{\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,\tilde{L}_\gamma\tilde{L}_{\alpha-\gamma}\right)^2
\end{aligned}$$

Hence,

$$\tilde{L}_\alpha\le\frac{c_b}{\sqrt{\nu}}\left(1+\frac{1}{\sqrt{\nu-4c_bc'}}\right)\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,\tilde{L}_\gamma\tilde{L}_{\alpha-\gamma}$$

Let $L_\alpha=\frac{1}{\sqrt{\alpha!}}\tilde{L}_\alpha$. Then

$$L_\alpha\le B_1\sum_{0<\gamma<\alpha}L_\gamma L_{\alpha-\gamma}$$

where B_1 depends on ν and c'. By the Catalan numbers method as discussion in Appendix 6,

$$\|u_\alpha\|_{L^\infty(0,T;V)}+\|\Delta u_\alpha\|_{L^2(0,T;H)}\le\sqrt{\alpha!}\mathcal{C}_{|\alpha|-1}\binom{|\alpha|}{\alpha}B_1^{|\alpha|-1}K_1^{|\alpha|}$$

and the statement of the Proposition holds with q_1 satisfying

$$B_1^2K_1^2 2^{5-q_1}\sum_{i=1}^{\infty}i^{1-q_1}=1.$$

(ii) We now show the uniform boundedness of each mode u_α for all $t\ge 0$. For $\alpha=(0)$, this is shown in the estimates of (4.14a). For $|\alpha|=1$, $\alpha=\epsilon_l$, choose in (4.7b) the test function $v = (-\Delta)u_\alpha$,

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|u_{\epsilon_l}\|_V^2 + \nu|\Delta u_{\epsilon_l}|^2 &\le |b(u_{\epsilon_l}, u_0, \Delta u_{\epsilon_l})| + |b(u_0, u_{\epsilon_l}, \Delta u_{\epsilon_l})| + |\langle \mathfrak{u}_l g, \Delta u_{\epsilon_l}\rangle| \\
&\le 2c_b\|u_0\|_V\|u_{\epsilon_l}\|_V^{1/2}|\Delta u_{\epsilon_l}|^{3/2} + |\mathfrak{u}_l g|\,|\Delta u_{\epsilon_l}| \\
&\le \frac{\varepsilon}{2}|\Delta u_{\epsilon_l}|^2 + \frac{1}{2\varepsilon}\big(2c_b\|u_0\|_V\|u_{\epsilon_l}\|_V^{1/2}|\Delta u_{\epsilon_l}|^{1/2} + |\mathfrak{u}_l g|\big)^2 \\
&\le \frac{\varepsilon}{2}|\Delta u_{\epsilon_l}|^2 + \frac{2c_b^2\|u_0\|_V^2}{2\varepsilon}\|u_{\epsilon_l}\|_V|\Delta u_{\epsilon_l}| + \frac{1}{\varepsilon}|\mathfrak{u}_l g|^2 \\
&\le (\varepsilon)|\Delta u_{\epsilon_l}|^2 + \frac{2^3c_b^4\|u_0\|_V^4}{\varepsilon^3}\|u_{\epsilon_l}\|_V^2 + \frac{1}{\varepsilon}|\mathfrak{u}_l g|^2
\end{aligned}$$

Taking $\varepsilon = \frac{\nu}{2}$,

$$\frac{d}{dt}\|u_{\epsilon_l}\|_V^2 + \nu|\Delta u_{\epsilon_l}|^2 \le \frac{2^7c_b^4}{\nu^3}\|u_0\|_V^2\|u_{\epsilon_l}\|_V^2 + \frac{4}{\nu}|\mathfrak{u}_l g|^2$$

and from (4.3) and (4.14a),

$$\begin{aligned}
\frac{d}{dt}\|u_{\epsilon_l}\|_V^2 &\le \Big(\frac{2^7c_b^4c'^4}{\nu^3} - \nu\lambda_1\Big)\|u_{\epsilon_l}\|_V^2 + \frac{4}{\nu}|\mathfrak{u}_l g|^2 \\
&\le -\beta\|u_{\epsilon_l}\|_V^2 + \frac{4}{\nu}|\mathfrak{u}_l g|^2
\end{aligned}$$

where $\beta := -\big(\frac{2^7c_b^4c'^4}{\nu^3} - \nu\lambda_1\big) > 0$ by (A1$'$). By Gronwall's inequality,

$$\|u_{\epsilon_l}(T)\|_V^2 \le \int_0^T \frac{4}{\nu}|\mathfrak{u}_l g|^2 e^{-\beta(T-s)}ds \le \frac{4}{\nu\beta}\|\mathfrak{u}_l\|_{L^\infty}^2\|g\|_{L^\infty(0,\infty;H)}^2\left(1 - e^{-\beta T}\right)$$

for any $T > 0$. Also,

$$|\Delta u_{\epsilon_l}(t)|^2 \le \frac{2^7c_b^4c'^2}{\nu^4}\|u_{\epsilon_l}(t)\|_V^2 + \frac{4}{\nu^2}|\mathfrak{u}_l g(t)|^2.$$

It follows that

$$L_{\epsilon_l} := \sup_{t\ge 0}\left(\|u_{\epsilon_l}(t)\|_V + |\Delta u_{\epsilon_l}(t)|\right) \le K_2,$$

for all l, where the constant K_2 is independent of l and t.

For $|\alpha| \ge 2$, let $L_\alpha := \frac{1}{\sqrt{\alpha!}}\sup_{t\ge 0}(\|u_\alpha(t)\|_V + |\Delta u_\alpha(t)|$. Then

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\|u_\alpha\|_V^2 + \nu|\Delta u_\alpha|^2 \\
&\le |b(u_\alpha, u_0, \Delta u_\alpha)| + |b(u_0, u_\alpha, \Delta u_\alpha)| + \sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,|b(u_\gamma, u_{\alpha-\gamma}, \Delta u_\alpha)|
\end{aligned}$$

$$\leq 2c_b\|u_0\|_V\|u_\alpha\|_V^{1/2}|\Delta u_\alpha|^{3/2} + \sum_{0<\gamma<\alpha} \sqrt{\tbinom{\alpha}{\gamma}}\, c_b\|u_\gamma\|_V\, |\Delta u_{\alpha-\gamma}|\, |\Delta u_\alpha|.$$

By similar computations,

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\|u_\alpha\|_V^2 + \nu|\Delta u_\alpha|^2 \\
&\leq \frac{2^7 c_b^4}{\nu^3}\|u_0\|_V^4\|u_\alpha\|_V^2 + \frac{4c_b^2}{\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,\|u_\gamma\|_V\,|\Delta u_{\alpha-\gamma}|\right)^2 \\
&\leq \frac{2^7 c_b^4}{\nu^3}\|u_0\|_V^4\|u_\alpha\|_V^2 + \frac{4c_b^2}{\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\tbinom{\alpha}{\gamma}}\,\left(\sup_{s\geq 0}\|u_\gamma(s)\|_V\right)\left(\sup_{s\geq 0}|\Delta u_{\alpha-\gamma}(s)|\right)\right)^2
\end{aligned}$$

and so

$$\frac{d}{dt}\|u_\alpha\|_V^2 \leq -\beta\|u_\alpha\|_V^2 + \frac{4c_b^2}{\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\alpha!}L_\gamma L_{\alpha-\gamma}\right)^2.$$

By Gronwall's inequality and triangle inequality,

$$\begin{aligned}
\|u_\alpha(T)\|_V^2 &\leq \frac{4c_b^2}{\nu}\int_0^T\left(\sum_{0<\gamma<\alpha}\sqrt{\alpha!}L_\gamma L_{\alpha-\gamma}\, e^{-\beta(T-s)/2}\right)^2 ds \\
&\leq \frac{4c_b^2}{\nu}\left(\sum_{0<\gamma<\alpha}\sqrt{\alpha!}L_\gamma L_{\alpha-\gamma}\Big(\int_0^T e^{-\beta(T-s)}ds\Big)^{1/2}\right)^2
\end{aligned}$$

so

$$\frac{1}{\sqrt{\alpha!}}\sup_{T\geq 0}\|u_\alpha(T)\|_V \leq \frac{2c_b^2}{\sqrt{\nu\beta}}\sum_{0<\gamma<\alpha}L_\gamma L_{\alpha-\gamma}$$

We have also,

$$|\Delta u_\alpha(t)|^2 \leq \frac{2^7 c_b^4 c'^4}{\nu^4}\|u_\alpha(t)\|_V^2 + \frac{4c_b^2}{\nu^2}\left(\sum_{0<\gamma<\alpha}\sqrt{\alpha!}L_\gamma L_{\alpha-\gamma}\right)^2$$

for any $t \geq 0$.

Hence, it follows that

$$L_\alpha \leq B_2\sum_{0<\gamma<\alpha}L_\gamma L_{\alpha-\gamma}$$

where B_2 depends on ν, c' and c_b, but is independent of t.

By the Catalan method in Appendix 6,

$$\sup_{t\geq 0}\left(\|u_\alpha(t)\|_V + |\Delta u_\alpha(t)|\right) \leq \sqrt{\alpha!}\mathcal{C}_{|\alpha|-1}\binom{|\alpha|}{\alpha}B_2^{|\alpha|-1}K_2^{|\alpha|}$$

for $|\alpha| \geq 1$, and the statement of the Proposition holds with q_2 satisfying

$$B_2^2K_2^2 2^{5-q_2}\sum_{i=1}^{\infty} i^{1-q_2} = 1.$$

□

4. Long time convergence to the stationary solution

In this section, we study the solutions $u(t,x)$ of (4.1) and $\bar{u}(x)$ of (4.2) with $\sigma(t,x) = \bar{\sigma}(x) = 0$, and for simplicity consider the case with $f(t,x) = \bar{f}(x)$ and $g(t,x) = \bar{g}(x)$. We study the convergence of $u(t,x)$ to the stationary solution $\bar{u}(x)$ as $t \to \infty$, first in a weak sense (in a generalized space $\mathcal{D}'(H)$) with some exponential rate of convergence in each mode, then in a strong sense (in some Kondratiev space $\mathcal{S}_{-1,-q}(H)$) using a compact embedding argument. The latter proof, unfortunately, is does not provide a rate of convergence. For time-dependent f, g, similar results can be obtained under suitable assumptions, but the exponential convergence of each mode is not guaranteed.

Let $z(t) := u(t) - \bar{u}$. The propagator system for z is

$$z_{0,t} + B(u_0,u_0) - B(\bar{u}_0,\bar{u}_0) = \nu\Delta z_0 \tag{4.15a}$$

$$z_{\alpha,t} + A_0(t;u_\alpha) - \bar{A}_0(\bar{u}_\alpha) = -\sum_{0<\gamma<\alpha}\sqrt{\binom{\alpha}{\gamma}}\,\big(B(u_\gamma,u_{\alpha-\gamma}) - B(\bar{u}_\gamma,\bar{u}_{\alpha-\gamma})\big) \tag{4.15b}$$

with $z_\alpha(0,x) = u_\alpha(0,x) - \bar{u}_\alpha(x)$, $z|_{\partial D} = 0$ and $\operatorname{div} z_\alpha \equiv 0$, for all α.

Proposition 4.1. *Let $d = 2$. Assume (A0), (A0′), (A1), and assume*

$$\nu\left(\frac{\lambda_1}{c_2'}\right)^{3/4} > \frac{2}{\nu}|\bar{f}| + \frac{c_2^2}{\nu^5\lambda_1^{3/2}}|\bar{f}|^3 \tag{A3}$$

where c_2, c_2' are specific constants depending only on D.

Then the solution $u(t)$ of (4.1) converges in $\mathcal{D}'(H)$ to the solution $\bar{u}$ of (4.2),

$$u(t) \overset{\mathcal{D}'(H)}{\longrightarrow} \bar{u}, \qquad \text{as } t \to \infty.$$

Remark. In the following proof, all computations follow through even when $d = 3$. So, a similar statement to Proposition 4.1 can be made for $d = 3$, provided a strong solution $u(t)$ exists in $\mathcal{D}'(H^2 \cap V)$ for all $t > 0$, and the zero-th mode $u_0(t)$ satisfies the energy inequality [**59**]

$$\frac{1}{2}\frac{d}{dt}|u_0(t)|^2 + \nu\|u_0(t)\|_V^2 \leq \langle \bar{f}, u_0(t)\rangle.$$

Remark. If $f(t,x)$ and $g(t,x)$ depend on time, then an additional condition for the proposition to hold is that $f(t), g(t)$ converge to $\bar{f}, \bar{g}$ in H.

Proof. For $\alpha = (0)$, the convergence for the deterministic Navier-Stokes equation is well-known: if $u_0(t)$ is any weak solution of (4.7a) with initial condition $u_0(0) \in H$, then $u_0(t) \longrightarrow \bar{u}_{(0)}$ in H as $t \to \infty$, provided (A3) holds. Moreover, $|z_0(t)|$ decays exponentially,

$$|z_0(t)| \leq |z_0(0)|\, e^{-\bar{\nu}t}, \tag{4.16}$$

where $\bar{\nu} := \nu\lambda_1 - \frac{c_2'}{\nu^{1/3}}|\Delta \bar{u}_0|^{4/3} > 0$. (See e.g., Theorem 10.2 in [**59**]; the positivity of $\bar{\nu}$ follows from the fact that $|\Delta\bar{u}_0|$ can be majorized by the RHS of (A3).)

For $\alpha = \epsilon_l$, choosing the test function $v = z_{\epsilon_l}$ in the weak formulation of (4.15b),

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}|z_{\epsilon_l}|^2 + \nu\|z_{\epsilon_l}\|_V^2 \\
&\leq |b(z_{\epsilon_l}, \bar{u}_0, z_{\epsilon_l})| + |b(z_{\epsilon_l}, z_0, z_{\epsilon_l})| + |b(z_0, \bar{u}_{\epsilon_l}, z_{\epsilon_l})| + |b(\bar{u}_{\epsilon_l}, z_0, z_{\epsilon_l})| \\
&\leq c_b\|\bar{u}_0\|_V\|z_{\epsilon_l}\|_V^2 + c_b\|z_0\|_{L^\infty}|z_{\epsilon_l}|\,\|z_{\epsilon_l}\|_V + 2c_b|\Delta\bar{u}_{\epsilon_l}|\,|z_0|\,\|z_{\epsilon_l}\|_V \\
&\leq c_b\|\bar{u}_0\|_V\|z_{\epsilon_l}\|_V^2 + \frac{c_b^2}{2\varepsilon}\|z_0\|_{L^\infty}^2|z_{\epsilon_l}|^2 + \varepsilon\|z_{\epsilon_l}\|_V^2 + \frac{2c_b^2}{\varepsilon}|\Delta\bar{u}_{\epsilon_l}|^2|z_0|^2
\end{aligned}$$

where we have used the ε-inequality in the last line with any $0 < \varepsilon < \bar{\beta}$. So,

$$\frac{1}{2}\frac{d}{dt}|z_{\epsilon_l}|^2 + (\bar{\beta} - \varepsilon)\|z_{\epsilon_l}\|_V^2 \leq \frac{c_b^2}{2\varepsilon}\|z_0\|_{L^\infty}^2|z_{\epsilon_l}|^2 + \frac{2c_b^2}{\varepsilon}|\Delta\bar{u}_{\epsilon_l}|^2|z_0|^2. \tag{4.17}$$

Using the Poincare inequality (4.3) and taking $\varepsilon = \frac{\bar{\beta}}{2}$,

$$\frac{d}{dt}|z_{\epsilon_l}|^2 + \bar{\beta}\lambda_1 |z_{\epsilon_l}|^2 \leq \frac{2c_b^2}{\bar{\beta}}\|z_0\|_{L^\infty}^2 |z_{\epsilon_l}|^2 + \frac{8c_b^2}{\bar{\beta}}|\Delta \bar{u}_{\epsilon_l}|^2 |z_0|^2$$

For some appropriately chosen $t_0 \in (0, \infty)$ to be discussed next, we apply Gronwall's inequality,

$$|z_{\epsilon_l}(T)|^2 \leq e^{\int_{t_0}^T \varphi(t)dt}|z_{\epsilon_l}(t_0)|^2 + \int_{t_0}^T \psi_l(s) e^{\int_s^T \varphi(t)dt} ds$$

where

$$\varphi(t) = \frac{4c_b^2}{\bar{\beta}}\|z_0(t)\|_{L^\infty}^2 - \bar{\beta}\lambda_1,$$
$$\psi_l(t) = \frac{8c_b^2}{\bar{\beta}}|\Delta \bar{u}_{\epsilon_l}|^2 |z_0(t)|^2.$$

The t_0 is chosen large enough so that $\|z_0(t)\|_{L^\infty}^2 < \frac{\bar{\beta}^2\lambda_1}{4c_b^2}$ whenever $t \geq t_0$. Such t_0 exists, because by (4.14b) and the Sobolev embedding $z_0(t) \in C^{1/2}$ is Hölder continuous with exponent $\gamma < 1$ and $\sup_{t \geq \tau} \|z_0(t)\|_{C^\gamma} \leq c''$ is uniformly in t. Then due to (4.16), we deduce that in fact $z_0(t, \cdot) \longrightarrow 0$ uniformly on D as $t \to \infty$.

