Eidgenössische Technische Hochschule Zürich

Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Adaptive Galerkin approximation algorithms for partial differential equations in infinite dimensions*

Ch. Schwab† and E. Süli‡

Research Report No. 2011-69
November 2011

Revised: July 2012

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

*The research reported in this paper was carried out at the Hausdorff Institute for Mathematics (HIM), Bonn, during the HIM-Trimester on "High Dimensional Approximation", May–August 2011.

†The author was partially supported by the European Research Council (ERC) under the FP7 programme ERC AdG 247277.

‡Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK. The author was partially supported by the EPSRC Science and Innovation award to the Oxford Centre for Nonlinear PDE (EP/E035027/1).

ADAPTIVE GALERKIN APPROXIMATION ALGORITHMS FOR PARTIAL DIFFERENTIAL EQUATIONS IN INFINITE DIMENSIONS

CHRISTOPH SCHWAB AND ENDRE SÜLI

Abstract. Space-time variational formulations of infinite-dimensional Fokker–Planck (FP) and Ornstein–Uhlenbeck (OU) equations for functions on a separable Hilbert space H are developed. The well-posedness of these equations in the Hilbert space $\mathrm{L}^2(H,\mu)$ of functions on H, which are square-integrable with respect to a Gaussian measure μ on H, is proved. Specifically, for the infinite-dimensional FP equation, adaptive space-time Galerkin discretizations, based on a tensorized Riesz basis, built from biorthogonal piecewise polynomial wavelet bases in time and the Wiener–Hermite polynomial chaos in the Wiener–Itô decomposition of $\mathrm{L}^2(H,\mu)$, are introduced. The resulting space-time adaptive Wiener-Hermite polynomial Galerkin discretizations of the infinite-dimensional PDE are proved to converge quasioptimally in the sense that they produce sequences of finite-dimensional approximations which attain the best possible *algebraic* rates afforded by the tensor-products of multiresolution (wavelet) time-discretizations and of systems of tensorized Wiener–Hermite polynomial chaos expansions in $\mathrm{L}^2(H,\mu)$. Here, quasioptimality is understood with respect to the nonlinear, best N-term approximation benchmark of the solution. As a consequence, the proposed adaptive Galerkin solution algorithms exhibit dimension-independent performance, which is optimal with respect to the algebraic best N-term rate afforded by the solution and the regularity of the multiresolution (wavelet) time-discretizations in the finite-dimensional case, in particular. All constants in our error and complexity bounds are shown to be independent of the number of "active" coordinates identified by the proposed adaptive Galerkin approximation algorithms. The computational work and memory required by the proposed algorithms scale linearly with the support size of the coefficient vectors in the approximations, with dimension-independent constants.

1. Introduction

Partial differential equations in infinite dimensions arise in a number of relevant applications; most notably as forward and backward Kolmogorov equations for stochastic Partial Differential Equations (sPDEs for short), we refer to, e.g., [4, 3] and the references therein. Citing [9, p. X], "parabolic equations on Hilbert spaces appear in mathematical physics to model systems with infinitely many degrees of freedom. Typical examples are provided by spin configurations in statistical mechanics and by crystals in solid state theory. Infinite-dimensional parabolic equations provide an analytic description of infinite-dimensional diffusion processes in such branches of applied mathematics as population biology, fluid dynamics, and mathematical finance." In spite of their prominence in a range of relevant applications, the numerical solution of PDEs in infinite dimensions appears to have received only scant attention in the mathematical literature. Numerical approximations to such equations are mostly attempted by path simulations in the corresponding stochastic partial differential equation. Their path-wise solutions belong to function spaces over finite-dimensional domains and can be, therefore, approximated by standard discretization techniques, combined with Monte Carlo path sampling. In the present paper, we propose and analyze a novel, deterministic, adaptive, spectral-Galerkin approach to the construction of finite-dimensional numerical approximations to the deterministic forward equation in infinite-dimensional spaces,

The research reported in this paper was carried out at the Hausdorff Institute for Mathematics (HIM), Bonn, during the HIM-Trimester on "High Dimensional Approximation", May–August 2011.

C.S. was partially supported by the European Research Council (ERC) under the FP7 programme ERC AdG 247277. E.S. was partially supported by the EPSRC Science and Innovation award to the Oxford Centre for Nonlinear PDE (EP/E035027/1).

which exhibit certain optimality properties. The proposed approach is based on space-time variational formulations of these equations, which are posed in Gel'fand-triples of Sobolev spaces over a separable Hilbert H with respect to a Gaussian measure μ, and on the use of Riesz bases of these spaces, which have been developed for linear and nonlinear parabolic PDEs in finite dimensions in [19, 5, 14]. We use here the general approach developed by Cohen, Dahmen and DeVore in a series of papers (cf. [6, 7] and the references therein) to show the optimality of nonlinear, adaptive Galerkin approximations of elliptic operator equations on bounded domains in $\mathbb{R}^d$. These tools were extended in [19] to parabolic evolution equations, using multiresolution (wavelet) bases, again on bounded domains in $\mathbb{R}^d$. In this paper, we present a class of adaptive Galerkin discretizations of Fokker–Planck equations on an infinite-dimensional, separable Hilbert space H, equipped with a Gaussian measure μ. In contrast with the considerations in [6, 7, 19], we use a *spectral Galerkin method* on H here, based on Hermite polynomials of a sequence of Gaussian random variables on H, and, as in [19], a wavelet type Riesz basis with respect to the time variable. We prove stability and optimality of the adaptive space-time Galerkin discretizations thus obtained. This is achieved by verifying, for a particular class of second-order differential operators on H, numerical sparsity in terms of a Wiener polynomial chaos basis and the Riesz basis property on the separable Hilbert space $L^2(H,\mu)$. Our adaptive Galerkin discretization is shown to generate, in particular, a family of finite-dimensional discretizations of the Fokker–Planck equation with stability and, in a certain sense, optimal finite-dimensional truncations of the infinite-dimensional problem. The use of a spectral basis of Wiener polynomial chaos type obviates "meshing" the "domain" H of the parabolic equation.

The structure of this paper is as follows: in the next section, we first present two space-time variational formulations of the parabolic equation set in a domain of finite dimension d. We then prepare the corresponding formulation of an infinite-dimensional Fokker–Planck equation by recapitulating basic facts about Gaussian measures on separable Hilbert spaces. In Section 4, we prove well-posedness of the second of our two space-time variational formulations in the infinite-dimensional space $\mathrm{L}^2(H,\mu)$, where H denotes a separable Hilbert space and μ is a Gaussian measure on H. We shall establish, in particular, that the solution operator of the variational formulation is an isomorphism between suitable solution and data spaces. We then focus on the formulation and the analysis of space-time-adaptive Galerkin approximations. We outline the general principle in Section 5, where we introduce the idea of conversion of abstract, well-posed operator equations on separable Hilbert spaces to equivalent, bi-infinite matrix-vector problems in the sequence space $\ell^2(\mathbb{N})$. After reviewing some relevant facts concerning N-term approximations in $\ell^2(\mathbb{N})$, we then introduce abstract adaptive Galerkin approximation algorithms that *construct* sequences of N-term approximations, which, while not being best N-term approximations, are optimal in the sense that these approximations converge asymptotically at the rate afforded by the best N-term approximation, provided that certain conditions are met by the operators and the Riesz bases used to discretize them. In Section 8 the abstract concepts are specialized to infinite-dimensional Kolmogorov equations. Notably, a Wiener polynomial chaos type Riesz basis in $\mathrm{L}^2(H,\mu)$ and a wavelet basis in time is used as the basis for the Galerkin discretization. In Section 9 we verify the abstract assumptions for the specific equations of interest, and in Section 10 we consider the more general setting of nonsymmetric equations with drift, leading to our main result concerning optimality of adaptive Galerkin discretizations with dimension-independent bounds, in Section 11.

2. Space-time variational formulation of abstract parabolic problems

2.1. **Abstract parabolic equations.** Given $T>0$ and a Gel'fand triple

$$\mathcal{V} \hookrightarrow \mathcal{H} \cong \mathcal{H}^* \hookrightarrow \mathcal{V}^*, \tag{2.1}$$

for separable Hilbert spaces $\mathcal{H}$ and $\mathcal{V}$ over the field $\mathbb{R}$ of real numbers, with the continuous and dense embeddings signified by the symbol $\hookrightarrow$, we consider the abstract parabolic differential equation

$$\partial_t u + Au = f \ \text{ in } \mathcal{V}^* \text{ for } t \in (0,T), \tag{2.2}$$

where $A \in \mathcal{L}(\mathcal{V}, \mathcal{V}^*)$, with $\|A\|_{\mathcal{L}(\mathcal{V},\mathcal{V}^*)} = M_a > 0$, satisfies $A = A^*$, and we suppose that the following Garding inequality holds:

$$\exists m_a > 0 \quad \exists \kappa \geq 0 \quad \forall v \in \mathcal{V}: \qquad \mathfrak{a}(v,v) := {}_{\mathcal{V}^*}\langle Av, v\rangle_{\mathcal{V}} \geq m_a \|v\|_{\mathcal{V}}^2 - \kappa \|v\|_{\mathcal{H}}^2. \tag{2.3}$$

Here ${}_{\mathcal{V}^*}\langle\cdot,\cdot\rangle_{\mathcal{V}}$ denotes the duality pairing between $\mathcal{V}^*$, the dual space of $\mathcal{V}$, and the space $\mathcal{V}$. As our aim is to develop the numerical analysis of space-time adaptive Galerkin discretizations of infinite-dimensional parabolic problems, we begin, following [19], by presenting weak formulations that are amenable to adaptive Galerkin discretizations. To this end, we consider the Bochner spaces $\mathrm{L}^2(0,T;\mathcal{V})$, $\mathrm{L}^2(0,T;\mathcal{H})$ and $\mathrm{L}^2(0,T;\mathcal{V}^*)$ and we define

$$\mathrm{H}^1(0,T;\mathcal{V}) := \{u \in \mathrm{L}^2(0,T;\mathcal{V}) : u' \in \mathrm{L}^2(0,T;\mathcal{V})\}, \tag{2.4}$$

as well as

$$\mathrm{H}^1_{0,\{0\}}(0,T;\mathcal{V}) := \{u \in \mathrm{H}^1(0,T;\mathcal{V}) : u(0) = 0 \ \text{ in } \mathcal{V}\}, \tag{2.5a}$$

$$\mathrm{H}^1_{0,\{T\}}(0,T;\mathcal{V}) := \{u \in \mathrm{H}^1(0,T;\mathcal{V}) : u(T) = 0 \ \text{ in } \mathcal{V}\}. \tag{2.5b}$$

Here and throughout the rest of the paper u' will signify $\mathrm{d}u/\mathrm{d}t$ or $\partial u/\partial t$, depending on the context.

In the variational formulation of the parabolic problem, an important role is played by the space $\mathcal{X}$ defined by

$$\mathcal{X} := \mathrm{L}^2(0,T;\mathcal{V}) \cap \mathrm{H}^1(0,T;\mathcal{V}^*), \tag{2.6}$$

which we equip with the norm $\|\cdot\|_{\mathcal{X}}$ defined by

$$\|v\|_{\mathcal{X}} := \big(\|v\|^2_{\mathrm{L}^2(0,T;\mathcal{V})} + \|v'\|^2_{\mathrm{L}^2(0,T;\mathcal{V}^*)}\big)^{\frac{1}{2}}.$$

With $\mathcal{V}$, $\mathcal{H}$ and $\mathcal{V}^*$ as in the triple (2.1) the following continuous embedding holds:

$$\mathcal{X} \hookrightarrow \mathrm{C}([0,T];\mathcal{H}) \tag{2.7}$$

(in the sense that any $v \in \mathcal{X}$ is equal almost everywhere to a function that is uniformly continuous as a mapping from the nonempty closed interval $[0,T]$ of the real line into $\mathcal{H}$). Therefore, for $u \in \mathcal{X}$ and $0 \leq t \leq T$, the values $u(t)$ are well-defined in $\mathcal{H}$ and there exists a constant $C = C(T) > 0$ such that

$$\forall u \in \mathcal{X} \quad \forall t \in [0,T]: \quad \|u(t)\|_{\mathcal{H}} \leq C\|u\|_{\mathcal{X}}.$$

In particular, for $u \in \mathcal{X}$ the values $u(0)$ and $u(T)$ are well-defined in $\mathcal{H}$ and

$$\mathcal{X}_{0,\{0\}} := \{u \in \mathcal{X} : u(0) = 0 \text{ in } \mathcal{H}\}, \quad \mathcal{X}_{0,\{T\}} := \{u \in \mathcal{X} : u(T) = 0 \text{ in } \mathcal{H}\}$$

are closed linear subspaces of $\mathcal{X}$. Henceforth we shall write $\mathcal{Y} = \mathrm{L}^2(0,T;\mathcal{V})$ and we denote by $\mathcal{Y}^*$ the dual space of $\mathcal{Y}$, which is isomorphic to $\mathrm{L}^2(0,T;\mathcal{V}^*)$ identifying $\mathrm{L}^2(0,T;\mathcal{H})$ with its own dual.

2.2. **First space-time variational formulation.** We consider (2.2) in the special case when

$$Lu := \partial_t u + Au = f \in \mathcal{Y}^* \text{ in } (0,T), \ \ u(0) = u_0 \in \mathcal{H}, \tag{2.8}$$

in conjunction with a suitable homogenous boundary condition incorporated in the (domain of) definition of the linear operator A. We begin by considering the case when the initial condition is also homogeneous, i.e., $u_0 = 0 \in \mathcal{H}$.

The first space-time variational formulation of the parabolic problem (2.8) is based on the bilinear form

$$(u,v) \in \mathcal{X} \times \mathcal{Y} \mapsto \mathfrak{B}(u,v) := \int_0^T \Big({}_{\mathcal{V}^*}\langle u', v\rangle_{\mathcal{V}} + \mathfrak{a}(u,v)\Big)\,\mathrm{d}t \in \mathbb{R}. \tag{2.9}$$

Then, the space-time weak formulation of the parabolic problem (2.8) with homogeneous initial condition $u_0 = 0$ in $\mathcal{H}$ reads as follows: given $f \in \mathcal{Y}^*$, find $u \in \mathcal{X}_{0,\{0\}}$ such that

$$\mathfrak{B}(u,v) = f(v) \quad \forall v \in \mathcal{Y}. \tag{2.10}$$

2.3. **Second space-time variational formulation.** In (2.10), the initial condition $u(0) = 0$ was incorporated in the definition of the function space $\mathcal{X}_{0,\{0\}}$ in which the solution to the problem was sought. To accommodate a nonhomogeneous initial condition, one may proceed in (at least) two different ways. If

$$u(0) = u_0 \neq 0 \ \text{ in } \mathcal{H}, \tag{2.11}$$

then one can for example first subtract from u a function $u_p \in \mathcal{X}$ such that $u_p|_{t=0} = u_0$; we may, in particular, choose for this purpose

$$u_p = \mathrm{e}^{-t}u_0, \ \text{for } u_0 \in \mathcal{V} \subset \mathcal{H}. \tag{2.12}$$

The disadvantage of this approach is that $u_0 \in \mathcal{V}$ is needed (instead of $u_0 \in \mathcal{H}$), which can be viewed as an unnecessarily restrictive demand on the regularity of the initial datum u_0.

Alternatively, in order to relax the regularity requirement $u_0 \in \mathcal{V}$ to $u_0 \in \mathcal{H}$, one may impose (2.11) weakly, either by enforcing it via a multiplier as in [19] or by a space-time variational formulation, which incorporates it as a *natural boundary condition* as follows: in (2.10) we integrate the time derivative by parts using that

$$\int_0^T {}_{\mathcal{V}^*}\langle u', v\rangle_{\mathcal{V}} \,\mathrm{d}t = -\int_0^T {}_{\mathcal{V}^*}\langle v', u\rangle_{\mathcal{V}} \,\mathrm{d}t + (u, v)|_0^T, \quad u, v \in \mathrm{L}^2(0,T;\mathcal{V}) \cap \mathrm{H}^1(0,T;\mathcal{V}^*). \tag{2.13}$$

If $u(0) = u_0 \neq 0$ in $\mathcal{H}$, we require that

$$v(T) = 0. \tag{2.14}$$

Thus, (2.13) and (2.14) lead to the weak formulation (2.17) below. To state it, we recall the spaces $\mathcal{Y} = \mathrm{L}^2(0,T;\mathcal{V})$ and

$$\mathcal{X} = \mathrm{L}^2(0,T;\mathcal{V}) \cap \mathrm{H}^1(0,T;\mathcal{V}^*) \tag{2.15}$$

and the subspaces (cf. also (2.5))

$$\mathcal{X}_{0,\{0\}} := \{u \in \mathcal{X} : \ u(0) = 0 \ \text{ in } \mathcal{H}\}, \quad \mathcal{X}_{0,\{T\}} := \{u \in \mathcal{X} : \ u(T) = 0 \ \text{ in } \mathcal{H}\}, \tag{2.16}$$

equipped with the norm of $\mathcal{X}$; we shall write $\|\cdot\|_{\mathcal{X}_{0,\{0\}}}$ and $\|\cdot\|_{\mathcal{X}_{0,\{T\}}}$ to indicate that the norm of $\mathcal{X}$ is applied to an element of $\mathcal{X}_{0,\{0\}}$ and $\mathcal{X}_{0,\{T\}}$, respectively. Thanks to the continuous embedding (2.7), $\mathcal{X}_{0,\{0\}}$ and $\mathcal{X}_{0,\{T\}}$ are closed, linear subspaces of $\mathcal{X}$; in particular the expression (2.13) is meaningful for $u, v \in \mathcal{X}$. The variational form of the parabolic problem (2.8) with weak enforcement of the initial condition, which we shall also refer to as the *space-time adjoint weak formulation*, then reads: given $u_0 \in \mathcal{H}$ and $f \in \mathcal{X}^*_{0,\{T\}}$, find

$$u \in \mathcal{Y} : \quad \mathfrak{B}^*(u, v) = \ell^*(v) \quad \forall v \in \mathcal{X}_{0,\{T\}}. \tag{2.17}$$

Here the bilinear form $\mathfrak{B}^*(\cdot,\cdot)$ is given by

$$\mathfrak{B}^*(u, v) := \int_0^T \big(-{}_{\mathcal{V}^*}\langle v', u\rangle_{\mathcal{V}} + \mathfrak{a}(u, v)\big) \,\mathrm{d}t \tag{2.18}$$

and the linear functional ℓ^* is defined by

$$\ell^*(v) := {}_{\mathcal{X}^*}\langle f, v\rangle_{\mathcal{X}} + (u_0, v(0))_{\mathcal{H}}. \tag{2.19}$$

For future reference we collect some simple properties of the functional ℓ^* in a proposition.

