Deterministic Simulation of Multi-Beaded Models of Dilute Polymer Solutions

Leonardo Figueroa
Worcester College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in Numerical Analysis

Trinity Term 2011

Abstract

We study the convergence of a nonlinear approximation method introduced in the engineering literature for the numerical solution of a high-dimensional Fokker–Planck equation featuring in Navier–Stokes–Fokker–Planck systems that arise in kinetic models of dilute polymers. To do so, we build on the analysis carried out recently by Le Bris, Lelièvre and Maday (Const. Approx. 30: 621–651, 2009) in the case of Poisson's equation on a rectangular domain in $\mathbb{R}^2$, subject to a homogeneous Dirichlet boundary condition, where they exploited the connection of the approximation method with the greedy algorithms from nonlinear approximation theory explored, for example, by DeVore and Temlyakov (Adv. Comput. Math. 5:173–187, 1996). We extend the convergence analysis of the pure greedy and orthogonal greedy algorithms considered by Le Bris, Lelièvre and Maday to the technically more complicated situation of the elliptic Fokker–Planck equation, where the role of the Laplace operator is played out by a high-dimensional Ornstein–Uhlenbeck operator with unbounded drift, of the kind that appears in Fokker–Planck equations that arise in bead-spring chain type kinetic polymer models with finitely extensible nonlinear elastic potentials, posed on a high-dimensional Cartesian product configuration space $\mathsf{D} = D_1 \times \cdots \times D_N$ contained in $\mathbb{R}^{Nd}$, where each set D_i, $i = 1, \ldots, N$, is a bounded open ball in $\mathbb{R}^d$, $d = 2, 3$. We exploit detailed information on the spectral properties and elliptic regularity of the Ornstein–Uhlenbeck operator to give conditions on the true solution of the Fokker–Planck equation which guarantee certain rates of convergence of the greedy algorithms. We extend the analysis to discretized versions of the greedy algorithms.

Acknowledgments

General. First and foremost, I thank God for being with me and Claudia during all this time, including this thesis preparation time. Then, my lovely Claudia, of course. It makes me so happy to be your husband.

I also want to thank most unreservedly my supervisor, Professor Endre Süli, who besides being a fantastic mathematician is really a pleasure to work and talk with. Being his student has truly been a privilege.

I would also like to thank my colleagues, particularly Bernhard Langwallner, Jaroslav Fowkes and Hao Wang, with whom I talked the most, for all those little conversations which I really enjoyed.

Thanks as well to my family back home. You are amazing!

Idea contributions. I am grateful to Professor Marco Marletta (Cardiff University) for helpful suggestions regarding the Liouville transformation.

Funding. This research was supported by a doctoral scholarship from the Chilean government's *Comisión Nacional de Investigación Científica y Tecnológica*.

Table of Contents

List of Figures

List of Algorithms and Hypotheses

List of Symbols

In this work and in this list in particular we use the symbol $\cdot$, with or without subscripts, as a placeholder. Standard objects (e.g., $\mathrm{C}^1(\cdot)$, $\mathbb{R}$, ∇) are omitted, as are objects whose scope is very limited, such as those appearing in only one proof.

Symbol	Description	First mention
a	Configurational Fokker–Planck bilinear form	(2.2)
a_r, $a_{(r,s)}$	Boundary-describing mapping	Definition 2.16
A	Rouse (spring interaction) matrix	(1.6)
A_r, $A_{(r,s)}$	Boundary-describing transformation	Definition 2.16
$\mathcal{A}_1$, $\mathcal{A}_{1,\boldsymbol{l}}$	Spaces of fast convergence for the greedy algorithms	(3.30), (4.17)
$\mathcal{A}$	Functional-valued function based on a	Subsection 3.1.1
b, b_i	Squared ratio of the maximal extension of a spring and its characteristic length	(1.13), Subsection 1.1.3
B	Ball / Matrix with entries $B_{ij} = \delta_{ij}\sqrt{b_i}$	(1.11)/ Subsection 4.1.6
$\boldsymbol{B}_\nu$	Brownian force acting on the ν-th bead	(1.4)
$\mathcal{B}_1$	$\mathsf{M}\,\mathcal{A}_1$	(3.31)
c	Coefficient of zero-order term in bilinear form a	(2.2)
d	Dimension of the physical space	Subsection 1.1.2
$\mathfrak{d}$	Distance-to-the-boundary function	Lemma 2.10, Hypothesis E
D_i	Single spring configuration domain: $B(0, \sqrt{b_i})$	Subsection 1.1.3
D	Full configuration domain	Subsection 1.1.3
$\mathfrak{D}$	Dictionary	Definition 3.11, (3.25)
$\mathfrak{D}_{\boldsymbol{l}}$	Truncated dictionary	(4.15)
$e_n^{(i)}$, $e_{\boldsymbol{n}}$	Eigenfunctions in $\mathrm{H}^1_{M_i}(D_i)$, $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$	(3.33), (3.37)
$\tilde{e}_{\boldsymbol{n}}$	Eigenfunctions in $\mathrm{H}(\mathsf{D};\mathsf{M})$	Subsection 4.1.5
f	Right-hand side functional of Fokker–Planck equation	(2.1)
f_n	Residual after the n-th iteration of Algorithm I and Algorithm II	Algorithm I, Algorithm II
$\hat{f}_n$	Residual after the n-th iteration of Algorithm V and Algorithm VI	Algorithm V, Algorithm VI
$\boldsymbol{f}$	External body force acting on the fluid	(1.1)
$\hat{F}$	${}^{L^{-1}(\cdot)}\!/\!_{\sqrt{\cdot}}$	Subsection 2.2.2

$\boldsymbol{F}$, $\boldsymbol{F}_i$	Spring forces	Subsection 1.1.2, Subsection 1.1.3
G	Matrix connecting the bead position and spring connector vectors / Greedy term	Subsection 1.1.2/ Definition 3.11
h_i	Ratio of a Maxwellian and its associated power function	Hypothesis E
H	Spring constant	Subsection 1.1.2
$\mathrm{H}^{\cdot_1}_{\cdot_2}(\cdot_3)$	$\cdot_2$-weighted Sobolev space of order $\cdot_1$ on the domain $\cdot_3$; $\cdot_1$ might be a net	Section 1.4, (3.39)
$\mathrm{H}^{\cdot_1,\mathrm{mix}}_{\cdot_2}(\cdot_3)$	Higher-mixed-derivatives $\cdot_2$-weighted Sobolev space of order $\cdot_1$ on the domain $\cdot_3$	Section 1.4
$\tilde{\mathrm{H}}^{\cdot,\mathrm{mix}}_{\mathsf{M}}$	Second order operator-based Sobolev-like spaces	(3.78), (3.79),
$\hat{\mathrm{H}}_{\boldsymbol{l}}$	Tensor-product finite-dimensional subspace of $\mathrm{H}(\mathsf{D};\mathsf{M})$	(4.2)
$\hat{\mathrm{H}}^{(i)}_{l}$	Finite-dimensional subspace of $\mathrm{H}(D_i;M_i)$	Subsection 4.1.1
$\mathrm{H}(\mathsf{D};\mathsf{M})$	Weighted Sobolev-like space	Subsection 2.1.1
$\mathrm{H}^{\cdot}(\mathsf{D};\mathsf{M})$	$\mathsf{M}\,\mathrm{H}^{\cdot}_{\mathsf{M}}(\mathsf{D})$	(3.39)
$\mathrm{H}(D_i;M_i)$	Single-spring weighted Sobolev-like space	Subsection 2.1.1
$\mathfrak{H}$	Generic Hilbert space	Definition 3.11, (3.24)
I_i	Shorthand for the inner product of $\mathrm{L}^2_{1/M_i}(D_i)$	(4.67)
$\boldsymbol{I}$	Identity tensor	(1.2)
$\mathcal{I}^{(i)}$	Mass matrix	(4.74)
$J^{(\cdot_1)}_{\cdot_2}$	Jacobi polynomial of parameter $\cdot_1$ and index $\cdot_2$	Subsection 4.1.6, (4.37)
k_B	Boltzmann constant	Subsection 1.1.2
K_i	$\mathrm{L}^2_{1/M_i}(D_i)$ semi-inner-product	(4.67)
$\mathcal{K}$, $\tilde{\mathcal{K}}$	Diffusion bilinear forms	(1.28)
$\mathcal{K}^{(i)}$	Stiffness matrix	(4.74)
$J_\cdot$	Energy functional $\varphi \mapsto \frac{1}{2}a(\varphi,\varphi) - \cdot(\varphi)$	(3.1)
$\mathcal{J}_\cdot$	Energy functional with respect to factor functions and source term $\cdot$	Subsection 4.2.1
l_0	Characteristic length-scale of a spring	Subsection 1.1.2
L	Langevin function	(1.11)
L_0	Characteristic macroscopic length	Subsection 1.1.2
$L_\cdot$, $\hat{L}_\cdot$	Weighted second order differential operators	(3.76), (3.77)
$\mathrm{L}^2_\cdot$	$\cdot$-weighted Lebesgue space of order of integrability 2	Section 1.4
M_i	Partial Maxwellians	(1.24)
M	(Full) Maxwellian	(1.25)
n_p	Molecule concentration	(1.8)
$\mathfrak{n}_{\boldsymbol{x}}$, $\mathfrak{n}_{\boldsymbol{q}_i}$	Unit outward normal vectors defined on the boundary of Ω, D_i	Subsection 1.1.3
N	Number of springs	Subsection 1.1.2
N_g	Number of tensor-product terms in source term	(4.64)
p	Fluid pressure	(1.2)

$P_{\boldsymbol{l}}$, $P_{\boldsymbol{l}}^{(\boldsymbol{\alpha})}$	Projection operators	(4.23), (4.41)
$\mathbb{P}_l$, $\mathbb{P}_{\boldsymbol{l}}$	Spaces of polynomials	Subsection 4.1.6
$q_{\max}$	Maximal extension of the springs	(1.10)
$\boldsymbol{q}$, $\boldsymbol{q}_i$	Ensemble of configuration vectors, i-th configuration vector	Subsection 1.1.2
$\boldsymbol{q}^{(k)}$	Point for 1/M-weighted quadrature over D	Subsection 1.1.4
Q_{D}, Q_Ω	Number of quadrature points over D, Ω	Subsection 1.1.4
$r_n^{(i)}$	Factor function computed by Algorithm I or Algorithm II	(3.3), (3.4)
$\hat{r}_n^{(i)}$	Factor function computed by Algorithm V or Algorithm VI	(4.4), (4.5)
$\boldsymbol{r}$, $\boldsymbol{r}_\nu$, $\boldsymbol{r}_c$	Ensemble of bead positions, position of the ν-th bead, center of mass of the beads	Subsection 1.1.2
R	Greedy residual	Definition 3.11
T	Absolute temperature	Subsection 1.1.2
T_{end}	Length of the time interval of interest	Subsection 1.1.2
T_i, $\tilde{T}_i$	Convection bilinear forms	(4.67)
$\mathcal{T}$, $\tilde{\mathcal{T}}$	Spatio-configurational convection bilinear forms	(1.28)
$\mathcal{T}^{(i)}$, $\tilde{\mathcal{T}}^{(i)}$	Convection Gram matrices	(4.74)
$\boldsymbol{u}$	Fluid velocity	(1.1)
U_0	Characteristic macroscopic velocity	Subsection 1.1.2
U, U_i	Spring potentials	Subsection 1.1.2, Subsection 1.1.3
$U_r^{\cdot}$, $U_{(r,s)}^{\cdot}$	Shifted boundary strip-describing set	(2.25)
$\hat{U}$	Antiderivative of $\hat{F}$ with $\hat{U}(0) = 0$	Subsection 2.2.2
$\mathrm{V}^{\cdot 1}_{\rho^{\cdot 2}}(\Lambda)$	Non-uniformly weighted Sobolev space	(4.32)
$\mathrm{V}^{\cdot 1}_{\rho^{\cdot 2}}(\Lambda^N)$	Multi-dimensional non-uniformly weighted Sobolev spaces	(4.39)
$\mathrm{V}^{\cdot,\mathrm{mix}}_{\rho^{\cdot 2}}(\Lambda^N)$	Higher-mixed-derivatives Multi-dimensional non-uniformly weighted Sobolev spaces	(4.40)
$w_{\mathsf{D}}^{(k)}$	Weight for 1/M-weighted quadrature over D	Subsection 1.1.4
$w_\Omega^{(k)}$	Weight for quadrature over Ω	Subsection 1.1.4
$\boldsymbol{W}_\nu$	Wiener process corresponding to the ν-th bead	Subsection 1.1.2
Wi	Weissenberg number	Subsection 1.1.2
$\boldsymbol{x}$	Generic spatial variable	Subsection 1.1.2
$\boldsymbol{x}^{(k)}$	Point for quadrature on Ω	Subsection 1.1.4
Z_i	Normalization factor for partial Maxwellians	(1.24)
$\mathcal{Z}_\theta$	Approximability space with associated rate θ	(4.21)
α_i	Exponent associated to a Maxwellian	Hypothesis E
$\alpha^{(n)}$	Coefficients computed by the n-th iteration of Algorithm II	(3.5)
$\hat{\alpha}^{(n)}$	Coefficients computed by the n-th iteration of Algorithm VI	(4.6)
γ_i	Margin of power-like behavior of a Maxwellian	Hypothesis E
δ_n	Error after the n-th iteration of Algorithm I or Algorithm II	(3.6)

$\hat{\delta}_n$	Error after the n-th iteration of Algorithm V or Algorithm VI	(4.7)
$\tilde{\Delta}_r$, $\tilde{\Delta}_{(r,s)}$	Boundary-describing-function domain	Definition 2.16
ζ	Drag coefficient	(1.4)
ζ_{R}	Riemann's zeta function	(3.50)
μ_s	Solvent dynamic viscosity	(1.2)
λ	Characteristic relaxation time of a spring	Subsection 1.1.2
$\lambda_n^{(i)}$, $\lambda_{\boldsymbol{n}}$	Eigenvalues	(3.33), (3.37)
$\lambda_{\min}$, $\lambda_{\max}$	Smallest and largest eigenvalues of the matrix A	Subsection 1.1.2
Λ	$(-1,1)$	Subsection 4.1.6
Λ_r, $\Lambda_{(r,s)}$	Boundary-describing set	Subsection 2.3.1
ρ	Fluid density / Real function $y \mapsto 1-y^2$	(1.1)/ Subsection 4.1.6
$\boldsymbol{\sigma}$	Fluid internal stress tensor	(1.1)
$\boldsymbol{\tau}$	Polymeric extra stress	(1.2)
$\mathrm{T}^{(\cdot)}$, $\tilde{\mathrm{T}}^{(\cdot)}$	Nets	(3.40), (4.49)
$\Upsilon^{(\cdot)}$, $\tilde{\Upsilon}^{(\cdot)}$	Nets	(3.48), (4.49)
ψ	Solution to the full / configurational Fokker–Planck equation	(1.22) / (2.1)
ψ_0	Initial condition for the full Fokker–Planck equation	(1.22)
ψ^{beads}	Probability density function with respect to the bead positions	Subsection 1.1.2
ψ^n, $\psi^{n+\frac{1}{2}}$	Approximations of ψ after temporal discretization	Subsection 1.1.4
Ω	Spatial domain / generic domain for the study of eigenvalues	Subsection 1.1.2, Subsection 2.2.3
supp	Support	Section 1.4
$\Subset$	Compact-embedding relation	Section 1.4
$[\cdot]$	Set $\{1,\dots,\cdot\} \subset \mathbb{N}$	Section 1.4
$\cdot'$	Vector resulting from removing the last component of a vector $\cdot$	Subsection 2.3.1
$\times$, $\bigtimes$	Cartesian product	Section 1.4
$\otimes$, $\bigotimes$	Tensor product	Section 1.4

CHAPTER 1

Introduction

1.1. Origin of the Fokker–Planck equation

1.1.1. Dilute polymer solutions. The simulation of dilute polymer solutions is important in a number of industrial contexts. However, their rheological behavior is complex and their simulation difficult. Thus, every practical attempt at simulating their flow must compromise between fidelity to the underlying physical reality and computational feasibility.

A polymer is a chain-like macromolecule composed of repeating units, called monomers, bound together by chemical bonds. A dilute polymer solution occurs when these long molecules are dispersed in a solvent in such a way that their mutual interaction is negligible compared to their interaction with the solvent [**SC88**].

The polymer molecules can exert an enormous influence on the behavior of their solvent and make it depart ostensibly from what can be expected from a Newtonian fluid [**Jos90**]. In particular the polymer molecules can retain some memory of their previous configurations as time progresses giving rise to elastic effects.

Each polymer molecule is immensely complicated in the way in which its many constituent parts respond to the movement of each other and of the surrounding solvent particles. The range of both spatial and time scales that a detailed description would entail is huge [**GP02**]. Moreover, there are—essentially random—thermal effects in play.

As the underlying physics is so complicated and the aims of the practitioners are so varied it doesn't come as a surprise that there exists a rich hierarchy of models of the polymer molecules themselves or of their interaction with the solvent. These models span all the way from the atomistic to the fully macroscopic [**Keu00**]. Whether or not a model is sensible for a given application will depend on which characteristics of the system are deemed important and on what spatial and temporal scales they are sought for.

In the context where the Fokker–Planck equation we will study arises the primary interest lies in the bulk form of certain features of the solution, namely velocity and, possibly, internal stresses. Hence, those variables are modeled as functions of a spatial continuum. The microscopical polymer conformations are of interest only inasmuch as they affect those variables of primary interest. Here is where one of the challenges of dilute polymer simulation presents itself, namely, that as pointed out before, the conformation of the polymer molecules does indeed affect the target variables, making the problem a multiscale one in a natural way.

One way to account feasibly for the activity of the polymer molecules within the framework of continuum mechanics is by the so-called *micro-macro* method [**Keu04**]. It entails using some coarse-grained kinetic theory model of the polymer molecule conformations and simulating their evolution under the relevant mechanical and thermal forces while using standard approximation techniques for the momentum and continuity equations of the flow [**Keu00**]. This can be done directly from the underlying stochastic model using techniques such as CONNFFESSIT [**LÖ93, FLÖ95**], which simulates ensembles of idealized polymers using Monte Carlo methods. Alternatively, it can be done via a Fokker–Planck equation [**Loz03**], in which a probability density function for the molecular conformations is approximated instead using deterministic numerical schemes. Be it in one way or the other, the polymeric contribution to the internal stresses of the solution is computed as an aggregate of the contributions of individual molecules. Kinetic models of dilute polymers have a rich hierarchy of their own [**BCAH87**] and give practitioners a great deal of flexibility at the time of deciding how many and which of the molecular degrees of freedom to retain.

The widely used alternative is a fully macroscopic approach, wherein closed form constitutive laws model the contribution of the polymer molecules to the internal stresses of the flow. While some of these constitutive laws were devised with little more than data-fitting in mind, many of them are derived from coarse-grained models of the polymer molecules via closure approximations. However, not every coarse-grained molecular model admits an equivalent closure approximation. Thus, the highly desirable flexibility with which kinetic theory incorporates microscopic phenomena is not retained by macroscopic constitutive laws. In fact, even when there exists a true equivalence between a kinetic theory model and its closure approximation, the numerical scheme resulting from the latter may turn out to be inferior to the one stemming from the former [**Loz03**].

1.1.2. Mathematical model. In this subsection we present the larger mathematical model from which the elliptic Fokker–Planck equation (2.1)—whose approximation is our actual object of study—will be distilled.

As it transpires from the above discussion, we model the flow velocity and internal stresses as functions of a material continuum, while using a kinetic theory model for the microscopic polymer molecule conformations. The tools of statistical mechanics will link the two spatio-temporal scales.

We start with the incompressible Navier–Stokes equations of flow that express the conservation of linear and angular momentum and mass within the fluid (for a derivation see,

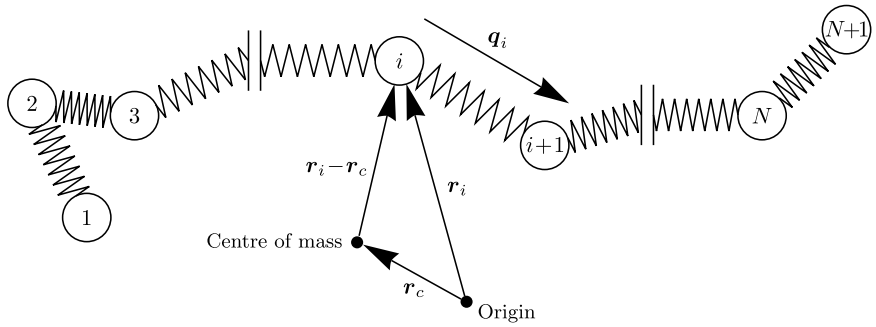

FIGURE 1.1. Rouse chain with N springs and $N+1$ beads. Adapted from Figure 11.4-1 of [**BCAH87**].

e.g., [**Bat67**]). These read

$$\rho \frac{D\boldsymbol{u}}{Dt} = \operatorname{div}_{\boldsymbol{x}}(\boldsymbol{\sigma}) + \boldsymbol{f}, \tag{1.1a}$$

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\mathrm{T}}, \quad \text{and} \tag{1.1b}$$

$$\operatorname{div}_{\boldsymbol{x}}(\boldsymbol{u}) = 0, \tag{1.1c}$$

where the (vector, tensor, vector) fields $\boldsymbol{u}$, $\boldsymbol{\sigma}$ and $\boldsymbol{f}$ stand for the velocity, internal stress and external body force, respectively, while the scalar ρ stands for the density of the fluid, which is assumed constant in both space and time. $\frac{D\boldsymbol{u}}{Dt} = \frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla_{\boldsymbol{x}})\boldsymbol{u}$ is the material derivative of $\boldsymbol{u}$. These equations are posed on a temporal-spatial domain of the form $(0, T_{\mathrm{end}}] \times \Omega$, with $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, being the *spatial* domain—the subscript $\boldsymbol{x}$ on differential operators is to emphasize that they have to be taken with respect to Ω. The equations (1.1) must be complemented with suitable initial and boundary conditions.

Now, we model the internal stress tensor as being the sum of a macroscopic Newtonian and a non-Newtonian part due to the effects of the microscopic polymer conformations:

$$\boldsymbol{\sigma} = -p\boldsymbol{I} + \mu_s \left(\nabla_{\boldsymbol{x}} \boldsymbol{u} + (\nabla_{\boldsymbol{x}} \boldsymbol{u})^{\mathrm{T}} \right) + \boldsymbol{\tau}, \tag{1.2}$$

where the scalar field p is the pressure, $\boldsymbol{I}$ is the identity tensor and μ_s the (constant) solvent dynamic viscosity and $\boldsymbol{\tau}$ the polymeric extra stress. It is worth noting that, as both the Newtonian and (as will be seen in equation (1.9)) non-Newtonian contributions to $\boldsymbol{\sigma}$ are symmetric, equation (1.1b) is automatically satisfied.

We will only give a scant derivation of the Fokker–Planck equation and refer the interested reader to [**BS07**], [**DLO07**], [**BCAH87**] and [**Kne08**], which the discussion that follows is based upon.

The polymer molecules are modeled as *Rouse chains*, which consist of a linear arrangement of $N+1$ beads with mass m each joined by N massless elastic springs. A model molecule in the case $N = 1$, corresponding to a single spring connecting two beads, is known as a

dumbbell. Whatever $N \in \mathbb{N}$ is, the state of a molecule in such an arrangement is completely determined by the position of each of its $N+1$ beads, denoted by $\boldsymbol{r}_\nu$ (see Figure 1.1).

Equivalently the state of a molecule can be described by the center of mass of the bead system, $\boldsymbol{r}_c = (N+1)^{-1} \sum_{\nu\in[N+1]} \boldsymbol{r}_\nu$ (here we are using the notation $[k]$ to denote the set of natural numbers $\{1,\dots,k\}$ for all $k \in \mathbb{N}$), together with the N connector vectors $\boldsymbol{q}_i := \boldsymbol{r}_{i+1} - \boldsymbol{r}_i$. For the sake of convenience we introduce the notation $\boldsymbol{r}$ in lieu of $(\boldsymbol{r}_1,\dots,\boldsymbol{r}_{N+1})$ and, analogously, $\boldsymbol{q}$ in place of $(\boldsymbol{q}_1,\dots,\boldsymbol{q}_N)$.

Each spring exerts an elastic conservative force on the beads it connects along the corresponding connector vector and has a magnitude that depends isotropically on it. We model that dependence by a *spring force* function $\boldsymbol{F}\colon D \to \mathbb{R}^d$, which has the form $\boldsymbol{F}(\boldsymbol{p}) = H\,U'(\frac{1}{2}\,|\boldsymbol{p}|^2)\boldsymbol{p}$, where $D \subseteq \mathbb{R}^d$ is either a 0-centered open ball or the full of $\mathbb{R}^d$ that contains all the admissible connector vectors, H is a spring constant, and U is a *spring potential*, which we take normalized as

$$U'(0) = 1. \tag{1.3}$$

It is immediate that $\boldsymbol{F}(\boldsymbol{p}) = -\boldsymbol{F}(-\boldsymbol{p})$.

In the absence of external forces and neglecting inertial effects the Langevin equation for the ν-th bead in this model reads

$$\zeta\,(\,\mathrm{d}\boldsymbol{r}_\nu - \boldsymbol{u}(\boldsymbol{r}_\nu,\cdot)\,\mathrm{d}t) = \boldsymbol{B}_\nu\,\mathrm{d}t + \sum_{i=1}^{N} G_{\nu i}\boldsymbol{F}(\boldsymbol{q}_i)\,\mathrm{d}t. \tag{1.4}$$

Here ζ is a drag coefficient, the $\boldsymbol{B}_\nu$ denote Brownian forces acting on each bead and the $(N+1)\times N$ matrix G is defined by $G_{\nu i} := \delta_{\nu i} - \delta_{\nu,i+1}$; that is,

$$G = \begin{pmatrix} 1 & & \\ -1 & \ddots & \\ & \ddots & 1 \\ & & -1 \end{pmatrix} \in \mathbb{R}^{(N+1)\times N}.$$

This is an equivalent way of expressing that the ν-th bead is pulled by the $(\nu-1)$-th spring in the $-\boldsymbol{q}_{\nu-1} = \boldsymbol{r}_{\nu-1} - \boldsymbol{r}_\nu$ direction and by the ν-th spring in the $\boldsymbol{q}_\nu = \boldsymbol{r}_{\nu+1} - \boldsymbol{r}_\nu$ direction with proper provision for the beads at the extremes. Note that $\boldsymbol{q}_i = -\sum_{\nu\in[N+1]} G_{\nu i}\boldsymbol{r}_\nu$.

Each Brownian force $\boldsymbol{B}_\nu$ is defined by a d-component vectorial Wiener process $\boldsymbol{W}_\nu$ via $\boldsymbol{B}_\nu\,\mathrm{d}t = \sqrt{2k_BT\zeta}\;\mathrm{d}\boldsymbol{W}_\nu$, where k_B is the Boltzmann constant and T is the absolute temperature.

Next, we define the position-space probability density function as the nonnegative function $\psi^{\text{beads}}\colon \mathbb{R}^{d(N+1)+1} \to \mathbb{R}$ such that, given a Borel set $\mathcal{A}$ of the position-space $\mathbb{R}^{d(N+1)}$,

$$\int_{\mathcal{A}} \psi^{\text{beads}}(\boldsymbol{r},t)\,\mathrm{d}\boldsymbol{r}$$

is the expected number of idealized molecules at time t having bead positions in $\mathcal{A}$. Using arguments from the analysis of stochastic processes and from statistical physics (cf. [**BCAH87**,

§17–18], [**Kne08**, Subsection 1.3.1]) it can be shown that, as a consequence of (1.4), the evolution of ψ^{beads} is governed by the equation

$$\frac{\partial \psi^{\text{beads}}}{\partial t} + \sum_{\nu=1}^{N+1} \operatorname{div}_{\boldsymbol{r}_\nu} \left(\boldsymbol{u}(\boldsymbol{r}_\nu, \cdot)\psi^{\text{beads}} + \frac{1}{\zeta} \sum_{i=1}^{N} G_{\nu i} \boldsymbol{F}(\boldsymbol{q}_i)\psi^{\text{beads}} \right) = \sum_{\nu=1}^{N+1} \frac{k_B T}{\zeta} \Delta_{\boldsymbol{r}_\nu} \psi^{\text{beads}},$$

which, mediating the change of variables $\psi(\boldsymbol{r}_c, \boldsymbol{q}) := \psi^{\text{beads}}(\boldsymbol{r}(\boldsymbol{r}_c, \boldsymbol{q}))$, can be expressed as

$$\begin{aligned} \frac{\partial \psi}{\partial t} + \operatorname{div}_{\boldsymbol{r}_c} \left(\frac{1}{N+1} \sum_{\nu=1}^{N+1} \boldsymbol{u}(\boldsymbol{r}_\nu, \cdot)\psi \right) & \\ - \sum_{i=1}^{N} \operatorname{div}_{\boldsymbol{q}_i} \left(\left[\sum_{\nu=1}^{N+1} G_{\nu i} \boldsymbol{u}(\boldsymbol{r}_\nu, \cdot) + \frac{1}{\zeta} \sum_{j=1}^{N} A_{ij} \boldsymbol{F}(\boldsymbol{q}_j) \right] \psi \right) & \\ = \frac{k_B T}{\zeta(N+1)} \Delta_{\boldsymbol{r}_c} \psi + \frac{k_B T}{\zeta} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \operatorname{div}_{\boldsymbol{q}_i} \nabla_{\boldsymbol{q}_j} \psi, & \quad (1.5) \end{aligned}$$

where $A = G^{\mathrm{T}} G$ is known as the Rouse matrix. In explicit form,

$$A = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{N \times N}. \quad (1.6)$$

The matrix A is symmetric and positive-definite and we denote its minimal and maximal eigenvalues by $\lambda_{\min}$ and $\lambda_{\max}$, respectively.

Next, we make the local homogeneity assumption, which states that, on the length scale of the molecule, $\boldsymbol{u}$ is a linear function of its spatial variable. That is, $\boldsymbol{u}(\boldsymbol{r}_c + \boldsymbol{p}, t) \approx \boldsymbol{u}(\boldsymbol{x}, t) + \nabla_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}, t)\boldsymbol{p}$, where we have explicitly identified the generic spatial variable $\boldsymbol{x}$ with the center of mass of the bead system. Equation (1.5) turns into the Fokker–Planck equation

$$\begin{aligned} \frac{\partial \psi}{\partial t} + \operatorname{div}_{\boldsymbol{x}} (\boldsymbol{u}\psi) + \sum_{i=1}^{N} \operatorname{div}_{\boldsymbol{q}_i} \left((\nabla_{\boldsymbol{x}} \boldsymbol{u})\, \boldsymbol{q}_i \psi - \frac{1}{\zeta} \sum_{j=1}^{N} A_{ij} \boldsymbol{F}(\boldsymbol{q}_j)\psi \right) & \\ = \frac{k_B T}{\zeta(N+1)} \Delta_{\boldsymbol{x}} \psi + \frac{k_B T}{\zeta} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \operatorname{div}_{\boldsymbol{q}_i} \nabla_{\boldsymbol{q}_j} \psi. & \end{aligned} \quad (1.7)$$

At this stage it is worth noting that ψ must be interpreted as the distribution function such that $\int_{\mathcal{A}} \psi(\boldsymbol{x}, \boldsymbol{q}, t)\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{q}$ is the number of molecules whose ensembles of center of mass and configuration vectors lie in the set $\mathcal{A} \subseteq \Omega \times D^N$ at time t. In particular,

$$n_p(\boldsymbol{x}, t) = \int \psi(\boldsymbol{x}, \boldsymbol{q}, t)\, \mathrm{d}\boldsymbol{q} \quad (1.8)$$

is the molecule concentration at the point $\boldsymbol{x}$ and time t and has units of length to the minus d. From now on we assume that there are no concentration gradients; i.e., n_p is constant

throughout the spatial domain. As follows from [**Kne08**, Lemma 1.3], (1.7) ensures that n_p is constant throughout time too.

The internal configuration of the polymer molecules contributes to the polymeric extra tensor $\boldsymbol{\tau}$ in two distinct ways (in the presence of external forces there would be a third contribution; see [**BCAH87**, §13.3]). First, there is a contribution from the force of tension or compression transmitted by each of the springs as it straddles across planes in the solution, given by

$$\int_{D^N} \boldsymbol{q}_i \boldsymbol{F}(\boldsymbol{q}_i)^{\mathrm{T}} \psi(\boldsymbol{x}, \boldsymbol{q}, t)\, \mathrm{d}\boldsymbol{q} \qquad \forall\, i \in [N].$$

In the second place, there is a contribution from the momentum transported by each of the beads crossing the same planes. Under the equilibration in momentum space assumption this contribution is

$$-n_p k_B T \boldsymbol{I} \qquad \forall\, \nu \in [N{+}1].$$

Therefore,

$$\boldsymbol{\tau}(\boldsymbol{x}, t) = \sum_{i=1}^{N} \int_{D^N} \boldsymbol{q}_i \boldsymbol{F}(\boldsymbol{q}_i)^{\mathrm{T}} \psi(\boldsymbol{x}, \boldsymbol{q}, t)\, \mathrm{d}\boldsymbol{q} - (N{+}1) n_p k_B T \boldsymbol{I}, \tag{1.9}$$

which is called the Kramers expression. Note that $\boldsymbol{\tau}$ is a symmetric tensor due to the fact that the connector forces act along the connector vectors.

In this work we focus on spring force laws which only allow the springs to extend up to a certain maximal extension. Moreover, the force laws under consideration enforce this finite extensibility by making the springs store amounts of energy that diverge to positive infinity as the spring approaches this maximal extension. Mathematically, we express this as

$$\lim_{s \to \left(\frac{1}{2} q_{\max}^2\right)_-} U(s) = \infty, \tag{1.10}$$

where $q_{\max}$ is the maximal extension of the springs—we recall that the force has the form $\boldsymbol{F}(\boldsymbol{p}) = HU'(\frac{1}{2}|\boldsymbol{p}|^2)\boldsymbol{p}$. The rationale behind this requirement is that these properties are borne by the *Inverse Langevin force law*, which is given by

$$D = B(0, q_{\max}) \subset \mathbb{R}^d \qquad \text{and} \qquad \boldsymbol{F}(\boldsymbol{p}) = \frac{H}{3} \frac{L^{-1}\left(\frac{|\boldsymbol{p}|}{q_{\max}}\right)}{|\boldsymbol{p}|/q_{\max}} \boldsymbol{p}, \tag{1.11}$$

where L is the Langevin Function $t \mapsto \coth(t) - 1/t$—the fact that this Force law does indeed come from a potential which obeys the properties (1.3) and (1.10) will be shown later in Subsection 2.2.2. By $B(0, q_{\max})$ we mean, as usual, the ball centered in 0 with radius $q_{\max}$. The Inverse Langevin force law, in true coarse-graining form, can be obtained from a limit process involving polymer models consisting of a chain of freely joined bead-rod models where the number of conjoined rods tends to infinity while overall mean-length of the chain is being kept constant (see [**KG42**]). As the Inverse Langevin force law and its potential are hard to

manipulate, a number of approximations have been suggested in the literature, some of which will be presented later in Subsection 1.1.3.

Remark 1.1. An important spring force model, which is excluded from our considerations, is the simple *Hookean model* described by

$$D = \mathbb{R}^d, \qquad U(s) = s \qquad \text{and} \qquad \boldsymbol{F}(\boldsymbol{p}) = H\boldsymbol{p}.$$

However, in many practically relevant flow regimes the physically unrealistic allowance of the Hookean model for indefinitely extended springs outweighs its mathematical convenience.

Another force model which is also out of the scope of this work is the *Linear locked model* [**TS71**] described by

$$D = B(0, q_{\max}) \subset \mathbb{R}^d, \qquad U(s) = s \qquad \text{and} \qquad \boldsymbol{F}(\boldsymbol{p}) = H\boldsymbol{p};$$

i.e., a truncation of the Hookean model. Here, in violation of (1.10), the maximal extension is attainable with finite energy.

We define non-dimensionalized (hatted) variables in terms of their non-hatted counterparts as

$$\boldsymbol{x} = L_0\hat{\boldsymbol{x}}, \quad \boldsymbol{q}_i = l_0\hat{\boldsymbol{q}}_i, \quad \boldsymbol{u} = U_0\hat{\boldsymbol{u}}, \quad t = L_0/U_0\hat{t}, \quad \text{and} \quad \psi = \hat{\psi}\, n_p/l_0^{Nd},$$

where $l_0 := \sqrt{k_BT/H}$ is the characteristic length-scale of a spring and L_0 and U_0 are the characteristic macroscopic length and velocity, respectively. Using these relations the Fokker–Planck equation (1.7) can be recast as

$$\frac{U_0}{L_0}\frac{\partial\hat{\psi}}{\partial\hat{t}} + \frac{U_0}{L_0}\operatorname{div}_{\hat{\boldsymbol{x}}}(\hat{\boldsymbol{u}}\hat{\psi}) + \sum_{i=1}^{N}\operatorname{div}_{\hat{\boldsymbol{q}}_i}\left(\frac{U_0}{L_0}(\nabla_{\hat{\boldsymbol{x}}}\hat{\boldsymbol{u}})\hat{\boldsymbol{q}}_i\hat{\psi} - \frac{1}{4\lambda}\sum_{j=1}^{N}A_{ij}\hat{\boldsymbol{F}}(\hat{\boldsymbol{q}}_j)\hat{\psi}\right)$$
$$= \frac{1}{4\lambda(N+1)}\left(\frac{l_0}{L_0}\right)^2\Delta_{\hat{\boldsymbol{x}}}\hat{\psi} + \frac{1}{4\lambda}\sum_{i=1}^{N}\sum_{j=1}^{N}A_{ij}\,\operatorname{div}_{\hat{\boldsymbol{q}}_i}\nabla_{\hat{\boldsymbol{q}}_j}\hat{\psi},$$

where $\lambda := \zeta/(4H)$ is the characteristic relaxation time of a spring, $\hat{U}(s) = l_0^{-2}U(l_0^2 s)$, $\hat{\boldsymbol{F}}(\hat{\boldsymbol{q}}) = \hat{U}'(\frac{1}{2}\,|\hat{\boldsymbol{q}}|^2)\hat{\boldsymbol{q}} = (Hl_0)^{-1}\boldsymbol{F}(l_0\hat{\boldsymbol{q}})$ and the spatial, configurational and temporal variables $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{q}}_i$ and $\hat{t}$ now live in $\hat{\Omega} = \Omega/L_0$, $\hat{D} = D/l_0$ and $\hat{T}_{\text{end}} = U_0T_{\text{end}}/L_0$, respectively.

Defining the non-dimensional Weissenberg number as $\mathrm{Wi} := \lambda U_0/L_0$ (that is, the ratio of the microscopic to macroscopic time scales) and discarding the hats we have

$$\frac{\partial\psi}{\partial t} + \operatorname{div}_{\boldsymbol{x}}(u\psi) + \sum_{i=1}^{N}\operatorname{div}_{\boldsymbol{q}_i}\left((\nabla_{\boldsymbol{x}}\boldsymbol{u})\boldsymbol{q}_i\psi - \frac{1}{4\mathrm{Wi}}\sum_{j=1}^{N}A_{ij}\boldsymbol{F}(\boldsymbol{q}_j)\psi\right)$$
$$= \frac{1}{4\mathrm{Wi}(N+1)}\left(\frac{l_0}{L_0}\right)^2\Delta_{\boldsymbol{x}}\psi + \frac{1}{4\mathrm{Wi}}\sum_{i=1}^{N}\sum_{j=1}^{N}A_{ij}\,\operatorname{div}_{\boldsymbol{q}_i}\nabla_{\boldsymbol{q}_j}\psi, \quad (1.12)$$

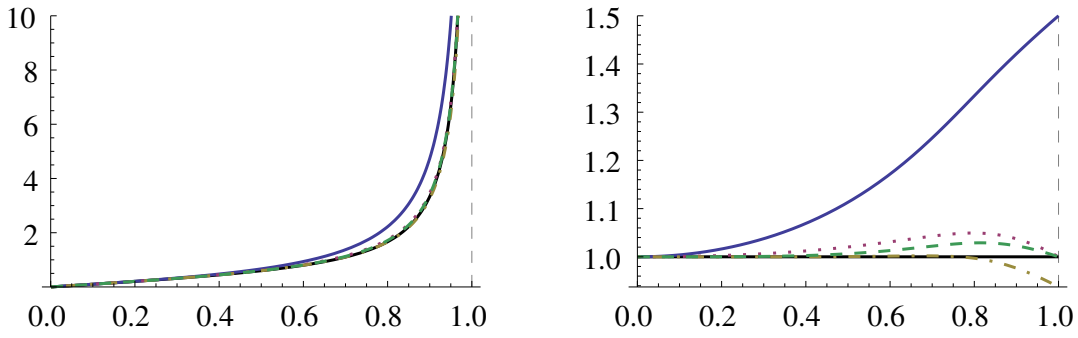

FIGURE 1.2. *Left*: Plots of $|\boldsymbol{p}|/\sqrt{b_i}$ versus $|\boldsymbol{F}_i(\boldsymbol{p})|/\sqrt{b_i}$ where $\boldsymbol{F}_i$ corresponds to the Inverse Langevin (continuous black), FENE (continuous blue), CPAIL (red dotted), TEAIL (olive dot-dashed) and CP (green dashed) force laws. They are barely distinguishable but for the FENE case. *Right*: Plots of $|\boldsymbol{p}|/\sqrt{b_i}$ versus $|\boldsymbol{F}_i(\boldsymbol{p})|/|\boldsymbol{F}(\boldsymbol{p})|$, where $\boldsymbol{F}_i$ corresponds to the Inverse Langevin, FENE, CPAIL, TEAIL and CP force laws (using the same color- and line-codings as in the previous plot) and $\boldsymbol{F}$ corresponds to the Inverse Langevin force law.

From this non-dimensionalization procedure also arises the important non-dimensional parameter

$$b := \frac{q_{\max}^2 H}{k_B T} = \frac{q_{\max}^2}{l_0^2}, \tag{1.13}$$

which, as the formula above makes explicit, measures how the maximal admissible extension compares with the characteristic microscopic length-scale l_0 of a spring. Having defined b we can express the non-dimensionalized single-spring configuration domain as $D = \{\boldsymbol{p} \in \mathbb{R}^d\colon\ |\boldsymbol{p}| < \sqrt{b}\}$ now. After non-dimensionalization the conservation of the solute concentration is manifested as

$$\int_{D^N} \psi(\boldsymbol{x}, \boldsymbol{q}, t)\, \mathrm{d}\boldsymbol{q} = 1, \tag{1.14}$$

and the transformed Kramers expression (1.9) adopts the form

$$\boldsymbol{\tau}(\boldsymbol{x}, t) = n_p k_B T \sum_{i=1}^{N} \int_{D^N} \boldsymbol{q}_i \boldsymbol{F}(\boldsymbol{q}_i)^{\mathrm{T}} \psi(\boldsymbol{x}, \boldsymbol{q}, t)\, \mathrm{d}\boldsymbol{q} - (N+1) n_p k_B T \boldsymbol{I}. \tag{1.15}$$

1.1.3. Fokker–Planck equation. So far, we have assumed that all the springs obey the same force law. However, this is not necessary. Thus, at negligible extra cost, we will allow different springs obey the same force law with different parameter or different force laws altogether.

Therefore, the (now potentially different) spring forces in the model are given by functions $\boldsymbol{F}_i\colon D_i \to \mathbb{R}^d$, which have the form $\boldsymbol{F}_i(\boldsymbol{p}) = U_i'(\frac{1}{2}|\boldsymbol{p}|^2)\boldsymbol{p}$, $\boldsymbol{p} \in D_i := B(0, \sqrt{b_i}) \subset \mathbb{R}^d$, $b_i > 0$, $i \in [N]$, and the spring potentials $U_i\colon [0, b_i/2) \to \mathbb{R}$ are such that $U_i(s) \to \infty$ as $s \to b_i/2_-$. It follows that $\boldsymbol{F}_i(\boldsymbol{p}) = -\boldsymbol{F}_i(-\boldsymbol{p})$ for all $\boldsymbol{p} \in D_i$. The generic configurational variable $\boldsymbol{q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N)$ now lives in the configuration domain $\mathsf{D} := D_1 \times \cdots \times D_N \subset \mathbb{R}^{Nd}$.

Typical force law examples include the *FENE (Finitely Extensible Nonlinear Elastic) model* [**War72**] with

$$U_i(s) = -\frac{b_i}{2}\ln\left(1-\frac{2s}{b_i}\right) \qquad \text{and} \qquad \boldsymbol{F}_i(\boldsymbol{q}_i) = \frac{1}{1-\frac{|\boldsymbol{q}_i|^2}{b_i}}\boldsymbol{q}_i, \tag{1.16}$$

Cohen's Padé approximant to the Inverse Langevin (CPAIL) model [**Coh91**] with

$$U_i(s) = \frac{s}{3} - \frac{b_i}{3}\ln\left(1-\frac{2s}{b_i}\right) \qquad \text{and} \qquad \boldsymbol{F}_i(\boldsymbol{q}_i) = \frac{1-\frac{|\boldsymbol{q}_i|^2}{3b_i}}{1-\frac{|\boldsymbol{q}_i|^2}{b_i}}\boldsymbol{q}_i \tag{1.17}$$

and *Treloar's empirical approximant to the Inverse Langevin model (TEAIL) model* [**Tre54**, Appendix 2] with

$$U_i(s) = \frac{5b_i}{16}\arctan\left(\frac{4s}{5b_i+2s}\right) - \frac{5b_i}{32}\ln\left(\frac{5\left(1-\frac{2s}{b_i}\right)^2}{5+\frac{4s}{b_i}+\left(\frac{2s}{b_i}\right)^2}\right)$$
$$\text{and} \qquad \boldsymbol{F}_i(\boldsymbol{q}_i) = \frac{1}{1-\left(\frac{3}{5}\frac{|\boldsymbol{q}_i|^2}{b_i}+\frac{1}{5}\frac{|\boldsymbol{q}_i|^4}{b_i^2}+\frac{1}{5}\frac{|\boldsymbol{q}_i|^6}{b_i^3}\right)}\boldsymbol{q}_i. \tag{1.18}$$

We note that all three of these force laws are approximations to the *Inverse Langevin force law* [**KG42**], already presented in (1.11), whose non-dimensionalized form is

$$\boldsymbol{F}_i(\boldsymbol{q}_i) = \frac{\sqrt{b_i}}{3}L^{-1}\left(\frac{|\boldsymbol{q}_i|}{\sqrt{b_i}}\right)\frac{\boldsymbol{q}_i}{|\boldsymbol{q}_i|}; \tag{1.19}$$

we recall that the Langevin function L is defined by $L(t) := \coth(t) - 1/t$ on $[0,\infty)$ (more precisely, defined by the formula in $(0,\infty)$ and continuously extended to 0 at 0). As L is strictly monotonic increasing on $[0,\infty)$ and tends to 1 as its argument tends to ∞, it follows that the function $|\boldsymbol{q}_i| \in [0,\sqrt{b_i}) \mapsto L^{-1}(|\boldsymbol{q}_i|/\sqrt{b_i}) \in [0,\infty)$ is strictly monotonic increasing, with a vertical asymptote at $|\boldsymbol{q}_i| = \sqrt{b_i}$. We say more about the Inverse Langevin force law in Subsection 2.2.2.

Remark 1.2. The $(1,2)$-, $(3,2)$-, and $(1,6)$-Padé approximants around 0 to $t \mapsto \frac{1}{3}L^{-1}(t)$ are, respectively,

$$t \mapsto \frac{t}{1-\frac{3}{5}t^2}, \qquad t \mapsto \frac{1-\frac{12}{35}t^2}{1-\frac{33}{35}t^2}t, \qquad \text{and} \qquad t \mapsto \frac{t}{1-\left(\frac{3}{5}t^2+\frac{36}{175}t^4+\frac{108}{875}t^6\right)}.$$

In order to ensure that the singularity lies where it should (at $t=1$) one can do the approximations $\frac{3}{5} \approx 1$, in the first case, $\frac{12}{35} \approx \frac{1}{3}$ and $\frac{33}{35} \approx 1$, in the second, and $\frac{36}{175} \approx \frac{1}{5}$ and $\frac{108}{875} \approx \frac{1}{5}$, in the third, which produce the FENE model (1.16), the CPAIL model (1.17) and the TEAIL

model (1.18), respectively—indeed, that is exactly what Cohen does in [**Coh91**] to produce the CPAIL model and, as shown in [**HS02**], it is a way to produce the TEAIL model.

An alternative to this perturbation of Padé approximations is the use of constrained Padé approximations which incorporate from the outset the singularity at $t = 1$. In order to preserve the oddness of the inverse Langevin function L^{-1} (that is, $L^{-1}(-t) = -L^{-1}(t)$, which is directly inherited from the corresponding property of the direct Langevin function L), it is better to constrain the Padé approximation by enforcing the presence of the factor $(1 - t^2)$ in the denominator of the approximant. As it will become apparent in due course, it is also desirable that the approximant shares with the Inverse Langevin function the limit

$$\lim_{t\to 1_-}\left[(1-t^2)\frac{1}{3}L^{-1}(t)\right] = \frac{2}{3}. \tag{1.20}$$

Performing a thus constrained Padé approximation of the Inverse Langevin Force results, in the $(3, 2)$-case, exactly in the CPAIL model (1.17). This is not surprising, for the adjustment of the coefficients in the numerator of the unconstrained $(3, 2)$-Padé approximant was performed by Cohen in order to ensure the reproduction of (1.20). The constrained $(5, 2)$-Padé approximation produces a barely more complicated force law:

$$U_i(s) = \frac{s}{3} - \frac{s^2}{15b_i} - \frac{b_i}{3}\ln\left(1 - \frac{2s}{b_i}\right) \qquad \text{and} \qquad \boldsymbol{F}_i(\boldsymbol{q}_i) = \frac{1 - \frac{2|\boldsymbol{q}_i|^2}{5b_i} + \frac{|\boldsymbol{q}_i|^4}{15b_i^2}}{1 - \frac{|\boldsymbol{q}_i|^2}{b_i}}\boldsymbol{q}_i. \tag{1.21}$$

This force law, which we will simply call *Constrained Padé (CP)*, has a singularity at $|\boldsymbol{q}_i| = \sqrt{b_i}$ of the same order and comes from an approximant that reproduces (1.20) by design and its approximation order is two orders higher than that of the CPAIL model (1.17) in the vicinity of $\boldsymbol{q}_i = \boldsymbol{0}$. One can produce many similar force laws by varying the polynomial orders of the numerator and the denominator of the sought-after constrained Padé approximant.

Recapping, we have, from (1.12), that the Fokker–Planck equation under consideration for the probability density function ψ has the following form (see also [**BS07, BS08, BS09, BS11a, BS11b**]):

$$\frac{\partial\psi}{\partial t} + \operatorname{div}_{\boldsymbol{x}}(\boldsymbol{u}\psi) + \sum_{i=1}^{N}\operatorname{div}_{\boldsymbol{q}_i}\left[(\nabla_{\boldsymbol{x}}\boldsymbol{u})\boldsymbol{q}_i\psi - \frac{1}{4\mathrm{Wi}}\sum_{j=1}^{N}A_{ij}(\boldsymbol{F}_j(\boldsymbol{q}_j)\psi + \nabla_{\boldsymbol{q}_j}\psi)\right]$$
$$= \frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\Delta_{\boldsymbol{x}}\psi, \quad (\boldsymbol{x},\boldsymbol{q},t) \in \Omega\times\mathsf{D}\times(0,T_{\mathrm{end}}], \tag{1.22a}$$

which is complemented by initial and no-flux boundary conditions

$$\psi(\cdot,\cdot,0) = \psi_0, \qquad (\boldsymbol{x},\boldsymbol{q}) \in \Omega\times\mathsf{D}, \tag{1.22b}$$

$$\frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\nabla_{\boldsymbol{x}}\psi\cdot\mathfrak{n}_{\boldsymbol{x}} = 0, \qquad (\boldsymbol{x},\boldsymbol{q},t) \in \partial\Omega\times\mathsf{D}\times(0,T_{\mathrm{end}}], \tag{1.22c}$$

and

$$\left[(\nabla_{\boldsymbol{x}}\boldsymbol{u})\boldsymbol{q}_i\psi - \frac{1}{4\mathrm{Wi}}\sum_{j=1}^{N} A_{ij}(\boldsymbol{F}_j(\boldsymbol{q}_j)\psi + \nabla_{\boldsymbol{q}_j}\psi)\right]\cdot \mathfrak{n}_{\boldsymbol{q}_i} = 0, \quad i \in [N],$$

$$(\boldsymbol{x},\boldsymbol{q},t) \in \Omega \times \partial\mathsf{D} \times (0,T_{\mathrm{end}}], \tag{1.22d}$$

where $\mathfrak{n}_{\boldsymbol{x}}$ and $\mathfrak{n}_{\boldsymbol{q}_i}$ are unit outward normal vector defined (a.e. with respect to the surface measure) on $\partial\Omega$ and ∂D_i, $i \in [N]$, respectively. We remark that the boundary condition (1.22d) is an ensemble of N boundary conditions, which collectively account for the full $(Nd-1)$-dimensional measure of $\partial\mathsf{D}$.

The choice of the boundary condition (1.22d) is motivated by the following: we want to make sure that the evolution of ψ is compatible with (1.14); that is,

$$\tilde{\rho}(\boldsymbol{x},t) := \int_{D_N} \psi(\boldsymbol{x},\boldsymbol{q},t)\,\mathrm{d}\boldsymbol{q} = 1$$

for all $(\boldsymbol{x},t) \in \Omega \times [0,T_{\mathrm{end}}]$. Assuming that $\tilde{\rho}(\boldsymbol{x},0) = 1$ for all $\boldsymbol{x} \in \Omega$, integrating (1.22a) and (1.22c) over D^N, using the divergence theorem and the boundary condition (1.22d) we deduce that

$$\frac{\partial\tilde{\rho}}{\partial t} + \mathrm{div}_{\boldsymbol{x}}(\boldsymbol{u}\tilde{\rho}) = \frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\Delta_{\boldsymbol{x}}\tilde{\rho} \tag{1.23}$$

in $\Omega \times (0,T_{\mathrm{end}}]$ subject to $\tilde{\rho}(\boldsymbol{x},0) \equiv 1$ and the boundary condition $\frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\nabla_{\boldsymbol{x}}\tilde{\rho}\cdot\mathfrak{n}_{\boldsymbol{x}} = 0$ on $\partial\Omega \times (0,T_{\mathrm{end}}]$. As (1.23) subject to the stated initial and boundary conditions has a unique solution, and the constant function 1 clearly is a solution, it follows that $\tilde{\rho}$ is identically 1 in the entirety of $\Omega \times [0,T_{\mathrm{end}}]$, and so (1.14) is preserved.

We define the partial Maxwellians M_i and the (full) Maxwellian M by

$$M_i(\boldsymbol{p}) := \frac{1}{Z_i}\exp\left(-U_i\left(\tfrac{1}{2}|\boldsymbol{p}|^2\right)\right), \quad \boldsymbol{p} \in D_i, \quad i \in [N]; \tag{1.24}$$

$$\mathsf{M}(\boldsymbol{q}) := \prod_{i=1}^{N} M_i(\boldsymbol{q}_i), \quad \boldsymbol{q} \in \mathsf{D}; \tag{1.25}$$

that is (in the notation which will be introduced in Section 1.4), $\mathsf{M} = \bigotimes_{i\in[N]} M_i$. Here, each Z_i is a positive constant chosen so that $\int_{D_i} M_i(\boldsymbol{p})\,\mathrm{d}\boldsymbol{p} = 1$ (in the cases of interest we will be able to do so because of the adoption of Hypothesis A in Chapter 2). Thereby, $\int_{\mathsf{D}} \mathsf{M}(\boldsymbol{q})\,\mathrm{d}\boldsymbol{q} = 1$. The fact that the Maxwellian factorizes—which comes from the fact that the energy stored in the chain is the sum of the potential energies stored in each spring—will be crucial throughout the rest of this work. For a start, this fact allows us to write

$$\boldsymbol{F}_j(\boldsymbol{q}_j)\psi + \nabla_{\boldsymbol{q}_j}\psi = \psi\nabla_{\boldsymbol{q}_j}U_j(\tfrac{1}{2}|\boldsymbol{q}_j|^2) + \nabla_{\boldsymbol{q}_j}\psi = \mathsf{M}\nabla_{\boldsymbol{q}_j}\left(\frac{\psi}{\mathsf{M}}\right), \tag{1.26}$$

whence we have got rid of the unbounded drift terms of the form $\operatorname{div}_{\boldsymbol{q}_i}(\boldsymbol{F}_j(\boldsymbol{q}_j))$ present in (1.22a) by trading them off with a singular diffusion term. This procedure is known as a *Kolmogorov symmetrization* and results in an *Ornstein–Uhlenbeck* operator.

Multiplying (1.22a) by φ/M, using (1.26) and (formally) integrating by parts, the corresponding weak form of (1.22) is: Find $\psi = \psi(\boldsymbol{x}, \boldsymbol{q}, t)$ such that

$$\int_{\Omega\times\mathsf{D}} \left\{ \frac{\partial\psi}{\partial t}\frac{\varphi}{\mathsf{M}} + \operatorname{div}_{\boldsymbol{x}}(\boldsymbol{u}\psi)\frac{\varphi}{\mathsf{M}} - \sum_{i=1}^{N}\left[(\nabla_{\boldsymbol{x}}\boldsymbol{u})\boldsymbol{q}_i\psi - \frac{1}{4\mathrm{Wi}}\sum_{j=1}^{N} A_{ij}\mathsf{M}\nabla_{\boldsymbol{q}_j}\left(\frac{\psi}{\mathsf{M}}\right)\right]\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\varphi}{\mathsf{M}}\right) + \frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\nabla_{\boldsymbol{x}}\psi\cdot\nabla_{\boldsymbol{x}}\varphi\frac{1}{\mathsf{M}}\right\} = 0 \quad (1.27)$$

for all $\varphi = \varphi(\boldsymbol{x}, \boldsymbol{q})$ in a suitable function space.

For the sake of convenience we define the following bilinear forms:

$$\tilde{\mathcal{T}}(\boldsymbol{u};\sigma,\tau) := \int_{\Omega\times\mathsf{D}} \operatorname{div}_{\boldsymbol{x}}(\boldsymbol{u}\sigma)\frac{\tau}{\mathsf{M}}, \qquad \tilde{\mathcal{K}}(\sigma,\tau) := \frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\int_{\Omega\times\mathsf{D}} \nabla_{\boldsymbol{x}}\sigma\cdot\nabla_{\boldsymbol{x}}\tau\frac{1}{\mathsf{M}}, \quad (1.28a)$$

$$\mathcal{T}(\boldsymbol{u};\sigma,\tau) := -\int_{\Omega\times\mathsf{D}}\sum_{i=1}^{N}(\nabla_{\boldsymbol{x}}\boldsymbol{u})\boldsymbol{q}_i\sigma\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\tau}{\mathsf{M}}\right), \quad (1.28b)$$

$$\mathcal{K}(\sigma,\tau) := \frac{1}{4\mathrm{Wi}}\int_{\Omega\times\mathsf{D}}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}\mathsf{M}\nabla_{\boldsymbol{q}_j}\left(\frac{\sigma}{\mathsf{M}}\right)\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\tau}{\mathsf{M}}\right). \quad (1.28c)$$

Then, (1.27) can be written concisely as

$$\left\langle\frac{\partial\psi}{\partial t}, \varphi/\mathsf{M}\right\rangle + \tilde{\mathcal{T}}(\boldsymbol{u};\psi,\varphi) + \tilde{\mathcal{K}}(\psi,\varphi) + \mathcal{T}(\boldsymbol{u};\psi,\varphi) + \mathcal{K}(\psi,\varphi) = 0 \quad (1.29)$$

for all $\varphi = \varphi(\boldsymbol{x}, \boldsymbol{q})$ in a suitable function space. We note that $\tilde{\mathcal{T}}$ and $\tilde{\mathcal{K}}$ involve partial derivatives of their arguments with respect to the spatial variable $\boldsymbol{x}$ only. Analogously, $\mathcal{T}$ and $\mathcal{K}$ involve partial derivatives of their arguments with respect to the configuration space variable $\boldsymbol{q}$ only. This observation motivates the use of the alternating direction scheme based on operator splitting whose informal description is given in the next subsection.

1.1.4. Alternating direction scheme. In this subsection we present an alternating-direction scheme which justifies focusing on a simplified form of the Fokker–Planck equation (1.22) which is only defined on the configuration domain D and is elliptic.

Let Δt be such that $M := T_{\mathrm{end}}/\Delta t \in \mathbb{N}$ and define $t^n := n\Delta t$ for $n \in \{0, \dots, M\}$ and $t^{n+1/2} := (n + \frac{1}{2})\Delta t$ for $n \in \{0, \dots, M-1\}$. We will consider the following *alternating-direction* semidiscretization of (1.27): We initialize the scheme by defining $\psi^0 := \psi_0$; for $n \in \{0, \dots, M-1\}$ and then define the 'intermediate' function $\psi^{n+1/2}$ and the approximation

ψ^{n+1} to $\psi(t^{n+1},\cdot,\cdot)$, respectively, by

$$\left\langle \frac{\psi^{n+1/2}-\psi^n}{\Delta t/2}, \frac{\varphi}{\mathsf{M}} \right\rangle + \tilde{\mathcal{T}}(\boldsymbol{u}(\cdot,t^{n+1});\psi^{n+1/2},\varphi) + \tilde{\mathcal{K}}(\psi^{n+1/2},\varphi) = -\mathcal{T}(\boldsymbol{u}(\cdot,t^n);\psi^n,\varphi) - \mathcal{K}(\psi^n,\varphi) \tag{1.30a}$$

and

$$\left\langle \frac{\psi^{n+1}-\psi^{n+1/2}}{\Delta t/2}, \frac{\varphi}{\mathsf{M}} \right\rangle + \mathcal{K}(\psi^{n+1},\varphi) = -\mathcal{T}(\boldsymbol{u}(\cdot,t^n);\psi^n,\varphi) - \tilde{\mathcal{T}}(\boldsymbol{u}(\cdot,t^{n+1});\psi^{n+1/2},\varphi) - \tilde{\mathcal{K}}(\psi^{n+1/2},\varphi), \tag{1.30b}$$

for all $\varphi = \varphi(\boldsymbol{x},\boldsymbol{q})$ in a suitable function space. In (1.30a) the spatial bilinear forms $\tilde{\mathcal{T}}$ and $\tilde{\mathcal{K}}$ are treated implicitly while the configuration space bilinear forms $\mathcal{T}$ and $\mathcal{K}$ are treated explicitly. In (1.30b) the spatial bilinear forms $\tilde{\mathcal{T}}$ and $\tilde{\mathcal{K}}$ and the configuration space bilinear form $\mathcal{T}$ associated with the drag term are treated explicitly, while the bilinear form $\mathcal{K}$ is treated implicitly.