Consequently, we have that $\sup_{t \geq t_0} \varphi(t) < 0$. Set $\bar{\varphi} > 0$ satisfying

$$2\bar{\varphi} < \min\left\{ -\sup_{t \geq t_0} \varphi(t),\, 2\bar{\nu} \right\}.$$

Obviously, $\exp\big\{ \int_{t_0}^T \varphi(t)dt \big\} \leq \exp\big\{ -2\bar{\varphi}(T - t_0) \big\}$. Moreover, from (4.16),

$$|\psi_l(t)| \leq \frac{8c_b^2}{\bar{\beta}}|\Delta \bar{u}_{\epsilon_l}|^2 |z_0(t_0)|^2 e^{-2\bar{\nu}(t-t_0)} =: C_{\psi_l} e^{-2\bar{\nu}(t-t_0)} \longrightarrow 0$$

decays exponentially as $t \to \infty$. Combining these results,

$$\begin{aligned} |z_{\epsilon_l}(T)|^2 &\leq e^{-2\bar{\varphi}(T-t_0)}|z_{\epsilon_l}(t_0)|^2 + \int_{t_0}^T C_{\psi_l} e^{-2\bar{\nu}(s-t_0)} e^{-2\bar{\varphi}(T-s)} ds \\ &\leq e^{-2\bar{\varphi}(T-t_0)}|z_{\epsilon_l}(t_0)|^2 + \frac{C_{\psi_l}}{2(\bar{\nu} - \bar{\varphi})}\left(e^{-2\bar{\phi}(T-t_0)} e^{-2\bar{\nu}(T-t_0)} \right) \longrightarrow 0 \end{aligned}$$

as $T \to \infty$. (In the first term, $|z_{\epsilon_l}(t_0)|^2$ has been shown to be finite for any finite t_0.) Since $\bar{\varphi} < \bar{\nu}$,

$$|z_{\epsilon_l}(T)|^2 \leq \left(|z_{\epsilon_l}(t_0)|^2 + \frac{C_{\psi_l}}{2(\bar{\nu} - \bar{\varphi})} \right) e^{-2\bar{\varphi}(T-t_0)} =: K_{\epsilon_l}^2 e^{-2\bar{\varphi}(T-t_0)} \tag{4.18}$$

for $T \geq t_0$. K_{ϵ_l} does not depend on T.

For $|\alpha| \geq 2$, we prove by induction. Fix α, and assume the induction hypothesis that:

> For each $0 < \gamma < \alpha$, for $T \geq t_0$,
>
> $$|z_\gamma(T)| \leq K_\gamma e^{-2^{1-|\gamma|}\bar{\varphi}(T-t_0)} \longrightarrow 0 \tag{4.19}$$
>
> as $T \to \infty$, where K_γ does not depend on T.

We want to show that (4.19) also holds for α.

From (4.15b) with test function $v = z_\alpha$,

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}|z_\alpha|^2 + \nu|\nabla z_\alpha|^2 \\
&\leq |b(z_\alpha, \bar{u}_0, z_\alpha)| + |b(z_\alpha, z_0, z_\alpha)| + |b(z_0, \bar{u}_\alpha, z_\alpha)| + |b(\bar{u}_\alpha, z_0, z_\alpha)| \\
&\quad + \sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}} \left(|b(z_\gamma, z_{\alpha-\gamma}, z_\alpha)| + |b(z_\gamma, \bar{u}_{\alpha-\gamma}, z_\alpha)| + |b(\bar{u}_\gamma, z_{\alpha-\gamma}, z_\alpha)|\right)
\end{aligned}$$

Similar to (4.17), using the ε-inequality with any $0 < \varepsilon < \bar{\beta}/2$,

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}|z_\alpha|^2 + (\bar{\beta} - 2\varepsilon)\|z_\alpha\|_V^2 \leq{}& \frac{c_b^2}{2\varepsilon}\|z_0\|_{L^\infty}^2|z_\alpha|^2 + \frac{2c_b^2}{\varepsilon}|\Delta\bar{u}_\alpha|^2|z_0|^2 \\
&+ \frac{c_b^2}{4\varepsilon}\Big(\sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}} \left(\|z_{\alpha-\gamma}\|_V + 2\|\bar{u}_{\alpha-\gamma}\|_V\right)|z_\gamma|_{1/2}\Big)^2
\end{aligned}$$

Using the Poincare inequality and taking $\varepsilon = \bar{\beta}/4$,

$$\begin{aligned}
\frac{d}{dt}|z_\alpha(t)|^2 &\leq \Big(\frac{4c_b^2}{\bar{\beta}}\|z_0\|_{L^\infty}^2 - \lambda_1\bar{\beta}\Big)|z_\alpha|^2 + \frac{16c_b^2}{\bar{\beta}}|\Delta\bar{u}_\alpha|^2|z_0|^2 \\
&\quad + \frac{2c_b^2}{\bar{\beta}}\Big(\sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}} \left(\|z_{\alpha-\gamma}\|_V + 2\|\bar{u}_{\alpha-\gamma}\|_V\right)|z_\gamma|^{1/2}\|z_\gamma\|_V^{1/2}\Big)^2 \\
&\leq \varphi(t)|z_\alpha(t)|^2 + \psi_\alpha(t)
\end{aligned}$$

where now

$$\begin{aligned}
\psi_\alpha(t) ={}& \frac{16c_b^2}{\bar{\beta}}|\Delta\bar{u}_\alpha|^2|z_0(t)|^2 \\
&+ \frac{2c_b^2}{\bar{\beta}}\left(\sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}} \left(\|z_{\alpha-\gamma}(t)\|_V + 2\|\bar{u}_{\alpha-\gamma}\|_V\right)^2\|z_\gamma(t)\|_V\right)\left(\sum_{0<\gamma<\alpha} \sqrt{\binom{\alpha}{\gamma}}\,|z_\gamma(t)|\right)
\end{aligned}$$

From the hypothesis (4.19),

$$|\psi_\alpha(t)| \leq C_{\psi_\alpha} e^{-2\bar{\nu}(t-t_0)} + \tilde{C}_{\psi_\alpha} \Big(\sum_{0<\gamma<\alpha} \sqrt{\tbinom{\alpha}{\gamma}}\, K_\gamma e^{-2^{-|\gamma|} 2\bar{\varphi}(t-t_0)} \Big)$$

where

$$\begin{aligned} C_{\psi_\alpha} &= \frac{16 c_b^2}{\bar{\beta}} \|\bar{u}_\alpha\|_{H^2}^2 |z_0(t_0)|^2, \\ \tilde{C}_{\psi_\alpha} &= \frac{2c_b^2}{\bar{\beta}} \left(\sum_{0<\gamma<\alpha} \sqrt{\tbinom{\alpha}{\gamma}} \left(\sup_{s\geq 0} \|z_{\alpha-\gamma}(s)\|_V + 2\|\bar{u}_{\alpha-\gamma}\|_V \right)^2 \sup_{s \geq 0} \|z_\gamma(s)\|_V \right), \end{aligned}$$

and $C_{\phi_\alpha}, \tilde{C}_{\phi_\alpha}$ do not depend on t. By Gronwall's inequality,

$$\begin{aligned} |z_\alpha(T)|^2 &\leq e^{-\bar{\varphi}(T-t_0)} |z_\alpha(t_0)|^2 + \int_{t_0}^{T} \psi_\alpha(s) e^{-\bar{\varphi}(T-s)} ds \\ &\leq e^{-\bar{\varphi}(T-t_0)} |z_\alpha(t_0)|^2 + \frac{C_{\psi_\alpha}}{2(\bar{\nu} - \bar{\varphi})} e^{-2\bar{\varphi}(T-t_0)} + \tilde{C}_{\psi_\alpha} \sum_{0<\gamma<\alpha} \sqrt{\tbinom{\alpha}{\gamma}}\, K_\gamma \frac{e^{-2^{1-|\gamma|}\bar{\varphi}(T-t_0)}}{1 - 2^{-|\gamma|}} \\ &\leq K_\alpha^2 e^{-2^{1-(|\alpha|-1)}\bar{\varphi}(T-t_0)} \end{aligned}$$

where K_α does not depend on T. Hence,

$$|z_\alpha(T)| \leq K_\alpha e^{-2^{1-|\alpha|}\bar{\varphi}(T-t_0)} \tag{4.20}$$

for all $T \geq t_0$. It follows that (4.19) holds also for α, and the result follows. □

We proceed to deduce the long time convergence of $u(t)$ in some Kondratiev space $\mathcal{S}_{-1,-q}(H)$. The manner of estimates in Proposition 4.1 is not directly suited for applying the Catalan numbers method. Instead, we will use a compact embedding type argument in the following lemma to show the result.

LEMMA 4.2. *Let $u^k \in \mathcal{S}_{-1,-q}(V)$ be a sequence satisfying*

$$\sum_\alpha \frac{r^\alpha}{\alpha!} \Big(\sup_k \|u_\alpha^k\|_V^2 \Big) < \infty,$$

that is, satisfying $\{u^k\} \in \mathcal{S}_{-1,-q}(\ell^\infty(V))$.

Then there exists a subsequence $\tilde{k}_N$ such that $u^{\tilde{k}_N}$ converges in $\mathcal{D}'(H)$ to some $\bar{u} \in \mathcal{D}'(H)$. Furthermore, if $\bar{u} \in \mathcal{S}_{-1,-q}(H)$, then the convergence is in $\mathcal{S}_{-1,-q}(H)$.

Proof. The convergence in $\mathcal{D}'(H)$ will follow easily from the fact that V is compactly embedded in H. Let $\mathcal{J}_N = \{\alpha \in \mathcal{J} : |\alpha| \leq N, \text{ and } \alpha_i = 0 \text{ for } i > N\}$. Since $\sup_k \|u_0^k\|_V < \infty$, there exists a subsequence $\{k_j^0\}_{j=1}^\infty$ such that $\|u_0^k - \bar{u}_0\|_H \to 0$ for some $\bar{u}_0 \in H$. Iteratively, for each N, there exists further subsequences $\{k_j^N\}_{j=1}^\infty \subset \{k_j^{N-1}\}_{j=1}^\infty$ such that for every $\alpha \in \mathcal{J}_N$,

$$\|u_\alpha^k - \bar{u}_\alpha\|_H \to 0$$

for some $\bar{u}_\alpha \in H$. In particular, for each N, we can find j_N such that

$$\|u_\alpha^{k_{j_N}^N} - \bar{u}_\alpha\|_H \leq N^{-1}, \quad \text{for all } \alpha \in \mathcal{J}_N.$$

Consequently, choose the subsequence $\tilde{k}_N := k_{j_N}^N$ and we have found the limit $\bar{u} = \sum_\alpha \bar{u}_\alpha \xi_\alpha$. It follows that $u^{\tilde{k}_N} \to \bar{u}$ in $\mathcal{D}'(H)$.

Now suppose $\bar{u} \in \mathcal{S}_{-1,-q}(H)$. Let $\varepsilon > 0$ be arbitrary. For any N,

$$\|u^{\tilde{k}_N} - \bar{u}\|_{-1,-q;H}^2 = \sum_{\alpha \in \mathcal{J}_N} \frac{r^\alpha}{\alpha!} \|u^{\tilde{k}_N} - \bar{u}\|_H^2 + \sum_{\alpha \notin \mathcal{J}_N} \frac{r^\alpha}{\alpha!} \|u^{\tilde{k}_N} - \bar{u}\|_H^2 = (I) + (II)$$

By our special choice of $\tilde{k}_N$, there exists N_I such that

$$(I) \leq \sum_{\alpha \in \mathcal{J}_N} \frac{r^\alpha}{\alpha!} N^{-2} < \frac{\varepsilon}{2} \quad \text{whenever } N > N_I.$$

From the hypothesis of the lemma, there exists N_{II} such that

$$(II) \leq 2 \sum_{\alpha \notin \mathcal{J}_N} \frac{r^\alpha}{\alpha!} \Big(\sup_k \|u^k\|_V^2 \Big) + 2 \sum_{\alpha \notin \mathcal{J}_N} \frac{r^\alpha}{\alpha!} \|\bar{u}\|_H^2 < \frac{\varepsilon}{2} \quad \text{whenever } N > N_{II}.$$

Thus, $\|u^{\tilde{k}_N} - \bar{u}\|_{-1,-q;H}^2 < \varepsilon$ whenever $N > \max\{N_I, N_{II}\}$. □

The hypothesis in Lemma 4.2 is stronger than requiring $u^k \in l^\infty(\mathcal{S}_{-1,-q}(V))$, thus it is a weaker statement of what might be construed as a compact embedding result for Kondratiev spaces. It is not shown whether $\mathcal{S}_{-1,-q}(V)$ is compactly embedded in $\mathcal{S}_{-1,-q}(H)$. Nonetheless, it is sufficient for our purposes.

Corollary 4.3. *Let $d = 2$. Assume the hypotheses of Propositions 2.3 and 3.3(ii). Then, for the solutions $u(t)$ and $\bar{u}$ of* (4.1), (4.2), *we have that*

$$u(t) \longrightarrow \bar{u} \quad \textit{in } \mathcal{S}_{-1,-q}(H), \ \textit{as } t \to \infty,$$

for $q > \max\{q_0, q_2\}$, where q_0, q_2 are the numbers from Propositions 2.3, 3.3.

Proof. In the proof of Proposition 3.3, we have in fact shown that $u(t)$ belongs to the space $\mathcal{S}_{-1,-q}(L^\infty([0,\infty);V))$. Taking any sequence of times, $t_k \to \infty$, the sequence $\{u(t_k)\}$ satisfies the hypothesis of Lemma 4.2. So, there exists a subsequence of $u(t_k)$ converging in $\mathcal{S}_{-1,-q}(H)$ to $\bar{u}$. This is true for any sequence $\{t_k\}$, hence $u(t) \longrightarrow \bar{u}$ in $\mathcal{S}_{-1,-q}(H)$ as $t \to \infty$. □

5. Finite Approximation by Wiener Chaos Expansions

In this section, we study the accuracy of the Galerkin approximation of the solutions of the quantized stochastic Navier-Stokes equations. The goal is to quantify the convergence rate of approximate solutions obtained from a finite truncation of the Wiener chaos expansion, where the convergence is in a suitable Kondratiev space. In relation to being a numerical approximation, quantifying the truncation error is the first step towards understanding the error from the full discretization of the quantized stochastic Navier-Stokes equation.