Proposition 2.1. *For $f \in \mathcal{X}^*_{0,\{T\}} \simeq \mathrm{L}^2(I;\mathcal{V}^*) + (\mathrm{H}^1_{0,\{T\}})^*(I;\mathcal{V}) \simeq \mathrm{L}^2(I) \otimes \mathcal{V}^* + (\mathrm{H}^1_{0,\{T\}}(I))^* \otimes \mathcal{V}$ and, for any $u_0 \in \mathcal{H}$, the functional ℓ^* in* (2.19) *is linear and continuous on $\mathcal{X}_{0,\{T\}}$, i.e., there exists a constant $C > 0$ such that*

$$\forall v \in \mathcal{X}_{0,\{T\}} : \quad |\ell^*(v)| \leq C\big(\|f\|_{\mathcal{X}^*_{0,\{T\}}} + \|u_0\|_{\mathcal{H}}\big)\, \|v\|_{\mathcal{X}_{0,\{T\}}}. \tag{2.20}$$

With all preparatory considerations in place, we are now ready to discuss the question of well-posedness of the variational formulations (2.10) and (2.17).

2.4. **Well-Posedness.** By proceeding analogously as in the proof of [19, Theorem 5.1], we arrive at the following result.

Theorem 2.2. *Suppose that $\mathcal{V}$ and $\mathcal{H}$ are separable Hilbert spaces over the field $\mathbb{R}$ such that $\mathcal{V} \hookrightarrow \mathcal{H} \simeq \mathcal{H}^* \hookrightarrow \mathcal{V}^*$ with continuous and dense embeddings. Assume further that the bilinear form $\mathfrak{a}(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is continuous, i.e.,*

$$\exists M_a > 0 \quad \forall w, v \in \mathcal{V} : \quad |\mathfrak{a}(w,v)| \leq M_a \|v\|_{\mathcal{V}} \|w\|_{\mathcal{V}}, \tag{2.21}$$

and coercive in the sense that it satisfies a Garding inequality, i.e., there exist $m_a > 0$ and $\kappa \geq 0$ such that

$$\forall v \in \mathcal{V} : \quad \mathfrak{a}(v,v) \geq m_a \|v\|_{\mathcal{V}}^2 - \kappa \|v\|_{\mathcal{H}}^2. \tag{2.22}$$

Then, there exists a positive constant $C = C(m_a, M_a, \kappa, T)$ such that the bilinear form $\mathfrak{B}(\cdot,\cdot)$, as defined in (2.9), satisfies the following inequalities, referred to as the inf-sup condition:

$$\inf_{u \in \mathcal{X}_{0,\{0\}} \setminus \{0\}} \sup_{v \in \mathcal{Y} \setminus \{0\}} \frac{\mathfrak{B}(u,v)}{\|u\|_{\mathcal{X}} \|v\|_{\mathcal{Y}}} \geq C \tag{2.23}$$

and

$$\forall v \in \mathcal{Y} \setminus \{0\} : \sup_{u \in \mathcal{X}_{0,\{0\}}} \mathfrak{B}(u,v) > 0. \tag{2.24}$$

Furthermore, there exists a positive constant $C = C(m_a, M_a, \kappa, T)$, such that $\mathfrak{B}^*(\cdot,\cdot)$ in (2.18) satisfies the following inf-sup condition:

$$\inf_{u \in \mathcal{Y} \setminus \{0\}} \sup_{v \in \mathcal{X}_{0,\{T\}} \setminus \{0\}} \frac{\mathfrak{B}^*(u,v)}{\|u\|_{\mathcal{Y}} \|v\|_{\mathcal{X}_{0,\{T\}}}} \geq C \tag{2.25}$$

and

$$\forall v \in \mathcal{X}_{0,\{T\}} \setminus \{0\} : \quad \sup_{u \in \mathcal{Y}} \mathfrak{B}^*(u,v) > 0. \tag{2.26}$$

Moreover, the bilinear forms $\mathfrak{B}(\cdot,\cdot)$ and $\mathfrak{B}^*(\cdot,\cdot)$ defined in (2.9) *and in* (2.17)*, respectively, induce boundedly invertible, linear operators $B \in \mathcal{L}(\mathcal{X}_{0,\{0\}}, \mathcal{Y}^*)$ and $B^* \in \mathcal{L}(\mathcal{Y}, (\mathcal{X}_{0,\{T\}})^*)$, where the spaces $\mathcal{X}_{0,\{0\}}$, $\mathcal{X}_{0,\{T\}}$ and $\mathcal{Y}$ are defined in* (2.15), (2.16).

The stability conditions satisfied by $\mathfrak{B}^*(\cdot,\cdot)$ imply the following result concerning unique solvability of the corresponding variational formulation (2.17). An analogous result holds for (2.10), thanks to the stability properties of $\mathfrak{B}(\cdot,\cdot)$.

Theorem 2.3. *For every $u_0 \in \mathcal{H}$ and for every $f \in \mathcal{X}_{0,\{T\}}^* = \mathrm{L}^2(I;\mathcal{V}^*) + \mathrm{H}_{0,\{T\}}^{-1}(I;\mathcal{V})$, there exists a unique weak solution $u \in \mathcal{Y} = \mathrm{L}^2(0,T;\mathcal{V})$ of (2.8) in the sense that u satisfies (2.17). Moreover, the operator $L : \mathcal{Y} \to \mathcal{X}_{0,\{T\}}^*$ is an isomorphism.*

The proof of this theorem is analogous to that of [19, Theorem 5.1], and is therefore omitted. We are now ready to consider a concrete realization of this abstract theoretical framework, which concerns parabolic PDEs in infinite dimensions.

3. Gaussian measures

Suppose that H is a separable Hilbert space with norm $\|\cdot\|_H$ and inner product $(\cdot,\cdot)_H$ over the field of real numbers. We denote by $\mathrm{L}(H)$ the space of all bounded linear operators from H into H, equipped with the associated operator norm $\|\cdot\|_{\mathrm{L}(H)}$, and we denote by $\mathrm{L}^+(H)$ the subset of H consisting of all nonnegative, symmetric and bounded, linear operators on H. Finally, $\mathfrak{B}(H)$ will signify the σ-algebra of all Borel subsets of H. We recall that a bounded linear operator $R \in \mathrm{L}(H)$ is said to be *trace-class* if there exist two sequences $(a_k)_{k=1}^\infty$ and $(b_k)_{k=1}^\infty$ in H such that

$$Rh = \sum_{k=1}^\infty (h, a_k)_H b_k, \quad h \in H, \tag{3.1}$$

and

$$\sum_{k=1}^\infty \|a_k\|_H \, \|b_k\|_H < \infty.$$

The set of all elements of $\mathrm{L}(H)$ that are trace-class will be denoted by $\mathrm{L}_1(H)$. We note that if $R \in \mathrm{L}_1(H)$, then R is a compact linear operator on H. The set $\mathrm{L}_1(H)$, endowed with the usual operations of addition and scalar multiplication, is a Banach space with the norm

$$\|R\|_{\mathrm{L}_1(H)} := \inf\left\{\sum_{k=1}^{\infty}\|a_k\|_H\,\|b_k\|_H \,:\, Rh=\sum_{k=1}^{\infty}(h,a_k)_H b_k,\quad h\in H,\quad (a_k)_{k=1}^{\infty},\,(b_k)_{k=1}^{\infty}\subset H\right\}.$$

Assuming that $R \in \mathrm{L}_1(H)$, the *trace* $\mathrm{Tr}\,R$ of R is defined by the formula

$$\mathrm{Tr}\,R=\sum_{k=1}^{\infty}(Re_k,e_k)_H,$$

where at this stage $(e_k)_{k=1}^{\infty}$ is any complete orthonormal basis in H. In particular if $R \in \mathrm{L}_1(H)$ is expressed by (3.1), then

$$\mathrm{Tr}\,R=\sum_{k=1}^{\infty}(a_k,b_k)_H.$$

The definition of *trace* being independent of the choice of the basis, we have that $|\mathrm{Tr}R| \leq \| R\|_{\mathrm{L}_1(H)}$. The next proposition (cf. Pietsch [17]) sharpens these statements further.

Proposition 3.1. *Suppose that R is a compact selfadjoint linear operator on a separable Hilbert space H, and that $(\lambda_k)_{k=1}^{\infty}$ are its eigenvalues (repeated according to their multiplicity).*

Then, $R \in \mathrm{L}_1(H)$ if, and only if, $\sum_{k=1}^{\infty}|\lambda_k|<+\infty$. Furthermore, $\|R\|_{\mathrm{L}_1(H)}=\sum_{k=1}^{\infty}|\lambda_k|$ and $\mathrm{Tr}\,R=\sum_{k=1}^{\infty}\lambda_k$. In particular if $\lambda_k\geq 0$ for all $k\in\mathbb{N}$, then $\mathrm{Tr}\,R=\|R\|_{\mathrm{L}_1(H)}$.

Let $\mathbb{R}^{\mathbb{N}}$ denote linear space of all sequences $x=(x_k)_{k=1}^{\infty}$ of real numbers, equipped with the metric

$$\varrho(x,y):=\sum_{k=1}^{\infty}2^{-k}\frac{|x_k-y_k|}{1+|x_k-y_k|},\quad x,y\in\mathbb{R}^{\mathbb{N}}.$$

Let further $\ell^2(\mathbb{N})$ denote the Hilbert space of all sequences $x=(x_k)_{k=1}^{\infty}\in\mathbb{R}^{\infty}$ such that

$$\|x\|_{\ell^2(\mathbb{N})}:=\left(\sum_{k=1}^{\infty}|x_k|^2\right)^{1/2}<+\infty;$$

the inner product on $\ell^2(\mathbb{N})$ is defined by $(x,y)_{\ell^2(\mathbb{N})}:=\sum_{k=1}^{\infty}x_k y_k$, for $x,y\in\ell^2(\mathbb{N})$.

Let us further consider, for $a\in\mathbb{R}$ and for $\lambda>0$, the Gaussian measure on $\mathbb{R}$ with mean a and standard deviation λ defined by

$$N_{a,\lambda}(\mathrm{d}x):=\frac{1}{\sqrt{2\pi\lambda}}\,\mathrm{e}^{-\frac{(x-a)^2}{2\lambda}}\,\mathrm{d}x.$$

The following result is stated as Theorem 1.2.1 in [9].

Theorem 3.2. *Suppose that $a\in H$ and $Q\in\mathrm{L}_1^+(H)$. Then, there exists a unique probability measure μ on $(H,\mathfrak{B}(H))$ such that*

$$\int_H \mathrm{e}^{\imath(h,x)_H}\,\mu(\mathrm{d}x)=\mathrm{e}^{\imath(a,h)_H}\mathrm{e}^{-\frac{1}{2}(Qh,h)_H},\quad h\in H.$$

Moreover, μ is the restriction to H (identified with the Hilbert space $\ell^2(\mathbb{N})$) of the product measure

$$\bigotimes_{k=1}^{\infty}\mu_k=\bigotimes_{k=1}^{\infty}N_{a_k,\lambda_k},$$

defined on $(\mathbb{R}^{\mathbb{N}},\mathfrak{B}(\mathbb{R}^{\mathbb{N}}))$.

We shall write $\mu=N_{a,Q}$, and we call a the *mean* and the trace-class operator Q the *covariance operator* of μ. The measure μ will be referred to as a *Gaussian measure on H* with mean a and covariance operator Q. If the law of a random variable is a Gaussian measure, then the random variable is said to be *Gaussian*. In particular, Theorem 3.2 implies that a random variable X with

values in H is Gaussian if, and only if, for any $h \in H$ the real-valued random variable $(h, X)_H$ is Gaussian.

For $\mu = N_{a,Q}$, we denote by $\mathrm{L}^2(H,\mu)$ the Hilbert space of all square-integrable (equivalence classes of) functions from H into $\mathbb{R}$ with inner product

$$(u,v)_{\mathrm{L}^2(H,\mu)} = \int_H u(x)\, v(x)\, \mu(\mathrm{d}x), \quad u, v \in \mathrm{L}^2(H,\mu)$$

and norm

$$\|u\|_{\mathrm{L}^2(H,\mu)} = (u,u)^{\frac{1}{2}}_{\mathrm{L}^2(H,\mu)}, \quad u \in \mathrm{L}^2(H,\mu).$$

Analogously, we shall denote by $\mathrm{L}^2(H,\mu;H)$ the Hilbert space of all square-integrable (equivalence classes of) functions from H into H with inner product

$$(u,v)_{\mathrm{L}^2(H,\mu;H)} = \int_H (u(x), v(x))_H\, \mu(\mathrm{d}x), \quad u, v \in \mathrm{L}^2(H,\mu;H)$$

and norm

$$\|u\|_{\mathrm{L}^2(H,\mu;H)} = (u,u)^{\frac{1}{2}}_{\mathrm{L}^2(H,\mu;H)}, \quad u \in \mathrm{L}^2(H,\mu;H).$$

Throughout the rest of the paper, $\mu = N_Q := N_{0,Q}$, where $Q \in \mathrm{L}_1^+(H)$. Moreover, we shall assume that $\mathrm{Ker}(Q) = \{0\}$. We shall also suppose that there exists a complete orthonormal system $(e_k)_{k=1}^\infty$ in H and a sequence $(\lambda_k)_{k=1}^\infty$ of positive real numbers, the eigenvalues of Q, such that $Qe_k = \lambda_k e_k$, $k \in \mathbb{N}$. The eigenvalues λ_k are assumed to be enumerated in decreasing order (and repeated according to their multiplicity), and accumulating only at zero. For $x \in H$, we shall then write $x_k = (x, e_k)_H$, $k \in \mathbb{N}$.

The subspace $Q^{1/2}(H) = \{Q^{1/2}h : h \in H\}$ is called the *reproducing kernel* of the Gaussian measure $N_Q := N_{0,Q}$. If $\mathrm{Ker}(Q) = \{0\}$ as has been assumed, then $Q^{1/2}(H)$ is dense in H. In fact, if $x_0 \in H$ is such that $(Q^{1/2}h, x_0)_H = 0$ for all $h \in H$, then $Q^{1/2}x_0 = 0$, and therefore $Qx_0 = 0$, which implies that $x_0 = 0$.

For $Q \in \mathrm{L}_1^+(H)$ such that $\mathrm{Ker}(Q) = \{0\}$, we introduce the isomorphism W from H into $\mathrm{L}^2(H, N_Q)$ as follows: for $f \in Q^{1/2}(H)$ let $W_f \in \mathrm{L}^2(H; N_Q)$ be defined by

$$W_f(x) = (Q^{-1/2}f, x)_H, \quad x \in H. \tag{3.2}$$

Let $\mu = N_Q$. We define Hermite polynomials in $\mathrm{L}^2(H,\mu)$. Let us consider to this end the set Γ of all mappings $\gamma : n \in \mathbb{N} \to \gamma_n \in \{0\} \cup \mathbb{N}$, such that $|\gamma| := \sum_{k=1}^\infty \gamma_k < +\infty$. Clearly $\gamma \in \Gamma$ if, and only if, $\gamma_n = 0$ for all, except possibly finitely many, $n \in \mathbb{N}$. For any $\gamma \in \Gamma$ we define the *Hermite polynomial*

$$H_\gamma(x) := \prod_{k \geq 1} H_{\gamma_k}(W_{e_k}(x)), \quad x \in H \tag{3.3}$$

where the factors in the product on the right-hand side are defined by

$$H_n(\xi) = \frac{(-1)^n}{\sqrt{n!}}\, \mathrm{e}^{\frac{\xi^2}{2}}\, \frac{\mathrm{d}^n}{\mathrm{d}\xi^n}\left(\mathrm{e}^{-\frac{\xi^2}{2}}\right), \quad \xi \in \mathbb{R}, \quad n \in \{0\} \cup \mathbb{N}; \tag{3.4}$$

Note that $H_0 \equiv 1$, so that for each $\gamma \in \Gamma$, the countable product $H_\gamma(x)$ contains only finitely many nontrivial factors and is, therefore, well-defined. Moreover, with the Gaussian measure $\mu = N_Q$ being a countable product measure, we have that

$$\forall \gamma, \gamma' \in \Gamma : \quad (H_\gamma, H_{\gamma'})_{\mathrm{L}^2(H,\mu)} = \delta_{\gamma,\gamma'}. \tag{3.5}$$

We shall denote by $E(H)$ the linear space spanned by all exponential functions, that is all functions $\varphi : x \in H \mapsto \varphi(x) \in \mathbb{R}$ of the form

$$\varphi(x) = \mathrm{e}^{(h,x)_H}, \quad h \in H.$$

It follows from Proposition 1.2.5 in Da Prato and Zabczyk [9] that $E(H)$ is dense in $\mathrm{L}^2(H,\mu)$. On account of the separability of H, $\mathrm{L}^2(H,\mu)$ is separable.