Let $\left((\boldsymbol{q}^{(k)}, w_{\mathsf{D}}^{(k)})\colon k \in [Q_{\mathsf{D}}]\right)$ and $\left((\boldsymbol{x}^{(k)}, w_{\Omega}^{(k)})\colon k \in [Q_{\Omega}]\right)$ be $1/\mathsf{M}$- and 1-weighted quadrature rules on D and Ω, respectively. We then approximate (1.30a) by performing numerical integration over the configuration space, which results in

$$\begin{aligned}
&\sum_{k=1}^{Q_{\mathsf{D}}} w_{\mathsf{D}}^{(k)} \int_\Omega \frac{\psi^{n+1/2}(\cdot,\boldsymbol{q}^{(k)}) - \psi^n(\cdot,\boldsymbol{q}^{(k)})}{\Delta t/2} \varphi(\cdot,\boldsymbol{q}^{(k)}) \\
&\qquad + \sum_{k=1}^{Q_{\mathsf{D}}} w_{\mathsf{D}}^{(k)} \int_\Omega \operatorname{div}_{\boldsymbol{x}} \left(\boldsymbol{u}(\cdot,t^{n+1})\psi^{n+1/2}(\cdot,\boldsymbol{q}^{(k)})\right) \varphi(\cdot,\boldsymbol{q}^{(k)}) \\
&\qquad + \sum_{k=1}^{Q_{\mathsf{D}}} w_{\mathsf{D}}^{(k)} \frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)} \int_\Omega \nabla_{\boldsymbol{x}} \psi^{n+1/2}(\cdot,\boldsymbol{q}^{(k)}) \cdot \nabla_{\boldsymbol{x}} \varphi(\cdot,\boldsymbol{q}^{(k)}) \\
&\approx \sum_{k=1}^{Q_{\mathsf{D}}} w_{\mathsf{D}}^{(k)} \int_\Omega \sum_{i=1}^{N} \mathsf{M}(\boldsymbol{q}^{(k)}) (\nabla_{\boldsymbol{x}} \boldsymbol{u}(\cdot,t^n)) \boldsymbol{q}_i^{(k)} \psi^n(\cdot,\boldsymbol{q}^{(k)}) \cdot \nabla_{\boldsymbol{q}_i} \left(\frac{\varphi}{\mathsf{M}}\right)\Big|_{(\cdot,\boldsymbol{q}^{(k)})} \\
&\qquad - \sum_{k=1}^{Q_{\mathsf{D}}} w_{\mathsf{D}}^{(k)} \frac{1}{4\mathrm{Wi}} \int_\Omega \sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij} \mathsf{M}(\boldsymbol{q}^{(k)}) \nabla_{\boldsymbol{q}_j} \left(\frac{\psi^n}{\mathsf{M}}\right)\Big|_{(\cdot,\boldsymbol{q}^{(k)})} \cdot \nabla_{\boldsymbol{q}_i} \left(\frac{\varphi}{\mathsf{M}}\right)\Big|_{(\cdot,\boldsymbol{q}^{(k)})},
\end{aligned}$$

for all $\varphi = \varphi(\boldsymbol{x},\boldsymbol{q})$ in a suitable function space. Here, the symbol $\approx$ denotes equality up to quadrature errors. By selecting Q_{D} linearly independent functions $\zeta_{(m)}$, $m \in [Q_{\mathsf{D}}]$, of $\boldsymbol{q} \in \mathsf{D}$ such that $\zeta_{(m)}(\boldsymbol{q}^{(k)}) = \delta_{km}$, $k,m \in [Q_{\mathsf{D}}]$, and taking successively $\varphi = \varphi_{(m)}$, where

$\varphi_{(m)}(\boldsymbol{x},\boldsymbol{q}) := \chi(\boldsymbol{x})\zeta_{(m)}(\boldsymbol{q})$, in the equality above, we obtain a total of Q_{D} independent variational problems, each posed over the d-dimensional domain Ω, of the form:

$$\begin{aligned}
&\frac{1}{\Delta t/2}\int_\Omega \psi^{n+1/2}(\cdot,\boldsymbol{q}^{(m)})\chi + \int_\Omega \operatorname{div}_{\boldsymbol{x}}\left(\boldsymbol{u}(\cdot,t^{n+1})\psi^{n+1/2}(\cdot,\boldsymbol{q}^{(m)})\right)\chi \\
&\qquad\qquad + \frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\int_\Omega \nabla_{\boldsymbol{x}}\psi^{n+1/2}(\cdot,\boldsymbol{q}^{(m)})\cdot\nabla_{\boldsymbol{x}}\chi \\
&\approx \frac{1}{\Delta t/2}\int_\Omega \psi^{n}(\cdot,\boldsymbol{q}^{(m)})\chi \\
&\quad + \frac{1}{w_{\mathsf{D}}^{(m)}}\sum_{k=1}^{Q_{\mathsf{D}}} w_{\mathsf{D}}^{(k)}\Bigg[\int_\Omega \sum_{i=1}^{N}\mathsf{M}(\boldsymbol{q}^{(k)})(\nabla_{\boldsymbol{x}}\boldsymbol{u}(\cdot,t^n))\boldsymbol{q}_i^{(k)}\psi^n(\cdot,\boldsymbol{q}^{(k)})\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\zeta_{(m)}}{\mathsf{M}}\right)\Bigg|_{(\cdot,\boldsymbol{q}^{(k)})}\chi \\
&\qquad - \frac{1}{4\mathrm{Wi}}\int_\Omega\sum_{i=1}^{N}\sum_{j=1}^{N}A_{ij}\mathsf{M}(\boldsymbol{q}^{(k)})\nabla_{\boldsymbol{q}_j}\left(\frac{\psi^n}{\mathsf{M}}\right)\Bigg|_{(\cdot,\boldsymbol{q}^{(k)})}\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\zeta_{(m)}}{\mathsf{M}}\right)\Bigg|_{(\cdot,\boldsymbol{q}^{(k)})}\chi\Bigg] \\
&\qquad\qquad =: \mathfrak{M}_{(m)}(\psi^n;\chi) \qquad \forall\, m\in[Q_{\mathsf{D}}], \quad (1.31)
\end{aligned}$$

for all $\chi = \chi(\boldsymbol{x})$ in a suitable function space, where each $\mathfrak{M}_{(m)}(\psi^n;\cdot)$, $m\in[Q_{\mathsf{D}}]$, is a linear functional. Thus, (1.31) amounts to solving Q_{D} mutually independent linear convection-diffusion problems over Ω. We refer to these well-studied problems as the *spatial part* of (the weak form of) the Fokker–Planck equation.

In turn, we can approximate (1.30b) by performing numerical quadrature over Ω, resulting in

$$\begin{aligned}
&\sum_{k=1}^{Q_\Omega} w_\Omega^{(k)}\int_{\mathsf{D}}\frac{\psi^{n+1}(\boldsymbol{x}^{(k)},\cdot)-\psi^{n+1/2}(\boldsymbol{x}^{(k)},\cdot)}{\Delta t/2}\frac{\varphi(\boldsymbol{x}^{(k)},\cdot)}{\mathsf{M}} \\
&\qquad - \sum_{k=1}^{Q_\Omega} w_\Omega^{(k)}\int_{\mathsf{D}}\sum_{i=1}^{N}(\nabla_{\boldsymbol{x}}\boldsymbol{u}(\boldsymbol{x}^{(k)},t^n))\boldsymbol{q}_i\psi^n(\boldsymbol{x}^{(k)},\cdot)\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\varphi(\boldsymbol{x}^{(k)},\cdot)}{\mathsf{M}}\right) \\
&\quad + \sum_{k=1}^{Q_\Omega} w_\Omega^{(k)}\frac{1}{4\mathrm{Wi}}\int_{\mathsf{D}}\sum_{i=1}^{N}\sum_{j=1}^{N}A_{ij}\mathsf{M}\nabla_{\boldsymbol{q}_j}\left(\frac{\psi^{n+1}(\boldsymbol{x}^{(k)},\cdot)}{\mathsf{M}}\right)\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\varphi(\boldsymbol{x}^{(k)},\cdot)}{\mathsf{M}}\right) \\
&\qquad \approx -\sum_{k=1}^{Q_\Omega} w_\Omega^{(k)}\int_{\mathsf{D}}\operatorname{div}_{\boldsymbol{x}}\left(\boldsymbol{u}(\cdot,t^{n+1})\psi^{n+1/2}\right)\Big|_{(\boldsymbol{x}^{(k)},\cdot)}\varphi(\boldsymbol{x}^{(k)},\cdot)\frac{1}{\mathsf{M}} \\
&\qquad\qquad - \sum_{k=1}^{Q_\Omega} w_\Omega^{(k)}\frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\int_{\mathsf{D}}\nabla_{\boldsymbol{x}}\psi^{n+1/2}\Big|_{(\boldsymbol{x}^{(k)},\cdot)}\cdot\nabla_{\boldsymbol{x}}\varphi|_{(\boldsymbol{x}^{(k)},\cdot)}\frac{1}{\mathsf{M}},
\end{aligned}$$

for all $\varphi = \varphi(\boldsymbol{x},\boldsymbol{q})$ in a suitable function space. By selecting Q_Ω linearly independent functions $\chi_{(m)}$, $m\in[Q_\Omega]$, of $\boldsymbol{x}\in\Omega$ such that $\chi_{(m)}(\boldsymbol{x}^{(k)}) = \delta_{km}$, $k,m\in[Q_\Omega]$, and taking successively $\varphi = \varphi_{(m)}$, where $\varphi_{(m)}(\boldsymbol{x},\boldsymbol{q}) := \chi_{(m)}\zeta(\boldsymbol{q})$, in the equality above, we obtain a total of Q_Ω

independent variational problems over the Nd-dimensional domain D of the form:

$$\begin{aligned}
&\frac{1}{\Delta t/2}\int_{\mathsf{D}}\psi^{n+1}(\boldsymbol{x}^{(m)},\cdot)\frac{\zeta}{\mathsf{M}}+\frac{1}{4\mathrm{Wi}}\int_{\mathsf{D}}\sum_{i=1}^{N}\sum_{j=1}^{N}A_{ij}\mathsf{M}\nabla_{\boldsymbol{q}_j}\left(\frac{\psi^{n+1}(\boldsymbol{x}^{(m)},\cdot)}{\mathsf{M}}\right)\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\zeta}{\mathsf{M}}\right)\\
&\quad\approx\Bigg[\frac{1}{\Delta t/2}\int_{\mathsf{D}}\psi^{n+1/2}(\boldsymbol{x}^{(m)},\cdot)\frac{\zeta}{\mathsf{M}}+\int_{\mathsf{D}}\sum_{i=1}^{N}(\nabla_{\boldsymbol{x}}\boldsymbol{u}(\boldsymbol{x}^{(m)},t^n))\boldsymbol{q}_i\psi^{n}(\boldsymbol{x}^{(m)},\cdot)\cdot\nabla_{\boldsymbol{q}_i}\left(\frac{\zeta}{\mathsf{M}}\right)\\
&\qquad\qquad-\int_{\mathsf{D}}\operatorname{div}_{\boldsymbol{x}}\left(\boldsymbol{u}(\cdot,t^{n+1})\psi^{n+1/2}\right)\Big|_{(\boldsymbol{x}^{(m)},\cdot)}\frac{\zeta}{\mathsf{M}}\\
&\qquad-\frac{1}{w_\Omega^{(m)}}\sum_{k=1}^{Q_\Omega}w_\Omega^{(k)}\frac{(l_0/L_0)^2}{4\mathrm{Wi}(N+1)}\int_{\mathsf{D}}\nabla_{\boldsymbol{x}}\psi^{n+1/2}\Big|_{(\boldsymbol{x}^{(k)},\cdot)}\cdot\nabla_{\boldsymbol{x}}\chi_{(m)}\Big|_{(\boldsymbol{x}^{(k)},\cdot)}\frac{\zeta}{\mathsf{M}}\Bigg]\\
&\qquad\qquad=:\mathfrak{N}_{(m)}(\psi^{n+1/2};\zeta)\qquad\forall\, m\in[Q_\Omega],\quad(1.32)
\end{aligned}$$

for all $\zeta=\zeta(\boldsymbol{q})$ in a suitable function space, where each $\mathfrak{N}_{(m)}(\psi^{n+1/2};\cdot)$, $m\in[Q_\Omega]$, is a linear functional. Thus, (1.32) amounts to solving $[Q_\Omega]$ mutually independent linear elliptic variational problems, each posed on the high-dimensional configuration domain $\mathsf{D}=D_1\times\cdots\times D_N\subset\mathbb{R}^{Nd}$. For that reason we refer to them as the *configurational part* of (the weak form of) the Fokker–Planck equation.

It is the approximate solution by greedy algorithms of problems such as those described in (1.32) that this work is concerned with.

Remark 1.3. For a complete analysis of alternating direction schemes such as the one derived above we refer to [**Kne08**, Section 3.3], where the corresponding scheme is analyzed in the dumbbell case $N=1$.

1.2. Literature

The problems described in (1.32), despite being independent of the spatial variable and elliptic in the configuration variable, are hard enough to be of interest on their own. However, from the point of view of engineering and scientific applications, there is much interest on the more demanding task of approximating the situation in which the Navier–Stokes (1.1) system feeds the velocity $\boldsymbol{u}$ into the full Fokker–Planck equation (1.22a) and, conversely, the probability density function ψ of the latter takes part in driving the Navier–Stokes, via the Kramers expression (1.15) and the stress description (1.2). What results is a fully coupled Navier–Stokes–Kramers–Fokker–Planck system, which is posed on the Cartesian product of a time domain $(0,T_{\mathrm{end}}]$, a spatial domain $\Omega\subset\mathbb{R}^d$ and the high-dimensional configuration domain $\mathsf{D}\subset\mathbb{R}^{Nd}$.

Needless to say, the fully-coupled problem is computationally very demanding. Consequently, it has only been recently that practitioners have ventured into the numerical approximation of time-dependent, non-homogeneous micro-macro models of dilute polymers in complex geometries. And still, the vast majority of the literature on the subject concerns

itself with a number of simplified systems where one or more of the simplifying assumptions of steady state, globally homogeneous flow, no convection in space, have been made.

1.2.1. Stochastic approximation. The polymer model under consideration is originally posed as a stochastic differential equation (in our setting, (1.4) would be the starting point). Therefore, the use of Monte Carlo methods to approximate the evolution of the microscopic configuration of the model molecules is very natural. The polymeric extra stress tensor $\boldsymbol{\tau}$, which couples the microscopic model with the macroscopic flow equations (1.1) and (1.2), must be computed by averaging a functional akin to the Kramers expression (1.9) over many realizations of the microscopic equations (see [**OP02**, Section 11.4] or [**Keu04**, Section 4]).

Laso and Öttinger [**LÖ93**], with their *CONNFFESSIT* (acronym for "Calculation of Non-Newtonian Flow: Finite Elements and Stochastic Simulation Technique") scheme, pioneered the coupling of a stochastic form of the dynamic equations for a polymer model and macroscopic flow equations; for the latter, they used standard Finite Element methods. As it is to be expected, the stochastic error in the computation of the polymeric extra stress tensor decays at a rate of the form $C\,k^{-1/2}$, where k is the number of computed trajectories, with the constant C being proportional to the square root of the variance of the configuration vectors $\boldsymbol{q}_i$. Because of this, a number of variance reduction techniques have been proposed (see, for example, [**MÖ96, ÖvdBH97, JBL04**]).

Another important improvement on the original CONNFFESSIT approach is the use of Brownian configuration fields instead of ensembles of particles [**HvHvdB97, ÖvdBH97**]. On top of reducing the variance, this procedure allows for constructing approximations to the polymeric extra stress tensor $\boldsymbol{\tau}$ that are differentiable without post-processing. The importance of this becomes apparent on noticing that, as expressed in (1.1a) and (1.2), it is the divergence of $\boldsymbol{\tau}$ which drives the macroscopic flow.

We close our discussion of stochastic approximation methods by pointing out that these methods scale well with respect to the number of degrees of freedom of the molecular model, which make them the best option for chain models with a large number of springs. Now, the $\mathcal{O}(k^{-1/2})$ convergence rate cannot be improved upon with purely stochastic approaches. This is what motivates the switch to the Fokker–Planck equation (1.22a), which is deterministic in nature, and might thus be approximated using methods with faster convergence rates.

1.2.2. Deterministic Fokker–Planck approximation. Below, we mention some recent contributions that aim for the approximation of the Fokker–Planck equation ((1.12) itself or closely related ones) using deterministic methods.

C. Chauvière, A. Lozinski and collaborators have used spectral methods extensively for the configurational part of the Fokker–Planck equation. For example, in [**Loz03**] Lozinski considers both Hookean and FENE dumbbells in a dilute solution and approximates the configurational part using, as mentioned, spectral methods, and the spatial part using the

spectral element method of Chauvière and Owens [**CO01**]. He also studies and simulates some micro-macro models of non-diluted (concentrated or melt) polymers. There are also direct comparisons between deterministic Fokker–Planck approximation schemes and their stochastic counterparts. Besides that, it is here where we see a comparison between micro-macro simulations of Hookean dumbbells and the purely macroscopic simulation of its closure, namely, the Oldroyd-B fluid model (see [**BCAH87**] for the derivation of the latter from the former). It is seen that the micro-macro approach is the more efficient and robust of the two. See also [**LC03**]. The singularity of the FENE potential at the right end of its domain compels Lozinski and collaborators to search for a normalized version of ψ with a factor of the form

$$(1 - |\boldsymbol{q}|^2/b)^{-s} \tag{1.33}$$

where s is a parameter chosen on computational grounds (see [**CL04a**], [**CL04b**]). It is interesting to remark that in some of their work they use a three-dimensional configuration space while assuming a two-dimensional spatial domain, which is another manifestation of the modeling flexibility of the micro-macro approach.

In [**Kne08**] D. Knezevic gives a detailed and rigorous treatment of the full Fokker–Planck system arising from dilute solutions of polymers modeled as FENE dumbbells and goes on briefly to consider the coupled Navier–Stokes–Kramers–Fokker–Planck system. To cope with the unbounded drift in the Fokker–Planck equation he performs a Kolmogorov symmetrization on it which has the result of replacing the abovementioned unbounded drift term with a weighted diffusion term, which is the approach we adopted as well (cf. (1.26)). Then he goes on to discretize the configurational part of this symmetrized equation using spectral methods. After that, he carefully justifies the operator splitting procedure in a semi-discrete setting and subsequently introduces a finite element method to approximate the spatial part combined with a quadrature-based scheme for the configurational part and therefore he engages the full Fokker–Planck equation (1.12). He considers the choice of function spaces and the approximation properties of the schemes involved. Also, he explains the natural parallelism of his approach and exhibits numerical results obtained by exploiting this trait. Finally, he considers the full Navier–Stokes–Fokker–Planck system. He extends many of his results based on the Kolmogorov symmetrization to the normalized formulation of Chauvière and Lozinski mentioned in the previous paragraph. See also [**KS09a**] and [**KS09b**].

In [**DLY05**] Q. Du, C. Liu and P. Yu suggest a finite difference scheme for (1.12) in the dumbbell case. This scheme preserves the positivity of ψ and its integral with respect to the configuration space. In order to deal with the singular nature of the FENE potential these authors also use a normalization factor which is a special case of the one used in the abovementioned contribution by Chauvière and Lozinski. When it comes to their numerical scheme they discretize it up to some radius smaller than $\sqrt{b}$ and impose a no-flux boundary condition at the resulting artificial boundary. They warn that for large velocity gradients

the probability density function ψ tends to concentrate on parts of the boundary of the configuration domain, so they turn to Monte Carlo methods in that case.

In [**HO06**] C. Helzel and F. Otto simulated a Stokes–Fokker–Planck system for rod-like polymers (Doi model, see [**BCAH87**]) using finite volume and finite difference methods in two spatial dimensions and two configurational dimensions, respectively, plus time. Their choice of kinetic theory model does not exhibit a singular behavior as the FENE Rouse model, and in particular the FENE dumbbell, do. However they consider dilute solutions, concentrated solutions and melts, and perform stability analyzes that, among other things, indicate whether the assumption of stationarity is reasonable for some flow regimes.

In [**Nay98**] R. Nayak solved the Fokker–Planck equation arising from the Doi model in complex flow conditions using a combination of Daubechies wavelet-based Galerkin method for the configurational part and a discontinuous Galerkin method for the spatial part. Apart from performing a stability analysis she proceeds to couple her combined solver of the Fokker–Planck equation with a finite element scheme for the unsteady Navier–Stokes equations. She also took advantage of the highly parallel nature of the resulting combined schemes.

The important issue of proving that a given method does indeed converge to the true solution is difficult to tackle in the case of the fully coupled Navier–Stokes–Kramers–Fokker–Planck systems; indeed, it was only recently that a proof of the well-posedness of that system (and hence, the existence of a unique solution to approximate) was given in [**BS11a**]. After that, in [**BS11b**], J. Barrett and E. Süli proved the convergence of a Finite Element approximation scheme for the Navier–Stokes–Kramers–Fokker–Planck system. Previously, such a proof existed for the *corotational* case [**BS09**]; that is, the case in which $\nabla_{\boldsymbol{x}}\boldsymbol{u} + (\nabla_{\boldsymbol{x}}\boldsymbol{u})^{\mathrm{T}} = 0$.

1.2.3. Simulation of the Rouse model with $N > 1$. The literature discussed above concerned itself with low-dimensional configuration domains such as d-dimensional balls or their boundaries, $d \in \{2, 3\}$. Here we will mention two approaches that have been proposed in the case of configuration domains of higher dimension.

In [**DLO07**] P. Delaunay, A. Lozinski and R.G. Owens proposed the use of a sparse tensor product basis for a Galerkin method for the solution of the configurational part of the Fokker–Planck equation in two spatial dimensions. They use the first N_l eigenfunctions of the dumbbell ($N = 1$) version of a modified (in order to deal with the symmetries and boundary conditions of the problem) configuration-space Fokker–Planck operator. Labeling the eigenpairs as $(\lambda_k, \tilde{\varphi}_k)$, where the real parts of the eigenvalues λ_k are monotonically increasing with respect to k, they construct the sparse tensor product basis as $\{\tilde{\varphi}_{i_1}(\boldsymbol{q}_1)\otimes\cdots\otimes\tilde{\varphi}_{i_N}(\boldsymbol{q}_N) \colon \boldsymbol{i} \in \boldsymbol{J}\}$. The set of multi-indices $\boldsymbol{J}$ is a subset of the full set $[N_l]^N$ of possible multi-indices chosen so as to exclude the products of highly oscillating eigenfunctions. N_l is not arbitrary, for it depends on a level of resolution l in a way that respects an introduced hierarchy of the $\tilde{\varphi}_i$, much in the way that standard finite element sparse grid methods [**BG04**] enrich their univariate basis with all basis functions with the same oscillation rate at each step. The authors report

that their method is competitive for $N \leq 3$ and moderate Weissenberg numbers ($\mathrm{Wi} \approx 1$). Interesting as this is, we won't comment further on this approach.

Another approach, which will be the focus of this work, was recently proposed in the engineering literature in a succession of papers by Ammar, Mokdad, Chinesta, Keunings and collaborators [**AMCK06, AMCK07, AND$^+$10, GACC10, CALK11**] under the names *Separated Representation* and *Proper Generalized Decomposition*. A variant with a discretization based on spectral methods instead of the finite element methods preferred by Ammar et al. was presented by Leonenko and Phillips [**LP09**]. A similar method was considered independently by Nouy [**Nou07, Nou08**] and Nouy & Le Maître [**NLM09**] under the name *Power type Generalized Spectral Decomposition*, for the numerical solution of stochastic partial differential equations. However, the historical roots of the technique can be traced back to the work of Schmidt [**Sch07**]. Ammar et al. and Nouy report that the algorithm performs well in numerical experiments and comment that it extends to a large variety of partial differential equations.

In the simplified mathematical setting of Poisson's equation $-\Delta u = f$ posed on the rectangular domain $\Omega = \Omega_x \times \Omega_y$, where Ω_x and Ω_y are bounded open subintervals of $\mathbb{R}$, subject to a homogeneous Dirichlet boundary condition on $\partial\Omega$, the convergence of a modified Separated Representation algorithm was shown in a recent paper by Le Bris, Lelièvre and Maday [**LBLM09**], by drawing on connections with greedy algorithms from nonlinear approximation theory (cf. DeVore and Temlyakov [**DT96**]). In [**LBLM09**], the solution was represented as a sum

$$u(x,y) = \sum_{n\geq 1} r_n(x)\, s_n(y) \tag{1.34}$$

by iteratively determining functions $x \in \Omega_x \mapsto r_n(x)$ and $y \in \Omega_y \mapsto s_n(y)$, $n \geq 1$, such that for all n, the product $(x,y) \in \Omega \mapsto r_n(x)\, s_n(y)$ is the best approximation in the norm of the Sobolev space $\mathrm{H}^1_0(\Omega)$ to the solution $(x,y) \in \Omega \mapsto v(x,y)$ of the Poisson equation

$$-\Delta v(x,y) = f(x,y) + \Delta \left(\sum_{k\leq n-1} r_k(x) s_k(y) \right),$$

subject to a homogeneous Dirichlet boundary condition, in terms of a single tensor product $r(x)\, s(y)$; Le Bris, Lelièvre and Maday thus show that it is possible to give a sound mathematical basis to the algorithm proposed by Ammar et al., provided that one considers a variational form of the approach that manipulates global minimizers of Dirichlet energies instead of stationary points to the associated Euler–Lagrange equations (in the follow-up paper [**CEL11**] by Cancès, Ehrlacher and Lelièvre it was further shown that one can also work with local—yet still energy-decreasing—minimizers provided that one keeps within the two-fold tensor product setting of (1.34)). In order to reformulate the approach in such a variational setting, the arguments in [**LBLM09**] crucially rely on the fact that the Laplace operator is self-adjoint,

and as noted by the authors of [**LBLM09**], the analysis does not apply exactly to the actual implementation of the method as described in the papers by Ammar et al., where stationary points of the Euler–Lagrange equations associated with the Dirichlet energies are computed instead—hence the need of modifying the algorithm proposed by Ammar et al. Indeed, since global minimizers of Dirichlet energies in the approach of Le Bris, Lelièvre and Maday on the one hand and stationary points of the associated Euler–Lagrange equations in the approach of Ammar et al. on the other are each sought in *nonlinear* manifolds embedded in a Sobolev space, rather than over the entire Sobolev space (which is a normed *linear* space), the two approaches are not equivalent. The authors of [**LBLM09**] also comment that: "Likewise, it is unclear to us how to provide a mathematical foundation of the approach for nonvariational situations, such as an equation involving a differential operator that is not self-adjoint." This latter remark is particularly pertinent in the context of Fokker–Planck equations for kinetic bead-spring chain models for dilute polymers, of the kind considered by Ammar et al., where the differential operator in configuration space featuring in the Fokker–Planck equation, a generalized Ornstein–Uhlenbeck operator, is a non-self-adjoint elliptic operator with a drift term that involves an unbounded potential.

It is this last point that the present work is aimed at addressing.

1.3. Objectives and structure of this work

We perform a nonlinear approximation of the analytical solution $\psi\colon (\boldsymbol{q}_1, \dots, \boldsymbol{q}_N) \in \mathsf{D} \mapsto \psi(\boldsymbol{q}_1, \dots, \boldsymbol{q}_N)$ to the high-dimensional degenerate[1] elliptic boundary-value problem (1.32) on $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ by Separated Representations of the form

$$\sum_{k=1}^{K} \prod_{i=1}^{N} \psi_k^{(i)}(\boldsymbol{q}_i),$$

where the factors $\psi_k^{(i)}$, $k \in [K]$, are defined on the d-dimensional domain D_i, $i \in [N]$. Instead of being selected from an *a priori* fixed set, the factors $\psi_k^{(i)}$, $i \in [N]$, are obtained, N at a time, for each $k \in [K]$, as the best approximation (in a sense to be made precise in Subsection 3.1.1) among all possible such factors. The (potentially large) number of terms K is likewise not fixed in advance, but is the outcome of a termination criterion.

The rest of this work is organized as follows: First, we close this chapter by introducing some notation and other conventions in Section 1.4.

In Chapter 2 we focus on the functional-analytic context of our analysis of the Separated Representation strategy. In Section 2.1 we give a variational formulation of the elliptic configuration Fokker–Planck equation (1.32) (Subsection 2.1.1) and present some basic results

[1] In this work we use *degenerate* to refer to the fact that the problem involves unbounded drifts; it should not be construed as to imply any loss (in the right functional-analytic setting—cf. Subsection 2.1.1) of the elliptic character of the problem.

on Maxwellian-weighted Sobolev spaces (Subsection 2.1.2). In Section 2.2 we obtain compact embedding and density of compactly-supported smooth functions results for both the force laws whose potential we know explicitly (Subsection 2.2.1) and for the Inverse Langevin force law (Subsection 2.2.2) and then study the asymptotic behavior of the eigenvalue problems associated with the Maxwellian-weighted Sobolev spaces (Subsection 2.2.3). In Section 2.3 we introduce and develop the notion of bounded Lipschitz domain (Subsection 2.3.1), and show constructively that this property is preserved by the Cartesian product operation—in an simplified setting first (Subsection 2.3.2), and in its full form later (Subsection 2.3.3)—; then we show that the obtained construction can be used to preserve under tensor products certain decay-near-the boundary property of weight functions defined on bounded Lipschitz domains (Subsection 2.3.4). In Section 2.4 we prove a number of basic results concerning tensor products of functions in weighted Sobolev spaces (Subsection 2.4.1), prove a nice result on the tensorization of a compact embedding property (Subsection 2.4.2) and use the construction of the previous section to prove the density of smooth functions in a class of weighted Sobolev spaces with Cartesian-product-induced corners (Subsection 2.4.3).

Chapter 3 concerns itself with a formulation of the Separated Representation strategy in a continuous setting. In Section 3.1 two abstract algorithms are introduced for our degenerate setting (Subsection 3.1.1) and their correctness is proved (Subsection 3.1.2). Then, in Section 3.2 we obtain a number of auxiliary results (Subsection 3.2.1) which are then used to prove the convergence of the abstract algorithms (Subsection 3.2.2) and to show that they are greedy algorithms in the sense of nonlinear approximation theory (Subsection 3.2.3); subsequently, we introduce rates of convergence for the algorithms provided that the true solution to the Fokker–Planck equation under consideration is a member of certain abstract space (Subsection 3.2.4). At an abstract level, an important part of our arguments in the first two sections of this chapter follow those of [**LBLM09**]; however, the degenerate nature of the Fokker–Planck operator complicates considerably the verification of the basic results which feed into the abstract results. In Section 3.3 we seek to give characterizations of subspaces of this abstract space in more easily verifiable terms; first, we describe how the eigenfunctions and eigenvalues of a tensor-product Maxwellian-weighted linear operator relate to their counterparts with respect to the operators associated to the factor (or partial) Maxwellians (Subsection 3.3.1). With this information we can describe subspaces of the space of guaranteed rates of convergence for the abstract greedy algorithms in terms of weighted summability of the Fourier expansions of their members (Subsection 3.3.2). Then, using delicate elliptic regularity results for the associated Ornstein–Uhlenbeck operators, we present subspaces that are defined in terms of weighted-Sobolev regularity (Subsection 3.3.3)

In Chapter 4 we study variants of the abstract greedy algorithms of the previous chapter in which the search manifolds are tensor products of finite-dimensional linear spaces, while

still measuring the error in the original norm. In Section 4.1, after introducing some notation (Subsection 4.1.1) we present what we call discrete greedy algorithms (Subsection 4.1.2) and prove a number of their properties, including correctness, convergence, identification as greedy algorithms in the sense of nonlinear approximation theory and convergence rates (Subsection 4.1.3). This convergence rates, however, can only be justified in the very improbable case of the true solution lying in the precise finite dimensional space of tensor products in which the iterates generated by the discrete greedy algorithms live. Therefore, we go on to describe some theoretical arguments which allow for guaranteeing convergence rates for solutions in certain spaces which are more general, but quite abstract (Subsection 4.1.4). Then, at some level paralleling the developments in the previous chapter, we show that if the discrete greedy algorithms are based on finite-dimensional subspaces based on eigenfunctions of the single-domain Ornstein–Uhlenbeck operators (Subsection 4.1.5) or on polynomials multiplied by the corresponding Maxwellian function (Subsection 4.1.6), familiarly-defined subspaces of the abstract spaces of guaranteed rates of convergence can be described explicitly. Then, in Section 4.2 we describe the gap between the discrete greedy algorithms and practical implementations of the Separated Representation strategy (Subsection 4.2.1), describe a simple inner iteration to help bridge this gap (Subsection 4.2.2) and then illustrate the behavior of the resulting procedure on a very simple numerical example (Subsection 4.2.4).

In Chapter 5 we give some conclusions and describe some potential avenues of further work.

At last, Appendix A, contains some auxiliary results on distributions and variational eigenvalue problems.

1.4. On notation and other conventions

We denote by $[k]$ the integer interval $\{i \in \mathbb{N} \colon 1 \leq i \leq k\}$. We shall denote sequences and arrangements of elements a_i indexed by indices i in an index set $\mathcal{I}$ by $(a_i \colon i \in \mathcal{I})$.

We shall write $\boldsymbol{q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N) \in D_1 \times \cdots \times D_N = \times_{i\in[N]} D_i =: \mathsf{D}$. Given N real-valued functions f_i, each defined on the corresponding set D_i, we denote by $\bigotimes_{i\in[N]} f_i$ their *tensor product*; i.e., the function

$$\boldsymbol{q} \in \mathsf{D} \mapsto \prod_{i\in[N]} f_i(\boldsymbol{q}_i).$$

We extend this notation in three ways. Firstly, as the tensor-product operation is order-dependent, we will use subscripts on the $\otimes$ and the $\bigotimes$ signs to denote where on $\boldsymbol{q} \in \mathsf{D}$ the function, or functions, following them act; e.g., $\bigotimes_{i\in[N]\setminus\{j\}} f_i \otimes_j f_j$ evaluated on $\boldsymbol{q} \in \mathsf{D}$ is

$$\prod_{i\in[N]\setminus\{j\}} f_i(\boldsymbol{q}_i)\, f_j(\boldsymbol{q}_j).$$

Secondly, we will use the same notation for the sets resulting from the tensor products of members of function spaces: suppose that F_i is a nonempty set of real-valued functions

defined on D_i, $i \in [N]$; we then write $\bigotimes_{i\in[N]} F_i := \{\bigotimes_{i\in[N]} f_i \colon f_i \in F_i,\ i \in [N]\}$. Thirdly, if exactly one of the factors is vector-valued, the products involving it at the time of evaluation must be interpreted as scalar-vector products implying that the resulting tensor product will be vector-valued too.

It is worth noting that there exists another notion of tensor product of Hilbert spaces, whose output is another Hilbert space (see, for example, [**RS80**, Section 2.4]). We will not make use of this notion of tensor-product in the present work.

The symbol $\Subset$ will stand for the compact embedding relation. The support of a real-valued function f will be denoted by $\operatorname{supp}(f)$.

Given a nonempty open set $E \subset \mathbb{R}^n$ and $k \in \mathbb{N}_0$ we denote by $\mathrm{C}^k(E)$ the space of real-valued functions defined on E whose partial derivatives up to order k are continuous. Also,

$$\mathrm{C}(E) := \mathrm{C}^0(E), \quad \mathrm{C}^\infty(E) := \cap_{k=0}^\infty \mathrm{C}^k(E),$$
$$\mathrm{C}^k_0(E) = \{f \in \mathrm{C}^k(E) \colon \operatorname{supp}(f) \Subset E\}, \quad k \in \mathbb{N}_0 \cup \{\infty\}.$$

The sets $\mathrm{C}^k(\overline{E})$, for $k \in \mathbb{N}_0$, $\mathrm{C}(\overline{E})$ and $\mathrm{C}^\infty(\overline{E})$ are the analogues of the corresponding sets defined above, where instead of mere continuity, uniform continuity and boundedness are demanded.

Given a measurable and almost everywhere positive real-valued function w defined on an open set $E \subset \mathbb{R}^n$; i.e., a *weight*, we denote by $\mathrm{L}^2_w(E)$ the Lebesgue space of square-integrable functions with respect to the weight w, equipped with its usual norm,

$$\|\varphi\|_{\mathrm{L}^2_w(E)} := \left(\int_E |\varphi|^2 \, w \right)^{1/2}.$$

We also define the w-weighted Sobolev spaces

$$\mathrm{H}^m_w(E) := \left\{ \varphi \in \mathrm{L}^2_w(E) \cap \mathrm{L}^1_{\mathrm{loc}}(E) \colon \partial_\alpha \varphi \in \mathrm{L}^2_w(E),\ |\alpha| \le m \right\},$$

where $|\alpha| := |\alpha|_1 = \sum_{i\in[n]} \alpha_i$, equipped with the norm defined by the expression

$$\|\varphi\|^2_{\mathrm{H}^m_w(E)} := \sum_{|\alpha|\le m} \|\partial_\alpha \varphi\|^2_{\mathrm{L}^2_w(E)}.$$

We will also find occasion to use the higher-mixed-derivatives, or (simply) 'mix' w-weighted Sobolev spaces

$$\mathrm{H}^{m,\mathrm{mix}}_w(E) := \left\{ \varphi \in \mathrm{L}^2_w(E) \cap \mathrm{L}^1_{\mathrm{loc}}(E) \colon \partial_\alpha \varphi \in \mathrm{L}^2_w(E),\ |\alpha|_\infty \le m \right\},$$

where $|\alpha|_\infty := \max_{i\in[n]} \alpha_i$, equipped with the norm defined by

$$\|\varphi\|^2_{\mathrm{H}^{m,\mathrm{mix}}_w(E)} := \sum_{|\alpha|_\infty\le m} \|\partial_\alpha \varphi\|^2_{\mathrm{L}^2_w(E)}.$$

As customary, we omit w from the subscript of the spaces described above in the case $w \equiv 1$. Given two weights w_1 and w_2 defined on an open set E we will say that they are *equivalent* if there exist positive constants c_1 and c_2 such that, almost everywhere in E,

$$c_1\, w_1 \leq w_2 \leq c_2\, w_1.$$

Later, we will introduce a number of hypotheses (the first of them is Hypothesis A in Subsection 2.1.1). The need for them is so recurrent in this work that we adopt as a convention that, once one of them is introduced, it is assumed to hold for the rest of the document.

We have expounded on the origin of the Fokker–Planck equation and its distillation to the configuration-space-only and elliptic form whose approximation we will consider. We also gave an overview of the literature on the numerical approximation of the Fokker–Planck equation.

In the following chapter we will focus on the functional-analytic aspects of the Fokker–Planck equation (1.32).

CHAPTER 2

Maxwellian-weighted Sobolev spaces

In this chapter we will discuss the functional-analytic setting of the Fokker–Planck equation (1.32) and prove a number of results on certain weighted Sobolev spaces which, it will transpire, are the natural setting for our problem. In Cartesian-product domains such as D some standard techniques for the analysis for weighted Sobolev spaces are not applicable because the product domain preserves a limited amount of regularity from its factor domains. We will find that the use of *ad hoc* methods tailored to weights that have the tensor-product structure is fruitful. Most of these results will find applications in Chapter 3 and Chapter 4, although some are shown exclusively because they are of interest on their own.

2.1. Function spaces

2.1.1. Variational formulation. The form of the problem (1.32) motivates us to consider the linear elliptic variational problem

$$a(\psi, \varphi) = f(\varphi), \tag{2.1}$$

posed on the high-dimensional configuration domain $\mathsf{D} = D_1 \times \cdots \times D_N \subset \mathbb{R}^{Nd}$, where

$$a(\psi, \varphi) := \frac{1}{4\mathrm{Wi}} \int_{\mathsf{D}} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \mathsf{M} \nabla_{\boldsymbol{q}_j} \left(\frac{\psi}{\mathsf{M}} \right) \cdot \nabla_{\boldsymbol{q}_i} \left(\frac{\varphi}{\mathsf{M}} \right) + c \int_{\mathsf{D}} \frac{\psi \varphi}{\mathsf{M}}, \tag{2.2}$$

the parameter c is positive and f is a linear functional. The natural function space associated with problem (2.1) is

$$\mathrm{H}(\mathsf{D}; \mathsf{M}) := \left\{ \varphi \in \mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D}) \cap \mathsf{M}\, \mathrm{L}^1_{\mathrm{loc}}(\mathsf{D}) \colon \nabla_{\boldsymbol{q}_i}(\varphi/\mathsf{M}) \in [\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})]^d \quad \forall\, i \in [N] \right\},$$

equipped with the norm

$$\|\varphi\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} := \left(\|\varphi\|^2_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} + \sum_{i=1}^{N} \left\| \nabla_{\boldsymbol{q}_i}(\varphi/\mathsf{M}) \right\|^2_{[\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})]^d} \right)^{1/2}.$$

An easy to prove manifestation of this naturality is the fact that, for all $\psi, \varphi \in \mathrm{H}(\mathsf{D}; \mathsf{M})$,

$$a(\psi, \varphi) \le \max\left(\frac{\lambda_{\max}}{4\mathrm{Wi}}, c \right) \|\psi\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} \|\varphi\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} \quad \text{and} \quad a(\varphi, \varphi) \ge \min\left(\frac{\lambda_{\min}}{4\mathrm{Wi}}, c \right) \|\varphi\|^2_{\mathrm{H}(\mathsf{D};\mathsf{M})}. \tag{2.3}$$

The spaces $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ and $\mathrm{H}(\mathsf{D};\mathsf{M})$ are isometrically isomorphic to, respectively, $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ and $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ via the relations

$$\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D}) = \mathsf{M}\,\mathrm{L}^2_{\mathsf{M}}(\mathsf{D}), \quad \|\cdot\|_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} = \big\|\mathsf{M}^{-1}\cdot\big\|_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})}, \tag{2.4a}$$

$$\mathrm{H}(\mathsf{D};\mathsf{M}) = \mathsf{M}\,\mathrm{H}^1_{\mathsf{M}}(\mathsf{D}), \quad \|\cdot\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} = \big\|\mathsf{M}^{-1}\cdot\big\|_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})}. \tag{2.4b}$$

Later, we will make use of the spaces $\mathrm{H}(D_i;M_i)$, $i\in[N]$, each of which is the i-th partial Maxwellian analogue of $\mathrm{H}(\mathsf{D};\mathsf{M})$. That is,

$$\mathrm{H}(D_i;M_i) := \Big\{\varphi\in \mathrm{L}^2_{1/M_i}(D_i)\cap M_i\,\mathrm{L}^1_{\mathrm{loc}}(D_i)\colon \nabla(\varphi/M_i)\in[\mathrm{L}^2_{M_i}(D_i)]^d\Big\},$$

where we have chosen to write $\nabla(\varphi/M_i)$ instead of $\nabla_{\boldsymbol{q}_i}(\varphi/M_i)$ because this is a more natural notation when considering $\mathrm{H}(D_i;M_i)$ in isolation. We equip $\mathrm{H}(D_i;M_i)$ with the norm

$$\|\varphi\|_{\mathrm{H}(D_i;M_i)} := \Big(\,\|\varphi\|^2_{\mathrm{L}^2_{1/M_i}(D_i)} + \|\nabla(\varphi/M_i)\|^2_{[\mathrm{L}^2_{M_i}(D_i)]^d}\,\Big)^{1/2}.$$

Remark 2.1.

(1) For $i\in[N]$, $\mathrm{H}(D_i;M_i)$ is exactly $\mathrm{H}(\mathsf{D};\mathsf{M})$ if $N=1$ and $\mathsf{M}=M_i$. None of the results involving $\mathrm{H}(\mathsf{D};\mathsf{M})$ appearing below depend on restrictions on N and thereby remain valid for $\mathrm{H}(D_i;M_i)$. Just like (2.4), $\varphi\mapsto M_i\varphi$ is an isometric isomorphism between $\mathrm{L}^2_{M_i}(D_i)$ and $\mathrm{L}^2_{1/M_i}(D_i)$ and between $\mathrm{H}^1_{M_i}(D_i)$ and $\mathrm{H}(D_i;M_i)$.

(2) The definitions above can be extended to open subsets of D and of the D_i, $i\in[N]$, in the usual way.

Before listing our structural hypotheses and proving the properties we need of $\mathrm{H}(\mathsf{D};\mathsf{M})$ we fully state the weak formulation of our model problem:

Given $f\in\mathrm{H}(\mathsf{D};\mathsf{M})'$, find $\psi\in\mathrm{H}(\mathsf{D};\mathsf{M})$ such that

$$a(\psi,\varphi) = f(\varphi) \qquad \forall\,\varphi\in\mathrm{H}(\mathsf{D};\mathsf{M}). \tag{2.5}$$

We adopt the following structural hypotheses.

Hypothesis A. *For each $i\in[N]$, the spring potential U_i belongs to $\mathrm{C}^1([0,\frac{b_i}{2}))$, where $b_i>0$, and satisfies $\lim_{s\to b_i/2_-}U(s)=\infty$.*

Immediate consequences of Hypothesis A are that $\mathsf{M}\in\mathrm{C}(\overline{\mathsf{D}})\cap\mathrm{C}^1(\mathsf{D})$ and that, for all $K\Subset\mathsf{D}$, there exist positive constants c_K and C_K such that $c_K\le\mathsf{M}(\boldsymbol{q})\le C_K$, for all $\boldsymbol{q}\in K$.

Hypothesis B. *For each $i\in[N]$, $\mathrm{H}^1_{M_i}(D_i)$ is compactly embedded in $\mathrm{L}^2_{M_i}(D_i)$.*

Remark 2.2. It is easy to check that springs obeying the FENE (1.16), CPAIL (1.17), TEAIL (1.18) and CP (1.21) force models comply with Hypothesis A. The corresponding fact for springs obeying the Inverse Langevin (1.19) force models is shown in Subsection 2.2.2.

In Step 1 of section A.1 of [**BS08**] it is proved that springs obeying the FENE model (1.16) satisfy Hypothesis B, under the condition $b_i \geq 2$. The corresponding result for springs obeying the CPAIL model with $b_i \geq 3$, the TEAIL model with $b_i \geq 16/5$, and CP model with $b_i \geq 3$ is shown in Lemma 2.6 in Subsection 2.2.1; the case of springs obeying the Inverse Langevin force law is proved in Subsection 2.2.2.

2.1.2. Basic properties of Maxwellian-weighted Sobolev spaces.

Lemma 2.3. $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$, $\mathrm{H}^m_{\mathsf{M}}(\mathsf{D})$ *for* $m \in \mathbb{N}$, $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ *and* $\mathrm{H}(\mathsf{D};\mathsf{M})$ *are separable Hilbert spaces.*

Proof. The operation $\varphi \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}) \mapsto \varphi/\sqrt{\mathsf{M}}$ defines an isometric isomorphism between $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ and $\mathrm{L}^2(\mathsf{D})$. Therefore the first space inherits its separability from the latter. On noting that $\mathsf{M}^{-1} \in \mathrm{L}^1_{\mathrm{loc}}(\mathsf{D})$, Theorem 1.11 of [**KO84**] guarantees the completeness of $\mathrm{H}^m_{\mathsf{M}}(\mathsf{D})$ (this source actually states the result for the case $m = 1$ only; however, the proof carries over to higher m in this single-weight case) and thus, $\mathrm{H}^m_{\mathsf{M}}(\mathsf{D})$ is separable by an argument along the lines of [**AF03**, ¶3.5]. The spaces $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ and $\mathrm{H}(\mathsf{D};\mathsf{M})$ inherit these properties via the isometric isomorphism (2.4). Finally, as their respective norms obey the parallelogram law, these spaces are Hilbert spaces. □

Lemma 2.4. *The following inclusions hold:*

(a) $\mathrm{C}^1_0(\mathsf{D}) \subset \mathrm{H}(\mathsf{D};\mathsf{M})$*;*
(b) $\mathrm{C}^1(\overline{\mathsf{D}}) \subset \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$*.*

Proof. Let $\varphi \in \mathrm{C}^1_0(\mathsf{D})$ and $K := \mathrm{supp}(\varphi) \Subset \mathsf{D}$. Then, the membership of φ into $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ follows from

$$\int_{\mathsf{D}} \varphi^2 \frac{1}{\mathsf{M}} = \int_K \varphi^2 \frac{1}{\mathsf{M}} \leq |K| \sup_{\boldsymbol{q}\in K} \frac{\varphi(\boldsymbol{q})^2}{\mathsf{M}(\boldsymbol{q})} < \infty,$$

which, in turn, stems from the fact that M is positively bounded from below on each compact subset of D. Similarly, for all $K' \Subset \mathsf{D}$,

$$\int_{K'} \left|\frac{\varphi}{\mathsf{M}}\right| \,\mathrm{d}\boldsymbol{q} \leq \left|K' \cap K\right| \sup_{\boldsymbol{q}\in K'\cap K} \frac{|\varphi(\boldsymbol{q})|}{\mathsf{M}(\boldsymbol{q})} < \infty$$

on account of which $\varphi \in \mathsf{M}\,\mathrm{L}^1_{\mathrm{loc}}(\mathsf{D})$. The latter implies that φ/M defines a regular distribution in the usual way. Then, for each $i \in [N]$, $\nabla_{\boldsymbol{q}_i}(\varphi/\mathsf{M})$ exists as a distribution and coincides with the classical i-th component gradient of φ/M, which belongs to $[\mathrm{C}(\mathsf{D})]^d$ because of Hypothesis A. Then,

$$\int_{\mathsf{D}} \left|\nabla_{\boldsymbol{q}_i}\Big(\frac{\varphi}{\mathsf{M}}\Big)\right|^2 \mathsf{M} \leq |K| \sup_{\boldsymbol{q}\in K} \left|\nabla_{\boldsymbol{q}_i}\Big(\frac{\varphi(\boldsymbol{q})}{\mathsf{M}(\boldsymbol{q})}\Big)\right|^2 \mathsf{M}(\boldsymbol{q}) < \infty$$

and that proves (a).

Part (b) is an immediate consequence of the membership of M in $\mathrm{L}^1(\mathsf{D})$. □

2.2. Properties of partial Maxwellian-weighted Sobolev spaces

2.2.1. Sobolev spaces weighted with Maxwellians in explicit form. We shall derive some key properties of the function spaces associated with the CPAIL force model (1.17), the TEAIL force model (1.18) and the CP force model (1.21) using the corresponding properties of the function spaces associated with the FENE force model (1.16). For this we will make use of the basic result that follows.

Proposition 2.5. *Suppose w_1 and w_2 are equivalent weights defined on an open set E. Then, for any $m \in \mathbb{N}$, $\mathrm{H}^m_{w_1}(E)$ and $\mathrm{H}^m_{w_2}(E)$ are algebraically and topologically the same space and so are $\mathrm{H}^{m,\mathrm{mix}}_{w_1}(E)$ and $\mathrm{H}^{m,\mathrm{mix}}_{w_2}(E)$.*

Proof. From the definition we gave of the equivalence of weights in Section 1.4, we know that there exist positive constants c_1 and c_2 such that, for any measurable function $g\colon E \to \mathbb{R}$, $c_1 \int_E g^2 w_1 \leq \int_E g^2 w_2 \leq c_2 \int_E g^2 w_1$. Thus, $\mathrm{L}^2_{w_1}(E)$ and $\mathrm{L}^2_{w_2}(E)$ are the same set and have, as normed spaces, equivalent norms. As, for $j \in \{1,2\}$, the $\mathrm{H}^m_{w_j}(E)$ and $\mathrm{H}^{m,\mathrm{mix}}_{w_j}(E)$ spaces (resp. their norms) are defined in terms of membership (resp. the norms) of their derivatives in $\mathrm{L}^2_{w_j}(E)$, we obtain the desired result. □

Let $b \geq 3$. It follows from (1.16), (1.17) and (1.24) that the Maxwellian M_{C} associated to a spring obeying the CPAIL model with parameter b and the Maxwellian M_{F} associated to a spring obeying the FENE model with parameter $2b/3$ are, respectively,

$$M_{\mathrm{C}}(\boldsymbol{p}) = Z_{\mathrm{C}} \exp(-|\boldsymbol{p}|^2/6) \left(1 - \frac{|\boldsymbol{p}|^2}{b}\right)^{b/3}, \qquad \boldsymbol{p} \in D_{\mathrm{C}} = B\left(0, \sqrt{b}\right) \subset \mathbb{R}^d \tag{2.6}$$

and

$$M_{\mathrm{F}}(\boldsymbol{p}) = Z_{\mathrm{F}} \left(1 - \frac{|\boldsymbol{p}|^2}{2b/3}\right)^{b/3}, \qquad \boldsymbol{p} \in D_{\mathrm{F}} = B\left(0, \sqrt{2b/3}\right) \subset \mathbb{R}^d,$$

where Z_{C} and Z_{F} are positive constants whose specific values are of no particular relevance below. Let us denote by T the invertible map $\boldsymbol{p} \in D_{\mathrm{C}} \mapsto \sqrt{2/3}\,\boldsymbol{p} \in D_{\mathrm{F}}$. On defining $\tilde{M}\colon D_{\mathrm{C}} \to \mathbb{R}$ via $\tilde{M} := M_{\mathrm{F}} \circ T$ we find that $\tilde{M}$ and M_{C} are equivalent weights. Then, Proposition 2.5 implies that $\mathrm{H}^1_{M_{\mathrm{C}}}(D_{\mathrm{C}})$ and $\mathrm{H}^1_{\tilde{M}}(D_{\mathrm{C}})$ (the latter is well-defined since $\tilde{M}^{-1}$ inherits from M_{C}^{-1} its $\mathrm{L}^1_{\mathrm{loc}}(D_{\mathrm{C}})$ regularity—thereby falling under the hypotheses of [**KO84**, Theorem 1.11]) are algebraically and topologically the same space. The same is true of the pairs of spaces given by $\mathrm{L}^2_{M_{\mathrm{C}}}(D_{\mathrm{C}})$ and $\mathrm{L}^2_{\tilde{M}}(D_{\mathrm{C}})$ and $\mathrm{H}(M_{\mathrm{C}}; D_{\mathrm{C}})$ and $\mathrm{H}(\tilde{M}; D_{\mathrm{C}})$.

Now, T and T^{-1} are $[\mathrm{C}^\infty(\overline{D_{\mathrm{C}}})]^d$ and $[\mathrm{C}^\infty(\overline{D_{\mathrm{F}}})]^d$ functions, respectively. Then, an argument analogous to Lemma A.4 leads to the fact that the composition with T^{-1} is a well-defined, invertible, linear and bounded operator between $\mathrm{H}^1_{\tilde{M}}(D_{\mathrm{C}})$ and $\mathrm{H}^1_{M_{\mathrm{F}}}(D_{\mathrm{F}})$ and also between $\mathrm{L}^2_{\tilde{M}}(D_{\mathrm{C}})$ and $\mathrm{L}^2_{M_{\mathrm{F}}}(D_{\mathrm{F}})$, and its inverse is the composition with T. By (2.4), composition with T^{-1} is also such an operator between $\mathrm{H}(D_{\mathrm{C}}; \tilde{M})$ and $\mathrm{H}(D_{\mathrm{F}}; M_{\mathrm{F}})$ having as its inverse the composition with T.

We can thus use the connection between the M_{F}-weighted spaces and the $\tilde{M}$-weighted spaces and the connection between the latter and the M_{C}-weighted spaces to state that

$$\mathrm{H}^1_{M_{\mathrm{F}}}(D_{\mathrm{F}}) \Subset \mathrm{L}^2_{M_{\mathrm{F}}}(D_{\mathrm{F}}) \implies \mathrm{H}^1_{M_{\mathrm{C}}}(D_{\mathrm{C}}) \Subset \mathrm{L}^2_{M_{\mathrm{C}}}(D_{\mathrm{C}})$$

and

$$\overline{\mathrm{C}^\infty_0(D_{\mathrm{F}})}^{\mathrm{H}(D_{\mathrm{F}};M_{\mathrm{F}})} = \mathrm{H}(D_{\mathrm{F}};M_{\mathrm{F}}) \implies \overline{\{f\circ T\colon f\in \mathrm{C}^\infty_0(D_{\mathrm{F}})\}}^{\mathrm{H}(D_{\mathrm{C}};M_{\mathrm{C}})} = \mathrm{H}(D_{\mathrm{C}};M_{\mathrm{C}}).$$

As $2b/3 \geq 2$, the statements on the left-hand side of the above implications hold (as noted in Remark 2.2 and Remark 3.9); consequently, so do the statements on each right-hand side. By noting that, on account of its infinite differentiability, the composition with T maps $\mathrm{C}^\infty_0(D_{\mathrm{F}})$ into $\mathrm{C}^\infty_0(D_{\mathrm{C}})$ and that M_{C} itself is a $\mathrm{C}^\infty(D_{\mathrm{C}})$ function, we have the following lemma:

Lemma 2.6. *Let $M\colon D\to\mathbb{R}$ be the Maxwellian associated to a spring obeying the CPAIL force model* (1.17) *with parameter $b\geq 3$, the TEAIL force model* (1.18) *with parameter $b\geq 16/5$ or the CP force model* (1.21) *with parameter $b\geq 3$. Then,*

(1) The compact embedding $\mathrm{H}^1_M(D)\Subset\mathrm{L}^2_M(D)$ holds.
(2) The set $\mathrm{C}^\infty_0(D)$ is dense in $\mathrm{H}(D;M)$.

Proof. The proof for the CPAIL force model is contained in the discussion that precedes this lemma. Now, it is easy, yet somewhat laborious, to use (1.18) and (1.24) to prove that the Maxwellian associated with the TEAIL force model with parameter b is equivalent as a weight to the function

$$\boldsymbol{p}\in D\mapsto \left(1-\frac{|\boldsymbol{p}|^2}{b}\right)^{5b/16}.$$

Hence, the same arguments that gave us the result for the CPAIL force model apply to the TEAIL force model with the stated restriction on its parameter b. The proof in the case of springs obeying the CP force model is completely analogous to the CPAIL case, so we make no further comment. □

2.2.2. Sobolev spaces weighted with Inverse Langevin Maxwellians. First of all, we need to know whether the Inverse Langevin force law defined in (1.19), comes from a potential as described in the text of Subsection 1.1.3 and its Hypothesis A; i.e., whether, given some $b_i>0$, there exists some $\mathrm{C}^1([0,b_i/2))$ potential U_i, diverging towards plus infinity as its argument tends to $b_i/2$ from the left, such that

$$U_i'(\tfrac{1}{2}\,|\boldsymbol{q}_i|^2)\boldsymbol{q}_i = \frac{\sqrt{b_i}}{3}L^{-1}\left(\frac{|\boldsymbol{q}_i|}{\sqrt{b_i}}\right)\frac{\boldsymbol{q}_i}{|\boldsymbol{q}_i|};$$

we recall that the Langevin function $L(t)$ is defined by $\coth(t)-1/t$ in $(0,\infty)$ and is continuously extended to $t=0$ as $L(0)=0$.

For this, we prove that $\hat{F}\colon[0,1)\to\mathbb{R}$ defined by $\hat{F}(s)=L^{-1}(\sqrt{s})/\sqrt{s}$ is the derivative of a $\mathrm{C}^1([0,1))$ function which tends to ∞ as its argument tends to 1 from the left. As L^{-1} is

continuous in $[0,1)$, the $\mathrm{C}([0,1))$ regularity of $\hat{F}$ will be guaranteed by the finiteness of

$$\lim_{y\to 0_+} \frac{L^{-1}(y)}{y} = \lim_{t\to 0_+} \frac{t}{L(t)} = \lim_{t\to 0_+} \frac{t}{\coth(t) - 1/t} = 3, \tag{2.7}$$

which comes about via the right-continuity of the Langevin function L. We will need the auxiliary result that follows.

Proposition 2.7.

$$L(t) \leq \frac{t}{1+t} \qquad \forall\, t \in [0,\infty). \tag{2.8}$$

Proof. At $t = 0$, both sides evaluate to 0. When $t > 0$, from the truncated-series inequality $2t^2 + 2t + 1 \leq e^{2t}$ we obtain, successively

$$2e^{-t}t^2 + 2e^{-t}t + e^{-t} \leq e^t,$$
$$(e^t + e^{-t})(t^2 + t) \leq (e^t - e^{-t})(t^2 + t + 1)$$

and

$$\coth(t) \leq \frac{t^2+t+1}{t^2+t} = \frac{t^2}{t^2+t} + \frac{t+1}{t^2+t} = \frac{t}{1+t} + \frac{1}{t}.$$

□

Using the change of variable $s = t/(1+t) \in [0,1)$ in (2.8), we obtain

$$L\left(\frac{s}{1-s}\right) \leq s \qquad \text{and thus} \qquad \frac{s}{1-s} \leq L^{-1}(s).$$

Dividing the last inequality by s, if $s \in (0,1)$, and using the limit (2.7) if $s = 0$, we obtain that, for $s \in [0,1)$,

$$\hat{F}(s) = \frac{L^{-1}(\sqrt{s})}{\sqrt{s}} \geq \frac{1}{1-\sqrt{s}}. \tag{2.9}$$

Then, from the continuity of $\hat{F}$ and the comparison (2.9), it follows that

$$\hat{U}\colon s \mapsto \int_0^s \hat{F}(s')\,\mathrm{d}s'$$

is the sought-after $\mathrm{C}^1([0,1))$ function diverging to ∞ at 1, and the $\mathrm{C}^1([0,b_i/2))$ potential diverging at $b_i/2$ is given by the rescaling

$$U_i(s) = \frac{b_i}{6}\hat{U}\left(\frac{2s}{b_i}\right). \tag{2.10}$$

We will need the following auxiliary result, which an analogue of (2.9), in order to obtain a useful upper bound on U_i.

Proposition 2.8.

$$\hat{F} \leq \frac{1}{\sqrt{s}-s} \qquad \forall\, s \in (0,1). \tag{2.11}$$

Proof. A straightforward consequence of the fact that $\sinh(t) \leq \cosh(t)$ is

$$1 - \frac{1}{t} \leq L(t) \qquad \forall\, t \in (1, \infty). \tag{2.12}$$

Using change of variable $s = (t-1)/t \in (0,1)$ and applying L^{-1} to both sides of the resulting inequality we obtain

$$L^{-1}(s) \leq \frac{1}{1-s},$$

whence the result. □

Let M_i be the Maxwellian weight associated with the potential U_i of (2.10), according to the definition given in (1.24). From the relation (2.9) it transpires that, for $\boldsymbol{p} \in D_i = B(0, \sqrt{b_i})$,

$$\begin{aligned} Z_i M_i(\boldsymbol{p}) = \exp(-U_i(\tfrac{1}{2}|\boldsymbol{p}|^2)) &= \exp\left(-\frac{b_i}{6}\hat{U}\left(\frac{|\boldsymbol{p}|^2}{b_i}\right)\right) \\ &\leq \exp\left(-\frac{b_i}{6}\int_0^{|\boldsymbol{p}|^2/b_i} \frac{1}{1-\sqrt{s'}}\,\mathrm{d}s'\right) = \exp\left(\frac{b_i}{3}\left[\frac{|\boldsymbol{p}|}{\sqrt{b_i}} + \log\left(1 - \frac{|\boldsymbol{p}|}{\sqrt{b_i}}\right)\right]\right) \\ &= \exp\left(\frac{\sqrt{b_i}|\boldsymbol{p}|}{3}\right)\left(1 - \frac{|\boldsymbol{p}|}{\sqrt{b_i}}\right)^{b_i/3}. \end{aligned}$$

Here, Z_i is a positive constant which, besides existing, is of no interest to us. Similarly, starting with (2.11), we can obtain, for $\boldsymbol{p} \in D_i$,

$$Z_i M_i(\boldsymbol{p}) \geq \exp\left(-\frac{b_i}{6}\int_0^{|\boldsymbol{p}|^2/b_i} \frac{1}{\sqrt{s'} - s'}\,\mathrm{d}s'\right) = \left(1 - \frac{|\boldsymbol{p}|}{\sqrt{b_i}}\right)^{b_i/3}$$

Therefore, the Maxwellian associated with the Inverse Langevin force law can be bounded from above and below by the CPAIL Maxwellian with the same parameter b_i (cf. (2.6); in other words, these two Maxwellians are equivalent. As Hypothesis B is a topological statement involving Maxwellian-weighted Sobolev spaces, Proposition 2.5 and the fact that CPAIL Maxwellians with parameter $b_i \geq 3$ obey Hypothesis B (proved in Lemma 2.6) implies that Inverse Langevin Maxwellians with parameter $b_i \geq 3$ obey Hypothesis B as well. For the very same reason, $\mathrm{C}_0^\infty(D_i)$ is dense in $\mathrm{H}(D_i; M_i)$ if M_i comes from the Inverse Langevin Force law with parameter $b_i \geq 3$.