In what follows, we will consider the truncation error estimates for the steady solution $\bar{u}$. Recall the estimate (4.12) for $|\Delta\bar{u}|$: for $r_\alpha^2 = \frac{(2\mathbb{N})^{-q\alpha}}{\alpha!}$, with $q > q_0$, we have

$$r_\alpha^2 |\Delta\bar{u}_\alpha|^2 \le \mathcal{C}_{|\alpha|-1}^2 \binom{|\alpha|}{\alpha} (2\mathbb{N})^{(1-q)\alpha} B_0^{-2} (B_0 K)^{2|\alpha|}.$$

This estimate arose from the method of rescaling via Catalan numbers, and will be the estimate we use for the convergence analysis. For the time-dependent equation, similar analysis can be performed using the analogous Catalan rescaled estimate, and will not be shown.

Let $\mathcal{J}_{M,P} = \{\alpha : |\alpha| \le P, \dim(\alpha) \le M\}$, where M, P may take value ∞. The projection of $\bar{u}$ into $\mathrm{span}\{\xi_\alpha, \alpha \in \mathcal{J}_{M,P}\}$ is $\bar{u}^{M,P} = \sum_{\alpha \in \mathcal{J}_{M,P}} \bar{u}_\alpha \xi_\alpha$.

Then the error $e = \bar{u} - \bar{u}^{M,P}$ can be written as

$$\begin{aligned} |\Delta e|^2 &= \sum_{\alpha \in \mathcal{J} \setminus \mathcal{J}_{M,P}} r_\alpha^2 |\Delta\bar{u}_\alpha|^2 \\ &= \sum_{|\alpha|=P+1}^{\infty} r_\alpha^2 |\Delta\bar{u}_\alpha|^2 + \sum_{\{|\alpha| \le P,\, |\alpha_{\le M}| < |\alpha|\}} r_\alpha^2 |\Delta\bar{u}_\alpha|^2 \end{aligned}$$

$$= \underbrace{\sum_{|\alpha|=P+1}^{\infty} r_\alpha^2|\Delta\bar{u}_\alpha|^2}_{(IV)} + \underbrace{\sum_{|\alpha|=1}^{P} \underbrace{\sum_{i=0}^{|\alpha|-1} \underbrace{\sum_{|\alpha_{\leq M}|=i} r_\alpha^2|\Delta\bar{u}_\alpha|^2}_{(I)}}_{(II)}}_{(III)}$$

We define the following values

$$\hat{Q} := 2^{1-q}B_0^2K^2\sum_{i=1}^{\infty} i^{1-q},$$

$$\hat{Q}_{\leq M} := 2^{1-q}B_0K^2\sum_{i=1}^{M} i^{1-q}, \qquad \hat{Q}_{>M} := 2^{1-q}B_0K^2\sum_{i=M+1}^{\infty} i^{1-q}.$$

In particular, the term $\hat{Q}_{>M}$ decays on the order of M^{2-q}.

We proceed to estimate the terms (I)-(IV), by similar computations to Wan et al. For fixed $1 \leq p \leq P$, $|\alpha| = p$, and fixed $i < p$,

$$\begin{aligned}(I) &\leq \mathcal{C}_{p-1}^2 B_0^{-2} \sum_{|\alpha_{\leq M}|=i,\, |\alpha_{>M}|=p-i} \binom{|\alpha|}{\alpha}(2\mathbb{N})^{(1-q)\alpha}(B_0K)^{2p} \\ &= \mathcal{C}_{p-1}^2 B_0^{-2}\binom{p}{i}\hat{Q}_{\leq M}^{i}\hat{Q}_{>M}^{p-i}\end{aligned}$$

Then for fixed $1 \leq p \leq P$, $|\alpha| = p$,

$$\begin{aligned}(II) = \sum_{i=0}^{p-1}(I) &\leq \mathcal{C}_{p-1}^2 B_0^{-2}\sum_{i=0}^{p-1}\binom{p}{i}\hat{Q}_{\leq M}^{i}\hat{Q}_{>M}^{p-i} \\ &= \mathcal{C}_{p-1}^2 B_0^{-2}(\hat{Q}^p - \hat{Q}_{\leq M}^p)\end{aligned}$$

And finally,

$$\begin{aligned}(III) = \sum_{|\alpha|=1}^{P}(II) &\leq \sum_{p=1}^{P}\mathcal{C}_{p-1}^2 B_0^{-2}(\hat{Q}^p - \hat{Q}_{\leq M}^p) \\ &\leq \frac{1}{B_0^2}(\hat{Q} - \hat{Q}_{\leq M}) + \frac{1}{16\pi B_0^2}\sum_{p=2}^{P}\frac{2^{4p}}{(p-1)^3}(\hat{Q}^p - \hat{Q}_{\leq M}^p)\end{aligned}$$

Since $\hat{Q}^p - \hat{Q}_{\leq M}^p \leq p\hat{Q}^{p-1}(\hat{Q} - \hat{Q}_{\leq M})$ by the mean value theorem for $x \mapsto x^p$,

$$(III) \leq \frac{1}{B_0^2}\hat{Q}_{>M} + \frac{1}{16\pi B_0^2}\hat{Q}_{>M}\sum_{p=2}^{P}\frac{p2^{4p}\hat{Q}^{p-1}}{(p-1)^3}$$

$$\leq \frac{1}{B_0^2}\hat{Q}_{>M} + \frac{1}{\pi B_0^2}\hat{Q}_{>M}\sum_{p=2}^{P}\frac{p(2^4\hat{Q})^{p-1}}{(p-1)^3}$$

$$\leq \frac{1}{B_0^2}\hat{Q}_{>M}\sum_{p=0}^{P-1}(2^4\hat{Q})^p$$

To estimate Term (IV),

$$(IV) \leq \sum_{p=P+1}^{\infty}\sum_{|\alpha|=p}\mathcal{C}_{p-1}^2 B_0^{-2}(2^{1-q}B_0^2K^2)^p\binom{|\alpha|}{\alpha}(\mathbb{N})^{(1-q)\alpha}$$

$$= B_0^{-2}\sum_{p=P+1}^{\infty}\mathcal{C}_{p-1}^2(2^{1-q}B_0^2K^2)^p\Big(\sum_{i\geq 1}i^{1-q}\Big)^p$$

$$\leq B_0^{-2}\sum_{p=P+1}^{\infty}\frac{2^{4(p-1)}}{\pi(p-1)^3}\hat{Q}^p \leq \frac{1}{16\pi B_0^2}\frac{(2^4\hat{Q})^{P+1}}{1-2^4\hat{Q}}$$

Putting the estimates together,

$$|\Delta e|^2 \leq C\big((2^4\hat{Q})^{P+1} + M^{2-q}\big)$$

Notice the condition $2^4\hat{Q} < 1$ in (4.13), which ensured summability of the weighted norm of the solution, is of course a required assumption for the convergence of the error estimate.

6. The Catalan numbers method

The Catalan numbers method was used in the preceding sections to derive estimates for the norms in Kondratiev spaces. This method was previously described in [**34**, **55**], but we restate it here just for the record.

Lemma 6.1. *Suppose L_α are a collection of positive real numbers indexed by $\alpha \in \mathcal{J}$, satisfying*

$$L_\alpha \leq B\sum_{0<\gamma<\alpha}L_\gamma L_{\alpha-\gamma}.$$

Then

$$L_\alpha \leq \mathcal{C}_{|\alpha|-1}B^{|\alpha|-1}\binom{|\alpha|}{\alpha}\prod_i L_{\epsilon_i}^{\alpha_i}$$

for all α, where $\mathcal{C}_n$ are the Catalan numbers.

Proof. The result is clearly true for $\alpha = \epsilon_i$. By induction, let $|\alpha| \geq 2$, and suppose the result is true for all $\gamma < \alpha$. Then

$$\begin{aligned} L_\alpha &\leq \sum_{0<\gamma<\alpha} \mathcal{C}_{|\gamma|-1}\mathcal{C}_{|\alpha-\gamma|-1} B^{|\alpha|-1} \binom{|\gamma|}{\gamma}\binom{|\alpha-\gamma|}{\alpha-\gamma}\Big(\prod_i L_{\epsilon_i}^{\alpha_i}\Big) \\ &= \sum_{n=1}^{|\alpha|-1} \sum_{\substack{0<\gamma<\alpha \\ |\gamma|=n}} \mathcal{C}_{n-1}\mathcal{C}_{|\alpha|-n-1} \frac{n!}{\gamma!}\frac{(|\alpha|-n)!}{(\alpha-\gamma)!}\, B^{|\alpha|-1}\Big(\prod_i L_{\epsilon_i}^{\alpha_i}\Big) \\ &= \sum_{n=1}^{|\alpha|-1} \mathcal{C}_{n-1}\mathcal{C}_{|\alpha|-n-1} \underbrace{\sum_{\substack{0<\gamma<\alpha \\ |\gamma|=n}} \binom{|\alpha|}{n}^{-1}\binom{\alpha}{\gamma}}_{(*)} \frac{|\alpha|!}{\alpha!} B^{|\alpha|-1}\Big(\prod_i L_{\epsilon_i}^{\alpha_i}\Big) \end{aligned}$$

We claim that $(*) = 1$, for any α and any $n < |\alpha|$. Indeed, let $K_\alpha = (k_1, \ldots, k_{|\alpha|})$ be the characteristic set of α. Each summand in $(*)$ is

$$\left(\frac{|\alpha|!}{\alpha!}\right)^{-1} \frac{n!}{\gamma!}\frac{(|\alpha|-n)!}{(\alpha-\gamma)!}$$

The term $\frac{|\alpha|!}{\alpha!}$ is the number of distinct permutations of K_α, whereas the term $\frac{n!}{\gamma!}\frac{(|\alpha|-n)!}{(\alpha-\gamma)!}$ is the number of distinct permutations of K_α where only $K_\gamma, K_{\alpha-\gamma}$ has been permuted within themselves. On the other hand, the latter term is the number of distinct permutations of K_α corresponding to a particular γ, where the correspondence of a permutation of K_α to a $\gamma \in \{\gamma : 0 < \gamma < \alpha, |\gamma| = n\}$ can be made by taking K_γ to be the first n entries of that permutation of K_α. Thus, each summand in $(*)$ is the relative frequency of γ over all distinct permutations of K_α, and hence their sum must equal 1.

To complete the proof, using the recursion property of the Catalan numbers,

$$\begin{aligned} L_\alpha &\leq \sum_{n=1}^{|\alpha|-1} \mathcal{C}_{n-1}\mathcal{C}_{|\alpha|-n-1} \left(\frac{|\alpha|!}{\alpha!}\right) B^{|\alpha|-1} \prod_i L_{\epsilon_i}^{\alpha_i} \\ &= \mathcal{C}_{|\alpha|-1}\left(\frac{|\alpha|!}{\alpha!}\right) B^{|\alpha|-1} \prod_i L_{\epsilon_i}^{\alpha_i}. \end{aligned}$$

□

If L_α satisfies the hypothesis of Lemma 6.1, and if $L_{\epsilon_i} \leq K$ for all i, then for $r = (2\mathbb{N})^{-q}$,

$$\sum_{|\alpha|=n} r^\alpha L_\alpha^2 \leq \sum_{|\alpha|=n} \mathcal{C}_{n-1}^2 B^{2(|\alpha|-1)} K^{2|\alpha|} \binom{|\alpha|}{\alpha} (2\mathbb{N})^{(1-q)\alpha}$$

$$
\begin{aligned}
&= B^{-2}\mathcal{C}_{n-1}^2\big(B^2K^22^{1-q}\big)^n \sum_{|\alpha|=n} \binom{|\alpha|}{\alpha}\mathbb{N}^{(1-q)\alpha} \\
&= B^{-2}\mathcal{C}_{n-1}^2\big(B^2K^22^{1-q}\big)^n \Big(\sum_{i=1}^{\infty} i^{(1-q)}\Big)^n
\end{aligned}
$$

For large n, the Catalan numbers behave asymptotically like $\mathcal{C}_n \sim \frac{2^{2n}}{\sqrt{\pi}n^{3/2}}$. Hence, the sum $\sum_{n=0}^{\infty}\sum_{|\alpha|=n} r^{\alpha}L_{\alpha}^2$ converges for any $q > \max\{q_0, 2\}$, where q_0 satisfies

$$
B^2K^22^{5-q_0}\sum_{i=1}^{\infty} i^{(1-q_0)} = 1.
$$

CHAPTER 5

Randomization of Incoherent Forcing for Improvement of Energy Approximations

1. Introduction

In this chapter, we consider a linear SPDE

$$\frac{\partial}{\partial t} v = \mathcal{A} v + \dot{W}_Q(x), \quad x \in U,\ t > 0, \tag{5.1}$$

and a system of deterministic PDEs

$$\frac{\partial}{\partial t} v_i(x,t) = \mathcal{A} v_i(x,t) + \rho_i e_i(x), \quad x \in U,\ t > 0, \quad \text{for } i = 1, 2, \ldots, \infty, \tag{5.2}$$

where $U \subset \mathbb{R}^d$ is an open bounded domain, $\mathcal{A}$ is a linear partial differential operator, $\{e_i,\, i \geq 1\}$ is an orthonormal basis in $L_2(U)$, and $\dot{W}(x)$ is a weighted spatial noise, given by

$$\dot{W}(x) = \sum_{i \geq 1} \sigma_i e_i(x) \xi_i$$

with $\{\xi_i,\, i \geq 1\}$ being a set of independent Gaussian random variables and $\{\sigma_i,\, i \geq 1\}$ being a set of nonnegative weights (see (2.3)). If all $\sigma_i = 1$, $\dot{W}(x)$ is a standard spatial white noise; this case is presented in [**43**]. We assume that the initial conditions in (5.1) and (5.2) are zero. In fact, we recall from Definition 1.5, and the discussion therein, that (5.1) and (5.2) are equivalent in that

$$v_i(x,t) = \mathbb{E}[v(x,t)\xi_i] \quad \text{and} \quad v(x,t) = \sum_{i \geq 1} v_i(x,t) \left(\dot{W}, e_i\right)_{L_2(U)}.$$

The equivalence of (5.1) and (5.2) is a very simple implication of the Wiener chaos expansion for SPDEs. System (5.2) is the propagator system for (5.1). Under very general assumptions, a solution of one of the two equations exists and is unique if and only if the other has a unique solution (see [**49**] and Theorem 3.2). Moreover, if $\sum_{i \geq 1} \sigma_i^2 < \infty$, then the solution is in L_2, and if $\sum_{i\geq 1}\sigma_i^2 = \infty$, then the solution is found in a Sobolev space with a negative index.

The energy of a solution u of (5.1) is defined by

$$\mathcal{E}[v(t)] := \mathbb{E}\|v(\cdot,t)\|^2_{L_2(U)} = \sum_{i\geq 1}\|v_i(\cdot,t)\|^2_{L_2(U)}. \tag{5.3}$$

Clearly, it is independent of the choice of the basis $\{e_i,\, i\geq 1\}$.

Our main goal is to identify suitable bases $\{e_i,\, i\geq 1\}$ as well as estimators $\hat{v}^{(n)}(x,t) = \sum_{i=1}^n v_i(x,t)\sigma_i\xi_i$ such that the energy of $\hat{v}^{(n)}(x,t)$ efficiently approximates $\mathcal{E}[v(t)]$. For a finite N-dimensional noise $\dot{W}_N(x)$, we want to study the behavior of the estimators as $N\to\infty$.