For any $k \in \mathbb{N}$ we consider the partial derivative in the direction e_k (with e_k as above), defined by

$$\mathrm{D}_k\varphi(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\varphi(x + \varepsilon e_k) - \varphi(x)\right), \quad x \in H,\ \varphi \in E(H).$$

Thus, if $\varphi(x) = \mathrm{e}^{(h,x)_H}$ with $h \in H$, then clearly

$$\mathrm{D}_k\varphi(x) = \mathrm{e}^{(h,x)_H}\, h_k, \quad \text{where } h_k = (h, e_k)_H.$$

Let Λ_0 denote the linear span of $\{H_\gamma \otimes e_k : \gamma \in \Gamma,\, k \in \mathbb{N}\}$, with H_γ and with e_k as above and $\mu = N_Q$, and define the linear operator

$$\mathrm{D} : E(H) \subset \mathrm{L}^2(H,\mu) \to \mathrm{L}^2(H,\mu;H) \quad \text{by} \quad \mathrm{D}\varphi(x) := \sum_{k=1}^{\infty} \mathrm{D}_k\varphi(x)\, e_k, \quad x \in H\,.$$

Thanks to Proposition 9.2.2 in Da Prato and Zabczyk [9], D_k is a closable linear operator for all $k \in \mathbb{N}$. If φ belongs to the domain of the closure of D_k, which we shall still denote by D_k, we shall say that $\mathrm{D}_k\varphi$ belongs to $\mathrm{L}^2(H,\mu)$.

Analogously, by Proposition 9.2.4 in Da Prato and Zabczyk [9], D is a closable linear operator. If φ belongs to the domain of the closure of D, which we shall still denote by D, we shall say that $\mathrm{D}\varphi$ belongs to $\mathrm{L}^2(H,\mu;H)$.

For $\mu = N_Q$, let us denote by $\mathrm{W}^{1,2}(H,\mu)$ the linear space of all functions $\varphi \in \mathrm{L}^2(H,\mu)$ such that $\mathrm{D}\varphi \in \mathrm{L}^2(H,\mu;H)$, equipped with the inner product

$$(\varphi,\psi)_{\mathrm{W}^{1,2}(H,\mu)} := (\varphi,\psi)_{\mathrm{L}^2(H,\mu)} + \int_H (\mathrm{D}\varphi(x), \mathrm{D}\psi(x))_H\, \mu(\mathrm{d}x)$$

and norm $\|\varphi\|_{\mathrm{W}^{1,2}(H,\mu)} = (\varphi,\varphi)^{\frac{1}{2}}_{\mathrm{W}^{1,2}(H,\mu)}$. Then, the *Sobolev space* $\mathrm{W}^{1,2}(H,\mu)$ is complete, and is therefore a separable Hilbert space.

For any $\varphi \in \mathrm{L}^2(H,\mu)$, we have that

$$\varphi = \sum_{\gamma\in\Gamma} \varphi_\gamma H_\gamma, \quad \text{where } \varphi_\gamma := (\varphi, H_\gamma)_{\mathrm{L}^2(H,\mu)}.$$

The next theorem, proved independently by Da Prato, Malliavin and Nualart [8] and Peszat [16] provides an analogous characterization of functions belonging to $\mathrm{W}^{1,2}(H,\mu)$ in terms of the complete orthonormal basis $(H_\gamma)_{\gamma\in\Gamma}$.

Theorem 3.3. *A function $\varphi \in \mathrm{L}^2(H,\mu)$ belongs to $\mathrm{W}^{1,2}(H,\mu)$ if, and only if,*

$$\sum_{\gamma\in\Gamma\setminus\{0\}} \langle\gamma,\lambda^{-1}\rangle |\varphi_\gamma|^2 < +\infty, \tag{3.6}$$

where $\varphi_\gamma := (\varphi, H_\gamma)_{\mathrm{L}^2(H,\mu)}$, $\langle\gamma,\lambda^{-1}\rangle := \sum_{k=1}^{\infty} \gamma_k\lambda_k^{-1}$, and $(\lambda_k)_{k=1}^{\infty}$ is the sequence of (positive) eigenvalues (repeated according to their multiplicity) of the covariance operator $Q \in \mathrm{L}_1^+(H)$, $\mathrm{Ker}(Q) = \{0\}$. Moreover, if (3.6) holds, then

$$\|\varphi\|^2_{\mathrm{W}^{1,2}(H,\mu)} = \|\varphi\|^2_{\mathrm{L}^2(H,\mu)} + \sum_{\gamma\in\Gamma\setminus\{0\}} \langle\gamma,\lambda^{-1}\rangle |\varphi_\gamma|^2.$$

Identifying $\mathrm{L}^2(H,\mu)$ with its own dual $\mathrm{L}^2(H,\mu)^*$, we obtain

$$\varphi \in (\mathrm{W}^{1,2}(H,\mu))^* \iff \sum_{\gamma\in\Gamma\setminus\{0\}} (\langle\gamma,\lambda^{-1}\rangle)^{-1} |\varphi_\gamma|^2 < +\infty. \tag{3.7}$$

Finally, the embedding of $\mathrm{W}^{1,2}(H,\mu)$ into $\mathrm{L}^2(H,\mu)$ is compact.

For a proof of these statements we refer, for example, to [9, Theorem 9.2.12]. As $E(H)$ is dense in both $\mathrm{L}^2(H,\mu)$ and $\mathrm{W}^{1,2}(H,\mu)$, the embedding of $\mathrm{W}^{1,2}(H,\mu)$ into $\mathrm{L}^2(H,\mu)$ is also dense.

4. Fokker–Planck equation in countably many dimensions

For $k \in \mathbb{N}$, let us denote by D_k the set $\mathbb{R}^d$ equipped with the Gaussian measure

$$\mu_k(\mathrm{d}q_k) = N_{a_k,\Sigma_k}(\mathrm{d}q_k) = \frac{1}{(2\pi)^{d/2}[\det(\Sigma_k)]^{1/2}} \exp\left(-\tfrac{1}{2}(q_k - a_k)^\top \Sigma_k^{-1}(q_k - a_k)\right) \mathrm{d}q_k$$

with *mean* $a_k \in \mathbb{R}^d$ and positive definite *covariance matrix* $\Sigma_k \in \mathbb{R}^{d\times d}$. We shall assume henceforth that $a_k = 0$ for all $k \in \mathbb{N}$ and that the covariance operator Q, represented by the infinite block-diagonal matrix $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2, \dots)$, with $d \times d$ diagonal blocks Σ_k, $k = 1, 2, \dots$, is trace-class. We define

$$D := \bigtimes_{k=1}^{\infty} D_k$$

so that

$$q := (q_1^\top, q_2^\top, \dots)^\top \in D, \quad q_k \in D_k, \quad k = 1, 2, \dots.$$

We equip the domain D with the product measure

$$\mu := \bigotimes_{k=1}^{\infty} \mu_k = \bigotimes_{k=1}^{\infty} N_{0,\Sigma_k}.$$

Let $\mathbf{A} = (A_{ij})_{i,j=1}^{\infty} \in \mathbb{R}^{\infty\times\infty}$ be a symmetric bi-infinite matrix, i.e., $A_{ij} = A_{ji}$ for all $i, j \in \mathbb{N}$. Suppose further that there exists a real number $\gamma_0 > 0$ such that

$$\sum_{i,j=1}^{\infty} A_{ij}\,\xi_i\xi_j \geq \gamma_0 \|\xi\|_{\ell^2}^2 \quad \text{for all} \quad \xi = (\xi_i)_{i=1}^{\infty} \in \ell^2(\mathbb{N}) \tag{4.1}$$

and a real number $\gamma_1 > 0$ such that

$$\left| \sum_{i,j=1}^{\infty} A_{ij}\,\xi_i\eta_j \right| \leq \gamma_1 \|\xi\|_{\ell^2}\,\|\eta\|_{\ell^2} \quad \text{for all} \quad \xi = (\xi_i)_{i=1}^{\infty} \in \ell^2(\mathbb{N}) \text{ and } \eta = (\eta_i)_{i=1}^{\infty} \in \ell^2(\mathbb{N}). \tag{4.2}$$

Example 4.1. *As an example of an bi-infinite matrix* $\mathbf{A}$ *that satisfies* (4.1) *and* (4.2)*, we mention*

$$\mathbf{A}[\epsilon] = \mathrm{tridiag}\{(\epsilon_j, 1, \epsilon_j) : j = 1, 2, \dots\}, \tag{4.3}$$

where the sequence $\epsilon = \{\epsilon_j\}_{j\geq 1}$ *is assumed to satisfy* $\|\epsilon\|_\infty < 1/2$*. Then, the matrix* $\mathbf{A}[\epsilon]$ *in* (4.3) *satisfies* (4.1) *with* $\gamma_0 = 1 - 2\|\epsilon\|_\infty$ *and* (4.2) *with* $\gamma_1 = 1 + 2\|\epsilon\|_\infty$*.*

Example 4.2. *An important application of our methods are Fokker–Planck equations with state spaces of high, but finite dimension. Such equations arise, among others, in connection with meso-scale descriptions of polymeric fluids. We refer to* [1, 2] *and the references therein. Here, our results apply if we consider* $\mathbf{A}$ *of block-diagonal from* $\mathbf{A} = \mathrm{diag}\{\mathbf{A}_{11}, \mathbf{A}_{22}\}$*, where* $\mathbf{A}_{11} \in \mathbb{R}^{d\times d}$ *is symmetric, positive definite, and* $\mathbf{A}_{22}$ *is as in Example 4.1. Then,* (4.1) *and* (4.2) *hold with constants* $0 < \gamma_0 \leq \gamma_1 < \infty$ *depending on the spectrum of* $\mathbf{A}_{11}$ *and on* d*, however.*

We now define $\mathcal{H} := \mathrm{L}^2(D;\mu)$ and $\mathcal{V} := \mathrm{W}^{1,2}(D,\mu)$. Thanks to Theorem 3.3 and the subsequent discussion, $\mathcal{V} \hookrightarrow \mathcal{H}$, with continuous, dense and compact embedding. In order to state the space-time variational formulation of the infinite-dimensional Fokker–Planck equation under consideration here, we again write $\mathcal{Y} = \mathrm{L}^2(0,T;\mathcal{V})$ and, guided by the abstract framework in Sections 2.2, 2.3, we consider the function space

$$\mathcal{X} = \mathrm{L}^2(0,T;\mathcal{V}) \cap \mathrm{H}^1(0,T;\mathcal{V}^*) \tag{4.4}$$

and its subspaces $\mathcal{X}_{0,\{0\}}$, $\mathcal{X}_{0,\{T\}}$ as in (2.16), equipped with the norm of $\mathcal{X}$, which are closed due to (2.7).

With these spaces, the *space-time adjoint weak formulation of the infinite-dimensional Fokker–Planck equation* reads as follows: given $\widehat{\psi}_0 \in \mathcal{H}$ and $g \in \mathcal{X}_{0,\{T\}}^*$, find

$$\widehat{\psi} \in \mathcal{Y}: \quad \mathfrak{B}^*(\widehat{\psi}, \widehat{\varphi}) = \ell^*(\widehat{\varphi}) \quad \forall \widehat{\varphi} \in \mathcal{X}_{0,\{T\}}, \tag{4.5}$$

where the bilinear form $\mathfrak{B}^*(\cdot,\cdot) : \mathcal{Y} \times \mathcal{X}_{0,\{T\}} \to \mathbb{R}$ is defined by

$$\mathfrak{B}^*(\widehat{\psi}, \widehat{\varphi}) := \int_0^T \big(-{}_{\mathcal{V}^*}\langle \widehat{\varphi}', \widehat{\psi}\rangle_{\mathcal{V}} + \mathfrak{a}(\widehat{\psi}, \widehat{\varphi}) \big)\,\mathrm{d}t, \tag{4.6}$$

with

$$\mathfrak{a}(\widehat{\psi}, \widehat{\varphi}) = \sum_{i,j=1}^{\infty} A_{ij} (\nabla_{q_i}\widehat{\psi}, \nabla_{q_j}\widehat{\varphi})_{[\mathrm{L}^2(D,\mu)]^d}, \tag{4.7}$$

and the linear functional ℓ^* is defined by

$$\ell^*(\widehat{\varphi}) := (\widehat{\psi}_0, \widehat{\varphi}(0))_{\mathrm{L}^2(D,\mu)} + \int_0^T {}_{\mathcal{V}^*}\langle g(t), \hat{\varphi}\rangle_{\mathcal{V}}\, \mathrm{d}t. \tag{4.8}$$

The well-posedness of the infinite-dimensional Fokker–Planck equation is now an immediate consequence of Theorem 2.25 and Theorem 2.3.

Theorem 4.3. *For the bilinear form $\mathfrak{B}^*(\cdot,\cdot)$ in (4.6) there exists a positive constant C such that*

$$\inf_{\widehat{\psi}\in\mathcal{Y}\setminus\{0\}} \sup_{\widehat{\varphi}\in\mathcal{X}_{0,\{T\}}\setminus\{0\}} \frac{\mathfrak{B}^*(\widehat{\psi},\widehat{\varphi})}{\|\widehat{\psi}\|_{\mathcal{Y}}\|\widehat{\varphi}\|_{\mathcal{X}_{0,\{T\}}}} \geq C \tag{4.9}$$

and

$$\forall \widehat{\varphi} \in \mathcal{X}_{0,\{T\}} \setminus \{0\} : \qquad \sup_{\widehat{\psi}\in\mathcal{Y}} \mathfrak{B}^*(\widehat{\psi},\widehat{\varphi}) > 0. \tag{4.10}$$

Furthermore, for each $\widehat{\psi}_0 \in \mathcal{H}$ and each $g \in \mathcal{X}^*_{0,\{T\}}$, there exists a unique $\widehat{\psi} \in \mathcal{Y} = \mathrm{L}^2(0,T;\mathcal{V})$ that satisfies (4.5); and the linear operator $B^* \in \mathcal{L}(\mathcal{Y}, \mathcal{X}^*_{0,\{T\}})$ induced by $\mathfrak{B}^*(\cdot,\cdot)$ is an isomorphism.

Following [19], we shall next develop a class of adaptive Galerkin discretization algorithms for (4.5), along the lines of adaptive wavelet discretizations of boundedly invertible operator equations considered in [6, 7]. These algorithms exhibit, in particular, *"stability by adaptivity"*, i.e., their stability follows directly from the stability (4.9), (4.10) of the continuous, infinite-dimensional problem through suitable *Riesz bases* of the spaces $\mathcal{Y}$ and $\mathcal{X}_{0,\{T\}}$, which we shall construct. Notably, in the present context, these algorithms are *dimensionally robust by design*: as we shall prove, they deliver a sequence of approximate solutions with finitely supported coefficient vectors, i.e., with only finitely many variables q_k being 'active' in each iteration. We will establish certain optimality properties for these finitely supported Galerkin solutions, with all constants in the error and complexity bounds being absolute, i.e., independent of the number of active variables. We first develop the necessary concepts in an abstract setting, before applying them to the Fokker–Planck equation (4.5) in space-time variational form.

5. Well-posed operator equations as bi-infinite matrix problems

Let us denote for a moment by $\mathcal{X}, \mathcal{Y}$ generic separable Hilbert spaces over $\mathbb{R}$, and let us assume that we have available a Riesz basis $\Psi^{\mathcal{X}} = \{\psi^{\mathcal{X}}_\lambda : \lambda \in \nabla_{\mathcal{X}}\}$ for $\mathcal{X}$, meaning that the *synthesis operator*

$$s_{\Psi^{\mathcal{X}}} : \ell^2(\nabla_{\mathcal{X}}) \to \mathcal{X} : \mathbf{c} \mapsto \mathbf{c}^\top \Psi^{\mathcal{X}} := \sum_{\lambda\in\nabla_{\mathcal{X}}} c_\lambda \psi^{\mathcal{X}}_\lambda$$

is boundedly invertible. By identifying $\ell^2(\nabla_{\mathcal{X}})$ with its dual, the adjoint of $s_{\Psi^{\mathcal{X}}}$, known as the *analysis operator*, is

$$s^*_{\Psi^{\mathcal{X}}} : \mathcal{X}^* \to \ell^2(\nabla_{\mathcal{X}}) : g \mapsto [g(\psi^{\mathcal{X}}_\lambda)]_{\lambda\in\nabla_{\mathcal{X}}}.$$

Similarly, let $\Psi^{\mathcal{Y}} = \{\psi^{\mathcal{Y}}_\lambda : \lambda \in \nabla_{\mathcal{Y}}\}$ be a Riesz basis for $\mathcal{Y}$, with synthesis operator $s_{\Psi^{\mathcal{Y}}}$ and its adjoint $s^*_{\Psi^{\mathcal{Y}}}$.