Proposition 2.9. *The potential U_i associated to the Inverse Langevin force law (cf.* (2.10)*) is monotonic increasing and convex.*

Proof. As U_i is a linear rescaling of $\hat{U}$, it is enough to prove that $\hat{U}$ is monotonic increasing and convex. As $\hat{U}(s) = \int_0^s \hat{F}(s')\,\mathrm{d}s'$, $s \in [0,1)$, its first derivative is $\hat{F} = L^{-1}(\sqrt{\cdot})/\sqrt{\cdot}$, which is non-negative. We will now prove that $\hat{F}$ is monotonic increasing, whence the convexity of $\hat{U}$ will follow.

The first step is showing that the Langevin function L is concave in $(0,\infty)$. Indeed,

$$L''(t) = 2\left(\frac{\cosh(t)}{\sinh(t)^3} - \frac{1}{t^3}\right) \leq 0,$$

which can be easily yet laboriously proved by comparing the series expansions of $t \mapsto t^3\cosh(t)$ and $t \mapsto \sinh(t)^3$. By continuity, L is concave in $[0,\infty)$ as well.

Now, let us take $0 < s_1 < s_2 < 1$ and let $t_i := L^{-1}(\sqrt{s_i})$, $i \in \{1,2\}$; note that since both L and the square-root function are strictly monotonic increasing, we have that $0 < t_1 < t_2$. The concavity of L, then, allows for

$$t_1 = \frac{t_2 - t_1}{t_2}0 + \frac{t_1}{t_2}t_2 \implies L(t_1) \geq \frac{t_2 - t_1}{t_2}L(0) + \frac{t_1}{t_2}L(t_2) = \frac{t_1}{t_2}L(t_2)$$
$$\implies \frac{t_1}{L(t_1)} \leq \frac{t_2}{L(t_2)} \implies \frac{L^{-1}(\sqrt{s_1})}{\sqrt{s_1}} \leq \frac{L^{-1}(\sqrt{s_2})}{\sqrt{s_2}};$$

that is, $\hat{F}$ is monotonic increasing in $(0,1)$, which implies that $\hat{U}$ in convex in $[0,1)$. □

2.2.3. Eigenvalue asymptotics for partial Maxwellian-weighted operators. We will need to know the asymptotic behavior of the eigenvalues of Maxwellian-weighted eigenvalue problems. More precisely, we need to know if Weyl's law is satisfied by these problems.

Lemma 2.10. *Let $\Omega \subset \mathbb{R}^d$ be a bounded and convex domain of class C^3 and let $w \in \mathrm{C}^2(\Omega)$ be a positive function such that $\mathrm{C}^2_0(\Omega)$ is dense in $\mathrm{H}^1_w(\Omega)$ and $\mathrm{H}^1_w(\Omega) \Subset \mathrm{L}^2_w(\Omega)$. We further assume that*

(1) $\inf_{\boldsymbol{p}\in\Omega} Q_1(\boldsymbol{p}) > -\infty$, or
(2) there exists a $\Theta > 0$ such that $\gamma_\Theta := \inf_{\boldsymbol{p}\in\Omega} \mathfrak{d}(\boldsymbol{p})^2 Q_\Theta(\boldsymbol{p}) \in (-1/4, 0]$,

where

$$Q_\Theta := \Theta - w^{-1/2}\operatorname{div}(w\nabla w^{-1/2})$$

and $\mathfrak{d}$ is the distance-to-the-boundary function in Ω.

Let $(\lambda_n : n \in \mathbb{N})$ be the (ordered, with repetitions according to multiplicity) sequence of eigenvalues of the problem: Find $\lambda \in \mathbb{R}$ and $u \in \mathrm{H}^1_w(\Omega) \setminus \{0\}$ such that

$$\langle u, v\rangle_{\mathrm{H}^1_w(\Omega)} = \lambda\langle u, v\rangle_{\mathrm{L}^2_w(\Omega)} \qquad \forall\, v \in \mathrm{H}^1_w(\Omega). \tag{2.13}$$

Then, there exist positive numbers c_1 and c_2 and a natural number n_0 such that

$$n \geq n_0 \implies c_1 n^{2/d} \leq \lambda_n \leq c_2 n^{2/d}. \tag{2.14}$$

Proof. Let, for $\Theta > 0$, $(\lambda_{\Theta,n} : n \in \mathbb{N})$ be the (ordered, with repetitions according to multiplicity) sequence of of eigenvalues of the shifted problem: Find $\lambda^\Theta \in \mathbb{R}$ and $u \in \mathrm{H}^1_w(\Omega) \setminus \{0\}$ such that

$$\langle u, v\rangle_{\mathrm{H}^1_w(\Omega),\Theta} := \langle \nabla u, \nabla v\rangle_{[\mathrm{L}^2_w(\Omega)]^d} + \Theta\langle u, v\rangle_{\mathrm{L}^2_w(\Omega)} = \lambda^\Theta\langle u, v\rangle_{\mathrm{L}^2_w(\Omega)} \quad \forall\, v \in \mathrm{H}^1_w(\Omega). \tag{2.15}$$

By the hypotheses of the lemma the existence and the accumulation at ∞ only of the $\lambda_{\Theta,n}$ is guaranteed via Lemma A.5 in Appendix A. It further follows from the spectral theory of self-adjoint compact operators that $\lambda_{\Theta,n}$ can be characterized by the Courant–Fischer–Weyl min-max principle:

$$\lambda_{\Theta,n} = \min_{\substack{\dim(S)=n \\ S\subset \mathrm{H}^1_w(\Omega)}} \max_{z\in S\setminus\{0\}} \frac{\langle z,z\rangle_{\mathrm{H}^1_w(\Omega),\Theta}}{\langle z,z\rangle_{\mathrm{L}^2_w(\Omega)}} = \inf_{\substack{\dim(S)=n \\ S\subset \mathrm{C}^2_0(\Omega)}} \sup_{z\in S\setminus\{0\}} \frac{\langle z,z\rangle_{\mathrm{H}^1_w(\Omega),\Theta}}{\langle z,z\rangle_{\mathrm{L}^2_w(\Omega)}}, \tag{2.16}$$

the second equality being a consequence of the density of $\mathrm{C}^2_0(\Omega)$ in $\mathrm{H}^1_w(\Omega)$ (cf. [**Dav95**, Theorem 4.5.3]). Note that when $\Theta = 1$ the problem (2.15) and the problem (2.13) coincide (and so do the sequences $(\lambda_{\Theta,n}\colon n\in\mathbb{N})$ and $(\lambda_n\colon n\in\mathbb{N})$).

Let $L := w^{-1/2} \in \mathrm{C}^2(\Omega)$, let z be an arbitrary $\mathrm{C}^2_0(\Omega)$ function and let $y := L^{-1}z$. Then,

$$\begin{aligned}
\|z\|^2_{\mathrm{H}^1_w(\Omega),\Theta} &= \int_\Omega \left(|\nabla(Ly)|^2 + \Theta(Ly)^2\right) L^{-2} \\
&= \int_\Omega |\nabla y|^2 + \int_\Omega \left(\Theta + L^{-2}\,|\nabla L|^2\right) y^2 + \int_\Omega 2yL^{-1}\nabla L\cdot\nabla y \\
&= \int_\Omega |\nabla y|^2 + \int_\Omega \left(\Theta + L^{-2}\,|\nabla L|^2\right) y^2 + \int_\Omega L^{-1}\nabla L\cdot\nabla(y^2) \\
&= \int_\Omega |\nabla y|^2 + \int_\Omega \left[\Theta + L^{-2}\,|\nabla L|^2 - \operatorname{div}\left(L^{-1}\nabla L\right)\right] y^2 \\
&= \int_\Omega |\nabla y|^2 + \int_\Omega \left[\Theta - L\,\operatorname{div}(L^{-2}\nabla L)\right] y^2 \\
&= \int_\Omega |\nabla y|^2 + \int_\Omega Q_\Theta\, y^2.
\end{aligned}$$

Similarly, $\|z\|^2_{\mathrm{L}^2_w(\Omega)} = \|y\|^2_{\mathrm{L}^2(\Omega)}$. As $z\in\mathrm{C}^2_0(\Omega)$ is arbitrary and $z\mapsto L^{-1}z$ is a bijection of $\mathrm{C}^2_0(\Omega)$ into itself, (2.16) begets

$$\lambda_{\Theta,n} = \inf_{\substack{\dim(S)=n \\ S\subset \mathrm{C}^2_0(\Omega)}} \sup_{y\in S\setminus\{0\}} \frac{\|\nabla y\|^2_{[\mathrm{L}^2(\Omega)]^d} + \int_\Omega Q_\Theta\, y^2}{\|y\|^2_{\mathrm{L}^2(\Omega)}}. \tag{2.17}$$

If condition (1) holds, there must exist a $\Theta > 0$ such that $Q_\Theta \geq 0$ in Ω. For such a Θ, of course, $\int_\Omega Q_\Theta y^2 \geq 0$. On the other hand, if condition (2) is met, then with the particular Θ given in the condition we have that

$$\int_\Omega Q_\Theta\, y^2 \geq \gamma_\Theta \int_\Omega \frac{y^2}{\mathfrak{d}^2} \geq \frac{\gamma_\Theta}{4}\,\|\nabla y\|^2_{[\mathrm{L}^2_w(\Omega)]^d},$$

the last inequality being a multi-dimensional Hardy inequality (see, e.g., [**MMP98**, Theorem 11], bearing in mind that γ_Θ has been assumed to be nonpositive). In either case, we can

write

$$\lambda_{\Theta,n} \geq \inf_{\substack{\dim(S)=n \\ S\subset \mathrm{C}^2_0(\Omega)}} \sup_{y\in S\setminus\{0\}} \frac{\alpha \left\|\nabla y\right\|^2_{[\mathrm{L}^2(\Omega)]^d}}{\left\|y\right\|^2_{\mathrm{L}^2(\Omega)}}, \tag{2.18}$$

where

$$0 < \alpha := \begin{cases} 1 & \text{if condition (1) holds,} \\ (1+\gamma_\Theta/4) & \text{if condition (2) holds.} \end{cases}$$

The C^3 regularity of $\partial\Omega$ implies the existence of an $\varepsilon_0 \in (0,1)$ such that for each $\varepsilon \in (0,\varepsilon_0)$ there exists a subdomain $\Omega_\varepsilon \Subset \Omega$ that is also of class C^3 and has measure $(1-\varepsilon)\left|\Omega\right|$. Fixing $\varepsilon \in (0,\varepsilon_0)$, the fact that the extensions by zero of functions in $\mathrm{C}^2_0(\Omega_\varepsilon)$ form a subspace of $\mathrm{C}^2_0(\Omega)$ and (2.16) imply that the eigenvalues of the unshifted problem (2.13) can be bounded from above according to

$$\lambda_n \leq \inf_{\substack{\dim(S)=n \\ S\subset \mathrm{C}^2_0(\Omega_\varepsilon)}} \sup_{z\in S\setminus\{0\}} \frac{\langle z,z\rangle_{\mathrm{H}^1_w(\Omega_\varepsilon)}}{\langle z,z\rangle_{\mathrm{L}^2_w(\Omega_\varepsilon)}}. \tag{2.19}$$

Now, the right-hand side of (2.18) and the right-hand side of (2.19) are precisely the n-th eigenvalue associated with the (variational form of the) problem

$$-\alpha\Delta y = \mu y \ \text{ in } \Omega, \qquad y = 0 \text{ on } \partial\Omega$$

and the problem

$$-\operatorname{div}(w\nabla y) + wy = \nu w y \ \text{ in } \Omega_\varepsilon, \qquad y = 0 \text{ on } \partial\Omega_\varepsilon,$$

respectively. These standard eigenvalue problems obey Weyl's law (this results from the fairly general Theorem 2.4 of [**Cla67**] with input from the regularity result in [**Bro61**, Theorem 2.4]—alternatively, see [**CH53**, §VI.4.4]); that is,

$$\lim_{\mu\to\infty} \frac{\#\{n\in\mathbb{N}\colon \mu_n \leq \mu\}}{\mu^{d/2}} = \frac{\alpha^{-d/2}\left|\Omega\right|}{(2\sqrt{\pi})^d\,\Gamma(1+d/2)} = \alpha^{-d/2}C > 0, \tag{2.20a}$$

$$\lim_{\nu\to\infty} \frac{\#\{n\in\mathbb{N}\colon \nu_n \leq \nu\}}{\nu^{d/2}} = \frac{\left|\Omega_\varepsilon\right|}{(2\sqrt{\pi})^d\,\Gamma(1+d/2)} = (1-\varepsilon)C > 0, \tag{2.20b}$$

where $C := \left|\Omega\right|\left((2\sqrt{\pi})^d\Gamma(1+d/2)\right)^{-1}$. Particularizing these limits to $\mu=\mu_n$ and $\nu=\nu_n$ they turn into statements about the rate of growth of the eigenvalues themselves, as opposed to the counting functions. That is,

$$\lim_{n\to\infty} \mu_n/n^{2/d} = \alpha C^{-2/d} \qquad \text{and} \qquad \lim_{n\to\infty} \nu_n/n^{2/d} = (1-\varepsilon)^{-2/d}C^{-2/d}.$$

From the definition of the shifted eigenvalue problem (2.15), for any Θ, it is immediate that

$$\lambda_{\Theta,n} = \lambda_n + \Theta - 1 \qquad \forall\, n\in\mathbb{N}.$$

We then deduce, via the inequalities (2.18) and (2.19), that the asymptotic bounds (2.14) hold. □

Remark 2.11.

(1) It follows from the proof of Lemma 2.10 that, if condition (1) holds, the constants c_1 and c_2 of (2.14) can be taken arbitrarily close to $C^{-2/d}$ and, consequently, to each other.

(2) One might relax the condition of convexity of the domain in Lemma 2.10 at the possible cost of having a stricter lower bound for γ_Θ in condition (2), as the constant for the Hardy inequality might deteriorate. The C^3 regularity condition on the domain can be drastically relaxed (see, for example [**BS70**]); however, the literature tends to force one to choose at most two among readability, the size of the class of problems covered, and frugality in terms of hypotheses. For our purposes, the statement in Lemma 2.10 suffices.

(3) The transformation of the last Rayleigh quotient in (2.16) into the simpler one in (2.17) by means of rescalings of both the eigenfunctions and their argument we did in Lemma 2.10 is an instance of what is called the Liouville Transformation technique in the theory of Sturm-Liouville problems (see [**Eve05**, §7]).

Corollary 2.12. *The eigenvalues of the eigenvalue problem* (3.33) *associated with both the FENE model* (1.16) *and the CPAIL model* (1.17) *obey* (2.14) *if their parameter b_i is greater than* 2 *and* 3*, respectively.*

Proof. We shall apply Lemma 2.10. For both the FENE and CPAIL models the domains (being balls) and their associated Maxwellian weights are regular enough. The compact embedding and density hypotheses are satisfied in the parameter ranges under consideration (cf. Hypothesis B, Remark 2.2, Remark 3.9 and (2.4)). It only remains to prove condition (1) or condition (2).

From (1.16) and (1.24) it follows that the Maxwellian associated to the FENE potential is

$$M_i(\boldsymbol{p}) = Z_i^{-1}\left(1 - |\boldsymbol{p}|^2/b_i\right)^{b_i/2}, \qquad \boldsymbol{p} \in B(0, \sqrt{b_i}), \tag{2.21}$$

where Z_i is a positive constant. A direct calculation returns that with this weight the quantity Q_Θ defined in Lemma 2.10 is

$$Q_\Theta(\boldsymbol{p}) = \Theta + \left(\frac{1}{4} - \frac{1}{b_i}\right)|\boldsymbol{p}|^2\left(1 - \frac{|\boldsymbol{p}|^2}{b_i}\right)^{-2} - \frac{d}{2}\left(1 - \frac{|\boldsymbol{p}|^2}{b_i}\right)^{-1}.$$

In this form, it is readily apparent that Q_1 is bounded from below in its domain $B(0, \sqrt{b_i})$ (i.e., 1 holds) if $b_i > 4$. From the fact that $\mathfrak{d}(\boldsymbol{p}) = \sqrt{b} - |\boldsymbol{p}|$ for all $\boldsymbol{p}$ in the domain under consideration it is easy to see that $\mathfrak{d}^2 Q_\Theta$ is always bounded from below and uniformly continuous up to the boundary. If $b_i \in (2, 4]$, Q_Θ is never bounded from below, so it takes negative values and thus the infimum of $\mathfrak{d}^2 Q_\Theta$ is strictly less than zero. As $\mathfrak{d}^2$ is continuous and positive within the domain yet zero at its boundary, the existence of a Θ that makes case

(2) hold is equivalent to demanding that

$$\lim_{|\boldsymbol{p}|\to\sqrt{b_i}_-} \mathfrak{d}(\boldsymbol{p})^2 Q_1(\boldsymbol{p}) \in (-1/4, 0].$$

As in the range $b_i \in (2,4]$ that limit is $b_i(b_i/4-1)/4$ we see that the condition (2) holds there.

Analogously, (1.17) and (1.24) imply that the Maxwellian associated to the CPAIL potential is

$$M_i(\boldsymbol{p}) = Z_i^{-1} \exp\left(-|\boldsymbol{p}|^2/6\right) \left(1 - |\boldsymbol{p}|^2/b_i\right)^{b_i/3}, \qquad \boldsymbol{p} \in B(0, \sqrt{b_i}), \tag{2.22}$$

with Z_i a positive constant. Again, a direct calculation yields

$$Q_\Theta(\boldsymbol{p}) = \Theta - \frac{d}{6} + \frac{|\boldsymbol{p}|^2}{36} + \left(\frac{1}{9} - \frac{2}{3b_i}\right)|\boldsymbol{p}|^2 \left(1 - \frac{|\boldsymbol{p}|^2}{b_i}\right)^{-2} - \left(\frac{d}{3} - \frac{|\boldsymbol{p}|^2}{9}\right)\left(1 - \frac{|\boldsymbol{p}|^2}{b_i}\right)^{-1}.$$

By arguments similar to those given when considering the FENE potential, we have that condition (1) holds if $b_i > 6$ or if $b_i = 6$ and $d = 2$; and that condition (2) holds if $b_i \in (3,6]$. □

If two weights w and $\tilde{w}$ defined on a domain Ω are equivalent—that is, there exist two positive constants c_1 and c_2 such that $c_1\, w \le \tilde{w} \le c_2\, w$—a number of consequences follow immediately. As shown in Proposition 2.5, $\mathrm{L}^2_w(\Omega)$ and $\mathrm{L}^2_{\tilde{w}}(\Omega)$ on the one hand and $\mathrm{H}^1_w(\Omega)$ and $\mathrm{H}^1_{\tilde{w}}(\Omega)$ on the other will be one and the same algebraically and topologically. In particular, the hypotheses of Lemma A.5 will be met by the eigenvalue problem

$$\langle e, v\rangle_{\mathrm{H}^1_w(\Omega)} = \lambda \langle e, v\rangle_{\mathrm{L}^2_w(\Omega)} \qquad \forall\, v \in \mathrm{H}^1_w(\Omega)$$

if, and only if, they are met by the eigenvalue problem

$$\langle e, v\rangle_{\mathrm{H}^1_{\tilde{w}}(\Omega)} = \tilde{\lambda} \langle e, v\rangle_{\mathrm{L}^2_{\tilde{w}}(\Omega)} \qquad \forall\, v \in \mathrm{H}^1_{\tilde{w}(\Omega)}.$$

The inf-sup characterization (cf. (2.16)) of the successive eigenvalues of both problems allow for the bounds

$$\frac{c_1}{c_2}\lambda_n \le \tilde{\lambda}_n \le \frac{c_2}{c_1}\lambda_n.$$

That is, the bounds (2.14) will hold for one set of eigenvalues if, and only if, they hold for the other. This allows for establishing the following sufficiency condition for weights defined on two- or three-dimensional balls, which is in most cases much easier to test than the conditions of Lemma 2.10.

Lemma 2.13. *Let Ω be an open ball in two or three dimensions and let w be a positive and continuous weight defined on Ω with the property*

$$\sigma_1 \mathfrak{d}(\boldsymbol{p})^\alpha \le w(\boldsymbol{p}) \le \sigma_2 \mathfrak{d}(\boldsymbol{p})^\alpha,$$

where $\mathfrak{d}$ is the distance-to-the-boundary function, for all $\boldsymbol{p} \in \Omega$ such that $\mathfrak{d}(\boldsymbol{p}) < \delta$, for some exponent $\alpha > 1$, for some margin $\delta > 0$ and some positive constants σ_1 and σ_2.

Then, the eigenvalues of the problem

$$\langle e, v\rangle_{\mathrm{H}^1_w(\Omega)} = \lambda \langle e, v\rangle_{\mathrm{L}^2_w(\Omega)} \qquad \forall\, v \in \mathrm{H}^1_w(\Omega)$$

obey the two-sided bounds (2.14).

Proof. If the radius of the ball happens to be $\sqrt{2\alpha}$ the conditions on w force it to be comparable to the FENE Maxwellian (2.21) and so the result follows from the above discussion. Otherwise, one just needs to rescale the domain; this will effect a fixed linear transformation on the eigenvalues, but will not affect the validity of the bounds (2.14) (the constants involved will change, though). □

Corollary 2.14. *The eigenvalues of the eigenvalue problem* (3.33) *associated with the TEAIL model* (1.18) *with parameter* $b_i > 16/5$*, the CP model* (1.21) *with parameter* $b_i > 3$ *or the Inverse Langevin model* (1.19) *with parameter* $b_i > 3$ *obey* (2.14).

Proof. As stated in the proof of Lemma 2.6, the Maxwellian weight associated with the TEAL force model with parameter b_i is equivalent to $\boldsymbol{p} \mapsto (1 - |\boldsymbol{p}|^2/b_i)^{5b_i/16}$. As

$$1 - \frac{|\boldsymbol{p}|^2}{b_i} = \left(1 + \frac{|\boldsymbol{p}|}{\sqrt{b_i}}\right)\left(1 - \frac{|\boldsymbol{p}|}{\sqrt{b_i}}\right) = \left(1 + \frac{|\boldsymbol{p}|}{\sqrt{b_i}}\right)\frac{1}{\sqrt{b_i}}\mathfrak{d}(\boldsymbol{p})$$

gives us the desired result for the TEAIL force model.

Now, the Maxwellian associated with the CP force model is (modulo a multiplicative constant)

$$\boldsymbol{p} \in D_i \mapsto \exp\left(\frac{|\boldsymbol{p}|^4}{60b_i} - \frac{|\boldsymbol{p}|^2}{6}\right)\left(1 - \frac{|\boldsymbol{p}|^2}{b_i}\right)^{b_i/3}.$$

Therefore, it is equivalent to the Maxwellian associated with the CPAIL force model (cf. (2.6) with the same parameter b_i. As shown in Subsection 2.2.2, the Maxwellian associated with the Inverse Langevin force law is also equivalent to the CPAIL Maxwellian with the same parameter. Therefore, in view of the discussion that precedes Lemma 2.13 and Corollary 2.12, result is valid for the CP force model and the Inverse Langevin force model in the stated parameter range. □

Remark 2.15. The eigenvalue problem (3.33) associated with either the FENE or the CPAIL model falls within what is called *weak degeneracy* case in the Russian spectral theory literature; i.e., problems of the form: Given $\Omega \subset \mathbb{R}^d$, find $(\lambda, u) \in \mathbb{R} \times (\mathrm{H}^1_{\mathfrak{d}^\alpha}(\Omega) \setminus \{0\})$ such that

$$\int_\Omega (A\nabla u \cdot \nabla v + h\, u\, v)\, \mathfrak{d}^\alpha = \lambda \int_\Omega b\, u\, v\, \mathfrak{d}^\beta \qquad \forall\, v \in \mathrm{H}^1_{\mathfrak{d}^\alpha}(\Omega), \tag{2.23}$$

where $\alpha - \beta < 2/d$ (see [**VS74**, §1] for the precise statement, which includes additional conditions on Ω, A, h, b, α and β). As, in the FENE and CPAIL versions of (3.33), the same weight (the associated Maxwellian) appears in both the left- and right-hand side bilinear forms, and, in both cases, that weight is bounded from above and below by powers of $\mathfrak{d}$ (cf.

(2.21), (2.22)), it turns out that our problem is equivalent to a problem of the form (2.23) with $\alpha - \beta = 0$.

The result, according to [**VS74**, Theorem 1.1] and assuming that $b \geq 0$ is that

$$\lim_{\lambda\to\infty} \lambda^{-d/2}\#\{n \in \mathbb{N}\colon \lambda_n < \lambda\} = \frac{1}{(2\sqrt{\pi})^d\Gamma(1+d/2)} \int_\Omega \frac{\mathfrak{d}^{-(\alpha-\beta)d/2} b^{d/2}}{\sqrt{\det(A)}} \tag{2.24}$$

(compare this with (2.20); note also that in [**VS74**] the statement is made in terms of what in our notation is $1/\lambda$). The problem with this particular source is that, for a proof, it remits the reader to either one of two publications. The first, [**BS72**] proves related yet not directly applicable results—there is a gap that needs to be bridged by means, perhaps elementary, that are unknown to us. We have not been able to get hold of the second, [**Taš75**] by G. M. Taščijan (also romanized as Tashchiyan). However, the latter is also cited in [**Taš81**, Theorem 1], where a generalization of (2.24) is proved, under the condition (in our notation) $d > 2$.

2.3. Cartesian product of Lipschitz domains

It is usually necessary, in order to be able to prove non-trivial results on weighted Sobolev spaces, to know something about the regularity of the domains they are defined on. In our case, we deal with a Cartesian-product domain $\mathsf{D} = D_1 \times \cdots \times D_N$, which does not inherit the full smoothness of its factor domains—indeed, D has corners.

It turns out (cf. Lemma 2.20) that the right regularity setting (i.e., the amount of regularity that D *does* inherit) is given by the notion of (bounded) Lipschitz domain, which we recall below.

2.3.1. Lipschitz domains. We start by introducing some notation. Given a vector $q \in \mathbb{R}^n$, for some $n \in \mathbb{N}$, $n \geq 2$, we will write $q' := (q_1, \ldots, q_{n-1})$. Whenever we have an invertible function $A_i\colon \mathbb{R}^n \to \mathbb{R}^n$, an open set $\tilde{\Delta}_i \subset \mathbb{R}^{n-1}$ and a function $a_i\colon \tilde{\Delta}_i \to \mathbb{R}$ sharing the same index i, we will write

$$U_i^I := A_i^{-1}\big(\{q \in \mathbb{R}^n\colon q' \in \tilde{\Delta}_i, q_n - a_i(q') \in I\}\big), \quad I \subset \mathbb{R}. \tag{2.25}$$

We will deal with sets I of the form $(-\beta, 0)$, $(0, \beta)$ and $\{0\}$; this last case, $U_i^{\{0\}}$, will be denoted by Λ_i.

Definition 2.16. *We say that a bounded open set $E \subset \mathbb{R}^n$ is a Lipschitz domain if, and only if, there exists a positive integer m, a positive real number β and, for $r \in [m]$, rigid transformations $A_r\colon \mathbb{R}^n \to \mathbb{R}^n$, open sets $\tilde{\Delta}_r \subset \mathbb{R}^{n-1}$ and Lipschitz functions $a_r\colon \tilde{\Delta}_r \to \mathbb{R}$ such that:*

$$\partial E = \bigcup_{r=1}^m \Lambda_r,$$

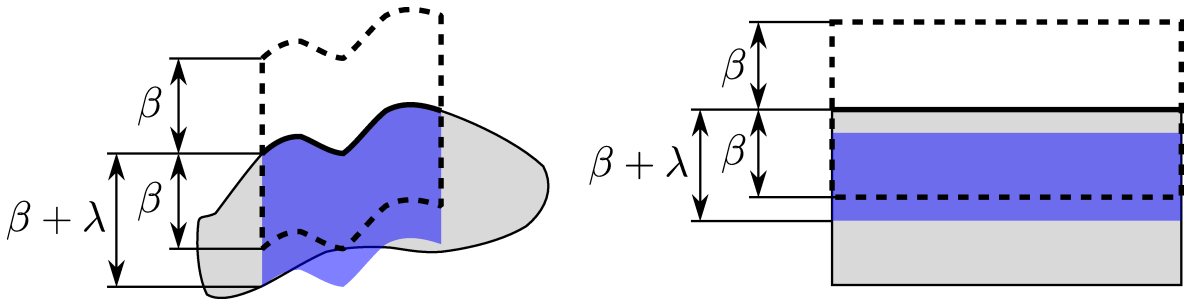

FIGURE 2.1. Problematic Lipschitz representations. *Left*: The margin β is the supremum among the margins admissible according to Definition 2.16—if β is enlarged by any positive λ the resulting undershot area, $U_r^{(-(\beta+\lambda),0)}$ (in blue), will grow beyond the confinement of the domain. *Right*: Whatever the additional margin $\lambda > 0$ is, the shifted undershot area $U_r^{(-(\beta+\lambda),-\lambda)}$ (in blue) will be contained in the domain but not compactly so.

and, for all $r \in [m]$,

$$U_r^{(-\beta,0)} \subset E \quad \textit{and} \quad U_r^{(0,\beta)} \subset \overline{E}^c.$$

We will then say that $\{(A_r, \tilde{\Delta}_r, a_r)\}_{r=1}^m$ *is a Lipschitz representation of the bounded set* E *with margin* β.

Remark 2.17.

(1) The notion of Lipschitz domain of the definition above is equivalent to the notion of a domain satisfying the strong local Lipschitz condition according to [**AF03**, ¶4.9].

(2) If a set has a Lipschitz representation with a certain margin β, any smaller yet still positive real number in $(0, \beta)$ will also be an admissible margin for the same Lipschitz representation. Therefore we can always assume that, given a Lipschitz representation of a bounded set, its margin β has been picked so that there exists some $\lambda^* > 0$ such that $\beta + \lambda^*$ is still an admissible margin; i.e., we avoid the situation depicted in the left part of Figure 2.1. In what follows we will always do so.

(3) Another situation that we want to avoid is depicted in the right part of Figure 2.1; in simple terms, we want to be able to shift the undershot region away from the boundary so that the resulting region does not touch the boundary of the domain. We give a precise description of this aim and show that it is always attainable in the following proposition.

Proposition 2.18. *Let* $E \subset \mathbb{R}^n$ *be a bounded Lipschitz domain. Then it has a Lipschitz representation* $\{(A_r, \tilde{\Delta}_r, a_r)\}_{r=1}^m$ *with associated margin* β *such that, for* $\lambda > 0$ *and small enough, and* $r \in [m]$,

$$U_r^{(-\beta-\lambda,-\lambda)} \Subset E. \tag{2.26}$$

Proof. As E is a bounded Lipschitz domain it has a Lipschitz representation $\{(A_r, \tilde{\Delta}_r, a_r)\}_{r=1}^m$ with margin $\beta > 0$. If (2.26) holds for this Lipschitz representation, we are done. Otherwise, for $r \in [m]$ and $\delta > 0$ let us consider $\tilde{\Delta}_{r,\delta} := \{q' \in \tilde{\Delta} \colon \operatorname{dist}(q', \partial\tilde{\Delta}_r) > \delta\}$. Then, $U_{r,\delta}^{(-\beta-\lambda,-\lambda)} \Subset E$ for $\lambda > 0$ and small enough and where we have defined $A_{r,\delta}$ as A_r and $a_{r,\delta}$ as $a_r|_{\tilde{\Delta}_{r,\delta}}$. We would have what we need if all the properties of a Lipschitz representation of E were retained by $\{(A_{r,\delta}, \tilde{\Delta}_{r,\delta}, a_{r,\delta})\}_{r=1}^m$. This is the case, with the same margin, except, perhaps, for the property that demands that $\partial E = \bigcup_{r=1}^m \Lambda_{r,\delta}$. We will now show that there must exist some $\delta > 0$ such that this last property is also retained.

Let $x \in \partial E$ and $\mathcal{I}_x$ the set of all $r \in [m]$ such that $x \in \Lambda_r$, which is not empty. Given $r \in \mathcal{I}_x$, there exists $\delta_{r,x} > 0$ such that $x \in \Lambda_{r,\delta_{r,x}}$, as $\tilde{\Delta}_r$ is open. Letting $\delta_x := \max\{\delta_{r,x} \colon r \in \mathcal{I}_x\}$, we have that

$$x \in \bigcup_{r=1}^m \Lambda_{r,\delta_x} \subset \bigcup_{r=1}^m U_{r,\delta_x}^{(-\beta,\beta)} =: U_{\delta_x}$$

So ∂E is covered by the family of open sets $\{U_{\delta_x}\}_{x\in\partial E}$. Then, there exists a finite subfamily of those open sets that covers ∂E. Choosing δ as the smallest δ_x associated with members of the finite subfamily of U_{δ_x} obtained above, $\partial E = \bigcup_{r=1}^m \Lambda_{r,\delta}$ is retained. □

Remark 2.19. In the light of the previous result, we can always assume that (2.26) holds and so we will.

2.3.2. Two semi-infinite intervals. Let $E = (-\infty, 0)$. This is not a bounded Lipschitz domain in the sense given by Definition 2.16. It does not only fail to be bounded. As $n-1 = 0$, we can't define non-empty open sets $\tilde{\Delta}_r \subset \mathbb{R}^{n-1}$ to serve as domains of a boundary-describing functions[1]. Its Cartesian product with itself, however, does happen to have a boundary interesting enough to provide some insight into more complicated—and definition-complying—cases.

Letting $a_1 = 0$ and $\beta = 1$ we have that

$$\Lambda_1 := U_1^{\{0\}} = \partial E, \qquad U_1^{(-\beta,0)} \subset E, \qquad U_1^{(0,\beta)} \subset \overline{E}^c,$$

where

$$U_1^I := \{q \in \mathbb{R} \colon q - a_1 \in I\}, \quad I \subset \mathbb{R}.$$

Let us choose β' and γ such that $0 < \gamma < \beta' \leq (1 - 1/\sqrt{2})\beta$. We observe that $U_1^{(-\beta',0)}$, together with $U_0 := \{q \in E \colon \operatorname{dist}(q, \partial E) > \gamma\}$, forms a partition of E.

From the definition of a boundary and the properties of the Cartesian product we have that

$$\begin{aligned} \partial(E \times E) &= (\partial E \times E) \cup (\partial E \times \partial E) \cup (E \times \partial E) \\ &= (\Lambda_1 \times U_0) \cup \Big((\Lambda_1 \times U_1^{(-\beta',0)}) \cup (\Lambda_1 \times \Lambda_1) \cup (U_1^{(-\beta',0)} \times \Lambda_1) \Big) \cup (U_0 \times \Lambda_1). \end{aligned} \tag{2.27}$$

[1]It is possible to extend Definition 2.16 to one-dimensional domains. Doing so, however, requires an amount of special handling that we prefer to avoid, as this case is not directly relevant to our purposes.

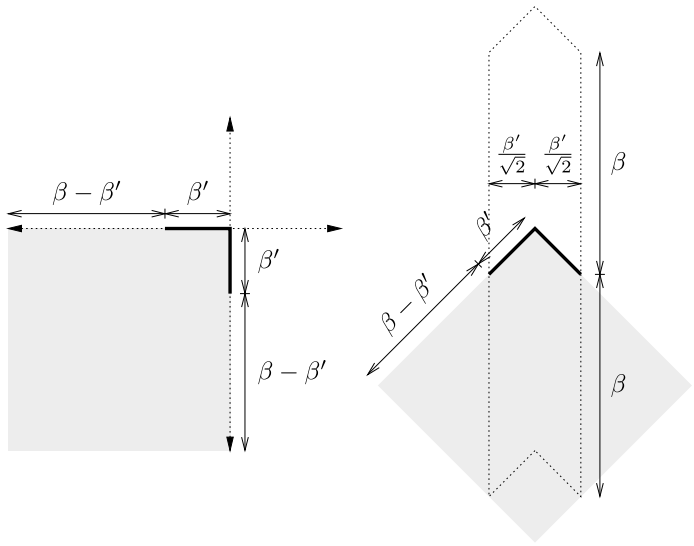

FIGURE 2.2. Illustration of the argument used to represent corners. After a rigid transformation the marked part of the boundary is the graph of a Lipschitz function which by being undershot (resp. overshot) less than β renders points strictly in the inside (resp. outside) of $E \times E$.

Let us consider the first piece identified in the partition described in (2.27). It turns out that it is readily described by $A_{1,0}\colon (q,p) \in \mathbb{R}^2 \to (-p,q) \in \mathbb{R}^2$, $\tilde{\Delta}_{1,0} := -U_0$ and $a_{1,0}\colon p \in \tilde{\Delta}_{1,0} \to a_1 \in \mathbb{R}$ as

$$\Lambda_1 \times U_0 = A_{1,0}^{-1}\left(\{(p,q)\colon p \in \tilde{\Delta}_{1,0}, q = a_{1,0}(p)\}\right).$$

Further, we get that

$$U_{1,0}^{(-\beta,0)} := A_{1,0}^{-1}\left(\{(p,q)\colon p \in \tilde{\Delta}_{1,0}, q - a_{1,0}(p) \in (-\beta,0)\}\right) = U_1^{(-\beta,0)} \times U_0 \subset E \times E,$$
$$U_{1,0}^{(0,\beta)} := A_{1,0}^{-1}\left(\{(p,q)\colon p \in \tilde{\Delta}_{1,0}, q - a_{1,0}(p) \in (0,\beta)\}\right) = U_1^{(\beta,0)} \times U_0 \subset \overline{E \times E}^c.$$

The third piece can be described analogously.

The description of the second piece of the boundary identified in (2.27) is a bit more complicated as it needs to involve a rotation that isn't merely a shuffling of components with perhaps some changes of sign and we have to deal with the corner. By defining $A_{1,1}\colon (q,p) \in \mathbb{R}^2 \to (q-p, q+p)/\sqrt{2}$, $\tilde{\Delta}_{1,1} := ((-\beta' + a_1 - a_1)/\sqrt{2}, (\beta' + a_1 - a_1)/\sqrt{2})$ and the Lipschitz function

$$a_{1,1}\colon t \in \tilde{\Delta}_{1,1} \to \frac{a_1 + a_1}{\sqrt{2}} - \left|t - \frac{a_1 - a_1}{\sqrt{2}}\right|,$$

where the unsimplified parts were left as signs of things to come, we have:

$$\begin{aligned}
&A_{1,1}^{-1}\left(\{(t,y)\colon t\in\tilde{\Delta}_{1,1}, y=a_{1,1}(t)\}\right)\\
&=\{(q,p)\in\mathbb{R}^2\colon q-p\in(-\beta',\beta'), q+p=2a_1-|q-p|\}\\
&=\{(q,p)\in\mathbb{R}^2\colon q-p\in(-\beta',\beta'),\\
&\qquad\qquad (q>p\wedge q=a_1)\vee(q=p\wedge q=a_1)\vee(q<p\wedge p=a_1)\}\\
&=(\Lambda_1\times U_1^{(-\beta',0)})\cup(\Lambda_1\times\Lambda_1)\cup(U_1^{(-\beta',0)}\times\Lambda_1).
\end{aligned}$$

The same kind of calculation renders

$$\begin{aligned}
&A_{1,1}^{-1}\left(\{(t,y)\colon t\in\tilde{\Delta}_{1,1}, y-a_{1,1}(t)\in(-\beta,0)\}\right)\\
&=\{(q,p)\in\mathbb{R}^2\colon q-p\in(-\beta',\beta'), q+p-2a_1+|q-p|\in(-\sqrt{2}\beta,0)\}\\
&=\{(q,p)\in\mathbb{R}^2\colon q-p\in(0,\beta'), q-a_1\in(-\beta/\sqrt{2},0)\}\\
&\quad\cup\{(q,p)\in\mathbb{R}^2\colon q-p\in\{0\}, q-a_1\in(\beta/\sqrt{2},0)\}\\
&\quad\cup\{(q,p)\in\mathbb{R}^2\colon q-p\in(-\beta',0), p-a_1\in(-\beta/\sqrt{2},0)\}\\
&\subset(U_1^{(-\beta/\sqrt{2},0)}\times U_1^{(-\beta,0)})\cup(U_1^{(-\beta/\sqrt{2},0)}\times U_1^{(-\beta/\sqrt{2},0)})\cup(U_1^{(-\beta,0)}\times U_1^{(-\beta/\sqrt{2},0)})\\
&\subset E\times E
\end{aligned}$$

thanks to the fact that $\beta/\sqrt{2}+\beta'\leq\beta$.

In an analogous way we can prove that

$$A_{1,1}^{-1}\left(\{(t,y)\colon t\in\tilde{\Delta}_{1,1}, y-a_{1,1}(t)\in(0,\beta)\}\right)\subset\overline{E\times E}^c.$$

2.3.3. Any two bounded Lipschitz domains. Let $E_1\subset\mathbb{R}^{n_1}$ and $E_2\subset\mathbb{R}^{n_2}$ be bounded Lipschitz domains. Then, according to Definition 2.16, for $i=1,2$, there exist $m_i\in\mathbb{N}$, $\beta_i>0$, and for $r\in[m_i]$ rigid transformations $A_{i;r}$, open sets $\tilde{\Delta}_{i;r}\subset\mathbb{R}^{n_i-1}$ and Lipschitz functions $a_{i;r}\colon\tilde{\Delta}_{i,j}\to\mathbb{R}$ such that the boundary of E_i can be written as

$$\partial E_i=\bigcup_{r=1}^{m_i}\Lambda_{i;r}$$

and

$$U_{i;r}^{(-\beta_i,0)}\subset E_i\quad\text{and}\quad U_{i;r}^{(0,\beta_i)}\subset\overline{E_i}^c,\ r\in[m_i].$$

Let $n:=n_1+n_2$, $\beta:=\min(\beta_1,\beta_2)$ and β' and γ be chosen so that $0<\gamma<\beta'\leq(1-1/\sqrt{2})\beta$. For $i=1,2$, let $U_{i;0}:=\{x\in E_i\colon\operatorname{dist}(x,\partial E_i)>\gamma\}$. We have that

$$E_i:=U_{i;0}\cup\bigcup_{r=1}^{m_i}U_{i;r}^{(-\beta',0)}.$$

Then, the above representations of ∂E_i and E_i allow for writing the boundary of $E_1 \times E_2$ as

$$\partial(E_1\times E_2) = \bigcup_{r=1}^{m_1}(\Lambda_{1;r}\times U_{2;0}) \\ \cup \bigcup_{r=1}^{m_1}\bigcup_{s=1}^{m_2}\Big((\Lambda_{1;r}\times U_{2;s}^{(-\beta',0)})\cup(\Lambda_{1;r}\times\Lambda_{2;s})\cup(U_{1;r}^{(-\beta',0)}\times\Lambda_{2;s})\Big)\cup\bigcup_{s=1}^{m_2}(U_{1;0}\times\Lambda_{2;s}). \quad (2.28)$$

Let us consider a piece of the form $\Lambda_{1;r}\times U_{2;0}$ out of the decomposition (2.28) and define the invertible transformation $A_{(r,0)}\colon (q,p)\in\mathbb{R}^n \to (p, A_{1;r}(q))\in\mathbb{R}^n$, the open set $\tilde\Delta_{(r,0)} := U_{2;0}\times\tilde\Delta_{1;r}\subset\mathbb{R}^{n-1}$ and the Lipschitz function $a_{(r,0)}\colon (p,q')\in\tilde\Delta_{(r,0)}\to a_{1;r}(q')\in\mathbb{R}$. We note that $A_{(r,0)}$ is affine and has its Jacobian determinant identical to either 1 or -1. These objects beget, for any $I\subset\mathbb{R}$:

$$\begin{aligned} U_{(r,0)}^I &= A_{(r,0)}^{-1}\Big(\{(p,q)\colon (p,q')\in\tilde\Delta_{(r,0)}, q_{n_1}-a_{(r,0)}(p,q')\in I\}\Big)\\ &= \{(A_{1;r}^{-1}(q),p)\colon q'\in\tilde\Delta_{1;r}, q_{n_1}-a_{1;r}(q')\in I, p\in U_{2;0}\}\\ &= U_{1;r}^I\times U_{2;0}. \end{aligned}$$

Therefore,

$$\Lambda_{(r,0)} := U_{(r,0)}^{\{0\}} = \Lambda_{1;r}\times U_{2;0},$$

$$U_{(r,0)}^{(-\beta,0)} = U_{1;r}^{(-\beta,0)}\times U_{2;0}\subset E_1\times E_2 \quad\text{and}\quad U_{(r,0)}^{(0,\beta)} = U_{1;r}^{(0,\beta)}\times U_{2;0}\subset \overline{E_1\times E_2}^c.$$

Thus, we have covered the piece of $\partial(E_1\times E_2)$ we were interested in, in a way that is good enough for our purposes, as we will see below.

Pieces of the form $U_{1;0}\times\Lambda_{2;s}$ can be described analogously.

Let us consider now a piece of the form $(\Lambda_{1;r}\times U_{2;s}^{(-\beta',0)})\cup(\Lambda_{1;r}\times\Lambda_{2;s})\cup(U_{1;r}^{(-\beta',0)}\times\Lambda_{2;s})$. Let us define

$$A_{(r,s)}\colon (q,p)\in\mathbb{R}^n\to\Big(q',p',\frac{q_{n_1}-p_{n_2}}{\sqrt{2}},\frac{q_{n_1}+p_{n_2}}{\sqrt{2}}\Big),$$

$$\tilde\Delta_{(r,s)} := \Big\{(q',p',t)\in\mathbb{R}^{n-1}\colon q'\in\tilde\Delta_{1;r}, p'\in\tilde\Delta_{2;s}, t-\frac{a_{1;r}(q')-a_{2;s}(p')}{\sqrt{2}}\in\Big(\frac{-\beta'}{\sqrt{2}},\frac{\beta'}{\sqrt{2}}\Big)\Big\},$$

$$a_{(r,s)}\colon (q',p',t)\in\tilde\Delta_{(r,s)}\to\frac{a_{1;r}(q')+a_{2;s}(p')}{\sqrt{2}}-\Big|t-\frac{a_{1;r}(q')-a_{2;s}(p')}{\sqrt{2}}\Big|\in\mathbb{R}.$$

Again, $A_{(r,s)}$ is an affine transformation with Jacobian determinant identical to either 1 or -1, $\tilde\Delta_{(r,s)}$ is an open subset of $\mathbb{R}^{n-1}$ (for it is the inverse image of $(-\beta'/\sqrt{2},\beta'/\sqrt{2})$ under a Lipschitz—and hence continuous—function) and $a_{(r,s)}$ is Lipschitz.

Now, given $I\subset\mathbb{R}$ let us characterize the points $(q,p)\in U_{(r,s)}^I$; that is, those points in $\mathbb{R}^n$ whose image under $A_{(r,s)}$, which we denote by (q',p',t,y), complies with

$$(q',p',t)\in\tilde\Delta_{(r,s)},\quad y-a_{(r,s)}\in I.$$

Note that q' as the ensemble of the first $n_1 - 1$ components of q and q' as a component of (q', p', t, y) are one and the same, due to the definition of $A_{(r,s)}$; similarly for p'. The above memberships translate into

$$q' \in \tilde{\Delta}_{1;r}, \quad p' \in \tilde{\Delta}_{2;s}, \quad q_{n_1} - a_{1;r}(q') - (p_{n_2} - a_{2;s}(p')) \in (-\beta', \beta'),$$
$$q_{n_1} - a_{1;r}(q') + p_{n_2}(p') - a_{2;s}(p') + |q_{n_1} - a_{1;r}(q') - (p_{n_2} - a_{2;s}(p'))| \in \sqrt{2}I.$$

Then, decomposing $(-\beta', \beta')$ into its negative, zero and positive parts we have that

$$\begin{aligned}\left(q_{n_1} - a_{1;r}(q') \in \frac{I}{\sqrt{2}}, p_{n_2} - a_{2;s}(p') \in \left(\frac{\inf I}{\sqrt{2}} - \beta', \frac{\sup I}{\sqrt{2}}\right)\right) \\ \vee \left(q_{n_1} - a_{1;r}(q') \in \frac{I}{\sqrt{2}}, p_{n_2} - a_{2;s}(p') \in \frac{I}{\sqrt{2}}\right) \\ \vee \left(p_{n_2} - a_{2;s}(p') \in \frac{I}{\sqrt{2}}, q_{n_1} - a_{1;r}(q') \in \left(\frac{\inf I}{\sqrt{2}} - \beta', \frac{\sup I}{\sqrt{2}}\right)\right).\end{aligned}$$

Importantly, these inclusions, together with $(q', p') \in \tilde{\Delta}_{1;r} \times \tilde{\Delta}_{2;s}$, characterize $U^I_{(r,s)}$ if I is a singleton. Therefore,

$$\Lambda_{(r,s)} = (\Lambda_{1;r} \times U^{(-\beta',0)}_{2;s}) \cup (\Lambda_{1;r} \times \Lambda_{2;s}) \cup (U^{(-\beta',0)}_{1;r} \times \Lambda_{2;s}),$$
$$U^{(-\beta,0)}_{(r,s)} \subset (U^{(-\beta/\sqrt{2},0)}_{1;r} \times U^{(-\beta,0)}_{2;s}) \cup (U^{(-\beta/\sqrt{2},0)}_{1;r} \times U^{(-\beta/\sqrt{2},0)}_{2;s}) \cup (U^{(-\beta,0)}_{1;r} \times U^{(-\beta/\sqrt{2},0)}_{2;s}),$$
$$U^{(0,\beta)}_{(r,s)} \subset (U^{(0,\beta/\sqrt{2})}_{1;r} \times U^{(-\beta',\beta/\sqrt{2})}_{2;s}) \cup (U^{(0,\beta/\sqrt{2})}_{1;r} \times U^{(0,\beta/\sqrt{2})}_{2;s}) \cup (U^{(-\beta',\beta/\sqrt{2})}_{1;r} \times U^{(0,\beta/\sqrt{2})}_{2;s}),$$

where the fact that $\beta/\sqrt{2} + \beta' \leq \beta$ has been used. In this way $\Lambda_{(r,s)}$ is exactly the part of the boundary of $E_1 \times E_2$ we are interested on, $U^{(-\beta,0)}_{(r,s)}$ is contained in $E_1 \times E_2$ and $U^{(0,\beta)}_{(r,s)}$ is contained in $\overline{E_1 \times E_2}^c$.

Recapping, for $(r, s) \in (\{0, \ldots, m_1\} \times \{0, \ldots, m_2\}) \setminus \{(0, 0)\}$, we have affine functions $A_{(r,s)}$, open sets $\tilde{\Delta}_{(r,s)}$, and Lipschitz functions $a_{(r,s)} \colon \tilde{\Delta}_{(r,s)} \to \mathbb{R}$ and a number $\beta > 0$ such that

$$\partial(E_1 \times E_2) = \bigcup_{\substack{r \in \{0,\ldots,m_1\} \\ s \in \{0,\ldots,m_2\} \\ (r,s) \neq (0,0)}} \Lambda_{(r,s)}, \quad U^{(-\beta,0)}_{(r,s)} \subset E_1 \times E_2 \quad \text{and} \quad U^{(0,\beta)}_{(r,s)} \subset \overline{E_1 \times E_2}^c;$$

that is, we have proved that $E_1 \times E_2$ is a bounded Lipschitz domain.

Finally we note that, as E_1 and E_2 need not have the same dimensionality, this argument also builds a Lipschitz description of a Cartesian product of any finite number of bounded Lipschitz domains. So, we have a constructive proof of:

Lemma 2.20. *Let, for $l \in [L]$, $\Omega_l \subset \mathbb{R}^{n_l}$ be a bounded Lipschitz domain. Then, the Cartesian product*

$$\bigtimes_{l=1}^{L} \Omega_l \subset \mathbb{R}^{\sum_{l=1}^{L} n_l}$$

is a bounded Lipschitz domain as well.

Remark 2.21. An alternative proof of the Lipschitz-regularity of the Cartesian product of bounded Lipschitz domains follows by combining Theorem 3.1 in the Ph.D. Thesis of Reinhard Hochmuth: *Randwertproblem einer nicht hypoelliptischen linearen partiellen Differentialgleichung. Dissertation, Freie Universitat Berlin, 1989*, which implies that the Cartesian product of a finite number of bounded domains, each satisfying the uniform cone property, is a bounded domain satisfying the uniform cone property, and Theorem 1.2.2.2 in the book of Grisvard [**Gri85**], which states that a bounded open set in $\mathbb{R}^n$ has the uniform cone property if, and only if, its boundary is Lipschitz.

2.3.4. Application to the tensor product of decreasing functions. Let us suppose now that there exist, for $i = 1, 2$, positive functions $w_i \in \mathrm{C}(\overline{E_i})$ and constants $\beta_i^* \in (0, \beta_i]$ such that, for all $r \in [m_i]$,

$$w_i(A_{i;r}^{-1}(q', q_{n_i} - \lambda)) \geq w_i(A_{i;r}^{-1}(q', q_{n_i})) \tag{2.29}$$

whenever

$$q' \in \tilde{\Delta}_{i;r} \quad \text{and} \quad -\beta_i^* < (q_{n_i} - \lambda) - a_{i;r}(q') < q_{n_i} - a_{i;r}(q') < 0.$$

In other words, the function $w_i \circ A_{1;r}^{-1}$ is monotonic decreasing with respect to its last component within a band of width β_i^* immediately below the graph of $a_{i;r}$ with respect to $\tilde{\Delta}_{i;r}$.

We would like to show that a monotonicity such as described in (2.29) still holds for the function $w := w_1 \otimes w_2$ defined over $E_1 \times E_2$ by $w(q, p) := w_1(q)w_2(p)$ with respect to the Lipschitz description of $\partial(E_1 \times E_2)$ that we constructed in the previous subsection.

Let $\beta^* := \min(\beta_1^*, \beta_2^*)$. We pick the constant β' of the previous subsection (which is used to describe the corner regions of $\partial(E_1 \times E_2)$) so that $0 < \beta' \leq (1 - 1/\sqrt{2})\beta^*$. We can do so because $\beta^* \leq \beta = \min(\beta_1, \beta_2)$.

Let us first consider a part of $\partial(E_1 \times E_2)$ of the form $\Lambda_{1;r} \times U_{2;0}$ for some $r \in [m_1]$, which were described above by the corresponding $A_{(r,0)}$, $\tilde{\Delta}_{(r,0)}$ and $a_{(r,0)}$. If $(p, q') \in \tilde{\Delta}_{(r,0)}$ and q_{n_1} and λ satisfy

$$-\beta^* < q_{n_1} - \lambda - a_{(r,0)}(p, q') < q_{n_1} - a_{(r,0)}(p, q') < 0,$$

we have that

$$-\beta_1^* < q_{n_1} - \lambda - a_{1;r}(q') < q_{n_1} - a_{1;r}(q') < 0 \quad \text{and} \quad p \in U_{2;0}.$$

So,

$$w(A^{-1}_{(r,0)}(p,q',q_{n_1}-\lambda)) = w_1(A^{-1}_{1;r}(q',q_{n_1}-\lambda))w_2(p) \\ \geq w_1(A^{-1}_{1;r}(q',q_{n_1}))w_2(p) = w(A^{-1}_{(r,0)}(p,q',q_{n_1})).$$

This works analogously for the parts of $\partial(E_1 \times E_2)$ with the form $U_{1;0} \times \Lambda_{2;s}$ for some $s \in [m_2]$.

Let us now consider a part of $\partial(E_1 \times E_2)$ described by $A_{(r,s)}$, $\tilde{\Delta}_{(r,s)}$ and $a_{(r,s)}$ for some $(r,s) \in ([m_1] \times [m_2]) \setminus \{(0,0)\}$. Then, for $(q',p',t) \in \tilde{\Delta}_{(r,s)}$ and y and λ satisfying

$$-\beta^* < y - \lambda - a_{(r,s)}(q',p',t) < y - a_{(r,s)}(q',p',t) < 0,$$

we have, after some algebra and using the fact that $\beta' + \beta^*/\sqrt{2} \leq \beta^*$, that

$$-\beta_1^* < \frac{y+t}{\sqrt{2}} - \frac{\lambda}{\sqrt{2}} - a_{1;r}(q') < \frac{y+t}{\sqrt{2}} - a_{1;r}(q') < 0,$$
$$-\beta_2^* < \frac{y-t}{\sqrt{2}} - \frac{\lambda}{\sqrt{2}} - a_{2;s}(p') < \frac{y-t}{\sqrt{2}} - a_{2;s}(p') < 0.$$

Thus

$$\begin{aligned} &w(A^{-1}_{(r,s)}(q',p',t,y-\lambda)) \\ &\quad = w_1\Big(A^{-1}_{1;r}\big(q',(y+t-\lambda)/\sqrt{2}\big)\Big)w_2\Big(A^{-1}_{2;s}\big(p',(y-t-\lambda)/\sqrt{2}\big)\Big) \\ &\quad \geq w_1\Big(A^{-1}_{1;r}\big(q',(y+t)/\sqrt{2}\big)\Big)w_2\Big(A^{-1}_{2;s}\big(p',(y-t)/\sqrt{2}\big)\Big) \\ &\quad = w(A^{-1}_{(r,s)}(q',p',t,y)), \end{aligned}$$

and so, we have proved that the property (2.29) is preserved by the tensor product of the w_i.

We conclude this subsection with two remarks. The first is that w inherits the properties of uniform continuity (i.e., $w \in \mathrm{C}(\overline{E_1 \times E_2})$) and positivity from its factors. The second is that this construction, as the one in the previous subsection, will work for tensor products of more than two functions that comply with the property described in (2.29).

2.4. Tensorization of properties of weighted Sobolev spaces

2.4.1. General properties of tensor products.

Lemma 2.22. *Suppose that $T \in \mathcal{D}'(\mathsf{D})$ is a distribution such that*

$$T\left(\bigotimes_{i=1}^{N} \varphi^{(i)}\right) = 0 \qquad \forall\, (\varphi^{(1)},\dots,\varphi^{(N)}) \in \bigtimes_{i\in[N]} \mathrm{C}_0^\infty(D_i).$$

Then, $T = 0$ in $\mathcal{D}'(\mathsf{D})$.

Further, for any ensemble of sequences of distributions $\left(R_n^{(i)} : n \geq 1\right)$, $i \in [N]$, *with* $R_n^{(i)} \in \mathcal{D}'(D_i)$ *and such that* $\lim_{n\to\infty} R_n^{(i)} = R^{(i)}$ *in* $\mathcal{D}'(D_i)$ *for* $i \in [N]$, *we have that*

$$\lim_{n\to\infty} \bigotimes_{i\in[N]} R_n^{(i)} = \bigotimes_{i\in[N]} R^{(i)} \quad \textit{in} \quad \mathcal{D}'(\mathsf{D}).$$

Proof. These are standard results from the theory of distributions, so we omit the proofs and refer the reader to Section 1.3.2 of the book of Vladimirov [**Vla02**], for example. □

Lemma 2.23. *The following statements hold:*

(1) For any ensemble $r^{(i)} \in \mathrm{H}(D_i; M_i)$, $i \in [N]$, $\bigotimes_{i\in[N]} r^{(i)} \in \mathrm{H}(\mathsf{D}; \mathsf{M})$.

(2) Suppose that $r^{(i)} : D_i \to \mathbb{R}$, $i \in [N]$, *are measurable functions. Then, the next two statements are equivalent:*

(a) $r^{(i)} \in \mathrm{H}(D_i; M_i) \setminus \{0\}$ *for all* $i \in [N]$;

(b) $\bigotimes_{i\in[N]} r^{(i)} \in \mathrm{H}(\mathsf{D}; \mathsf{M}) \setminus \{0\}$.

Proof. (1) It is immediate from the factorization of M that $\bigotimes_{i=1}^N r^{(i)}$ belongs to $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$. Thanks to Lemma 2.22, the identity

$$\nabla_{\boldsymbol{q}_j} \left(\frac{\bigotimes_{i=1}^N r^{(i)}}{\mathsf{M}} \right) = \bigotimes_{\substack{i=1 \\ i\neq j}}^N \left(\frac{r^{(i)}}{M_i} \right) \otimes_j \nabla \left(\frac{r^{(j)}}{M_j} \right) \tag{2.30}$$

holds in the distributional sense. Then, as $r^{(i)}/M_i \in \mathrm{L}^2_{M_i}(D_i)$ for $i \in [N] \setminus \{j\}$, and $\nabla(r^{(j)}/M_j) \in [\mathrm{L}^2_{M_j}(D_j)]^d$, the factorization of the Maxwellian M allows for stating that, for $j \in [N]$,

$$\nabla_{\boldsymbol{q}_j} (\bigotimes_{i=1}^N r^{(i)}/\mathsf{M}) \in [\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})]^d.$$

That completes the proof of Part (1).

(2) We shall prove the second part by showing that (b) is both necessary and sufficient for (a).

(a) $\implies$ (b): This is immediate from the first part and the fact that the tensor product of the $r^{(i)}$, $i \in [N]$, cannot be null if none of its factors is.

(b) $\implies$ (a): Suppose that $\bigotimes_{i=1}^N r^{(i)} \in \mathrm{H}(\mathsf{D}; \mathsf{M}) \setminus \{0\}$; then, because of the tensor-product structure of M, the positivity of M_i on compact subsets of D_i for $i \in [N]$ and Fubini's theorem, $r^{(i)} \in M_i \, \mathrm{L}^1_{\mathrm{loc}}(D_i) \cap \mathrm{L}^2_{1/M_i}(D_i)$, $i \in [N]$. Hence, each $r^{(i)}/M_i$ defines a regular distribution in $\mathcal{D}'(D_i)$. Again, Lemma 2.22 makes (2.30) valid and thus,

$$\left\| \bigotimes_{i=1}^N r^{(i)} \right\|^2_{\mathrm{H}(\mathsf{D};\mathsf{M})} = \prod_{i=1}^N \left\| r^{(i)} \right\|^2_{\mathrm{L}^2_{1/M_i}(D_i)} + \sum_{j=1}^N \left(\prod_{\substack{i=1 \\ i\neq j}}^N \left\| r^{(i)} \right\|^2_{\mathrm{L}^2_{1/M_i}(D_i)} \right) \left\| \nabla(r^{(j)}/M_j) \right\|^2_{[\mathrm{L}^2_{M_j}(D_i)]^d}. \tag{2.31}$$

Now, none of the $r^{(i)}$ can be null (otherwise their tensor product would be null). On combining this with their $1/M_i$-weighted square integrability, the identity (2.31) renders

$$\left\|\nabla(r^{(i)}/M_i)\right\|^2_{[\mathrm{L}^2_{M_i}(D_i)]^d} < \infty \qquad \text{for all } i \in [N].$$

Hence $r^{(i)} \in \mathrm{H}(D_i; M_i) \setminus \{0\}$ for $i \in [N]$. □

2.4.2. Tensorization of compactness embeddings.

Lemma 2.24. *The space* $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ *is compactly embedded in* $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ *and* $\mathrm{H}(\mathsf{D};\mathsf{M})$ *is compactly embedded in* $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$.