Getting a little bit ahead of the story, we remark that, while the energy $\mathcal{E}[v(t)]$ does not depend on the choice of the basis, the rate of convergence of the approximate energy $\sum_{i=1}^n \|v_i(\cdot,t)\|^2_{L_2(U)}$ does and, sometimes, does so quite substantially.

Approximating the energy $\|v(\cdot,t)\|^2_{L_2(U)}$ for system (5.2), and similar systems, requires solving a large number of PDEs that differ only by the forcing terms. For example, the problem of efficient approximation of the energy comes up in the modeling of wave propagation with incoherent sources [**40**], which appear in a wide range of problems in optics, such as those related to diffuse light [**71**]. Some popular examples include the Raman photonic crystal spectrometer [**51**], which is used to measure spatially incoherent light in environmental and biological sensing, as well as fluorescent or bioluminescent tomography [**66**], which has been used successfully to achieve in-vivo functional imaging in cancer research and drug monitoring. In modeling the performance of new designs for photonic crystal spectrometers, one has to compute the solutions of Maxwell equations, which govern the light propagation in the spectrometer, with spatially incoherent sources $f(x)$. Similarly, current models in fluorescent tomography are based on solving a diffusion approximation of the well-known radiative transport equation, and due to the random phase value it is again natural to model the incoherent fluorescent light source by point sources. Therefore, engineers routinely model incoherence by solving very large systems of equations, each of them excited by a point mass function $f_i(x) = f_i\delta_{x_i}(x)$, $i = 1,\ldots,N$.

On one hand, the incoherence property is well modelled by point sources in (5.2). On the other hand, the sheer number of required point sources sets a computational roadblock.

To mitigate the aforementioned numerical complications, it was proposed in [**4**, **5**] to circumvent the local scale problem by replacing the localized forcing terms with a new global scale forcing that efficiently consolidates most of the energy into just a few terms. This was implemented by replacing multiple Maxwell equations with point sources by a single Maxwell equation driven by white noise $\dot{W}_N(x) = \sum_{i=1}^{N} \xi_i n_i(x)$, where $\{\xi_i,\, i = 1, 2, \ldots, N\}$ were independent standard Gaussian random variables and $\{n_i,\, i = 1, 2, \ldots, N\}$ was a subset of a trigonometric basis. The numerical simulations presented in [**4**, **5**] demonstrate a dramatic reduction in computational complexity in evaluating the energy $\|v(\cdot, t)\|_{L_2}^2$ while maintaining a similar level of accuracy of energy approximation. However, papers [**4**, **5**] were not concerned with rigorous theoretical explanations of the validity of the proposed algorithm and the potential scope of its applicability.

Thus, we present here a rigorous approach to the problem of efficient approximation of the energy (5.3) for systems of fairly general evolution equations (5.2).

In section 3, we compare the efficiency of the small scale (point forcing) basis and the "large scale" $\mathcal{A}$-eigenfunction basis, and we deduce our main result—the approximation of the energy using the latter basis yields a *1st order improvement* over the former (see Theorem 3.1). In fact, we will show that the number of expansion terms under the eigenfunction basis is $\mathcal{O}(1)$ in N, whereas under the point forcing basis it is $\mathcal{O}(N)$. In section 4, we show numerical results for the one-dimensional heat equation under the point forcing and cosine bases that corroborate the theoretical results, and we also show results for the convection-diffusion equations that suggest the applicability of this method to a broader class of parabolic equations.

We remark that the change of basis method is not the only way to tackle the deterministic system. The key point is the randomization of the system (5.2) to the SPDE (5.1), which can then be handled by various methods, such as WCE or Monte Carlo simulation. While there are numerous works in the literature studying such equations with additive noise, most of these works use a single choice of basis, which is usually a generic basis in the case of white noise, or the basis derived from the Karhunen–Loève expansion (e.g., [**13**, **20**]). We point out that, at least in the case of a self-adjoint operator $\mathcal{A}$, the choice of new basis should be related to the eigenfunctions of $\mathcal{A}$ (see section 3), rather than to the basis arising from the Karhunen–Loève expansion of the noise. Interestingly, [**11**, **22**] specifically chose to use a basis similar to the point forcing basis, but this was only to expedite the use of the finite element method. In [**11**], the stochastic term was handled by Monte Carlo simulation, and L_2-convergence properties of the solutions were studied. To the best of our knowledge, direct comparison of two bases has not received as much attention.

2. Change of Wiener chaos basis

We introduce the framework that will lead up to the proposed change of Wiener chaos basis idea. Let $U \subset \mathbb{R}^d$ be an open bounded domain. Let $-\mathcal{A}$ be a positive definite self-adjoint elliptic operator of order $2m$, equipped with either periodic or zero Dirichlet boundary conditions. (In the case of periodic boundary conditions, the domain U will be a torus $\mathbb{T}^d$.) We assume the dimensionality condition

$$2m/d > 1/2. \tag{5.4}$$

It is well known that $-\mathcal{A}$ has eigenfunctions $\{\mathfrak{m}_i\}$ that form an orthonormal basis in $L_2(U)$, and the corresponding eigenvalues $\{\lambda_i\}$ behave asymptotically as [**58**]

$$\lambda_i \sim i^{2m/d}. \tag{5.5}$$

We will refer to $\{\mathfrak{m}_i\}$ as the $\mathcal{A}$-*eigenfunction basis* in $L_2(U)$.

As an unbounded positive definite self-adjoint operator on $L_2(U)$, $-\mathcal{A}$ has a well-defined square root $\Lambda = \sqrt{-\mathcal{A}}$, which has domain $\mathcal{D}(\Lambda) = H^m_{\text{per}}$ or H^m_0. Then Λ induces a Hilbert scale which we denote by $H^\gamma_{\mathcal{A}}$, $\gamma \in \mathbb{R}$, with norms

$$\|\phi\|^2_{H^\gamma_{\mathcal{A}}} = \sum_{j=1}^{\infty} \left(\lambda_j^{1/2m}\right)^{2\gamma} \phi_j^2 \tag{5.6}$$

for ϕ of the form $\phi = \sum_{j=1}^{J} \phi_j \mathfrak{m}_j$, for some $J \in \mathbb{N}$ [**41**]. $H^\gamma_{\mathcal{A}}$ is the closure of the set of such ϕ in the norm $\|\cdot\|_{H^\gamma_{\mathcal{A}}}$. It can be shown that $H^\gamma_{\mathcal{A}}$ is equivalent to the usual Sobolev scale. In particular, the norm $\|\cdot\|_{H^{-2m}_{\mathcal{A}}}$ is equivalent to the Sobolev norm

$$\|\phi\|_{H^{-2m}} := \sup_{\psi \in H^{2m}_{\cdot}} \frac{|\langle \phi, \psi \rangle_{H^{-2m}, H^{2m}}|}{\|\psi\|_{H^{2m}}},$$

where we denoted $H^{2m}_{\cdot} = H^{2m}_{\text{per}}$ or H^{2m}_0 in the case of periodic or zero Dirichlet boundary conditions, respectively.

In order to define a localized basis, we consider a partition of the domain U. Let $N < \infty$ be arbitrary. Let $\mathcal{I} = \{I_i,\, i = 1, \dots, N\}$ be a partition of U into (small) nonoverlapping subsets with Lebesgue measure $|I_i| \sim 1/N$. We assume the family $\mathcal{I}$ is *quasi-uniform* in N. That is, there exist constants ρ_1, ρ_2 such that

$$\max_i r_i \leq (\rho_1 |U|)^{1/d} N^{-1/d},$$
$$\min_i \varepsilon_i \geq (\rho_2 |U|)^{1/d} N^{-1/d},$$

where $r_i = \operatorname{diam}(I_i)$ and ε_i is the radius of the largest sphere B_i contained in I_i. The quasi-uniform assumption implies nondegeneracy, i.e., that there exists ρ_3 such that $2\varepsilon_i \geq \rho_3 r_i^{(N)}$ for all i, N. It then follows that

$$\rho_- |U| N^{-1} \leq \min_i |I_i| \leq \max_i |I_i| \leq \rho_+ |U| N^{-1}$$

and

$$\tilde{\rho}_- (r_i)^d \leq |I_i| \leq \tilde{\rho}_+ (r_i)^d,$$

and hence

$$\varepsilon_i \sim r_i \sim N^{-1/d}.$$

We are now ready to introduce the two bases $\{n_i\}$ and $\{m_i\}$ that will be the focus of our comparative analysis.

(1) *Point forcing basis*:

$$n_i(x) = \frac{1}{\sqrt{|I_i|}} \mathbf{1}_{I_i}(x) \quad \text{for } i = 1, \dots, N, \tag{5.7}$$

and $\{n_i\}_{i=N+1}^{\infty}$ is any basis in $\mathcal{S}_N^{\perp}$, where $\mathcal{S}_N = \operatorname{span}\{n_i,\, i = 1, \dots, N\}$.

(2) *(Discrete) eigenfunction basis in $\mathcal{S}_N$*:

$$\begin{aligned} m_1 &= \mathfrak{m}_1, \\ m_i &= \frac{1}{Z_i} \left(\mathcal{P}_N \mathfrak{m}_i - \sum_{j=1}^{i-1} (\mathcal{P}_N \mathfrak{m}_i, m_j) m_j \right), \quad i = 2, \dots, N, \end{aligned} \tag{5.8}$$

where $\mathcal{P}_N$ is the L_2 projection onto $\mathcal{S}_N$ and Z_i is the normalization constant. In other words, $\{m_i\}$ is the Gram–Schmidt orthonormalization of the L_2 projections of the first N eigenfunction basis elements onto $\mathcal{S}_N$.

Many quantities considered in this chapter, such as the definitions of the two bases, depend on the parameter N. The limit as $N \to \infty$ is an object of study. However, in the rest of chapter, we will suppress explicitly writing this dependence on N if no ambiguity arises.

Define the Gaussian noise $\dot{W}_Q(x)$ on $L_2(U)$ by the Wiener chaos expansion

$$\dot{W}_Q(x) := \sum_{i \geq 1} \sigma_i n_i(x) \eta_i \tag{5.9}$$

where $\eta_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $\sigma_i \geq 0$, and the covariance operator Q^2 is defined by $Qn_i = \sigma_i n_i$ for $i = 1, 2, \ldots$ (see (2.3)). We do not restrict Q^2 to being a nuclear operator, but in the case where we desire $\dot{W}_Q$ to be a finite N-dimensional noise, we will assume that Range $Q \subset \mathcal{S}_N$. In this case, $\sigma_i = \sigma_i^{(N)}$ are nonzero only for $i = 1, \ldots, N$.

We consider the equation

$$\frac{\partial v}{\partial t} = \mathcal{A}v + \dot{W}_Q(x)G(t) \tag{5.10}$$

with zero initial conditions and either periodic or zero Dirichlet boundary conditions.[1] Here, $G(t)$ is a bounded function on $[0, T]$ satisfying

$$\frac{C_{G1}}{\lambda_j} \leq \int_0^t e^{-\lambda_j(t-s)} G(s) ds \leq \frac{C_{G2}}{\lambda_j}, \qquad \text{for } t \in (0, T], \tag{5.11}$$

for $j = 1, 2, \ldots$, where the constants C_{G1}, C_{G2} are independent of j and N, and C_{G2} is independent of t.

At this point, we introduce the related equation driven by an infinite dimensional Gaussian noise, which will be used for the error analysis in section 3.1. We assume for the sequence $\{\sigma_i^{(N)}, i = 1, \ldots, N\}$ that $\sup_N \sup_{i \leq N} \sigma_i^{(N)} < \infty$, and

$$\sigma_i^{(N)} \overset{N\to\infty}{\longrightarrow} \sigma_i^* \quad \text{uniformly for } i \leq N.$$

That is, $\forall \epsilon > 0$, $\exists N_0$ such that if $N > N_0$, then $|\sigma_i^{(N)} - \sigma_i^*| < \epsilon$, $\forall i \leq N$. For simplicity, we assume $\sigma_i^* = 1$ for all $i = 1, 2, \ldots$, but the results can be extended to any bounded sequence σ_i^*. The uniform convergence of $\{\sigma_i\}$ makes it possible to study the asymptotic behaviour of (5.10) through the related limiting SPDE driven by an infinite dimensional

[1] For simplicity, we will always assume zero initial conditions and periodic or zero Dirichlet boundary conditions, even when not explicitly stated. We also always take $x \in U$ and $t \in (0, T]$ for arbitrary $T < \infty$.

noise. To this end, we consider the SPDE with Gaussian white noise on $L_2(U)$, with $Q = I$ and $\dot{W}(x) = \sum_{i \geq 1} \xi_i \mathfrak{m}_i(x)$,

$$\frac{\partial u^*}{\partial t} = \mathcal{A}u^* + \dot{W}(x)G(t). \tag{5.12}$$

Equation (5.12) is solved in the triple $H^{-m} \hookrightarrow L_2 \hookrightarrow H^m$ and should be understood in the weak sense. Its propagator system is

$$\frac{\partial \hat{u}_i^*}{\partial t} = \mathcal{A}\hat{u}_i^* + \mathfrak{m}_i(x)G(t). \tag{5.13}$$

The equivalence of the propagator system to the weak solution can be shown. Moreover, there exists a solution u^* such that $u^*(t) \in L_2(\Omega; L_2(U))$ for each $t \in (0, T]$, and the energy $\mathcal{E}[u^*] := \|u^*(t)\|^2_{L_2(\Omega;L_2(U))}$ at any fixed $t \in (0, T]$ is finite. (See section 3.1.)

The framework to allow us to change the basis of the Wiener Chaos expansion is elementary. Direct computation gives that

$$Qm_j = \sum_{k=1}^{N} \Sigma_{jk} m_k, \quad \text{where } \Sigma_{jk} = \sum_{i=1}^{N} \sigma_i (n_i, m_j)(n_i, m_k).$$

The uniform convergence of $\{\sigma_i\}$ implies that $\Sigma_{jk} \longrightarrow \delta_{jk}$ as $N \to \infty$. We will write Σ_j in place of Σ_{jj}. Then there are two equivalent WCEs for $\dot{W}_Q$,

$$\dot{W}_Q(x) = \sum_{i=1}^{N} n_i(x)\sigma_i\eta_i = \sum_{i=1}^{N} m_i(x)\Sigma_i\xi_i,$$

where

$$\xi_i = \Sigma_i^{-1} \sum_{k=1}^{N} \sigma_k(n_k, m_i)\eta_k.$$

The ξ_is are identically distributed standard Gaussian random variables, but in general they are not independent. The covariance matrix $\rho = (\rho_{ij})_{i,j=1}^{N}$ is

$$\rho_{ij} := \mathbb{E}[\xi_i\xi_j] = \frac{\sum_{k=1}^{N} \Sigma_{ik}\Sigma_{kj}}{\Sigma_i\Sigma_j}.$$

Clearly, ρ is symmetric positive definite for each N, and we have $\rho_{ij} \longrightarrow \delta_{ij}$ as $N \to \infty$. In the case that $\sigma_i \equiv 1$ for all $i = 1, \ldots, N$, the ξ_is are i.i.d. standard Gaussian random variables, and the relationship between $\{\xi_i\}$ and $\{\eta_i\}$ reduces to the usual change of basis formula,

$$\xi_i = \sum_{i=1}^{N} (m_i, n_i)\eta_i. \tag{5.14}$$

We remark that the second expansion in (2) is, strictly speaking, not a Wiener chaos expansion, because the ξ_is are not orthogonal in $L_2(\Omega)$. It is a standard exercise to transform the expansion into an orthogonal expansion by a linear transformation of the ξ_is. However, we will not do that here, but instead just work directly with the linearly independent expansion (2).