Now let $B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}^*)$ be boundedly invertible. Then also its adjoint $B^* \in \mathcal{L}(\mathcal{Y}, \mathcal{X}^*)$ is boundedly invertible and, for every $f \in \mathcal{Y}^*$, $f^* \in \mathcal{X}^*$ the *operator equations*

$$Bu = f, \qquad B^*u^* = f^*$$

admit unique solutions $u \in \mathcal{X}$, $u^* \in \mathcal{Y}$. Writing $u = s_{\Psi^{\mathcal{X}}}\mathbf{u}$ and $u^* = s_{\Psi^{\mathcal{Y}}}\mathbf{u}^*$, these operator equations are *equivalent* to *bi-infinite matrix-vector problems*

$$\mathbf{B}\mathbf{u} = \mathbf{f}, \quad \mathbf{B}^*\mathbf{u}^* = \mathbf{f}^*, \tag{5.1}$$

where the "load vectors" $\mathbf{f} = s^*_{\Psi^{\mathcal{Y}}} f = [f(\psi^{\mathcal{Y}}_\lambda)]_{\lambda\in\nabla_{\mathcal{Y}}} \in \ell^2(\nabla_{\mathcal{Y}})$, $\mathbf{f}^* = s^*_{\Psi^{\mathcal{X}}} f = [f(\psi^{\mathcal{X}}_\lambda)]_{\lambda\in\nabla_{\mathcal{X}}} \in \ell^2(\nabla_{\mathcal{X}})$; and the *"stiffness"* or *system matrices* $\mathbf{B}$ and $\mathbf{B}^*$ given by

$$\mathbf{B} = s^*_{\Psi^{\mathcal{Y}}} B s_{\Psi^{\mathcal{X}}} = [(B\psi^{\mathcal{X}}_\mu)(\psi^{\mathcal{Y}}_\lambda)]_{\lambda\in\nabla_{\mathcal{Y}},\mu\in\nabla_{\mathcal{X}}} \in \mathcal{L}(\ell^2(\nabla_{\mathcal{X}}), \ell^2(\nabla_{\mathcal{Y}})),$$

and with $\mathbf{B}^*$ defined analogously, are boundedly invertible. We consider the associated bilinear forms

$$\mathfrak{B}(\cdot,\cdot)\ :\ \mathcal{X}\times\mathcal{Y}\to\mathbb{R}:(w,v)\mapsto(Bw)(v),\quad \mathfrak{B}^*(\cdot,\cdot)\ :\ \mathcal{Y}\times\mathcal{X}\to\mathbb{R}:(v,w)\mapsto(B^*v)(w),$$

and introduce the notations

$$\mathbf{B}=\mathfrak{B}(\Psi^{\mathcal{X}},\Psi^{\mathcal{Y}}),\quad \mathbf{f}=f(\Psi^{\mathcal{Y}}),\quad \mathbf{B}^*=\mathfrak{B}^*(\Psi^{\mathcal{Y}},\Psi^{\mathcal{X}}),\quad \mathbf{f}^*=f^*(\Psi^{\mathcal{X}}).$$

With the *Riesz constants*

$$\Lambda^{\mathcal{X}}_{\Psi^{\mathcal{X}}}:=\|s_{\Psi^{\mathcal{X}}}\|_{\ell^2(\nabla_{\mathcal{X}})\to\mathcal{X}}=\sup_{0\neq\mathbf{c}\in\ell^2(\nabla_{\mathcal{X}})}\frac{\|\mathbf{c}^\top\Psi^{\mathcal{X}}\|_{\mathcal{X}}}{\|\mathbf{c}\|_{\ell^2(\nabla_{\mathcal{X}})}},$$

$$\lambda^{\mathcal{X}}_{\Psi^{\mathcal{X}}}:=\|s^{-1}_{\Psi^{\mathcal{X}}}\|^{-1}_{\mathcal{X}\to\ell^2(\nabla_{\mathcal{X}})}=\inf_{0\neq\mathbf{c}\in\ell^2(\nabla_{\mathcal{X}})}\frac{\|\mathbf{c}^\top\Psi^{\mathcal{X}}\|_{\mathcal{X}}}{\|\mathbf{c}\|_{\ell^2(\nabla_{\mathcal{X}})}},$$

and $\Lambda^{\mathcal{Y}}_{\Psi^{\mathcal{Y}}}$ and $\lambda^{\mathcal{Y}}_{\Psi^{\mathcal{Y}}}$ defined analogously, we obviously have that

$$\|\mathbf{B}\|_{\ell^2(\nabla_{\mathcal{X}})\to\ell^2(\nabla_{\mathcal{Y}})}\le\|B\|_{\mathcal{X}\to\mathcal{Y}^*}\Lambda^{\mathcal{X}}_{\Psi^{\mathcal{X}}}\Lambda^{\mathcal{Y}}_{\Psi^{\mathcal{Y}}},\tag{5.2}$$

$$\|\mathbf{B}^{-1}\|_{\ell^2(\nabla_{\mathcal{Y}})\to\ell^2(\nabla_{\mathcal{X}})}\le\frac{\|B^{-1}\|_{\mathcal{Y}^*\to\mathcal{X}}}{\lambda^{\mathcal{X}}_{\Psi^{\mathcal{X}}}\lambda^{\mathcal{Y}}_{\Psi^{\mathcal{Y}}}}.\tag{5.3}$$

6. Best N-term approximations and approximation classes

We say that $\mathbf{u}_N\in\ell^2(\nabla_{\mathcal{X}})$ is a *best N-term approximation* of $\mathbf{u}\in\ell^2(\nabla_{\mathcal{X}})$, if it is the best possible approximation to u in the norm of $\ell^2(\nabla_{\mathcal{X}})$, given a budget of N coefficients. Determining the best N-term approximation $\mathbf{u}_N$ of $\mathbf{u}$ requires searching the infinite vector $\mathbf{u}$, which is not feasible, i.e., actually locating $\mathbf{u}_N$ may not be practically feasible. Nevertheless, rates of convergence afforded by best N-term approximations serve as a *benchmark* for concrete numerical schemes. To this end, we collect all $\mathbf{u}\in\ell^2(\nabla_{\mathcal{X}})$ admitting a best N-term approximation converging with rate $s>0$ in the approximation class $\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}})):=\big\{\mathbf{v}\in\ell^2(\nabla_{\mathcal{X}}):\|\mathbf{v}\|_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))}<\infty\big\}$, where

$$\|\mathbf{v}\|_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))}:=\sup_{\varepsilon>0}\varepsilon\times[\min\{N\in\mathbb{N}_0:\|\mathbf{v}-\mathbf{v}_N\|_{\ell^2(\nabla_{\mathcal{X}})}\le\varepsilon\}]^s,$$

which consists of all $\mathbf{v}\in\ell^2(\nabla_{\mathcal{X}})$ whose best N-term approximations converge to $\mathbf{v}$ with rate s.

Generally, best N-term approximations cannot be realized in practice, in particular not in situations where the vector $\mathbf{u}$ to be approximated is only defined implicitly as the solution of an bi-infinite matrix-vector problem, such as (5.1). We will now design, following [19], a space-time adaptive Galerkin discretization method for the infinite-dimensional Fokker–Planck equation (4.5) that produces a sequence of approximations to $\mathbf{u}$, which, whenever $\mathbf{u}\in\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))$ for some $s>0$, converge to $\mathbf{u}$ *with this rate* $s>0$ in computational complexity that is linear with respect to N, the cardinality of the support of the "active" coefficients in the computed, finitely supported approximation to $\mathbf{u}$.

7. Adaptive Galerkin methods

Let $s>0$ be such that $\mathbf{u}\in\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))$. In [6] and [7], adaptive wavelet Galerkin methods for solving elliptic operator equations (5.1) were introduced; the methods considered in both papers are iterative methods. To be able to bound their complexity, one needs a suitable bound on the complexity of an approximate matrix-vector product in terms of the prescribed tolerance. We formalize this idea through the notion of s^*-admissibility.

Definition 7.1. *The bi-infinite matrices $\mathbf{B}\in\mathcal{L}(\ell^2(\nabla_{\mathcal{X}}),\ell^2(\nabla_{\mathcal{Y}}))$, $\mathbf{B}^*\in\mathcal{L}(\ell^2(\nabla_{\mathcal{Y}}),\ell^2(\nabla_{\mathcal{X}}))$ are* s^*-admissible if there exist routines

$$\mathbf{APPLY}_{\mathbf{B}}[\mathbf{w},\varepsilon]\to\mathbf{z},\quad \mathbf{APPLY}_{\mathbf{B}^*}[\tilde{\mathbf{w}},\varepsilon]\to\tilde{\mathbf{z}}$$

that yield, for any prescribed tolerance $\varepsilon>0$ and any finitely supported $\mathbf{w}\in\ell^2(\nabla_{\mathcal{X}})$ and $\tilde{\mathbf{w}}\in\ell^2(\nabla_{\mathcal{Y}})$, finitely supported vectors $\mathbf{z}\in\ell^2(\nabla_{\mathcal{Y}})$ and $\tilde{\mathbf{z}}\in\ell^2(\nabla_{\mathcal{X}})$ such that

$$\|\mathbf{B}\mathbf{w}-\mathbf{z}\|_{\ell^2(\nabla_{\mathcal{Y}})}+\|\mathbf{B}^*\tilde{\mathbf{w}}-\tilde{\mathbf{z}}\|_{\ell^2(\nabla_{\mathcal{X}})}\le\varepsilon$$

and for which, for any $\bar{s} \in (0, s^*)$, there exists an *admissibility constant* $a_{\mathbf{B},\bar{s}}$ such that

$$\#\mathrm{supp}\,\mathbf{z} \leq a_{\mathbf{B},\bar{s}}\,\varepsilon^{-1/\bar{s}}\|\mathbf{w}\|^{1/\bar{s}}_{\mathcal{A}^{\bar{s}}_\infty(\ell^2(\nabla_\mathcal{X}))},$$

and likewise for $\tilde{\mathbf{z}}$, $\tilde{\mathbf{w}}$. The number of arithmetic operations and storage locations used by the call $\mathbf{APPLY}_\mathbf{B}[\mathbf{w},\varepsilon]$ is bounded by some absolute multiple of

$$a_{\mathbf{B},\bar{s}}\,\varepsilon^{-1/\bar{s}}\|\mathbf{w}\|^{1/\bar{s}}_{\mathcal{A}^{\bar{s}}_\infty(\ell^2(\nabla_\mathcal{X}))} + \#\mathrm{supp}\,\mathbf{w} + 1.$$

The design of $\mathbf{APPLY}_\mathbf{B}[\mathbf{w},\varepsilon]$ and $\mathbf{APPLY}_{\mathbf{B}^*}[\tilde{\mathbf{w}},\varepsilon]$ for the system matrices $\mathbf{B}$ and $\mathbf{B}^*$ arising from the infinite-dimensional Fokker–Planck equation (4.5) is the major technical building block in the adaptive Galerkin discretization of (4.5).

In order to approximate $\mathbf{u}$ one should be able to approximate $\mathbf{f}^*$. Throughout this paper we shall assume the availability of the following routine.

$\mathbf{RHS}_{\mathbf{f}^*}[\varepsilon] \to \mathbf{f}^*_\varepsilon$: *For given $\varepsilon > 0$, the routine yields a finitely supported $\mathbf{f}^*_\varepsilon \in \ell^2(\nabla_\mathcal{X})$ with*

$$\|\mathbf{f}^* - \mathbf{f}^*_\varepsilon\|_{\ell^2(\nabla_\mathcal{X})} \leq \varepsilon \quad \textit{and} \quad \#\mathrm{supp}\,\mathbf{f}^*_\varepsilon \lesssim \min\{N : \|\mathbf{f}^* - \mathbf{f}^*_N\| \leq \varepsilon\},$$

with the number of arithmetic operations and storage locations used by the call $\mathbf{RHS}_{\mathbf{f}^*}[\varepsilon]$ being bounded by some absolute multiple of $\#\mathrm{supp}\,\mathbf{f}^*_\varepsilon + 1$.

The availability of the routines $\mathbf{APPLY}_\mathbf{B}$ and $\mathbf{RHS}_\mathbf{f}$ has the following implications.

Proposition 7.2. *Let $\mathbf{B}$ in (5.1) be s^*-admissible. Then, for any $\bar{s} \in (0, s^*)$, we have that*

$$\|\mathbf{B}\|_{\mathcal{A}^{\bar{s}}_\infty(\ell^2(\nabla_\mathcal{X}))\to\mathcal{A}^{\bar{s}}_\infty(\ell^2(\nabla_\mathcal{Y}))} \leq a^{\bar{s}}_{\mathbf{B},\bar{s}}.$$

For $\mathbf{z}_\varepsilon := \mathbf{APPLY}_\mathbf{B}[\mathbf{w},\varepsilon]$, we have that

$$\|\mathbf{z}_\varepsilon\|_{\mathcal{A}^{\bar{s}}_\infty(\ell^2(\nabla_\mathcal{Y}))} \leq a^{\bar{s}}_{\mathbf{B},\bar{s}}\|\mathbf{w}\|_{\mathcal{A}^{\bar{s}}_\infty(\ell^2(\nabla_\mathcal{X}))}.$$

Analogous statements hold for $\mathbf{B}^*$ in (5.1).

A proof of Proposition 7.2 can be given along the lines of the arguments presented in [6, 7]. Using the definition of $\mathcal{A}^s_\infty(\ell^2(\nabla_\mathcal{Y}))$ and the properties of $\mathbf{RHS}_\mathbf{f}$, we have the following corollary.

Corollary 7.3. *Suppose that, in (5.1), $\mathbf{B}^*$ is s^*-admissible and $\mathbf{u} \in \mathcal{A}^s_\infty(\ell^2(\nabla_\mathcal{Y}))$ for $s < s^*$; then, for $\mathbf{f}^*_\varepsilon = \mathbf{RHS}_{\mathbf{f}^*}[\varepsilon]$,*

$$\#\mathrm{supp}\,\mathbf{f}^*_\varepsilon \lesssim a_{\mathbf{B}^*,s}\,\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_\mathcal{Y}))},$$

with the number of arithmetic operations and storage locations used by the call $\mathbf{RHS}_{\mathbf{f}^*}[\varepsilon]$ being bounded by some absolute multiple of

$$a_{\mathbf{B}^*,s}\,\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_\mathcal{Y}))} + 1.$$

Remark 7.4. *Besides $\|\mathbf{f}^* - \mathbf{f}^*_\varepsilon\|_{\ell^2(\nabla_\mathcal{Y})} \leq \varepsilon$, the complexity bounds in Corollary 7.3, with $a_{\mathbf{B}^*,s}$ signifying a constant that depends only on $\mathbf{B}^*$ and s, are essential for the use of $\mathbf{RHS}_{\mathbf{f}^*}$ in the adaptive Galerkin methods.*

The following corollary of Proposition 7.2 from [19] can be used for example for the construction of valid **APPLY** and **RHS** routines in case the adaptive Galerkin algorithms are applied to a preconditioned system.

Corollary 7.5. *Suppose that $\mathbf{B}^* \in \mathcal{L}(\ell^2(\nabla_\mathcal{Y}), \ell^2(\nabla_\mathcal{X}))$, $\mathbf{C} \in \mathcal{L}(\ell^2(\nabla_\mathcal{X}), \ell^2(\nabla_\mathcal{Z}))$ are both s^*-admissible; then, so is $\mathbf{CB}^* \in \mathcal{L}(\ell^2(\nabla_\mathcal{Y}), \ell^2(\nabla_\mathcal{Z}))$. A valid routine $\mathbf{APPLY}_{\mathbf{CB}^*}$ is*

$$[\mathbf{w},\varepsilon] \mapsto \mathbf{APPLY}_\mathbf{C}\big[\mathbf{APPLY}_{\mathbf{B}^*}[\mathbf{w},\varepsilon/(2\|\mathbf{C}\|)],\varepsilon/2\big], \tag{7.1}$$

with admissibility constant $a_{\mathbf{CB}^*,\bar{s}} \lesssim a_{\mathbf{B}^*,\bar{s}}(\|\mathbf{C}\|^{1/\bar{s}} + a_{\mathbf{C},\bar{s}})$ for $\bar{s} \in (0, s^*)$.

For some $s^* > s$, let $\mathbf{C} \in \mathcal{L}(\ell^2(\nabla_\mathcal{Y}), \ell^2(\nabla_\mathcal{Z}))$ be s^*-admissible. Then, for

$$\mathbf{RHS}_{\mathbf{Cf}^*}[\varepsilon] := \mathbf{APPLY}_\mathbf{C}[\mathbf{RHS}_{\mathbf{f}^*}[\varepsilon/(2\|\mathbf{C}\|)],\varepsilon/2], \tag{7.2}$$

we have that

$$\#\mathrm{supp}\,\mathbf{RHS}_{\mathbf{Cf}^*}[\varepsilon] \lesssim a_{\mathbf{B}^*,s}(\|\mathbf{C}\|^{1/s} + a_{\mathbf{C},s})\,\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_\mathcal{Y}))}$$

and $\|\mathbf{Cf}^* - \mathbf{RHS}_{\mathbf{Cf}^*}[\varepsilon]\|_{\ell^2(\nabla_{\bar{\mathcal{Z}}})} \le \varepsilon$*, with the number of arithmetic operations and storage locations used by the call* $\mathbf{RHS}_{\mathbf{Cf}^*}[\varepsilon]$ *being bounded by some absolute multiple of*

$$a_{\mathbf{B}^*,s}(\|\mathbf{C}\|^{1/s} + a_{\mathbf{C},s})\,\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))} + 1.$$

Remark 7.6. *The properties of* $\mathbf{RHS}_{\mathbf{Cf}^*}$ *stated in Corollary 7.5 show that* $\mathbf{RHS}_{\mathbf{Cf}^*}$ *is a valid routine for approximating* $\mathbf{Cf}^*$ *in the sense of Remark 7.4.*

In the particular case when $\mathbf{B}$ is *symmetric positive definite*, i.e., $\nabla_{\mathcal{X}} = \nabla_{\mathcal{Y}}$ and $\mathbf{B} = \mathbf{B}^* > 0$, the two Galerkin methods considered in the papers [6, 7] were shown to be *quasi-optimal* in the following sense.

Theorem 7.7. *Suppose that, in (5.1),* $\mathbf{B}^*$ *is* s^**-admissible; then, for any* $\varepsilon > 0$ *the two adaptive Galerkin methods from* [6, 7] *produce an approximation* $\mathbf{u}_\varepsilon$ *to* $\mathbf{u}$ *with* $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell^2(\nabla_{\mathcal{Y}})} \le \varepsilon$*. Suppose that, in* (5.1)*, for some* $s > 0$ *we have that* $\mathbf{u} \in \mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))$*; then* $\#\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))}$ *and if, moreover,* $s < s^*$*, then the number of arithmetic operations and storage locations required by a call of either of these adaptive solvers with tolerance* $\varepsilon > 0$ *is bounded by some multiple of*

$$\varepsilon^{-1/s}(1 + a_{\mathbf{B},s})\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))} + 1.$$

The multiples depend on s *only when* s *tends to* 0 *or* ∞*, and on* $\|\mathbf{B}\|$ *and* $\|\mathbf{B}^{-1}\|$ *when they tend to infinity.*

The method from [7] is based on a damped Richardson iteration for the approximate solution of $\mathbf{Bu} = \mathbf{f}$, where the required residual computations are approximated using calls of $\mathbf{APPLY}_{\mathbf{B}}$ and $\mathbf{RHS}_{\mathbf{f}}$ within tolerances that decrease linearly with the iteration counter.