Proof. Throughout this proof we will assume, for ease of exposition, that $N = 2$; the argument carries over to higher N without difficulties. Let $u \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$. As by (1.25) $\mathsf{M} = M_1 \otimes M_2$, it follows from Fubini's theorem that, for almost all $\boldsymbol{q}_1 \in D_1$,

$$u(\boldsymbol{q}_1, \cdot) \in \mathrm{L}^1_{\mathrm{loc}}(D_2) \quad \text{and} \quad \partial_\alpha u(\boldsymbol{q}_1, \cdot) \in \mathrm{L}^2_{M_2}(D_2),$$

where α is any multi-index in $[\mathbb{N}_0]^d$ with $0 \le |\alpha| \le 1$. Fubini's theorem, again, ensures that, given $\varphi_2 \in \mathrm{C}^\infty_0(D_2)$ and $\alpha_2 \in [\mathbb{N}_0]^d$, $0 \le \alpha_2 \le 1$,

$$\int_{D_1} \left[(-1)\int_{D_2} u(\boldsymbol{q}_1, \cdot)\partial_{\alpha_2}\varphi_2\right]\varphi_1 \,\mathrm{d}\boldsymbol{q}_1 = \int_{D_1}\left[\int_{D_2} \partial_{(0,\alpha_2)} u(\boldsymbol{q}_1, \cdot)\varphi_2\right]\varphi_1 \,\mathrm{d}\boldsymbol{q}_1,$$

for all $\varphi_1 \in \mathrm{C}^\infty_0(D_1)$. Therefore, $\partial_{\alpha_2}[u(\boldsymbol{q}_1, \cdot)] = \partial_{(0,\alpha_2)} u(\boldsymbol{q}_1, \cdot)$ in the weak sense on D_2 for almost all $\boldsymbol{q}_1 \in D_1$. As $\partial_{(0,\alpha_2)} u(\boldsymbol{q}_1, \cdot)$ lies in $\mathrm{L}^2_{M_2}(D_2)$ for almost all $\boldsymbol{q}_1 \in D_1$ we have that

$$u(\boldsymbol{q}_1, \cdot) \in \mathrm{H}^1_{M_2}(D_2) \quad \text{for almost all } \boldsymbol{q}_1 \in D_1. \tag{2.32}$$

In the same way it can be proved that

$$u(\cdot, \boldsymbol{q}_2) \in \mathrm{H}^1_{M_1}(D_1) \quad \text{for almost all } \boldsymbol{q}_2 \in D_2.$$

Let us define, for $i \in \{1, 2\}$, the sequence $\big(D_{i,(n)} \colon n \ge 1\big)$ of bounded and proper subsets of D_i by $D_{i,(n)} := B\big(0, \frac{\sqrt{b_i}n}{n+1}\big)$. Then,

$$D_{i,(n)} \subset D_{i,(n+1)},\ n \in \mathbb{N}, \qquad \bigcup_{n=1}^{\infty} D_{i,(n)} = D_i \qquad \text{and} \quad \mathrm{H}^1_{M_i}(D_{i,(n)}) \Subset \mathrm{L}^2_{M_i}(D_{i,(n)}).$$

This last relation is a consequence of the corresponding relation for the unweighted case, $\mathrm{H}^1(D_{i,(n)}) \Subset \mathrm{L}^2(D_{i,(n)})$—in turn a consequence of the boundedness and Lipschitz continuity of $D_{i,(n)}$—on account of the existence of positive lower and upper bounds for M_i on $D_{i,(n)}$, whereupon, via Proposition 2.5, there is algebraic and topological equivalence between $\mathrm{H}^1_{M_i}(D_{i,(n)})$ and $\mathrm{H}^1(D_{i,(n)})$ and between $\mathrm{L}^2_{M_i}(D_{i,(n)})$ and $\mathrm{L}^2(D_{i,(n)})$.

Letting, for $n \in \mathbb{N}$, $\mathsf{D}_{(n)} := \times_{i=1}^2 D_{i,(n)} \subsetneq \mathsf{D}$, the above properties get inherited:

$$\mathsf{D}_{(n)} \subset \mathsf{D}_{(n+1)},\ n \in \mathbb{N}, \qquad \bigcup_{n=1}^{\infty} \mathsf{D}_{(n)} = \mathsf{D} \qquad \text{and} \quad \mathrm{H}^1_{\mathsf{M}}(\mathsf{D}_{(n)}) \Subset \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}_{(n)}).$$

The third statement follows from the fact that the $\mathsf{D}_{(n)}$, being Cartesian products of bounded Lipschitz domains, are also bounded Lipschitz domains (this is given by Lemma 2.20 or Remark 2.21[2]). Let us define $D_i^{(n)} := D_i \setminus D_{i,(n)}$ and $\mathsf{D}^{(n)} := \mathsf{D} \setminus \mathsf{D}_{(n)}$. Thanks to [**OK90**, Theorem 17.6], the above compact embeddings on members of a nested covering imply the following characterizations (the first, for $i \in \{1,2\}$):

$$\mathrm{H}^1_{M_i}(D_i) \Subset \mathrm{L}^2_{M_i}(D_i) \iff \lim_{n\to\infty} \sup_{u\in \mathrm{H}^1_{M_i}(D_i)\setminus\{0\}} \int_{D_i^{(n)}} u^2 M_i \,/\, \|u\|^2_{\mathrm{H}^1_{M_i}(D_i)} = 0, \tag{2.33}$$

$$\mathrm{H}^1_{\mathsf{M}}(\mathsf{D}) \Subset \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}) \iff \lim_{n\to\infty} \sup_{u\in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})\setminus\{0\}} \int_{\mathsf{D}^{(n)}} u^2 \mathsf{M} \,/\, \|u\|^2_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})} = 0. \tag{2.34}$$

From Hypothesis B, the left-hand side of (2.33) holds; hence, its right-hand side also holds. Using (2.32) and (2.33) with $i = 2$, we deduce that for each $\varepsilon > 0$ there exists an $\tilde{n} = \tilde{n}(\varepsilon) \in \mathbb{N}$ such that $n \geq \tilde{n}$ implies

$$\begin{aligned}
\int_{D_1\times D_2^{(n)}} u^2 \mathsf{M} &= \int_{D_1} \left[\int_{D_2^{(n)}} u^2(\boldsymbol{q}_1,\cdot) M_2 \right] M_1(\boldsymbol{q}_1)\,\mathrm{d}\boldsymbol{q}_1 \\
&\leq \varepsilon \int_{D_1} \|u(\boldsymbol{q}_1,\cdot)\|^2_{\mathrm{H}^1_{M_2}(D_2)}\, M_1(\boldsymbol{q}_1)\,\mathrm{d}\boldsymbol{q}_1 \\
&= \varepsilon \int_{D_1} \left[\int_{D_2} u^2(\boldsymbol{q}_1,\cdot) M_2 + \int_{D_2} \left|\nabla_{\boldsymbol{q}_2} u(\boldsymbol{q}_1,\cdot)\right|^2 M_2 \right] M_1(\boldsymbol{q}_1)\,\mathrm{d}\boldsymbol{q}_1 \\
&\leq \varepsilon \,\|u\|^2_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})}\,.
\end{aligned}$$

An analogous result can be proved for the M-weighted integral of u^2 on $D_1^{(n)} \times D_2$. Then, since $\mathsf{D}^{(n)} = (D_1 \times D_2^{(n)}) \cup (D_1^{(n)} \times D_2)$ and $u \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ is arbitrary, the right-hand side of (2.34) holds; hence, so does its left-hand side.

Finally, the embedding $\mathrm{H}(\mathsf{D};\mathsf{M}) \Subset \mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ follows directly from the embedding $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D}) \Subset \mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ on account of the isometric isomorphism (2.4). □

2.4.3. Tensorization of the density of smooth functions. We start with a lemma concerning the density of smooth functions in weighted Sobolev spaces defined on Lipschitz domains.

Lemma 2.25.

(1) Let $E \subset \mathbb{R}^n$ be a bounded Lipschitz domain with Lipschitz representation given by $\{(A_r, \tilde{\Delta}_r, a_r)\}_{r=1}^m$ with margin β and let $w \in \mathrm{C}(\overline{E})$ be positive in E and complying with the following condition: There exist $c_0, \beta^* \in \mathbb{R}$ such that $0 < c_0 < \beta^* \leq \beta$ and,

[2]A third proof, valid in the special case of the domain $\mathsf{D}_{(n)}$ consists of noting that, as a Cartesian product of bounded open convex sets, $\mathsf{D}_{(n)}$ is a bounded open convex set in $\mathbb{R}^n$ (cf. [**HUL01**], p. 23), and then applying Corollary 1.2.2.3 in Grisvard [**Gri85**], which states that a bounded open convex set in $\mathbb{R}^n$ has a Lipschitz boundary.

for all $r \in [m]$,

$$w(A_r^{-1}(q', q_n - \lambda)) \geq \frac{1}{4} w(A_r^{-1}(q', q_n)) \quad \textit{for} \quad \begin{cases} \lambda \in (0, \beta^* - c_0], \\ q' \in \tilde{\Delta}_r, \\ q_n \in (a_r(q') - c_0, a_r(q')). \end{cases} \tag{2.35}$$

Then, $\mathrm{C}^\infty(\overline{E})$ *is dense in* $\mathrm{H}^1_w(E)$.

(2) The set $\mathrm{C}^\infty(\overline{\mathsf{D}})$ *is dense in* $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$.

(3) The set $\mathsf{M}\,\mathrm{C}^\infty(\overline{\mathsf{D}})$ *is dense in* $\mathrm{H}(\mathsf{D};\mathsf{M})$.

Remark 2.26. The result of the part (1) of Lemma 2.25 is valid, in particular, if 1/4 is replaced by 1 in equation (2.35); i.e., if w is monotonic decreasing in a strip of width β^* along the boundary of E described by its Lipschitz representation.

Proof of Lemma 2.25. The argument for the first part closely follows the proof of Theorem 1.1 on p. 307 in the paper of Nečas [**Neč62**]. We start by emphasizing that, as $w^{-1} \in \mathrm{L}^1_{\mathrm{loc}}(E)$, $\mathrm{H}^1_w(E)$ is indeed a Banach space and, via part (b) of Lemma 2.4, $\mathrm{C}^\infty(\overline{E}) \subset \mathrm{H}^1_w(E)$. Since the Jacobian matrix of each mapping A_r is orthogonal with determinant 1, its specific choice does not affect the argument below. We shall therefore assume for ease of exposition that A_r is the identity mapping. Thus, for example, we shall write $u(q', q_n)$ instead of $u(A_r^{-1}(q', q_n))$.

There exist functions $\varphi_r \in \mathrm{C}^\infty_0(U_r^{(-\beta,\beta)})$, with $0 \leq \varphi_r \leq 1$, $r \in [m]$, and a function $\varphi_{m+1} \in \mathrm{C}^\infty_0(E)$, with $0 \leq \varphi_{m+1} \leq 1$, such that we have the partition of unity

$$\sum_{i=1}^{m+1} \varphi_r \equiv 1 \quad \text{on} \quad \overline{E}. \tag{2.36}$$

Now, given $u \in \mathrm{H}^1_w(E)$, let $u_r := u\varphi_r$ for $r \in [m+1]$; clearly, $u_r \in \mathrm{H}^1_w(E)$. We define $u_{r,\lambda}$ by $u_{r,\lambda}(q', q_n) := u_r(q', q_n - \lambda)$ for $r \in [m]$ and $u_{m+1,\lambda} := u_{m+1}$. We begin by showing that

$$\lim_{\lambda \to 0_+} u_{r,\lambda} = u_r \quad \text{in } \mathrm{H}^1_w(E), \quad r \in [m+1]. \tag{2.37}$$

For $r = m+1$, this is immediate. For a given $r \in [m]$, let g_r signify the function u_r or any of its first partial derivatives, and define $V_r := U_r^{(-\beta,0)}$. Clearly, $V_r \subset U_r^{(-\beta,\beta)}$. By the triangle

inequality, with $q = (q', q_n)$,

$$\begin{aligned}
&\left[\int_{V_r} |g_r(q', q_n) - g_r(q', q_n - \lambda)|^2 \, w(q) \, \mathrm{d}q\right]^{\frac{1}{2}} \\
&\quad = \left[\int_{V_r} |g_r(q', q_n) w(q)^{\frac{1}{2}} - g_r(q', q_n - \lambda) w(q)^{\frac{1}{2}}|^2 \, \mathrm{d}q\right]^{\frac{1}{2}} \\
&\quad \le \left[\int_{V_r} |g_r(q', q_n) w(q', q_n)^{\frac{1}{2}} - g_r(q', q_n - \lambda) w(q', q_n - \lambda)^{\frac{1}{2}}|^2 \, \mathrm{d}q\right]^{\frac{1}{2}} \\
&\qquad + \left[\int_{V_r} |g_r(q', q_n - \lambda)|^2 \, |w(q', q_n - \lambda)^{\frac{1}{2}} - w(q', q_n)^{\frac{1}{2}}|^2 \, \mathrm{d}q\right]^{\frac{1}{2}} \\
&\quad =: \mathrm{T}_1 + \mathrm{T}_2.
\end{aligned} \tag{2.38}$$

We begin by considering T_2. Let $\varepsilon > 0$ and $\lambda \in (0, \beta^* - c_0]$; then, from the hypotheses,

$$|w(q', q_n - \lambda)^{\frac{1}{2}} - w(q', q_n)^{\frac{1}{2}}|^2 \le w(q', q_n - \lambda) \quad \text{for} \quad \begin{cases} q' \in \tilde{\Delta}_r, \\ q_n \in (a_r(q') - c_0, a_r(q')). \end{cases}$$

Hence, and by the absolute continuity of the Lebesgue integral, there exists $c_1 \in (0, c_0]$ small enough and independent of λ such that

$$\begin{aligned}
&\int_{\tilde{\Delta}_r} \int_{a_r(q') - c_1}^{a_r(q')} |g_r(q', q_n - \lambda)|^2 \, |w(q', q_n - \lambda)^{\frac{1}{2}} - w(q', q_n)^{\frac{1}{2}}|^2 \, \mathrm{d}q_n \, \mathrm{d}q' \\
&\qquad\qquad \le \int_{\tilde{\Delta}_r} \int_{a_r(q') - c_1}^{a_r(q')} |g_r(q', q_n - \lambda)|^2 \, w(q', q_n - \lambda) \, \mathrm{d}q_n \, \mathrm{d}q' < \frac{1}{2}\varepsilon^2.
\end{aligned} \tag{2.39}$$

Now for $c_1 > 0$ fixed, the uniform continuity of w and the assumption of (2.26) guarantee the existence of $\lambda > 0$ sufficiently small such that

$$\begin{aligned}
&\int_{\tilde{\Delta}_r} \int_{a_r(q') - \beta}^{a_r(q') - c_1} |g_r(q', q_n - \lambda)|^2 \, |w(q', q_n - \lambda)^{\frac{1}{2}} - w(q', q_n)^{\frac{1}{2}}|^2 \, \mathrm{d}q_n \, \mathrm{d}q' \\
&\quad \le \max_{\substack{q' \in \tilde{\Delta}_r \\ -\beta \le q_n - a_r(q') \le -c_1}} \left|1 - \frac{w(q', q_n)^{\frac{1}{2}}}{w(q', q_n - \lambda)^{\frac{1}{2}}}\right|^2 \\
&\qquad \times \int_{\tilde{\Delta}_r} \int_{a_r(q') - \beta}^{a_r(q') - c_1} |g_r(q', q_n - \lambda)|^2 \, w(q', q_n - \lambda) \, \mathrm{d}q_n \, \mathrm{d}q' < \frac{1}{2}\varepsilon^2.
\end{aligned} \tag{2.40}$$

Summing (2.39) and (2.40) and taking the square root of both sides of the resulting inequality, we deduce that for any $\varepsilon > 0$ there exists $\lambda > 0$ such that

$$\left[\int_{V_r} |g_r(q', q_n - \lambda)|^2 \, |w(q', q_n - \lambda)^{\frac{1}{2}} - w(q', q_n)^{\frac{1}{2}}|^2 \, \mathrm{d}q\right]^{\frac{1}{2}} < \varepsilon.$$

Hence,

$$\lim_{\lambda\to 0_+}\left[\int_{V_r}|g_r(q',q_n-\lambda)|^2\,|w(q',q_n-\lambda)^{\frac{1}{2}}-w(q',q_n)^{\frac{1}{2}}|^2\,\mathrm{d}q\right]^{\frac{1}{2}}=0. \tag{2.41}$$

That concludes the analysis of term T_2.

Concerning T_1, continuity in the L^2-norm of the translation operator implies that

$$\lim_{\lambda\to 0_+}\left[\int_{V_r}|g_r(q',q_n)w(q',q_n)^{\frac{1}{2}}-g_r(q',q_n-\lambda)w(q',q_n-\lambda)^{\frac{1}{2}}|^2\,\mathrm{d}q\right]^{\frac{1}{2}}=0. \tag{2.42}$$

Finally, (2.41) and (2.42) imply (2.37).

Having shown (2.37), it now suffices to prove that each of the functions $u_{r,\lambda}$, $r\in[m+1]$, is a limit in $\mathrm{H}^1_w(E)$ of $\mathrm{C}^\infty(\overline{E})$ functions. To this end, we notice that due to (2.26) and the fact that $\mathrm{supp}(\varphi_{m+1})\Subset E$, the positivity of w renders $u_{r,\lambda}\in\mathrm{H}^1(E)$. As E is a bounded Lipschitz domain, $\mathrm{C}^\infty(\overline{E})$ is dense in $\mathrm{H}^1(E)$. Thus, each $u_{r,\lambda}$ is the limit in $\mathrm{H}^1(E)$ of a sequence $(u_{r,\lambda,k})_{k\geq 1}$ of $\mathrm{C}^\infty(\overline{E})$ functions. On noting that $w\in\mathrm{C}(\overline{E})\subset\mathrm{L}^\infty(E)$, we have that these limits can be taken in $\mathrm{H}^1_w(E)$. Then, from

$$\|u-\sum_{r=1}^{m+1}u_{r,\lambda,k}\|_{\mathrm{H}^1_w(E)}\leq\sum_{r=1}^{m+1}\Big(\|u_r-u_{r,\lambda}\|_{\mathrm{H}^1_w(E)}+\|u_{r,\lambda}-u_{r,\lambda,k}\|_{\mathrm{H}^1_w(E)}\Big)$$

we find that any $u\in\mathrm{H}^1_w(E)$ can be approximated arbitrarily closely by $\mathrm{C}^\infty(\overline{E})$ functions in the $\mathrm{H}^1_w(E)$ norm.

The second part follows from noting that each of the partial Maxwellians M_i and their corresponding partial Maxwellian weighted Sobolev spaces $\mathrm{H}^1_{M_i}(D_i)$, $i\in[N]$, comply with the hypotheses of the first part. To show that this is the case we start by fixing $i\in[N]$ and considering a point $\boldsymbol{x}$ in ∂D_i. Then, there exists a rotation $A_{\boldsymbol{x}}$ that applied to $\boldsymbol{x}$ gives $\sqrt{b_i}\boldsymbol{e}_d$ (the member of $\mathbb{R}^d$ whose only nonzero entry is the last, and is $\sqrt{b_i}$). Defining the open set $\tilde{\Delta}_{\boldsymbol{x}}:=B(0,\sqrt{b_i}/2)\subset\mathbb{R}^{d-1}$ and the Lipschitz function $a_{\boldsymbol{x}}\colon y\in\tilde{\Delta}_{\boldsymbol{x}}\to\sqrt{b_i-|y|^2}\in\mathbb{R}$ we have that $\Lambda_{\boldsymbol{x}}\subset\partial D_i$. Further, we have that

$$|A^{-1}_{\boldsymbol{x}}(q',t)|<|A^{-1}_{\boldsymbol{x}}(q',y)|\quad\text{for}\quad\begin{cases}q'\in\tilde{\Delta}_{\boldsymbol{x}},\\ t,y\in(a_{\boldsymbol{x}}(q')-\sqrt{b_i}/2,a_{\boldsymbol{x}}(q')+\sqrt{b_i}/2),\\ t<y.\end{cases} \tag{2.43}$$

Then, it follows immediately that $U^{(-\sqrt{b_i}/2,0)}_{\boldsymbol{x}}\subset D$ and $U^{(0,\sqrt{b_i}/2)}_{\boldsymbol{x}}\subset\big(\overline{D_i}\big)^c$. Noting that there exists a finite set $\mathcal{I}\subset\partial D_i$ so that $\partial D_i=\bigcup_{\boldsymbol{x}\in\mathcal{I}}\Lambda_{\boldsymbol{x}}$ we have that $\{(A_{\boldsymbol{x}},\tilde{\Delta}_{\boldsymbol{x}},a_{\boldsymbol{x}})\}_{\boldsymbol{x}\in\mathcal{I}}$ is a Lipschitz representation of D_i with margin $\sqrt{b_i}/2$.

Now, as the potential U_i is a monotonic increasing function, which tends to infinity as its argument tends to $b_i/2$ from below, the partial Maxwellian M_i belongs to $\mathrm{C}(\overline{D_i})$, is positive in D_i and obeys

$$M_i(\boldsymbol{p})\geq M_i(\tilde{\boldsymbol{p}})\quad\text{if}\quad|\boldsymbol{p}|\leq|\tilde{\boldsymbol{p}}|,$$

for all $\boldsymbol{p}$, $\tilde{\boldsymbol{p}} \in D_i$. Combining this with (2.43) gives

$$M_i(A_{\boldsymbol{x}}^{-1}(q',t)) \geq M_i(A_{\boldsymbol{x}}^{-1}(q',y)) \quad \text{for} \quad \begin{cases} \boldsymbol{x} \in \mathcal{I}, \\ q' \in \tilde{\Delta}_{\boldsymbol{x}}, \\ t, y \in (a_{\boldsymbol{x}}(q') - \sqrt{b_i}/2, a_{\boldsymbol{x}}(q')), \\ t < y. \end{cases}$$

As such a property holds for all the M_i, $i \in [N]$, we can use the results of Subsection 2.3.3 and Subsection 2.3.4 to deduce that there exists a Lipschitz representation $\{(A_r, \tilde{\Delta}_r, a_r)\}_{r=1}^m$ of the full configuration space D with margin $\beta := \min_{i\in[N]} \sqrt{b_i}/2$, and that, with respect to that Lipschitz description, the tensor product Maxwellian M, positive and belonging to $\mathrm{C}(\overline{\mathsf{D}})$, is monotonic decreasing. That is, for $r \in [m]$,

$$\mathsf{M}(A_r^{-1}(q',t)) \geq \mathsf{M}(A_r^{-1}(q',y)) \quad \text{for} \quad \begin{cases} q' \in \tilde{\Delta}_i \subset \mathbb{R}^{dN-1}, \\ t, y \in (a_i(q') - \beta, a_i(q')), \\ t < y. \end{cases}$$

As the condition above implies the condition (2.35), the desired result follows.

Finally, the third part comes from the relation (2.4) between $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ and $\mathrm{H}(\mathsf{D};\mathsf{M})$. □

Having identified the weighted Sobolev spaces that naturally feature in the analysis of the Fokker–Planck equation under consideration, we proved a number of basic results on them under the assumption of some structural hypotheses. We also went on to show that weights associated with concrete force laws used in the literature do indeed satisfy those structural hypotheses. We also studied some geometrical aspects of the Cartesian product of Lipschitz domains and showed how they affect certain basic properties of Sobolev spaces with tensor-product weights based on these Cartesian-product domains.

In the next chapter we will introduce the Separated Representation strategy and exploit their identification as Greedy Algorithms in the sense of the theory of nonlinear approximation.

CHAPTER 3

Continuous Separated Representation

In this chapter we will study two algorithms that describe the Separated Representation strategy as applied to the Fokker–Planck equation under consideration. Using the notion of greedy algorithm from the theory of nonlinear approximation it is possible to obtain certain convergence rates for the algorithms as long as the true solution lies in certain space with a very abstract definition. Up to this stage, the main arguments follow closely those found in [**LBLM09**] for the approximation of Poisson operators using the Separated Representation strategy, although some important details are particular to our degenerate case. Then, we then proceed to partially characterize the abstract space of guaranteed convergence rates in more familiar terms; namely, weighted summability of squared Fourier coefficients and weighted-Sobolev-type regularity.

3.1. Greedy algorithms

3.1.1. Two algorithms. The existence of a unique weak solution to (2.1) is an immediate consequence of the Lax–Milgram theorem via the facts that $\mathrm{H}(\mathsf{D};\mathsf{M})$ is a Hilbert space (cf. Lemma 2.3) and a is a bounded and coercive bilinear form on $\mathrm{H}(\mathsf{D};\mathsf{M})$ (cf. (2.3)). By virtue of the Riesz representation theorem, there exists a bounded linear operator $\mathcal{A}\colon \mathrm{H}(\mathsf{D};\mathsf{M}) \to \mathrm{H}(\mathsf{D};\mathsf{M})'$, defined by

$$(\mathcal{A}\psi)(\varphi) = a(\psi,\varphi) \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}).$$

Thanks to the symmetry of a, the weak formulation (2.1) can be restated as the following, equivalent, energy minimization problem:

$$\psi := \operatorname*{arg\,min}_{\varphi\in\mathrm{H}(\mathsf{D};\mathsf{M})} J_f(\varphi) \qquad \text{where} \qquad J_f(\varphi) := \frac{1}{2}\,a(\varphi,\varphi) - f(\varphi). \tag{3.1}$$

We observe that, with $\psi \in \mathrm{H}(\mathsf{D};\mathsf{M})$ as in (3.1),

$$J_f(\varphi) = \frac{1}{2}\,a(\varphi-\psi,\varphi-\psi) - \frac{1}{2}\,a(\psi,\psi) \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}). \tag{3.2}$$

Following the work of Le Bris, Lelièvre and Maday [**LBLM09**] concerning the numerical solution of high-dimensional Poisson equations, we consider two abstract algorithms.

Algorithm I *(Pure Greedy Algorithm).*

1. `Define:` $f_0 := f \in \mathrm{H}(\mathsf{D};\mathsf{M})'$.
2. `For` $n \geq 1$ `do:`

2.1 Find $r_n^{(i)} \in \mathrm{H}(D_i; M_i)$, $i \in [N]$, such that

$$(r_n^{(1)}, \dots, r_n^{(N)}) \in \operatorname*{arg\,min}_{(s^{(1)},\dots,s^{(N)}) \in \bigtimes_{i=1}^N \mathrm{H}(D_i;M_i)} \frac{1}{2} a\left(\bigotimes_{i=1}^N s^{(i)}, \bigotimes_{i=1}^N s^{(i)}\right) - f_{n-1}\left(\bigotimes_{i=1}^N s^{(i)}\right). \tag{3.3}$$

2.2 Define

$$f_n := f_{n-1} - \mathcal{A}\left(\bigotimes_{i=1}^N r_n^{(i)}\right).$$

2.3 If $\|f_n\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} \geq \mathtt{TOL}$, then proceed to iteration $n+1$; else, stop.

Algorithm II *(Orthogonal Greedy Algorithm)*.

1. Define: $f_0 := f \in \mathrm{H}(\mathsf{D};\mathsf{M})'$.
2. For $n \geq 1$ do:

2.1 Find $r_n^{(i)} \in \mathrm{H}(D_i; M_i)$, $i \in [N]$, such that

$$(r_n^{(1)}, \dots, r_n^{(N)}) \in \operatorname*{arg\,min}_{(s^{(1)},\dots,s^{(N)}) \in \bigtimes_{i=1}^N \mathrm{H}(D_i;M_i)} \frac{1}{2} a\left(\bigotimes_{i=1}^N s^{(i)}, \bigotimes_{i=1}^N s^{(i)}\right) - f_{n-1}\left(\bigotimes_{i=1}^N s^{(i)}\right). \tag{3.4}$$

2.2 Solve the following Galerkin problem on the span of $\left(\bigotimes_{i=1}^N r_k^{(i)} : k \in [n]\right)$:

$$\alpha^{(n)} := \operatorname*{arg\,min}_{\beta \in \mathbb{R}^n} \left\{ \frac{1}{2} a\left(\sum_{k=1}^n \beta_k \bigotimes_{i=1}^N r_k^{(i)}, \sum_{k=1}^n \beta_k \bigotimes_{i=1}^N r_k^{(i)}\right) - f\left(\sum_{k=1}^n \beta_k \bigotimes_{i=1}^N r_k^{(i)}\right) \right\}. \tag{3.5}$$

2.3 Define

$$f_n := f - \mathcal{A}\left(\sum_{k=1}^n \alpha_k^{(n)} \bigotimes_{i=1}^N r_k^{(i)}\right) \in \mathrm{H}(\mathsf{D};\mathsf{M})'.$$

2.4 If $\|f_n\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} \geq \mathtt{TOL}$, then proceed to iteration $n+1$; else, stop.

The approximations to the true solution ψ given by the above algorithms at iteration n are

$$\sum_{k=1}^n \bigotimes_{i=1}^N r_k^{(i)} \qquad \text{and} \qquad \sum_{k=1}^n \alpha_k^{(n)} \bigotimes_{i=1}^N r_n^{(i)}$$

for Algorithm I and Algorithm II, respectively. For future reference, we define $\delta_n \in \mathrm{H}(\mathsf{D};\mathsf{M})$ as the unique solution of the problem

$$a(\delta_n, \varphi) = f_n(\varphi) \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}).$$

Clearly, for all n up to the (existing or not) termination of the corresponding algorithm,

$$\delta_n = \begin{cases} \delta_{n-1} - \bigotimes_{i=1}^N r_n^{(i)} & \text{for the Pure Greedy Algorithm,} \\ \psi - \sum_{k=1}^n \alpha_k^{(n)} \bigotimes_{i=1}^N r_k^{(i)} & \text{for the Orthogonal Greedy Algorithm,} \end{cases} \tag{3.6}$$

where $\psi =: \delta_0$ is the unique solution of (3.1); i.e., the δ_n are the errors at the n-th iteration. Proving the convergence of the algorithms amounts to showing that the sequences $(\delta_n \colon n \geq 0)$ corresponding to the two algorithms converge to 0 in $\mathrm{H}(\mathsf{D};\mathsf{M})$.

3.1.2. Correctness of the algorithms. The proof of the correctness of Algorithm I (resp. Algorithm II) amounts to showing that, given $f_{n-1} \in \mathrm{H}(\mathsf{D};\mathsf{M})'$ (resp. $(f_{n-1}, \alpha^{(n-1)}) \in \mathrm{H}(\mathsf{D};\mathsf{M})' \times \mathbb{R}^{n-1}$), the loop 2 (resp. 2) returns a well-defined member of $\mathrm{H}(\mathsf{D};\mathsf{M})'$ (resp. $\mathrm{H}(\mathsf{D};\mathsf{M})' \times \mathbb{R}^n$).

We start by observing that, due to the first part of Lemma 2.23, the set of N-way tensor products of ensembles of functions $\mathrm{H}(D_i; M_i)$, $i \in [N]$, is a subset of $\mathrm{H}(\mathsf{D};\mathsf{M})$, thereby rendering the minimization *problems* (3.3) and (3.4) sound. However, the existence of *solutions* $(r_n^{(1)}, \ldots, r_n^{(N)})$ to these problems is quite another matter: it will be proved using Lemma 3.1 and Theorem 3.2 below.

Lemma 3.1. *Suppose that $f \in \mathrm{H}(\mathsf{D};\mathsf{M})' \setminus \{0\}$ and consider the functional J_f, as in* (3.1)*. Then, there exists $(r^{(1)}, \ldots, r^{(N)})$ in $\times_{i=1}^N \mathrm{H}(D_i; M_i)$ such that*

$$J_f\left(\bigotimes_{i=1}^N r^{(i)}\right) < 0.$$

Proof. This proof is based on the proof of Lemma 3 of [**LBLM09**]. Consider any functional $f \in \mathrm{H}(\mathsf{D};\mathsf{M})' \setminus \{0\}$ and assume that the thesis is false; i.e., $J_f\left(\bigotimes_{i=1}^N r^{(i)}\right) \geq 0$ for all ensembles $(r^{(1)}, \ldots, r^{(N)}) \in \times_{i=1}^N \mathrm{H}(D_i; M_i)$; then,

$$\frac{1}{2}\, a\left(\bigotimes_{i=1}^N r^{(i)}, \bigotimes_{i=1}^N r^{(i)}\right) \geq f\left(\bigotimes_{i=1}^N r^{(i)}\right) \qquad \forall\ (r^{(1)}, \ldots, r^{(N)}) \in \bigtimes_{i=1}^N \mathrm{H}(D_i; M_i).$$

Given a particular ensemble $(r^{(1)}, \ldots, r^{(N)}) \in \times_{i=1}^N \mathrm{H}(D_i; M_i)$, we can replace $r^{(1)}$ with $\varepsilon r^{(1)}$ and, by virtue of the bilinearity of a and the linearity of f we obtain

$$\frac{1}{2}\, \varepsilon^2 a\left(\bigotimes_{i=1}^N r^{(i)}, \bigotimes_{i=1}^N r^{(i)}\right) \geq \varepsilon\, f\left(\bigotimes_{i=1}^N r^{(i)}\right). \tag{3.7}$$

By combining the inequalities resulting from dividing both sides of (3.7) by positive ε and taking the one-sided limit $\varepsilon \to 0_+$ and from dividing (3.7) by a negative ε and taking the one-sided limit $\varepsilon \to 0_-$ we get that

$$f\left(\bigotimes_{i=1}^N r^{(i)}\right) = 0.$$

As this is valid for any ensemble $(r^{(1)}, \ldots, r^{(N)}) \in \times_{i=1}^N \mathrm{H}(D_i; M_i)$, Lemma 2.4 implies that it is valid, in particular, for any ensemble $(r^{(1)}, \ldots, r^{(N)}) \in \times_{i=1}^N \mathrm{C}_0^\infty(D_i)$, whence Lemma 2.22 implies that $f = 0$. As this contradicts the hypotheses of the lemma, its thesis holds. □

We are now in a position to prove the existence of solutions to problems (3.3) and (3.4).

Theorem 3.2. *Given $f_{n-1} \in \mathrm{H}(\mathsf{D};\mathsf{M})'$, each of the problems* (3.3) *and* (3.4) *has a solution.*

Proof. Since problems (3.3) and (3.4) are completely analogous, it suffices to consider one of them—say, (3.3). Then, as $(0,\dots,0)$ is a solution of (3.3) and (3.4) when $f_{n-1}=0$, we assume from now on that $f_{n-1}\neq 0$.

By (3.2) and the coerciveness of a, $J_{f_{n-1}}(\varphi)\geq -\frac{1}{2}a(\psi,\psi)$ for all $\varphi\in \mathrm{H}(\mathsf{D};\mathsf{M})$, where ψ is the unique solution of (2.5) in $\mathrm{H}(\mathsf{D};\mathsf{M})$ when $f=f_{n-1}$. As, by Lemma 2.23, the N-way tensor product of functions in $\mathrm{H}(D_i;M_i)$, $i\in[N]$, is a subset of $\mathrm{H}(\mathsf{D};\mathsf{M})$, $J_{f_{n-1}}$ is bounded from below over that manifold. That is,

$$\mathfrak{m}:=\inf_{(s^{(1)},\dots,s^{(N)})\in\bigtimes_{i=1}^{N}\mathrm{H}(D_i;M_i)} J_{f_{n-1}}\left(\bigotimes_{i=1}^{N}s^{(i)}\right)>-\infty. \tag{3.8}$$

It follows from Lemma 3.1 that $\mathfrak{m}<0$. Our aim is to show that the infimum $\mathfrak{m}$ is attained at an element of the form $\bigotimes_{i=1}^{N}r^{(i)}$ with $(r^{(1)},\dots,r^{(N)})\in\bigtimes_{i=1}^{N}(\mathrm{H}(D_i;M_i)\setminus\{0\})$.

From (3.8), there exists a sequence $\left(\bigotimes_{i\in[N]}r_k^{(i)}\colon k\geq 1\right)$ of N-way tensor products of functions in $\mathrm{H}(D_i;M_i)$, $i\in[N]$, such that

$$\lim_{k\to\infty}J_{f_{n-1}}\left(\bigotimes_{i=1}^{N}r_k^{(i)}\right)=\mathfrak{m}.$$

On noting that, from the definition of a in (2.2),

$$\begin{aligned}
J_{f_{n-1}}(\varphi)&=\frac{1}{2}\,a(\varphi-\psi,\varphi-\psi)-\frac{1}{2}\,a(\psi,\psi)\\
&=\frac{1}{4}a(\varphi,\varphi)+\left[\frac{1}{4}a(\varphi,\varphi)-a(\varphi,\psi)+a(\psi,\psi)\right]-a(\psi,\psi)\\
&\geq\frac{1}{4}\,a(\varphi,\varphi)-a(\psi,\psi)\\
&\geq\frac{1}{4}\min\left(\frac{\lambda_{\min}}{4\mathrm{Wi}},c\right)\,\|\varphi\|^2_{\mathrm{H}(\mathsf{D};\mathsf{M})}-a(\psi,\psi)
\end{aligned}$$

for all $\varphi\in\mathrm{H}(\mathsf{D};\mathsf{M})$ it follows, with $\varphi=\bigotimes_{i\in[N]}r_k^{(i)}$, that the sequence $\left(\bigotimes_{i\in[N]}r_k^{(i)}\colon k\geq 1\right)$ is bounded in $\mathrm{H}(\mathsf{D};\mathsf{M})$; in other words, there exists $C>0$ such that (cf. (2.31)):

$$\left\|\bigotimes_{i=1}^{N}r_k^{(i)}\right\|^2_{\mathrm{H}(\mathsf{D};\mathsf{M})}=\prod_{i=1}^{N}\left\|r_k^{(i)}\right\|^2_{\mathrm{L}^2_{1/M_i}(D_i)}+\sum_{j=1}^{N}\left(\prod_{\substack{i=1\\ i\neq j}}^{N}\left\|r_k^{(i)}\right\|^2_{\mathrm{L}^2_{1/M_i}(D_i)}\right)\left\|\nabla(r_k^{(j)}/M_j)\right\|^2_{[\mathrm{L}^2_{M_j}(D_j)]^d}\leq C \tag{3.9}$$

for all $k\geq 1$. Since the value of $\bigotimes_{i\in[N]}r_k^{(i)}$ is unaltered by multiplying the first $N-1$ factors by positive constants $c_{1,k},\dots,c_{N-1,k}$, respectively, and dividing the final factor by the product $c_{1,k}\cdots c_{N-1,k}$, we can assume without loss of generality that

$$\|r_k^{(i)}\|^2_{\mathrm{L}^2_{1/M_i}(D_i)}=1,\qquad i\in[N-1]. \tag{3.10}$$

Thus, it follows from (3.9) that

$$\left\|r_k^{(N)}\right\|^2_{\mathrm{L}^2_{1/M_N}(D_N)}+\left\|r_k^{(N)}\right\|^2_{\mathrm{L}^2_{1/M_N}(D_N)}\sum_{j=1}^{N-1}\left\|\nabla(r_k^{(j)}/M_j)\right\|^2_{[\mathrm{L}^2_{M_j}(D_j)]^d}+\left\|\nabla(r_k^{(N)}/M_N)\right\|^2_{[\mathrm{L}^2_{M_N}(D_N)]^d}\leq C. \quad (3.11)$$

Since the sequence $\left(\bigotimes_{i\in[N]} r_k^{(i)}\colon k\geq 1\right)$ is bounded in $\mathrm{H}(\mathsf{D};\mathsf{M})$, and $\mathrm{H}(\mathsf{D};\mathsf{M})$ is a Hilbert space, and therefore reflexive, the sequence has a weakly convergent subsequence in $\mathrm{H}(\mathsf{D};\mathsf{M})$, which we denote by $\left(\bigotimes_{i\in[N]} r_{\phi(k)}^{(i)}\colon k\geq 1\right)$; we denote its weak limit by $r\in\mathrm{H}(\mathsf{D};\mathsf{M})$. Since $J_{f_{n-1}}$ is convex on $\mathrm{H}(\mathsf{D};\mathsf{M})$ and continuous (and thereby also semicontinuous) in the strong topology of $\mathrm{H}(\mathsf{D};\mathsf{M})$, it is weakly lower-semicontinuous on $\mathrm{H}(\mathsf{D};\mathsf{M})$. Hence

$$J_{f_{n-1}}(r)\leq\liminf_{k\to\infty} J_{f_{n-1}}\left(\bigotimes_{i=1}^N r_{\phi(k)}^{(i)}\right)=\lim_{k\to\infty} J_{f_{n-1}}\left(\bigotimes_{i=1}^N r_k^{(i)}\right)=\mathfrak{m}<0.$$

Thus we deduce that $r\neq 0$ (as $r=0$ would imply that $J_{f_{n-1}}(r)=0$); hence, $r\in\mathrm{H}(\mathsf{D};\mathsf{M})\backslash\{0\}$.

According to (3.10) and (3.11) each subsequence $\left(r_{\phi(k)}^{(i)}\colon k\geq 1\right)$, is bounded in the respective space $\mathrm{L}^2_{1/M_i}(D_i)$, for $i\in[N]$. Then, $\left(r_{\phi(k)}^{(i)}\colon k\geq 1\right)$ has a weakly convergent subsequence in $\mathrm{L}^2_{1/M_i}(D_i)$, say $\left(r_{\phi'(k)}^{(i)}\colon k\geq 1\right)$, for $i\in[N]$; let us denote by $r^{(i)}\in\mathrm{L}^2_{1/M_i}(D_i)$ the corresponding weak limits:

$$\lim_{k\to\infty}\int_{D_i} r_{\phi'(k)}^{(i)}\varphi\frac{1}{M_i}=\int_{D_i} r^{(i)}\varphi\frac{1}{M_i}\qquad\forall\,\varphi\in\mathrm{L}^2_{1/M_i}(D_i),\quad i\in[N]. \quad (3.12)$$

As by Lemma 2.4, $\mathrm{C}_0^\infty(D_i)\subset\mathrm{H}(D_i;M_i)\subset\mathrm{L}^2_{1/M_i}(D_i)$, (3.12) is valid, in particular, for all $\varphi\in\mathrm{C}_0^\infty(D_i)$. Thus, $\left(r_{\phi'(k)}^{(i)}/M_i\colon k\geq 1\right)$ converges to $r^{(i)}/M_i$ in $\mathcal{D}'(D_i)$ for $i\in[N]$. Hence, by Lemma 2.22,

$$\lim_{k\to\infty}\bigotimes_{i=1}^N\frac{r_{\phi'(k)}^{(i)}}{M_i}=\bigotimes_{i=1}^N\frac{r^{(i)}}{M_i}=\frac{\bigotimes_{i=1}^N r^{(i)}}{\mathsf{M}}\quad\text{in }\mathcal{D}'(\mathsf{D}). \quad (3.13)$$

As all the distributions involved in (3.13) are regular distributions, (3.13) is equivalent to

$$\lim_{k\to\infty}\int_{\mathsf{D}}\frac{\bigotimes_{i=1}^N r_{\phi'(k)}^{(i)}}{\mathsf{M}}\varphi=\int_{\mathsf{D}}\frac{\bigotimes_{i=1}^N r^{(i)}}{\mathsf{M}}\varphi\quad\forall\,\varphi\in\mathrm{C}_0^\infty(\mathsf{D}). \quad (3.14)$$

Now, for all $\varphi\in\mathrm{C}_0^\infty(\mathsf{D})$, the functional defined on $\mathrm{H}(\mathsf{D};\mathsf{M})$ by $\xi\in\mathrm{H}(\mathsf{D};\mathsf{M})\mapsto\int_{\mathsf{D}}\xi\varphi\frac{1}{\mathsf{M}}$ is a member of $\mathrm{H}(\mathsf{D};\mathsf{M})'$. Then, the weak convergence of the sequence $\left(\bigotimes_{i\in[N]} r_{\phi(k)}^{(i)}\colon k\geq 1\right)$ to r in $\mathrm{H}(\mathsf{D};\mathsf{M})$ gives

$$\lim_{k\to\infty}\int_{\mathsf{D}}\frac{\bigotimes_{i=1}^N r_{\phi'(k)}^{(i)}}{\mathsf{M}}\varphi=\int_{\mathsf{D}}\frac{r}{\mathsf{M}}\varphi\quad\forall\,\varphi\in\mathrm{C}_0^\infty(\mathsf{D}) \quad (3.15)$$

for its subsequence. We have then that $\left(\mathsf{M}^{-1}\bigotimes_{i\in[N]} r_{\phi'(k)}^{(i)}\colon k\geq 1\right)$ converges in $\mathcal{D}'(\mathsf{D})$ to both $\mathsf{M}^{-1}\bigotimes_{i\in[N]} r^{(i)}$ and $\mathsf{M}^{-1}r$. However, $\mathcal{D}'(\mathsf{D})$ is also a Hausdorff topological space. Thus,

the two limits have to coincide. That is,

$$\frac{r}{\mathsf{M}} = \frac{\bigotimes_{i=1}^{N} r^{(i)}}{\mathsf{M}} \qquad \text{in } \mathcal{D}'(\mathsf{D}).$$

Hence, also, $r = \bigotimes_{i=1}^{N} r^{(i)}$ almost everywhere. As $r \in \mathrm{H}(\mathsf{D};\mathsf{M}) \setminus \{0\}$ and has a tensor-product structure, the second part of Lemma 2.23 implies that $r^{(i)} \in \mathrm{H}(D_i; M_i) \setminus \{0\}$ for $i \in [N]$. Now,

$$J_{f_{n-1}}\left(\bigotimes_{i=1}^{N} r^{(i)}\right) = J_{f_{n-1}}(r) \leq \mathfrak{m}.$$

Recalling the definition of $\mathfrak{m}$ from (3.8), we have thus shown that the infimum in (3.8) is attained at $\bigotimes_{i=1}^{N} r^{(i)}$. Thus, $(r^{(1)}, \ldots, r^{(N)}) \in \times_{i=1}^{N}(\mathrm{H}(D_i; M_i) \setminus \{0\})$ is a solution to problem (3.3). That completes the proof. □

Remark 3.3. The proof of Theorem 3.2 follows the structure of the proof of Proposition 1 of [**LBLM09**]. However, the adaptation of the distributional arguments made there to our Maxwellian-weighted setting is delicate; hence our detailed presentation of the proof.

Having proved the existence of solutions to the minimization problems (3.3) of Algorithm I and (3.4) of Algorithm II, establishing the correctness of what is left of the algorithms is straightforward. The Galerkin problem in step `2.2` of Algorithm II is well-defined and has a unique solution for each $n \geq 1$, because it is equivalent to the minimization of a coercive quadratic form over a finite-dimensional linear space. Then, at last, the definition of the n-th residual in step `2.2` of Algorithm I and in step `2.3` of Algorithm II are correct on noting that $\mathcal{A}$ maps $\mathrm{H}(\mathsf{D};\mathsf{M})$ into $\mathrm{H}(\mathsf{D};\mathsf{M})'$.

In the next section we establish the convergence of the sequences generated by the two algorithms, which will then imply that both algorithms will terminated for any fixed tolerance `TOL` after a `TOL`-dependent number of steps.

3.2. Convergence

3.2.1. Euler–Lagrange equations.

Lemma 3.4. *Local minimizers* $(r_n^{(1)}, \ldots, r_n^{(N)})$ *of the minimization problems* (3.3) *or* (3.4) *satisfy the following Euler–Lagrange equation (system): For all ensembles* $(s^{(1)}, \ldots, s^{(N)})$ *in* $\times_{i\in[N]} \mathrm{H}(D_i; M_i)$,

$$a\left(\bigotimes_{i=1}^{N} r_n^{(i)}, \sum_{j=1}^{N} \bigotimes_{\substack{i=1\\ i\neq j}}^{N} r_n^{(i)} \otimes_j s^{(j)}\right) = f_{n-1}\left(\sum_{j=1}^{N} \bigotimes_{\substack{i=1\\ i\neq j}}^{N} r_n^{(i)} \otimes_j s^{(j)}\right). \tag{3.16}$$

From this, it follows that, for both the Pure Greedy Algorithm (Algorithm I) and the Orthogonal Greedy Algorithm (Algorithm II):

$$a\left(\bigotimes_{i=1}^{N} r_n^{(i)}, \bigotimes_{i=1}^{N} r_n^{(i)}\right) = a\left(\delta_{n-1}, \bigotimes_{i=1}^{N} r_n^{(i)}\right). \tag{3.17}$$

Proof. Let $(r_n^{(1)}, \dots, r_n^{(N)})$ be a solution to the minimization problem (3.3) or (3.4). Then, given any $(s^{(1)}, \dots, s^{(N)})$, (3.16) is an equivalent way of writing that the derivative of

$$J_{f_{n-1}}\left(\bigotimes_{i=1}^{N}\left(r_n^{(i)} + \varepsilon s^{(i)}\right)\right)$$

with respect to ε is zero when evaluated at $\varepsilon = 0$. Since, by hypothesis, $(r_n^{(1)}, \dots, r_n^{(N)})$ is a local minimizer of $J_{f_{n-1}}$ and $\varepsilon \mapsto \mathfrak{J}_n(\varepsilon) := J_{f_{n-1}}\left(\bigotimes_{i=1}^{N}\left(r_n^{(i)} + \varepsilon s^{(i)}\right)\right)$ is regular enough, the fact that $\mathfrak{J}_n'(0) = 0$ implies that (3.16) holds.

Setting $(s^{(1)}, \dots, s^{(N)}) = (r_n^{(1)}, \dots, r_n^{(N)})$ in (3.16) and combining the resulting equality with the fact that

$$a\left(\delta_{n-1}, \bigotimes_{i=1}^{N} r_n^{(i)}\right) = f_{n-1}\left(\bigotimes_{i=1}^{N} r_n^{(i)}\right)$$

we obtain (3.17). □

Remark 3.5.

(1) The above lemma only states that local minima of the minimization problem (3.3) and (3.4) satisfy the Euler–Lagrange equation (3.16). The converse might be false.
(2) In what follows we make liberal use of the norm $\|\cdot\|_a := a(\cdot,\cdot)^{1/2}$ on $\mathrm{H}(\mathsf{D};\mathsf{M})$, which is something that, on account of its equivalence with $\|\cdot\|_{\mathrm{H}(\mathsf{D};\mathsf{M})}$, makes no difference when making topological statements (such as convergence).

The following lemma corresponds to Lemma 6 of [**LBLM09**].

Lemma 3.6. *Suppose $f_{n-1} \neq 0$ and let $(r_n^{(1)}, \dots, r_n^{(N)})$ be a global minimizer for the minimization problem* (3.3) *of the Algorithm I or for the minimization problem* (3.4) *of the Algorithm II. Then,*

$$\left\|\bigotimes_{i=1}^{N} r_n^{(i)}\right\|_a = \frac{a\left(\delta_{n-1}, \bigotimes_{i=1}^{N} r_n^{(i)}\right)}{\left\|\bigotimes_{i=1}^{N} r_n^{(i)}\right\|_a} = \sup_{s \in \bigotimes_{i\in[N]} \mathrm{H}(D_i;M_i)\setminus\{0\}} \frac{a\left(\delta_{n-1}, s\right)}{\|s\|_a}. \tag{3.18}$$

If $f_{n-1} = 0$, the equality between the left-most and the right-most expressions in (3.18) *is still valid.*

Proof. We start by considering the case $f_{n-1} \neq 0$. Then, the first equality in (3.18) comes directly from (3.17) in Lemma 3.4 and the fact that $\|r_n\|_a \neq 0$, which is guaranteed by Lemma 3.1. Now, analogously to (3.2), $J_{f_{n-1}}$ can be written as

$$J_{f_{n-1}}(\varphi) = \frac{1}{2}a(\varphi - \delta_{n-1}, \varphi - \delta_{n-1}) - \frac{1}{2}a(\delta_{n-1}, \delta_{n-1}) \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}).$$

Combining this representation of $J_{f_{n-1}}$ with the fact that $r_n := \bigotimes_{i\in[N]} r_n^{(i)}$ minimizes $J_{f_{n-1}}$ among the members of $\bigotimes_{i\in[N]} \mathrm{H}(D_i; M_i)$ and the first equality of (3.18), according to which

$a\left(\delta_{n-1}, r_n\right) = \|r_n\|_a^2$, we have, for all $s \in \bigotimes_{i\in[N]} \mathrm{H}(D_i; M_i) \setminus \{0\}$, that

$$\left\|\delta_{n-1} - \frac{a\left(\delta_{n-1}, r_n\right)}{\|r_n\|_a^2} r_n\right\|_a^2 = \|\delta_{n-1} - r_n\|_a^2 \leq \left\|\delta_{n-1} - \frac{a(\delta_{n-1}, s)}{\|s\|_a^2} s\right\|_a^2.$$

Therefore,

$$\frac{a(\delta_{n-1}, r_n)^2}{a(r_n, r_n)} \geq \frac{a(\delta_{n-1}, s)^2}{a(s, s)}.$$

Taking the supremum over $s \in \bigotimes_{i\in[N]} \mathrm{H}(D_i; M_i) \setminus \{0\}$ and noting that r_n is an admissible s we get the second equality in (3.18). The statement concerning the $f_{n-1} = 0$ case is trivially true. □

3.2.2. Convergence.

Theorem 3.7. *The Pure Greedy Algorithm (Algorithm I) converges to the solution ψ to* (2.5)*; that is, for any* $\mathtt{TOL} > 0$ *there exists some iteration number n such that*

$$\|f_n\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} = \|\delta_n\|_a < \mathtt{TOL}.$$

Proof. This proof is a refinement of the proof of Theorem 1 of [**LBLM09**].

Let $\left((r_n^{(1)}, \ldots, r_n^{(N)})\colon n \geq 1\right)$ be a sequence in $\times_{i=1}^N \mathrm{H}(D_i; M_i)$ returned by the Pure Greedy Algorithm. Then, from (3.6) and (3.17) in Lemma 3.4 we obtain

$$a\left(\delta_n, \bigotimes_{i=1}^N r_n^{(i)}\right) = 0$$

and then

$$\|\delta_{n-1}\|_a^2 = \left\|\delta_n + \bigotimes_{i=1}^N r_n^{(i)}\right\|_a^2 = \|\delta_n\|_a^2 + \left\|\bigotimes_{i=1}^N r_n^{(i)}\right\|_a^2.$$

Hence the sequence $(\|\delta_n\|_a : n \geq 0)$ is nonnegative and monotonic nonincreasing, and therefore converges in $\mathbb{R}$; by summing the above expression over n we then deduce that

$$\sum_{n=1}^{\infty} a\left(\bigotimes_{i=1}^N r_n^{(i)}, \bigotimes_{i=1}^N r_n^{(i)}\right) < \infty. \tag{3.19}$$

Let us define the function $\phi\colon \mathbb{N} \to \mathbb{N}$ recursively by $\phi(1) := 1$ and

$$\phi(k) := \min\left\{n \in \mathbb{N}\colon n > \phi(k-1) \quad \text{and} \quad \left\|\bigotimes_{i=1}^N r_n^{(i)}\right\|_a \leq \left\|\bigotimes_{i=1}^N r_{\phi(k-1)}^{(i)}\right\|_a\right\}, \quad k \geq 2.$$

From (3.19) the function ϕ is well-defined and strictly monotonic increasing. Hence, it is suitable for defining subsequences. As each $\left(r_{\phi(n)}^{(1)}, \ldots, r_{\phi(n)}^{(N)}\right)$ is a global minimizer to the

problem (3.3) with the instance $f_{\phi(n)-1}$, via (3.6) and Lemma 3.6 we have, for $1 \leq m \leq n$,

$$\begin{aligned}
\left\|\delta_{\phi(n)-1} - \delta_{\phi(m)-1}\right\|_a^2 &= \left\|\delta_{\phi(n)-1}\right\|_a^2 + \left\|\delta_{\phi(m)-1}\right\|_a^2 - 2a\left(\delta_{\phi(n)-1}, \delta_{\phi(n)-1} + \sum_{k=\phi(m)}^{\phi(n)-1} \bigotimes_{i=1}^{N} r_k^{(i)}\right) \\
&= \left\|\delta_{\phi(m)-1}\right\|_a^2 - \left\|\delta_{\phi(n)-1}\right\|_a^2 - 2\sum_{k=\phi(m)}^{\phi(n)-1} a\left(\delta_{\phi(n)-1}, \bigotimes_{i=1}^{N} r_k^{(i)}\right) \\
&\leq \left\|\delta_{\phi(m)-1}\right\|_a^2 - \left\|\delta_{\phi(n)-1}\right\|_a^2 + 2\sum_{k=\phi(m)}^{\phi(n)-1} \left\|\bigotimes_{i=1}^{N} r_k^{(i)}\right\|_a \left\|\bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}\right\|_a \\
&\leq \left\|\delta_{\phi(m)-1}\right\|_a^2 - \left\|\delta_{\phi(n)-1}\right\|_a^2 + 2\sum_{k=\phi(m)}^{\phi(n)-1} \left\|\bigotimes_{i=1}^{N} r_k^{(i)}\right\|_a^2 .
\end{aligned} \tag{3.20}$$

From the convergence of $\left(\left\|\delta_{\phi(n)-1}\right\|_a : n \geq 1\right)$ in $\mathbb{R}$ and (3.19), we deduce that the sequence $\left(\delta_{\phi(n)-1} \colon n \geq 1\right)$ is a Cauchy sequence in $\mathrm{H}(\mathsf{D};\mathsf{M})$ and thus converges to some $\delta_\infty \in \mathrm{H}(\mathsf{D};\mathsf{M})$. Another consequence of the global optimality of each $(r_n^{(1)}, \ldots, r_n^{(N)})$ is: For all ensembles $(s^{(1)}, \ldots, s^{(N)})$ in $\times_{i\in[N]} \mathrm{H}(D_i; M_i)$ and $n \geq 1$,

$$\begin{aligned}
\frac{1}{2}a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - a\left(\delta_{\phi(n)-1}, \bigotimes_{i=1}^{N} s^{(i)}\right) &\geq J_{f_{\phi(n)-1}}\left(\bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}\right) \\
&= \frac{1}{2}a\left(\bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}, \bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}\right) - f_{\phi(n)-1}\left(\bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}\right) \\
&= -\frac{1}{2}a\left(\bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}, \bigotimes_{i=1}^{N} r_{\phi(n)}^{(i)}\right).
\end{aligned}$$

Taking the limit as n tends to infinity at both ends, and noting that by (3.19) the right-hand side of the last inequality converges to 0, we obtain

$$\frac{1}{2}a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - a\left(\delta_\infty, \bigotimes_{i=1}^{N} s^{(i)}\right) \geq 0.$$

Thus, Lemma 3.1 implies that $\delta_\infty = 0$. Hence the sequence $\left(\left\|\delta_{\phi(n)-1}\right\| : n \geq 1\right)$ converges to zero as n tends to infinity. As the sequence $(\|\delta_n\|_a : n \geq 0)$ is monotonic nonincreasing and $(\phi(n) - 1 \colon n \geq 1)$ is a monotonic increasing infinite sequence in $\mathbb{N}$, it follows that the full sequence $(\|\delta_n\| : n \geq 1)$ converges to the common limit in $\mathbb{R}$: $0 = \|\delta_\infty\|_a$, giving

$$\lim_{n\to\infty} \delta_n = 0 \quad \text{in} \quad \mathrm{H}(\mathsf{D};\mathsf{M}).$$

□

The following corollary is a direct consequence of Theorem 3.7 and will prove useful later on.

Corollary 3.8. *Suppose that F_i is a dense subset of $\mathrm{H}(D_i;M_i)$ for $i\in[N]$. Then, the span of $\bigotimes_{i\in[N]}F_i$ is dense in $\mathrm{H}(\mathsf{D};\mathsf{M})$.*

Proof. Let $\tau\in\mathrm{H}(\mathsf{D};\mathsf{M})$. Applying Theorem 3.7 to the case in which the right-hand side functional $f\in\mathrm{H}(\mathsf{D};\mathsf{M})'$ of problem (2.5) is $\varphi\mapsto a(\tau,\varphi)$ (i.e., the $\mathrm{H}(\mathsf{D};\mathsf{M})$ *approximation* problem) it follows that τ can be approximated arbitrarily closely by finite sums of the form $\sum_{m\in[M]}\bigotimes_{i\in[N]}r_m^{(i)}$, where $r_m^{(i)}\in\mathrm{H}(D_i;M_i)$ for $m\in[M]$ and $i\in[N]$. Thus, if we can show that $\bigotimes_{i\in[N]}F_i$ is dense in the manifold $\bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)$, our desired result will stand.

Let, then, $r^{(i)}\in\mathrm{H}(D_i;M_i)$, for $i\in[N]$. From the density of F_i in $\mathrm{H}(D_i;M_i)$ for each $i\in[N]$, there exists a sequence $\left(r_n^{(i)}:n\geq1\right)$ in F_i, which converges to $r^{(i)}$ in $\mathrm{H}(D_i;M_i)$. Now,

$$\bigotimes_{i=1}^{N}r^{(i)}-\bigotimes_{i=1}^{N}r_n^{(i)}=\sum_{k=1}^{N}\bigotimes_{i=1}^{N}t_{n,k}^{(i)},\qquad\text{where}\qquad t_{n,k}^{(i)}:=\begin{cases}r_n^{(i)} & \text{if } i>k,\\ r^{(i)}-r_n^{(i)} & \text{if } i=k,\\ r^{(i)} & \text{if } i<k.\end{cases}$$

Then, (cf. (2.31)),

$$\left\|\bigotimes_{i=1}^{N}r^{(i)}-\bigotimes_{i=1}^{N}r_n^{(i)}\right\|_{\mathrm{H}(\mathsf{D};\mathsf{M})}^2\leq\sum_{k=1}^{N}\left[\prod_{i=1}^{N}\left\|t_{n,k}^{(i)}\right\|_{\mathrm{L}^2_{1/M_i}(D_i)}^2+\sum_{j=1}^{N}\prod_{\substack{i=1\\i\neq j}}^{N}\left\|t_{n,k}^{(i)}\right\|_{\mathrm{L}^2_{1/M_i}(D_i)}^2\left\|\nabla(t_{n,k}^{(j)}/M_j)\right\|_{[\mathrm{L}^2_{M_j}(D_j)]^d}^2\right].$$

As each product term on the right-hand side above consists of $N-1$ bounded factors and one vanishing factor as $n\to\infty$, the full expression tends to zero as n tends to infinity and, therefore, so does the left-hand side. The desired result follows. □

Remark 3.9. Suppose that, for each $i\in[N]$,

$$\mathrm{C}_0^\infty(D_i)\text{ is dense in }\mathrm{H}(D_i;M_i).\tag{3.21}$$

Then, as

$$\operatorname{span}\left(\bigotimes_{i=1}^{N}\mathrm{C}_0^\infty(D_i)\right)\subset\mathrm{C}_0^\infty(\mathsf{D})\subset\mathrm{H}(\mathsf{D};\mathsf{M}),$$

we have, thanks to Corollary 3.8, that

$$\mathrm{C}_0^\infty(\mathsf{D})\text{ is dense in }\mathrm{H}(\mathsf{D};\mathsf{M}).\tag{3.22}$$

Springs obeying the FENE model (1.16) comply with (3.21) under the condition $b_i\geq2$ as is proved in Remark 3.7 of [**Mas08**]. In turn, springs obeying the CPAIL model (1.17) with parameter $b_i\geq3$, the TEAIL model (1.18) with parameter $b_i\geq16/5$ or the CP model (1.21) with parameter $b_i\geq3$ comply with (3.21) as it is shown in Lemma 2.6 in Subsection 2.2.1. Finally, springs obeying the Inverse Langevin model (1.19) with parameter $b_i\geq3$ are shown

to comply with (3.21) in Subsection 2.2.2. So, if each of the partial Maxwellians M_i, $i \in [N]$, that constitute M obeys any of these five models, (3.22) holds.

Interesting as (3.22) is, we make no use of it in the main body of this work and that is why we shall not adopt (3.21) as a hypothesis on a par with hypotheses A and B above or hypotheses C, D and E below. However, we do use (3.21) as an ingredient in the proof of the compliance of FENE and CPAIL spring potentials with Hypothesis C of Section 3.3 (cf. Corollary 2.12 in Subsection 2.2.3).