By the change of basis formula, we write the solution of (5.10) in two expansions

$$u(x,t) = \sum_{i=1}^{N} \hat{v}_i(x,t)\eta_i = \sum_{i=1}^{N} \hat{u}_i(x,t)\xi_i. \tag{5.15}$$

Multiplying both sides of (5.10) by η_i or ξ_i, and taking expectation yields two equivalent propagator systems

$$\frac{\partial}{\partial t}\hat{v}_i = \mathcal{A}\hat{v}_i + \sigma_i n_i(x) G(t) \tag{5.16a}$$

$$\sum_{j=1}^{N} \rho_{ji} \frac{\partial}{\partial t}\hat{u}_j = \sum_{j=1}^{N} \rho_{ji} \left(\mathcal{A}\hat{u}_j + \Sigma_j m_j(x) G(t) \right) \tag{5.16b}$$

for $i = 1, \ldots, N$. Since ρ is invertible, (5.16b) reduces to a simpler system

$$\frac{\partial}{\partial t}\hat{u}_i = \mathcal{A}\hat{u}_i + \Sigma_i m_i(x) G(t) \tag{5.16b$'$}$$

Then the energy of (5.10), $\mathcal{E}[u] := \mathbb{E}\|u\|_{L^2}^2$, can be computed from the solutions of either system (5.16a) or (5.16b′) by a simple algebraic formula

$$\mathcal{E}[u] = \sum_{i=1}^{N} \|\hat{v}_i\|_{L_2}^2 = \sum_{i=1}^{N} (\hat{u}_i, \hat{u}_j)\rho_{ij} \tag{5.17}$$

In order to reduce the computational cost of computing the solutions of all N equations in the system (5.16a) or (5.16b′), we approximate the energy of the N-system by the energy of a truncated system. Truncating the systems (5.16) to $n < N$ equations means to consider the systems

$$\frac{\partial}{\partial t}\hat{v}_i = \mathcal{A}\hat{v}_i + \sigma_i n_i(x)G(t), \quad \text{for } i = 1, \dots, n \tag{5.18a}$$

$$\frac{\partial}{\partial t}\hat{u}_i = \mathcal{A}\hat{u}_i + \Sigma_i m_i(x)G(t), \quad \text{for } i = 1, \dots, n. \tag{5.18b}$$

System (5.18a) is the propagator system of

$$\frac{\partial}{\partial t}v^{(n)} = \mathcal{A}v^{(n)} + \dot{W}_{\mathcal{P}_n Q}(x)G(t) \tag{5.19}$$

where $\mathcal{P}_n$ the projection into span$\{n_i,\, i = 1, \dots, n\}$. System (5.18b) is the related system to

$$\frac{\partial}{\partial t}u^{(n)} = \mathcal{A}u^{(n)} + \dot{Z}_n(x)G(t) \tag{5.20}$$

where $\dot{Z}_n(x) = \sum_{i=1}^{n} m_i(x)\Sigma_i\xi_i$. Obviously, (5.19) and (5.20) are different SPDEs with different energies,

$$\mathcal{E}[v^{(n)}] = \sum_{i=1}^{n} \|\hat{v}_i\|_{L_2}^2 \;\neq\; \mathcal{E}[u^{(n)}] = \sum_{i,j=1}^{n} (\hat{u}_i, \hat{u}_j)\rho_{ij}.$$

The energies $\mathcal{E}[v^{(n)}]$ and $\mathcal{E}[u^{(n)}]$ will be taken as an approximation to the true energy $\mathcal{E}[u]$.

The absolute and relative errors of the approximations will be denoted as

$$R[v^{(n)}] := \mathcal{E}[u] - \mathcal{E}[v^{(n)}] = \sum_{i=n+1}^{N} \|\hat{v}_i\|_{L_2}^2, \quad \text{and} \quad \bar{R}[v^{(n)}] = \frac{R[v^{(n)}]}{\mathcal{E}[u]}$$
$$R[u^{(n)}] := \mathcal{E}[u] - \mathcal{E}[u^{(n)}] = \sum_{i=n+1}^{N} \|\hat{u}_i\|_{L_2}^2, \quad \text{and} \quad \bar{R}[u^{(n)}] = \frac{R[u^{(n)}]}{\mathcal{E}[u]}$$

for $n \leq N$. We will compare the performance of the two bases using the relative error of the approximate energy. Given an allowable relative error r, let

$$n_P := \inf\{n : \bar{R}[v^{(n)}] < r\} \quad \text{and} \quad n_E := \inf\{n : \bar{R}[u^{(n)}] < r\}. \tag{5.21}$$

be the minimum truncation sizes that achieves the relative error r. Define the *improvement* of the eigenfunction basis over the point forcing basis as

$$n_P/n_C.$$

The improvement is an indication of the computational savings of using the eigenfunction basis for the relative error r.

3. Comparative error analysis and 1st order improvement

In the foregoing section, all the quantities depend on the number N of subdivisions of U. In this section, we study the asymptotic behavior as $N \to \infty$. We will formulate precise bounds on the relative error and compare the asymptotic behavior of the two bases as $N \to \infty$.

The main goal of this section is to show the 1*st order improvement* of the change of basis method, in the sense of the following theorem.

THEOREM 3.1. *Given a relative error $r \in (0,1)$, we have, at worst,* 1st order improvement *as $N \to \infty$.*

More precisely, there exist constants $0 < C_{0,min} < C_{0,max} \leq 1$, depending on r but independent of N, such that for every $C_0 \in [C_{0,min}, C_{0,max})$ there exists $N_0 = N_0(C_0) > 0$ such that

$$\frac{n_P}{n_E} \geq C_0 N$$

whenever $N > N_0$. Moreover, $N_0 \to \infty$ as $C_0 \uparrow C_{0,max}$.

Obviously, 1st order improvement is the best one can hope for, simply because $n_P \leq N$ and $n_E \geq 1$, so that $n_P/n_E \leq N$. The result of Theorem 3.1 states that the constant in front of the 1st order improvement can vary in an interval, with a larger constant holding for larger N.

A big part of the proof of Theorem 3.1 involves studying the decay in n of the relative errors $\bar{R}[v^{(n)}]$ and $\bar{R}[u^{(n)}]$. Theorem 3.1 follows easily from two key facts: first, that the relative error $\bar{R}[v^{(n)}]$ for the point forcing basis decays no faster than linearly; second, that the relative error $\bar{R}[u^{(n)}]$ for the eigenfunction basis decays no slower than superlinearly, on the order of $n^{-\alpha}$ with $\alpha > 0$. (See Figures 1 and 2 for illustration and motivation.) We make these two statements more precise in the following two propositions.

PROPOSITION 3.2. *For the solution of* (5.12), *define the relative error by*

$$\bar{R}[u^{*,(n)}] := \frac{\sum_{i=n+1}^{\infty} \|\hat{u}_i^*\|_{L_2}^2}{\mathcal{E}[u^*]}$$

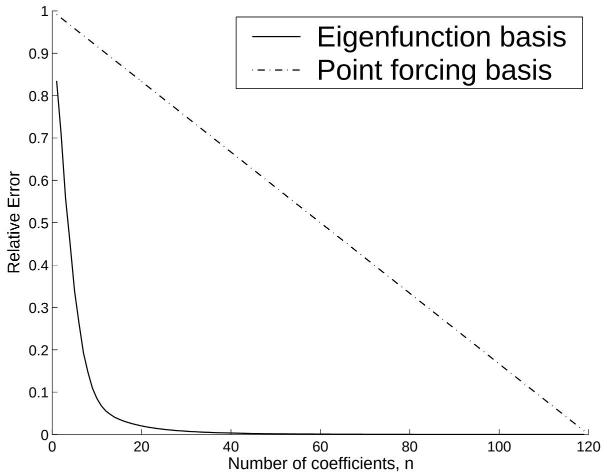

FIGURE 1. Relative errors incurred $\bar{R}[u^{(n)}]$ when the system is truncated to n coefficients, under the point forcing basis (dotted line) and the eigenfunction (cosine) basis (solid line). The convection-diffusion equation was used to produce this data.

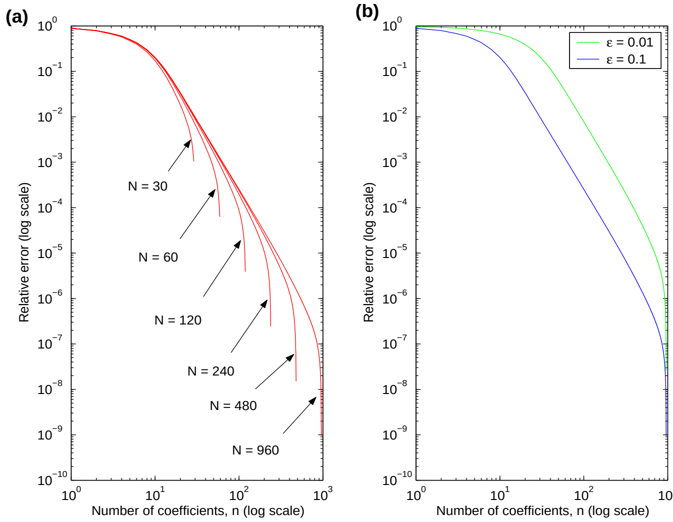

FIGURE 2. (a) Relative errors on log-log axes for increasing values of N, under the cosine basis for the heat equation. (b) Relative errors for two values of diffusion coefficients $\epsilon = 0.1, 0.01$. The graph for $\epsilon = 0.01$ lies above the graph for $\epsilon = 0.1$.

for $n = 1, 2, \ldots$. *Then*

$$\bar{R}[u^{*,(n)}] \sim n^{-4m/d+1}. \tag{5.22}$$

Given a relative error r*,*

$$n_0 := \inf\left\{n : \bar{R}[u^{*,(n)}] < r\right\} \sim r^{\frac{d}{d-4m}} \tag{5.23}$$

as $r \downarrow 0$.

Proposition 3.3. *There exists a constant* C *independent of* n *and* N *such that*

$$\bar{R}[v^{(n)}] \geq L(n) = 1 - nCN^{-1}, \tag{5.24}$$

where $L(n)$ *is a straight line passing through the point* $(0,1)$ *and with slope* $-CN^{-1}$, *which tends to* 0 *as* $N \to \infty$.

To show the decay behavior of the relative errors, we will focus on finding bounds on the L_2 norms of the solution modes $\hat{u}_i$ and $\hat{v}_i$. In order to be useful for explaining this contrasting behavior of the two bases, the bounds need to be sensitive to the localness or globalness of the basis and should provide accurate bounds on the solution modes. Error bounds involving $\|n_i\|^2_{L_2}$ and $\|m_i\|^2_{L_2}$ are clearly insensitive to the choice of basis, since both norms equal 1. Standard methods for estimating the time evolution of $\|u(t)\|^2_{L_2}$, such as those involving Gronwall's inequality, may also be inadequate. A case in point is the following.

Suppose $u(t)$ solves the heat equation

$$\frac{\partial u}{\partial t} = \Delta u + f(x), \quad x \in [0, X],$$

with zero initial conditions and periodic boundary conditions (cf. section 4.1). Also assume that $u(t)$ has periodic derivative. Then

$$\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|^2_{L_2} &= \int_U u(x,t)u_t(x,t)\,dx \\
&\leq -\|u_x(t)\|^2_{L_2} + \|u(t)\|_{H^1}\,\|f\|_{H^{-1}} \\
&\leq -\|u_x(t)\|^2_{L_2} + \left(\|u(t)\|^2_{L_2} + \|u_x(t)\|^2_{L_2}\right) + \frac{1}{4}\|f\|^2_{H^{-1}} \\
&= \|u(t)\|^2_{L_2} + \frac{1}{4}\|f\|^2_{H^{-1}}.
\end{aligned}$$

Thus, by Gronwall's inequality,

$$\|u(t)\|^2_{L_2} \leq \frac{te^{2t}}{4}\,\|f\|^2_{H^{-1}} = C(t)\|f\|^2_{H^{-1}}.$$

Using n_i or the cosine basis m_i in place of f, the energy of each mode is bounded by

$$\begin{aligned}\|\hat{u}_i(t)\|_{L_2}^2 &\leq C(t)\|m_i\|_{H^{-1}}^2 \leq C(t)\left(\frac{X}{(i-1)\pi}\right)^2,\\ \|\hat{v}_i(t)\|_{L_2}^2 &\leq C(t)\|n_i\|_{H^{-1}}^2 \leq C(t).\end{aligned}$$

Then the absolute error of the energy estimate of an n-equation truncated system (for fixed N) decays on the order of

$$R[u^{(n)}] \sim \mathcal{O}\left(\frac{1}{n} - \frac{1}{N}\right), \quad R[v^{(n)}] \sim \mathcal{O}\left(\frac{N-n}{N}\right).$$

The estimate for $R[v^{(n)}]$ is consistent with numerical results, but we will establish this result in more generality. But the estimate for $R[u^{(n)}]$ is merely an upper bound and does not predict the actual $\mathcal{O}\left(n^{-3}\right)$ decay (see Figure 2).

3.1. Fourier techniques for the eigenfunction basis. The Fourier expansion is an effective technique for obtaining exact error estimates. We begin by considering the limiting infinite dimensional equation (5.12) with covariance operator $Q = I$, and its propagator system (5.13). The solution of (5.12) has the Wiener chaos expansion:

$$u^*(t) = \sum_{i=1}^{\infty} \hat{u}_i^*(t)\xi_i = \sum_{i,j=1}^{\infty} \hat{u}_{ij}^*(t)\mathfrak{m}_j\xi_i,$$

where $\hat{u}_{ij}^* = (\hat{u}_i^*, \mathfrak{m}_j)$ are the Fourier coefficients of $\hat{u}_i^*$ with respect to the $\mathcal{A}$-eigenfunction basis in $L_2(U)$. Note that only the low order modes $\{\xi_i\mathfrak{m}_j\}_{i,j\geq 1}$ are nonzero because the noise appears additively. From the propagator system (5.13), the Fourier coefficients solve the decoupled system of ODEs

$$\begin{cases}\frac{\mathrm{d}}{\mathrm{d}t}\hat{u}_{ij}^* = -\lambda_j\hat{u}_{ij}^* + \delta_{ij}G(t),\\ \hat{u}_{ij}^*(0) = 0\end{cases} \tag{5.25}$$

for $i, j = 1, 2, \ldots$. The solution satisfies

$$\delta_{ij}\frac{C_{G1}}{\lambda_j} \leq \hat{u}_{ij}^*(t) = \delta_{ij}\int_0^t e^{-\lambda_j(t-s)}G(s)\,ds \leq \delta_{ij}\frac{C_{G2}}{\lambda_j}.$$

In other words, all the energy is concentrated on the modes $\{\xi_i \mathfrak{m}_i\}$, and

$$C_{G1}^2 \sum_{i=1}^{\infty} \frac{1}{\lambda_i^2} \leq \mathbb{E}\|u^*(t)\|_{L_2(U)}^2 \leq C_{G2}^2 \sum_{i=1}^{\infty} \frac{1}{\lambda_i^2},$$

where the summations converge because of the asymptotic behavior of the eigenvalues (5.5) and the dimensionality condition (5.4). It follows that the dimensionality condition (5.4) is necessary and sufficient for $u^*(t) \in L_2(\Omega; L_2(U))$ to be square integrable.