The method from [6] produces a sequence $\Xi_0 \subset \Xi_1 \subset \cdots \subset \nabla_{\mathcal{X}}$, together with a corresponding sequence of (approximate) finitely supported Galerkin solutions $\mathbf{u}_i \in \ell^2(\Xi_i)$. The coefficients of approximate residuals $\mathbf{f} - \mathbf{Bu}_i$ are used to steer the expansion of the sets $\{\Xi_i\}_{i\ge 0}$.

Both methods rely on a recurrent coarsening of the approximation vectors, where small coefficients are removed in order to keep an optimal balance between accuracy and support length; in [12], a modification of the algorithm introduced in [6] was proposed, which does not require the coarsening step.

The s^*-admissibility of $\mathbf{B}$ can be expected, since the structure of $\mathbf{B}$ in the multiresolution Riesz-bases in $[0,T]$ and in Wiener–Hermite polynomial chaos bases of $L^2(H;\mu)$ is, as we will show in what follows, close to that of a computable sparse matrix.

Definition 7.8. $\mathbf{B} \in \mathcal{L}(\ell^2(\nabla_{\mathcal{X}}), \ell^2(\nabla_{\mathcal{Y}}))$ *is* s^*-computable *if, for each* $N \in \mathbb{N}$*, there exists a* $\mathbf{B}^{[N]} \in \mathcal{L}(\ell^2(\nabla_{\mathcal{X}}), \ell^2(\nabla_{\mathcal{Y}}))$ *having in each column at most* N *nonzero entries whose joint computation takes an absolute multiple of* N *operations, such that the* computability constants

$$c_{\mathbf{B},\bar{s}} := \sup_{N\in\mathbb{N}} N\|\mathbf{B} - \mathbf{B}^{[N]}\|^{1/\bar{s}}_{\ell^2(\nabla_{\mathcal{X}})\to\ell^2(\nabla_{\mathcal{Y}})} \tag{7.3}$$

are finite for any $\bar{s} \in (0, s^*)$*. The notion of* s^**-computability of* $\mathbf{B}^*$ *is defined analogously.*

Proposition 7.9. *An* s^**-computable* $\mathbf{B}$ *is* s^**-admissible. Moreover, for* $\bar{s} < s^*$*,* $a_{\mathbf{B},\bar{s}} \lesssim c_{\mathbf{B},\bar{s}}$ *where the constant in this bound depends only on* $\bar{s} \downarrow 0$*,* $\bar{s} \uparrow s^*$*, and on* $\|\mathbf{B}\| \to \infty$*.*

This is proved by constructing a suitable $\mathbf{APPLY}_{\mathbf{B}}$ routine as in [6, §6.4] (a log factor in the complexity bound there due to sorting was removed later by application of an approximate sorting, see [11] and the references there).

Remark 7.10. *Theorem 7.7 requires that* $\mathbf{B}$ *is* s^**-admissible for an* $s^* > s$ *when* $\mathbf{u} \in \mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))$*. Generally this value of* s *is unknown, and so the condition on* s^* *should be interpreted in the sense that* s^* *has to be larger than any* s *for which membership of the solution* $\mathbf{u}$ *in* $\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))$ *can be expected.*

The approach from [7] applies also to the saddle point variational principle (4.5) whenever one has a linearly convergent stationary iterative scheme available for the matrix-vector problem $\mathbf{B}^*\mathbf{u} = \mathbf{f}^*$. There is, unfortunately, no such scheme available for a general boundedly invertible $\mathbf{B}^*$.

In particular, for the stiffness matrices $\mathbf{B}^*$ resulting from the space-time saddle-point formulation (4.5) of the infinite-dimensional Fokker–Planck equation no directly applicable scheme is available. In [7], adaptive Galerkin discretizations for such problems are proposed for the *normal equation*

$$\mathbf{B}\mathbf{B}^*\mathbf{u} = \mathbf{B}\mathbf{f}^*. \tag{7.4}$$

By Theorem 4.3, the operator $\mathbf{B}\mathbf{B}^* \in \mathcal{L}(\ell^2(\nabla_{\mathcal{Y}}), \ell^2(\nabla_{\mathcal{Y}}))$ is boundedly invertible, symmetric positive definite, with

$$\|\mathbf{B}\mathbf{B}^*\|_{\ell^2(\nabla_{\mathcal{Y}})\to\ell^2(\nabla_{\mathcal{Y}})} \le \|\mathbf{B}^*\|^2_{\ell^2(\nabla_{\mathcal{Y}})\to\ell^2(\nabla_{\mathcal{X}})},\ \|(\mathbf{B}\mathbf{B}^*)^{-1}\|_{\ell^2(\nabla_{\mathcal{Y}})\to\ell^2(\nabla_{\mathcal{Y}})} \le \|(\mathbf{B}^*)^{-1}\|^2_{\ell^2(\nabla_{\mathcal{X}})\to\ell^2(\nabla_{\mathcal{Y}})}.$$

Now let $\mathbf{u} \in \mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{X}}))$, and for some $s^* > s$, let $\mathbf{B}$ and $\mathbf{B}^*$ be s^*-admissible. By Corollary 7.5, with $\mathbf{B}^*$ in place of $\mathbf{C}$, a valid $\mathbf{RHS}_{\mathbf{B}\mathbf{f}^*}$ routine is given by (7.2), and $\mathbf{B}\mathbf{B}^*$ is s^*-admissible with a valid $\mathbf{APPLY}_{\mathbf{B}\mathbf{B}^*}$ routine given by (7.1). In the context of (4.5), one execution of $\mathbf{APPLY}_{\mathbf{B}\mathbf{B}^*}$ corresponds to one (approximate) sweep over the "primal" problem, followed by one (approximate) sweep over the dual problem, respectively. A combination of Theorem 7.7 and Corollary 7.5 yields the following result, obtained in [19], which will be the basis for our Wiener–Hermite polynomial chaos discretization.

Theorem 7.11. *For any $\varepsilon > 0$, the adaptive wavelet methods from* [7] *or from* [6] *and* [12] *applied to the normal equations* (7.4) *using the above $\mathbf{APPLY}_{\mathbf{B}\mathbf{B}^*}$ and $\mathbf{RHS}_{\mathbf{B}f^*}$ routines produce an approximation $\mathbf{u}_\varepsilon$ to $\mathbf{u}$ with $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell^2(\nabla_{\mathcal{Y}})} \le \varepsilon$.*

Suppose that for some $s > 0$, $\mathbf{u} \in \mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))$; then $\#\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))}$, with the constant in this bound only being dependent on s when it tends to 0 or ∞, and on $\|\mathbf{B}^*\|$ and $\|(\mathbf{B}^*)^{-1}\|$ when they tend to infinity.

Suppose that $s < s^*$; then, the number of arithmetic operations and storage locations required by a call of either of these adaptive wavelet methods with tolerance ε is bounded by some multiple of

$$1 + \varepsilon^{-1/s}(1 + a_{\mathbf{B}^*,s}(1 + a_{\mathbf{B},s}))\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))}$$

where this multiple only depends on s when it tends to 0 or ∞, and on $\|\mathbf{B}^*\|$ and $\|(\mathbf{B}^*)^{-1}\|$ when they tend to infinity.

8. Infinite-dimensional Fokker–Planck equation as bi-infinite matrix vector equation

We apply the foregoing abstract concepts to the space-time variational formulation (4.5) of the infinite-dimensional Fokker–Planck equation. To construct Riesz bases for $\mathcal{X}_{0,\{T\}}$ and $\mathcal{Y}$, we use that

$$\mathcal{X}_{0,\{T\}} = (\mathrm{L}^2(I)\otimes\mathcal{V}) \cap (\mathrm{H}^1_{0,\{T\}}(I)\otimes\mathcal{V}^*) \quad \text{and} \quad \mathcal{Y} = \mathrm{L}^2(I)\otimes\mathcal{V}$$

with the spaces $\mathcal{X}_{0,\{T\}}$ as defined in (4.4), and $\mathcal{Y} = \mathrm{L}^2(0,T;\mathcal{V})$; recall that $\mathcal{H} = \mathrm{L}^2(D,\mu)$ and $\mathcal{V} = \mathrm{W}^{1,2}(D,\mu)$, with $\mathcal{V}$ continuously, densely and compactly embedded into $\mathcal{H}$. Let

$$\Upsilon = \{H_\gamma : \gamma \in \Gamma\} \subset \mathcal{V}\,, \quad \text{with} \quad H_\gamma(q) = \prod_{k\ge1} H_{\gamma_k}\left(\frac{q_k}{\sqrt{\lambda_k}}\right),$$

denote the Wiener–Hermite polynomial chaos basis (3.3). By Theorem 3.3, Υ is a *normalized Riesz basis* for $[\mathrm{L}^2(H,\mu)]^d$, which, when rescaled by the scaling sequence $\{(\langle\gamma,\lambda^{-1}\rangle_*)^{-1/2} : \gamma\in\Gamma\}$ in $\mathcal{V}$ and by the scaling sequence $\{(\langle\gamma,\lambda^{-1}\rangle_*)^{1/2} : \gamma\in\Gamma\}$ in $\mathcal{V}^*$, is a Riesz basis for $\mathcal{V}$ and $\mathcal{V}^*$, respectively; here the scaling factors are defined by

$$\langle\gamma,\lambda^{-1}\rangle_* := \begin{cases} \langle\gamma,\lambda^{-1}\rangle & \text{if } \gamma\in\Gamma\setminus\{0\},\\ 1 & \text{if } \gamma = 0; \end{cases} \tag{8.1}$$

see (3.6), (3.7) and also (9.1), (9.2) ahead. In the time interval $I = (0,T)$, we denote by

$$\Theta = \{\theta_\lambda : \lambda\in\nabla_t\} \subset \mathrm{H}^1_{0,\{T\}}(I)$$

a collection of functions indexed by the countable set ∇_t that is a *normalized Riesz basis for* $\mathrm{L}^2(I)$, *which, when rescaled in* $\mathrm{H}^1(I)$, *is a Riesz basis for* $\mathrm{H}^1_{0,\{T\}}(I)$. It then follows from [13, Prop. 1 and 2] that the collection $\Theta \otimes \Upsilon$, normalized in $\mathcal{X}$, i.e., the collection

$$\left\{ (t,q) \mapsto \frac{\theta_\lambda(t) H_\gamma(q)}{\sqrt{\langle \gamma, \lambda^{-1}\rangle_* + \|\theta_\lambda\|^2_{\mathrm{H}^1(I)}/\langle \gamma, \lambda^{-1}\rangle_*}} : (\lambda,\gamma) \in \nabla_{\mathcal{X}} := \nabla_t \times \Gamma \right\},$$

is a Riesz basis for $\mathcal{X}_{0,\{T\}}$, and that $\Theta \otimes \Upsilon$ normalized in $\mathcal{Y}$, i.e., the collection

$$\Big\{ (t,q) \mapsto \frac{\theta_\lambda(t) H_\gamma(q)}{\|H_\gamma\|_{\mathcal{V}}} : (\lambda,\gamma) \in \nabla_{\mathcal{Y}} := \nabla_t \times \Gamma \Big\},$$

is a Riesz basis for $\mathcal{Y}$. Moreover, denoting the Riesz basis for $\mathcal{V}^*$ consisting of the collection Υ normalized in $\mathcal{V}^*$ by $[\Upsilon]_{\mathcal{V}^*}$, and similarly for the other collections and spaces, with the notations introduced in Sect. 5, we have that

$$\Lambda^{\mathcal{X}}_{[\Theta\otimes\Upsilon]_{\mathcal{X}}} \le \max(\Lambda^{\mathrm{L}^2(I)}_{\Theta} \Lambda^{\mathcal{V}}_{[\Upsilon]_{\mathcal{V}}}, \Lambda^{\mathrm{H}^1(I)}_{[\Theta]_{\mathrm{H}^1(I)}} \Lambda^{\mathcal{V}^*}_{[\Upsilon]_{\mathcal{V}^*}}), \tag{8.2}$$

$$\lambda^{\mathcal{X}}_{[\Theta\otimes\Upsilon]_{\mathcal{X}}} \ge \min(\lambda^{\mathrm{L}^2(I)}_{\Theta} \lambda^{\mathcal{V}}_{[\Upsilon]_{\mathcal{V}}}, \lambda^{\mathrm{H}^1(I)}_{[\Theta]_{\mathrm{H}^1(I)}} \lambda^{\mathcal{V}^*}_{[\Upsilon]_{\mathcal{V}^*}}), \tag{8.3}$$

$$\Lambda^{\mathcal{Y}}_{[\Theta\otimes\Upsilon]_{\mathcal{Y}}} \le \Lambda^{\mathrm{L}^2(I)}_{\Theta} \Lambda^{\mathcal{V}}_{[\Upsilon]_{\mathcal{V}}}, \tag{8.4}$$

$$\lambda^{\mathcal{Y}}_{[\Theta\otimes\Upsilon]_{\mathcal{Y}}} \ge \lambda^{\mathrm{L}^2(I)}_{\Theta} \lambda^{\mathcal{V}}_{[\Upsilon]_{\mathcal{V}}}. \tag{8.5}$$

Denoting by $\|\Upsilon\|_{\mathcal{V}}$ the infinite diagonal matrix with diagonal entries $\|H_\gamma\|_{\mathcal{V}}$ where $\gamma \in \Gamma$, and similarly for the other collections and spaces, the stiffness or system matrix $\mathbf{B}^*$ corresponding to the variational form (4.5) and the Riesz bases $[\Theta \otimes \Upsilon]_{\mathcal{Y}}$, $[\Theta \otimes \Upsilon]_{\mathcal{X}}$ for $\mathcal{Y}$ and $\mathcal{X}_{0,\{T\}}$ is given by the bi-infinite matrix

$$\begin{aligned}\mathbf{B}^* &= \mathfrak{B}^*([\Theta \otimes \Upsilon]_{\mathcal{Y}}, [\Theta \otimes \Upsilon]_{\mathcal{X}}) \\ &= \left[\mathrm{Id}_t \otimes \|\Upsilon\|_{\mathcal{V}}^{-1}\right] \circ \left[-(\Theta,\Theta')_{\mathrm{L}^2(I)} \otimes (\Upsilon,\Upsilon)_{\mathcal{H}} + \int_I \mathfrak{a}(\Theta\otimes\Upsilon, \Theta\otimes\Upsilon)\,\mathrm{d}t \right] \circ \left[\|\Theta\otimes\Upsilon\|_{\mathcal{X}}^{-1}\right],\end{aligned}$$

where Id_t denotes the identity operator with respect to the t variable and the symbol $\circ$ signifies composition of operators.

Writing the solution u of (4.5) as $u = \mathbf{u}^\top [\Theta \otimes \Sigma]_{\mathcal{X}}$, we deduce that $\mathbf{u}$ is the solution of the bi-infinite matrix-vector equation $\mathbf{B}^*\mathbf{u} = \mathbf{f}^*$ with

$$\mathbf{f}^* := \left[(\widehat{\psi}_0, [\Upsilon]_{\mathcal{H}})_{\mathcal{H}} \right]. \tag{8.6}$$

Introducing the infinite diagonal matrices

$$\mathbf{D}_1 := (\|\Theta\|_{\mathrm{H}^1(I)} \otimes \|\Upsilon\|_{\mathcal{V}^*}) \|\Theta\otimes\Upsilon\|_{\mathcal{X}}^{-1} \quad \text{and} \quad \mathbf{D}_2 := (\mathrm{Id}_t \otimes \|\Upsilon\|_{\mathcal{V}}) \|\Theta\otimes\Upsilon\|_{\mathcal{X}}^{-1},$$

both being diagonal matrices with entries in modulus less than 1, the bi-infinite matrix operator $\mathbf{B}^*$ can be written as

$$\mathbf{B}^* = \left[(-[\Theta]_{\mathrm{H}^1_{0,\{T\}}(I)^*}, \Theta')_{\mathrm{L}^2(I)} \otimes ([\Upsilon]_{\mathcal{V}^*}, [\Upsilon]_{\mathcal{V}})_{\mathcal{H}}\, \mathbf{D}_1 + \int_I \mathfrak{a}(\Theta\otimes[\Upsilon]_{\mathcal{V}}, \Theta\otimes[\Upsilon]_{\mathcal{V}})\,\mathrm{d}t\, \mathbf{D}_2 \right]. \tag{8.7}$$

9. s^*-admissibility of $\mathbf{B}$ and $\mathbf{B}^*$ from (8.7) and of its adjoint

By Theorem 7.9, the s^*-admissibility of $\mathbf{B}$ and $\mathbf{B}^*$ follows from their s^*-computability. To verify the s^*-computability of tensor products of possibly bi-infinite matrices, it suffices to analyze the s^*-computability of the factors, according to the following result from [19, 15].

Proposition 9.1. *Let, for some* $s^* > 0$, $\mathbf{D}$ *and* $\mathbf{E}$ *be* s^**-computable. Then,*

(a) $\mathbf{D} \otimes \mathbf{E}$ *is* s^**-computable with computability constant satisfying, for* $0 < \bar{s} < \tilde{s} < s^*$, $c_{\mathbf{D}\otimes\mathbf{E},\bar{s}} \lesssim (c_{\mathbf{D},\tilde{s}}\, c_{\mathbf{E},\tilde{s}})^{\tilde{s}/\bar{s}}$*; and*

(b) *for any* $\varepsilon \in (0, s^*)$, $\mathbf{D} \otimes \mathbf{E}$ *is* $(s^*-\varepsilon)$*–computable, with computability constant* $c_{\mathbf{D}\otimes\mathbf{E},\bar{s}}$ *satisfying, for* $0 < \bar{s} < s^* - \varepsilon < \tilde{s} < s^*$, $c_{\mathbf{D}\otimes\mathbf{E},\bar{s}} \lesssim \max(c_{\mathbf{D},\tilde{s}}, 1) \max(c_{\mathbf{E},\tilde{s}}, 1)$.

The constants absorbed in the $\lesssim$ symbol in the bounds on the computability constants in (a) and (b) are only dependent on $\tilde{s} \downarrow 0$, $\tilde{s} \to \infty$ and $\tilde{s} - \bar{s} \downarrow 0$.