Theorem 3.10. *The Orthogonal Greedy Algorithm (Algorithm II) converges to the solution ψ to problem (2.5); that is, for any $\mathtt{TOL} > 0$ there exists some iteration number n such that*

$$\|f_n\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} = \|\delta_n\|_a < \mathtt{TOL}.$$

Proof. This proof is based on the proof of Theorem 2 of [**LBLM09**]. We first note that due to (3.6), the optimality of $\alpha^{(n)}$ in (3.5) and the optimality of $(r_n^{(1)}, \dots, r_n^{(N)})$ in (3.4) (via Lemma 3.4),

$$\|\delta_n\|_a^2 = \left\| \psi - \sum_{k=1}^{n} \alpha_k^{(n)} \bigotimes_{i=1}^{N} r_k^{(i)} \right\|_a^2 \leq \left\| \delta_{n-1} - \bigotimes_{i=1}^{N} r_n^{(i)} \right\|_a^2 = \|\delta_{n-1}\|_a^2 - \left\| \bigotimes_{i=1}^{N} r_n^{(i)} \right\|_a^2. \tag{3.23}$$

Thus, just like in the proof of Theorem 3.7, we have that the sequence of norms $(\|\delta_n\|_a : n \geq 0)$ is monotonic decreasing and thus convergent and that $\sum_{n\geq 1} a\left(\bigotimes_{i\in[N]} r_n^{(i)}, \bigotimes_{i\in[N]} r_n^{(i)}\right) < \infty$. As $(\delta_n \colon n \geq 0)$ is a bounded sequence in the reflexive space $\mathrm{H}(\mathsf{D};\mathsf{M})$, a weakly convergent subsequence $(\delta_{n_m} \colon m \geq 1)$ can be extracted; we denote the weak limit by δ_∞. From the optimality of $(r_{n_m+1}^{(1)}, \dots r_{n_m+1}^{(N)})$ with respect to problem (3.4) it follows by Lemma 3.4 that, for all $(s^{(1)}, \dots, s^{(N)}) \in \times_{i\in[N]} \mathrm{H}(D_i; M_i)$,

$$\frac{1}{2} a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - a\left(\delta_{n_m}, \bigotimes_{i=1}^{N} s^{(i)}\right) \geq -\frac{1}{2} a\left(\bigotimes_{i=1}^{N} r_{n_m+1}^{(i)}, \bigotimes_{i=1}^{N} r_{n_m+1}^{(i)}\right).$$

On taking the limit $m \to \infty$ at both sides we obtain

$$\frac{1}{2} a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - a\left(\delta_\infty, \bigotimes_{i=1}^{N} s^{(i)}\right) \geq 0$$

whence, via Lemma 3.1, $\delta_\infty = 0$. However, from the Galerkin orthogonality associated with problem (3.5), $a(\psi - \delta_{n_m}, \delta_{n_m}) = 0$. That is, $\|\delta_{n_m}\|_a^2 = a(\psi, \delta_{n_m})$. Hence, $\lim_{m\to\infty} \|\delta_{n_m}\|_a^2 = \lim_{m\to\infty} a(\psi, \delta_{n_m}) = a(\psi, \delta_\infty) = 0$. As the full sequence of norms $(\|\delta_n\|_a : n \geq 0)$ is monotonic decreasing, the full sequence $(\delta_n \colon n \geq 0)$ converges strongly to 0 in $\mathrm{H}(\mathsf{D};\mathsf{M})$. □

3.2.3. General Greedy Algorithms. As the authors of [**LBLM09**] recognized in their analysis of the Separated Representation strategy for the Poisson equation, the notion of greedy algorithm from the theory of nonlinear approximation opens the door to a discussion about rates of convergence for the algorithms with respect to the number of steps. The

same can be done in our Maxwellian-weighted setting, as long as we first ensure that the Algorithm I and Algorithm II are indeed greedy algorithms in this theoretical sense.

Definition 3.11. *Given a Hilbert space $\mathfrak{H}$, a* dictionary *is a set $\mathfrak{D} \subset \mathfrak{H}$ whose elements have unit $\mathfrak{H}$-norm and obey*

$$g \in \mathfrak{D} \implies -g \in \mathfrak{D}.$$

Given $f \in \mathfrak{H}$, let $g(f)$ be a member of $\mathfrak{D}$ which maximizes $g \mapsto \langle f, g\rangle_{\mathfrak{H}}$—such a maximizer is assumed to exist. We further define

$$G(f) := \langle f, g(f)\rangle_{\mathfrak{H}}\, g(f)$$

and

$$R(f) := f - G(f).$$

Based on Definition 3.11, the following two algorithms are defined in [**DT96**]

Algorithm III *(General Pure Greedy Algorithm).* `Input: Some` $f \in \mathfrak{H}$.

`1. Define:` $R_0 := f$ `and` $G_0 := 0$.
`2. For` $n \geq 1$ `do:`
`2.1 Obtain`
$$g(R_{n-1}) \in \arg\max_{\tilde{g}\in\mathfrak{D}} \langle R_{n-1}, \tilde{g}\rangle_{\mathfrak{H}}.$$
`2.2 Set` $G_n := G_{n-1} + G(R_{n-1}) = G_{n-1} + \langle R_{n-1}, g(R_{n-1})\rangle_{\mathfrak{H}}\, g(R_{n-1})$.
`2.3 Set` $R_n := f - G_n = R(R_{n-1})$.

Algorithm IV *(General Orthogonal Greedy Algorithm).* `Input: Some` $f \in \mathfrak{H}$.

`1. Define:` $R_0 := f$ `and` $G_0 := 0$.
`2. For` $n \geq 1$ `do:`
`2.1 Obtain`
$$g(R_{n-1}) \in \arg\max_{\tilde{g}\in\mathfrak{D}} \langle R_{n-1}, \tilde{g}\rangle_{\mathfrak{H}}.$$
`2.2 Set` $H_n := \operatorname{span}\{g(R_0), \ldots, g(R_{n-1})\}$.
`2.3 Set` $G_n(f) := P_{H_n}(f)$`; i.e., the projection of` f `on` H_n.
`2.4 Set` $R_n := f - G_n(f)$.

The assumption of the existence of a maximizer $g(f)$ of $g \mapsto \langle f, g\rangle_{\mathfrak{H}}$ made in Definition 3.11 can be tricky to satisfy and depends heavily on what additional structure the dictionary $\mathfrak{D}$ has. No uniqueness of $g(f)$ is assumed, so $f \mapsto g(f)$ must be seen less as a function than as a selection procedure.

The purpose of Algorithm III and Algorithm IV is to construct approximations to f which have, at iteration n, the form

$$G_n = \sum_{k=0}^{n-1} G(R_k) = \sum_{k=0}^{n-1} \langle R_k, g(R_k)\rangle_{\mathfrak{H}}\, g(R_k)$$

and

$$G_n = \sum_{k=0}^{n-1} \alpha_k^{(n)} g(R_k),$$

respectively, with $\alpha^{(n)} \in \mathbb{R}^n$ given by the condition that G_n is the projection of f on the space spanned by the $G(R_k)$, respectively. The R_k are the successive residuals, which, in this approximation problem setting, are also the errors.

As their names suggest, Algorithm I and Algorithm II fall into this abstract setting. Indeed, let

$$\mathfrak{H} = \mathrm{H}(\mathsf{D};\mathsf{M})' \tag{3.24}$$

equipped with the inner product induced by the a-inner product of $\mathrm{H}(\mathsf{D};\mathsf{M})$; that is, for all $g_1,\, g_2 \in \mathrm{H}(\mathsf{D};\mathsf{M})'$,

$$\langle g_1, g_2\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'} := a(\gamma_1, \gamma_2)$$

where, for $j \in \{1,2\}$, γ_j is defined as the unique solution in $\mathrm{H}(\mathsf{D};\mathsf{M})$ to the variational problem

$$a(\gamma_j, \varphi) = g_j(\varphi) \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}).$$

Also, we fix the dictionary according to

$$\mathfrak{D} = \left\{ g \in \mathrm{H}(\mathsf{D};\mathsf{M})' \colon g = a\,(s,\cdot)\,,\ s \in \bigotimes_{i=1}^{N} \mathrm{H}(D_i;M_i),\ \|g\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} = 1 \right\}, \tag{3.25}$$

and let f (the member of $\mathfrak{H}$ to be approximated) be f (the right-hand side functional of (2.1)). In order to make apparent the connection between the functional-minimizing procedure of Algorithm I and Algorithm II and the inner-product-maximizing procedure of Algorithm III and Algorithm IV we need the proposition that follows.

Proposition 3.12. *Let $\tilde{f} \in \mathrm{H}(\mathsf{D};\mathsf{M})' \setminus \{0\}$. Then,*

$$r \in \operatorname*{arg\,min}_{t \in \bigotimes_{i\in[N]} \mathrm{H}(D_i;M_i)} J_{\tilde{f}}(t) \implies a\left(\frac{r}{\|r\|_a}, \cdot\right) \in \operatorname*{arg\,max}_{\tilde{g}\in\mathfrak{D}} \langle \tilde{f}, \tilde{g}\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'} \tag{3.26}$$

and

$$a(s,\cdot) \in \operatorname*{arg\,max}_{\tilde{g}\in\mathfrak{D}} \langle \tilde{f}, \tilde{g}\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'} \implies \langle \tilde{f}, a(s,\cdot)\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'}\, s \in \operatorname*{arg\,min}_{t \in \bigotimes_{i\in[N]} \mathrm{H}(D_i;M_i)} J_{\tilde{f}}(t). \tag{3.27}$$

Proof. We recall first that, from Lemma 3.1, r in (3.26) is not zero, so it can be normalized. Let $\tilde{\psi}$ be the unique solution in $\mathrm{H}(\mathsf{D};\mathsf{M})$ to

$$a(\tilde{\psi}, \varphi) = \tilde{f}(\varphi) \quad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}).$$

Then, the $\mathrm{H}(\mathsf{D};\mathsf{M})'$-inner product between $\tilde{f}$ and a generic member of $\mathrm{H}(\mathsf{D};\mathsf{M})'$ of the form $\tilde{g} = a(\tilde{s},\cdot)$ is simply $a(\tilde{\psi},\tilde{s})$. As for these functionals $\|\tilde{g}\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} = \|\tilde{s}\|_a$, the implications (3.26) and (3.27) are equivalent to

$$r \in \operatorname*{arg\,min}_{t\in\bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)} J_{\tilde{f}}(t) \implies \frac{r}{\|r\|_a} \in \operatorname*{arg\,max}_{\substack{\tilde{s}\in\bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)\\ \|\tilde{s}\|_a=1}} a(\tilde{\psi},\tilde{s}) \tag{3.28}$$

and

$$s \in \operatorname*{arg\,max}_{\substack{\tilde{s}\in\bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)\\ \|\tilde{s}\|_a=1}} a(\tilde{\psi},\tilde{s}) \implies a(\tilde{\psi},s)s \in \operatorname*{arg\,min}_{t\in\bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)} J_{\tilde{f}}(t), \tag{3.29}$$

respectively.

The truth of (3.28) is readily apparent on account of Lemma 3.6, and we have (3.26). Proving (3.29) is slightly more involved; we will prove it by proving its contraposition. So, let us take any unit a-norm $s \in \bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)$ such that

$$a(\tilde{\psi},s)s \notin \operatorname*{arg\,min}_{t\in\bigotimes_{i\in[N]}\mathrm{H}(D_i;M_i)} J_{\tilde{f}}(t)$$

and let r^* be a global minimizer of $J_{\tilde{f}}$ on $\bigotimes_{i=1}^N \mathrm{H}(D_i;M_i)$, whose existence is guaranteed by Theorem 3.2. Then,

$$\frac{1}{2}a(r^*,r^*) - \tilde{f}(r^*) < \frac{1}{2}a\left(a(\tilde{\psi},s)s, a(\tilde{\psi},s)s\right) - \tilde{f}\left(a(\tilde{\psi},s)s\right),$$

which because of $a(r^*,r^*) = a(\tilde{\psi},r^*) = f(r^*)$ (cf. Lemma 3.4), $\tilde{f}(s) = a(\tilde{\psi},s)$ and the unit a-norm of s results in

$$a(\tilde{\psi},s)^2 < a(\tilde{\psi},r^*) = \|r^*\|_a\, a\left(\tilde{\psi},\frac{r^*}{\|r^*\|_a}\right) = \frac{a(\tilde{\psi},r^*)}{\|r^*\|_a} a\left(\tilde{\psi},\frac{r^*}{\|r^*\|_a}\right) = a\left(\tilde{\psi},\frac{r^*}{\|r^*\|_a}\right)^2.$$

Therefore, s is not a global maximizer of $a(\tilde{\psi},\cdot)$ among the unit a-norm members of $\mathrm{H}(D_i;M_i)$. Thus, we have proved the contraposition of (3.29), and so (3.29) itself and the equivalent form (3.29). □

Proposition 3.12 (plus obvious arguments in case the relevant functional is zero) allows for stating that the approximants $\sum_{k=1}^N \bigotimes_{i\in[N]} r_n^{(i)}$ and $\sum_{k=1}^N \alpha_k^{(n)} \bigotimes_{i=1}^N r_k^{(i)}$ (resp. the errors δ_n) of Algorithm I and Algorithm II are connected to the approximants G_n (resp. the residuals R_n) of Algorithm III and Algorithm IV via the a-based Riesz isometry between $\mathrm{H}(\mathsf{D};\mathsf{M})'$ and $\mathrm{H}(\mathsf{D};\mathsf{M})$. Therefore, beyond the non-essential addition of a termination criterion to Algorithm II and Algorithm I, the correspondence is established.

3.2.4. Rate of convergence. The theory of nonlinear approximation introduced in Subsection 3.2.3 provides us with some estimates of the rate of convergence of Algorithm I and Algorithm II.

Following [**DT96**] we introduce the space

$$\mathcal{A}_1 := \bigcup_{M>0} \overline{\mathcal{A}_1^o(M)}, \tag{3.30a}$$

where

$$\mathcal{A}_1^o(M) := \Bigg\{ \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M})\colon \varphi = \sum_{k\in\Lambda} c_k w_k,\ w_k \in \bigotimes_{i=1}^{N} \mathrm{H}(D_i;M_i),\ \|w_k\|_a = 1, \\ |\Lambda| < \infty \quad \text{and} \quad \sum_{k\in\Lambda} |c_k| \le M \Bigg\}, \tag{3.30b}$$

equipped with the norm

$$\|\varphi\|_{\mathcal{A}_1} := \inf\left\{ M > 0\colon \varphi \in \overline{\mathcal{A}_1^o(M)} \right\}. \tag{3.30c}$$

The importance of this space becomes apparent in the light of the following two theorems, which we cite below.

Theorem 3.13. *If the solution ψ of* (2.5) *is a member of $\mathcal{A}_1$, the n-th error δ_n of the Pure Greedy Algorithm (Algorithm I) satisfies*

$$\|\delta_n\|_a \le \|\psi\|_{\mathcal{A}_1}\, n^{-1/6}.$$

Proof. This follows from a direct application of Theorem 3.6 of [**DT96**]. □

Theorem 3.14. *If the solution ψ of* (2.5) *is a member of $\mathcal{A}_1$, the n-th error δ_n of the Orthogonal Greedy Algorithm (Algorithm II) satisfies*

$$\|\delta_n\|_a \le \|\psi\|_{\mathcal{A}_1}\, n^{-1/2}.$$

Proof. This follows from a direct application of Theorem 3.7 of [**DT96**]. □

Remark 3.15.

(1) We have chosen to set $\mathcal{A}_1$ as a subset of $\mathrm{H}(\mathsf{D};\mathsf{M})$ instead of (equivalently, via the Riesz isometry) $\mathrm{H}(\mathsf{D};\mathsf{M})'$, the latter being what a straightforward translation of the definition of $\mathcal{A}_1$ in [**DT96**] would entail.

(2) Pure Greedy Algorithm-based approximations such as Algorithm I have been proved to obey the slightly improved rate (see [**Tem08**, Remark 2.3.11] and references therein)

$$\|\delta_n\|_a \le 4\,\|\psi\|_{\mathcal{A}_1}\, n^{-11/62}.$$

(3) In [**CEL11**, Theorem 4.1] it is shown that the convergence of the Orthogonal Greedy Algorithm takes place exponentially fast if the factor spaces and the full ansatz space (in our setting the $\mathrm{H}(D_i;M_i)$ and $\mathrm{H}(\mathsf{D};\mathsf{M})$, respectively) are finite-dimensional.

We note that $\mathcal{A}_1$ will remain the same space if in its definition—in (3.30b), in particular—we replace the energy norm $\|\cdot\|_a$ with the standard norm of $\mathrm{H}(\mathsf{D};\mathsf{M})$, as these two norms are

equivalent. Then, $\varphi \in \mathrm{H}(\mathsf{D};\mathsf{M})$ will be a member of $\mathcal{A}_1$ if, and only if, there exists some M^* such that, for all $\varepsilon > 0$, there exists some $\chi_\varepsilon \in \mathrm{H}(\mathsf{D};\mathsf{M})$ that satisfies

$$\|\varphi - \chi_\varepsilon\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} \leq \varepsilon, \quad \chi_\varepsilon = \sum_{k\in\Lambda^{(\varepsilon)}} c_k^{(\varepsilon)} w_k^{(\varepsilon)}, \quad |\Lambda^{(\varepsilon)}| < \infty, \quad \sum_{k\in\Lambda^{(\varepsilon)}} |c_k^{(\varepsilon)}| \leq M^*;$$

and, for $k \in \Lambda^{(\varepsilon)}$,

$$\|w_k^{(\varepsilon)}\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} = 1 \quad \text{and} \quad w_k^{(\varepsilon)} \in \bigotimes_{i=1}^{N} \mathrm{H}(D_i; M_i).$$

By virtue of the isometric isomorphism described in (2.4), the above relations imply

$$\left\|\mathsf{M}^{-1}\varphi - \mathsf{M}^{-1}\chi_\varepsilon\right\|_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})} \leq \varepsilon, \quad \mathsf{M}^{-1}\chi_\varepsilon = \sum_{k\in\Lambda^{(\varepsilon)}} c_k^{(\varepsilon)} \mathsf{M}^{-1} w_k^{(\varepsilon)},$$

and, for $k \in \Lambda^{(\varepsilon)}$,

$$\|\mathsf{M}^{-1} w_k^{(\varepsilon)}\|_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})} = 1 \quad \text{and} \quad \mathsf{M}^{-1} w_k^{(\varepsilon)} \in \bigotimes_{i=1}^{N} \mathrm{H}^1_{M_i}(D_i),$$

the last relation being a consequence of the tensor-product structure of the Maxwellian M. Thus we have shown that $\mathsf{M}^{-1}\varphi \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ can be approximated to within any positive tolerance ε in the norm of $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ by finite linear combinations of normalized members of $\bigotimes_{i\in[N]} \mathrm{H}^1_{M_i}(D_i)$ with the coefficients of the linear combinations having their absolute sum bounded by M^*. In other words, the membership of $\varphi \in \mathcal{A}_1$ implies the membership of $\mathsf{M}^{-1}\varphi$ in the $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$-based analogue of $\mathcal{A}_1$, namely,

$$\mathcal{B}_1 := \bigcup_{M>0} \overline{\mathcal{B}_1^o(M)}, \tag{3.31a}$$

where

$$\mathcal{B}_1^o(M) := \Bigg\{ \varphi \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})\colon \varphi = \sum_{k\in\Lambda} c_k w_k,\ w_k \in \bigotimes_{i=1}^{N} \mathrm{H}^1_{M_i}(D_i),\ \|w_k\|_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})} = 1, \\ |\Lambda| < \infty \quad \text{and} \quad \sum_{k\in\Lambda} |c_k| \leq M \Bigg\}, \tag{3.31b}$$

and equipped with the norm

$$\|\varphi\|_{\mathcal{B}_1} := \inf\left\{ M > 0\colon \varphi \in \overline{\mathcal{B}_1^o(M)} \right\}. \tag{3.31c}$$

In a completely analogous way, the membership of $\mathsf{M}^{-1}\varphi$ in $\mathcal{B}_1$ implies the membership of φ in $\mathcal{A}_1$. We then have the relations

$$\mathcal{A}_1 = \mathsf{M}\,\mathcal{B}_1, \qquad \|\cdot\|_{\mathcal{A}_1} = \left\|\mathsf{M}^{-1}\cdot\right\|_{\mathcal{B}_1}, \tag{3.32}$$

where the last equality follows from the fact that the coefficients of the approximations to φ are the same as the coefficients of the corresponding approximations to $\mathsf{M}^{-1}\varphi$.

As the definition of $\mathcal{A}_1$ given in (3.30) is fairly abstract, it is of interest to have conditions in terms of regularity that guarantee membership in $\mathcal{A}_1$ analogous to the conditions provided in [**LBLM09**, Remark 4] for the Separated Representation strategy applied to the Laplacian defined on a tensor product of one-dimensional domains. This is the theme of the next section. Because of the identity (3.32), we can pose the problem in terms of membership in the $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$-based $\mathcal{B}_1$ instead with no loss of generality and with a substantial gain in succinctness of exposition; thus we shall henceforth phrase our results in terms of $\mathcal{B}_1$ rather than $\mathcal{A}_1$.

3.3. Characterization of subspaces of rapidly converging solutions

3.3.1. Eigenvalues. The hypotheses of Lemma A.5 are satisfied by the eigenvalue problems

$$\langle e^{(i)}, \varphi\rangle_{\mathrm{H}^1_{M_i}(D_i)} = \lambda^{(i)}\langle e^{(i)}, \varphi\rangle_{\mathrm{L}^2_{M_i}(D_i)} \qquad \forall\, \varphi \in \mathrm{H}^1_{M_i}(D_i), \tag{3.33}$$

(for $i \in [N]$ here and in what follows), and

$$\langle e, \varphi\rangle_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})} = \lambda\langle e, \varphi\rangle_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} \qquad \forall\, \varphi \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D}), \tag{3.34}$$

whence their solutions do have the distribution, orthogonality and spanning properties stated in that lemma (the hypothesis $\overline{V} = H$, which is not discussed elsewhere, follows from the density of infinitely differentiable and compactly supported functions in any weighted L^2 space). In particular, they have sequences of solutions (eigenpairs) $\left((\lambda_n^{(i)}, e_n^{(i)})\colon n \in \mathbb{N}\right)$ and $((\lambda_n, e_n)\colon n \in \mathbb{N})$, respectively, with

$$\varphi \in \mathrm{L}^2_{M_i}(D_i) \quad \text{and} \quad \sum_{n=1}^{\infty} \lambda_n^{(i)} \langle \varphi, e_n^{(i)}\rangle^2_{\mathrm{L}^2_{M_i}(D_i)} < \infty \iff \varphi \in \mathrm{H}^1_{M_i}(D_i), \tag{3.35}$$

and

$$\varphi \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}) \quad \text{and} \quad \sum_{n=1}^{\infty} \lambda_n \langle \varphi, e_n\rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} < \infty \iff \varphi \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D}). \tag{3.36}$$

Next, we exploit the special tensor-product structure of the full Maxwellian M to characterize the eigenpairs of its associated eigenvalue problem (3.34) in terms of the eigenpairs of the eigenvalue problem (3.33) associated to the partial Maxwellians M_i.

Lemma 3.16. *The net* $\left((\lambda_{\boldsymbol{n}}, e_{\boldsymbol{n}})\colon \boldsymbol{n} = (n_1, \ldots, n_N) \in \mathbb{N}^N\right)$ *is a complete system of solutions of the eigenvalue problem* (3.34)*, where*

$$\lambda_{\boldsymbol{n}} := 1 + \sum_{i=1}^{N} (\lambda_{n_i}^{(i)} - 1) \qquad \textit{and} \qquad e_{\boldsymbol{n}} := \bigotimes_{i=1}^{N} e_{n_i}^{(i)}. \tag{3.37}$$

Proof. Given $\tau = \bigotimes_{i\in[N]} \tau^{(i)} \in \bigotimes_{i\in[N]} \mathrm{C}_0^\infty(D_i)$, we have that

$$\begin{aligned}\langle e_{\boldsymbol{n}}, \tau\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})} &= \langle e_{\boldsymbol{n}}, \tau\rangle_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} + \sum_{j=1}^{N} \left\langle \nabla e^{(j)}_{n_j}, \nabla \tau^{(j)} \right\rangle_{[\mathrm{L}^2_{M_j}(D_j)]^d} \prod_{\substack{i=1\\ i\neq j}}^{N} \left\langle e^{(i)}_{n_i}, \tau^{(i)} \right\rangle_{\mathrm{L}^2_{M_i}(D_i)} \\ &= \langle e_{\boldsymbol{n}}, \tau\rangle_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} + \sum_{j=1}^{N} (\lambda^{(j)}_{n_j} - 1) \left\langle e^{(j)}_{n_j}, \tau^{(j)} \right\rangle_{\mathrm{L}^2_{M_j}(D_j)} \prod_{\substack{i=1\\ i\neq j}}^{N} \left\langle e^{(i)}_{n_i}, \tau^{(i)} \right\rangle_{\mathrm{L}^2_{M_i}(D_i)} \\ &= \lambda_{\boldsymbol{n}} \langle e_{\boldsymbol{n}}, \tau\rangle_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})}\,.\end{aligned}$$

Since the span of $\bigotimes_{i=1}^N \mathrm{H}(D_i; M_i)$ is dense in $\mathrm{H}(\mathsf{D};\mathsf{M})$ (as is readily seen from Corollary 3.8 and (2.4)), the equality of the first and the last expression in the chain of equalities above is valid for all $\tau \in \mathrm{H}(\mathsf{D};\mathsf{M})$. Hence, $(\lambda_{\boldsymbol{n}}, e_{\boldsymbol{n}})$ is an eigenpair of (3.34). Further, we deduce from the chain of equalities above that $e_{\boldsymbol{n}}$ is orthogonal to $e_{\boldsymbol{m}}$ in both $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ and $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ if $\boldsymbol{n} \neq \boldsymbol{m}$.

From (A.5) in Lemma A.5, for $i \in [N]$,

$$\overline{\operatorname{span}\left(e^{(i)}_n \colon n \geq 1\right)} = \mathrm{H}^1_{M_i}(D_i).$$

Hence, invoking Corollary 3.8 and (2.4) we obtain that

$$\overline{\bigotimes_{i=1}^{N} \operatorname{span}\left(e^{(i)}_n \colon n \geq 1\right)} \subset \overline{\operatorname{span}\left(e_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right)} = \mathrm{H}^1_{\mathsf{M}}(\mathsf{D}).$$

Thus, $\left(e_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ forms an orthogonal system that spans $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$. Therefore, via Theorem VI.9 of [**Bre83**], all the eigenpairs of the (full) Maxwellian eigenvalue problem (3.34) have the form $(\lambda_{\boldsymbol{n}}, e_{\boldsymbol{n}})$ as given in (3.37) (modulo linear combinations of eigenfunctions belonging to the same eigenspace). □

It follows from Lemma 3.16 that the eigenvalues and eigenfunctions of (3.34) are more naturally indexed by $\mathbb{N}^N$ than by $\mathbb{N}$; in what follows, we shall refrain from indexing *contra natura*.

3.3.2. Characterization via summability of Fourier coefficients. As the sequence $\left((\lambda_{\boldsymbol{n}}, e_{\boldsymbol{n}}) \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ is a complete system of eigenpairs of (3.34) (because of Lemma 3.16), (A.5) in Lemma A.5 ensures that, for all $\tau \in \mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$,

$$\tau = \sum_{\boldsymbol{n}\in\mathbb{N}^N} \left\langle \tau, \frac{e_{\boldsymbol{n}}}{\sqrt{\lambda_{\boldsymbol{n}}}} \right\rangle_{\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})} \frac{e_{\boldsymbol{n}}}{\sqrt{\lambda_{\boldsymbol{n}}}} = \sum_{\boldsymbol{n}\in\mathbb{N}^N} \sqrt{\lambda_{\boldsymbol{n}}}\, \langle \tau, e_{\boldsymbol{n}}\rangle_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} \frac{e_{\boldsymbol{n}}}{\sqrt{\lambda_{\boldsymbol{n}}}} \qquad \text{in } \mathrm{H}^1_{\mathsf{M}}(\mathsf{D}).$$

Hence, given the tensor-product structure of the $e_{\boldsymbol{n}}$ and the unit $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$-norm of the $e_{\boldsymbol{n}}/\sqrt{\lambda_{\boldsymbol{n}}}$, we can guarantee that $\tau \in \mathcal{B}_1$ (cf. (3.31)) if

$$\sum_{\boldsymbol{n}\in\mathbb{N}^N} \sqrt{\lambda_{\boldsymbol{n}}} |\langle \tau, e_{\boldsymbol{n}}\rangle_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})}| < \infty.$$

In turn, this holds if

$$A := \sum_{\boldsymbol{n}\in\mathbb{N}^N} \frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}} < \infty \qquad \text{and} \qquad B := \sum_{\boldsymbol{n}\in\mathbb{N}^N} \sigma_{\boldsymbol{n}} \langle \tau, e_{\boldsymbol{n}} \rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} < \infty, \tag{3.38}$$

where $\left(\sigma_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ is a sequence of positive real numbers that are to be chosen below. We note that the requirement of B being finite can be seen—for $\sigma_{\boldsymbol{n}} = \lambda_{\boldsymbol{n}}$, for example, this is certainly the case, as follows from (3.36)—as a regularity requirement on τ. Thus, there is a trade-off in (3.38) between the requirement that the $\sigma_{\boldsymbol{n}}$ grow fast enough to ensure the finiteness of A and the desirability of the $\sigma_{\boldsymbol{n}}$ growing slow enough to avoid demanding more regularity than necessary of the functions τ for which B is finite.

As a first step in formalizing the above we consider, given a net $\Sigma = \left(\sigma_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ with entries in $\mathbb{R}_{>0}$, the space of all those $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ functions for which the term B, as defined in (3.38), is finite; thus, we define

$$\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D}) := \left\{ \varphi \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}) \colon \sum_{\boldsymbol{n}\in\mathbb{N}^N} \sigma_{\boldsymbol{n}} \langle \varphi, e_{\boldsymbol{n}} \rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} < \infty \right\} \tag{3.39a}$$

and equip it with the norm

$$\|\varphi\|_{\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})} := \left(\sum_{\boldsymbol{n}\in\mathbb{N}^N} \sigma_{\boldsymbol{n}} \langle \varphi, e_{\boldsymbol{n}} \rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} \right)^{1/2}. \tag{3.39b}$$

It is readily seen that, if there exists a $\sigma > 0$ with $\sigma_{\boldsymbol{n}} \geq \sigma$ for all $\boldsymbol{n} \in \mathbb{N}^N$, then $\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$ is a separable Hilbert space that is continuously embedded in $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$. Further, if there exists a $\sigma' > 0$ such that $\sigma_{\boldsymbol{n}} \geq \sigma' \lambda_{\boldsymbol{n}}$ for all $\boldsymbol{n} \in \mathbb{N}^N$, then $\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$ is continuously embedded in $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ and, due to Lemma 2.24, it is compactly embedded in $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$.

At this stage we could just choose Σ to be, e.g., $\sigma_{\boldsymbol{n}} = \lambda_{\boldsymbol{n}} \|\boldsymbol{n}\|_2^{\alpha}$ for some $\alpha > N$ and an application of a multiple series version of the integral test for convergence (see, for example, [**GL10**, Proposition 7.57]) would render the sum A in (3.38) finite. However, the resulting space $\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$ would then still have quite an abstract description. What we therefore wish to do instead is to choose each $\sigma_{\boldsymbol{n}}$ as a suitable polynomial function of the $(\lambda^{(1)}, \ldots, \lambda^{(N)})$. Then, under certain reasonable conditions, which we will make explicit below, we shall be able to characterize the resulting space in terms of regularity properties. One of these conditions has to do with the fact that we can only know that A of (3.38) is finite, with $\sigma_{\boldsymbol{n}}$ as a certain polynomial function of the $\lambda_{\boldsymbol{n}}$, if we have some information about the asymptotic behavior of the $\lambda_{\boldsymbol{n}}$. Consequently, we adopt the following hypothesis.

Hypothesis C. *For each $i \in [N]$ there exist positive real numbers $c_1^{(i)}$ and $c_2^{(i)}$ and $n^{(i)} \in \mathbb{N}$ such that $n \geq n^{(i)}$ implies*

$$c_1^{(i)} n^{2/d} \leq \lambda_n^{(i)} \leq c_2^{(i)} n^{2/d},$$

where $\lambda_n^{(i)}$ is the n-th member of the (ordered, with repetitions according to multiplicity) sequence of eigenvalues of (3.33) *and d is the common dimension of the single-spring configuration domains D_i.*

Remark 3.17. Hypothesis C basically consists of assuming that, to some extent, the eigenvalues of the problem (3.33) behave like the eigenvalues of a regular elliptic operator such as the Poisson operator. We proved in Corollary 2.12 and Corollary 2.14 in Subsection 2.2.3 that if the partial Maxwellian M_i comes from either the FENE model (1.16) with parameter $b_i > 2$, the CPAIL model (1.17) with parameter $b_i > 3$, the TEAIL model (1.18) with parameter $b_i > 16/5$, the CP model (1.21) with parameter $b_i > 3$ or the Inverse Langevin model (1.19) with parameter $b_i > 3$, Hypothesis C holds.

Theorem 3.18. *Let* $\mathrm{T}^{(l)} = \left(\tau_{\boldsymbol{n}}^{(l)} : \boldsymbol{n} \in \mathbb{N}^N\right)$ *be defined by*

$$\tau_{\boldsymbol{n}}^{(l)} := \left(\sum_{i=1}^{N} \lambda_{n_i}^{(i)}\right)^l + \prod_{i=1}^{N} \left(\lambda_{n_i}^{(i)}\right)^l \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N. \tag{3.40}$$

Then,

$$\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(d+1)}}(\mathsf{D}) \subset \mathcal{B}_1. \tag{3.41}$$

Proof. According to the previous discussion, the stated inclusion will hold once we have shown that the infinite sum over $\boldsymbol{n} \in \mathbb{N}^N$ of $\lambda_{\boldsymbol{n}}/\tau_{\boldsymbol{n}}^{(d+1)}$ converges; i.e., that A in (3.38) is finite. To prove this, we start by noting that, modulo a decrease of $c_1^{(i)}$ and an increase of $c_2^{(i)}$, we can take $n^{(i)} = 1$ in Hypothesis C as a consequence of all the $\lambda_n^{(i)}$ being positive; we do so from now on. This, together with (3.37) and Hypothesis C, renders the chain of inequalities

$$\frac{\lambda_{\boldsymbol{n}}}{\tau_{\boldsymbol{n}}^{(d+1)}} \leq \frac{\sum_{i=1}^{N} \lambda_{n_i}^{(i)}}{\left(\sum_{i=1}^{N} \lambda_{n_i}^{(i)}\right)^{d+1} + \prod_{i=1}^{N} \left(\lambda_{n_i}^{(i)}\right)^{d+1}} \leq C \frac{\sum_{i=1}^{N} n_i^{2/d}}{\left(\sum_{i=1}^{N} n_i^{2/d}\right)^{d+1} + \prod_{i=1}^{N} \left(n_i^{2/d}\right)^{d+1}} \tag{3.42}$$

valid for all $\boldsymbol{n} \in \mathbb{N}^N$ and some $C > 0$ that depends on the $c_1^{(i)}$, the $c_2^{(i)}$, N and d only. Clearly, it will be enough to show that the right-most expression in (3.42) results in a convergent series.

At this stage we note that we can use an already mentioned multiple series version of the integral test for convergence to state that the infinite sum of the right-most expressions in (3.42) will converge if, and only if, the integral

$$I_1 := \int_{[1,\infty)^N} \frac{\sum_{i=1}^{N} x_i^{2/d}}{\left(\sum_{i=1}^{N} x_i^{2/d}\right)^{d+1} + \prod_{i=1}^{N} \left(x_i^{2/d}\right)^{d+1}} \,\mathrm{d}x$$

is finite. To check the latter, it is convenient do introduce the transformation

$$\begin{aligned} x_1 &= r\cos(\phi_1)^d, \\ x_2 &= r\sin(\phi_1)^d\cos(\phi_2)^d, \\ x_3 &= r\sin(\phi_1)^d\sin(\phi_2)^d\cos(\phi_3)^d, \\ &\vdots \\ x_{N-1} &= r\sin(\phi_1)^d\cdots\sin(\phi_{N-2})^d\cos(\phi_{N-1})^d, \\ x_N &= r\sin(\phi_1)^d\cdots\sin(\phi_{N-2})^d\sin(\phi_{N-1})^d, \end{aligned} \tag{3.43}$$

which maps $[1,\infty)\times[0,\pi/2]\times\cdots\times[0,\pi/2]$ into $\mathbb{R}^N_{\geq 0}\setminus B_{2/d}(0,1)$. Here we denote by $B_{2/d}(\hat{x},R)$ the ball centered at $\hat{x}$ and with generalized radius R in the $2/d$-quasinorm (norm if $d\leq 2$) of $\mathbb{R}^N$; i.e., $\{x\in\mathbb{R}^N\colon \sum_{i=1}^N |x_i-\hat{x}_1|^{2/d}\leq R^{2/d}\}$.

The Jacobian determinant of the transformation (3.43) is $r^{N-1}g(\phi_1,\dots,\phi_{N-1})$, where

$$g(\phi_1,\dots,\phi_{N-1}) := \prod_{i=1}^{N-1}\left[\left(\sin(\phi_i)^d\right)^{N-1-i} w_d(\phi_i)\right]$$

and

$$w_d(\theta) = \cos(\theta)^d(\sin(\theta)^d)' - \sin(\theta)^d(\cos(\theta)^d)' = d\cos(\theta)^{d-1}\sin(\theta)^{d-1}.$$

Changing variables in the integral defining I_1 according to (3.43) and enlarging the domain of integration to one that is more convenient, we obtain that the finiteness of I_1 is implied by the finiteness of

$$I_2 := \int_0^{\pi/2}\cdots\int_0^{\pi/2}\int_1^{\infty} \frac{r^{2/d+N-1}g(\phi_1,\dots,\phi_{N-1})}{\left(r^{2/d}\right)^{d+1} + r^{2/d\times N\times(d+1)}f(\phi_1,\dots,\phi_{N-1})}\,\mathrm{d}r\,\mathrm{d}\phi_{N-1}\cdots\,\mathrm{d}\phi_1,$$

where f is defined by

$$f(\phi_1,\dots,\phi_{N-1}) := \left[\prod_{i=1}^N x_i^{2/d}\right]^{d+1} \Big/ r^{2/d\times N\times(d+1)} = \prod_{i=1}^{N-1}\left[\cos(\phi_i)^{2(d+1)}\sin(\phi_i)^{2(N-i)(d+1)}\right].$$

Now, let $p\in(1,\infty)$ and let q be its Hölder conjugate. Setting

$$\begin{aligned} a &= p^{1/p} r^{\frac{2(d+1)}{dp}}, \\ b &= q^{1/q} r^{\frac{2N(d+1)}{dq}} f(\phi_1,\dots,\phi_{N-1})^{1/q}, \end{aligned}$$

we can use Young's inequality $ab \leq a^p/p + b^q/q$ with $p, q \in (1, \infty)$ and $a, b \geq 0$, to bound

$$\begin{aligned} I_2 &\leq C_{d,N,p} \int_0^{\pi/2} \cdots \int_0^{\pi/2} \int_1^{\infty} \frac{r^{2/d+N-1} g(\phi_1, \ldots, \phi_{N-1})}{r^{\frac{2(d+1)}{dp} + \frac{2N(d+1)}{dq}} f(\phi_1, \ldots, \phi_{N-1})^{1/q}} \, \mathrm{d}r \, \mathrm{d}\phi_{N-1} \cdots \, \mathrm{d}\phi_1 \\ &= C_{d,N,p} \int_1^{\infty} r^{\frac{2}{d}+N-1-\frac{2(d+1)}{dp}-\frac{2N(d+1)}{dq}} \, \mathrm{d}r \\ &\quad \times \prod_{i=1}^{N-1} \int_0^{\pi/2} \cos(\phi_i)^{d-1-\frac{2(d+1)}{q}} \sin(\phi_i)^{d(N-1-i)+d-1-\frac{2(N-i)(d+1)}{q}} \, \mathrm{d}\phi_i, \end{aligned} \tag{3.44}$$

where $C_{d,N,p} = d^{N-1} p^{-1/p} q^{-1/q}$. The radial univariate integral in (3.44) will be finite if the exponent on r is less than -1; that is,

$$p > \frac{2(N-1)(d+1)}{N(d+2)-2}. \tag{3.45}$$

In order to tackle the angular integrals in (3.44) we note that if $\alpha > -1$ and $\beta > -1$ (see [**AR10**, equation 5.14.2]),

$$\int_0^{\pi/2} \sin(\theta)^{\alpha} \cos(\theta)^{\beta} \, \mathrm{d}\theta = \frac{1}{2} \int_0^1 t^{\frac{\alpha-1}{2}} (1-t)^{\frac{\beta-1}{2}} \, \mathrm{d}s = \frac{\Gamma\left(\frac{\alpha+1}{2}\right) \Gamma\left(\frac{\beta+1}{2}\right)}{2\,\Gamma\left(\frac{\alpha+\beta+2}{2}\right)} < \infty. \tag{3.46}$$

By (3.46), the angular integrals in (3.44) will be finite if $d - 1 - 2(d+1)/q > -1$ and, for $i \in [N-1]$, $d(N-1-i) + d - 1 - 2(N-i)(d+1)/q > -1$; this reduces to

$$p < \frac{2(d+1)}{d+2}. \tag{3.47}$$

As both $d \geq 1$ and $N \geq 1$, conditions (3.45) and (3.47) can be satisfied simultaneously; hence, there exists a value of $p \in \left(\max\left(1, \frac{2(N-1)(d+1)}{N(d+2)-2}\right), \frac{2(d+1)}{d+2}\right)$ such that the bound in (3.44) is nontrivial and thus I_2 is finite and our desired result holds. □

Remark 3.19. By replacing the functions $\cos(\cdot)^d$ and $\sin(\cdot)^d$ by the functions

$$\operatorname{sign}(\cos(\cdot)) \, |\cos(\cdot)|^d \qquad \text{and} \qquad \operatorname{sign}(\sin(\cdot)) \, |\sin(\cdot)|^d,$$

respectively, the transformation (3.43), extended to $[0, \infty) \times [0, \pi] \times \cdots \times [0, \pi] \times [0, 2\pi]$, ranges over the whole of $\mathbb{R}^N$. We give a graphical depiction of the transformation in Figure 3.1.

For later reference we introduce another family of weights that also produces subspaces of $\mathcal{B}_1$.

Theorem 3.20. *Let* $\Upsilon^{(l)} = \left(\upsilon_{\boldsymbol{n}}^{(l)} : \boldsymbol{n} \in \mathbb{N}^N\right)$ *be defined by*

$$\upsilon_{\boldsymbol{n}}^{(l)} := \left(\sum_{i=1}^{N} \lambda_{n_i}^{(i)}\right)^l \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N. \tag{3.48}$$

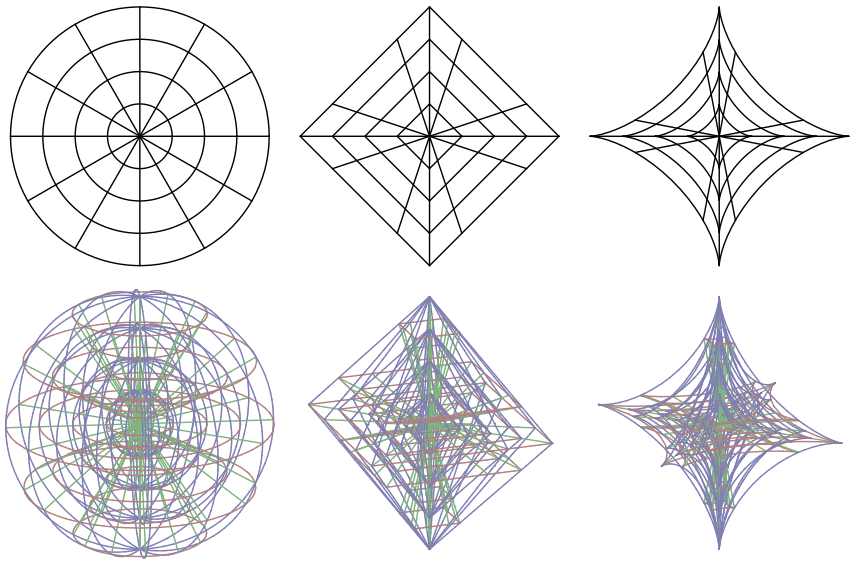

FIGURE 3.1. Coordinate lines corresponding to the transformation (3.43) with $N = 2$ (top row), $N = 3$ (bottom row) and d equal to 1, 2 and 3 (left, center and right columns, respectively). In all the plots the radial variable r was truncated at 1.

Then,

$$\mathrm{H}^{\Upsilon^{(l)}}_{\mathsf{M}}(\mathsf{D}) \subset \mathcal{B}_1 \quad \textit{if} \quad l > 1 + \tfrac{1}{2}Nd. \tag{3.49}$$

Proof. The proof of this theorem is similar to the proof of Theorem 3.18, except this time the result hinges on the finiteness of the integral

$$\int_0^{\pi/2} \cdots \int_0^{\pi/2} \int_1^{\infty} r^{(2/d)(1-l)+N-1} g(\phi_1, \ldots, \phi_{N-1}) \, \mathrm{d}r \, \mathrm{d}\phi_{N-1} \cdots \, \mathrm{d}\phi_1,$$

with g again being r^{1-N} times the Jacobian determinant of the transformation of (3.43). As $d \geq 1$ and $N \geq 1$, finiteness of this integral follows from the condition $(2/d)(1-l)+N-1 < -1$, and hence the stated result. □

Remark 3.21. It is easy enough to see, from the proofs of Theorem 3.18 and Theorem 3.20, that the inclusions (3.41) and (3.49) actually are embeddings of spaces. It would be of interest to have information on the norm of those embeddings beyond the mere acknowledgement of their finiteness; their dependence on N is of particular interest.

In the case of (3.41), the dependence on N of the norm of the embedding is, essentially[1], that of square root of $\sum_{\boldsymbol{n} \in \mathbb{N}^N} \lambda_{\boldsymbol{n}} / \tau_{\boldsymbol{n}}^{(d+1)}$. Now, the dependence on N of that quantity is not easy to estimate directly. The reason is that the multiple series version of the integral test of convergence we use (i.e., [**GL10**, Proposition 7.57]) does not directly bound the multiple series by the associated multiple integral. Instead, this integral test of convergence gives,

[1]The dependence on N of the constant C of (3.42) can be suppressed.

beside the interdependence of the finiteness of the sum and the integral, the composite bound

$$\sum_{\boldsymbol{n}\in\mathbb{N}^N} f(\boldsymbol{n}) \le \int_{[1,\infty)^N} f(\boldsymbol{n})\,\mathrm{d}\boldsymbol{n} + \sum_{\substack{\boldsymbol{n}\in\mathbb{N}^N\\ \boldsymbol{n}\not\geq 2}} f(\boldsymbol{n}),$$

where f, the function being summed and integrated, is monotonic decreasing (in all directions). Then, besides the difficulty given by the fact that the resulting integral evaluates to an expression depending on the additional parameter p in an algebraically complicated way, there is the issue of estimating the sum over $\{\boldsymbol{n} \in \mathbb{N}^N \colon \boldsymbol{n} \not\geq 2\}$, which only slightly less complicated than estimating the original sum.

One alternative is circumventing the use of the integral test and replacing the weights $\tau_{\boldsymbol{n}}^{(d+1)}$ defined in (3.40) with the smaller weights $\hat{\tau}_{\boldsymbol{n}}^{(d+1)} := \left(\prod_{i\in[N]} n_i\right)^{d+1}$. Then, it is straightforward to compute the corresponding sum and show that $\mathrm{H}_{\mathsf{M}}^{\hat{\mathrm{T}}^{(d+1)}}(\mathsf{D})$ is also embedded in $\mathcal{B}_1$ with a norm bounded by an N-independent constant times

$$N^{1/2}\,\zeta_{\mathrm{R}}\left(\frac{2(d+1)}{d}\right)^{\frac{N-1}{2}}, \tag{3.50}$$

where ζ_{R} stands for Riemann's zeta function. As $\hat{\tau}_{\boldsymbol{n}}^{(d+1)} \le \tau_{\boldsymbol{n}}^{(d+1)}$ for $\boldsymbol{n} \in \mathbb{N}^N$, we have that the embedding of $\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(d+1)}}(\mathsf{D})$ in $\mathcal{B}_1$ has its norm bounded by the same quantity. Moreover, the important results of the next subsection involving $\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(d+1)}}(\mathsf{D})$, namely, Lemma 3.29 and Theorem 3.30, remain valid if the net $\hat{\mathrm{T}}^{(d+1)} := \left(\hat{\tau}_{\boldsymbol{n}}^{(d+1)} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ is used instead of the net $\mathrm{T}^{(d+1)} := \left(\tau_{\boldsymbol{n}}^{(d+1)} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$. However, we cannot use this technique to discard the possibility of the norm of the embedding of $\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(d+1)}}(\mathsf{D})$ growing at a rate slower than that of (3.50).

The question in the case of the embedding (3.49) has a different character. Indeed, the parameter l of the weights $v_{\boldsymbol{n}}^{(l)}$ itself depends on N; hence, depending on how closely one chooses l from its lower bound $1 + \frac{1}{2}Nd$, one can get abitrarily large embedding norm growth rates with respect to N.

The definition of $\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(d+1)}}(\mathsf{D})$ given by (3.40) is certainly less abstract than the definition of $\mathcal{B}_1$ (given in (3.31)). However, we can describe subspaces of the former space in even less abstract terms by showing that certain regularity conditions translate into summability conditions expressed in terms of Fourier coefficients, such as those that define $\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(d+1)}}(\mathsf{D})$ (cf. (3.39a)). In order to understand the appropriate regularity requirements for this purpose, we need to study the regularity properties of certain degenerate elliptic operators in Maxwellian-weighted Sobolev spaces.

3.3.3. Characterization via regularity. We start by adopting two further hypotheses.

Hypothesis D. *For $i \in [N]$ the spring potential U_i is monotonic increasing and convex.*

Hypothesis E. *For $i \in [N]$ there exists a distance $\gamma_i \in (0, \sqrt{b_i})$, an exponent $\alpha_i > 1$ and a function $h_i \in \mathrm{C}^3([0,\gamma_i])$ that is positive on $[0,\gamma_i]$, such that*

$$M_i(\boldsymbol{p}) = h_i(\mathfrak{d}_i(\boldsymbol{p}))\,\mathfrak{d}_i(\boldsymbol{p})^{\alpha_i}$$

for all $\boldsymbol{p} \in D_i$ such that $\mathfrak{d}_i(\boldsymbol{p}) \in (0,\gamma_i)$, where $\mathfrak{d}_i$ is the distance-to-the-boundary function in D_i.

Remark 3.22. Hypothesis D can be regarded as a strengthening of Hypothesis A. It is easy to check, by direct calculation, that springs obeying the FENE model (1.16), the CPAIL model (1.17), the TEAIL model (1.18) or the CP model (1.21) comply with it. The corresponding result for the Inverse Langevin model Equation 1.19 was proved in (2.9) in Subsection 2.2.2.

With Hypothesis E we are restricting ourselves, essentially, to power weights. The compliance of the FENE, the CPAIL, the TEAIL and CP force models with this hypothesis is also easy to check if their parameter b_i is strictly greater than 2, 3, 15/6 and 3, respectively. We do not know if Hypothesis E is satisfied by the Maxwellian associated with the Inverse Langevin force law (1.19).

Lemma 3.23. *For $i \in [N]$,*

(a) the space $\mathrm{C}_0^\infty(D_i)$ is dense in $\mathrm{H}^1_{M_i}(D_i)$;
(b) the space $\mathrm{C}^\infty(\overline{D_i})$ is dense in $\mathrm{H}^m_{M_i}(D_i)$, for $m \in \mathbb{N}$.

Proof. In Proposition 9.10 (resp. Theorem 7.2) of [**Kuf85**] the result (a) (resp. (b)) is stated for weights that are powers greater than 1 (resp. greater or equal than 0) of the distance-to-the-boundary function; the bilateral boundedness of the function h_i by positive constants, implied by Hypothesis E, extends the statement to our case. □

The additional requirements on the potentials U_i, $i \in [N]$, and the preceding lemma allow us to prove a first elliptic regularity result.

Lemma 3.24. *Let $i \in [N]$. If $g \in \mathrm{L}^2_{M_i}(D_i)$, the solution $z \in \mathrm{H}^1_{M_i}(D_i)$ of*

$$\langle z, \varphi\rangle_{\mathrm{H}^1_{M_i}(D_i)} = \langle g, \varphi\rangle_{\mathrm{L}^2_{M_i}(D_i)} \qquad \forall\, \varphi \in \mathrm{H}^1_{M_i}(D_i) \tag{3.51}$$

obeys the regularity estimate

$$\|z\|_{\mathrm{H}^2_{M_i}(D_i)} + \left\| \frac{1}{M_i} \operatorname{div}(M_i \nabla z) \right\|_{\mathrm{L}^2_{M_i}(D_i)} \leq C_i \, \|g\|_{\mathrm{L}^2_{M_i}(D_i)}$$

for some $C_i > 0$.

Proof. From Hypothesis A and Hypothesis D we know that the function $V_i \colon D_i \to \mathbb{R}$ defined by $V_i(\boldsymbol{p}) := \frac{1}{2} U_i(\frac{1}{2}\,|\boldsymbol{p}|^2) \in \mathbb{R}$ is convex and diverges to infinity as its argument approaches the boundary of D_i from within. Then, it follows from Theorem 3.4 of [**DPL04**] and the density of $\mathrm{C}_0^\infty(D_i)$ in $\mathrm{H}^1_{M_i}(D_i)$ given in part (a) of Lemma 3.23 that there exists a unique solution $\tilde{z}$

in $\{u \in \mathrm{H}^2_{M_i}(D_i)\colon \nabla V_i \cdot \nabla u \in \mathrm{L}^2_{M_i}(D_i)\}$ to the $\mathrm{L}^2_{M_i}(D_i)$ equation

$$\tfrac{1}{2}\tilde{z} - \tfrac{1}{2}\Delta\tilde{z} + \nabla V_i \cdot \nabla \tilde{z} = \tfrac{1}{2}g, \tag{3.52}$$

and it obeys the estimates

$$\|\tilde{z}\|_{\mathrm{L}^2_{M_i}(D_i)} \leq 2\left\|\tfrac{1}{2}g\right\|_{\mathrm{L}^2_{M_i}(D_i)} = \|g\|_{\mathrm{L}^2_{M_i}(D_i)}\,, \tag{3.53a}$$

$$\|\nabla\tilde{z}\|_{[\mathrm{L}^2_{M_i}(D_i)]^d} \leq 2\sqrt{2}\left\|\tfrac{1}{2}g\right\|_{\mathrm{L}^2_{M_i}(D_i)} = \sqrt{2}\,\|g\|_{\mathrm{L}^2_{M_i}(D_i)}\,, \tag{3.53b}$$

$$\|\nabla\nabla\tilde{z}\|_{[\mathrm{L}^2_{M_i}(D_i)]^{d\times d}} \leq 4\left\|\tfrac{1}{2}g\right\|_{\mathrm{L}^2_{M_i}(D_i)} = 2\,\|g\|_{\mathrm{L}^2_{M_i}(D_i)}\,. \tag{3.53c}$$

The regularity of M_i and $\tilde{z}$ admits the use of the Leibniz formula for the product of a regular distribution and a continuously differentiable function provided in Lemma A.2 in Section A.1. We can then write $M_i\Delta\tilde{z} - 2M_i\nabla V_i \cdot \nabla\tilde{z} = \operatorname{div}(M_i\nabla\tilde{z})$ (for this we have used that M_i is proportional to $\exp(-2V_i)$ (cf. (1.24)). Plugging this into (3.52) and (3.53a) gives

$$\left\|\frac{1}{M_i}\operatorname{div}(M_i\nabla\tilde{z})\right\|_{\mathrm{L}^2_{M_i}(D_i)} \leq \|g\|_{\mathrm{L}^2_{M_i}(D_i)} + \|\tilde{z}\|_{\mathrm{L}^2_{M_i}(D_i)} \leq 2\,\|g\|_{\mathrm{L}^2_{M_i}(D_i)}\,. \tag{3.54}$$

Multiplying (3.52) by $2M_i$, using the Leibniz formula for the product of a regular distribution and a continuously differentiable function again, and testing with any $\varphi \in \mathrm{C}^\infty_0(D_i) \subset \mathrm{L}^2_{M_i}(D_i)$, we find that

$$\int_{D_i}\tilde{z}\varphi M_i + \int_{D_i}(\nabla\tilde{z}\cdot\nabla\varphi)M_i = \int_{D_i} g\varphi.$$

It follows from the density of $\mathrm{C}^\infty_0(D_i)$ in $\mathrm{H}^1_{M_i}(D_i)$ and the uniqueness of the solution z of (3.51) that $z = \tilde{z}$ and hence (3.53) and (3.54) give the desired result. □

In order to obtain an iterated elliptic regularity result, we need the technical lemma that follows.

Lemma 3.25 (Hardy inequalities)**.** *Let $H > 0$. Then, there exists $C_H > 0$ such that*

$$\int_0^H \frac{1}{y^2}\left(\int_0^y f(s)\,\mathrm{d}s\right)^2 \mathrm{d}y \leq C_H\int_0^H f(s)^2\,\mathrm{d}s \qquad \forall\, f \in \mathrm{L}^1((0,H)). \tag{3.55}$$

If $\alpha > 1$, then there exists $C_{H,\alpha}$ such that

$$\int_0^H y^{\alpha-2}f(y)^2\,\mathrm{d}y \leq C_{H,\alpha}\int_0^H y^\alpha\left[f(y)^2 + f'(y)^2\right]\mathrm{d}y \qquad \forall\, f \in \mathrm{H}^1_{(\cdot)^\alpha}((0,H)). \tag{3.56}$$

Proof. The inequality (3.55) is a direct consequence of the standard Hardy inequality (the $H = \infty$ case); see, for example, [**DiB02**, Proposition VIII.18.1]. Alternatively, see [**OK90**, Theorem 1.14] for a very general form, which encompasses (3.55).

To prove (3.56) we will use a procedure inspired by the proof of Theorem 8.2 of [**Kuf85**]. The first ingredient is the inequality

$$\int_0^H y^{\alpha-2}f(y)^2\,\mathrm{d}y \leq C_1\int_0^H y^\alpha f'(y)^2\,\mathrm{d}y$$

valid for all f in $\mathrm{C}^1([0,H])$ such that $f(H)=0$ (see, e.g., [**OK90**, Example 6.8.ii]). Let now φ_0 and φ_1 in $\mathrm{C}^1([0,H])$ form a partition of unity subordinate to the covering $H = (0,2H/3)\cup(H/3,H)$. Then, given any $f\in\mathrm{C}^1([0,H])$, let $f_0:=\varphi_0 f$ and $f_1:=\varphi_1 f$. Using the above inequality, the validity of (3.56) for $\mathrm{C}^1([0,H])$ functions follows from

$$\begin{aligned}
\|f\|_{\mathrm{L}^2_{(\cdot)^{\alpha-2}}((0,H))} &\le \|f_0\|_{\mathrm{L}^2_{(\cdot)^{\alpha-2}}((0,2H/3))} + \|f_1\|_{\mathrm{L}^2_{(\cdot)^{\alpha-2}}((H/3,H))}\\
&\le C_1^{1/2}\left\|f_0'\right\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,2H/3))} + \left\|(\cdot)^{-1}f_1\right\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((H/3,H))}\\
&\le C_1^{1/2}\left\|\varphi_0 f'\right\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,2H/3))} + C_1^{1/2}\left\|\varphi_0' f\right\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,2H/3))} + \frac{3}{H}\|f_1\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((H/3,H))}\\
&\le C_2\left\|f'\right\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,2H/3))} + C_3\|f\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,2H/3))} + C_4\|f\|_{\mathrm{L}^2_{(\cdot)^{\alpha}}((H/3,H))}\\
&\le C_5\left(\|f\|^2_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,H))} + \left\|f'\right\|^2_{\mathrm{L}^2_{(\cdot)^{\alpha}}((0,H))}\right)^{1/2}.
\end{aligned}$$

The validity of the inequality for all $f\in\mathrm{H}^1_{(\cdot)^\alpha}((0,H))$ is then a consequence of the density of $\mathrm{C}^1([0,H])$ functions in $\mathrm{H}^1_{(\cdot)^\alpha}((0,H))$, the completeness of $\mathrm{L}^2_{(\cdot)^{\alpha-2}}((0,H))$ and the continuity of the injection of that latter space into $\mathrm{L}^2_{(\cdot)^\alpha}((0,H))$. □

We shall now iterate Lemma 3.24: extra regularity for g implies extra regularity for z.

Lemma 3.26. *Let $i\in[N]$ and $g\in\mathrm{H}^2_{M_i}(D_i)$. Then, the solution $z\in\mathrm{H}^1_{M_i}(D_i)$ of*

$$\langle z,\varphi\rangle_{\mathrm{H}^1_{M_i}(D_i)} = \langle g,\varphi\rangle_{\mathrm{L}^2_{M_i}(D_i)} \qquad \forall\,\varphi\in\mathrm{H}^1_{M_i}(D_i) \tag{3.57}$$

obeys the regularity estimate

$$\|z\|_{\mathrm{H}^4_{M_i}(D_i)} \le C_i\,\|g\|_{\mathrm{H}^2_{M_i}(D_i)},$$

for some $C_i>0$.

Proof. The core of this proof is based on Lemmas 3.1 and 3.3 of [**Fre87**]. As their adaptation to our geometry is nontrivial, we give a detailed argument. Note that in this proof we shall omit the spring index i in order to avoid cluttering the notation.

Part 1: We start by describing a change of coordinates and how (3.51) transforms under it.

Given $\boldsymbol{p}\in\mathbb{R}^d$, let $\boldsymbol{p}'$ denote $(p_1,\dots,p_{d-1})\in\mathbb{R}^{d-1}$. Let ζ be some constant in $(0,1)$ and let us define, for $\varepsilon\in(0,\zeta]$, the sets

$$\tilde{U}_\varepsilon := P'_\varepsilon\times(0,\varepsilon\gamma) \qquad\text{and}\qquad U_\varepsilon := S(\tilde{U}_\varepsilon),$$

where γ is the distance (with its spring index omitted) mentioned in Hypothesis E and

$$P'_\varepsilon := \begin{cases} \varepsilon(-\pi/2,\pi/2) & \text{if } d=2,\\ \varepsilon(-1,1)\times\varepsilon(-\pi/2,\pi/2) & \text{if } d=3 \end{cases}$$

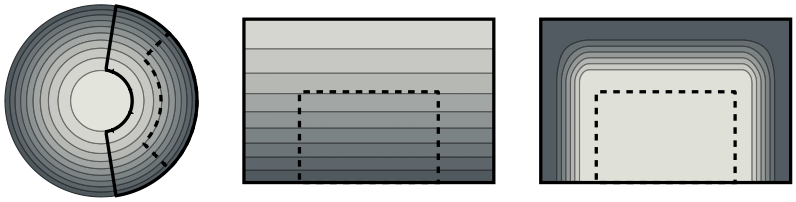

FIGURE 3.2. Illustration of the construction used in the proof of Lemma 3.26. From left to right: Contour plot of M on D with U_ζ and U_ε enclosed by the thick continuous and dashed lines, respectively; contour plot of $\tilde{M}$ on $\tilde{U}_\zeta$ with $\tilde{U}_\varepsilon$ enclosed by the thick dashed line; contour plot of an admissible $\tilde{\omega}$ on $\tilde{U}_\zeta$ with $\tilde{U}_\varepsilon$ enclosed by the thick dashed line.

and $S\colon \tilde{U}_\zeta \to U_\zeta$ is defined by the formula

$$S(\boldsymbol{p}) = \begin{cases} (\sqrt{b} - p_2)\,(\cos(p_1),\ \sin(p_1)) & \text{if } d = 2, \\ (\sqrt{b} - p_3)\left(\sqrt{1-p_1^2}\cos(p_2),\ \sqrt{1-p_1^2}\sin(p_2),\ p_1\right) & \text{if } d = 3, \end{cases} \qquad \forall\, \boldsymbol{p} \in \tilde{U}_\zeta.$$

Note that if $0 < \varepsilon_1 < \varepsilon_2 \leq \zeta$ then $U_{\varepsilon_1} \subset U_{\varepsilon_2} \subset D$ and $\tilde{U}_{\varepsilon_1} \subset \tilde{U}_{\varepsilon_2}$. Having its domain and defining formula carefully crafted for the purpose, the transformation S turns out to be invertible, orientation-preserving and $\mathrm{C}^\infty(\overline{\tilde{U}_\zeta})$-regular. All of this is easy to see if one takes into account that S is a variant of the polar (resp. spherical) to Cartesian coordinate transformation if $d = 2$ (resp. $d = 3$) with the radial variable being measured from the boundary of D and increasing towards its center. We denote the inverse of S by T; it, too, has uniformly bounded derivatives of all orders.