The consequence of this computation is that we have precise asymptotic estimates for the truncation error:

$$R[u^{*,(n)}] := \sum_{i=n+1}^{\infty} \|\hat{u}_i^*\|_{L_2}^2 = \sum_{i=n+1}^{\infty} (\hat{\hat{u}}_{ii}^*)^2 \sim n^{-4m/d+1}. \tag{5.26}$$

The asymptotics in (5.26) obviously hold for $\bar{R}[u^{*,(n)}] = R[u^{*,(n)}]/\mathcal{E}[u^*]$ as well, and we obtain (5.22) in Proposition 3.2. Equation (5.23) then follows by taking the function inverse of the asymptotic bounds in (5.22). Thus,

$$C r^{\frac{d}{d-4m}} \leq n_0 \leq C' r^{\frac{d}{d-4m}} \tag{5.27}$$

for r sufficiently small.

We now look at the finite dimensional model, $N < \infty$, with nuclear Q and $\dot{W}_Q$ defined in (5.9). $\dot{W}_Q$ can be viewed as a finite dimensional approximation of $\dot{W}$, and we have that $\mathcal{E}[u] \to \mathcal{E}[u^*]$.

Instead of (5.25), the relevant system of ODEs for (5.10) corresponding to the discrete eigenfunction basis in $\mathcal{S}_N$ comes from (5.16b′):,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} \hat{\hat{u}}_{ij} = -\lambda_j \hat{\hat{u}}_{ij} + \Sigma_i(m_i, \mathfrak{m}_j) G(t), \\ \hat{\hat{u}}_{ij}(0) = 0 \end{cases} \tag{5.28}$$

for $i = 1, \ldots, N$, $j = 1, 2, \ldots$. The solution is

$$\hat{\hat{u}}_{ij}(t) = \Sigma_i(m_i, \mathfrak{m}_j) \int_0^t e^{-\lambda_j (t-s)} G(s)\, ds.$$

The behavior of the truncation error of system (5.28) can be expected to be close to that of (5.25). Indeed, since the Gram–Schmidt orthonormalized elements m_i in (5.8) are finite sums of projections $\mathcal{P}_N \mathfrak{m}_k$, and since $\mathcal{P}_N \mathfrak{m}_k \longrightarrow \mathfrak{m}_k$ in L_2 as $N \to \infty$, it follows that

$m_i \longrightarrow \mathfrak{m}_i$ in L_2 as $N \to \infty$ for each $i = 1, 2, \ldots$. For any fixed j, $(m_i, \mathfrak{m}_j) \longrightarrow \delta_{ij}$ also. Thus, by (5.11) and the dominated convergence theorem,

$$\begin{aligned}(\hat{u}_i, \hat{u}_{i'})\rho_{ii'} &= \sum_{j\geq 1} \rho_{ii'}\Sigma_i\Sigma_{i'}(m_i, \mathfrak{m}_j)(m_{i'}, \mathfrak{m}_j)\left(\int_0^t e^{-\lambda_j(t-s)}G(s)ds\right)^2 \\ &\longrightarrow \delta_{ii'}\sum_{j\geq 1}\delta_{ij}\delta_{i'j}\left(\int_0^t e^{-\lambda_j(t-s)}G(s)ds\right)^2 = \delta_{ii'}\|\hat{u}_i^*\|^2_{L_2(U)}, \quad \text{as } N \to \infty\end{aligned}$$

for all $i, i' = 1, 2, \ldots$. Since $\mathcal{E}[u] \longrightarrow \mathcal{E}[u^*]$, it follows that

$$R[u^{(n)}] = \mathcal{E}[u] - \mathcal{E}[u^{(n)}] \overset{N\to\infty}{\longrightarrow} R[u^{*,(n)}] = \mathcal{E}[u^*] - \mathcal{E}[u^{*,(n)}] \tag{5.29}$$

for every n. In particular, we deduce that for fixed truncation size n, the relative error incurred by truncating the large size N system must tend to the relative error incurred by truncating the infinite system.

We do not assert that $\bar{R}[u^{(n)}] = \mathcal{O}\left(n^{-\frac{4m}{d}+1}\right)$. Nonetheless, similar to n_0 in (5.23), an asymptotic result for the minimum truncation size to achieve relative error r for the discrete eigenfunction basis, easily follows.

Proposition 3.4. *Let $n_E = n_E(N, r)$ be defined as in (5.21). Then there is some $r^* \in (0, 1)$ such that for any relative error $r < r^*$, there exists $N(r)$ such that*

$$Cr^{\frac{d}{d-4m}} \leq n_E \leq C'r^{\frac{d}{d-4m}}$$

whenever $N \geq N(r)$. The constants C, C' are independent of r, N.

Proof. From (5.27), there exists r^* such that

$$Cr^{\frac{d}{d-4m}} \leq n_0(r) \leq C'r^{\frac{d}{d-4m}}$$

whenever $r < r^*$. Fix $\delta \in (0, 1)$. For any $r < r^*/(1+\delta)$, choose $N(r)$ such that

$$\left|\bar{R}[u^{(n)}] - \bar{R}[u^{*,(n)}]\right| < \delta r$$

holds for both $n = n_0((1+\delta)r)$ and $n = n_0((1-\delta)r)$. Then

$$\bar{R}[u^{(n)}]|_{n=n_0((1+\delta)r)} > r \quad \text{and} \quad \bar{R}[u^{(n)}]|_{n=n_0((1-\delta)r)} < r$$

and

$$n_0((1+\delta)r) < n_E \le n_0((1-\delta)r).$$

Hence

$$C(1+\delta)^{\frac{d}{d-4m}} r^{\frac{d}{d-4m}} \le n_E \le C'(1-\delta)^{\frac{d}{d-4m}} r^{\frac{d}{d-4m}},$$

and the result follows with $r^*/(1+\delta)$ in place of r^*. $\square$

3.2. $\boldsymbol{H^{-2m}}$ norm estimates for the point forcing basis. For the error analysis for the point forcing basis, similar computations for the system of ODEs for the point forcing basis coming from (5.16b′) show that $\hat{\hat{v}}_{ij} := (\hat{v}_i, \mathfrak{m}_j)$ satisfies

$$\hat{\hat{v}}_{ij}(t) = \sigma_i(n_i, \mathfrak{m}_j) \int_0^t e^{-\lambda_j(t-s)} G(s)\, ds.$$

Clearly, the energy of the system is *not* concentrated on $\{\hat{\hat{v}}_{ii},\, i = 1, \dots, N\}$, and

$$\|\hat{v}_i\|_{L_2}^2 = \sum_{j=1}^{\infty} \sigma_i^2 (n_i, \mathfrak{m}_j)^2 \left(\int_0^t e^{-\lambda_j(t-s)} G(s)\, ds \right)^2. \tag{5.30}$$

We have the following lemma.

Lemma 3.5. *Let n_i be defined in (5.7) for $i = 1, \dots, N$. Then we have the bounds*

$$C_1 N^{-2m/d} \le \|n_i\|_{H^{-2m}} \le C_2 N^{-1/2},$$

where C_1, C_2 are independent of i and N.

Proof. For the lower bound, consider the mollifier ζ_ε with support in $B(0,\varepsilon)$, and let α_i be the center of the largest sphere B_i contained in I_i with radius ε_i. Then, denoting $H^{2m}_{\cdot} = H^{2m}_{\text{per}}$ or H^{2m}_0,

$$\begin{aligned}
\|n_i\|_{H^{-2m}} &= \sup_{\psi \in H^{2m}_{\cdot}(U)} \frac{|\langle n_i, \psi \rangle|}{\|\psi\|_{H^{2m}(U)}} \\
&\ge \|\zeta_{\varepsilon_i}(\cdot - \alpha_i)\|_{H^{2m}(U)}^{-1} \int_U n_i(x) \zeta_{\varepsilon_i}(x - \alpha_i)\, dx \\
&= \|\zeta_{\varepsilon_i}\|_{H^{2m}(\mathbb{R}^d)}^{-1} (n_i * \zeta_{\varepsilon_i})(\alpha_i) \ge C N^{-2m/d}.
\end{aligned}$$

The last inequality holds because it can be computed that $\|\zeta_{\varepsilon_i}\|_{H^{2m}} \sim \varepsilon_i^{-(2m+d/2)}$ and $(n_i * \zeta_{\varepsilon_i})(\alpha_i) = n_i(\alpha_i) \sim N^{1/2} \sim \varepsilon_i^{-d/2}$.

For the upper bound,

$$\begin{aligned}\|n_i\|_{H^{-2m}} = \sup_{\psi\in H^{2m}_{\cdot}} \frac{|\langle n_i,\psi\rangle|}{\|\psi\|_{H^{2m}}} &\le |I_i|^{1/2} \sup_{\psi\in H^{2m}_{\cdot}} \frac{\frac{1}{|I_i|}\int_{I_i}|\psi|\,dx}{\|\psi\|_{H^{2m}}}\\ &\le CN^{-1/2},\end{aligned}$$

where the C is independent of i and N, since by the Sobolev embedding every ψ belonging to $H_0^{2m}(U)$ or $H_{\mathrm{per}}^{2m}(U)$ also belongs to $C^{0,1/2}(\bar{U})$. □

Corollary 3.6. *For each $i = 1,\ldots,N$, we have the bounds*

$$C_3N^{-4m/d} \le \sigma_i^{-2}\|\hat{v}_i\|_{L_2}^2 \le C_4N^{-1},$$

where C_3, C_4 are independent of i and N.

Proof. From the definition of the $H^\gamma_{\mathcal{A}}$ norm, (5.6),

$$\|\hat{v}_i\|_{L_2}^2 \ge \sum_{j=1}^{\infty}\left(\sigma_i(n_i,\mathfrak{m}_j)\frac{C_{G1}}{\lambda_j}\right)^2 = \sigma_i^2C_{G1}^2\|n_i\|^2_{H_{\mathcal{A}}^{-2m}}$$

and similarly

$$\|\hat{v}_i\|_{L_2}^2 \le \sum_{j=1}^{\infty}\left(\sigma_i(n_i,\mathfrak{m}_j)\frac{C_{G2}}{\lambda_j}\right)^2 = \sigma_i^2C_{G2}^2\|n_i\|^2_{H_{\mathcal{A}}^{-2m}}.$$

The result follows by the equivalence of the $H^\gamma_{\mathcal{A}}$ norms and the Sobolev norms, and from Lemma 3.5. □

The lower bound in Corollary 3.6 gives another way to see that the solution of the finite system will not converge to a square integrable of the infinite system if the dimensionality condition (5.4) is not met. This lower bound also gives a lower bound for the relative error of the point forcing basis. Interestingly, a more informative lower bound on the relative error can be derived from the upper bound in Corollary 3.6:

$$\begin{aligned}\bar{R}[v^{(n)}] &= \frac{\mathcal{E}[u]-\sum_{i=1}^n\|\hat{v}_i\|_{L_2}^2}{\mathcal{E}[u]}\\ &\ge \frac{\mathcal{E}[u]-C_4N^{-1}n\big(\sup_N\sup_{i\le N}\sigma_i^2\big)}{\mathcal{E}[u]} \qquad (5.31)\\ &= 1-n\frac{C_5}{N\mathcal{E}[u]} =: L(n).\end{aligned}$$

Since the constant C_4 is independent of n and N, the relative error is bounded from below by a straight line $L(n)$ passing through the point $(0, 1)$, and with slope $-C_5/(N\mathcal{E}[u])$ which tends to 0 as $N \to \infty$. We have just shown Proposition 3.3.

We now prove Theorem 3.1.

Proof of Theorem 3.1. From (5.31), the linear lower bound $L(n)$ attains relative error r for $n \geq n_L$, where

$$n_L = \frac{(1-r)\mathcal{E}[u]}{C_5} N = \tilde{C}(N)N$$

is the value such that $L(n_L) = r$. Then, since $\bar{R}[u^{(n)}] \geq L(n)$,

$$n_P \geq n_L = \tilde{C}(N)N.$$

$\tilde{C}(N)$ depends on N because $\mathcal{E}[u]$ depends on N. We next show a series of estimates for $\tilde{C}(N)$ to remove the dependence on N. First, $\mathcal{E}[u_{\{N\}}] \geq \mathcal{E}[u_{\{N=1\}}]$ for all N, so

$$\tilde{C}(N) \geq \tilde{C}(1)$$

for all N. Now, since $\mathcal{E}[u] \to \mathcal{E}[u^*]$, for any $\epsilon \in (0, \mathcal{E}[u^*] - \mathcal{E}[u_{\{N=1\}}])$, there exists $N(\epsilon)$ such that $\mathcal{E}[u] \geq \mathcal{E}[u^*] - \epsilon$. So

$$\tilde{C}(N) \geq \frac{(1-r)(\mathcal{E}[u^*] - \epsilon)}{C_4}$$

whenever $N > N(\epsilon)$. Denote $\tilde{C}(\infty) = \frac{(1-r)\mathcal{E}[u^*]}{C_5}$. As ϵ ranges from 0 to $\mathcal{E}[u^*] - \mathcal{E}[u_{\{N=1\}}]$, the right-hand side of the last inequality ranges from $\tilde{C}(\infty)$ to $\tilde{C}(1)$. Clearly, $N(\epsilon)$ increases to ∞ as $\epsilon \downarrow 0$. In other words, for any $C \in [\tilde{C}(1), \tilde{C}(\infty))$, there exists $N(C)$ such that

$$n_P \geq \tilde{C}(N)N \geq CN$$

whenever $N > N(C)$. Moreover, $N(C)$ increases to ∞ as $C \uparrow \tilde{C}(\infty)$.

For n_E, recall $n_0 = \inf\{n : \bar{R}[u^{*,(n)}] < r\}$, (5.23). Let $\epsilon_0 = r - \bar{R}[u^{*,(n_0)}] > 0$. From (5.29), $\bar{R}[u^{(n_0)}] \to \bar{R}[u^{*,(n_0)}]$ as $N \to \infty$. So there exists $N(n_0) > 0$ such that

$$\bar{R}[u^{(n_0)}] < \bar{R}[u^{*,(n_0)}] + \epsilon_0 = r$$

whenever $N > N(n_0)$. Hence, $n_E \leq n_0$ if $N > N(n_0)$.

Combining the two inequalities for n_P, n_E,

$$\frac{n_P}{n_E} \geq \frac{CN}{n_0} = C_0 N$$

whenever $N > N_0(C_0) := \max\{N(C), N(n_0)\}$. Hence, $C_0 \in [C(1)/n_0, C(\infty)/n_0)$ and $N_0 \to \infty$ as $C_0 \uparrow C(\infty)/n_0$. □

The next result gives upper and lower bounds on the improvement in terms of the relative error r.

Corollary 3.7. *There exist $r^* \in (0,1)$ and constants $0 < C_{*,min} < C_{*,max} \leq C_* \leq 1$ such that, for every $r < r^*$ and every $C_0 \in [C_{*,min}, C_{*,max})$, there exists $N_0 = N_0(r, C_0) > 0$ such that*

$$C_0 r^{-\frac{d}{d-4m}} N \leq \frac{n_P}{n_E} \leq C_* r^{-\frac{d}{d-4m}} N$$

whenever $N > N_0$. Moreover, $N_0 \to \infty$ as $C_0 \uparrow C_{*,max}$ or as $r \downarrow 0$.

Proof. In the proof of Theorem 3.1, the inequalities hold if we replace $\tilde{C}(N)$ with $\tilde{C}_*(N) := \frac{(1-r^*)\mathcal{E}[u]}{C_5}$ so that, for any $C \in [\tilde{C}_*(1), \tilde{C}_*(\infty))$, there exists $N(C)$ such that

$$n_P \geq \tilde{C}_*(N)N \geq CN$$

whenever $N > N(C)$. Also $N(C)$ increases to ∞ as $C \uparrow \tilde{C}_*(\infty)$. From Proposition 3.4,

$$\frac{n_P}{n_E} \geq \frac{CN}{C' r^{\frac{d}{d-4m}}} = C_0 N r^{-\frac{d}{d-4m}}$$

whenever $N > N_0(r, C_0)$.