In view of the representation (8.7) of $\mathbf{B}^*$, using Corollary 7.5 and Proposition 9.1, part (a), for proving s^*-admissibility of $\mathbf{B}$ and $\mathbf{B}^*$ it suffices to show that $(\Upsilon, \Upsilon)_{\mathcal{H}}$ and $\int_I \mathfrak{a}(\Theta \otimes [\Upsilon]_{\mathcal{V}}, \Theta \otimes [\Upsilon]_{\mathcal{V}})\,\mathrm{d}t$ and its adjoint are s^*-admissible. To prove this, by Theorem 7.9 we need to verify that these objects are s^*-computable. This will follow from the fact that $([\Theta]'_{\mathrm{H}^1(I)}, \Theta)_{\mathrm{L}^2(I)}$, $(\Theta, [\Theta]'_{\mathrm{H}^1(I)})_{\mathrm{L}^2(I)}$, $([\Upsilon]_{\mathcal{V}^*}, [\Upsilon]_{\mathcal{V}})_{\mathcal{H}}$ and $([\Upsilon]_{\mathcal{V}}, [\Upsilon]_{\mathcal{V}^*})_{\mathcal{H}}$ are s^*-computable and, in view of the definition of $\mathfrak{a}(\cdot,\cdot)$ in (4.6), from a sparsity assumption on the coefficient matrix $\mathbf{A}$.

9.1. **Choice of Riesz bases Θ and Υ.** We have already assumed that $\Theta = \{\theta_\lambda : \lambda \in \nabla_t\}$ is a normalized Riesz basis of $\mathrm{L}^2(I)$ that, when rescaled in $\mathrm{H}^1(I)$ is a Riesz basis for $\mathrm{H}^1_{0,\{T\}}(I)$. We will now select the basis Θ to be a *wavelet basis* that satisfies certain additional assumptions. Specifically, we shall assume that the basis functions θ_λ of Θ are:

- (t1) *local*, i.e., $\sup_{x\in[0,1],\ell\in\mathbb{N}_0} \#\{|\lambda| = \ell : x \in \operatorname{supp}\theta_\lambda\} < \infty$ and $\operatorname{supp}\theta_\lambda \lesssim 2^{-|\lambda|}$;
- (t2) *piecewise polynomial of degree d_t*, where by "piecewise" we mean that the singular support consists of a set of points whose cardinality is uniformly bounded;
- (t3) *globally continuous*, specifically $\|\theta_\lambda\|_{W^k_\infty(0,1)} \lesssim 2^{|\lambda|(\frac{1}{2}+k)}$ for $k \in \{0,1\}$;
- (t4) for $|\lambda| > 0$, have $\tilde{d}_t \geq d_t$ *vanishing moments*.

The assumptions (t1)–(t4) can be met by wavelet constructions (see, e.g. [5] and the references therein).

As a Riesz basis in D we choose the (countable) "polynomial chaos" basis $\Upsilon = \{H_\gamma : \gamma \in \Gamma\}$. According to Theorem 3.3, Υ is an orthonormal Riesz basis of $\mathrm{L}^2(D,\mu)$ that, rescaled in $\mathcal{V} = \mathrm{W}^{1,2}(D,\mu)$ or its dual $\mathcal{V}^*$, is a Riesz basis for these spaces, respectively. Indeed, on denoting by $\mathbf{D}^\lambda$ the diagonal matrix

$$\mathbf{D}^\lambda = \operatorname{diag}\{(\langle\gamma, \lambda^{-1}\rangle_*)^{1/2} : \gamma \in \Gamma\}, \tag{9.1}$$

we observe that the diagonal entries of $\mathbf{D}^\lambda$ relate to the $\mathcal{V}$ and $\mathcal{V}^*$ norms of $[\Upsilon]_{\mathcal{H}}$ as follows:

$$\mathbf{D}^\lambda \simeq \|[\Upsilon]_{\mathcal{H}}\|_{\mathcal{V}}, \quad (\mathbf{D}^\lambda)^{-1} \simeq \|[\Upsilon]_{\mathcal{H}}\|_{\mathcal{V}^*}\,. \tag{9.2}$$

Therefore, the collection $[\Upsilon]_{\mathcal{V}} := (\mathbf{D}^\lambda)^{-1}[\Upsilon]_{\mathcal{H}} = \{(D^\lambda_\gamma)^{-1}H_\gamma : \gamma \in \Gamma\}$ is a Riesz basis of $\mathcal{V} = \mathrm{W}^{1,2}(D,\mu)$. This follows readily from the $\mathrm{L}^2(D,\mu)$ orthonormality of Υ and from the identity (cf. [9, (9.2.11)])

$$\mathrm{D}_k H_\gamma = \gamma_k^{1/2}\lambda_k^{-1/2} H_{\gamma_k - 1}(W_{e_k}) H^{(k)}_\gamma, \qquad \text{for} \quad k = 1, 2, \ldots \quad \text{and} \quad \gamma \in \Gamma, \tag{9.3}$$

with the notational conventions $H_{-1}(W_{e_k}) := 0$ and $H^{(k)}_\gamma := \prod_{j\neq k} H_{\gamma_j}$ for $\gamma \in \Gamma$. By duality, the collection $[\Upsilon]_{\mathcal{V}^*} := \mathbf{D}^\lambda[\Upsilon]_{\mathcal{H}}$ is a Riesz basis of $\mathcal{V}^*$.

9.2. **s^*-computability of $([\Theta]'_{\mathrm{H}^1(I)}, \Theta)_{\mathrm{L}^2(I)}$ and of its adjoint.** By (t1), (t2) and (t4), for each $\lambda \in \nabla_t$ and $\ell \in \mathbb{N}_0$, the number of $\mu \in \nabla_t$ with $|\mu| = \ell$ and $\int_I \theta'_\lambda \theta_\mu\,\mathrm{d}t \neq 0$ or $\int_I \theta'_\mu \theta_\lambda\,\mathrm{d}t \neq 0$ is bounded, uniformly in λ and ℓ. Indeed, $\int_I \theta'_\lambda\theta_\mu\,\mathrm{d}t$ can only be nonzero when θ_μ does not vanish on the singular support of θ_λ, and using integration by parts, $\int_I \theta'_\mu\theta_\lambda\,\mathrm{d}t$ can only be nonzero when θ_μ does not vanish on the singular support of θ_λ or at ∂I.

As a consequence of Θ being of degree $d_t \geq 1$ we have that

$$2^{|\lambda|} = 2^{|\lambda|}\|\theta_\lambda\|_{\mathrm{L}^2(I)} = 2^{|\lambda|}\|(\mathrm{Id} - Q_{|\lambda|-1})\theta_\lambda\|_{\mathrm{L}^2(I)} \lesssim \|\theta_\lambda\|_{\mathrm{H}^1(I)}.$$

Here, for $m \in \mathbb{N}_0$, Q_m denotes the (Θ, Θ') biorthogonal projector onto $\operatorname{span}\{\theta_\lambda : |\lambda| \leq m\}$. Using (t1) and (t3), we infer that

$$\|\theta_\lambda\|^{-1}_{\mathrm{H}^1(I)}\Big|\int_I \theta'_\lambda\theta_\mu\,\mathrm{d}t\Big| \lesssim 2^{-|\lambda|}2^{-\max(|\lambda|,|\mu|)}2^{|\lambda|(\frac{1}{2}+1)}2^{\frac{1}{2}|\mu|} = 2^{-\frac{1}{2}\left||\lambda|-|\mu|\right|}. \tag{9.4}$$

Finally, we note that any entry of $([\Theta]'_{\mathrm{H}^1(I)}, \Theta)_{\mathrm{L}^2(I)}$ can be evaluated in closed form in $\mathcal{O}(1)$ work and memory. Schur's Lemma (cf. [18], p. 6, Par. 2) now implies that $([\Theta]'_{\mathrm{H}^1(I)}, \Theta)_{\mathrm{L}^2(I)}$ and $(\Theta, [\Theta]'_{\mathrm{H}^1(I)})_{\mathrm{L}^2(I)}$ are ∞-computable.

Note that from $\operatorname{diam}(\operatorname{supp}\theta_\lambda) \lesssim 2^{-|\lambda|}$ (cf. (t1)) and from and (t3) we deduce that $\|\theta_\lambda\|_{\mathrm{H}^1(I)} \lesssim 2^{|\lambda|}$, and thus that

$$\|\theta_\lambda\|_{\mathrm{H}^1(I)} \eqsim 2^{|\lambda|}.$$

Remark 9.2. *Suppose that, instead of* (t3)*, the θ_λ belong also to $\mathrm{C}^{r_t}(I)$ for some $r_t \in \mathbb{N}$ (necessarily with $r_t \le d_t - 2$), i.e., that $\|\theta_\lambda\|_{\mathrm{W}^{s,\infty}(0,1)} \lesssim 2^{|\lambda|(\frac{1}{2}+s)}$ for $s \in \{0, r_t+1\}$. Then, by subtracting a suitable polynomial of degree r_t from θ'_λ in* (9.4)*, and using that θ_μ has $\tilde{d}_t \ge d_t \ge r_t$ vanishing moments one deduces that, for $|\lambda| \le |\mu|$,*

$$\|\theta_\lambda\|^{-1}_{\mathrm{H}^1(I)} \left| \int_I \theta'_\lambda \theta_\mu \,\mathrm{d}t \right| \lesssim 2^{-\left| \,|\lambda|-|\mu|\, \right| (\frac{1}{2}+r_t)}.$$

Similarly, for $|\lambda| \ge |\mu|$, using integration by parts one obtains that

$$\|\theta_\lambda\|^{-1}_{\mathrm{H}^1(I)} \left| \int_I \theta'_\lambda \theta_\mu \,\mathrm{d}t \right| \lesssim 2^{-\left||\lambda|-|\mu|\right| (\frac{3}{2}+r_t)}$$

if $t \mapsto \theta_\lambda(t)\,\theta_\mu(t)$ vanishes on ∂I. Since the wavelets in time do not satisfy a homogeneous Dirichlet boundary condition, there are $\lambda, \mu \in \nabla_t$ with $|\lambda| \ge |\mu|$ for which $t \mapsto \theta_\lambda(t)\theta_\mu(t)$ does not vanish at the boundary. For such entries (9.4) *cannot be improved.*

Remark 9.3. *In reference* [5]*, multiresolutions $\Theta = (\theta_\lambda)_{\lambda\in\nabla_t}$ were constructed for which the matrices $([\Theta]'_{\mathrm{H}^1(I)}, \Theta)_{\mathrm{L}^2(I)}$ are* sparse without compression *and which satisfy a homogeneous Dirichlet boundary condition at $t = T$.*

9.3. **s^*-computability of $(\Upsilon,\Upsilon)_{\mathcal{H}}$ and $([\Upsilon]_{\mathcal{V}^*},[\Upsilon]_{\mathcal{V}})_{\mathcal{H}}$ and of its adjoint.** In view of the $\mathrm{L}^2(H,\mu)$-orthonormality (3.5) of the collection Υ of Hermite polynomials, both $(\Upsilon,\Upsilon)_{\mathcal{H}}$ and $([\Upsilon]_{\mathcal{V}^*},[\Upsilon]_{\mathcal{V}})_{\mathcal{H}}$ and their adjoints are diagonal bi-infinite matrices that are therefore ∞-computable.

9.4. **s^*-computability of $\int_I \mathfrak{a}(\Theta\otimes[\Upsilon]_{\mathcal{V}}, \Theta\otimes[\Upsilon]_{\mathcal{V}})\,\mathrm{d}t$ and of its adjoint.** We have that

$$\int_I \mathfrak{a}(\Theta\otimes[\Upsilon]_{\mathcal{V}}, \Theta\otimes[\Upsilon]_{\mathcal{V}})\,\mathrm{d}t = (\Theta,\Theta)_{\mathrm{L}^2(I)} \otimes \mathfrak{a}([\Upsilon]_{\mathcal{V}},[\Upsilon]_{\mathcal{V}}). \tag{9.5}$$

Since the form $\mathfrak{a}(\cdot,\cdot)$ defined in (4.7) is symmetric, by Proposition 9.1 it suffices to investigate s^*-computability of each of the two factors on the right-hand side of (9.5) in order to deduce s^*-computability of $\mathbf{B}$ and of $\mathbf{B}^*$.

As was noted in [19, Sec. 8.5], under the assumptions (t1)–(t4) above, $(\Theta,\Theta)_{\mathrm{L}^2(I)}$ is ∞-computable, so it remains to address the s^*-computability of $\mathfrak{a}([\Upsilon]_{\mathcal{V}},[\Upsilon]_{\mathcal{V}})$.

In view of the definition (4.7) of $\mathfrak{a}(\cdot,\cdot)$ in (4.6) and of Definition 7.8, s^*-computability of the bi-infinite matrix $\mathbf{G} = \mathfrak{a}([\Upsilon]_{\mathcal{V}},[\Upsilon]_{\mathcal{V}})$ depends on the structure of the bi-infinite matrix $\mathbf{A} = (A_{ij})_{i,j=1}^{\infty}$ in (4.6). Thanks to (9.5),

$$\mathfrak{a}([\Upsilon]_{\mathcal{V}},[\Upsilon]_{\mathcal{V}}) = (\mathbf{G}_{\gamma,\gamma'})_{\gamma,\gamma\in\Gamma},$$

where the entries $G_{\gamma\gamma'}$ are, for $\gamma,\gamma'\in\Gamma$, evaluated (assuming $d=1$ for ease of notation and using (9.3)) as follows:

$$\begin{aligned} G_{\gamma\gamma'} &= \sum_{i,j\ge1} A_{ij}(D^\lambda_\gamma)^{-1}(\mathrm{D}_{q_i}H_\gamma, \mathrm{D}_{q_j}H_{\gamma'})_{\mathrm{L}^2(D,\mu)}(D^\lambda_{\gamma'})^{-1} \\ &= \sum_{i,j\ge1} A_{ij}(D^\lambda_\gamma)^{-1}\left(\sqrt{\gamma_i/\lambda_i}\,H_{\gamma_i-1}H^{(i)}_\gamma, \sqrt{\gamma'_j/\lambda_j}\,H_{\gamma'_j-1}H^{(j)}_{\gamma'}\right)_{\mathrm{L}^2(D,\mu)}(D^\lambda_{\gamma'})^{-1} \\ &= (D^\lambda_\gamma)^{-1}(D^\lambda_{\gamma'})^{-1}\sum_{i,j\ge1} A_{ij}\sqrt{\frac{\gamma_i\gamma'_j}{\lambda_i\lambda_j}}\left(H_{\gamma_i-1}H^{(i)}_\gamma, H_{\gamma'_j-1}H^{(j)}_{\gamma'}\right)_{\mathrm{L}^2(D,\mu)}. \end{aligned} \tag{9.6}$$

To verify s^*-admissibility of the matrix $\mathbf{G}$, we impose additional hypotheses on the matrix $\mathbf{A}$. We consider two cases: $\mathbf{A}$ diagonal and $\mathbf{A}$ tridiagonal. First, consider the case $A_{ij} = \delta_{ij}$. Inserting

this into (9.6), we obtain from the $\mathrm{L}^2(D,\mu)$ orthonormality (3.5) of the collection $\{H_\gamma : \gamma \in \Gamma\}$ that

$$\begin{aligned} G_{\gamma\gamma'} &= (D_\gamma^\lambda)^{-1}(D_{\gamma'}^\lambda)^{-1} \sum_{i,j\geq 1} \delta_{ij} \sqrt{\frac{\gamma_i\gamma_j'}{\lambda_i\lambda_j}} \left(H_{\gamma_i-1}H_\gamma^{(i)}, H_{\gamma_j'-1}H_{\gamma'}^{(j)}\right)_{\mathrm{L}^2(D,\mu)} \\ &= (D_\gamma^\lambda)^{-1}(D_{\gamma'}^\lambda)^{-1} \sum_{i\geq 1} \sqrt{\frac{\gamma_i\gamma_i'}{\lambda_i^2}} \left(H_{\gamma_i-1}H_\gamma^{(i)}, H_{\gamma_i'-1}H_{\gamma'}^{(i)}\right)_{\mathrm{L}^2(D,\mu)} \\ &= \delta_{\gamma,\gamma'}(D_\gamma^\lambda)^{-1}(D_{\gamma'}^\lambda)^{-1} \sum_{i\geq 1} \sqrt{\frac{\gamma_i\gamma_i'}{\lambda_i^2}} \\ &= \delta_{\gamma,\gamma'}\,. \end{aligned}$$

For $A_{ij} = \delta_{ij}$ the bi-infinite matrix $\mathbf{G}$ is diagonal and, therefore, ∞-computable.