If f is a function with domain U_ζ we will write $\tilde{f} := f \circ S$. Then, $\varphi \in \mathrm{C}^\infty(\overline{U_\zeta}) \iff \tilde{\varphi} \in \mathrm{C}^\infty(\overline{\tilde{U}_\zeta})$. If m is a positive integer, part (b) of Lemma 3.23 states that $\mathrm{C}^\infty(\overline{D})$ is dense in $\mathrm{H}^m_M(D)$; as, for any $\varepsilon \in (0, \zeta]$, U_ε is regular enough (a Lipschitz domain), $\mathrm{C}^\infty(\overline{U_\varepsilon})$ is exactly the set of restrictions to U_ε of $\mathrm{C}^\infty(\overline{D})$ functions, whence $\mathrm{C}^\infty(\overline{U_\varepsilon})$ is dense in $\mathrm{H}^m_M(U_\varepsilon)$ as well. We also have, from Lemma A.4 in Section A.1, that $f \in \mathrm{H}^m_M(U_\varepsilon) \iff \tilde{f} \in \mathrm{H}^m_{\tilde{M}}(\tilde{U}_\varepsilon)$ and that

$$c_1(m)\|\tilde{f}\|_{\mathrm{H}^m_{\tilde{M}}(\tilde{U}_\varepsilon)} \leq \|f\|_{\mathrm{H}^m_M(U_\varepsilon)} \leq c_2(m)\|\tilde{f}\|_{\mathrm{H}^m_{\tilde{M}}(\tilde{U}_\varepsilon)} \qquad \forall\, f \in \mathrm{H}^m_M(U_\varepsilon), \tag{3.58}$$

where the positive constants c_1 and c_2 depend on m but can be chosen to be independent of ε.

As $\mathrm{C}^\infty(\overline{U_\zeta})$ is mapped by composition with S to $\mathrm{C}^\infty(\overline{\tilde{U}_\zeta})$ bijectively, it follows that $\mathrm{C}^\infty(\overline{\tilde{U}_\zeta})$ is dense in $\mathrm{H}^1_{\tilde{M}}(\tilde{U}_\zeta)$. The rules of calculus and the density of $\mathrm{C}^\infty(\overline{U_\zeta})$ and $\mathrm{C}^\infty(\overline{\tilde{U}_\zeta})$ functions in $\mathrm{H}^1_M(U_\zeta)$ and $\mathrm{H}^1_{\tilde{M}}(\tilde{U}_\zeta)$, respectively, give the equalities

$$\int_{U_\zeta} u\, v M = \int_{\tilde{U}_\zeta} \tilde{u}\, \tilde{v} \tilde{M} a \qquad \text{and} \qquad \int_{U_\zeta} \nabla u \cdot \nabla v M = \int_{\tilde{U}_\zeta} \nabla \tilde{u} A \nabla \tilde{v}^{\mathrm{T}} \tilde{M}, \tag{3.59a}$$

where we have used the shorthand notations

$$a = \det(\nabla S) \qquad \text{and} \qquad A = (\nabla S)^{-1}(\nabla S)^{-\mathrm{T}} \det(\nabla S). \tag{3.59b}$$

The first equality in (3.59a) is valid for u and v in $\mathrm{L}^2_M(U_\zeta)$ and the second for u and v in $\mathrm{H}^1_M(U_\zeta)$. Direct calculations give

$$A = A^{\mathrm{T}}, \qquad A_{k,d} = A_{d,k} = 0 \quad \forall\, k \in [d-1], \qquad A_{d,d} = a, \qquad \text{and} \qquad \partial_k a = 0 \quad \forall\, k \in [d-1], \tag{3.60}$$

which we will exploit later. The need for the last equality in (3.60) is the rationale behind taking the sine of the polar angle instead of the polar angle itself as the first argument of the transformation S in the case of $d = 3$. Additionally, by construction, $\tilde{M}$ is a function of the radial variable p_d only—namely, for all $\boldsymbol{p} \in \tilde{U}_\zeta$, $\tilde{M}(\boldsymbol{p}) = h(p_d)p_d^\alpha$, where h and α, with the spring index omitted, are those of Hypothesis E. Therefore,

$$\partial_k \tilde{M} = 0 \quad \forall\, k \in [d-1]. \tag{3.61}$$

Let us fix $\varepsilon \in (0, \zeta)$. For localization purposes we pick a $\mathrm{C}^\infty(U_\zeta)$ function ω with range $[0, 1]$, identically 1 in U_ε, with support bounded away from $\partial U_\zeta \setminus \partial D$ and such that $\partial_d \tilde{\omega}(\boldsymbol{p}) = 0$ if $\boldsymbol{p}$ lives within a finite distance $\gamma' > 0$ of $\partial\tilde{U}_\zeta \cap T(\partial D)$ (such a function is readily constructed as $\omega = \tilde{\omega} \circ T$ where $\tilde{\omega}(\boldsymbol{p}) = s(\boldsymbol{p}')t(p_d)$ and s and t are suitable mollified step functions). See Figure 3.2 for a depiction of the construction so far.

Now, as every member of $\mathrm{C}^\infty_0(\tilde{U}_\zeta)$ can be put in the form $\varphi \circ S$, where $\varphi \in \mathrm{C}^\infty_0(U_\zeta) \subset \mathrm{H}^1_M(U_\zeta)$, (3.57) and the equalities in (3.59) imply that $\tilde{z}$ obeys the distributional equation

$$-\operatorname{div}\left(\nabla\tilde{z}\, A\tilde{M}\right) + \tilde{z}a\tilde{M} = \tilde{g}\tilde{M}a \tag{3.62}$$

in $\tilde{U}_\zeta$.

Part 2: In this part we show that the relevant derivatives of $\tilde{z}$ in directions tangential to the radial (i.e., the d-th) coordinate possess additional regularity. The argument is a nontrivial adaptation of Lemma 3.1 of [**Fre87**].

Let $k \in [d-1]$. Then, using the Leibniz formula given by Lemma A.2 and applying some simple consequences of (3.60) and (3.61), the distributional equation (3.62) conduces to

$$\begin{aligned}-\operatorname{div}\left(\nabla(\partial_k\tilde{z})\, A\tilde{M}\right) + \partial_k\tilde{z}\, a\tilde{M} \\ = \partial_k\tilde{g}\, a\tilde{M} + \nabla\nabla\tilde{z}\colon \partial_k A\, \tilde{M} + \nabla\tilde{z}\cdot\operatorname{div}(\partial_k A)\tilde{M} + \nabla\tilde{M}\cdot(\nabla\tilde{z}\,\partial_k A),\end{aligned}$$

which serves as a stepping stone for

$$\begin{aligned}&-\operatorname{div}\left(\nabla(\tilde{\omega}\partial_k\tilde{z})\, A\tilde{M}\right) + \tilde{\omega}\partial_k\tilde{z}\, a\tilde{M} \\ &\quad= -\tilde{\omega}\operatorname{div}\left(\nabla(\partial_k\tilde{z})\, A\tilde{M}\right) + \tilde{\omega}\partial_k\tilde{z}\, a\tilde{M} - \nabla\tilde{\omega}\cdot\left(\nabla(\partial_k\tilde{z})\, A\tilde{M}\right) - \operatorname{div}\left(\partial_k\tilde{z}\,\nabla\tilde{\omega}\, A\tilde{M}\right) \\ &\quad= \tilde{\omega}\left[\partial_k\tilde{g}\, a\tilde{M} + \nabla\nabla\tilde{z}\colon \partial_k A\, \tilde{M} + \nabla\tilde{z}\cdot\operatorname{div}(\partial_k A)\tilde{M} + \nabla\tilde{M}\cdot(\nabla\tilde{z}\,\partial_k A)\right] \\ &\qquad - \nabla\tilde{\omega}\cdot\left(\nabla(\partial_k\tilde{z})\, A\tilde{M}\right) - \operatorname{div}\left(\partial_k\tilde{z}\,\nabla\tilde{\omega}\, A\tilde{M}\right)\end{aligned}$$

$$
\begin{aligned}
&= \tilde{\omega}\partial_k\tilde{g}\,a\tilde{M} + \tilde{\omega}\nabla\nabla\tilde{z}\colon \partial_k A\,\tilde{M} + \tilde{\omega}\nabla\tilde{z}\cdot\operatorname{div}(\partial_k A)\tilde{M} + \tilde{\omega}\nabla\tilde{M}\cdot(\nabla\tilde{z}\,\partial_k A)\\
&\qquad - 2\nabla\tilde{\omega}\cdot(\nabla(\partial_k\tilde{z})\,A\tilde{M}) - \partial_k\tilde{z}\,\operatorname{div}(\nabla\tilde{\omega}\,A\tilde{M})\\
&= \tilde{\omega}\partial_k\tilde{g}\,a\tilde{M} + \tilde{\omega}\nabla\nabla\tilde{z}\colon \partial_k A\,\tilde{M} + \tilde{\omega}\nabla\tilde{z}\cdot\operatorname{div}(\partial_k A)\tilde{M} + \tilde{\omega}\nabla\tilde{M}\cdot(\nabla\tilde{z}\,\partial_k A)\\
&\qquad\qquad - 2\nabla\tilde{\omega}\cdot(\nabla(\partial_k\tilde{z})\,A\tilde{M}) - \partial_k\tilde{z}\,\operatorname{div}(\nabla\tilde{\omega}\,A)\tilde{M} - \partial_k\tilde{z}\,\nabla\tilde{M}\cdot(\nabla\tilde{\omega}\,A). \qquad (3.63)
\end{aligned}
$$

We want to show that all the resulting terms are (the linear combination of) members of the space $a\tilde{M}\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\zeta) = \tilde{M}\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\zeta)$. Of the resulting seven terms above, the first three, the fifth and the sixth pose no problem, thanks to the regularity of g and Lemma 3.24. The fourth vanishes after making full use of the equalities in (3.60) and the sole dependence of $\tilde{M}$ on the radial variable—this is what the fourth equation in (3.60) is truly for. The membership of the seventh term in $\tilde{M}\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\zeta)$ stems from observing that

$$
\begin{aligned}
\left\|\partial_k\tilde{z}\,\nabla\tilde{M}\cdot(\nabla\tilde{\omega}\,A)/\tilde{M}\right\|_{\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\zeta)} &= \left\|\partial_k\tilde{z}\,\partial_d\tilde{M}\,\partial_d\tilde{\omega}\,a/\tilde{M}\right\|_{\mathrm{L}^2_{\tilde{M}}(P'_\zeta\times(\gamma',\zeta\gamma))}\\
&\leq \sup_{P'_\zeta\times(\gamma',\zeta\gamma)}\left(\partial_d\tilde{M}\,\partial_d\tilde{\omega}\,a/\tilde{M}\right)\|\partial_k\tilde{z}\|_{\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\zeta)} < \infty.
\end{aligned}
$$

Let $\hat{f}$, given some function f defined on $\tilde{U}_\zeta$, denote $f\circ T = f\circ S^{-1}$. Also, let $f_{(k)}$ denote the ratio of the right-hand side of (3.63) and $\tilde{M}a$. Then, (3.63) and the identities in (3.59) give

$$
-\operatorname{div}(M\nabla\widehat{\tilde{\omega}\partial_k\tilde{z}}) + \widehat{\tilde{\omega}\partial_k\tilde{z}}M = \widehat{f_{(k)}} \qquad (3.64)
$$

in U_ζ, with $\widehat{f_{(k)}} \in \mathrm{L}^2_M(U_\zeta)$. As the support of $\tilde{\omega}$ is bounded away from $\partial\tilde{U}_\zeta \setminus T(\partial D)$, we can extend $\widehat{\tilde{\omega}\partial_k\tilde{z}}$ and $\widehat{f_{(k)}}$ to the whole of D by zero while still satisfying (3.64). Then, Lemma 3.24 ensures that the extension of $\widehat{\tilde{\omega}\partial_k\tilde{z}}$ lives in $\mathrm{H}^2_M(D)$. It follows that $\tilde{\omega}\partial_k\tilde{z}$ lives in $\mathrm{H}^2_{\tilde{M}}(U_\zeta)$ and, consequently, $\partial_k\tilde{z} \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$.

This procedure can be iterated. Within U_ε the identity (3.63) particularizes to

$$
-\operatorname{div}\left(\nabla(\partial_k\tilde{z})\,A\tilde{M}\right) + \partial_k\tilde{z}\,a\tilde{M} = \partial_k g\,a\tilde{M} + \nabla\nabla\tilde{z}\colon\partial_k A\,\tilde{M} + \nabla\tilde{z}\cdot\operatorname{div}(\partial_k A)\tilde{M}.
$$

Let $g_{(k)} := \partial_k\tilde{g} + \nabla\nabla\tilde{z}\colon\partial_k A\,/a + \nabla\tilde{z}\cdot\operatorname{div}(\partial_k A)/a$ and let us redefine $\tilde{\omega}$ so that the role of $\tilde{U}_\zeta$ is now taken up by $\tilde{U}_\varepsilon$ and the role of the latter is taken up by $\tilde{U}_\delta$, where δ is some fixed number in $(0,\varepsilon)$. Thus, we can obtain an analogue to (3.63) for $\tilde{\omega}\partial_{l,k}\tilde{z}$, where $l,k\in[d-1]$:

$$
\begin{aligned}
&-\operatorname{div}(\nabla(\tilde{\omega}\partial_{l,k}\tilde{z})\,A\tilde{M}) + \tilde{\omega}\partial_{l,k}\tilde{z}\,a\tilde{M}\\
&\quad= \tilde{\omega}\partial_l g_{(k)}\,a\tilde{M} + \tilde{\omega}\nabla\nabla\partial_k\tilde{z}\colon\partial_l A\,\tilde{M} + \tilde{\omega}\nabla\partial_k\tilde{z}\cdot\operatorname{div}(\partial_l A)\tilde{M} + \tilde{\omega}\nabla\tilde{M}\cdot(\nabla(\partial_k\tilde{z})\,\partial_l A)\\
&\qquad\qquad - 2\nabla\tilde{\omega}\cdot(\nabla(\partial_{l,k}\tilde{z})\,A\tilde{M}) - \partial_{l,k}\tilde{z}\,\operatorname{div}(\nabla\tilde{\omega}\,A)\tilde{M} - \partial_{l,k}\tilde{z}\,\nabla M\cdot(\nabla\tilde{\omega}\,A).
\end{aligned}
$$

Analogously to the study of the first-order tangential derivatives we need all seven terms on the right-hand side of the above equation to belong to $a\tilde{M}\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\varepsilon) = \tilde{M}\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$ now. As, at this stage, we know that $\partial_k\tilde{z} \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$, the second, the third, the fifth and the sixth

term above pose no difficulties. The fourth term and the seventh term can be dealt with just as their counterparts in (3.63). When it comes to the first term, it is enough to show that $\partial_l g_{(k)} \in \mathrm{L}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$. Now,

$$\partial_l g_{(k)} = \partial_{l,k}\tilde{g} + \nabla\nabla\tilde{z}\colon \partial_l(\partial_k A\,/a) + \nabla\tilde{z}\cdot\partial_l(\mathrm{div}(\partial_k A)/a) + \partial_l\nabla\nabla\tilde{z}\colon(\partial_k A\,/a) + \partial_l\nabla\tilde{z}\cdot\mathrm{div}(\partial_k A)/a.$$

The first three terms above are clearly in $\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$—the first because of our hypotheses on g; so is the fifth, for the second derivatives of $\tilde{z}$ have the desired integrability. Finally, $\partial_l\nabla\nabla\tilde{z} \in \mathrm{L}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$ because $\partial_l\tilde{z} \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$, as shown above. Proceeding with the argument one finds, after localization, that $\partial_{k,l}\tilde{z} \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$. We mention in passing that by closely following the arguments above the linear operators $g \in \mathrm{H}^2_M(D) \mapsto \partial_l\tilde{z} \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\varepsilon)$ and $g \in \mathrm{H}^2_M(D) \mapsto \partial_{k,l}\tilde{z} \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ can be seen to be continuous; i.e., bounded.

Part 3: In this part we show the additional regularity of some derivatives of $\tilde{z}$ that involve the radial direction.

Expanding and rearranging the distributional equation (3.62), taking into account the sole dependence of $\tilde{M}$ on the last component of its argument and the properties of A given by (3.60) we get

$$-\frac{1}{\tilde{M}a}\partial_d(\partial_d\tilde{z}\,a\tilde{M}) = \tilde{g} - \tilde{z} + \frac{1}{a}\sum_{k=1}^{d-1}\sum_{j=1}^{d-1}(\partial_{j,k}\tilde{z}\,A_{j,k} + \partial_j\tilde{z}\,\partial_k A_{j,k}) =: f \tag{3.65}$$

in $\tilde{U}_\delta$. From the previous part of the proof and our assumptions on g we have that $f \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$. Multiplying the above expression by $\tilde{M}a$ and integrating with respect to the d-th variable we obtain

$$(\partial_d\tilde{z}\,a\tilde{M})[\boldsymbol{p}', p_d] - \lim_{s\to 0_+}(\partial_d\tilde{z}\,a\tilde{M})[\boldsymbol{p}', s] = \int_0^{p_d}(fa\tilde{M})[\boldsymbol{p}', s]\,\mathrm{d}s \tag{3.66}$$

for almost every $\boldsymbol{p}'$ in P'_δ. We note in passing that in this part of the proof we reserve square brackets for arguments of functions. Our first task is to show that the limit on the left-hand side of (3.66) vanishes. To this end, we first observe that, for $p_d, s \in (0, \delta\gamma)$,

$$p_d^{\alpha/2}\partial_d\tilde{z}[\boldsymbol{p}', p_d] = s^{\alpha/2}\partial_d\tilde{z}[\boldsymbol{p}', s] + \int_s^{p_d}\frac{\partial}{\partial\sigma}\left(\sigma^{\alpha/2}\partial_d\tilde{z}[\boldsymbol{p}', \sigma]\right)\mathrm{d}\sigma,$$

whence

$$p_d^{\alpha/2}\left|\partial_d\tilde{z}[\boldsymbol{p}', p_d]\right| \le s^{\alpha/2}\left|\partial_d\tilde{z}[\boldsymbol{p}', s]\right| + \left|\int_s^{p_d}\frac{\alpha}{2}\sigma^{\alpha/2-1}\partial_d\tilde{z}[\boldsymbol{p}', \sigma]\,\mathrm{d}\sigma\right| + \left|\int_s^{p_d}\sigma^{\alpha/2}\partial_{d,d}\tilde{z}[\boldsymbol{p}', \sigma]\,\mathrm{d}\sigma\right|.$$

Furthermore,

$$\begin{aligned}
&p_d^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', p_d]\right|^2 \\
&\le 3s^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 + \frac{3\alpha^2}{4}\left|\int_s^{p_d} \sigma^{\alpha/2-1}\partial_d \tilde{z}[\boldsymbol{p}', \sigma]\,\mathrm{d}\sigma\right|^2 + 3\left|\int_s^{p_d} \sigma^{\alpha/2}\partial_{d,d} \tilde{z}[\boldsymbol{p}', \sigma]\,\mathrm{d}\sigma\right|^2 \\
&\le 3s^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 + \frac{3\alpha^2}{4}\left|p_d - s\right|\int_s^{p_d} \sigma^{\alpha-2}\left|\partial_d \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma + 3\left|p_d - s\right|\int_s^{p_d} \sigma^{\alpha}\left|\partial_{d,d} \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma \\
&\le 3s^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 + \frac{3\alpha^2}{4}\delta\gamma\int_0^{\delta\gamma} \sigma^{\alpha-2}\left|\partial_d \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma + 3\delta\gamma\int_0^{\delta\gamma} \sigma^{\alpha}\left|\partial_{d,d} \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma.
\end{aligned}$$

Integrating this chain of inequalities with respect to s from 0 to $\delta\gamma$ and applying the Hardy inequality (3.56) stated in Lemma 3.25 we obtain

$$\begin{aligned}
&\delta\gamma\, p_d^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', p_d]\right|^2 \\
&\le 3\int_0^{\delta\gamma} s^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 \mathrm{d}s + \frac{3\alpha^2}{4}(\delta\gamma)^2\int_0^{\delta\gamma} \sigma^{\alpha-2}\left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 \mathrm{d}\sigma \\
&\qquad + 3(\delta\gamma)^2\int_0^{\delta\gamma} \sigma^{\alpha}\left|\partial_{d,d} \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma \\
&\le 3\int_0^{\delta\gamma} s^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 \mathrm{d}s + \frac{3\alpha^2 C_{H,\alpha}}{4}(\delta\gamma)^2\int_0^{\delta\gamma} \sigma^{\alpha}\left|\partial_{d,d} \tilde{z}[\boldsymbol{p}', s]\right|^2 \mathrm{d}\sigma \\
&\qquad + 3(\delta\gamma)^2\left(\frac{\alpha^2}{4} + C_{H,\alpha}\right)\int_0^{\delta\gamma} \sigma^{\alpha}\left|\partial_{d,d} \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma \\
&\le 3\int_0^{\delta\gamma} s^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', s]\right|^2 \mathrm{d}s + 3(\delta\gamma)^2\left(\frac{\alpha^2}{4} + \frac{C_{H,\alpha}}{4} + C_{H,\alpha}\right)\int_0^{\delta\gamma} \sigma^{\alpha}\left|\partial_{d,d} \tilde{z}[\boldsymbol{p}', \sigma]\right|^2 \mathrm{d}\sigma.
\end{aligned}$$

Then, dividing by $\delta\gamma$, integrating with respect to $\boldsymbol{p}'$ in P'_δ, using the bilateral boundedness of h and a by positive constants, and consolidating the constants, we get the trace-inequality-like bound

$$\int_{P'_\delta} p_d^\alpha \left|\partial_d \tilde{z}[\boldsymbol{p}', p_d]\right|^2 \mathrm{d}\boldsymbol{p}' \le C_1 \left\|\partial_d \tilde{z}\right\|^2_{\mathrm{H}^1_{\tilde{M}}(\tilde{U}_\delta)}. \tag{3.67}$$

Thus,

$$\int_{P'_\delta} \left|(\partial_d \tilde{z}\, a\tilde{M})[\boldsymbol{p}', p_d]\right| \mathrm{d}\boldsymbol{p}' \le C_2\, h[p_d]^{1/2} p_d^{\alpha/2}\left(\int_{P'_\delta} (|\partial_d \tilde{z}|^2\, \tilde{M}a)[\boldsymbol{p}]\,\mathrm{d}\boldsymbol{p}'\right)^{1/2} \to 0 \quad \text{as } p_d \to 0_+,$$

which implies the vanishing of the limit in (3.66).

Let us define $w\colon \tilde{U}_\delta \to \mathbb{R}$ by

$$w[\boldsymbol{p}] := \frac{\partial_d(\tilde{M}a)[\boldsymbol{p}]}{(\tilde{M}a)[\boldsymbol{p}]}\partial_d \tilde{z}[\boldsymbol{p}] = \frac{((ha)[p_d]p_d^\alpha)'}{(ha)[p_d]^2 p_d^{2\alpha}}\int_0^{p_d} (fa\tilde{M})[\boldsymbol{p}', s]\,\mathrm{d}s, \tag{3.68}$$

where we have taken the liberty of treating a as an univariate function, which it is in the algebraic sense. The equality is valid for almost every $\boldsymbol{p} \in \tilde{U}_\delta$. Note that w is a member

of $\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)$ because $\nabla M \cdot \nabla z/M \in \mathrm{L}^2_M(U_\delta)$; this, in turn, is a consequence of Lemma 3.24. We intend to show that $w \in \mathrm{H}^2_{\tilde{M}}(U_\delta)$. Let $(f_n \colon n \in \mathbb{N})$ be a sequence of $\mathrm{C}^\infty(\overline{\tilde{U}_\delta})$ functions converging to f in $\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ (its existence having been discussed in Part 1) and let

$$\begin{aligned} w_n[\boldsymbol{p}] &:= \frac{((ha)[p_d]p_d^\alpha)'}{(ha)[p_d]^2 p_d^{2\alpha}} \int_0^{p_d} (f_n a\tilde{M})[\boldsymbol{p}', s]\,\mathrm{d}s \\ &= \int_0^{p_d} \frac{(ha)'[p_d]p_d + \alpha(ha)[p_d]}{(ha)[p_d]^2} \left(\frac{s}{p_d}\right)^\alpha \frac{(haf_n)[\boldsymbol{p}', s]}{p_d}\,\mathrm{d}s \qquad (3.69) \\ &= \mathrm{aux}[p_d] \int_0^1 \xi^\alpha (haf_n)[\boldsymbol{p}', p_d\xi]\,\mathrm{d}\xi. \end{aligned}$$

Here we have written $\mathrm{aux}[p_d]$ in place of the first fraction in the second integral and denoted the function $\boldsymbol{p} \in \tilde{U}_\zeta \mapsto h(p_d) \in \mathbb{R}$ by h as well. The second equality comes via the change of variable $\xi = s/p_d$. As the function h and the determinant a have uniform C^3 and C^∞ regularity in U_δ, the function $\mathrm{aux} \in \mathrm{C}^2(\overline{\tilde{U}_\delta})$ and w_n is twice continuously differentiable in the d-th direction.

Differentiating the last integral representation of w_n with respect to its d-th variable twice and then reversing the change of variable we obtain

$$\begin{aligned} \partial_{d,d} w_n[\boldsymbol{p}] &= \sum_{k=0}^2 \binom{2}{k} \frac{\partial^{2-k}\mathrm{aux}}{\partial p_d^{2-k}}[p_d] \int_0^1 \partial_d^k (haf_n)[\boldsymbol{p}', p_d\xi]\,\xi^{\alpha+k}\,\mathrm{d}\xi \\ &= \sum_{k=0}^2 \binom{2}{k} \frac{\partial^{2-k}\mathrm{aux}}{\partial p_d^{2-k}}[p_d] \int_0^{p_d} \partial_d^k (haf_n)[\boldsymbol{p}', s] \left(\frac{s}{p_d}\right)^{\alpha+k} \frac{1}{p_d}\,\mathrm{d}s, \end{aligned}$$

whence, as $s/p_d \in (0,1)$ if $s \in (0, p_d)$,

$$\begin{aligned} p_d^{\alpha/2}\,|\partial_{d,d} w_n[\boldsymbol{p}]| &\le \frac{1}{p_d} \int_0^{p_d} \left(\sum_{k=0}^2 \binom{2}{k} \left| \frac{\partial^{2-k}\mathrm{aux}}{\partial p_d^{2-k}}[p_d]\,\partial_d^k (haf_n)[\boldsymbol{p}', s] \right| \left(\frac{s}{p_d}\right)^{\alpha/2+k} \right) s^{\alpha/2}\,\mathrm{d}s \\ &\le \frac{1}{p_d} \int_0^{p_d} \left(\sum_{k=0}^2 \binom{2}{k} \left| \frac{\partial^{2-k}\mathrm{aux}}{\partial p_d^{2-k}}[p_d]\,\partial_d^k (haf_n)[\boldsymbol{p}', s] \right| \right) s^{\alpha/2}\,\mathrm{d}s \\ &\le \frac{C_3}{p_d} \int_0^{p_d} \left(|(haf_n)|^2 + |\partial_d(haf_n)|^2 + |\partial_{d,d}(haf_n)|^2 \right)^{1/2} [\boldsymbol{p}', s] s^{\alpha/2}\,\mathrm{d}s \end{aligned}$$

for some $C_3 > 0$ independent of $\boldsymbol{p} = (\boldsymbol{p}', p_d)$. We square the resulting inequality, integrate it with respect to p_d from 0 to $\delta\gamma$, use the Hardy inequality (3.55) in Lemma 3.25 and note yet again the bilateral boundedness of h and a by positive constants to obtain

$$\begin{aligned} &\int_0^{\delta\gamma} p_d^\alpha (ha)[p_d]\,|\partial_{d,d} w_n[\boldsymbol{p}]|^2\,\mathrm{d}p_d \\ &\qquad \le C_4 \int_0^{\delta\gamma} \Big(|(haf_n)|^2 + |\partial_d(haf_n)|^2 + |\partial_{d,d}(haf_n)|^2 \Big)[\boldsymbol{p}]\, s^\alpha (ha)[p_d]\,\mathrm{d}p_d, \end{aligned}$$

where C_4 is still independent of $\boldsymbol{p}' \in P'_\delta$. Integrating this with respect to $\boldsymbol{p}' \in P'_\delta$, using the regularity of h and a and taking into account that $(\tilde{M}a)[\boldsymbol{p}] = (ha)[p_d]p_d^\alpha$ for all $\boldsymbol{p} \in \tilde{U}_\delta$ one gets

$$\|\partial_{d,d} w_n\|_{\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq C_5 \, \|f_n\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \, .$$

This argument can be carried over to all derivatives of order less than or equal to two of w_n (including zeroth order derivatives of w_n, meaning w_n itself). The result is

$$\|w_n\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq C_6 \, \|f_n\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \, .$$

As $\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ is a Hilbert space, there exists a subsequence $\big(w_{\phi(n)} \colon n \geq 1\big)$ with a weak limit $w^* \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$. By the continuity of the injection of $\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ into $\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)$, w^* is also the weak limit of the $w_{\phi(n)}$ in $\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)$.

Now, given any $\chi \in \mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)$,

$$\begin{aligned}
&\left\| \frac{\mathrm{aux}[\cdot_d]}{\cdot_d} \int_0^{\cdot_d} \left(\frac{s}{\cdot_d}\right)^\alpha (ha\chi)[\cdot', s] \, \mathrm{d}s \right\|^2_{\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)} \\
&\qquad = \int_{\tilde{U}_\delta} \left(\frac{\mathrm{aux}[p_d]}{p_d} \int_0^{p_d} \left(\frac{s}{p_d}\right)^\alpha (ha\chi)[\boldsymbol{p}', s] \, \mathrm{d}s \right)^2 \tilde{M}[\boldsymbol{p}] \, \mathrm{d}\boldsymbol{p} \\
&\qquad = \int_{\tilde{U}_\delta} \frac{\mathrm{aux}[p_d]^2}{p_d^2} \left(\int_0^{p_d} \left(\frac{s}{p_d}\right)^{\alpha/2} s^{\alpha/2} (ha\chi)[\boldsymbol{p}', s] \, \mathrm{d}s \right)^2 h[p_d] \, \mathrm{d}\boldsymbol{p} \\
&\qquad \leq C_7 \int_{P'_\delta} \int_0^{\delta\gamma} \frac{1}{p_d^2} \left(\int_0^{p_d} s^{\alpha/2} (ha\chi)[\boldsymbol{p}', s] \, \mathrm{d}s \right)^2 \mathrm{d}p_d \, \mathrm{d}\boldsymbol{p}' \\
&\qquad \leq C_8 \int_{P'_\delta} \int_0^{\delta\gamma} s^\alpha \left| (ha\chi)[\boldsymbol{p}', s] \right|^2 \mathrm{d}s \, \mathrm{d}\boldsymbol{p}' \\
&\qquad \leq C_9 \, \|\chi\|^2_{\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)} \, .
\end{aligned}$$

Hence, the operation that defines w (resp. w_n) in terms of f (resp. f_n) in (3.68) (resp. (3.69)) is a bounded map from $\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)$ to itself. Therefore, $\lim_{n\to\infty} f_n = f$ in $\mathrm{L}^2_{\tilde{M}}(\tilde{U}_\delta)$ implies $\lim_{n\to\infty} w_n = w$ in the same space. Thus, w and the weak limit w^* have to be the same measurable function and so $w \in \mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$. We get the bound

$$\|w\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq \liminf_{n\to\infty} \|w_{\phi(n)}\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq C_6 \|f_{\phi(n)}\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} .$$

As (with no loss of generality) we can assume that the f_n are scaled so that their $\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ norm is identically equal to the same norm of f, it follows that

$$\|w\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq C_6 \, \|f\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \, .$$

From (3.65) and (3.68),

$$-\partial_{d,d}\tilde{z} = f + \frac{\partial_d(\tilde{M}a)}{\tilde{M}a}\partial_d\tilde{z} = f + w,$$

whence $\|\partial_{d,d}\tilde{z}\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq (1 + C_6)\,\|f\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)} \leq C_{10}\,\|\tilde{g}\|_{\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)}$. We know from the previous part that all second derivatives of $\tilde{z}$ that do not involve the d-th direction have $\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ norms bounded by the $\mathrm{H}^2_{\tilde{M}}(\tilde{U}_\delta)$ norm of $\tilde{g}$. This and the corresponding result for $\partial_{d,d}\tilde{z}$ is enough to be able to bound all derivatives of $\tilde{z}$ of order less than or equal to four, and thus deduce that

$$\|\tilde{z}\|_{\mathrm{H}^4_{\tilde{M}}(\tilde{U}_\delta)} \leq C_{11}\,\|g\|_{\mathrm{H}^2_M(D)}$$

or, equivalently in the light of (3.58), that

$$\|z\|_{\mathrm{H}^4_M(U_\delta)} \leq C_{12}\,\|g\|_{\mathrm{H}^2_M(U_\delta)}\,. \tag{3.70}$$

Part 4: By modifying the transformation S one can get a localized bound of the form (3.70) for any origin-centered rotation of U_δ. It follows that (3.70) remains valid (with some other constant C_{12}) if we replace U_δ by the annulus/spherical shell $\{\boldsymbol{p} \in D\colon\ |\boldsymbol{p}| > \sqrt{b} - \delta\gamma\}$.

Let D_0 be the ball $B(0, \sqrt{b} - \delta\gamma/2) \Subset D$. As $\mathrm{C}^\infty_0(D_0) \subset \mathrm{C}^\infty_0(D)$, we have that

$$\langle z, \varphi\rangle_{\mathrm{H}^1_M(D_0)} = \langle g, \varphi\rangle_{\mathrm{L}^2_M(D_0)} \qquad \forall\, \varphi \in \mathrm{C}^\infty_0(D_0).$$

The existence of a positive infimum of M in D_0 implies that z is the weak solution to a regular (i.e., uniformly) elliptic problem in D_0 with $\mathrm{H}^2(D_0)$ right-hand side. It follows, via the $\mathrm{C}^{2,1}(D_0)$ regularity of M (see, e.g., [**GT01**, Theorem 8.10]), that for some $C_{13} > 0$,

$$\|z\|_{\mathrm{H}^4_M(D_0')} \leq C_{13}\,\|g\|_{\mathrm{H}^2_M(D_0)} \leq C_{13}\,\|g\|_{\mathrm{H}^2_M(D)}\,,$$

with $D_0' := B(0, \sqrt{n} - 3\delta\gamma/4) \Subset D_0$.

Combining this last estimate with the result in the aforementioned annulus/spherical shell (which in union with D_0' covers D), we obtain the desired global bound. □

The following statement is an almost trivial corollary of Lemma 3.26, yet it is a true iterate of Lemma 3.24 in the sense that the hypothesis on the right-hand side function is the thesis on the solution in Lemma 3.24. This makes it suitable for the arguments that will be used in the proof of Lemma 3.28.

Lemma 3.27. *Let $i \in [N]$, suppose that $g \in \mathrm{H}^2_{M_i}(D_i)$ and that $M_i^{-1}\,\mathrm{div}(M_i\nabla g) \in \mathrm{L}^2_{M_i}(D_i)$. Then, the solution $z \in \mathrm{H}^1_{M_i}(D_i)$ of*

$$\langle z, \varphi\rangle_{\mathrm{H}^1_{M_i}(D_i)} = \langle g, \varphi\rangle_{\mathrm{L}^2_{M_i}(D_i)} \qquad \forall\, \varphi \in \mathrm{H}^1_{M_i}(D_i)$$

obeys the regularity estimate

$$\|z\|_{\mathrm{H}^4_{M_i}(D_i)} + \left\|\frac{1}{M_i}\operatorname{div}(M_i\nabla z)\right\|_{\mathrm{H}^2_{M_i}(D_i)} + \left\|\frac{1}{M_i}\operatorname{div}\left(M_i\nabla\left[\frac{1}{M_i}\operatorname{div}(M_i\nabla z)\right]\right)\right\|_{\mathrm{L}^2_{M_i}(D_i)} \leq C_i\left(\|g\|_{\mathrm{H}^2_{M_i}(D_i)} + \left\|\frac{1}{M_i}\operatorname{div}(M_i\nabla g)\right\|_{\mathrm{L}^2_{M_i}(D_i)}\right)$$

for some $C_i > 0$.

Proof. This follows directly from Lemma 3.26 on noting that $M_i^{-1}\operatorname{div}(M_i\nabla z) = g - z$ in the distributional sense first, and then in the sense of measurable functions. □

Lemma 3.28. *Let $i \in [N]$. The following statements of equivalence hold:*

$$\tau \in \mathrm{H}^2_{M_i}(D_i) \quad \textit{and} \quad \frac{1}{M_i}\operatorname{div}(M_i\nabla\tau) \in \mathrm{L}^2_{M_i}(D_i) \iff \tau \in \mathrm{L}^2_{M_i}(D_i) \quad \textit{and} \quad \sum_{n=1}^{\infty}(\lambda_n^{(i)})^2\langle\tau, e_n^{(i)}\rangle^2_{\mathrm{L}^2_{M_i}(D_i)} < \infty; \tag{3.71}$$

and

$$\tau \in \mathrm{H}^4_{M_i}(D_i),\ \frac{1}{M_i}\operatorname{div}(M_i\nabla\tau) \in \mathrm{H}^2_{M_i}(D_i) \quad \textit{and} \quad \frac{1}{M_i}\operatorname{div}\left(M_i\nabla\left[\frac{1}{M_i}\operatorname{div}(M_i\nabla\tau)\right]\right) \in \mathrm{L}^2_{M_i}(D_i) \iff \tau \in \mathrm{L}^2_{M_i}(D_i) \quad \textit{and} \quad \sum_{n=1}^{\infty}(\lambda_n^{(i)})^4\langle\tau, e_n^{(i)}\rangle^2_{\mathrm{L}^2_{M_i}(D_i)} < \infty. \tag{3.72}$$

Proof. We will omit the spring index when proving (3.71) and (3.72). We start by denoting by L the operator that associates each $\varphi \in \mathrm{W}^{2,1}_{\mathrm{loc}}(D)$ with the distribution $M^{-1}\operatorname{div}(M\nabla\varphi)$ (this is a well-defined distribution because of the regularity of φ and M; cf. Lemma A.2 in Section A.1). We also write $\hat{L} := -L + I$ where I is the operator that associates each distribution with itself.

Let us define the Hilbert spaces (and associated norms)

$$\tilde{\mathrm{H}}^2_M(D) := \left\{\varphi \in \mathrm{H}^2_M(D)\colon L(\varphi) \in \mathrm{L}^2_M(D)\right\}, \qquad \|\varphi\|^2_{\tilde{\mathrm{H}}^2_M(D)} := \|\varphi\|^2_{\mathrm{H}^2_M(D)} + \|L(\varphi)\|^2_{\mathrm{L}^2_M(D)} \tag{3.73}$$

and

$$\tilde{\mathrm{H}}^4_M(D) := \left\{\varphi \in \mathrm{H}^4_M(D)\colon L(\varphi) \in \tilde{\mathrm{H}}^2_M(D)\right\}, \tag{3.74a}$$

$$\|\varphi\|^2_{\tilde{\mathrm{H}}^4_M(D)} := \|\varphi\|^2_{\mathrm{H}^4_M(D)} + \|L(\varphi)\|^2_{\tilde{\mathrm{H}}^2_M(D)} = \|\varphi\|^2_{\mathrm{H}^4_M(D)} + \|L(\varphi)\|^2_{\mathrm{H}^2_M(D)} + \left\|L^2(\varphi)\right\|_{\mathrm{L}^2_M(D)}. \tag{3.74b}$$

Because of the definition of $\tilde{\mathrm{H}}^2_M(D)$, $\hat{L}\colon \mathrm{H}^2_M(D) \to \mathrm{L}^2_M(D)$ is a bounded linear operator. As for every $\varphi \in \mathrm{L}^2_M(D)$ the solution $z \in \mathrm{H}^1_M(D)$ to

$$\langle z, \psi\rangle_{\mathrm{H}^1_M(D)} = \langle f, \psi\rangle_{\mathrm{L}^2_M(D)} \qquad \forall\, \psi \in \mathrm{H}^1_M(D)$$

exists and, thanks to Lemma 3.24, is bounded in $\tilde{\mathrm{H}}^2_M(D)$, $\hat{L}^{-1}\colon \mathrm{L}^2_M(D) \to \tilde{\mathrm{H}}^2_M(D)$ is well-defined and bounded. Similarly, by the definition of $\tilde{\mathrm{H}}^4_M(D)$ and Lemma 3.27, $\hat{L}^2\colon \tilde{\mathrm{H}}^4_M(D) \to \mathrm{L}^2_M(D)$ is a bounded linear operator with a bounded inverse.

Let $\tau \in \tilde{\mathrm{H}}^2_M(D)$; i.e., τ complies with the left-hand side of (3.71). It then follows that $f_\tau := -L(\tau) + \tau \in \mathrm{L}^2_M(D)$ and Parseval's identity thus renders

$$\infty > \|f_\tau\|^2_{\mathrm{L}^2_M(D)} = \sum_{n\geq 1} \langle f_\tau, e_n\rangle^2_{\mathrm{L}^2_M(D)} = \sum_{n\geq 1} \langle \tau, e_n\rangle^2_{\mathrm{H}^1_M(D)} = \sum_{n\geq 1} \lambda_n^2 \langle \tau, e_n\rangle^2_{\mathrm{L}^2_M(D)},$$

where $\langle f_\tau, e_n\rangle_{\mathrm{L}^2_M(D)} = \langle \tau, e_n\rangle_{\mathrm{H}^1_M(D)}$ follows by the density of $\mathrm{C}^\infty_0(D)$ in $\mathrm{H}^1_M(D)$.

To prove the converse, note that the eigenfunctions e_n of (3.33) are solutions of $e_n = \hat{L}^{-1}(\lambda_n e_n)$, whence $\|e_n\|_{\tilde{\mathrm{H}}^2_M(D)} \leq C\,\|\lambda_n e_n\|_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} = C\lambda_n$. Consequently, the partial sums

$$\tau_k := \sum_{n=1}^{k} \langle \tau, e_n\rangle_{\mathrm{L}^2_M(D)} e_n$$

are members of $\tilde{\mathrm{H}}^2_M(D)$. Then, if $k \leq l$, the $\mathrm{L}^2_M(D)$-orthonormality of the e_n yields that

$$\left\|\hat{L}(\tau_l) - \hat{L}(\tau_k)\right\|^2_{\mathrm{L}^2_M(D)} = \left\|\sum_{n=k+1}^{l} \langle \tau, e_n\rangle_{\mathrm{L}^2_M(D)} \hat{L}(e_n)\right\|^2_{\mathrm{L}^2_M(D)} = \sum_{n=k+1}^{l} \lambda_n^2 \langle \tau, e_n\rangle^2_{\mathrm{L}^2_M(D)}.$$

As the sum $\sum_{n\geq 1} \lambda_n^2 \langle \tau, e_n\rangle^2_{\mathrm{L}^2_M(D)}$ is assumed to converge, the sequence $\Big(\hat{L}(\tau_k)\colon k \geq 1\Big)$ is a Cauchy sequence in $\mathrm{L}^2_M(D)$ and hence it converges to some $f^* \in \mathrm{L}^2_M(D)$. The continuity of $\hat{L}^{-1}$ implies that the sequence $(\tau_k\colon k \geq 1)$ converges in $\tilde{\mathrm{H}}^2_M(D)$ to $\hat{L}^{-1}(f^*)$. The same sequence converges in $\mathrm{L}^2_M(D)$ to τ. The continuity of the injection of $\mathrm{H}^2_M(D)$ into $\mathrm{L}^2_M(D)$ then implies that $\tau = \hat{L}^{-1}(f^*) \in \tilde{\mathrm{H}}^2_M(D)$. This completes the proof of (3.71).

Let us suppose now that τ in $\tilde{\mathrm{H}}^4_M(D)$; i.e., τ complies with the left-hand side of (3.72). It follows that $f_\tau := -L(\tau) + \tau \in \tilde{\mathrm{H}}^2_M(D)$ and $g_\tau := -L(f_\tau) + f_\tau \in \mathrm{L}^2_M(D)$. Parseval's identity gives

$$\begin{aligned}\infty > \|g_\tau\|^2_{\mathrm{L}^2_M(D)} = \sum_{n\geq 1} \langle g_\tau, e_n\rangle^2_{\mathrm{L}^2_M(D)} &= \sum_{n\geq 1} \langle f_\tau, e_n\rangle^2_{\mathrm{H}^1_M(D)} \\ &= \sum_{n\geq 1} \lambda_n^2 \langle f_\tau, e_n\rangle^2_{\mathrm{L}^2_M(D)} = \sum_{n\geq 1} \lambda_n^4 \langle \tau, e_n\rangle^2_{\mathrm{L}^2_M(D)},\end{aligned}$$

where the second equality follows, similarly as above, by the density of $\mathrm{C}^\infty_0(D)$ in $\mathrm{H}^1_M(D)$ thanks to the boosted regularity of f_τ. The latter also allows the use of (3.33) to obtain the third equality.

To prove the converse we first note that each e_n is a solution of $e_n = \hat{L}^{-2}(\lambda_n^2 e_n)$, whence $\|e_n\|_{\tilde{\mathrm{H}}^4_M(D)} \leq C\,\|\lambda_n e_n\|_{\mathrm{L}^2_M(D)} = C\lambda_n$. Thus, the partial sums τ_k are members of $\tilde{\mathrm{H}}^4_M(D)$ now. If $k \leq l$,

$$\left\|\hat{L}^2(\tau_l - \tau_k) - \hat{L}^2(\tau_k)\right\|^2_{\mathrm{L}^2_M(D)} = \left\|\sum_{n=k+1}^{l} \langle \tau, e_n\rangle_{\mathrm{L}^2_M(D)} \hat{L}^2(e_n)\right\|^2_{\mathrm{L}^2_M(D)} = \sum_{n=k+1}^{l} \lambda_n^4 \langle \tau, e_n\rangle^2_{\mathrm{L}^2_M(D)}. \tag{3.75}$$

The finiteness of the sum $\sum_{n\geq 1} \lambda_n^4 \langle \tau, e_n\rangle_{\mathrm{L}^2_M(D)}$ thus makes of $\left(\hat{L}^2(\tau_k)\colon k \geq 1\right)$ a Cauchy sequence, which by virtue of the completeness of $\mathrm{L}^2_M(D)$ converges to some $g^* \in \mathrm{L}^2_M(D)$. The continuity of $\hat{L}^{-2}$ implies that the τ_k converge to $\hat{L}^{-2}g^*$ in $\tilde{\mathrm{H}}^4_M(D)$. As the partial sums converge in $\mathrm{L}^2_M(D)$ to τ, $\tau = \hat{L}^{-2}g^* \in \tilde{\mathrm{H}}^4_M(D)$. We have thus proved (3.72). □

We intend to exploit the previous single-domain results in order to say something about the multi-domain case. To this end, we define, for $i \in [N]$, the distributional operators (already utilized in the proof of Lemma 3.28) $L_i\colon \{\varphi \in \mathrm{L}^1_{\mathrm{loc}}(\mathsf{D})\colon \nabla_{\boldsymbol{q}_i}\varphi \in [\mathrm{W}^{1,1}_{\mathrm{loc}}(\mathsf{D})]^d\} \to \mathcal{D}'(\mathsf{D})$ by

$$L_i(\varphi) := \frac{1}{M_i}\,\mathrm{div}_{\boldsymbol{q}_i}(M_i \nabla_{\boldsymbol{q}_i}\varphi) \qquad \forall\, \varphi \in \mathcal{D}'(\mathsf{D}). \tag{3.76}$$

We also define $\hat{L}_i := -L_i + I$, where I is the identity operator mapping $\mathcal{D}'(\mathsf{D})$ onto itself. An easily verifiable and important property of these operators is that, as long as their composition is well-defined, they commute with respect to their spring index. Hence, we can naturally use multi-indices in $\mathbb{N}^N$ to denote the repeated application of distinct L_i or $\hat{L}_i$:

$$L_\beta := L_1^{\beta_1} \circ \cdots \circ L_N^{\beta_N}, \qquad \hat{L}_\beta := \hat{L}_1^{\beta_1} \circ \cdots \circ \hat{L}_N^{\beta_N}, \tag{3.77}$$

where any zero among the β_i is assumed to give rise to the identity operator. Now, for standard (weak) derivatives in $\mathcal{D}'(\mathsf{D})$, the multi-indices belong to $\mathbb{N}^{Nd}$ and come naturally grouped in N length-d sub-multi-indices (one for each factor of the Cartesian product $\mathsf{D} = D_1 \times \cdots \times D_N$). With this in mind we introduce the function $|\cdot|_{\infty,1} : \mathbb{N}^{Nd} \to \mathbb{N}$ defined by

$$|\boldsymbol{\alpha}|_{\infty,1} = |(\alpha_1, \ldots, \alpha_N)|_{\infty,1} := \max_{i\in[N]} |\alpha_i|_1 = \max_{i\in[N]} |\alpha_i|\,;$$

that is, the maximum among the orders of the component single-domain multi-indices.

With this compact notation, we now define the Hilbert spaces (with corresponding norms)

$$\tilde{\mathrm{H}}^{2,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) := \left\{\varphi \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D})\colon \partial_{\boldsymbol{\alpha}} \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}),\ |\boldsymbol{\alpha}|_{\infty,1} \leq 2;\ L_\beta(\varphi) \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}),\ |\beta|_\infty = 1\right\}, \tag{3.78a}$$

$$\|\varphi\|^2_{\tilde{\mathrm{H}}^{2,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})} := \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^{Nd}\\ |\boldsymbol{\alpha}|_{\infty,1}\leq 2}} \|\partial_{\boldsymbol{\alpha}}\varphi\|^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} + \sum_{\substack{\beta\in\mathbb{N}^{d}\\ |\beta|_\infty = 1}} \|L_\beta(\varphi)\|^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} \tag{3.78b}$$

and

$$\tilde{\mathrm{H}}^{4,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) := \Big\{\varphi \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D})\colon \partial_{\boldsymbol{\alpha}} \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}),\ |\boldsymbol{\alpha}|_{\infty,1} \le 4;\ L_\beta(\varphi) \in \mathrm{H}^2_{\mathsf{M}}(\mathsf{D}),\ |\beta|_\infty = 1;\ L_\beta(\varphi) \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}),\ |\beta|_\infty = 2\Big\}, \tag{3.79a}$$

$$\|\varphi\|^2_{\tilde{\mathrm{H}}^{4,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})} := \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^{Nd}\\ |\boldsymbol{\alpha}|_{\infty,1}\le 4}} \|\partial_{\boldsymbol{\alpha}}\varphi\|^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} + \sum_{\substack{\beta\in\mathbb{N}^{d}\\ |\beta|_\infty = 1}} \|L_\beta(\varphi)\|^2_{\mathrm{H}^2_{\mathsf{M}}(\mathsf{D})} + \sum_{\substack{\beta\in\mathbb{N}^{d}\\ |\beta|_\infty = 2}} \|L_\beta(\varphi)\|^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})}. \tag{3.79b}$$

The following lemma holds.

Lemma 3.29. *For $m \in \{2,4\}$, $\tilde{\mathrm{H}}^{m,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) \subset \mathrm{H}^{\mathrm{T}^{(m)}}_{\mathsf{M}}(\mathsf{D})$.*

Proof. We recall that, by Lemma 3.16, $\big((\lambda_{\boldsymbol{n}}, e_{\boldsymbol{n}})\colon \boldsymbol{n}\in\mathbb{N}^N\big)$ as defined in (3.37) is a complete set of solutions of the M-weighted eigenvalue problem (3.34) and that the latter have a tensor-product structure. Also, by the definitions in (3.39), (3.40) and (3.48), $\mathrm{H}^{\mathrm{T}^{(m)}}_{\mathsf{M}}(\mathsf{D})$ is the space of $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ functions whose squared Fourier coefficients, weighted with the coefficients defined by

$$\tau^{(m)}_{\boldsymbol{n}} = \left(\sum_{i=1}^N \lambda^{(i)}_{n_i}\right)^m + \prod_{i=1}^N \left(\lambda^{(i)}_{n_i}\right)^m \qquad \forall\, \boldsymbol{n}\in\mathbb{N}^N,$$

have finite sum.

If $\varphi \in \tilde{\mathrm{H}}^{m,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$, one can apply to it each operator $\hat{L}_i$ a total of $m/2$ times *cumulatively* and land in $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$; i.e.,

$$\hat{L}_{(m/2,\dots,m/2)}(\varphi) \in \mathrm{L}^2_{\mathsf{M}}(\mathsf{D}).$$

By Parseval's identity,

$$\begin{aligned} \infty > \left\|\hat{L}_{(m/2,\dots,m/2)}\varphi\right\|^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} &= \sum_{\boldsymbol{n}\in\mathbb{N}^N} \left\langle \hat{L}_{(m/2,\dots,m/2)}(\varphi), e_{\boldsymbol{n}}\right\rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} \\ &= \sum_{\boldsymbol{n}\in\mathbb{N}^N} \prod_{i=1}^N (\lambda^{(i)}_{n_i})^m \langle\varphi, e_{\boldsymbol{n}}\rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})}, \end{aligned} \tag{3.80}$$

where the second equality is justified by the density of $\mathrm{C}^\infty_0(\mathsf{D})$ functions in $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$, the regularity of φ and the Cartesian product form of the domain D and the tensor-product form of the Maxwellian weight function M.

Now, the distributional identity

$$\mathsf{L}(\varphi) := \frac{1}{\mathsf{M}}\operatorname{div}(\mathsf{M}\nabla\varphi) = \sum_{i=1}^d \frac{1}{M_i}\operatorname{div}_{\boldsymbol{q}_i}\left(M_i\nabla_{\boldsymbol{q}_i}\varphi\right)$$

(here, on the right-hand side, the M_i actually represent the function that to each $\boldsymbol{q}\in\mathsf{D}$ associates $M_i(\boldsymbol{q}_i)$) implies that the conditions on $\tilde{\mathrm{H}}^m_{\mathsf{M}}(\mathsf{D})$ are sufficient in order to deduce that the $(m/2)$-fold application of the operator $\widehat{\mathsf{L}} := -\mathsf{L} + N(\cdot)$ to φ belongs to $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$. An

argument analogous to the one that led to (3.80) leads to

$$\infty > \left\| \widehat{\mathsf{L}}^{m/2} \varphi \right\|^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} = \sum_{\boldsymbol{n} \in \mathbb{N}} \left(\sum_{i=1}^{N} \lambda_n^{(i)} \right)^m \langle \varphi, e_{\boldsymbol{n}} \rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})}, \tag{3.81}$$

and so the lemma is proved. □

We recall that Theorem 3.18 gives a condition on the parameter of the moderately abstract space $\mathrm{H}^{\mathrm{T}^{(m)}}_{\mathsf{M}}(\mathsf{D})$ under which it becomes a subspace of the abstract space $\mathcal{B}_1$, which in turn is connected by (3.32) to the space $\mathcal{A}_1$ of fast convergence of the greedy algorithms (cf. Theorem 3.13 and Theorem 3.14). Then, from Lemma 3.29 it is apparent that the apparently less abstract (although, admittedly, that might lie in the eye of the beholder) space $\tilde{\mathrm{H}}^{m,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$ will be a subspace of $\mathcal{B}_1$ for a suitable choice of the parameter. This statement will be made more precise in the following theorem.

Theorem 3.30. *Let* $\mathrm{H}^{\mathrm{T}^{(d+1)}}_{\mathsf{M}}(\mathsf{D})$ *be defined according to* (3.40)*, where* $d \in \{2, 3\}$*, as elsewhere, is the common dimensionality of the Cartesian factors that make up* D*. Then, the following inclusions hold:*

$$\tilde{\mathrm{H}}^{4,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) \subset \mathrm{H}^{\mathrm{T}^{(d+1)}}_{\mathsf{M}}(\mathsf{D}) \subset \mathcal{B}_1.$$

Proof. Lemma 3.29 gives that $\tilde{\mathrm{H}}^{4,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) \subset \mathrm{H}^{\mathrm{T}^{(4)}}_{\mathsf{M}}(\mathsf{D})$. In the case of $d = 3$ the result follows immediately from Theorem 3.18. When $d = 2$, by virtue of the relation (cf. (3.40)),

$$\sum_{\boldsymbol{n} \in \mathbb{N}^N} \tau^{(4)}_{\boldsymbol{n}} \langle \varphi, e_{\boldsymbol{n}} \rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} < \infty \implies \sum_{\boldsymbol{n} \in \mathbb{N}^N} \tau^{(3)}_{\boldsymbol{n}} \langle \varphi, e_{\boldsymbol{n}} \rangle^2_{\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})} < \infty,$$

and one has $\mathrm{H}^{\mathrm{T}^{(4)}}_{\mathsf{M}}(\mathsf{D}) \subset \mathrm{H}^{\mathrm{T}^{(3)}}_{\mathsf{M}}(\mathsf{D})$; appealing again to Theorem 3.18 we then deduce the result. □

Remark 3.31. If the hypotheses we have been making throughout this work (i.e., Hypotheses A, B, C, D and E) are met, nothing in our arguments essentially restricts the results to the physically relevant cases $d = 2$ and $d = 3$. In particular, in the case of $d = 1$, the combination of Theorem 3.18 with Lemma 3.29 yields that

$$\tilde{\mathrm{H}}^{2,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) \subset \mathrm{H}^{\mathrm{T}^{(d+1)}}_{\mathsf{M}}(\mathsf{D}) \subset \mathcal{B}_1.$$

Sobolev spaces of dominating mixed smoothness akin to $\tilde{\mathrm{H}}^{2,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$ can also be shown to be subspaces of the regularity class $\mathcal{B}_1$ in the case of the classical Poisson problem studied in [**LBLM09**]: Find $\psi \in \mathrm{H}^1_0(\mathsf{D})$ (with the standard meaning of the Sobolev space $\mathrm{H}^1_0(\mathsf{D})$; i.e., the set of all elements of $\mathrm{H}^1(\mathsf{D})$ that have zero trace on $\partial\mathsf{D}$—not a zero-weighted Sobolev space!) such that

$$\langle \psi, \varphi \rangle_{\mathrm{H}^1_0(\mathsf{D})} = \langle f, \varphi \rangle_{\mathrm{L}^2(\mathsf{D})} \quad \forall\, \varphi \in \mathrm{H}^1_0(\mathsf{D}),$$

where $\mathsf{D} = D_1 \otimes \cdots \otimes D_N$ and each D_i, $i \in [N]$, is an open interval. The corresponding greedy algorithms seek approximations that are linear combinations of $\bigotimes_{i \in [N]} \mathrm{H}^1_0(D_i)$ functions. The argument of Theorem 3.18 above holds in this case without any change, and so, given that

the n-th eigenvalue of the corresponding analogue to the partial-domain eigenvalue problem (3.33) is proportional to n^2, we have that

$$\left\{ \varphi \in \mathrm{L}^2(\mathsf{D}) \colon \sum_{\boldsymbol{n} \in \mathbb{N}^N} \left[\left(\sum_{i=1}^{N} n_i^2 \right)^2 + \prod_{i=1}^{N} \left(n_i^2 \right)^2 \right] \langle \varphi, e_{\boldsymbol{n}} \rangle_{\mathrm{L}^2(\mathsf{D})}^2 < \infty \right\} \subset \mathcal{B}_1.$$

In this non-degenerate setting it is possible to identify the space on the left-hand side of the above expression as

$$\mathrm{H}^{2,\mathrm{mix}}(\mathsf{D}) \cap \mathrm{H}_0^1(\mathsf{D}) := \left\{ \varphi \in \mathrm{H}_0^1 \colon \partial_\alpha \varphi \in \mathrm{L}^2(\mathsf{D}),\ |\alpha|_\infty = \max_{1 \leq i \leq N} \alpha_i \leq 2 \right\}.$$

This characterization should be contrasted with the condition for membership in $\mathcal{A}_1$ (which is identical to $\mathcal{B}_1$ in this unweighted setting) derived in [**LBLM09**, Remark 4], which demands, instead, that the true solution belongs to $\mathrm{H}^m(\mathsf{D}) \cap \mathrm{H}_0^1(\mathsf{D})$, with $m > 1 + N/2$. In fact the characterization given in [**LBLM09**, Remark 4] can be generalized to the requirement that the exact solution belongs to $\mathrm{H}^m(\mathsf{D}) \cap \mathrm{H}_0^1(\mathsf{D})$, with $m > 1 + Nd/2$, when the factor domains are no longer one-dimensional but d-dimensional; and, due to Theorem 3.20, such a characterization in terms of standard Sobolev spaces (rather than spaces of dominating mixed smoothness) also has a counterpart in our degenerate setting.

So, which choice of space is 'best'? Membership in standard Sobolev spaces is not directly comparable with membership in spaces of dominating mixed smoothness. An attractive feature of spaces of dominating mixed smoothness is that their regularity index is independent of N and such spaces are therefore more naturally suited to (high-dimensional) tensor-product settings such as ours.

Even if Lemma 3.29 held with an equality of spaces, there would still be some slack in the case of $d = 2$. We have gone about obtaining elliptic regularity results by two degrees of differentiation at a time. Consequently, we have not defined anything we could label $\tilde{\mathrm{H}}^3_{M_i}(D_i)$ or $\tilde{\mathrm{H}}^{3,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$ while being consistent with the definitions we have given for the single-spring spaces $\tilde{\mathrm{H}}^2_{M_i}(D_i)$ in (3.73) and $\tilde{\mathrm{H}}^4_{M_i}(D_i)$ in (3.74), and with the multi-spring spaces $\tilde{\mathrm{H}}^{2,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$ in (3.78) and $\tilde{\mathrm{H}}^{4,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$ in (3.79). Given the presence of the second-order operators of the form $M_i^{-1} \operatorname{div}(M_i \nabla \cdot)$ and $\mathsf{M}^{-1} \operatorname{div}(\mathsf{M} \nabla \cdot)$ in the definition of these even-indexed spaces, it is not immediately clear what a suitable explicit definition of the analogous odd-indexed spaces might be.