Also from Proposition 3.4, and since $n_P \leq N$,

$$\frac{n_P}{n_E} \leq \frac{N}{C r^{\frac{d}{d-4m}}} = C_* N r^{-\frac{d}{d-4m}}.$$

□

If $r_1 < r_2$, then $r_1^{-\frac{d}{d-4m}} < r_2^{-\frac{d}{d-4m}}$, so Corollary 3.7 indicates that one would expect a slower convergence to 1st order improvement for a smaller relative error. This observation is in accordance with Trend (T3) in the numerical simulations. We also note that the interval endpoints in Theorem 3.1 and Corollary 3.7 are inversely proportional to C_4 from Corollary 3.6, which is in turn inversely proportional to the norm of $\mathcal{A}$. This point is corroborated by the numerical result that showed that the improvement is better for a larger diffusivity constant (cf. Trend (T1)).

3.3. The non–self-adjoint case. If $\mathcal{A}$ is not self-adjoint or not positive definite, precise bounds on the error decay such as those obtained in the previous section may not be readily available. But, under additional assumptions, we can still deduce certain asymptotic results similar to the positive definite self-adjoint case, including the result of 1st order improvement. To see this, let us consider again the SPDE (5.12) in the triple $H^m \hookrightarrow L_2 \hookrightarrow H^{-m}$, where we assume $\mathcal{A}$ is a $2m$th order non–self-adjoint elliptic operator. Also assume a more stringent dimensionality condition:

$$m/d > 1/2. \tag{5.32}$$

We decompose $\mathcal{A} = \mathcal{A}_0 + \mathcal{A}_1$ into the symmetric part $\mathcal{A}_0 = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*)$ and the skew-symmetric part $\mathcal{A}_1 = \frac{1}{2}(\mathcal{A} - \mathcal{A}^*)$, and we assume that $-\mathcal{A}_0$ is positive definite. Then $-\mathcal{A}_0$ generates an eigenfunction basis $\{\mathfrak{m}_i\}$ with eigenfunctions $\{\lambda_i\}$ satisfying (5.5). Similarly to (5.6), $\mathcal{A}_0$ defines a scale of Hilbert spaces $H^\gamma_{\mathcal{A}_0}$, with norm $\|\phi\|^2_{H^\gamma_{\mathcal{A}_0}} = \sum_{j=1}^\infty (\phi, \mathfrak{m}_j)^2 \lambda_j^{\gamma/m}$, that is equivalent to the Sobolev scale H^γ.

In the infinite dimensional case with white noise (5.12), the existence and uniqueness of the solution u^* is shown in [**57**] because the asymptotics of the eigenvalues (5.5) and the new dimensionality condition (5.32) imply that $\dot{W} \in L_2(\Omega; H^{-m}(U))$. Applying the usual deterministic parabolic estimates to the propagator system (5.13), we have that $\hat{u}^*_i(t)$ is continuous in t, and

$$\|\hat{u}^*_i(t)\|^2_{L_2(U)} \le C\|G\|^2_{L_2(0,T)}\|\mathfrak{m}_i\|^2_{H^{-m}} \le C'\|\mathfrak{m}_i\|^2_{H^{-m}_{\mathcal{A}_0}} \le C'\lambda_i^{-1}.$$

for all $t \in (0, T]$. Then we have a result analogous to (but weaker than) Proposition 3.2. For the error

$$R[u^{*,(n)}] := \sum_{i=n+1}^{\infty} \|\hat{u}^*_i\|^2_{L_2} \le Cn^{-2m/d+1}, \tag{5.33}$$

and for $n_0 := \min\{n : \bar{R}[u^{*,(n)}] < r\}$,

$$n_0 \le Cr^{\frac{d}{d-2m}}.$$

In the finite dimensional case (5.10), we again have $\mathcal{E}[u] \to \mathcal{E}[u^*]$. For the discrete eigenfunction basis,

$$\left| \|\hat{u}_i\|_{L_2}^2 - \|\hat{u}_i^*\|_{L_2}^2 \right| = \left| \|\hat{u}_i\|_{L_2} - \|\hat{u}_i^*\|_{L_2} \right| \left(\|\hat{u}_i\|_{L_2} + \|\hat{u}_i^*\|_{L_2} \right) \leq C \|\hat{u}_i - \hat{u}_i^*\|_{L_2}$$
$$\leq C' \|\Sigma_i m_i - \mathfrak{m}_i\|_{H^{-m}} \overset{N \to \infty}{\longrightarrow} 0$$

for each $i \leq N$, and so

$$\left| \mathcal{E}[u^{(n)}] - \mathcal{E}[u^{*,(n)}] \right| = \left| \sum_{i=1}^{n} \|\hat{u}_i\|_{L_2}^2 - \|\hat{u}_i^*\|_{L_2}^2 \right| \longrightarrow 0.$$

Hence, $\bar{R}[u^{(n)}] \longrightarrow \bar{R}[u^{*,(n)}]$ as $N \to \infty$ for each n. For the point forcing basis,

$$\|\hat{v}_i(t)\|_{L_2(U)}^2 \leq C\sigma_i^2 \|n_i\|_{H^{-m}}^2 \leq C' N^{-1},$$

where the last inequality follows by an argument similar to the upper bound in Lemma 3.5. The proof of Theorem 3.1 follows through identically, so the statement of 1st order improvement applies to the non–self-adjoint case as well, provided (5.32) holds.

However, this argument by parabolic estimates works only when (5.32) holds; the behavior when $1/4 < m/d \leq 1/2$, which was covered in the self-adjoint case, is not addressed here. This should not be a surprise because the parabolic estimates are essentially Gronwall-type estimates, which we have noted in the beginning of section 3 to give suboptimal error bounds. The main difference between the two analyses is the estimation of the forcing terms in the H^{-m} norm in the parabolic estimate case, rather than the H^{-2m} norm in the self-adjoint case. Hence, the parabolic estimates provide only upper bounds on $R[u^{*,(n)}]$ that are $\mathcal{O}\left(n^{-2m/d+1}\right)$, which is less favorable and less precise than the $o(n^{-4m/d+1})$ decay found in the self-adjoint case. Nonetheless, we conjecture that the asymptotic behavior of $R[u^{*,(n)}]$ should in principle be dominated by the self-adjoint part $\mathcal{A}_0$, even though this is not reflected with the parabolic estimates (see section 4.3).

4. Examples and simulations

The change of basis strategy is applied to some simple equations to illustrate the efficiency of the point forcing and cosine bases, (5.34), (5.35), for approximating the energy of the systems (5.16a,b). One of the equations considered is the heat equation, for which we will observe results that corroborate the analysis in section 3. Although the analysis is asymptotic in nature, the 1st order convergence is already clear even for not-too-large system sizes. We also present numerical results for convection-diffusion equations that share very similar comparative properties to the pure diffusion case and extend the discussion to the connection with the pure convection equation.

For our numerical simulations, we take the interval $U = [0, X]$, and we let $\mathcal{I}_N = \{I_i, i = 1, \dots, N\}$ be a uniform partition of U into intervals of length X/N. We consider the operator with $\mathcal{A} = \epsilon\Delta$ with periodic boundary conditions, whose eigenfunctions are the usual cosine basis. ϵ is a small diffusivity coefficient. The two bases on $\mathcal{S}_N := \mathrm{span}\{n_i, i = 1, \dots, N\}$ are the following:

(1) *Point forcing basis*:

$$n_i(x) = \sqrt{\frac{N}{X}}\mathbf{1}_{I_i}(x) \quad \text{for } i = 1, \dots, N. \tag{5.34}$$

(2) *Cosine basis in $\mathcal{S}_N$*: The eigenfunction basis in $L_2([0, X])$ is the usual cosine basis:

$$\begin{aligned} \mathfrak{m}_1(x) &= \sqrt{\frac{1}{X}}, \\ \mathfrak{m}_i(x) &= \sqrt{\frac{2}{X}}\cos\left(\frac{(i-1)\pi x}{X}\right), \quad i = 2, 3, \dots. \end{aligned}$$

Define the *cosine basis in $\mathcal{S}_N$* as the Gram–Schmidt orthonormalization of the L_2 projections of the first N cosine basis elements onto $\mathcal{S}_N$:

$$\begin{aligned} m_1 &= \mathfrak{m}_1, \\ m_i &= \frac{1}{Z_i}\left(\mathcal{P}_N\mathfrak{m}_i - \sum_{j=1}^{i-1}(\mathcal{P}_N\mathfrak{m}_i, m_j)m_j\right), \end{aligned} \tag{5.35}$$

where $\mathcal{P}_N$ is the L_2 projection onto $\mathcal{S}_N$ and Z_i is the normalization constant.

In this example, we take $G(t) = 1$, and take the covariance $Q = \mathcal{P}_N$, so that $\sigma_i \equiv \sigma_i^* \equiv 1$ for all $i = 1, 2, \dots$. Then $\Sigma_i = 1$ for all $i = 1, \dots, N$ and the two WCEs for $\dot{W}_N$ are

$$\dot{W}_N(x) = \sum_{i=1}^{N} n_i(x)\eta_i = \sum_{i=1}^{N} m_i(x)\xi_i$$

where η_i and ξ_i are related by the usual change of basis formula (5.14). As a side note, since $\dot{W}_N$ is a finite truncation of the white noise $\dot{W}$, it is well-known that we can give ξ_i

and η_i precise expressions:

$$\xi_i := \int_U m_i(x)\, dW(x) \quad \text{and} \quad \eta_i := \int_U n_i(x)\, dW(x),$$

where $W(x)$ is a Brownian motion on U and from which the change of basis formula can be checked by direct computation.

We study the equation

$$\frac{\partial u}{\partial t} = \epsilon \Delta u + \dot{W}_N(x) \tag{5.36}$$

with zero initial conditions and periodic boundary conditions. Equations (5.15), (5.16), and (5.17) hold.

Note that, strictly speaking, the analysis of section 3 does not apply to (5.36) because $-\Delta$ with periodic boundary conditions has an eigenvalue $\lambda_1 = 0$ and thus is not strictly positive definite. Nonetheless, we can still apply the ideas from section 3 to obtain analogous results for the error decay and 1st order improvement. Equation (5.11) holds for $j = 2, 3, \dots,$ while for $j = 1$, $\lambda_1 = 0$,

$$\int_0^t e^{-\lambda_1(t-s)}\, ds = t,$$

so equations (5.26), (5.27) and (5.29) hold also, as do Propositions 3.2 and 3.4. For the point forcing basis, the analogous result to Corollary 3.6 is

$$C_3 N^{-1} \leq \|\hat{v}_i\|_{L_2}^2 \leq C_4 N^{-1}.$$

Indeed, the lower bound is

$$\|\hat{v}_i(t)\|_{L_2}^2 \geq |\hat{\hat{v}}_{i,1}(t)|^2 = (n_i, \mathfrak{m}_1)^2 t^2 = N^{-1} t^2.$$

For the upper bound, we integrate by parts backwards twice to find

$$(n_i, \lambda_j^{-1} \mathfrak{m}_j) = (-1)^{j-1} \lambda_j^{-1} f_i'(X) + (f_i, \mathfrak{m}_j), \quad j \geq 2, \tag{5.37}$$

for some function $f_i(x)$ such that $f_i'' = n_i$. (This step takes the place of invoking the H^{-2} norm of n_i.) It can be directly computed that

$$f_i(x) = \begin{cases} 0, & x \leq \frac{(i-1)X}{N}, \\ \frac{1}{2}\sqrt{\frac{N}{X}}\left(x - \frac{(i-1)X}{N}\right), & \frac{(i-1)X}{N} < x \leq \frac{iX}{N}, \\ \sqrt{\frac{X}{N}}\left(x + \frac{(\frac{1}{2}-i)X}{N}\right)^2, & x > \frac{iX}{N}, \end{cases}$$

so $f_i'(X) = \sqrt{X/N}$ and $\|f_i\|_{L_2}^2 = \cdots \leq \frac{17X^4}{15}\frac{1}{N}$. Squaring (5.37) and summing over j,

$$\|n_i\|_{H^{-2}}^2 = \sum_{j\geq 1} \lambda^{-2}(n_i, m_j)^2 \leq 2f'(X)^2 \sum_{j\geq 1} \lambda_j^{-2} + 2\|f_i\|_{L_2}^2 \leq CN^{-1}$$

Hence $\|\hat{v}_i\|_{L_2}^2 \leq C\|n_i\|_{H^{-2}}^2 \leq CN^{-1}$.

4.1. Heat equation. For the heat equation (5.36), we show in Figure 2(a) the relative error of the truncated system under the cosine basis for different values of N. We observe that, for each n, the relative error increases pointwise to a limit as $N \to \infty$. We assume that the $N = 960$ error plot is representative of the error in the limit as $N \to \infty$, at least for n not near 960. When $n > 10$, the relative error decays linearly on the log-log axes, with a gradient of ≈ -3; i.e., $\bar{R}[u^{(n)}] \sim \mathcal{O}\left(n^{-3}\right)$. This same order of decay is seen for $\epsilon = 0.01$ only when $n > 40$ (Figure 2(b)), and the actual relative error is larger than for $\epsilon = 0.1$. Both these orders of decay are consistent with (5.26) when $m = d = 1$.

In contrast, the relative error decays linearly on the linear axes for the point forcing basis (cf. Figure 1) and does not exhibit the same limiting behavior as the error plots for the cosine basis do. In fact, in this case of periodic boundary conditions, the relative error plot is simply a straight line of slope $-N^{-1}$ joining the points $(0, 1)$ and $(N, 0)$ because the energy of each $\|\hat{v}_i\|_{L_2}^2$ is equal. For a given level of relative error and for large values of N, n_P for the point basis scales on the order of $\mathcal{O}(N)$, whereas n_E for the cosine basis scales with $\mathcal{O}(1)$. As a result, this implies the 1st order convergence seen in Table 1.

Table 1(b) shows the improvements of the cosine basis for 5% error. We highlight several trends.

(T1) For fixed N, the improvement increases for larger ϵ. This increase is most significant for large N.

TABLE 1. Improvement n_P/n_E in the number of basis elements required to attain 5% error.

	(a) Convection-diffusion equation			(b) Heat equation	
N	$\epsilon = 0.1$	$\epsilon = 0.01$	$\epsilon = 0$	$\epsilon = 0.1$	$\epsilon = 0.01$
30	2.6364	2.4167	2.4167	1.8125	1.0741
60	4.2846	4.1429	3.8000	3.4118	1.3571
120	8.7692	7.6000	4.7917	6.3889	2.3
240	17.5385	12.6667	8.4815	12.7222	4.3019
480	35.0769	21.7619	15.7241	25.3889	8.4444
960	70.2308	43.4762	30.4333	50.6667	16.8889

(T2) We have 1*st order improvement*: doubling N increases the improvement by a factor that approaches double as N becomes large.

(T3) 1st order improvement is seen for a smaller error of 1% (data not shown), but the convergence to 1st order improvement is slower.

Numerical scheme. The discontinuous Galerkin dG(1) scheme with a 2nd order Runge–Kutta time stepping scheme [**12**] was used in this computation. For each number N of forcing terms, we took N spatial grid points and used $X = 2\pi$, $T = 0.5$. The simulations were also done using a fixed number of grid points (960 grid points) for all values of N, but little difference was found in the quantitative and qualitative behaviors of the estimates.

4.2. Convection-diffusion equations. We applied the same change of basis method for the stochastic convection-diffusion equation

$$\frac{\partial}{\partial t}u + bu_x = \epsilon u_{xx} + \dot{W}_N \tag{5.38}$$

with zero initial conditions and periodic boundary conditions. We performed simulations with constant convection speed $b_0 = 1.47$ and small diffusive coefficients $\epsilon = 0.01, 0.1$.