We next turn to the tridiagonal matrices $\mathbf{A}$ presented in Example 4.1. Here,

$$A_{ij} = \delta_{ij} + \epsilon_i\delta_{i,j-1} + \epsilon_i\delta_{i,j+1}, \qquad i,j = 1,2,\ldots.$$

The diagonal term having been already discussed, by superposition we may confine ourselves to investigating the computability of $\mathbf{G}$ for $A_{ij} = \epsilon_i\delta_{i,j\pm1}$, $i,j = 1,2,\ldots$. Given $\gamma,\gamma' \in \Gamma$, we calculate, as before, that

$$\begin{aligned} G_{\gamma\gamma'} &= (D_\gamma^\lambda)^{-1}(D_{\gamma'}^\lambda)^{-1} \sum_{i,j\geq 1} \epsilon_i\delta_{i,j\pm1} \sqrt{\frac{\gamma_i\gamma_j'}{\lambda_i\lambda_j}} \left(H_{\gamma_i-1}H_\gamma^{(i)}, H_{\gamma_j'-1}H_{\gamma'}^{(j)}\right)_{\mathrm{L}^2(D,\mu)} \\ &= (D_\gamma^\lambda)^{-1}(D_{\gamma'}^\lambda)^{-1} \sum_{i\geq 1} \epsilon_i \sqrt{\frac{\gamma_i\gamma_{i\mp1}'}{\lambda_i\lambda_{i\mp1}}} \left(H_{\gamma_i-1}H_\gamma^{(i)}, H_{\gamma_{i\mp1}'-1}H_{\gamma'}^{(i\mp1)}\right)_{\mathrm{L}^2(D,\mu)}. \end{aligned} \tag{9.7}$$

We evaluate

$$\left(H_{\gamma_i-1}H_\gamma^{(i)}, H_{\gamma_{i\mp1}'-1}H_{\gamma'}^{(i\mp1)}\right)_{\mathrm{L}^2(D,\mu)} = \begin{cases} 1 & \text{if } \gamma_i' = \gamma_i - 1 \wedge \gamma_{i\mp1}' = \gamma_{i\mp1} + 1 \wedge \gamma_j' = \gamma_j,\ j \in \{i, i\mp1\}^c, \\ 0 & \text{otherwise.} \end{cases}$$

For $\gamma \in \Gamma$ let us define the support of γ by $\mathrm{supp}(\gamma) := \{i \in \mathbb{N} : \gamma_i \neq 0\}$. Then, $G_{\gamma\gamma'} = 0$ if $\mathrm{supp}(\gamma) \cap \mathrm{supp}(\gamma') = \emptyset$. Therefore $\mathbf{G}$ is sparse. If, on the other hand, $\mathrm{supp}(\gamma) \cap \mathrm{supp}(\gamma')$ is nonempty, then, for each $\gamma \in \Gamma$ and each $i \in \mathbb{N}$, the expression

$$\left(H_{\gamma_i-1}H_\gamma^{(i)}, H_{\gamma_{i\mp1}'-1}H_{\gamma'}^{(i\mp1)}\right)_{\mathrm{L}^2(D,\mu)}$$

is equal to 1 for exactly one $\gamma'(\gamma,i) \in \Gamma$. From the above calculations we observe that for this column index $\gamma'(\gamma,i)$, we have that $G_{\gamma,\gamma'(\gamma,i)} = \epsilon_i$; we thus deduce that for the Fokker–Planck equation (2.17) with tridiagonal coefficient matrices $\mathbf{A}$ as in Example 4.1, the bi-infinite matrix $\mathbf{G}$ is still highly sparse: it contains in each row with index $\gamma \in \Gamma$, for each ϵ_i with $i \in \mathrm{supp}(\gamma)$, exactly one off-diagonal entry ϵ_i in the column with index $\gamma'(\gamma,i)$.

We now refer to Definition 7.8 and verify condition (7.3): for $\bar{s} > 0$, we have

$$c_{\mathbf{G},\bar{s}} := \sup_{N\in\mathbb{N}} N\|\mathbf{G} - \mathbf{G}^{[N]}\|_{\ell^2(\Gamma)\to\ell^2(\Gamma)}^{1/\bar{s}} < \infty.$$

Since

$$\|\mathbf{G} - \mathbf{G}^{[N]}\|_{\ell^2(\Gamma)\to\ell^2(\Gamma)}^2 \leq \sup_{\gamma\in\Gamma} \sum_{\gamma'\in\Gamma} |G_{\gamma\gamma'} - G_{\gamma\gamma'}^{[N]}|^2,$$

we *define* $\mathbf{G}^{[N]}$ "row"-wise for $N \in \mathbb{N}$ as follows: if $N = 1$, we select $\mathbf{G}^{[N]}$ to be the diagonal part of $\mathbf{G}$. If $N > 1$, we define $\mathbf{G}^{[N]}$ to contain, in the off-diagonal of the "row" associated with index $\gamma \in \Gamma$, at most N nonzero elements $G_{\gamma\gamma'}$ where $\gamma' = \gamma'(\gamma,i)$ with the index i such that $i \in \{j : \gamma_j \neq 0\} \cap \{j : \epsilon_j^{[N-1]} \neq 0\}$. Here, for a given sequence $\epsilon \in \ell^2(\mathbb{N})$ in the definition (4.3) of the infinite, tridiagonal matrix $\mathbf{A}$ appearing in the bilinear form $\mathfrak{a}(\cdot,\cdot)$, we denote by $\epsilon^{[N]}$ its *best* $(N-1)$*-term approximation* in $\ell^2(\mathbb{N})$. Then, for all $N \in \mathbb{N}$ and for every $0 < p \leq 2$,

$$\|\epsilon - \epsilon^{[N]}\|_{\ell^2(\mathbb{N})} \leq N^{-(1/p-1/2)}\|\epsilon\|_{\ell^p(\mathbb{N})}$$

by Stechkin's Lemma[1]. It therefore follows from the definition (9.7) of the entries $G_{\gamma\gamma'}$ of $\mathbf{G}$ and from the above calculations that

$$\forall N \in \mathbb{N}: \quad \|\mathbf{G} - \mathbf{G}^{[N]}\|_{\ell^2(\Gamma)\to\ell^2(\Gamma)} \le \|\epsilon - \epsilon^{[N-1]}\|_{\ell^2(\mathbb{N})} \le 2^{1/p-1/2} N^{-(1/p-1/2)} \|\epsilon\|_{\ell^p(\mathbb{N})},$$

from which we deduce, with $C_p := 2^{(1/p-1/2)/\bar{s}}$, that

$$c_{\mathbf{G},\bar{s}} = \sup_{N\in\mathbb{N}} N \|\mathbf{G} - \mathbf{G}^{[N]}\|_{\ell^2(\Gamma)}^{1/\bar{s}} \le C_p \left(\sup_{N\in\mathbb{N}} N^{1-(1/p-1/2)/\bar{s}} \right) \|\epsilon\|_{\ell^p(\mathbb{N})}^{1/\bar{s}} = 2\,\|\epsilon\|_{\ell^p(\mathbb{N})}^{1/\bar{s}} < \infty, \tag{9.8}$$

provided that: $\epsilon \in \ell^p(\mathbb{N})$ with $0 < p < 2$ and $\bar{s} = \bar{s}(p)$ is chosen as

$$0 < \bar{s}(p) := 1/p - 1/2. \tag{9.9}$$

Referring to the definition of s^*-computability (cf. Definition 7.8), we infer that $\mathbf{G}$ is s^*-computable with any $0 < s^* \le \bar{s}(p)$ if the sequence ϵ in Example 4.1 belongs to $\ell^p(\mathbb{N})$ with some $0 < p < 2$, resp. with $s^* = 1/p - 1/2$ (this encompasses the previous case, if $p = 0$ is understood to indicate that ϵ is the zero sequence).

10. Equations with drift

10.1. **Bounded invertibility of B and B^* for nonsymmetric $\mathfrak{a}(\cdot,\cdot)$.** So far, we have assumed that the "spatial" differential operator is defined by the symmetric bilinear form $\mathfrak{a}(\cdot,\cdot)$. This is always possible when the vector function appearing in the coefficient of the drift term in the Fokker–Planck equation is the gradient of a spring potential, by introducing Maxwellian weighted spaces $\mathrm{L}^2_M(D)$ and $\mathrm{H}^1_M(D)$ and by rescaling the probability density function by the Maxwellian; see, for example, the formulation in reference [1]. In cases when the drift term cannot be removed by transformation to Maxwellian-weighted operators, however, nonsymmetric bilinear forms $\mathfrak{a}(\cdot,\cdot)$ must be considered. The bounded invertibility of the operator B corresponding to the bilinear form $\mathfrak{B}^*(\cdot,\cdot)$ in (4.6) must then be considered separately.

We now address the s^*-computability of the drift term. Let μ be the countable product of Gaussian measures μ_k, $k \ge 1$, on $\mathbb{R}^d$ with trace-class covariance operator Q. The new contribution added to the symmetric bilinear form $\mathfrak{a}(\cdot,\cdot)$ that was under consideration in previous sections is assumed, for a sequence $\sigma = (\sigma_k)_{k\ge1} \in [\ell^\infty(\mathbb{N})]^{d\times d}$, to be the bilinear form $\mathfrak{d}(\cdot,\cdot)$ defined by

$$\mathfrak{d}(\widehat{\psi},\widehat{\varphi}) = -\sum_{k\ge1} \left(\sigma_k q_k \widehat{\psi}, \mathrm{D}_k \widehat{\varphi} \right)_{\mathrm{L}^2(H,\mu)}. \tag{10.1}$$

Proposition 10.1. *Assume that $\sigma \in [\ell^\infty(\mathbb{N})]^{d\times d}$ and that the covariance operator Q of the Gaussian measure μ on H is trace-class. Then, $\mathfrak{d}(\cdot,\cdot) : \mathrm{W}^{1,2}(H,\mu) \times \mathrm{W}^{1,2}(H,\mu) \to \mathbb{R}$ is continuous and*

$$|\mathfrak{d}(\widehat{\psi},\widehat{\varphi})| \le \|\sigma\|_{[\ell^\infty(\mathbb{N})]^{d\times d}} \left(2\,\mathrm{Tr}\, Q \int_H |\widehat{\psi}(q)|^2 \mu(\mathrm{d}q) + 4\|Q\|^2 \int_H |\mathrm{D}\widehat{\psi}(q)|^2 \mu(\mathrm{d}q) \right)^{1/2} \|\mathrm{D}\widehat{\varphi}\|_{\mathrm{L}^2(H,\mu)}. \tag{10.2}$$

Proof. We write

$$\begin{aligned} |\mathfrak{d}(\widehat{\psi},\widehat{\varphi})| \quad &\le \| \quad \sigma\|_{[\ell^\infty(\mathbb{N})]^{d\times d}} \left(\sum_{k\ge1} \left\| |q_k|\, \widehat{\psi} \right\|^2_{\mathrm{L}^2(H,\mu)} \right)^{1/2} \left(\sum_{k'\ge1} \|\mathrm{D}_{k'}\widehat{\varphi}\|^2_{\mathrm{L}^2(H,\mu)} \right)^{1/2} \\ &= \quad \|\sigma\|_{[\ell^\infty(\mathbb{N})]^{d\times d}} \left(\sum_{k\ge1} \left\| |q_k|\, \widehat{\psi} \right\|^2_{\mathrm{L}^2(H,\mu)} \right)^{1/2} \|\mathrm{D}\widehat{\varphi}\|_{\mathrm{L}^2(H,\mu)}. \end{aligned}$$

[1]**Stechkin's Lemma.** Let $0 < p \le q \le \infty$ and assume that $\alpha = (\alpha_\gamma)_{\gamma\in\Gamma} \in \ell^p(\Gamma)$. For $N \ge 1$, let $\Gamma_N \subset \Gamma$ denote the set of indices corresponding to the N largest values of $|\alpha_\gamma|$. Then,

$$\Big(\sum_{\gamma\notin\Gamma_N} |\alpha_\gamma|^q \Big)^{1/q} \le N^{-r} \|\alpha\|_{\ell^p(\Gamma)}, \qquad \text{where } r := \tfrac{1}{p} - \tfrac{1}{q} \ge 0.$$

The proof is elementary; see, for example, [10].

Using [9, Prop. 9.2.10] and the assumption that $\widehat{\psi} \in \mathrm{W}^{1,2}(H,\mu)$, we deduce that, for a Gaussian measure μ with trace-class covariance Q,

$$\int_H \|q\|^2 |\widehat{\psi}(q)|^2 \mu(\mathrm{d}q) \le 2\,\mathrm{Tr}\,Q \int_H |\widehat{\psi}(q)|^2 \mu(\mathrm{d}q) + 4\|Q\|^2 \int_H |\mathrm{D}\widehat{\psi}(q)|^2 \mu(\mathrm{d}q).$$

This then yields the desired inequality (10.2). □

The bound (10.2) implies a Garding inequality for the Fokker–Planck operator with drift.

Proposition 10.2. *Assume that the covariance operator Q of the Gaussian measure μ is trace-class, and that the infinite coefficient matrix A satisfies* (4.1), (4.2). *Then, the bilinear form*

$$\mathfrak{a}(\cdot,\cdot) + \mathfrak{d}(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$$

is continuous and satisfies the Garding inequality (2.22) *in the triple $\mathcal{V} \subset \mathcal{H} \simeq \mathcal{H}^* \subset \mathcal{V}^*$. In particular, the space-time variational formulation* (4.5) *with the spatial bilinear form $\mathfrak{a}(\cdot,\cdot)+\mathfrak{d}(\cdot,\cdot)$ in place of $\mathfrak{a}(\cdot,\cdot)$ in* (4.6) *is well-posed and its bilinear form $\mathfrak{B}^*(\cdot,\cdot)$ induces a boundedly invertible operator $B^* \in \mathcal{L}(\mathcal{Y}, (\mathcal{X}_{0,\{T\}})^*)$.*

Proof. The continuity of the form is evident from the previous proposition and the continuity of $\mathfrak{a}(\cdot,\cdot)$. The Garding inequality follows from the coercivity of $\mathfrak{a}(\cdot,\cdot)$ on $\mathcal{V}\times\mathcal{V}$ and from the inequality (10.2) using a Cauchy inequality with $\varepsilon > 0$ sufficiently small. The bounded invertibility of $B^* \in \mathcal{L}(\mathcal{Y}, (\mathcal{X}_{0,\{T\}})^*)$ follows from (2.22) and from Theorem 2.2. □

10.2. **s^*-computability of $\int_I \mathfrak{d}(\Theta\otimes[\Upsilon]_{\mathcal{V}}, \Theta\otimes[\Upsilon]_{\mathcal{V}})\,\mathrm{d}t$ and of its adjoint.** As in Section 9.4, thanks to the independence of the sequence $\sigma \in [\ell^\infty(\mathbb{N})]^{d\times d}$ of t, we have that

$$\int_I \mathfrak{d}(\Theta\otimes[\Upsilon]_{\mathcal{V}}, \Theta\otimes[\Upsilon]_{\mathcal{V}})\,\mathrm{d}t = (\Theta,\Theta)_{\mathrm{L}^2(I)} \otimes \mathfrak{d}([\Upsilon]_{\mathcal{V}}, [\Upsilon]_{\mathcal{V}}).$$

As in the discussion of (9.5), the sparsity of the factor $(\Theta,\Theta)_{\mathrm{L}^2(I)}$ is considered in [19, Sec. 8.5]. Sparsity and s^*-computability of $\mathfrak{d}(\cdot,\cdot)$ are therefore determined by that of the infinite *drift matrix* $\mathbf{D} = \mathbf{D}[\sigma]$ defined via $\mathfrak{d}([\Upsilon]_{\mathcal{V}}, [\Upsilon]_{\mathcal{V}}) = (D_{\gamma,\gamma'})_{\gamma,\gamma'\in\Gamma}$. With the Hermite polynomial Riesz basis $[\Upsilon]_{\mathcal{V}}$ of $\mathrm{L}^2(H,\mu)$, the matrix entries $D_{\gamma\gamma'} = D_{\gamma\gamma'}[\sigma]$ are (assuming, once again, for the sake of simplicity of the exposition that $d = 1$; abbreviating D_{q_k} as D_k to simplify the notation; and recalling the definition (9.1) of D_γ^λ):

$$D_{\gamma\gamma'}[\sigma] = \sum_{k\ge1} \sigma_k (D_\gamma^\lambda)^{-1} (D_{\gamma'}^\lambda)^{-1} \left(\mathrm{D}_k(q_k H_\gamma), H_{\gamma'}\right)_{\mathrm{L}^2(H,\mu)}. \tag{10.3}$$

For the ensuing calculations, we define

$$\theta_{\gamma\gamma'}(k) := \left(\mathrm{D}_k(q_k H_\gamma), H_{\gamma'}\right)_{\mathrm{L}^2(H,\mu)}$$

and observe that

$$\theta_{\gamma\gamma'}(k) = (H_\gamma, H_{\gamma'})_{\mathrm{L}^2(H,\mu)} + (q_k \mathrm{D}_k H_\gamma, H_{\gamma'})_{\mathrm{L}^2(H,\mu)} = \delta_{\gamma\gamma'} + (q_k \mathrm{D}_k H_\gamma, H_{\gamma'})_{\mathrm{L}^2(H,\mu)}. \tag{10.4}$$

In order to calculate the second term on the right-hand side of (10.4), we note that, by (9.3),

$$\mathrm{D}_k H_\gamma(q) = \begin{cases} \sqrt{\frac{\gamma_k}{\lambda_k}}\, H_{\gamma-e_k}(q) & \text{if} \quad \gamma_k \ge 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $e_k \in \Gamma$ denotes the multi-index with entry 1 in position k and with zero entries in all other positions. Also, thanks to the three-term recurrence relation

$$q_k H_{\gamma_k}(q_k) = \sqrt{\gamma_k+1}\, H_{\gamma_k+1}(q_k) + \sqrt{\gamma_k}\, H_{\gamma_k-1}(q_k), \quad \gamma_k \in \mathbb{N}_0,$$

for the univariate polynomials (3.4) (see, (9.1.3) in [9]), with the notational convention that $H_m(q_k) \equiv 0$ when $m < 0$, we have that

$$q_k \mathrm{D}_k \left[H_{\gamma_k}\left(\frac{q_k}{\sqrt{\lambda_k}}\right)\right] = q_k \sqrt{\frac{\gamma_k}{\lambda_k}} H_{\gamma_k-1}\left(\frac{q_k}{\sqrt{\lambda_k}}\right) = \frac{\gamma_k}{\sqrt{\lambda_k}} H_{\gamma_k}\left(\frac{q_k}{\sqrt{\lambda_k}}\right) + \sqrt{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}} H_{\gamma_k-2}\left(\frac{q_k}{\sqrt{\lambda_k}}\right).$$

Consequently, upon multiplying both sides of the last identity with $H_\gamma^{(k)}$, we deduce that

$$q_k \mathrm{D}_k H_\gamma = \frac{\gamma_k}{\sqrt{\lambda_k}}\, H_\gamma + \sqrt{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}}\, H_{\gamma-2e_k},$$

which, upon substitution into (10.4) and then inserting the resulting expression into (10.3) yields, for $\gamma, \gamma' \in \Gamma$, that

$$D_{\gamma\gamma'}[\sigma] \;=\; \sum_{k\geq 1} \sigma_k \left(\left(1+\frac{\gamma_k}{\sqrt{\lambda_k}}\right)\delta_{\gamma\gamma'} + \sqrt{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}}\,\delta_{\gamma-2e_k,\gamma'} \right) \langle \gamma, \lambda^{-1}\rangle_*^{-1/2} \langle \gamma', \lambda^{-1}\rangle_*^{-1/2},$$

with the notational convention that $\delta_{\gamma-2e_k,\gamma'} = 0$ for all $\gamma, \gamma' \in \Gamma$ such that $\gamma - 2e_k \notin \Gamma$. We observe that $\mathbf{D}[\sigma]$ depends linearly on the sequence σ. The verification of the s^*-computability now proceeds analogously as in the case of $\mathbf{A}[\epsilon]$.