We shall address this question through the use of function space interpolation. We start with the fact that, given two nets of positive weights $\Sigma^{(i)} = \left(\sigma^{(i)}_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N \right)$, $i \in \{1, 2\}$, one can show that the (real) $(1/2, 2)$-interpolation space between them obeys

$$\left[\mathrm{H}^{\Sigma^{(1)}}_{\mathsf{M}}(\mathsf{D}), \mathrm{H}^{\Sigma^{(2)}}_{\mathsf{M}}(\mathsf{D}) \right]_{1/2,2} = \mathrm{H}^{\tilde{\Sigma}}_{\mathsf{M}}(\mathsf{D}),$$

where

$$\tilde{\Sigma} = \left(\tilde{\sigma}_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right) \qquad \text{and} \qquad \tilde{\sigma}_{\boldsymbol{n}} := \left(\sigma_{\boldsymbol{n}}^{(1)}\right)^{1/2} \left(\sigma_{\boldsymbol{n}}^{(2)}\right)^{1/2} \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N,$$

with equivalence of norms (the proof is a simple modification of the argument given in [**Tar07**, Chapter 23]). As it can be shown that there exist positive constants c_1 and c_2 such that

$$c_1 \tau_{\boldsymbol{n}}^{(3)} \leq \sqrt{\tau_{\boldsymbol{n}}^{(2)}} \sqrt{\tau_{\boldsymbol{n}}^{(4)}} \leq c_2 \tau_{\boldsymbol{n}}^{(3)} \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N,$$

it follows that

$$\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(3)}}(\mathsf{D}) = \left[\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(4)}}(\mathsf{D}), \mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(2)}}(\mathsf{D})\right]_{1/2,2}, \tag{3.82}$$

with equivalence of norms. Given that the inclusions in Lemma 3.29 are actually continuous embeddings, it follows that

$$\left[\tilde{\mathrm{H}}_{\mathsf{M}}^{4,\mathrm{mix}}(\mathsf{D}), \tilde{\mathrm{H}}_{\mathsf{M}}^{2,\mathrm{mix}}(\mathsf{D})\right]_{1/2,2} \subset \mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(3)}}(\mathsf{D}), \tag{3.83}$$

with continuous embedding. Since in the case of $d = 2$ we have that $\mathrm{H}_{\mathsf{M}}^{\mathrm{T}^{(3)}}(\mathsf{D}) \subset \mathcal{B}_1$, defining $\tilde{\mathrm{H}}_{\mathsf{M}}^{3,\mathrm{mix}}(\mathsf{D})$ as the interpolation space appearing on the left-hand side of the previous expression is an appealing idea.

We have established the convergence properties of the Separated Representation strategy as expressed in the abstract algorithms Algorithm I and Algorithm II and have characterized subspaces of the abstract space of fast convergence $\mathcal{B}_1$ in simpler terms. We also proved that having the solution live in a space of functions with controlled high mixed derivatives implies membership in $\mathcal{B}_1$ and, hence, guaranteed rates of convergence for the greedy algorithms.

In the next chapter we study variants of the greedy algorithms whose iterates are sought for in the tensor product of finite dimensional subspaces.

CHAPTER 4

Discrete Separated Representation

In this section we study the effect on the greedy algorithms of restricting the search space of each new tensor-product term to the tensor product of finite dimensional subspaces of the $\mathrm{H}(D_i; M_i)$. We will show that the algorithms will converge to the best approximation of the solution to the Fokker–Planck equation in the space spanned by the tensor products of finite dimensional subspaces used in the computation. Then, we will discuss how to guarantee rates of convergence for the discrete greedy algorithms and discuss the properties of two particular classes of finite dimensional subspaces of $\mathrm{H}(D_i; M_i)$. We finish with a short discussion on the practical implementation of discrete greedy algorithm and illustrate its behavior with a simple numerical experiment.

4.1. Discrete spaces

4.1.1. Finite dimensional subspaces. Let, for $i \in [N]$ and $l \in \mathbb{N}$, $\hat{\mathrm{H}}_l^{(i)}$ be a subspace of $\mathrm{H}(D_i; M_i)$ of dimension l, with the nesting property

$$l \leq m \implies \hat{\mathrm{H}}_l^{(i)} \subset \hat{\mathrm{H}}_m^{(i)}. \tag{4.1}$$

Then, for all $\boldsymbol{l} = (l_1, \ldots, l_N) \in \mathbb{N}^N$,

$$\hat{\mathrm{H}}_{\boldsymbol{l}} := \operatorname{span}\left(\bigotimes_{i \in [N]} \hat{\mathrm{H}}_{l_i}^{(i)} \right), \tag{4.2}$$

which has dimension $\prod_{i \in [N]} l_i$, is a subspace of $\mathrm{H}(\mathsf{D}; \mathsf{M})$ (cf. Lemma 2.23) and the nesting property

$$\boldsymbol{l} \leq \boldsymbol{m} \implies \hat{\mathrm{H}}_{\boldsymbol{l}} \subset \hat{\mathrm{H}}_{\boldsymbol{m}} \tag{4.3}$$

is inherited from (4.1).

Each sequence $\left(\hat{\mathrm{H}}_l^{(i)} : l \in \mathbb{N}\right)$ may be thought of as a sequence of finite element or spectral element subspaces of $\mathrm{H}(D_i; M_i)$.

Remark 4.1. Families of finite-dimensional subspaces of each $\mathrm{H}(D_i; M_i)$ used in practice may come naturally indexed by something other than their dimension—say, a meshsize. Even if they can be reindexed according to their dimension, the resulting families may not have a representative of every dimension in $\mathbb{N}$. The results of this section still hold if instead of $\mathbb{N}$ some infinite subset of it indexes each family of subsets of $\mathrm{H}(D_i; M_i)$, for $i \in [N]$. If we

don't assume this from the beginning, it is because it would introduce a further burden to an already heavy notation system.

4.1.2. Greedy algorithms on discrete spaces. We introduce the following discrete versions of the Pure Greedy Algorithm (Algorithm I) and the Orthogonal Greedy Algorithm (Algorithm II) for the approximation of the variational problem (2.5) on $\hat{\mathrm{H}}_{\boldsymbol{l}}$.

Algorithm V *(Discrete Pure Greedy Algorithm).*

1. Set $\hat{f}_0 := f \in \mathrm{H}(\mathsf{D};\mathsf{M})'$.
2. For $n \geq 1$ do:
   2.1 Find $\hat{r}_n^{(i)}$, $i \in [N]$ such that

$$(\hat{r}_n^{(1)},\dots,\hat{r}_n^{(N)}) \in \underset{(s_n^{(1)},\dots,s_n^{(N)})\in \bigtimes_{i=1}^N \hat{\mathrm{H}}_{l_i}^{(i)}}{\arg\min} \frac{1}{2} a\left(\bigotimes_{i=1}^N s^{(i)}, \bigotimes_{i=1}^N s^{(i)}\right) - \hat{f}_{n-1}\left(\bigotimes_{i=1}^N s^{(i)}\right). \tag{4.4}$$

   2.2 Define

$$\hat{f}_n := \hat{f}_{n-1} - \mathcal{A}\left(\bigotimes_{i=1}^N \hat{r}_n^{(i)}\right).$$

   2.3 If $\|\hat{f}_n|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}\|_{\hat{\mathrm{H}}_{\boldsymbol{l}}'} \geq \mathtt{TOL}$, then proceed to iteration $n+1$; else, stop.

Algorithm VI *(Discrete Orthogonal Greedy Algorithm).*

1. Set $\hat{f}_0 := f \in \mathrm{H}(\mathsf{D};\mathsf{M})'$.
2. For $n \geq 1$ do:
   2.1 Find $\hat{r}_n^{(i)}$, $i \in [N]$ such that

$$(\hat{r}_n^{(1)},\dots,\hat{r}_n^{(N)}) \in \underset{(s_n^{(1)},\dots,s_n^{(N)})\in \bigtimes_{i=1}^N \hat{\mathrm{H}}_{l_i}^{(i)}}{\arg\min} \frac{1}{2} a\left(\bigotimes_{i=1}^N s^{(i)}, \bigotimes_{i=1}^N s^{(i)}\right) - \hat{f}_{n-1}\left(\bigotimes_{i=1}^N s^{(i)}\right). \tag{4.5}$$

   2.2 Solve the following Galerkin problem on the span of $\left(\bigotimes_{i=1}^N \hat{r}_k^{(i)} : k \in [n]\right)$:

$$\hat{\alpha}^{(n)} := \underset{\beta\in\mathbb{R}^n}{\arg\min}\left\{\frac{1}{2}\, a\left(\sum_{k=1}^n \beta_k \bigotimes_{i=1}^N \hat{r}_k^{(i)}, \sum_{k=1}^n \beta_k \bigotimes_{i=1}^N \hat{r}_k^{(i)}\right) - f\left(\sum_{k=1}^n \beta_k \bigotimes_{i=1}^N \hat{r}_k^{(i)}\right)\right\}. \tag{4.6}$$

   2.3 Define

$$\hat{f}_n := f - \mathcal{A}\left(\sum_{k=1}^n \hat{\alpha}_k^{(n)} \bigotimes_{i=1}^N \hat{r}_k^{(i)}\right).$$

   2.4 If $\|\hat{f}_n|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}\|_{\hat{\mathrm{H}}_{\boldsymbol{l}}'} \geq \mathtt{TOL}$, then proceed to iteration $n+1$; else, stop.

The approximations to the true solution ψ given by the above algorithms at iteration n are

$$\sum_{k=1}^n \bigotimes_{i=1}^N \hat{r}_k^{(i)} \qquad \text{and} \qquad \sum_{k=1}^n \hat{\alpha}_k^{(n)} \bigotimes_{i=1}^N \hat{r}_n^{(i)}.$$

for Algorithm V and Algorithm VI, respectively. We denote by $\hat{\delta}_n \in \mathrm{H}(\mathsf{D};\mathsf{M})$ the error committed by the n-th iteration of either of the above algorithms (cf. (3.6)):

$$\hat{\delta}_n = \begin{cases} \hat{\delta}_{n-1} - \bigotimes_{i=1}^N \hat{r}_n^{(i)} & \text{for the Discrete Pure Greedy Algorithm,} \\ \psi - \sum_{k=1}^n \hat{\alpha}_k^{(n)} \bigotimes_{i=1}^N \hat{r}_k^{(i)} & \text{for the Discrete Orthogonal Greedy Algorithm,} \end{cases} \tag{4.7}$$

where $\hat{\delta}_0 := \psi$, the true solution to (2.5).

Remark 4.2. In [**CEL11**, Section 4] it is shown that greedy algorithms akin to Algorithm V, whose iterates lie on the span of the tensor product of finite dimensional subspaces, converge at an exponential rate under certain conditions that imply that the space in which the error is measured is finite dimensional as well. As we measure the error in $\mathrm{H}(\mathsf{D};\mathsf{M})$, which is infinite-dimensional, this result is not directly applicable to our case.

4.1.3. Properties of the Discrete Greedy Algorithms. We prove basic results on the minimization problems (4.4) of Algorithm V and (4.5) of Algorithm VI. All of the results in this subsection have direct counterparts in Chapter 3, which allows for abbreviating their proofs in varying degrees.

Lemma 4.3. *Suppose that $f \in \mathrm{H}(\mathsf{D};\mathsf{M})$ is such that $f|_{\hat{\mathrm{H}}_{\boldsymbol{l}}} \neq 0$. Then, there exists an ensemble $(r^{(1)},\dots,r^{(N)})$ in $\times_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$ such that*

$$J_f\left(\bigotimes_{i=1}^N r^{(i)}\right) < 0.$$

Proof. Let us assume that the thesis is false; i.e., $J_f\left(\bigotimes_{i\in[N]} r^{(i)}\right) \geq 0$ for all $(r^{(1)},\dots,r^{(N)}) \in \times_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$. Given any such ensemble and any $\varepsilon \in \mathbb{R}$, one can replace $r^{(1)}$ with $\varepsilon r^{(1)}$ and obtain

$$\frac{1}{2}\varepsilon^2 a\left(\bigotimes_{i=1}^N r^{(i)}, \bigotimes_{i=1}^N r^{(i)}\right) \geq \varepsilon f\left(\bigotimes_{i=1}^N r^{(i)}\right).$$

Then, using the same argument based in one-sided limits that was used in Lemma 3.1, we can conclude that $f\left(\bigotimes_{i\in[N]} r^{(i)}\right) = 0$ for all $(r^{(1)},\dots,r^{(N)}) \in \times_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$; hence, $f|_{\hat{\mathrm{H}}_{\boldsymbol{l}}} = 0$, which contradicts the hypotheses of the lemma. Therefore, the thesis must hold. □

The following theorem ensures that the minimization problems (4.4) of Algorithm V and (4.5) of Algorithm VI have solutions.

Theorem 4.4. *The minimization problems* (4.4) *and* (4.5) *have solutions.*

Proof. Similarly to (3.2), the fact that $\hat{\delta}_{n-1}$ satisfies

$$a(\hat{\delta}_{n-1},\varphi) = \hat{f}_{n-1}(\varphi) \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M})$$

allows for writing $J_{\hat{f}_{n-1}} = \frac{1}{2}a(\cdot - \hat{\delta}_{n-1}, \cdot - \hat{\delta}_{n-1}) - \frac{1}{2}a(\hat{\delta}_{n-1}, \hat{\delta}_{n-1})$. Therefore, the image of $\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$ by $J_{\hat{f}_{n-1}}$ is bounded from below. Hence,

$$\mathfrak{m} := \inf_{s\in\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}} J_{\hat{f}_{n-1}}(s) > -\infty$$

and there exists a sequence $(r_k \colon k \geq 1) = \left(\bigotimes_{i\in[N]} r_k^{(i)} \colon k \geq 1\right)$ of $\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$ functions such that

$$\lim_{k\to\infty} J_{\hat{f}_{n-1}}(r_k) = \mathfrak{m}.$$

Now, as for any $\varphi \in \mathrm{H}(\mathsf{D};\mathsf{M})$

$$J_{\hat{f}_{n-1}}(\varphi) \geq \frac{1}{4}\min\left(\frac{\lambda_{\min}}{4\mathrm{Wi}}, c\right)\|\varphi\|^2_{\mathrm{H}(\mathsf{D};\mathsf{M})} - a(\hat{\delta}_{n-1}, \hat{\delta}_{n-1}),$$

the sequence $(r_k \colon k \geq 1)$ is bounded in $\mathrm{H}(\mathsf{D};\mathsf{M})$. As $\operatorname{span}\left(\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}\right)$ is finite-dimensional, there exists a subsequence $\left(r_{\phi(k)} \colon k \geq 1\right)$ which converges in the norm $\mathrm{H}(\mathsf{D};\mathsf{M})$ to some $r \in \operatorname{span}\left(\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}\right)$. As $J_{\hat{f}_{n-1}}$ is continuous in $\mathrm{H}(\mathsf{D};\mathsf{M})$,

$$\mathfrak{m} = \lim_{k\to\infty} J_{\hat{f}_{n-1}}\left(r_{\phi(k)}\right) = J_{\hat{f}_{n-1}}(r).$$

Now, just as in the proof of Theorem 3.2, for $i \in [N]$, we can suppose the sequences of factors $\left(r_{\phi(k)}^{(i)} \colon k \geq 1\right)$ bounded in $\mathrm{H}(D_i; M_i)$, which, on account of the finite dimensionality of the $\hat{\mathrm{H}}_{l_i}^{(i)}$, implies the existence of subsequences $\left(r_{\phi'(k)}^{(i)} \colon k \geq 1\right)$ which converge to some $r^{(i)} \in \hat{\mathrm{H}}_{l_i}^{(i)}$ in the respective norms $\mathrm{H}(D_i; M_i)$. Then, as can be deducted from Lemma 2.22,

$$r = \bigotimes_{i=1}^{N} r^{(i)};$$

that is, r is a member of $\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$ because it has the necessary tensor-product structure. Hence, the infimum $\mathfrak{m}$ is attained in $\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$ and the theorem is proved. □

The proofs of Lemma 4.5 and Lemma 4.6 that follow are completely analogous to their counterparts in the continuous case, Lemma 3.4 and Lemma 3.6, respectively; so, we omit them.

Lemma 4.5. *Local minimizers* $(\hat{r}_n^{(1)}, \ldots, \hat{r}_n^{(N)})$ *of the minimization problems* (4.4) *and* (4.5) *satisfy the following Euler–Lagrange equation system: For all* $(s_n^{(1)}, \ldots, s_n^{(N)}) \in \times_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$,

$$a\left(\bigotimes_{i=1}^{N} \hat{r}_n^{(i)}, \sum_{j=1}^{N} \bigotimes_{\substack{i=1\\ i\neq j}}^{N} \hat{r}_n^{(i)} \otimes_j \hat{s}^{(j)}\right) = \hat{f}_{n-1}\left(\sum_{j=1}^{N} \bigotimes_{\substack{i=1\\ i\neq j}}^{N} \hat{r}_n^{(i)} \otimes_j \hat{s}^{(j)}\right). \tag{4.8}$$

From this, it follows that, for both the Discrete Pure Greedy Algorithm (Algorithm V) and the Discrete Orthogonal Greedy Algorithm (Algorithm VI):

$$a\left(\bigotimes_{i=1}^{N}\hat{r}_n^{(i)},\bigotimes_{i=1}^{N}\hat{r}_n^{(i)}\right)=a\left(\hat{\delta}_{n-1},\bigotimes_{i=1}^{N}\hat{r}_n^{(i)}\right). \tag{4.9}$$

Lemma 4.6. *Suppose* $\hat{f}_{n-1}\Big|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}\neq 0$ *and let* $(\hat{r}_n^{(1)},\ldots,\hat{r}_n^{(N)})$ *be a global minimizer for the minimization problem* (4.4) *of the Algorithm V or for the minimization problem* (4.5) *of the Algorithm VI. Then,*

$$\left\|\bigotimes_{i=1}^{N}\hat{r}_n^{(i)}\right\|_a=\frac{a\left(\hat{\delta}_{n-1},\bigotimes_{i=1}^{N}\hat{r}_n^{(i)}\right)}{\left\|\bigotimes_{i=1}^{N}\hat{r}_n^{(i)}\right\|_a}=\sup_{s\in\bigotimes_{i\in[N]}\hat{\mathrm{H}}_{l_i}^{(i)}\setminus\{0\}}\frac{a\left(\hat{\delta}_{n-1},s\right)}{\|s\|_a}. \tag{4.10}$$

If $\hat{f}_{n-1}\Big|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}=0$, *the equality between the left-most and the right-most expressions in* (4.10) *is still valid.*

The approximation to ψ computed in the Discrete Pure Greedy Algorithm (Algorithm V), namely

$$\sum_{k=1}^{n}\bigotimes_{i=1}^{N}\hat{r}_k^{(i)},$$

and its counterpart resulting from the Discrete Orthogonal Greedy Algorithm (Algorithm VI), namely

$$\sum_{k=1}^{n}\hat{\alpha}_k^{(n)}\bigotimes_{i=1}^{N}\hat{r}_k^{(i)},$$

are members of $\hat{\mathrm{H}}_{\boldsymbol{l}}$ for every iteration n of the respective algorithms. Therefore, one cannot expect these approximations to converge to ψ except in the unlikely case in which ψ happens to be a member of the finite dimensional space $\hat{\mathrm{H}}_{\boldsymbol{l}}$. However, these algorithms do converge to the a-projection of ψ on $\hat{\mathrm{H}}_{\boldsymbol{l}}$. This is proved in the two theorems that follow.

Theorem 4.7. *The Discrete Pure Greedy Algorithm (Algorithm V) converges to the* a*-projection of* ψ *(the true solution to* (2.5)*) onto* $\hat{\mathrm{H}}_{\boldsymbol{l}}$*; that is, for any* $\texttt{TOL}>0$ *there exists some iteration number* n *such that*

$$\|\hat{f}_n|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}\|_{\hat{\mathrm{H}}_{\boldsymbol{l}}'}=\left\|P^a_{\hat{\mathrm{H}}_{\boldsymbol{l}}}(\psi)-\sum_{k=1}^{n}\bigotimes_{i=1}^{N}\hat{r}_k^{(i)}\right\|_a<\texttt{TOL},$$

where $P^a_{\hat{\mathrm{H}}_{\boldsymbol{l}}}(\psi)$ *denotes the* a*-projection of* ψ *onto* $\hat{\mathrm{H}}_{\boldsymbol{l}}$.

Proof. Except at the very end, this proof follows closely the proof of Theorem 3.7. Indeed, just like in the proof of Theorem 3.7, but with (4.7) and (4.9) in Lemma 4.5 in place of, respectively, (3.6) and (3.17), it can be proved that the sequence $(\|\hat{\delta}_n\|_a\colon n\geq 0)$ is monotonic

nonincreasing (and hence, converging in $\mathbb{R}$) and that

$$\sum_{n=1}^{\infty} a\left(\bigotimes_{i=1}^{N} \hat{r}_n^{(i)}, \bigotimes_{i=1}^{N} \hat{r}_n^{(i)}\right) < \infty, \tag{4.11}$$

where $\left((\hat{r}_n^{(1)}, \ldots, \hat{r}_n^{(N)})\colon n \geq 1\right)$ is the sequence in $\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$ returned by the Discrete Pure Greedy Algorithm. We now define the function $\phi\colon \mathbb{N} \to \mathbb{N}$ recursively by $\phi(1) := 1$ and

$$\phi(k) := \min\left\{n \in \mathbb{N}\colon n > \phi(k-1) \quad \text{and} \quad \left\|\bigotimes_{i=1}^{N} \hat{r}_n^{(i)}\right\|_a \leq \left\|\bigotimes_{i=1}^{N} \hat{r}_{\phi(k-1)}^{(i)}\right\|_a\right\}, \quad k \geq 2.$$

From (4.11) the function ϕ is well-defined and strictly monotonic increasing. Hence, it is suitable for defining subsequences. Then, if $1 \leq m \leq n$,

$$\left\|\hat{\delta}_{\phi(n)-1} - \hat{\delta}_{\phi(m)-1}\right\|_a^2 \leq \left\|\hat{\delta}_{\phi(m)-1}\right\|_a^2 - \left\|\hat{\delta}_{\phi(n)-1}\right\|_a^2 + 2\sum_{k=\phi(m)}^{\phi(n)-1} \left\|\bigotimes_{i=1}^{N} \hat{r}_k^{(i)}\right\|_a^2, \tag{4.12}$$

which can be proved just like its counterpart (3.20) in the proof of Theorem 3.7, but using Lemma 4.6 instead of Lemma 3.6. It follows from the convergence of $\left(\|\hat{\delta}_n\|_a\colon n \geq 0\right)$ and (4.11) that $\left(\hat{\delta}_{\phi(n)-1}\colon n \geq 1\right)$ is a Cauchy sequence in $\mathrm{H}(\mathsf{D};\mathsf{M})$ and thus converges to some $\hat{\delta}_\infty \in \mathrm{H}(\mathsf{D};\mathsf{M})$. From the fact that each $(\hat{r}_n^{(1)}, \ldots, \hat{r}_n^{(N)})$ is a global minimizer of its respective problem (4.4), it follows, via the identity (4.9) in Lemma 4.5, that for all $(s^{(1)}, \ldots, s^{(N)}) \in \bigtimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$,

$$\begin{aligned}\frac{1}{2}a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - a\left(\hat{\delta}_{\phi(n)-1}, \bigotimes_{i=1}^{N} s^{(i)}\right) &\geq \frac{1}{2}a\left(\bigotimes_{i=1}^{N} \hat{r}_{\phi(n)}^{(i)}, \bigotimes_{i=1}^{N} \hat{r}_{\phi(n)}^{(i)}\right) - \hat{f}_{\phi(n)-1}\left(\bigotimes_{i=1}^{N} \hat{r}_{\phi(n)}^{(i)}\right) \\ &= -\frac{1}{2}a\left(\bigotimes_{i=1}^{N} \hat{r}_{\phi(n)}^{(i)}, \bigotimes_{i=1}^{N} \hat{r}_{\phi(n)}^{(i)}\right).\end{aligned}$$

Taking the limit as n tends to infinity results in

$$\frac{1}{2}a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - a\left(\hat{\delta}_\infty, \bigotimes_{i=1}^{N} s^{(i)}\right) \geq 0.$$

As this is valid for all $(s^{(1)}, \ldots, s^{(N)}) \in \bigtimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$, Lemma 4.3 gives that $\hat{\delta}_\infty$ is a-orthogonal to $\hat{\mathrm{H}}_{\boldsymbol{l}}$. As for all $n \geq 1$,

$$\hat{\delta}_{\phi(n)-1} = \psi - \sum_{k=1}^{\phi(n)-1} \bigotimes_{i=1}^{N} \hat{r}_k^{(i)},$$

the subsequence of partial sums $\left(\sum_{k=1}^{\phi(n)-1} \bigotimes_{i\in[N]} \hat{r}_k^{(i)}\colon n \geq 1\right)$ converges to $\psi - \hat{\delta}_\infty$ in $\mathrm{H}(\mathsf{D};\mathsf{M})$. As this subsequence is contained in the finite-dimensional space $\hat{\mathrm{H}}_{\boldsymbol{l}}$, its limit $\psi - \hat{\delta}_\infty$ is a member of $\hat{\mathrm{H}}_{\boldsymbol{l}}$ as well. Then,

$$a(\psi - (\psi - \hat{\delta}_\infty), \xi) = 0 \qquad \forall\, \xi \in \hat{\mathrm{H}}_{\boldsymbol{l}}$$

whence we can conclude that $\psi - \hat{\delta}_\infty$ is the best approximation of ψ in $\hat{\mathrm{H}}_{\boldsymbol{l}}$ in the energy norm. Now, we know that the full sequence of norms $\left(\|\hat{\delta}_n\|_a \colon n \geq 0\right)$ converges in $\mathbb{R}$. As one of its subsequences converges to $\|\hat{\delta}_\infty\|_a$, the full sequence has the same limit. Then,

$$\lim_{n\to\infty} \left\| \psi - \sum_{k=1}^{n} \hat{r}_k^{(i)} \right\|_a^2 = \lim_{n\to\infty} \|\hat{\delta}_n\|_a^2 = \|\hat{\delta}_\infty\|_a^2. \tag{4.13}$$

So, using that $\hat{\delta}_\infty$ is orthogonal to members of $\hat{\mathrm{H}}_{\boldsymbol{l}}$, and, in particular, to $\psi - \hat{\delta}_\infty$, we have

$$\begin{aligned}
\left\| \sum_{k=1}^{n} \bigotimes_{i=1}^{N} \hat{r}_k - (\psi - \hat{\delta}_\infty) \right\|_a^2 &= \left\| \sum_{k=1}^{n} \bigotimes_{i=1}^{N} \hat{r}_k - \psi \right\|_a^2 + 2a \left(\sum_{k=1}^{n} \bigotimes_{i=1}^{N} \hat{r}_k - \psi, \hat{\delta}_\infty \right) + \|\hat{\delta}_\infty\|_a^2 \\
&= \left\| \sum_{k=1}^{n} \bigotimes_{i=1}^{N} \hat{r}_k - \psi \right\|_a^2 + 2a \left((\psi - \hat{\delta}_\infty) - \psi, \hat{\delta}_\infty \right) + \|\hat{\delta}_\infty\|_a^2 \\
&= \left\| \sum_{k=1}^{n} \bigotimes_{i=1}^{N} \hat{r}_k - \psi \right\|_a^2 - \|\hat{\delta}_\infty\|_a^2.
\end{aligned}$$

Taking the limit as n tends to infinity at both ends of the above chain of equalities and using (4.13) we obtain that the full sequence of partial sums $\left(\sum_{k=1}^{n} \bigotimes_{i=1}^{N} \hat{r}_k \colon n \geq 1\right)$ tends to $\psi - \hat{\delta}_\infty$; i.e., the best approximation of ψ in $\hat{\mathrm{H}}_{\boldsymbol{l}}$ as measured in the a-norm. □

Theorem 4.8. *The Discrete Orthogonal Greedy Algorithm (Algorithm VI) converges to the a-projection of ψ (the true solution to (2.5)) onto $\hat{\mathrm{H}}_{\boldsymbol{l}}$; that is, for any* TOL > 0 *there exists some iteration number n such that*

$$\|\hat{f}_n|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}\|_{\hat{\mathrm{H}}'_{\boldsymbol{l}}} = \left\| P^a_{\hat{\mathrm{H}}_{\boldsymbol{l}}}(\psi) - \sum_{k=1}^{n} \hat{\alpha}_k^{(n)} \bigotimes_{i=1}^{N} \hat{r}_k^{(i)} \right\|_a < \texttt{TOL},$$

where $P^a_{\hat{\mathrm{H}}_{\boldsymbol{l}}}(\psi)$ denotes the a-projection of ψ onto $\hat{\mathrm{H}}_{\boldsymbol{l}}$.

Proof. Just like in the proof of Theorem 3.10, based on (4.7), the optimality of $\hat{\alpha}^{(n+1)}$ in (4.6), the optimality of $(\hat{r}_{n+1}^{(1)}, \ldots, \hat{r}_{n+1}^{(N)})$ in (3.4) and (4.9) in Lemma 4.5 we have

$$\|\hat{\delta}_{n+1}\|_a^2 \leq \|\hat{\delta}_n\|_a^2 - \left\| \bigotimes_{i=1}^{N} \hat{r}_{n+1}^{(i)} \right\|_a^2. \tag{4.14}$$

Suppose $\hat{\delta}_n$ is not a-orthogonal to $\hat{\mathrm{H}}_{\boldsymbol{l}}$. Then we know from Lemma 4.3 that $\bigotimes_{i\in[N]} \hat{r}_{n+1}^{(i)} \neq 0$; thus, as $\alpha^{(n)} \in \mathbb{R}^n$ and $\alpha^{(n+1)} \in \mathbb{R}^{n+1}$ are chosen in step 2.2 of Algorithm VI in such a way as to produce the best approximation to ψ in the span of $\left(\bigotimes_{i=1}^{N} \hat{r}_k^{(i)} \colon k \in [n]\right)$ and the span of $\left(\bigotimes_{i=1}^{N} \hat{r}_k^{(i)} \colon k \in [n+1]\right)$, respectively, (4.14) implies

$$\left\| \psi - \sum_{k=1}^{n+1} \alpha^{(n+1)} \bigotimes_{i=1}^{N} \hat{r}_k^{(i)} \right\|_a < \left\| \psi - \sum_{k=1}^{n} \alpha^{(n)} \bigotimes_{i=1}^{N} \hat{r}_k^{(i)} \right\|_a,$$

which is only possible if $\bigotimes_{i\in[N]} \hat{r}_{n+1}^{(i)}$ is linearly independent of the set of the previously computed $\bigotimes_{i\in[N]} \hat{r}_k^{(i)}$, $k \in [n]$.

Now, as $\hat{\mathrm{H}}_{\boldsymbol{l}}$ has finite dimension $\prod_{i\in[N]} l_i$, the algorithm is bound to produce, at an iteration $n+1$ with $n \leq \prod_{i\in[N]} l_i + 1$, a new basis member $\bigotimes_{i=1}^{N} \hat{r}_{n+1}^{(i)}$ that is not linearly independent from the previous ones. It then follows from the above discussion that the n-th error $\hat{\delta}_n$ is orthogonal to $\hat{\mathrm{H}}_{\boldsymbol{l}}$. From then on, the only global minimizers of (4.5) are those whose tensor product result in the zero function of $\mathrm{H}(\mathsf{D};\mathsf{M})$, whence the algorithm will have converged; i.e., $\hat{\delta}_m = \hat{\delta}_n$, for all $m \geq n$. As $\hat{\delta}_n$ is a-orthogonal to $\hat{\mathrm{H}}_{\boldsymbol{l}}$,

$$a\left(\psi - \sum_{k=1}^{n} \alpha_k^{(n)} \bigotimes_{i=1}^{N} \hat{r}_k^{(i)}, \xi\right) = 0 \qquad \forall\, \xi \in \hat{\mathrm{H}}_{\boldsymbol{l}}.$$

As the n-th approximate $\sum_{k=1}^{n} \alpha_k^{(n)} \bigotimes_{i=1}^{N} \hat{r}_k^{(i)}$ is a member of $\hat{\mathrm{H}}_{\boldsymbol{l}}$, we can conclude that the function the algorithm has converged to is the a-projection of ψ onto $\hat{\mathrm{H}}_{\boldsymbol{l}}$. □

We will now show that the Discrete Pure Greedy Algorithm (Algorithm V) and the Discrete Orthogonal Greedy Algorithm (Algorithm VI), like their continuous counterparts, are instances of the General Pure Greedy Algorithm (Algorithm III) and General Orthogonal Greedy Algorithm (Algorithm IV), respectively.

We start by defining, for $\boldsymbol{l} \in \mathbb{N}^N$,

$$\mathfrak{D}_{\boldsymbol{l}} := \left\{ g \in \mathrm{H}(\mathsf{D};\mathsf{M})' \colon g = a(s,\cdot),\ s \in \bigotimes_{i=1}^{N} \hat{\mathrm{H}}_{l_i}^{(i)},\ \|g\|_{\mathrm{H}(\mathsf{D};\mathsf{M})'} = 1 \right\}, \tag{4.15}$$

which we call a *truncated* dictionary and is a subset of the dictionary $\mathfrak{D}$ specified in (3.25). We can use this truncated dictionary instead of $\mathfrak{D}$ within Algorithm III and Algorithm IV while still seeking to approximate $f \in \mathfrak{H} = \mathrm{H}(\mathsf{D};\mathsf{M})'$. The step `2.1` of Algorithm III and the step `2.1` of Algorithm IV turn into finding some $g = g(R_{n-1}) \in \mathfrak{D}_{\boldsymbol{l}}$ such that

$$g(R_{n-1}) \in \operatorname*{arg\,max}_{\tilde{g}\in\mathfrak{D}_{\boldsymbol{l}}} \langle R_{n-1}, \tilde{g}\rangle_{\mathfrak{H}}, \tag{4.16}$$

where R_{n-1} is the residual given by the previous iteration (or the initialization, if $n = 1$). We have the following analogue to Proposition 3.12:

Proposition 4.9. *Let $\tilde{f} \in \mathrm{H}(\mathsf{D};\mathsf{M})' \setminus \{0\}$. Then,*

$$r \in \operatorname*{arg\,min}_{t\in\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}} \left[\frac{1}{2}a(t,t) - \tilde{f}(t)\right] \implies a\left(\frac{r}{\|r\|_a}, \cdot\right) \in \operatorname*{arg\,max}_{\tilde{g}\in\mathfrak{D}_{\boldsymbol{l}}} \langle \tilde{f}, \tilde{g}\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'}$$

and

$$a(s,\cdot) \in \operatorname*{arg\,max}_{\tilde{g}\in\mathfrak{D}_{\boldsymbol{l}}} \langle \tilde{f}, \tilde{g}\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'} \implies \langle \tilde{f}, a(s,\cdot)\rangle_{\mathrm{H}(\mathsf{D};\mathsf{M})'}\, s \in \operatorname*{arg\,min}_{t\in\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}} \left[\frac{1}{2}a(t,t) - \tilde{f}(t)\right].$$

Proof. The proof of this proposition is completely analogous to the proof of Proposition 3.12; thus, we omit it. □

Using Proposition 4.9, it follows that Algorithm V and Algorithm VI are instances of the Greedy Algorithms Algorithm III and Algorithm IV of the theory of nonlinear approximation. It is then relevant to consider an analogue to the abstract space $\mathcal{A}_1$ defined in (3.30):

$$\mathcal{A}_{1,\boldsymbol{l}} := \bigcup_{M>0} \overline{\mathcal{A}^o_{1,\boldsymbol{l}}(M)}, \tag{4.17a}$$

where the closure is with respect to the $\mathrm{H}(\mathsf{D};\mathsf{M})$ norm and

$$\mathcal{A}^o_{1,\boldsymbol{l}}(M) := \Bigg\{ \varphi \in \hat{\mathrm{H}}_{\boldsymbol{l}} \colon \varphi = \sum_{k\in\Lambda} c_k w_k,\ w_k \in \bigotimes_{i=1}^{N} \hat{\mathrm{H}}^{(i)}_{l_i},\ \|w_k\|_a = 1, \\ |\Lambda| < \infty, \quad \text{and} \quad \sum_{k\in\Lambda} |c_k| \le M \Bigg\}, \tag{4.17b}$$

equipped with the norm (again, the closure being with respect to the $\mathrm{H}(\mathsf{D};\mathsf{M})$ norm)

$$\|\varphi\|_{\mathcal{A}_{1,\boldsymbol{l}}} := \inf \left\{ M > 0 \colon \varphi \in \overline{\mathcal{A}^o_{1,\boldsymbol{l}}(M)} \right\}. \tag{4.17c}$$

Continuing the parallel with Subsection 3.2.4, the space $\mathcal{A}_{1,\boldsymbol{l}}$ lies at the core of rate-of-convergence theorems, which, like their counterparts Theorem 3.13 and Theorem 3.14, stem from the application of Theorems 3.6 and 3.7 of [**DT96**].

Theorem 4.10. *If the solution ψ of (2.5) is a member of $\mathcal{A}_{1,\boldsymbol{l}}$, the n-th error $\hat{\delta}_n$ of the Discrete Pure Greedy Algorithm (Algorithm V) satisfies*

$$\left\| \hat{\delta}_n \right\|_a \le \|\psi\|_{\mathcal{A}_{1,\boldsymbol{l}}}\, n^{-1/6}.$$

Theorem 4.11. *If the solution ψ of (2.5) is a member of $\mathcal{A}_{1,\boldsymbol{l}}$, the n-th error $\hat{\delta}_n$ of the Discrete Orthogonal Greedy Algorithm (Algorithm VI) satisfies*

$$\left\| \hat{\delta}_n \right\|_a \le \|\psi\|_{\mathcal{A}_{1,\boldsymbol{l}}}\, n^{-1/2}.$$

Now, as all members of $\hat{\mathrm{H}}_{\boldsymbol{l}}$ are finite linear combinations of functions in $\bigotimes_{i\in[N]} \hat{\mathrm{H}}^{(i)}_{l_i}$, it is clear that $\mathcal{A}_{1,\boldsymbol{l}} = \hat{\mathrm{H}}_{\boldsymbol{l}}$ in the algebraic sense. Therefore, the above theorems are of limited interest if, as in general, the true solution ψ to (2.5) does not happen to be a member of $\hat{\mathrm{H}}_{\boldsymbol{l}}$. In contrast, in the general case $\psi \in \mathrm{H}(\mathsf{D};\mathsf{M})$ we do have that the discrete greedy algorithms expounded in this chapter converge to the a-projection of ψ onto $\hat{\mathrm{H}}_{\boldsymbol{l}}$, but without an explicit rate of convergence (Theorem 4.7 and Theorem 4.8).

We would like to obtain analogues of Theorem 3.13 and Theorem 3.14 which remain relevant if the solution ψ is not a member of the finite-dimensional space $\hat{\mathrm{H}}_{\boldsymbol{l}}$. As the approximations generated by the discrete Greedy Algorithms are always members of $\hat{\mathrm{H}}_{\boldsymbol{l}}$, some measure of the distance between ψ and $\hat{\mathrm{H}}_{\boldsymbol{l}}$ must appear in the desired analogues. For the

Discrete Orthogonal Greedy Algorithm (Algorithm VI) such a result can indeed be obtained; this is the theme of the next subsection.

4.1.4. Spaces defined by approximability. The following result is an important consequence of the definition of $\mathcal{A}_{1,\boldsymbol{l}}$.

Proposition 4.12. *Let $\varphi \in \mathcal{A}_{1,\boldsymbol{l}}$ and let $\xi \in \mathrm{H}(\mathsf{D};\mathsf{M})$. Then,*

$$a(\xi,\varphi) \leq \|\varphi\|_{\mathcal{A}_{1,\boldsymbol{l}}} \sup_{\substack{s \in \bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)} \\ \|s\|_a = 1}} a(\xi, s). \tag{4.18}$$

Proof. Suppose first that $\varphi \in \mathcal{A}^o_{1,\boldsymbol{l}}(M)$ (cf. (4.17b)). Then,

$$\varphi = \sum_{k\in\Lambda} c_k w_k, \quad |\Lambda| < \infty, \quad w_k \in \bigotimes_{i=1}^{N} \hat{\mathrm{H}}_{l_i}^{(i)}, \quad \|w_k\|_a = 1, \quad \sum_{k\in\Lambda} |c_k| \leq M,$$

where Λ is a subset of some set of indices used to describe the unit-norm members of $\bigotimes_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$. It is immediate that $a(\xi,\varphi) \leq MS$, where we have used S to denote the supremum in (4.18). By continuity, this result extends to $f \in \overline{\mathcal{A}^o_{1,\boldsymbol{l}}(M)}$ (the closure being with respect to the $\mathrm{H}(\mathsf{D};\mathsf{M})$ norm). Then, by the definition of $\mathcal{A}_{1,\boldsymbol{l}}$ as the union of the $\overline{\mathcal{A}^o_{1,\boldsymbol{l}}(M)}$ with positive M and of the norm of $\mathcal{A}_{1,\boldsymbol{l}}$ as the infimum of those M for which the argument lives in $\overline{\mathcal{A}^o_{1,\boldsymbol{l}}(M)}$, the result takes the desired form. □

Remark 4.13. An analogue of Proposition 4.12 for which we have currently no need can be proved for $\mathcal{A}_1$. In that case, it is easier to work with the original definition of $\mathcal{A}_1$ we gave in (3.30), which is based on the a-norm (that is, the energy norm), instead of (topologically equivalently) on the standard $\mathrm{H}(\mathsf{D};\mathsf{M})$ norm.

Lemma 4.14. *Let $(a_n : n \neq 1)$ be a sequence of non-negative numbers satisfying $a_1 \leq A$ and*

$$a_{n+1} \leq a_n(1 - a_n/A), \quad n \geq 1,$$

for some $A > 0$. Then,

$$a_n \leq A/n, \quad n \geq 1.$$

Proof. This is Lemma 3.4 of [**DT96**]. □

The following theorem provides a bound on the error committed by the Discrete Orthogonal Greedy Algorithm if the solution ψ being approximated fails to lie in $\mathcal{A}_{1,\boldsymbol{l}} = \hat{\mathrm{H}}_{\boldsymbol{l}}$, as it is usually the case.

Theorem 4.15. *Let $\psi \in \mathrm{H}(\mathsf{D};\mathsf{M})$ and let $h \in \mathcal{A}_{1,\boldsymbol{l}}$. Then, the error $\hat{\delta}_n$ (cf. (4.7)) of the Discrete Orthogonal Greedy Algorithm (Algorithm VI) satisfies*

$$\left\|\hat{\delta}_n\right\|_a^2 \leq \|\psi - h\|_a^2 + 4\,\|h\|^2_{\mathcal{A}_{1,\boldsymbol{l}}}\, n^{-1}, \quad n \geq 1. \tag{4.19}$$

Proof. This result is essentially Theorem 2.3 of [**BCDD08**] and the proof is, *mutatis mutandis*, the same. We write it at length because the result is given there in a Hilbert-space approximation setting (which would correspond to approximation in $\mathrm{H}(\mathsf{D};\mathsf{M})'$ in our case) instead of our preferred PDE-solving point or view (which focuses on $\mathrm{H}(\mathsf{D};\mathsf{M})$).

We start from

$$\begin{aligned}\left\|\hat{\delta}_{n-1}\right\|_a^2 = a(\hat{\delta}_{n-1},\psi) &= a(\hat{\delta}_{n-1},h+\psi-h) \leq a(\hat{\delta}_{n-1},h) + \left\|\hat{\delta}_{n-1}\right\|_a \|\psi-h\|_a \\ &\leq \|h\|_{\mathcal{A}_{1,\boldsymbol{l}}} \sup_{\substack{s\in\bigotimes_{i\in[N]}\hat{\mathrm{H}}_{l_i}^{(i)} \\ \|s\|_a=1}} a(\hat{\delta}_{n-1},s) + \left\|\hat{\delta}_{n-1}\right\|_a \|\psi-h\|_a \\ &= \|h\|_{\mathcal{A}_{1,\boldsymbol{l}}} \left\|\bigotimes_{i=1}^N r_n^{(i)}\right\|_a + \left\|\hat{\delta}_{n-1}\right\|_a \|\psi-h\|_a \\ &\leq \|h\|_{\mathcal{A}_{1,\boldsymbol{l}}} \left\|\bigotimes_{i=1}^N r_n^{(i)}\right\|_a + \frac{1}{2}\left(\left\|\hat{\delta}_{n-1}\right\|_a^2 + \|\psi-h\|_a^2\right),\end{aligned}$$

for $n \geq 1$, which comes from the Galerkin orthogonality caused by the relaxation step `2.2` of Algorithm VI, Proposition 4.12 and Lemma 4.6. Rearranging, and using the shorthand $a_{n-1} = \left\|\hat{\delta}_{n-1}\right\|_a^2 - \|\psi-h\|_a^2$,

$$\frac{a_{n-1}}{2\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}} \leq \left\|\bigotimes_{i=1}^N r_n^{(i)}\right\|_a.$$

It is easy to see that if a_{n-1} is negative, the desired result will trivially hold from $n-1$ on. Suppose, in contrast, that it is non-negative. Then, we can square both sides of the above inequality and obtain

$$\frac{a_{n-1}^2}{4\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}^2} \leq \left\|\bigotimes_{i=1}^N r_n^{(i)}\right\|_a^2.$$

Combining this with (4.14) in the proof of Theorem 4.8, we obtain

$$\left\|\hat{\delta}_n\right\|_a^2 \leq \left\|\hat{\delta}_{n-1}\right\|_a^2 - \frac{a_{n-1}^2}{4\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}^2}.$$

If we substract $\|\psi-h\|_a^2$ from both sides we are left with

$$a_n \leq a_{n-1}\left(1 - \frac{a_{n-1}}{4\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}^2}\right). \tag{4.20}$$

If $a_0 \geq 4\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}^2$, (4.20) implies that $a_1 \leq 0$, which puts us is the easy situation described above and thus the result holds. Otherwise, $a_1 \leq a_0 \leq 4\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}^2$. Combining this with (4.20) makes the sequence $(a_n \colon n \geq 1)$ fall within the scope of Lemma 4.14 with $A = 4\|h\|_{\mathcal{A}_{1,\boldsymbol{l}}}^2$ and the desired result follows. □

Now we define a family of subspaces of $\mathrm{H}(\mathsf{D};\mathsf{M})$ in terms of the approximability of their members by elements from $\mathcal{A}_{1,\boldsymbol{l}}$ (such definitions are common practice in the literature of

approximation theory): Given some function $\theta\colon \mathbb{N}^N \to \mathbb{R}$ we define

$$\mathcal{Z}_\theta := \Big\{ \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M})\colon \forall\, \boldsymbol{l} \in \mathbb{N}^N\ \exists\, h = h(\boldsymbol{l}) \in \mathcal{A}_{1,\boldsymbol{l}}\colon \\ \|h\|_{\mathcal{A}_{1,\boldsymbol{l}}} \le C,\ \|\varphi - h\|_a \le C\theta(\boldsymbol{l}),\ C>0 \text{ independent of } \boldsymbol{l} \Big\} \quad (4.21a)$$

equipped with the norm

$$\|\varphi\|_{\mathcal{Z}_\theta} := \inf \Big\{ C>0\colon \forall\, \boldsymbol{l} \in \mathbb{N}^N\ \exists\, h = h(\boldsymbol{l}) \in \mathcal{A}_{1,\boldsymbol{l}}\colon \\ \|h\|_{\mathcal{A}_{1,\boldsymbol{l}}} \le C,\ \|\varphi - h\|_a \le C\theta(\boldsymbol{l}),\ C>0 \text{ independent of } \boldsymbol{l} \Big\}. \quad (4.21b)$$

One must think of θ as an approximation rate-describing function (indeed, it transpires from Corollary 4.16 below that one would at least want θ to tend to 0 as all the entries of its argument tend to ∞). Note that $\mathcal{Z}_\theta$ depends on the family of subspaces $\hat{\mathrm{H}}_{\boldsymbol{l}} \subset \mathrm{H}(\mathsf{D};\mathsf{M})$ indexed by $\boldsymbol{l} \in \mathbb{N}^N$, which, in turn, depends on the families of subspaces $\hat{\mathrm{H}}^{(i)}_{l_i} \subset \mathrm{H}(D_i;M_i)$, $i \in [N]$. The justification for the definition of the space $\mathcal{Z}_\theta$ is given by the corollary that follows.

Corollary 4.16. *If the true solution $\psi \in \mathrm{H}(\mathsf{D};\mathsf{M})$ of (2.5) lies in $\mathcal{Z}_\theta$, then the application of the $\hat{\mathrm{H}}_{\boldsymbol{l}}$-based Discrete Orthogonal Greedy Algorithm (Algorithm VI) results in an error bounded as*

$$\left\|\hat{\delta}_n\right\|_a \le 2\,\|\psi\|_{\mathcal{Z}_\theta}\left(n^{-1/2} + \theta(\boldsymbol{l})\right).$$

Proof. From Theorem 4.15 we have that

$$\left\|\hat{\delta}_n\right\|_a^2 \le \|\psi - h\|_a^2 + 4\,\|h\|^2_{\mathcal{A}_{1,\boldsymbol{l}}}\, n^{-1},$$

for any $h \in \mathcal{A}_{1,\boldsymbol{l}}$. Choosing $h = h(\boldsymbol{l})$ given in the definition of $\mathcal{Z}_\theta$, taking square roots and using the fact that the 2-norm in $\mathbb{R}^2$ is bounded by the 1-norm we obtain the desired result. □

The above corollary gives a composite rate of convergence of the Discrete Orthogonal Greedy Algorithm (Algorithm VI) with respect to the iteration number and the discretization parameter $\boldsymbol{l}$. Our next task is to identify families of finite-dimensional subspaces $\hat{\mathrm{H}}^{(i)}_{l_i} \subset \mathrm{H}(D_i;M_i)$ such that subspaces of $\mathcal{Z}_\theta$ can be described in a less abstract fashion. What we do next was inspired by the example given in equation (2.58) of [**BCDD08**].

4.1.5. Spectral bases. Let $\tilde{e}_{\boldsymbol{n}} := e_{\boldsymbol{n}}/\mathsf{M}$ for $\boldsymbol{n} \in \mathbb{N}^N$, where $\big(e_{\boldsymbol{n}}\colon \boldsymbol{n} \in \mathbb{N}^N\big)$ is the complete eigenfunction system described in Lemma 3.16 for the $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$-eigenvalue problem (3.34). As division by M defines an isometric isomorphism between $\mathrm{H}(\mathsf{D};\mathsf{M})$ and $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ on the one hand and between $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ and $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ on the other, if all four of these spaces are normed

with their standard norms, it follows that $\left(\tilde{e}_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ is an orthogonal and complete eigenfunction system for $\mathrm{H}(\mathsf{D};\mathsf{M})$ and thus

$$\varphi = \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \langle \tilde{e}_{\boldsymbol{n}}, \varphi\rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \tilde{e}_{\boldsymbol{n}} + \sum_{\boldsymbol{n}\not\leq\boldsymbol{l}} \langle \tilde{e}_{\boldsymbol{n}}, \varphi\rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \tilde{e}_{\boldsymbol{n}} \qquad \forall\, \varphi \in \mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D}) \supset \mathrm{H}(\mathsf{D};\mathsf{M}).$$

As M and each $e_{\boldsymbol{n}}$ have a tensor-product structure, so do the $\tilde{e}_{\boldsymbol{n}}$; indeed,

$$\tilde{e}_{\boldsymbol{n}} = \bigotimes_{i=1}^{N} \frac{e^{(i)}_{n_i}}{M_i},$$

where the $\left(e^{(i)}_n \colon n \in \mathbb{N}\right)$ are normalized and complete systems of eigenfunctions to the single-spring problems (3.33) and the M_i are the partial Maxwellians (1.24). In this subsection we explore the consequences of fixing the finite dimensional spaces $\hat{\mathrm{H}}^{(i)}_l \subset \mathrm{H}(D_i; M_i)$ according to

$$\hat{\mathrm{H}}^{(i)}_l = \operatorname{span}\left(e^{(i)}_n/M_i \colon n \leq l\right) \qquad \forall\, i \in [N], \tag{4.22}$$

whence

$$\hat{\mathrm{H}}_{\boldsymbol{l}} = \operatorname{span}\left(\tilde{e}_{\boldsymbol{n}} \colon \boldsymbol{n} \leq \boldsymbol{l}\right).$$

For any $\boldsymbol{l}$, we define the truncated-expansion operator $P_{\boldsymbol{l}} \colon \mathrm{H}(\mathsf{D};\mathsf{M}) \to \hat{\mathrm{H}}_{\boldsymbol{l}}$ by

$$P_{\boldsymbol{l}}(\varphi) := \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \langle \tilde{e}_{\boldsymbol{n}}, \varphi\rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \tilde{e}_{\boldsymbol{n}} \qquad \forall\, \varphi \in \mathrm{H}(\mathsf{D};\mathsf{M}). \tag{4.23}$$

For $\Sigma := \left(\sigma_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right)$ a net of positive real numbers, let $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$ denote the space $\mathsf{M}\,\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$, equipped with the norm $\|\cdot\|_{\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})} := \left\|\mathsf{M}^{-1}\cdot\right\|_{\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})}$. We recall that $\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$ was defined in (3.39) as the subspace of $\mathrm{L}^2_{\mathsf{M}}(\mathsf{D})$ consisting of functions whose squared Fourier coefficients, weighted by the $\sigma_{\boldsymbol{n}}$, have finite sum. As, by design, $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$ and $\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$ are isometrically isomorphic (via the operation $\varphi \mapsto \mathsf{M}^{-1}\varphi$; cf. (2.4)), equivalent characterizations of $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$ and its norm are

$$\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M}) = \left\{ \varphi \in \mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D}) \colon \sum_{\boldsymbol{n}\in\mathbb{N}^N} \sigma_{\boldsymbol{n}} \langle \varphi, \tilde{e}_{\boldsymbol{n}}\rangle^2_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} < \infty \right\}, \tag{4.24a}$$

and

$$\|\varphi\|_{\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})} = \left(\sum_{\boldsymbol{n}\in\mathbb{N}^N} \sigma_{\boldsymbol{n}} \langle \varphi, \tilde{e}_{\boldsymbol{n}}\rangle^2_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right)^{1/2}. \tag{4.24b}$$

Let us recall that $\big(\lambda_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\big)$ is the collection of eigenvalues of the problem (3.34); so $\|\tilde{e}_{\boldsymbol{n}}\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} = \lambda_{\boldsymbol{n}}$ for all $\boldsymbol{n} \in \mathbb{N}^N$. We have, on the one hand, the approximability estimate

$$\begin{aligned}
\|\varphi - P_{\boldsymbol{l}}(\varphi)\|_a &= \left\| \sum_{\boldsymbol{n} \not\leq \boldsymbol{l}} \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \tilde{e}_{\boldsymbol{n}} \right\|_a \leq C \left\| \sum_{\boldsymbol{n} \not\leq \boldsymbol{l}} \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \tilde{e}_{\boldsymbol{n}} \right\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} \\
&= C \left[\sum_{\boldsymbol{n} \not\leq \boldsymbol{l}} \left| \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right|^2 \lambda_{\boldsymbol{n}} \right]^{1/2} \\
&= C \left[\sum_{\boldsymbol{n} \not\leq \boldsymbol{l}} \sigma_{\boldsymbol{n}} \left| \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right|^2 \frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}} \right]^{1/2} \\
&\leq C \sup_{\boldsymbol{n} \not\leq \boldsymbol{l}} \sqrt{\frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}}} \, \|\varphi\|_{\mathrm{H}^\Sigma(\mathsf{D};\mathsf{M})} \, .
\end{aligned} \tag{4.25}$$

On the other hand, we have the $\mathcal{A}_{1,\boldsymbol{l}}$-stability estimate

$$\begin{aligned}
\|P_{\boldsymbol{l}}(\varphi)\|_{\mathcal{A}_{1,\boldsymbol{l}}} &= \left\| \sum_{\boldsymbol{n} \leq \boldsymbol{l}} \|\tilde{e}_{\boldsymbol{n}}\|_a \, \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \frac{\tilde{e}_{\boldsymbol{n}}}{\|\tilde{e}_{\boldsymbol{n}}\|_a} \right\|_{\mathcal{A}_{1,\boldsymbol{l}}} \leq \sum_{\boldsymbol{n} \leq \boldsymbol{l}} \|\tilde{e}_{\boldsymbol{n}}\|_a \left| \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right| \\
&\leq C \sum_{\boldsymbol{n} \leq \boldsymbol{l}} \|\tilde{e}_{\boldsymbol{n}}\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} \left| \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right| = C \sum_{\boldsymbol{n} \leq \boldsymbol{l}} \sqrt{\lambda_{\boldsymbol{n}}} \left| \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right| \\
&\leq C \left(\sum_{\boldsymbol{n} \leq \boldsymbol{l}} \frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}} \right)^{1/2} \left(\sum_{\boldsymbol{n} \leq \boldsymbol{l}} \sigma_{\boldsymbol{n}} \left| \langle \tilde{e}_{\boldsymbol{n}}, \varphi \rangle_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \right|^2 \right)^{1/2} \\
&\leq C \left(\sum_{\boldsymbol{n} \in \mathbb{N}^N} \frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}} \right)^{1/2} \|\varphi\|_{\mathrm{H}^\Sigma(\mathsf{D};\mathsf{M})} \, .
\end{aligned} \tag{4.26}$$

Therefore, as long as the condition

$$\sum_{\boldsymbol{n} \in \mathbb{N}^N} \frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}} < \infty \tag{4.27}$$

is fulfilled, one can show that

$$\mathrm{H}^\Sigma(\mathsf{D};\mathsf{M}) \subset \mathcal{Z}_\theta \qquad \text{with} \qquad \theta(\boldsymbol{l}) := \sup_{\boldsymbol{n} \not\leq \boldsymbol{l}} \sqrt{\frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}}}.$$

Indeed, given any $\varphi \in \mathrm{H}^\Sigma(\mathsf{D};\mathsf{M})$, a suitable function $h \in \mathcal{A}_{1,\boldsymbol{l}}$ that satisfies the conditions given in the definition (4.21) of $\mathcal{Z}_\theta$ is given by $P_{\boldsymbol{l}}(\varphi)$, due to (4.25) and (4.26) and the assumption of the condition (4.27). We can now state the lemma that follows.

Lemma 4.17. *Let $\Sigma = \big(\sigma_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\big)$ satisfy the condition* (4.27) *and suppose that the true solution ψ to the Fokker–Planck equation* (2.5) *lies in the space $\mathrm{H}^\Sigma(\mathsf{D};\mathsf{M})$ (equivalently,*

$\psi/\mathsf{M} \in \mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$*). If* $\hat{\mathrm{H}}_{\boldsymbol{l}}$ *is fixed according to*

$$\hat{\mathrm{H}}_{\boldsymbol{l}} = \operatorname{span}\left(\tilde{e}_{\boldsymbol{n}} \colon \boldsymbol{n} \leq \boldsymbol{l}\right) \qquad \forall\, \boldsymbol{l} \in \mathbb{N}^N,$$

the error $\hat{\delta}_n$ *committed at the end of the* n*-th iteration of the Discrete Orthogonal Greedy Algorithm based on* $\hat{\mathrm{H}}_{\boldsymbol{l}}$ *(Algorithm VI) obeys*

$$\left\|\hat{\delta}_n\right\|_a \leq C \left\|\psi\right\|_{\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})} \left(n^{-1/2} + \sup_{\boldsymbol{n} \not\leq \boldsymbol{l}} \sqrt{\frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}}}\right) \tag{4.28}$$

for some $C > 0$*.*

Proof. This is a consequence of the above discussion and Corollary 4.16. □

Next, we want to particularize Lemma 4.17 to function spaces based on the nets of weights studied in in Subsection 3.3.2; that is,

$$\mathrm{T}^{(k)} = \left(\tau^{(k)}_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right) \qquad \text{where} \qquad \tau^{(k)}_{\boldsymbol{n}} = \left(\sum_{i=1}^{N} \lambda^{(i)}_{n_i}\right)^k + \prod_{i=1}^{N} \left(\lambda^{(i)}_{n_i}\right)^k \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N,$$

and

$$\Upsilon^{(k)} = \left(\upsilon^{(k)}_{\boldsymbol{n}} \colon \boldsymbol{n} \in \mathbb{N}^N\right) \qquad \text{where} \qquad \upsilon^{(k)}_{\boldsymbol{n}} = \left(\sum_{i=1}^{N} \lambda^{(i)}_{n_i}\right)^k \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N,$$

where the $\lambda^{(i)}_{n_i}$ are eigenvalues of the single-spring problems (3.33). The following lemma gives the associated approximation rates.

Lemma 4.18. *Let* $\Sigma = \mathrm{T}^{(k)}$ *or* $\Sigma = \Upsilon^{(k)}$*. Then, if* $k \geq 1$*, there exists* $C > 0$ *such that*

$$\left\|\varphi - P_{\boldsymbol{l}}(\varphi)\right\|_a \leq C \left(\min_{1 \leq i \leq N} l_i\right)^{(1-k)/d} \left\|\varphi\right\|_{\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})} \qquad \forall\, \varphi \in \mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M}).$$

Proof. On account of the adoption of Hypothesis C, Lemma 3.16 and the positivity of all of the involved eigenvalues, there exist positive constants c_1 and c_2, and for $i \in [N]$, constants $c^{(i)}_1$ and $c^{(i)}_2$, such that

$$c^{(i)}_1 n^{2/d} \leq \lambda^{(i)}_n \leq c^{(i)}_2 n^{2/d} \qquad \forall\, n \in \mathbb{N}$$

and

$$c_1 \sum_{i=1}^{N} n_i^{2/d} \leq \lambda_{\boldsymbol{n}} \leq c_2 \sum_{i=1}^{N} n_i^{2/d} \qquad \forall\, \boldsymbol{n} \in \mathbb{N}^N.$$

Therefore, the supremum $\sup_{\boldsymbol{n} \not\leq \boldsymbol{l}} \sqrt{\frac{\lambda_{\boldsymbol{n}}}{\sigma_{\boldsymbol{n}}}}$ appearing in (4.25) can be bounded by a positive constant times the supremum among $\{\boldsymbol{n} \in \mathbb{N}^n \colon \boldsymbol{n} \not\leq \boldsymbol{l}\}$ of the square roots of

$$\frac{\sum_{i=1}^{N} n_i^{2/d}}{\left(\sum_{i=1}^{N} n_i^{2/d}\right)^k} \qquad \text{and} \qquad \frac{\sum_{i=1}^{N} n_i^{2/d}}{\left(\sum_{i=1}^{N} n_i^{2/d}\right)^k + \prod_{i=1}^{N} \left(n_i^{2/d}\right)^k}$$

in the case $\Sigma = \Upsilon^{(k)}$ and $\Sigma = \mathrm{T}^{(k)}$, respectively. In either case, the supremum is attained at a multi-index with all its entries equal to 1 but for the j-th, which is equal to $l_j + 1$, where, in turn, $j \in \arg\min_{1 \le j \le N} l_j$. The supremum of (the square root) of both expressions is of the order $l_j^{(1-k)/d}$, and thus the result. □

Remark 4.19. Despite $\mathrm{H}^{\mathrm{T}^{(k)}}(\mathsf{D};\mathsf{M})$ being a space of functions with more regularity than that assumed by the functions in $\mathrm{H}^{\Upsilon^{(k)}}(\mathsf{D};\mathsf{M})$ with the same parameter k, the rate of decay given in Lemma 4.18 is the same for both spaces.

The condition (4.27) on the net Σ is identical to the first of the two conditions in (3.38) in Subsection 3.3.2. Thus, we can re-use results obtained there for the families of weight nets $\mathrm{T}^{(k)}$ and $\Upsilon^{(k)}$. We obtain the following theorem.

Theorem 4.20. *Let the true solution ψ to the Fokker–Planck equation (2.5) lie in $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$, where Σ is either*

(1) $\mathrm{T}^{(k)}$ with $k \ge d+1$, or
(2) $\Upsilon^{(k)}$ with $k > 1 + \frac{1}{2}Nd$.