Figure 1 shows the behavior of the relative errors of the two bases on linear axes. Under the point forcing expansion, the relative error of the truncated system decays linearly in n, whereas the relative error under the cosine expansion decays superlinearly. The improvement is also found for varying sizes of the full system, $N = 30, 60, \ldots, 960$ (Table 1(a)).

4.3. Further remarks. As noted in section 3.3, it is not straightforward to deduce precise error estimates for general equations where $\mathcal{A}$ does not provide an eigenfunction basis. If the equation is simple enough, the error decay rate can be found from the explicit solution. In the case of (5.38),

$$\begin{aligned}\frac{\partial}{\partial t}\|\hat{u}_i\|_{L_2}^2 &= \int_U 2\hat{u}_i(-b\hat{u}_{i,x} + \epsilon\hat{u}_{i,xx} + m_i)\,dx \\ &= \int_U b_x\hat{u}_i^2 + 2\epsilon\hat{u}_i\hat{u}_{i,xx} + 2\hat{u}_i m_i\,dx.\end{aligned}$$

If $\epsilon = 0$,

$$\|\hat{u}_i(t)\|_{L_2}^2 = \|\hat{u}_i(0)\|_{L_2}^2 + 2\int_0^t\int_U \hat{u}_i m_i\,dx\,dt = 2\int_0^t (\hat{u}_i(\cdot,\tau), m_i)\,d\tau,$$

so the error of each mode depends only on the coefficients $\hat{\hat{u}}_{ii}(\tau) := (\hat{u}_i(\tau), m_i)$ up to time t. By explicitly solving the convection equation,

$$\hat{\hat{u}}_{ii}(t) = \frac{X}{(i-1)\pi b}\sin\left(\frac{(i-1)\pi tb}{X}\right)$$

and hence

$$\begin{aligned}R[u^{(n)}] = \sum_{i=n+1}^{N}\|\hat{u}_i\|_{L_2}^2 &= \sum_{i=n+1}^{N} 2\int_0^t \hat{\hat{u}}_{ii}^{(N)}\,d\tau \\ &= 2\sum_{i=n+1}^{N}\left(\frac{X}{(i-1)\pi b}\right)^2\left(1-\cos\frac{(i-1)\pi tc}{X}\right) \\ &\sim \mathcal{O}\left(\frac{1}{n}-\frac{1}{N}\right) \overset{N\to\infty}{\longrightarrow} \mathcal{O}\left(\frac{1}{n}\right).\end{aligned}$$

An approximately $\mathcal{O}\left(n^{-1}\right)$ decay for the pure diffusion case is seen in Figure 3—this is the decay rate predicted by the parabolic estimate analysis in section 3.3. If $\epsilon > 0$, the decay rate seems to be a hybrid between the convection and the diffusion parts—for small n, the $\mathcal{O}\left(n^{-1}\right)$ decay from the convection part dominates, while for large n the decay shows better agreement with the $\mathcal{O}\left(n^{-3}\right)$ decay from the diffusion part. Evidently, the analysis in section 3.3 is unable to capture the intermediate and asymptotic behaviors of the error decay for the convection-diffusion equation.

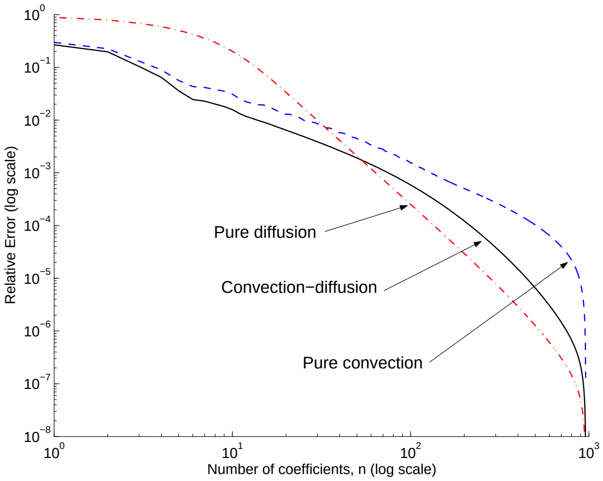

FIGURE 3. Log scale plots of the relative errors incurred by the truncation of the convection-diffusion system, as well as the pure diffusion and the pure convection systems. The convection and diffusion coefficients are $b = 6b_0$ and $\epsilon = 0.1$, respectively.

Bibliography

[1] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3):1005–1034 (electronic), 2007.

[2] Ivo Babuška, Raúl Tempone, and Georgios E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.

[3] Ivo Babuška, Raúl Tempone, and Georgios E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1251–1294, 2005.

[4] Majid Badieirostami, Ali Adibi, Hao-Min Zhou, and Shui-Nee Chow. Efficient modeling of spatially incoherence sources based on wiener chaos expansion method for the analysis of photonic crystal spectrometers. *Proc. SPIE Int. Soc. Op. Eng.*, pages 648018–1–8, 2007.

[5] Majid Badieirostami, Ali Adibi, Hao-Min Zhou, and Shui-Nee Chow. Wiener chaos expansion and simulation of electromagnetic wave propagation excited by a spatially incoherent source. *Multiscale Model. Simul.*, 8(2):591–604, 2009/10.

[6] A. Bensoussan and R. Temam. Équations stochastiques du type Navier-Stokes. *J. Functional Analysis*, 13:195–222, 1973.

[7] Fred Espen Benth and Jon Gjerde. Convergence rates for finite element approximations of stochastic partial differential equations. *Stochastics Stochastics Rep.*, 63(3-4):313–326, 1998.

[8] Fred Espen Benth and Thomas Gorm Theting. Some regularity results for the stochastic pressure equation of Wick-type. *Stochastic Anal. Appl.*, 20(6):1191–1223, 2002.

[9] R. H. Cameron and W. T. Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann. of Math. (2)*, 48:385–392, 1947.

[10] Yanzhao Cao. On convergence rate of Wiener-Ito expansion for generalized random variables. *Stochastics*, 78(3):179–187, 2006.

[11] Yanzhao Cao, Hongtao Yang, and Li Yin. Finite element methods for semilinear elliptic stochastic partial differential equations. *Numer. Math.*, 106(2):181–198, 2007.

[12] Bernardo Cockburn, George E. Karniadakis, and Chi-Wang Shu. The development of discontinuous Galerkin methods. In *Discontinuous Galerkin methods (Newport, RI, 1999)*, volume 11 of *Lect. Notes Comput. Sci. Eng.*, pages 3–50. Springer, Berlin, 2000.

[13] Arnaud Debussche and Jacques Printems. Weak order for the discretization of the stochastic heat equation. *Math. Comp.*, 78(266):845–863, 2009.

[14] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[15] Franco Flandoli. Dissipativity and invariant measures for stochastic Navier-Stokes equations. *NoDEA Nonlinear Differential Equations Appl.*, 1(4):403–423, 1994.

[16] Franco Flandoli and Dariusz Gatarek. Martingale and stationary solutions for stochastic Navier-Stokes equations. *Probab. Theory Related Fields*, 102(3):367–391, 1995.

[17] C. Foiaş. Statistical study of Navier-Stokes equations. I, II. *Rend. Sem. Mat. Univ. Padova*, 48:219–348 (1973); ibid. 49 (1973), 9–123, 1972.

[18] C. Foiaş and R. Temam. Homogeneous statistical solutions of Navier-Stokes equations. *Indiana Univ. Math. J.*, 29(6):913–957, 1980.

[19] Ciprian Foias, Ricardo M. S. Rosa, and Roger Temam. A note on statistical solutions of the three-dimensional Navier-Stokes equations: the stationary case. *C. R. Math. Acad. Sci. Paris*, 348(5-6):347–353, 2010.

[20] 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.

[21] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity. *SIAM J. Numer. Anal.*, 47(5):3624–3651, 2009.

[22] R. G. Ghanem and P. D. Spanos. Polynomial chaos in stochastic finite elements. *J. Appl. Mech.*, 57:197–202, 1990.

[23] Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[24] Colette Guillopé. Comportement à l'infini des solutions des équations de Navier-Stokes et propriété des ensembles fonctionnels invariants (ou attracteurs). *Ann. Inst. Fourier (Grenoble)*, 32(3):ix, 1–37, 1982.

[25] István Gyöngy and David Nualart. Implicit scheme for quasi-linear parabolic partial differential equations perturbed by space-time white noise. *Stochastic Process. Appl.*, 58(1):57–72, 1995.

[26] István Gyöngy and David Nualart. Implicit scheme for stochastic parabolic partial differential equations driven by space-time white noise. *Potential Anal.*, 7(4):725–757, 1997.

[27] Takeyuki Hida, Hui-Hsiung Kuo, Jürgen Potthoff, and Ludwig Streit. *White noise*, volume 253 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993. An infinite-dimensional calculus.

[28] Takeyuki Hida, Hui-Hsiung Kuo, Jürgen Potthoff, and Ludwig Streit. *White noise*, volume 253 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993. An infinite-dimensional calculus.

[29] H. Holden, T. Lindstrøm, B. Øksendal, J. Ubøe, and T.-S. Zhang. The pressure equation for fluid flow in a stochastic medium. *Potential Anal.*, 4(6):655–674, 1995.

[30] Helge Holden, Tom Lindstrøm, Bernt Øksendal, Jan Ubøe, and Tu Sheng Zhang. Stochastic boundary value problems: a white noise functional approach. *Probab. Theory Related Fields*, 95(3):391–419, 1993.

[31] Helge Holden, Bernt Øksendal, Jan Ubøe, and Tusheng Zhang. *Stochastic partial differential equations*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 1996. A modeling, white noise functional approach.

[32] Arnulf Jentzen and Peter Kloeden. Taylor expansions of solutions of stochastic partial differential equations with additive noise. *Ann. Probab.*, 38(2):532–569, 2010.

[33] Arnulf Jentzen and Peter E. Kloeden. Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 465(2102):649–667, 2009.

[34] S. Kaligotla and S. V. Lototsky. Wick product in stochastic burgers equation: a curse or a cure? *Preprint*, 2010.

[35] Gopinath Kallianpur and Jie Xiong. *Stochastic differential equations in infinite-dimensional spaces*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 26. Institute of Mathematical Statistics, Hayward, CA, 1995. Expanded version of the lectures delivered as part of the 1993 Barrett Lectures at the University of Tennessee, Knoxville, TN, March 25–27, 1993, With a foreword by Balram S. Rajput and Jan Rosinski.

[36] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[37] Yuri G. Kondratiev, Peter Leukert, and Ludwig Streit. Wick calculus in Gaussian analysis. *Acta Appl. Math.*, 44(3):269–294, 1996.

[38] Mihály Kovács, Stig Larsson, and Fredrik Lindgren. Strong convergence of the finite element method with truncated noise for semilinear parabolic stochastic equations with additive noise. *Numer. Algorithms*, 53(2-3):309–320, 2010.

[39] Mihály Kovács, Stig Larsson, and Fardin Saedpanah. Finite element approximation of the linear stochastic wave equation with additive noise. *SIAM J. Numer. Anal.*, 48(2):408–427, 2010.

[40] John D. Kraus and Daniel Fleisch. *Electromagnetics*. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill, New York, 1991.

[41] S. G. Kreĭn, Yu. Ī. Petunīn, and E. M. Semënov. *Interpolation of linear operators*, volume 54 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, R.I., 1982. Translated from the Russian by J. Szűcs.

[42] Hui-Hsiung Kuo. *White noise distribution theory*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996.

[43] C. Y. Lee, B. L. Rozovskii, and H. M. Zhou. Randomization of forcing in large systems of PDEs for improvement of energy estimates. *Multiscale Model. Simul.*, 8(4):1419–1438, 2010.

[44] Chia Ying Lee and Boris Rozovskii. A stochastic finite element method for stochastic parabolic equations driven by purely spatial noise. *Commun. Stoch. Anal.*, 4(2):271–297, 2010.

[45] S. V. Lototsky and B. L. Rozovskii. A unified approach to stochastic evolution equations using the Skorokhod integral. *Teor. Veroyatn. Primen.*, 54(2):288–303, 2009.

[46] S. V. Lototsky, B. L. Rozovskii, and D. Seleši. A note on generalized malliavin calculus. *Preprint*, 2010.

[47] Sergey Lototsky and Boris Rozovskii. Stochastic differential equations: a Wiener chaos approach. In *From stochastic calculus to mathematical finance*, pages 433–506. Springer, Berlin, 2006.

[48] Sergey V. Lototsky and Boris L. Rozovskii. Stochastic parabolic equations of full second order. In *Topics in stochastic analysis and nonparametric estimation*, volume 145 of *IMA Vol. Math. Appl.*, pages 199–210. Springer, New York, 2008.

[49] Sergey V. Lototsky and Boris L. Rozovskii. Stochastic partial differential equations driven by purely spatial noise. *SIAM J. Math. Anal.*, 41(4):1295–1322, 2009.

[50] Paul Malliavin. *Stochastic analysis*, volume 313 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997.

[51] Leonard Mandel and Emil. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge, UK, 1995.

[52] H. Manouzi. A finite element approximation of linear stochastic PDEs driven by multiplicative white noise. *Int. J. Comput. Math.*, 85(3-4):527–546, 2008.

[53] R. Mikulevicius and B. L. Rozovskii. Stochastic Navier-Stokes equations for turbulent flows. *SIAM J. Math. Anal.*, 35(5):1250–1310, 2004.

[54] R. Mikulevicius and B. L. Rozovskii. Global L_2-solutions of stochastic Navier-Stokes equations. *Ann. Probab.*, 33(1):137–176, 2005.

[55] R. Mikulevicius and B. L. Rozovskii. On quantized stochastic navier-stokes equations. *Preprint*, 2010.

[56] David Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[57] B. L. Rozovskiĭ. *Stochastic evolution systems*, volume 35 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990. Linear theory and applications to nonlinear filtering, Translated from the Russian by A. Yarkho.

[58] M. A. Shubin. *Pseudodifferential operators and spectral theory*. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1987. Translated from the Russian by Stig I. Andersson.

[59] Roger Temam. *Navier-Stokes equations and nonlinear functional analysis*, volume 66 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1995.

[60] Roger Temam. *Navier-Stokes equations*. AMS Chelsea Publishing, Providence, RI, 2001. Theory and numerical analysis, Reprint of the 1984 edition.

[61] Thomas Gorm Theting. Solving parabolic Wick-stochastic boundary value problems using a finite element method. *Stoch. Stoch. Rep.*, 75(1-2):49–77, 2003.

[62] Vidar Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[63] Gjermund Våge. Variational methods for PDEs applied to stochastic partial differential equations. *Math. Scand.*, 82(1):113–137, 1998.

[64] John B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.

[65] Xiaoliang Wan, Boris Rozovskii, and George Em Karniadakis. A stochastic modeling methodology based on weighted Wiener chaos and Malliavin calculus. *Proc. Natl. Acad. Sci. USA*, 106(34):14189–14194, 2009.

[66] R. Weissleder. Molecular imaging in cancer. *Science*, 312:1168–1171, 2006.

[67] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Comput. Methods Appl. Mech. Engrg.*, 191(43):4927–4948, 2002.

[68] Dongbin Xiu and George Em Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644 (electronic), 2002.

[69] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187(1):137–167, 2003.

[70] Yubin Yan. Galerkin finite element methods for stochastic parabolic partial differential equations. *SIAM J. Numer. Anal.*, 43(4):1363–1384 (electronic), 2005.

[71] A. Yodh and B. Chance. Spectropscopy and imaging with diffusing light. *Phys. Today*, 48:34–40, 1995.