We denote by $\sigma^{[N]}$ an N-term approximation of the sequence σ and observe that $\mathbf{D}^{[N]}[\sigma] := \mathbf{D}[\sigma^{[N]}]$ has at most $2N+1$ nonzero entries in each row with index $\gamma \in \Gamma$. To verify s^*-computability, by Definition 7.8 we must bound

$$c_{\mathbf{D},\bar{s}} = \sup_{N\in\mathbb{N}} N \left\| \mathbf{D}[\sigma] - \mathbf{D}^{[N]}[\sigma] \right\|^{1/\bar{s}}_{\ell^2(\Gamma)\to\ell^2(\Gamma)} \leq \sup_{N\in\mathbb{N}} (2N+1) \left\| \mathbf{D}[\sigma-\sigma^{[N]}] \right\|^{1/\bar{s}}_{\ell^2(\Gamma)\to\ell^2(\Gamma)}. \tag{10.5}$$

For a given fixed $\gamma \in \Gamma$ and for any $\gamma' \in \Gamma$ and any $k \in \mathbb{N}$, we have that

$$\theta_{\gamma\gamma'}(k) = \left(1+\frac{\gamma_k}{\sqrt{\lambda_k}}\right)\delta_{\gamma\gamma'} + \sqrt{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}}\,\delta_{\gamma-2e_k,\gamma'} = \begin{cases} 1+\frac{\gamma_k}{\sqrt{\lambda_k}} & \text{if } \gamma' = \gamma, \\ \sqrt{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}} & \text{if } \gamma' = \gamma - 2e_k \in \Gamma, \\ 0 & \text{otherwise.} \end{cases}$$

We note that since $\gamma' \in \Gamma$, and therefore all of its components are nonnegative, the second of these instances can only occur when $\gamma_k \geq 2$. Therefore, we can bound

$$\begin{aligned} \|\mathbf{D}[\sigma] - \mathbf{D}^{[N]}[\sigma]\|^2_{\ell^2(\Gamma)\to\ell^2(\Gamma)} &\leq \sup_{\gamma\in\Gamma} \sum_{\gamma'\in\Gamma} \left| D_{\gamma\gamma'}[\sigma-\sigma^{[N]}] \right|^2 \\ &= \sup_{\gamma\in\Gamma} \frac{1}{\langle\gamma,\lambda^{-1}\rangle_*} \sum_{\gamma'\in\Gamma} \frac{1}{\langle\gamma',\lambda^{-1}\rangle_*} \left| \sum_{k\geq 1} (\sigma_k - \sigma_k^{[N]})\, \theta_{\gamma\gamma'}(k) \right|^2. \end{aligned}$$

Fixing $\gamma \in \Gamma$ for now, we deduce that

$$\begin{aligned} \mathrm{S}_\gamma &:= \sum_{\gamma'\in\Gamma} \frac{1}{\langle\gamma',\lambda^{-1}\rangle_*} \left| \sum_{k\geq 1} (\sigma_k - \sigma_k^{[N]})\, \theta_{\gamma\gamma'}(k) \right|^2 \\ &\leq \|\sigma-\sigma^{[N]}\|^2_{\ell^1(\mathbb{N})} \left\{ \sup_{k\geq 1} \frac{\left(1+\frac{\gamma_k}{\sqrt{\lambda_k}}\right)^2}{\langle\gamma,\lambda^{-1}\rangle_*} + \sup_{k\geq 1\,:\,\gamma_k\geq 2} \frac{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}}{\langle\gamma-2e_k,\lambda^{-1}\rangle_*} \right\}. \end{aligned}$$

This implies that

$$\begin{aligned} \|\mathbf{D}-\mathbf{D}^{[N]}\|^2_{\ell^2(\Gamma)\to\ell^2(\Gamma)} &\leq \sup_{\gamma\in\Gamma} \tfrac{\mathrm{S}_\gamma}{\langle\gamma,\lambda^{-1}\rangle_*} \\ &\leq \sup_{\gamma\in\Gamma} \left\{ \sup_{k\geq 1} \tfrac{\left(1+\frac{\gamma_k}{\sqrt{\lambda_k}}\right)^2}{\langle\gamma,\lambda^{-1}\rangle_*^2} + \sup_{k\geq 1\,:\,\gamma_k\geq 2} \tfrac{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}}{\langle\gamma,\lambda^{-1}\rangle_*\langle\gamma-2e_k,\lambda^{-1}\rangle_*} \right\} \|\sigma-\sigma^{[N]}\|^2_{\ell^1(\mathbb{N})}. \end{aligned}$$

We bound the suprema in the curly brackets as follows. For $\gamma = 0$, the first supremum is equal to 1, while the second supremum is 0 by definition, as it is taken over an empty set. Let us now consider the nontrivial case, when $\gamma \in \Gamma \setminus \{0\}$. Since $|\gamma| < \infty$, the "support-set" of this index sequence γ, $S(\gamma) := \{j \in \mathbb{N} : \gamma_j \neq 0\}$, is finite. To bound the first term, we note that the finiteness

of $S(\gamma)$ implies that there exists $k^* \in N$ such that $\sup_k \gamma_k/\sqrt{\lambda_k} = \max_k \gamma_k/\sqrt{\lambda_k} = \gamma_{k^*}/\sqrt{\lambda_{k^*}} =: |\gamma/\sqrt{\lambda}|_\infty$, and we may bound

$$\begin{aligned}\sup_k \frac{\left(1+\frac{\gamma_k}{\sqrt{\lambda_k}}\right)^2}{\langle \gamma, \lambda^{-1}\rangle_*^2} &\leq 2\left(1+\left|\frac{\gamma}{\sqrt{\lambda}}\right|_\infty^2\right)\left(\sum_{j\geq 1}\frac{\gamma_j}{\lambda_j}\right)^{-2} = 2\left(1+\left|\frac{\gamma}{\sqrt{\lambda}}\right|_\infty^2\right)\left(\sum_{j\in S(\gamma)}\frac{\gamma_j}{\lambda_j}\right)^{-2}\\ &\leq 2\left(1+\left|\frac{\gamma}{\sqrt{\lambda}}\right|_\infty^2\right)\left(\frac{1}{\sqrt{\lambda_{k^*}}}\,\frac{\gamma_{k^*}}{\sqrt{\lambda_{k^*}}}\right)^{-2} = 2\,\lambda_{k^*}\,\frac{1+|\gamma/\sqrt{\lambda}|_\infty^2}{|\gamma/\sqrt{\lambda}|_\infty^2}\\ &= 2\,\lambda_{k^*}\left(\tfrac{1}{|\gamma/\sqrt{\lambda}|_\infty^2}+1\right) \leq 2\,\lambda_{k^*}\left(\tfrac{1}{|1/\sqrt{\lambda}|_\infty^2}+1\right) \leq 2\lambda_1(\lambda_1+1)\\ &\leq 4\,[\max(1,\lambda_1)]^2 = [2\,\max(1,\|Q\|)]^2 < \infty.\end{aligned}$$

The second term in the curly brackets is bounded analogously. Indeed, assuming that $\gamma - 2e_k \in \Gamma$ (for, else, the supremum in the second term is taken over an empty set and is therefore equal to 0 by definition), we have that

$$\begin{aligned}&\sup_{k\geq 1\,:\,\gamma_k\geq 2}\frac{\frac{\gamma_k(\gamma_k-1)}{\lambda_k}}{\langle\gamma,\lambda^{-1}\rangle_*\langle\gamma-2e_k,\lambda^{-1}\rangle_*} \leq \frac{|\gamma/\lambda|_\infty}{\langle\gamma,\lambda^{-1}\rangle_*}\sup_{k\geq 1\,:\,\gamma_k\geq 2}\frac{(\gamma_k-2)+1}{\langle\gamma-2e_k,\lambda^{-1}\rangle_*}\\ &\quad\leq \sup_{k\geq 1\,:\,\gamma_k\geq 2}\frac{(\gamma_k-2)+1}{\langle\gamma-2e_k,\lambda^{-1}\rangle_*} \leq \sup_{k\geq 1\,:\,\gamma_k\geq 2}\frac{\lambda_k\,\frac{\gamma_k-2}{\lambda_k}}{\langle\gamma-2e_k,\lambda^{-1}\rangle_*} + \sup_{k\geq 1\,:\,\gamma_k\geq 2}\frac{1}{\langle\gamma-2e_k,\lambda^{-1}\rangle_*}\\ &\quad\leq \lambda_1 + \max(1,\lambda_1) \leq 2\max(1,\|Q\|) < \infty.\end{aligned}$$

Inserting these bounds into (10.5), we infer that for $\sigma \in \ell^p(\mathbb{N})$ with some $0 < p < 1$ the bi-infinite drift matrix $\mathbf{D}[\sigma]$ is s^*-computable for any $0 < s^* \leq 1/p - 1$.

11. Optimality

Based on Proposition 9.1 and on the definition (8.7) of the bi-infinite matrix $\mathbf{B}^*$, we deduce from these observations our main result.

Theorem 11.1. *Consider the space-time variational formulation* (4.5) *of the infinite-dimensional Fokker–Planck equation with bilinear form* $\mathfrak{a}(\cdot,\cdot)$ *defined in* (4.7) *corresponding to the tridiagonal matrix* $\mathbf{A}$ *as in Example 4.1 based on the sequences* $\epsilon, \sigma \in \ell^p(\mathbb{N})$ *of elements* ϵ_i, σ_i *with* $0 < p < 1$.

Consider its representation $\mathbf{B}^*\mathbf{u} = \mathbf{f}^*$ *using a temporal wavelet basis* Θ *as in Sect. 9.1 and a spatial Hermite polynomial chaos basis* $\Upsilon = \{H_\gamma(q) : \gamma \in \Gamma, \quad q \in D\}$. *Then, for any* $\varepsilon > 0$, *the adaptive wavelet methods from* [7] *or* [6] *(and* [12]*) applied to the normal equation* (7.4) *with* $\mathbf{f}^*$ *and* $\mathbf{B}^*$ *as in (8.6) and (8.7), respectively, of the bi-infinite matrix representation of the Fokker–Planck equation* (2.17) *in countably many dimensions produce a Galerkin approximation* $\mathbf{u}_\varepsilon$ *with*

$$\|u - \mathbf{u}_\varepsilon^\top[\Theta\otimes\Upsilon]\|_{\mathcal{Y}} \eqsim \|\mathbf{u}-\mathbf{u}_\varepsilon\| \leq \varepsilon.$$

Suppose that for some $0 < s < \min\{d_t - 1, \bar{s}(p)\}$ *we have that* $\mathbf{u} \in \mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))$; *then,*

$$\operatorname{supp}\mathbf{u}_\varepsilon \lesssim \varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))}.$$

The number of arithmetic operations and storage locations required by one call of the space-time adaptive solver with tolerance ε *is bounded by some multiple of* $\varepsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_\infty(\ell^2(\nabla_{\mathcal{Y}}))} + 1$.

The above assertions remain valid when in (2.17) *the dimension is a finite number* Kd *(where* K *denotes the chain-length in mesoscopic models of dilute polymers in spatial dimension* $d = 2, 3$ *(see,* [2] *for further details), with the computability and admissibility constants* $c_{\mathbf{B}^*,s}$ *and* $a_{\mathbf{B}^*,s}$ *independent of the dimension* K.

We remark in closing that for matrices $\mathbf{A}$ as in Example 4.2 this result remains valid, however now with admissibility constants depending on the block $\mathbf{A}_{11}$ in an unspecific way. We also remark that the assumptions on the sequence σ could be slightly weakened, as the limit $1/p - 1$ of s^*-computability of $\mathbf{D}[\sigma]$ is larger than the value $\bar{s}(p)$ found in (9.9).

References

1. J.W. Barrett and E. Süli, *Existence and equilibration of global weak solutions to kinetic models for dilute polymers I: finitely extensible nonlinear bead-spring chains*, Math. Models Methods Appl. Sci. **21** (2011), no. 6, 1211–1289. MR MR2819196 (2012f:82067)
2. J.W. Barrett and E. Süli, *Existence and equilibration of global weak solutions to kinetic models for dilute polymers II: Hookean-type bead-spring chains*, Math. Models Methods Appl. Sci. **22** (2012), no. 5, 1–82, Extended version avialable from `http://arxiv.org/abs/1008.3052`.
3. V. Bogachev, G. Da Prato, and M. Röckner, *Existence and uniqueness of solutions for Fokker–Planck equations on Hilbert spaces*, J. Evol. Equ. **10** (2010), no. 3, 487–509. MR 2674056 (2011g:60107)
4. S. Cerrai, *Second Order PDE's in Finite and Infinite Dimension*, Lecture Notes in Mathematics, vol. 1762, Springer-Verlag, Berlin, 2001, A probabilistic approach. MR 1840644 (2002j:35327)
5. N. Chegini and R. Stevenson, *Adaptive wavelet schemes for parabolic problems: sparse matrices and numerical results*, SIAM J. Numer. Anal. **49** (2011), no. 1, 182–212. MR 2783222
6. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations – Convergence rates*, Math. Comp **70** (2001), 27–75.
7. ———, *Adaptive wavelet methods II - Beyond the elliptic case*, Found. Comput. Math. **2** (2002), no. 3, 203–245.
8. G. Da Prato, P. Malliavin, and D. Nualart, *Compact families of Wiener functionals*, C. R. Acad. Sci. Paris Sér. I Math. **315** (1992), no. 12, 1287–1291. MR 1194537 (94a:60085)
9. G. Da Prato and J. Zabczyk, *Second Order Partial Differential Equations in Hilbert Spaces*, London Mathematical Society Lecture Note Series, vol. 293, Cambridge University Press, Cambridge, 2002. MR 1985790 (2004e:47058)
10. R. DeVore, *Nonlinear approximation*, Acta Numer. **7** (1998), 51–150.
11. T.J. Dijkema, Ch. Schwab, and R. Stevenson, *An adaptive wavelet method for solving high-dimensional elliptic PDEs*, Constr. Approx. **30** (2009), no. 3, 423–455. MR 2558688 (2010m:65270)
12. T. Gantumur, H. Harbrecht, and R. Stevenson, *An optimal adaptive wavelet method without coarsening of the iterands*, Tech. Report 1325, Utrecht University, March 2005, To appear in *Math. Comp.*
13. M. Griebel and P. Oswald, *Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems*, Adv. Comput. Math. **4** (1995), no. 1–2, 171–206.
14. R. Guberović, Ch. Schwab, and R. Stevenson, *Space-time variational saddle point formulations of Stokes and Navier–Stokes equations*, Tech. report, Seminar for Applied Mathematics, ETH Zürich, http://www.sam.math.ethz.ch/reports/2011 (Submitted), 2011.
15. P.-A. Nitsche, *Best N term approximation spaces for tensor product wavelet bases*, Constr. Approx. **24** (2006), no. 1, 49–70. MR 2217525 (2007a:41032)
16. S. Peszat, *On a Sobolev space of functions of infinite number of variables*, Bull. Polish Acad. Sci. Math. **41** (1993), no. 1, 55–60. MR 1401886 (97m:46059)
17. A. Pietsch, *Nuclear Locally Convex Spaces*, Springer-Verlag, New York, 1972, Translated from the second German edition by William H. Ruckle, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 66. MR 0350360 (50 #2853)
18. I. Schur, *Bemerkungen zur Theorie der Beschränkten Bilinearformen mit unendlich vielen Veränderlichen*, J. reine angew. Math. **140** (1911), 1–28.
19. Ch. Schwab and R. Stevenson, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp. **78** (2009), no. 267, 1293–1318. MR 2501051 (2011a:65347)

Research Reports

No.	Authors/Title
11-69	*Ch. Schwab and E. Süli* Adaptive Galerkin approximation algorithms for partial differential equations in infinite dimensions
11-68	*A. Barth and A. Lang* Multilevel Monte Carlo method with applications to stochastic partial differential equations
11-67	*C. Effenberger and D. Kressner* Chebyshev interpolation for nonlinear eigenvalue problems
11-66	*R. Guberovic, Ch. Schwab and R. Stevenson* Space-time variational saddle point formulations of Stokes and Navier-Stokes equations
11-65	*J. Li, H. Liu and H. Sun* Enhanced approximate cloaking by SH and FSH lining
11-64	*M. Hansen and Ch. Schwab* Analytic regularity and best N-term approximation of high dimensional parametric initial value problems
11-63	*R. Hiptmair, G. Phillips and G. Sinha* Multiple point evaluation on combined tensor product supports
11-62	*J. Li, M. Li and S. Mao* Convergence analysis of an adaptive finite element method for distributed flux reconstruction
11-61	*J. Li, M. Li and S. Mao* A priori error estimates of a finite element method for distributed flux reconstruction
11-60	*H. Heumann and R. Hiptmair* Refined convergence theory for semi-Lagrangian schemes for pure advection
11-59	*V.A. Hoang and Ch. Schwab* N-term Galerkin Wiener chaos approximations of elliptic PDEs with lognormal Gaussian random inputs
11-58	*X. Claeys and R. Hiptmair* Electromagnetic scattering at composite objects: A novel multi-trace boundary integral formulation
11-57	*S. Mishra and L.V. Spinolo* Accurate numerical schemes for approximating intitial-boundary value problems for systems of conservation law