Further, let $\hat{\mathrm{H}}_{\boldsymbol{l}}$ be fixed according to

$$\hat{\mathrm{H}}_{\boldsymbol{l}} = \operatorname{span}(\tilde{e}_{\boldsymbol{n}} \colon \boldsymbol{n} \le \boldsymbol{l}) \qquad \forall\, \boldsymbol{l} \in \mathbb{N}^N.$$

Then, the error $\hat{\delta}_n$ committed at the end of the n-th iteration of the Discrete Orthogonal Greedy Algorithm based on $\hat{\mathrm{H}}_{\boldsymbol{l}}$ (Algorithm VI) obeys

$$\left\| \hat{\delta}_n \right\|_a \le C \left\| \psi \right\|_{\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})} \left(n^{-1/2} + \left(\min_{1 \le i \le N} l_i \right)^{(1-k)/d} \right) \tag{4.29}$$

for some $C > 0$.

Proof. From the proofs of Theorem 3.18 and Theorem 3.20, we know that the condition (4.27) is satisfied under the stated constraints on k. Therefore, we can appeal to Lemma 4.17 and, by way of the estimate given in Lemma 4.18, obtain the estimate (4.29) from (4.28). □

Remark 4.21.

(1) We gave a partial characterization of some $\mathrm{H}^{\mathrm{T}^{(k)}}_{\mathsf{M}}(\mathsf{D})$ spaces in terms of weighted Sobolev spaces of dominating mixed smoothness in Theorem 3.30 and Remark 3.31. Similar partial characterizations are possible for the $\mathrm{H}^{\Upsilon^{(k)}}_{\mathsf{M}}(\mathsf{D})$ spaces. As for any (real-valued, $\mathbb{N}^N$-indexed) net Σ the space $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$ was defined as $\mathsf{M}\,\mathrm{H}^{\Sigma}_{\mathsf{M}}(\mathsf{D})$, we immediately have partial characterizations of some of the $\mathrm{H}^{\mathrm{T}^{(k)}}_{\mathsf{M}}(\mathsf{D})$ and the $\mathrm{H}^{\Upsilon^{(k)}}_{\mathsf{M}}(\mathsf{D})$ spaces, in terms of which Theorem 4.20 was posed.
(2) The more regular the true solution ψ is (in this sense of summability of weighted squared Fourier coefficients), the faster the second term in the estimate (4.29) of Theorem 4.20 will decay. The first term, however, is not likewise affected by the regularity of ψ.

As in practice the family of eigenfunctions $\big(\tilde{e}_{\boldsymbol{n}}\colon \boldsymbol{n}\in\mathbb{N}^N\big)$ might not be at hand, it is of interest to analyze the behavior of the Discrete Orthogonal Algorithm based on more readily available subspaces of $\mathrm{H}(\mathsf{D};\mathsf{M})$. In the next subsection we study the effect of using subspaces of $\mathrm{H}(\mathsf{D};\mathsf{M})$ where the member functions are the product of the Maxwellian M and a polynomial; via (2.4), this means taking a subspaces of $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ consisting of polynomials.

4.1.6. Polynomial-based subspaces. For what follows, we restrict ourselves to the case $d=1$; that is, the case in which each domain D_i is an interval. Then, we consider setting, for $i\in[N]$,

$$\hat{\mathrm{H}}^{(i)}_l = M_i\,\mathbb{P}_l, \tag{4.30}$$

which results, according to the definition (4.2), in

$$\hat{\mathrm{H}}_{\boldsymbol{l}} = \mathsf{M}\,\mathbb{P}_{\boldsymbol{l}}. \tag{4.31}$$

Here, $\mathbb{P}_l$ denotes the space of univariate polynomials of degree less than or equal or equal to l, and $\mathbb{P}_{\boldsymbol{l}}$ denotes the space of N-variate polynomials where the degree with respect to the j-th variable is less than or equal than l_j, for $j\in[N]$.

Note that, as it is customary when using spaces of polynomials, we have switched to a zero-based indexing of the $\hat{\mathrm{H}}^{(i)}_l$ and $\hat{\mathrm{H}}_{\boldsymbol{l}}$; hence, their dimension is $l+1$ and $\prod_{i\in[N]}(l_i+1)$. This is of no special consequence.

In this context, for reasons that will become apparent below, the Jacobi polynomials prove to be a very useful tool. We start, then, by briefly reviewing those properties of the Jacobi polynomials we will need in the sequel. For a full definition we refer to [**BM97**, Section 19]. Let $\alpha>-1$, $\Lambda:=(-1,1)$ and let $\rho\colon\Lambda\to\mathbb{R}$ be defined by $\rho(y)=1-y^2$. Given $m\in\mathbb{N}$ and $\alpha>-1$ we define the non-uniformly weighted Sobolev space

$$\mathrm{V}^m_{\rho^\alpha}(\Lambda) := \Big\{ v \text{ measurable}\colon \partial_k v\in \mathrm{L}^2_{\rho^{\alpha+k}}(\Lambda),\ \text{for } 0\le k\le m\Big\}, \tag{4.32a}$$

which we equip with its natural norm:

$$\|\cdot\|_{\mathrm{V}^m_{\rho^\alpha}(\Lambda)} = \left[\sum_{k=0}^m \|\partial_k\cdot\|^2_{\mathrm{L}^2_{\rho^{\alpha+k}}(\Lambda)}\right]^{1/2}. \tag{4.32b}$$

Then, the sequence of Jacobi polynomials of (symmetric) parameter α, $\Big(J^{(\alpha)}_n\colon n\in\mathbb{N}_0\Big)$, is an orthogonal basis of $\mathrm{L}^2_{\rho^\alpha}(\Lambda)=\mathrm{V}^0_{\rho^\alpha}(\Lambda)$ and

$$\int_\Lambda {J^{(\alpha)}_n}'\varphi'\rho^{\alpha+1} = n(n+2\alpha+1)\int_\Lambda J^{(\alpha)}_n\varphi\rho^\alpha \qquad \forall\,\varphi\in\mathrm{V}^1_{\rho^\alpha}(\Lambda). \tag{4.33}$$

For any $\alpha>-1$ and $n\in\mathbb{N}_0$, the Jacobi polynomial $J^{(\alpha)}_n$ has degree n. Then, for all $l\in\mathbb{N}_0$, $\Big(J^{(\alpha)}_n\colon n\in\{0,\dots,l\}\Big)$ is a basis of $\mathbb{P}_l$. We will assume that the Jacobi polynomials come

normalized in the standard way—namely,

$$J_n^{(\alpha)}(1) = \frac{\Gamma(n+\alpha+1)}{n!\,\Gamma(\alpha+1)}.$$

Then, the Jacobi polynomials obey the following three identities, which we will exploit later:

$$J_n^{(\alpha)\prime} = \frac{n+2\alpha+1}{2} J_{n-1}^{(\alpha+1)} \qquad \forall\, n \in \mathbb{N}, \tag{4.34}$$

$$\left\| J_n^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = \frac{2^{2\alpha+1}\Gamma(n+\alpha+1)^2}{(2n+2\alpha+1)\,n!\,\Gamma(n+2\alpha+1)} \qquad \forall\, n \in \mathbb{N}_0, \tag{4.35}$$

and

$$\frac{n+2\alpha+1}{n+\alpha+1} J_{n+1}^{(\alpha)\,\prime} = (2n+2\alpha+1) J_n^{(\alpha)} + \frac{n+\alpha}{n+2\alpha} J_{n-1}^{(\alpha)\,\prime} \qquad \forall\, n \in \mathbb{N} \tag{4.36}$$

(cf. [**BM97**, equations (19.5), (19.8) and (19.11)]).

Lemma 4.22. *Let $\alpha > -1$ and let $\varphi \in \mathrm{V}^m_{\rho^\alpha}(\Lambda)$. Then, there exists a positive constant C, which only depends on α and m, such that*

$$\sum_{n=0}^{\infty} (n+1)^{2m} a_n^2 \le C \left\|\varphi\right\|^2_{\mathrm{V}^m_{\rho^\alpha}(\Lambda)} \qquad \textit{where} \qquad a_n := \left\langle \varphi, J_n^{(\alpha)} \right\rangle_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} \Big/ \left\| J_n^{(\alpha)} \right\|_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)}$$

(that is, the a_n are the coefficients of the $\mathrm{L}^2_{\rho^\alpha}(\Lambda)$ expansion of φ in terms of $\mathrm{L}^2_{\rho^\alpha}(\Lambda)$-normalized Jacobi polynomials with parameter α).

Proof. We start with the case $m = 1$. Therefore, let us suppose that we have $\varphi \in \mathrm{L}^2_{\rho^\alpha}(\Lambda)$ and $\varphi' \in \mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)$. Thus, one can expand φ' in terms of Jacobi polynomials with parameter $\alpha+1$ in the $\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)$ norm. Parseval's identity gives

$$\left\|\varphi'\right\|^2_{\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)} = \sum_{n=0}^{\infty} b_n^2 \qquad \text{where} \qquad b_n := \langle \varphi', J_n^{(\alpha+1)} \rangle_{\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)} \Big/ \left\| J_n^{(\alpha+1)} \right\|_{\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)}.$$

We can rephrase the identity (4.34) to give

$$J_n^{(\alpha+1)} = \frac{2}{n+2\alpha+2} J_{n+1}^{(\alpha)\,\prime} \qquad \forall\, n \in \mathbb{N}_0,$$

which allows for writing

$$\begin{aligned} \langle \varphi', J_n^{(\alpha+1)} \rangle_{\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)} &= \frac{2}{n+2\alpha+2} \left\langle \varphi', J_{n+1}^{(\alpha)\,\prime} \right\rangle_{\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)} \\ &= \frac{2(n+1)(n+2\alpha+2)}{n+2\alpha+2} \langle \varphi, J_{n+1}^{(\alpha)} \rangle_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = 2(n+1) \langle \varphi, J_{n+1}^{(\alpha)} \rangle_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} \end{aligned}$$

and

$$\langle J_n^{(\alpha+1)}, J_n^{(\alpha+1)} \rangle_{\mathrm{L}^2_{\rho^{\alpha+1}}(\Lambda)} = \frac{4(n+1)}{n+2\alpha+2} \langle J_{n+1}^{(\alpha)}, J_{n+1}^{(\alpha)} \rangle_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)}.$$

Therefore,

$$b_n = (n+1)^{1/2} (n+2\alpha+2)^{1/2} a_{n+1}$$

and

$$\|\varphi\|^2_{\mathrm{V}^1_{\rho^\alpha}(\Lambda)} = \sum_{n=0}^{\infty} \left(b_n^2 + a_n^2\right) = \sum_{n=0}^{\infty} (n+1)(n+2\alpha+2)a_{n+1}^2 + \sum_{n=0}^{\infty} a_n^2$$
$$= \sum_{n=0}^{\infty} \left[n(n+2\alpha+1)+1\right] a_n^2 \geq C_\alpha \sum_{n=0}^{\infty} (n+1)^2 a_n^2,$$

where $C_\alpha > 0$ depends on α only; thus, the result for $m = 1$. A simple induction argument extends this result to higher m. □

Let us define, for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in (-1, \infty)^N$ and $\boldsymbol{n} \in \mathbb{N}_0^N$, the tensorized Jacobi polynomials and weights

$$J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} := \bigotimes_{i=1}^{N} J_{n_i}^{(\alpha_i)} \quad \text{and} \quad \rho^{\boldsymbol{\alpha}} := \bigotimes_{i=1}^{N} \rho^{\alpha_i}. \tag{4.37}$$

Also, let $\boldsymbol{e}_j$ denote, for $j \in [N]$, the j-th canonical vector in $\mathbb{R}^N$; i.e., that vector of $\mathbb{R}^N$ whose entries are all 0 but for the j-th, which is 1. Straightforward consequences of these definitions and the corresponding univariate properties are that $\left(J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \colon \boldsymbol{n} \in \mathbb{N}_0^N\right)$ forms a complete orthogonal system of $\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$ and, whenever $\varphi \in \mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$ and $\partial_j \varphi \in \mathrm{L}^2_{\rho^{\boldsymbol{\alpha}+\boldsymbol{e}_j}}(\Lambda^N)$,

$$\left\langle \partial_j J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})}, \partial_j \varphi \right\rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}+\boldsymbol{e}_j}}(\Lambda^N)} = n_j(n_j + 2\alpha_j + 1) \left\langle J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})}, \varphi \right\rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}. \tag{4.38}$$

On defining the spaces

$$\mathrm{V}^m_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N) := \left\{ \varphi \text{ measurable} \colon \partial_\kappa \varphi \in \mathrm{L}^2_{\rho^{\boldsymbol{\alpha}+\kappa}}(\Lambda^N),\ \kappa \in \mathbb{N}_0^N,\ |\kappa|_1 \leq m \right\} \tag{4.39}$$

and

$$\mathrm{V}^{m,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N) := \left\{ \varphi \text{ measurable} \colon \partial_\kappa \varphi \in \mathrm{L}^2_{\rho^{\boldsymbol{\alpha}+\kappa}}(\Lambda^N),\ \kappa \in \mathbb{N}_0^N,\ |\kappa|_\infty \leq m \right\}, \tag{4.40}$$

equipped with their natural norms, we can generalize Lemma 4.22.

Lemma 4.23. *Let each component of $\boldsymbol{\alpha} \in \mathbb{R}^N$ be greater than -1.*

(1) If $\varphi \in \mathrm{V}^m_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$, then there exists a positive constant C, which only depends on $\boldsymbol{\alpha}$ and m, such that

$$\sum_{\boldsymbol{n} \in \mathbb{N}_0^N} \left[\sum_{i=1}^{N} (n_i + 1)^2 \right]^m a_{\boldsymbol{n}}^2 \leq C \|\varphi\|^2_{\mathrm{V}^m_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)},$$

where $a_{\boldsymbol{n}} := \langle \varphi, J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} / \|J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})}\|_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}$ for all $\boldsymbol{n}$ in $\mathbb{N}_0^N$.

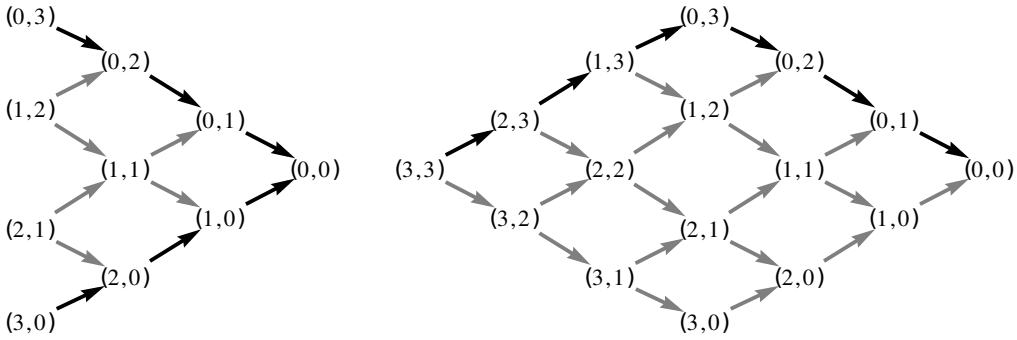

FIGURE 4.1. Differentiability Graphs. *Left*: Graph of derivatives of a function in $\mathrm{V}^m_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$—each vertex corresponds to a derivative and each edge points from a derivative to an antiderivative. *Right*: The corresponding graph for $\mathrm{V}^{m,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$. At both graphs $m = 3$ and $N = 2$. The darkened paths depict the paths taken by the induction procedure sketched in the proof of Lemma 4.23.

(2) If $\varphi \in \mathrm{V}^{m,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$, then there exists a positive constant C, which only depends on $\boldsymbol{\alpha}$ and m, such that

$$\sum_{\boldsymbol{n}\in\mathbb{N}_0^N} \left[\left[\sum_{i=1}^N (n_i+1)^2 \right]^m + \prod_{i=1}^N (n_i+1)^{2m} \right] a_{\boldsymbol{n}}^2 \leq C \left\| \varphi \right\|^2_{\mathrm{V}^{m,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)},$$

with the $a_{\boldsymbol{n}}$ defined in the same manner.

Proof. As this is almost completely analogous to the proof of Lemma 4.22 we will only remark on three aspects which distinguish the latter from this proof. The first is that here (4.38) takes the place of (4.33). The second is that in the proof of part (1), coefficients of the form $(n_i+1)^{2m}$ are generated by starting with $\partial_{m\boldsymbol{e}_i}\varphi$, going to $\partial_{(m-1)\boldsymbol{e}_i}\varphi$ and so on until φ gaining a factor of $(n_i+1)^2$ each time. As we can do this with all the coordinate directions, a factor of the form $\sum_{i\in[N]}(n_i+1)^{2m}$ is obtained; the preferred form $\left[\sum_{i\in[N]}(n_i+1)^2\right]^m$ follows by elementary means. The third aspect is that in the proof of part (2), the extra term $\prod_{i\in[N]}(n_i+1)^{2m}$ follows because this time we can start from $\partial_{(m,\ldots,m)}\varphi$ and go to φ obtaining the factors $(n_i+1)^{2m}$ sequentially (whereas they multiply each other) as opposed to in parallel (where we could only sum them). We give a graphic representation of these 'derivative-hopping' arguments in Figure 4.1. □

Remark 4.24. The 'derivative-hopping' argument given in Lemma 4.23 (on series expansions with respect to Jacobi polynomials) resembles the argument given in the proof of Lemma 3.29 (on series expansions with respect to eigenfunctions of Fokker–Planck operators). They differ, however, in that in this non-uniformly weighted setting we know enough of the basis functions—the Jacobi polynomials—that we can perform the analysis by one derivative at a time, as opposed to the one-second-order-operator at a time procedure of Lemma 3.29.

Given $\boldsymbol{\alpha} \in (-1,\infty)$ and $\boldsymbol{l} \in \mathbb{N}_0^N$, we denote by $P_{\boldsymbol{l}}^{(\boldsymbol{\alpha})}$ the orthogonal projector of $\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$ onto the space of N-variate polynomials of degree less than or equal to $\boldsymbol{l}$. As the Jacobi polynomials of parameter $\boldsymbol{\alpha}$ form a complete orthogonal system in $\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$, the projector $P_{\boldsymbol{l}}^{(\boldsymbol{\alpha})}$ has the explicit form

$$P_{\boldsymbol{l}}^{(\boldsymbol{\alpha})}(\varphi) = \sum_{\boldsymbol{n} \le \boldsymbol{l}} \frac{\left\langle \varphi, J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \right\rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}}{\left\langle J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})}, J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \right\rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \qquad \forall\, \varphi \in \mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N). \tag{4.41}$$

From the literature on Spectral Methods we have the following approximation result.

Lemma 4.25. *Let $\boldsymbol{\alpha} \in (-1,\infty)^N$. Then, there exists some $C > 0$ such that, for all $\boldsymbol{l} \in \mathbb{N}^N$ and $k \ge 2$,*

$$\left\| \varphi - P_{\boldsymbol{l}}^{(\boldsymbol{\alpha})} \right\|_{\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda)} \le C \left(\min_{1 \le i \le N} l_i \right)^{3/2-k} \|\varphi\|_{\mathrm{H}^k_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} \qquad \forall\, \varphi \in \mathrm{H}^k_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N).$$

Proof. In Theorem 2.4 of [**Guo00**] the univariate case is proved for φ in certain complicated functional space. Nonetheless, it is easy to prove that the univariate counterpart of $\mathrm{H}^{k+1}_{\rho^{\boldsymbol{\alpha}}}(\Lambda)$ is continuously embedded in that complicated space. Then, the extension to the multivariate case is a simple adaptation of the proof of Theorem 2.2 of [**CQ82**]. □

In order to produce bounds of the form (4.26) for projections on spaces of products of polynomials and the Maxwellian (as opposed to the tailored but in general unknown spectral basis of the $\tilde{e}_{\boldsymbol{n}}$), we need to know the rate of growth of the ratio

$$\left\| \nabla J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \right\|^2_{[\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)]^N} \Bigg/ \left\| J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \right\|^2_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} = \sum_{j=1}^{N} \left[\left\| J_{n_j}^{(\alpha_j)\prime} \right\|^2_{\mathrm{L}^2_{\rho^{\alpha_j}}(\Lambda)} \Bigg/ \left\| J_{n_j}^{(\alpha_j)} \right\|^2_{\mathrm{L}^2_{\rho^{\alpha_j}}(\Lambda)} \right] \tag{4.42}$$

with respect to $\boldsymbol{n}$. Of course, if on the right-hand side expression of (4.42) we were measuring each $J_{n_j}^{(\alpha_j)\prime}$ in the $\mathrm{L}^2_{\rho^{\alpha_j+1}}(\Lambda)$ norm (the 'right' norm), we know from (4.33) that the j-th ratio would be a quadratic function of n_j, as befits the eigenvalues of classic Sturm–Liouville problems such as the one underlying (4.33). However, we are using the $\mathrm{L}^2_{\rho^{\alpha_j}}(\Lambda)$ norm (the 'wrong' norm) instead and we will find another behavior.

Proposition 4.26. *Let $\alpha > -1$ and $n \in \mathbb{N}_0$. Then,*

$$\left\| J_n^{(\alpha)\prime} \right\|^2_{\mathrm{L}^2_{\rho^{\alpha}}(\Lambda)} = \frac{1}{2\alpha+2} n(n+2\alpha+1)(2n+2\alpha+1) \left\| J_n^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^{\alpha}}(\Lambda)}. \tag{4.43}$$

Also, if $\boldsymbol{\alpha} \in (-1,\infty)^N$ and $\boldsymbol{n} \in \mathbb{N}_0^N$,

$$\left\| \nabla J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \right\|^2_{[\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)]^N} = \left[\sum_{j=1}^{N} \frac{1}{2\alpha_j+2} n_j(n_j+2\alpha_j+1)(2n_j+2\alpha_j+1) \right] \left\| J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})} \right\|^2_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}. \tag{4.44}$$

Proof. The identity (4.44), on account of the tensor-product structure of $J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})}$ (cf. (4.37)), is a straightforward consequence of (4.43). So we only need to focus on the latter.

In order to prove the identity (4.43), we will proceed by induction. The case $n = 0$ is immediate and the case $n = 1$ is an easy consequence of (4.35) and the fact that $J_1^{(\alpha)}(t) = (\alpha + 1)t$, for $t \in \Lambda$. Let now $n \geq 2$ and let us suppose that the identity (4.43) is valid for $n - 1$; we shall deduce that it is then valid for $n + 1$. Firstly, we use (4.35) and the fact that $z^{-1}\Gamma(z + 1) = \Gamma(z)$ if $z > 0$ to produce

$$\left\| J_n^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = (n+1)\frac{(2n+2\alpha+3)(n+2\alpha+1)}{(2n+2\alpha+1)(n+\alpha+1)^2} \left\| J_{n+1}^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} \tag{4.45}$$

and

$$\left\| J_{n-1}^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = n(n+1)\frac{(n+2\alpha+1)(n+2\alpha)(2n+2\alpha+3)}{(n+\alpha+1)^2(2n+2\alpha-1)(n+\alpha)^2} \left\| J_{n+1}^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)}. \tag{4.46}$$

Then, taking the squared $\mathrm{L}^2_{\rho^\alpha}(\Lambda)$ norm of (4.36) and using the crucial fact that $J_n^{(\alpha)}$ is $\mathrm{L}^2_{\rho^\alpha}(\Lambda)$-orthogonal to ${J_{n-1}^{(\alpha)}}'$ (for the latter is a polynomial of degree at most $n-2$), we obtain

$$\frac{(n+2\alpha+1)^2}{(n+\alpha+1)^2} \left\| {J_{n+1}^{(\alpha)}}' \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = (2n+2\alpha+1)^2 \left\| J_n^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} + \frac{(n+\alpha)^2}{(n+2\alpha)^2} \left\| {J_{n-1}^{(\alpha)}}' \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)}. \tag{4.47}$$

As (4.43) is supposed to hold for $n - 1$,

$$\left\| {J_{n-1}^{(\alpha)}}' \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = \frac{1}{2\alpha+2}(n-1)(n+2\alpha)(2n+2\alpha-1) \left\| J_{n-1}^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)},$$

which, substituted into (4.47), together with (4.45) and (4.46), produces, after some elementary algebraic manipulations,

$$\left\| {J_{n+1}^{(\alpha)}}' \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)} = \frac{1}{2\alpha+1}(n+1)(n+2\alpha+2)(2n+2\alpha+3) \left\| J_{n+1}^{(\alpha)} \right\|^2_{\mathrm{L}^2_{\rho^\alpha}(\Lambda)}, \tag{4.48}$$

and hence (4.43). □

We will also need the following auxiliary result.

Proposition 4.27. *Let* $\tilde{\mathrm{T}}^{(k)} = \left(\tilde{\tau}_{\boldsymbol{n}}^{(k)} : \boldsymbol{n} \in \mathbb{N}^N \right)$ *and* $\tilde{\Upsilon}^{(k)} = \left(\tilde{\upsilon}_{\boldsymbol{n}}^{(k)} : \boldsymbol{n} \in \mathbb{N}^N \right)$ *be defined according to*

$$\tilde{\tau}_{\boldsymbol{n}}^{(k)} := \left(\sum_{i=1}^N n_i^2 \right)^k + \prod_{i=1}^N \left(n_i^2 \right)^k \qquad \text{and} \qquad \tilde{\upsilon}_{\boldsymbol{n}}^{(k)} := \left(\sum_{i=1}^N n_i^2 \right)^k. \tag{4.49}$$

Then,

$$k > 2 \implies \sum_{\boldsymbol{n} \in \mathbb{N}^N} \frac{\sum_{i=1}^N n_i^3}{\tilde{\tau}_{\boldsymbol{n}}^{(k)}} < \infty \tag{4.50}$$

and

$$k > \frac{3}{2} + \frac{N}{2} \implies \sum_{\boldsymbol{n} \in \mathbb{N}^N} \frac{\sum_{i=1}^N n_i^3}{\tilde{\upsilon}_{\boldsymbol{n}}^{(k)}} < \infty \tag{4.51}$$

Proof. The proof of (4.50) is a modification of the proof of Theorem 3.18. First, using a multivariate version of the integral test for convergence, the finiteness of the sum in (4.50) can be seen to be equivalent to the finiteness of the integral

$$\int_{[1,\infty)^N} \frac{\sum_{i=1}^N x_i^3}{\left(\sum_{i=1}^N x_i^2\right)^k + \prod_{i=1}^N \left(x_i^2\right)^k} \,\mathrm{d}x.$$

Using the fact that, in Euclidean spaces, $\|\cdot\|_3^3 \leq \|\cdot\|_2^3$ and the change of variable (3.43) with parameter $d = 1$ (resulting on an hyperspherical system of coordinates), the finiteness of the above integral is implied by the finiteness of the integral

$$\int_0^{\pi/2} \cdots \int_0^{\pi/2} \int_1^{\infty} \frac{r^3 \left[r^{N-1} \prod_{i=1}^{N-1} \sin(\phi_i)^{N-1-i}\right]}{r^{2k} + r^{2kN} \prod_{i=1}^{N-1} \left[\cos(\phi_i)^{2k} \sin(\phi_i)^{(N-i)2k}\right]} \,\mathrm{d}r \,\mathrm{d}\phi_{N-1} \cdots \,\mathrm{d}\phi_1,$$

which, given $p \in (1,\infty)$ with q being its Hölder conjugate, can be bounded, via an application of Young's inequality, by

$$p^{-1/p} q^{-1/q} \int_0^{\pi/2} \cdots \int_0^{\pi/2} \int_1^{\infty} \frac{r^{N+2} \prod_{i=1}^{N-1} \sin(\phi_i)^{N-1-i}}{r^{2k\left(\frac{1}{p}+\frac{N}{q}\right)} \prod_{i=1}^{N-1} \left[\cos(\phi_i)^{\frac{2k}{q}} \sin(\phi_i)^{\frac{(N-i)2k}{q}}\right]} \,\mathrm{d}r \,\mathrm{d}\phi_{N-1} \cdots \,\mathrm{d}\phi_1.$$

Continuing the analogy with the proof of Theorem 3.18, the above integral can be cast as a product of N univariate integrals with respect to r and each of the ϕ_i, $i \in [N-1]$, which will be finite if

$$N + 2 - 2k\left(\frac{1}{p} + \frac{N}{q}\right) < -1$$

and, for $i \in [N-1]$,

$$-\frac{2k}{q} > -1 \qquad \text{and} \qquad N - 1 - i - \frac{(N-i)2k}{q} > -1.$$

Using the assumption that $k > 2$, these conditions reduce to

$$p > \frac{2k(1-N)}{3 + N(1-2k)} \qquad \text{and} \qquad p < \frac{2k}{2k-1},$$

which, it is easy to show, can be satisfied simultaneously under the assumed restriction $k > 2$. This results in the finiteness of the above integrals and, ultimately, of the sum (4.50).

The proof of (4.51) is very similar, but simpler; so we omit it. □

Remark 4.28. When $d = 1$, the families of nets of the forms $\mathrm{T}^{(k)}$ and $\Upsilon^{(k)}$ defined in (3.40) and (3.48) are almost completely analogous to the families of the forms $\tilde{\mathrm{T}}^{(k)}$ and $\tilde{\Upsilon}^{(k)}$ defined in (4.49). The reason for introducing the new families is that the old involve the eigenvalues of the single-spring problems (3.33). Keeping those eigenvalues in the definition of the nets that appear in this subsection, where we are studying the effect on the Discrete Greedy Algorithm (Algorithm VI) of using bases other than those of the eigenfunctions of (3.33), would be misleading.

Having proved a number of auxiliary results, we will now describe how the results on the domain Λ^N and Jacobi weights for the form $\rho^{\boldsymbol{\alpha}}$ translate into the domain D and the Maxwellian weight M. So, let the matrix $B \in \mathbb{R}^{N\times N}$ be defined by $B_{ij} := \delta_{ij}\sqrt{b_i}$. The operation $\boldsymbol{s} \mapsto B\boldsymbol{s}$ defines a bijective map between $\Lambda^N = (-1,1)^N$ and $\mathsf{D} = \times_{i\in[N]}(-\sqrt{b_i},\sqrt{b_i})$.

A consequence of the adoption of Hypothesis E is that, for $i \in [N]$, there exist constants $c_1^{(i)}$ and $c_2^{(i)}$ such that

$$c_1^{(i)}\rho\big(\boldsymbol{q}_i/\sqrt{b_i}\big)^{\alpha_i} \le M_i(\boldsymbol{q}_i) \le c_2^{(i)}\rho\big(\boldsymbol{q}_i/\sqrt{b_i}\big)^{\alpha_i} \qquad \forall\, \boldsymbol{q}_i \in D_i = B\big(0,\sqrt{b_i}\big) = \big(-\sqrt{b_i},\sqrt{b_i}\big),$$

for some $\alpha_i > 1$. Note that this is valid for the whole of each domain D_i. Collecting the α_i in $\boldsymbol{\alpha} := (\alpha_1,\dots,\alpha_N) \in (1,\infty)^N$, we get that for $c_1 = \prod_{i\in[N]} c_1^{(i)} > 0$ and $c_2 = \prod_{i\in[N]} c_2^{(i)} > 0$,

$$c_1\rho(B^{-1}\boldsymbol{q})^{\boldsymbol{\alpha}} \le \mathsf{M}(\boldsymbol{q}) \le c_2\rho(B^{-1}\boldsymbol{q})^{\boldsymbol{\alpha}} \qquad \forall\, \boldsymbol{q} \in \mathsf{D}, \tag{4.52}$$

This informs the definitions

$$\varphi_n^{(i)} := M_i\, J_n^{(\alpha_i)}(\cdot/\sqrt{b_i}) \qquad \forall\,(i,n) \in [N]\times\mathbb{N}_0$$

and

$$\varphi_{\boldsymbol{n}} := \bigotimes_{i=1}^{N} \varphi_n^{(i)} = \mathsf{M}\, J_{\boldsymbol{n}}^{(\boldsymbol{\alpha})}(B^{-1}\cdot) \qquad \forall\, \boldsymbol{n} \in \mathbb{N}_0^N. \tag{4.53}$$

The $\varphi_n^{(i)}$ are products of a partial Maxwellian and an univariate polynomial of degree n and the $\varphi_{\boldsymbol{n}}$ are products of the full Maxwellian and a multivariate polynomial of (composite) degree $\boldsymbol{n}$. So, we have that the $\hat{\mathrm{H}}_l^{(i)}$ and $\hat{\mathrm{H}}_{\boldsymbol{l}}$, as defined in (4.30) and (4.31), are precisely

$$\hat{\mathrm{H}}_l^{(i)} = \operatorname{span}\Big(\varphi_n^{(i)} \colon n \in \{0,\dots,l\}\Big) \qquad \forall\,(i,l) \in [N]\times\mathbb{N}_0 \tag{4.54}$$

and

$$\hat{\mathrm{H}}_{\boldsymbol{l}} = \operatorname{span}\left(\varphi_{\boldsymbol{n}} \colon \boldsymbol{n} \in \bigtimes_{i=1}^{N}\{0,\dots,l_i\}\right) \qquad \forall\, \boldsymbol{l} \in \mathbb{N}_0^N. \tag{4.55}$$

We can now give the principal result of this subsection.

Theorem 4.29. *Let ψ be the true solution to the Fokker–Planck equation* (2.5) *and let $\hat{\mathrm{H}}_{\boldsymbol{l}}$ be fixed according to* (4.31) *(or, equivalently,* (4.55)*). If $\hat{\delta}_n$ is the error committed at the end of the n-th iteration of the Discrete Orthogonal Greedy Algorithm based on $\hat{\mathrm{H}}_{\boldsymbol{l}}$ (Algorithm VI), we have*

$$\mathsf{M}^{-1}\psi \in \mathrm{H}_{\mathsf{M}}^{k,\mathrm{mix}}(\mathsf{D}) \quad \textit{with} \quad k > 2$$
$$\implies \left\|\hat{\delta}_n\right\|_a \le C_1 \left\|\mathsf{M}^{-1}\psi\right\|_{\mathrm{H}_{\mathsf{M}}^{k,\mathrm{mix}}(\mathsf{D})} \left(n^{-1/2} + \left(\min_{1\le i\le N} l_i\right)^{3/2-k}\right) \tag{4.56}$$

for some $C_1 > 0$ independent of ψ, and

$$\mathsf{M}^{-1}\psi \in \mathrm{H}^k_{\mathsf{M}}(\mathsf{D}) \quad \textit{with} \quad k > \frac{3}{2} + \frac{N}{2}$$
$$\implies \left\|\hat{\delta}_n\right\|_a \leq C_2 \left\|\mathsf{M}^{-1}\psi\right\|_{\mathrm{H}^k_{\mathsf{M}}(\mathsf{D})} \left(n^{-1/2} + \left(\min_{1\leq i\leq N} l_i\right)^{3/2-k}\right) \tag{4.57}$$

for some $C_2 > 0$ independent of ψ.

Proof. We start with the proof of (4.56). So, we assume the left-hand expression of (4.56) and start by letting $\mathcal{B}\colon \mathrm{H}(\mathsf{D};\mathsf{M}) \to \mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$ be defined by

$$\mathcal{B}(\xi) := \frac{\xi(B\boldsymbol{\cdot})}{\mathsf{M}(B\boldsymbol{\cdot})} \qquad \forall\, \xi \in \mathrm{H}(\mathsf{D};\mathsf{M}).$$

It is easy to see, based of the bounds (4.52), that $\mathcal{B}$ is well-defined, linear and continuous operator with continuous inverse. With this definition and (4.52) we have,

$$\|\mathcal{B}(\xi)\|_{\mathrm{H}^{k,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} \leq C_3 \left\|\mathsf{M}^{-1}\xi\right\|_{\mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})} \qquad \forall\, \xi \in \mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}),$$

which, by straightforward embeddings, gives

$$\|\mathcal{B}(\xi)\|_{\mathrm{H}^{k}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} \leq C_3 \left\|\mathsf{M}^{-1}\xi\right\|_{\mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})} \qquad \forall\, \xi \in \mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}) \tag{4.58}$$

and

$$\|\mathcal{B}(\xi)\|_{\mathrm{V}^{k,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} \leq C_3 \left\|\mathsf{M}^{-1}\xi\right\|_{\mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})} \qquad \forall\, \xi \in \mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D}), \tag{4.59}$$

which come about easily under the observation that the action of $\mathcal{B}$ can be seen as the composition of the simple operations consisting of division by M and a rescaling of the argument.

Now, note that $\mathcal{B}^{-1}(\eta) = \mathsf{M}\,\eta(B^{-1}\boldsymbol{\cdot})$ for all $\eta \in \mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda)$; so, given $\boldsymbol{l} \in \mathbb{N}_0^N$, the operator

$$\tilde{P}_{\boldsymbol{l}} := \mathcal{B}^{-1} \circ P^{(\boldsymbol{\alpha})}_{\boldsymbol{l}} \circ \mathcal{B}$$

is a projector from $\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda)$ to $\hat{\mathrm{H}}_{\boldsymbol{l}} = \mathsf{M}\,\mathbb{P}_{\boldsymbol{l}}$. Further, a direct calculation makes it clear that the equality in (4.53) can be restated as

$$\mathcal{B}^{-1}\left(J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right) = \varphi_{\boldsymbol{n}} \qquad \forall\, \boldsymbol{n} \in \mathbb{N}_0^N. \tag{4.60}$$

Using the continuity of $\mathcal{B}$, the continuity of its inverse, the convergence result in Lemma 4.25 and (4.58), we have the approximability estimate

$$\begin{aligned}
\left\|\psi - \tilde{P}_{\boldsymbol{l}}(\psi)\right\|_a &\leq C_4 \left\|\psi - \tilde{P}_{\boldsymbol{l}}(\psi)\right\|_{\mathrm{H}(\mathsf{D};\mathsf{M})} \\
&\leq C_5 \left\|\mathcal{B}(\psi) - (\mathcal{B}\circ\tilde{P}_{\boldsymbol{l}})(\psi)\right\|_{\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} \\
&= C_5 \left\|\mathcal{B}(\psi) - P^{(\boldsymbol{\alpha})}_{\boldsymbol{l}}(B(\psi))\right\|_{\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}
\end{aligned} \tag{4.61}$$

$$\leq C_6 \left(\min_{1\leq i\leq N} l_i \right)^{3/2-k} \|\mathcal{B}(\psi)\|_{\mathrm{H}^k_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}$$
$$\leq C_7 \left(\min_{1\leq i\leq N} l_i \right)^{3/2-k} \left\|M^{-1}\psi\right\|_{\mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})}.$$

From Proposition 4.26 we have, for some $C_8 > 0$,

$$\frac{\left\|J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right\|^2_{\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}}{\left\|J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right\|^2_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \leq C_8 \sum_{i=1}^N (n_i+1)^3. \tag{4.62}$$

This is particularly easy to see if one uses the fact that $\boldsymbol{\alpha} \in (1,\infty)^N$ now (ultimately, because of Hypothesis E; this should be contrasted with the looser condition $\boldsymbol{\alpha} \in (-1,\infty)$ of some other results in this subsection).

Using the definition of the projector $P^{(\boldsymbol{\alpha})}_{\boldsymbol{l}}$ in (4.41), the identity (4.60), the estimate (4.62), part (2) of Lemma 4.23—which is where we use the restriction on k—and (4.50) of Proposition 4.27, we have a first $\mathcal{A}_{1,\boldsymbol{l}}$-stability estimate:

$$\begin{aligned}
\left\|\tilde{P}_{\boldsymbol{l}}(\psi)\right\|_{\mathcal{A}_{1,\boldsymbol{l}}} &= \left\| \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \frac{\langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}}{\langle J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}, J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \mathcal{B}^{-1}\left(J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right) \right\|_{\mathcal{A}_{1,\boldsymbol{l}}} \\
&= \left\| \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \frac{\langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}}{\langle J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}, J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \varphi_{\boldsymbol{n}} \right\|_{\mathcal{A}_{1,\boldsymbol{l}}} \\
&= \left\| \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \frac{\langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}}{\langle J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}, J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \|\varphi_{\boldsymbol{n}}\|_a \frac{\varphi_{\boldsymbol{n}}}{\|\varphi_{\boldsymbol{n}}\|_a} \right\|_{\mathcal{A}_{1,\boldsymbol{l}}} \\
&\leq \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \frac{\left|\langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}\right|}{\langle J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}, J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \|\varphi_{\boldsymbol{n}}\|_a \\
&\leq C_9 \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \frac{\left|\langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}\right|}{\left\|J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right\|_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \frac{\left\|J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right\|_{\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}}{\left\|J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right\|_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \\
&\leq C_{10} \sum_{\boldsymbol{n}\leq\boldsymbol{l}} \frac{\left|\langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}\right|}{\left\|J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}}\right\|_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \left[\sum_{i=1}^N (n_i+1)^3\right]^{1/2}
\end{aligned}$$

$$= C_{10} \sum_{\boldsymbol{n} \le \boldsymbol{l}} \left(\tilde{\tau}^{(k)}_{\boldsymbol{n}+\mathbf{1}} \right)^{1/2} \frac{\left| \langle \mathcal{B}(\psi), J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \rangle_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} \right|}{\left\| J^{(\boldsymbol{\alpha})}_{\boldsymbol{n}} \right\|_{\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)}} \frac{\left[\sum_{i=1}^N (n_i+1)^3 \right]^{1/2}}{\left(\tilde{\tau}^{(k)}_{\boldsymbol{n}+\mathbf{1}} \right)^{1/2}}$$

$$\le C_{11} \left\| \mathcal{B}(\psi) \right\|_{\mathrm{V}^{k,\mathrm{mix}}_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)} .$$

Here, we have used $\mathbf{1}$ to signify the multi-index of length N which has all of its entries equal to 1. Using (4.59), this turns into

$$\left\| \tilde{P}_{\boldsymbol{l}}(\psi) \right\|_{\mathcal{A}_{1,\boldsymbol{l}}} \le C_{12} \left\| \mathsf{M}^{-1} \psi \right\|_{\mathrm{H}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})} . \tag{4.63}$$

The approximability result (4.61) and the $\mathcal{A}_{1,\boldsymbol{l}}$-stability estimate (4.63) ensure (according to the definition of $\mathcal{Z}_\theta$ in (4.21)) that

$$\psi \in \mathcal{Z}_\theta \qquad \text{with} \qquad \theta(\boldsymbol{l}) := \left(\min_{1 \le i \le N} l_i \right)^{3/2-k} .$$

Then, Corollary 4.16 gives (4.56), the first of the desired results. The second of these, (4.57), is very similar, so we omit it. □

Remark 4.30. The main result of this subsection, Theorem 4.29, is given in terms of membership of $\mathsf{M}^{-1}\psi$ in certain Maxwellian-weighted Sobolev spaces. This should be contrasted with the main result of the previous subsection, Theorem 4.20, which is given in terms of membership of ψ in spaces of the form $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$ (that is, spaces described by weighted summability of Fourier coefficients; cf. (4.24)).

Now, as indicated in the first point of Remark 4.21, it is possible to give at least a partial characterization of the relevant $\mathrm{H}^{\Sigma}(\mathsf{D};\mathsf{M})$ spaces in terms of regularity requirements. Namely, for $k \in \{2,3,4\}$,

$$\mathrm{H}^{\mathrm{T}^{(k)}}(\mathsf{D};\mathsf{M}) \supseteq \mathsf{M}\, \tilde{\mathrm{H}}^{k,\mathrm{mix}}_{\mathsf{M}}(\mathsf{D})$$

(see (3.78), (3.79) and Lemma 3.29 for the cases $k = 2$ and $k = 4$ and (3.82) and (3.83) for the case $k = 3$; similar things can be proved for nets of the family $\Upsilon^{(k)}$). Therefore, Theorem 4.20 can, in some cases, be put in a form more in line with that of Theorem 4.29. This indicates that comparing Theorem 4.20 and Theorem 4.29 with the same parameter k is fair.

However, we cannot recast the regularity requirements of Theorem 4.29 in terms of summability of weighted squared coefficients of expansions with respect to some Hilbert basis (thus putting it in a form more in line with that of Theorem 4.20). The reason is that Lemma 4.25, on which Theorem 4.29 depends, has a regularity requirement as a hypothesis.

Another difference is that, compared with the approximation rate given in Lemma 4.18 (which feeds into Theorem 4.20) in the case $d = 1$, Lemma 4.25 (which, in turn, feeds into Theorem 4.29) has lost one half of a power of

$$\min_{1 \le i \le N} l_i$$

in decay rate with respect to $\boldsymbol{l}$. The reason behind this discrepancy is that, on the one hand, the truncation operators $P_{\boldsymbol{l}}$ defined in (4.23) are at the same time $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$- and $\mathrm{H}(\mathsf{D};\mathsf{M})$-orthogonal projectors, while, on the other hand the operators $P^{(\boldsymbol{\alpha})}_{\boldsymbol{l}}$ described in (4.41) are $\mathrm{L}^2_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$- and $\mathrm{V}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$-orthogonal projectors, but not $\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda^N)$-orthogonal projectors.

It is possible to define projector families with better approximation rates in $\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda)$ than the $P^{(\boldsymbol{\alpha})}_{\boldsymbol{l}}$ (see for example, [**GW04**, Lemma 2.2]). However, the relatively simple explicit form we have of the $P^{(\boldsymbol{\alpha})}_{\boldsymbol{l}}$ was for us essential in producing the necessary $\mathcal{A}_{1,\boldsymbol{l}}$-stability estimate (4.63) (for the Maxwellian-times-polynomials subspaces) using the template given by the derivation of the $\mathcal{A}_{1,\boldsymbol{l}}$-stability estimate (4.26) (for the span-of-eigenfunctions subspaces).

Perhaps the most important of the differences between Theorem 4.20 and Theorem 4.29 is that the latter requires a higher lower limit on the regularity level (in the notation of both results, a higher k). This—together with the difference in the approximability rates discussed in the previous paragraph—is a consequence of using the transformed eigenfunctions of the 'wrong' eigenvalue problem, (4.33), instead of the eigenfunctions of the relevant Fokker–Planck operators.

The very same reason that kept us from using the projectors with better approximation rates in $\mathrm{H}^1_{\rho^{\boldsymbol{\alpha}}}(\Lambda)$—the difficulty of extending $\mathcal{A}_{1,\boldsymbol{l}}$-stability results—has precluded adapting the analysis of this subsection to cases with $d \geq 2$. Indeed, when dealing with spherical domains—which are not true tensor-product domains—issues such as pole conditions (see, for example, [**Boy01**, Chapter 18]) need to be adressed and exacerbate the difficulty of studying the consequences of using bases other than the ones based on the eigenfunctions of the Fokker–Planck operators. It should be noted, however, that approximability results for operators with Maxwellian weights defined on discs and spheres are available (cf. [**KS09b**, Section 5]).

One aspect in which Theorem 4.20 and Theorem 4.29 have the same behavior is that the rate of convergence of the Discrete Greedy Algorithm (Algorithm VI) stays the same with respect to the iteration number n whatever the regularity level k is.

4.2. Numerical implementation

In this section we will comment on the computational behavior of an implementation of the Discrete Orthogonal Greedy Algorithm (Algorithm VI). Between the algorithm as it is described in Section 4.1 and the implementation we will describe there are some important differences, which we shall describe in the following subsection.

4.2.1. Gaps between theory and computation. The Discrete Orthogonal Greedy Algorithm (Algorithm VI) entails in its step `2.1` the global minimization of the non-convex

functional $\mathcal{J}_g\colon \times_{i\in[N]} \hat{\mathrm{H}}^{(i)}_{l_i} \to \mathbb{R}$ defined by

$$\mathcal{J}_g(s^{(1)},\dots,s^{(N)}) := \frac{1}{2}a\left(\bigotimes_{i=1}^{N} s^{(i)}, \bigotimes_{i=1}^{N} s^{(i)}\right) - g\left(\bigotimes_{i=1}^{N} s^{(i)}\right) \qquad \forall\,(s^{(1)},\dots,s^{(N)}) \in \bigtimes_{i=1}^{N} \hat{\mathrm{H}}^{(i)}_{l_i},$$

where g is some functional in $\mathrm{H}(\mathsf{D};\mathsf{M})'$. Indeed, at the n-th iteration of Algorithm VI, $g = \hat{f}_{n-1}$.

The domain of $\mathcal{J}_g$ is a $|\boldsymbol{l}|_1$-dimensional space. Under the plausible assumption of all the l_i being equal to a common $l \in \mathbb{N}$, this results in a space of dimension Nl, which grows very mildly with respect to N, the number of factor domains which constitute the domain $\mathsf{D} = D_1 \times \cdots \times D_N$, as compared with l^N, the dimension of $\hat{\mathrm{H}}_{\boldsymbol{l}}$. As it is, Algorithm VI simply assumes this global minimization is somehow performed exactly. By contrast, our computational implementation relies on approximating solutions to the associated Euler–Lagrange equation; this is the first gap between theory and computation.

However, the Euler–Lagrange equations associated with the minimization of $\mathcal{J}_g$ over $\times_{i=1}^{N} \hat{\mathrm{H}}^{(i)}_{l_i}$ can lend themselves very well to an alternating direction scheme—indeed, this, together with the relatively small dimensionality of $\times_{i\in[N]} \hat{\mathrm{H}}^{(i)}_{l_i}$, is the rationale behind the Separated Representation strategy. Therefore, at each iteration of the loop 2 of Algorithm VI, we will approximate a solution to the corresponding Euler–Lagrange equation using an inner iteration. We will describe it below in Subsection 4.2.2.

We also assume that the right-hand side g can be written as the finite sum of tensor products of $\mathrm{L}^2_{1/M_i}(D_i)$ functions. That is, we assume that g has the form

$$g = \sum_{k=1}^{N_g} \bigotimes_{i=1}^{N} g^{(i)}_k, \tag{4.64}$$

where $N_g \in \mathbb{N}$ and, for $(i,k) \in [N]\times[N_g]$, $g^{(i)}_k \in \mathrm{L}^2_{1/M_i}(D_i)$—we make the usual identification between $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ and its dual, which is a subset of the dual of $\mathrm{H}(\mathsf{D};\mathsf{M})$[1]. Recalling that g is a placeholder for $\hat{f}_{n-1}$, it transpires that (4.64) will hold if the right-hand side functional f of (2.5) is itself of the form (4.64). The assumption of (4.64) is a second gap between theory and computation.

Remark 4.31. In practice, if the right-hand side g fails to have the structure given in (4.64), a preliminary greedy procedure can be applied to it in order to approximate it with a finite sum of tensor products. We refer the interested reader to the surveys [**CEK+07**] and [**KB09**] in the discrete—i.e., array decomposition—case, which, in turn, informs procedures for the continuous—i.e., decomposition of functions defined on a Cartesian product of continua—case (see, e.g., [**HK07**]).

[1] If we merely assumed that each $g^{(i)}_k$ was a member of the larger space of functionals $\mathrm{H}(D_i;M_i)'$, there would be no guarantee of g being a member of $\mathrm{H}(\mathsf{D};\mathsf{M})'$.

Another difference between the Algorithm VI and its implementation has to do with the termination condition given in its step `2.4`; that is, $\|\hat{f}_n|_{\hat{\mathrm{H}}_{\boldsymbol{l}}}\|_{\hat{\mathrm{H}}'_{\boldsymbol{l}}} \geq \mathtt{TOL}$ for some tolerance $\mathtt{TOL} > 0$. The $\mathrm{H}(\mathsf{D};\mathsf{M})$-derived norm of $\hat{\mathrm{H}}'_{\boldsymbol{l}}$,

$$\|h\|_{\hat{\mathrm{H}}'_{\boldsymbol{l}}} := \sup_{\varphi \in \hat{\mathrm{H}}_{\boldsymbol{l}} \setminus \{0\}} \frac{|h(\varphi)|}{\|h\|_{\mathrm{H}(\mathsf{D};\mathsf{M})}} \qquad \forall\, h \in \hat{\mathrm{H}}'_{\boldsymbol{l}},$$

is hard to evaluate in general. What we do in our computations is simply setting the number of iterations in advance.

At this stage we choose to restrict ourselves to the $d = 1$ case, in which D is the Cartesian product of intervals, just as we did in Subsection 4.1.6. This greatly simplifies the exposition that follows and the implementation; however, it also precludes the numerical experiments from simulating situations of rheological interest.

4.2.2. Inner iteration. In order to describe the alternating direction scheme mentioned above we need to study the effect of the bilinear form a on functions with the tensor-product structure. It is easy to see that, given indices k and l in $[N]$ and ensembles $(u^{(1)}, \ldots, u^{(N)})$ and $(v^{(1)}, \ldots, v^{(N)})$ in $\times_{i\in[N]} \mathrm{H}^1_{M_i}(D_i)$, the Cartesian-product structure of the domain D and the tensor-product structure of the Maxwellian weight M (cf. (1.25)) give

$$\int_{\mathsf{D}} \left(\bigotimes_{i=1}^{N} u^{(i)}\right) \left(\bigotimes_{i=1}^{N} v^{(i)}\right) \mathsf{M} = \prod_{i=1}^{N} \int_{D_i} u^{(i)} v^{(i)} M_i, \tag{4.65a}$$

$$\int_{\mathsf{D}} \nabla_{\boldsymbol{q}_k} \left(\bigotimes_{i=1}^{N} u^{(i)}\right) \cdot \nabla_{\boldsymbol{q}_k} \left(\bigotimes_{i=1}^{N} v^{(i)}\right) \mathsf{M} = \left(\prod_{i\in[N]\setminus\{k\}} \int_{D_i} u^{(i)} v^{(i)} M_i\right) \int_{D_k} u^{(k)\prime} v^{(k)\prime} M_k \tag{4.65b}$$

and, if $k \neq l$,

$$\begin{aligned} \int_{\mathsf{D}} \nabla_{\boldsymbol{q}_l} \left(\bigotimes_{i=1}^{N} u^{(i)}\right) \cdot \nabla_{\boldsymbol{q}_k} \left(\bigotimes_{i=1}^{N} v^{(i)}\right) \mathsf{M} \\ = \left(\prod_{i\in[N]\setminus\{k,l\}} \int_{D_i} u^{(i)} v^{(i)} M_i\right) \left(\int_{D_l} u^{(l)\prime} v^{(l)} M_l\right) \int_{D_k} u^{(k)} v^{(k)\prime} M_k. \end{aligned} \tag{4.65c}$$

Note how the assumption of $d = 1$ has turned gradients into derivatives. Using (4.65) and the fact that, for $i \in [N]$, $\mathrm{H}(D_i; M_i) = M_i \,\mathrm{H}^1_{M_i}(D_i)$ (cf. (2.4)), we have that for all ensembles $(u^{(1)}, \ldots, u^{(N)})$ and $(v^{(1)}, \ldots, v^{(N)})$ in $\times_{i\in[N]} \mathrm{H}(D_i; M_i)$,

$$a\left(\bigotimes_{i=1}^{N} u^{(i)}, \bigotimes_{i=1}^{N} v^{(i)}\right)$$

$$= \sum_{k,l=1}^{N} \frac{A_{k,l}}{4\mathrm{Wi}} \begin{cases} \left(\prod_{i\in[N]\setminus\{k\}} I_i(u^{(i)}, v^{(i)}) \right) K_k(u^{(k)}, v^{(k)}) & \text{if } k = l \\ \left(\prod_{i\in[N]\setminus\{k,l\}} I_i(u^{(i)}, v^{(i)}) \right) \tilde{T}_l(u^{(l)}, v^{(l)})\, T_k(u^{(k)}, v^{(k)}) & \text{if } k \neq l \end{cases}$$

$$+ c \prod_{i=1}^{N} I_i(u^{(i)}, v^{(i)}) \quad (4.66)$$

where, for $i \in [N]$, I_i, K_i, T_i and $\tilde{T}_i$ are the bilinear forms in $\mathrm{H}(D_i; M_i) \times \mathrm{H}(D_i; M_i) \to \mathbb{R}$ defined by

$$\begin{aligned} I_i(\sigma, \tau) &= \langle \sigma, \tau \rangle_{\mathrm{L}^2_{1/M_i}(D_i)}, & K_i(\sigma, \tau) &= \int_{D_i} \left(\frac{\sigma}{M_i}\right)' \left(\frac{\tau}{M_i}\right)' M_i, \\ T_i(\sigma, \tau) &= \int_{D_i} \sigma \left(\frac{\tau}{M_i}\right)', \qquad \text{and} & \tilde{T}_i(\sigma, \tau) &= \int_{D_i} \left(\frac{\sigma}{M_i}\right)' \tau \end{aligned} \quad (4.67)$$

for all σ and τ in $\mathrm{H}(D_i; M_i)$. They are, respectively, the $\mathrm{L}^2_{1/M_i}(D_i)$ inner product, a stiffness bilinear form, an advection bilinear form and its transpose. Similarly, from the assumption of (4.64) we have that for all ensembles $(v^{(1)}, \ldots, v^{(N)}) \in \times_{i\in[N]} \mathrm{H}(D_i; M_i)$,

$$g\left(\bigotimes_{i=1}^{N} v^{(i)}\right) = \sum_{k=1}^{N_g} \prod_{i=1}^{N} I_i(g_k^{(i)}, v^{(i)}). \quad (4.68)$$

If some ensemble $(r^{(1)}, \ldots, r^{(N)}) \in \times_{i\in[N]} \hat{\mathrm{H}}^{(i)}_{l_i}$ minimizes $\mathcal{J}_f$, $\bigotimes_{i\in[N]} r^{(i)}$ minimizes J_f in $\bigotimes_{i=1}^{N} \hat{\mathrm{H}}^{(i)}_{l_i}$. Then, from Lemma 4.5, the ensemble of the $r^{(i)}$ satisfies the Euler–Lagrange equation

$$a\left(\bigotimes_{i=1}^{N} r^{(i)}, \sum_{j=1}^{N} \bigotimes_{\substack{i=1 \\ i\neq j}}^{N} r^{(i)} \otimes_j s^{(j)}\right) = g\left(\sum_{j=1}^{N} \bigotimes_{\substack{i=1 \\ i\neq j}}^{N} r^{(i)} \otimes_j s^{(j)}\right) \quad (4.69)$$

for all test ensembles $(s^{(1)}, \ldots, s^{(N)}) \in \times_{i\in[N]} \hat{\mathrm{H}}^{(i)}_{l_i}$. If for some fixed $m \in [N]$ we consider only test ensembles with $s^{(i)} = 0$ for $i \in [N] \setminus \{m\}$, the above equation reduces to

$$a\left(\bigotimes_{i=1}^{N} r^{(i)}, \bigotimes_{\substack{i=1 \\ i\neq m}}^{N} r^{(i)} \otimes_m s^{(m)}\right) = g\left(\bigotimes_{\substack{i=1 \\ i\neq m}}^{N} r^{(i)} \otimes_m s^{(m)}\right). \quad (4.70)$$

Using (4.66) and (4.68), (4.70) takes the form

$$\sum_{k,l=1}^{N} \frac{A_{k,l}}{4\mathrm{Wi}} \begin{cases} B_1 & \text{if } k = l, m \notin \{k,l\} \\ B_2 & \text{if } k \neq l, m \notin \{k,l\} \\ B_3 & \text{if } k = l = m \\ B_4 & \text{if } k \neq l, m = k \\ B_5 & \text{if } k \neq l, m = l \end{cases} + c\left[\prod_{i\in[N]\setminus\{m\}} I_i(r^{(i)}, r^{(i)})\right] I_m(r^{(m)}, s^{(m)})$$

$$= \sum_{k=1}^{N_g} \left[\prod_{i\in[N]\setminus\{m\}} I_i(g_k^{(i)}, r^{(i)})\right] I_m(g_k^{(m)}, s^{(m)}), \quad (4.71a)$$

where

$$B_1 = \left[\prod_{i\in[N]\setminus\{k,m\}} I_i(r^{(i)}, r^{(i)})\right] K_k(r^{(k)}, r^{(k)}) I_m(r^{(m)}, s^{(m)}), \quad (4.71b)$$

$$B_2 = \left[\prod_{i\in[N]\setminus\{k,l,m\}} I_i(r^{(i)}, r^{(i)})\right] \tilde{T}_l(r^{(l)}, r^{(l)}) T_k(r^{(k)}, r^{(k)}) I_m(r^{(m)}, s^{(m)}), \quad (4.71c)$$

$$B_3 = \left[\prod_{i\in[N]\setminus\{m\}} I_i(r^{(i)}, r^{(i)})\right] K_m(r^{(m)}, s^{(k)}), \quad (4.71d)$$

$$B_4 = \left[\prod_{i\in[N]\setminus\{l,m\}} I_i(r^{(i)}, r^{(i)})\right] \tilde{T}_l(r^{(l)}, r^{(l)}) T_m(r^{(m)}, s^{(m)}), \quad (4.71e)$$

$$B_5 = \left[\prod_{i\in[N]\setminus\{k,m\}} I_i(r^{(i)}, r^{(i)})\right] \tilde{T}_m(r^{(m)}, s^{(m)}) T_k(r^{(k)}, r^{(k)}). \quad (4.71f)$$

If for $i \in [N] \setminus \{m\}$, $r^{(i)}$ is fixed, the fact that $s^{(m)}$ is allowed to vary freely in $\hat{\mathrm{H}}_{l_m}^{(m)}$ allows for considering (4.71) a variational problem in the domain D_m with unknown $r^{(m)}$. As all the integrals in (4.71) are on the low dimensional domains D_i, the coefficients can be computed cheaply. We encode this procedure in the algorithm that follows, whose purpose is approximating a solution to the Euler–Lagrange equation (4.69) and takes in the implementation the place the minimization step `2.1` has in Algorithm VI.

Algorithm VII *(Inner Iteration)***.**

`1. Initialize` $(r_0^{(1)}, \ldots, r_0^{(N)}) \in \times_{i\in[N]} \hat{\mathrm{H}}_{l_i}^{(i)}$.

`2. For` $\tilde{n} \geq 1$ `do:`

`2.1 For` $m \in [N]$ `do:`

`2.1.1 For` $i \in [N] \setminus \{m\}$ `set` $r^{(i)} := r_{\tilde{n}-1}^{(i)}$

`2.1.2 Solve for` $r^{(m)}$ `in` (4.71) `and store the result in` $r_{\tilde{n}}^{(m)}$.

`2.2 If`

$$\left\| \bigotimes_{i=1}^{N} r_{\tilde{n}}^{(i)} - \bigotimes_{i=1}^{N} r_{\tilde{n}-1}^{(i)} \right\|_{\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})} \geq \texttt{TOL},$$

`then proceed to the inner iteration` $\tilde{n}+1$`; else, stop and return the ensemble` $(r_{\tilde{n}}^{(1)}, \dots, r_{\tilde{n}}^{(N)})$.

We choose to use the $\mathrm{L}^2_{1/\mathsf{M}}(\mathsf{D})$ norm in the termination test of the inner iteration because it is simple and fast to evaluate.

4.2.3. Implementation notes. We start by noting that, for $i \in [N]$, the nesting property (4.1) allows for defining a correspondingly nested family of bases of the finite dimensional subspaces $\hat{\mathrm{H}}_l^{(i)} \subset \mathrm{H}(D_i; M_i)$. Indeed, it is possible to define a single sequence $\left(\varphi_k^{(i)} \colon k \in \mathbb{N}\right)$ of $\mathrm{H}(D_i; M_i)$ functions such that, for all $l \in \mathbb{N}$, its truncation

$$\Phi_l^{(i)} := \left(\varphi_k^{(i)} \colon k \in [l]\right) \tag{4.72}$$

forms a basis of $\hat{\mathrm{H}}_l^{(i)}$. In particular, the bases described in (4.22) in Subsection 4.1.5 and in (4.30) in Subsection 4.1.6 have this form.

So, for $(i, l) \in [N] \times \mathbb{N}$, let us set $\vec{\cdot} \colon \hat{\mathrm{H}}_l^{(i)} \to \mathbb{R}^l$ as the invertible mapping that to each member of $\hat{\mathrm{H}}_l^{(i)}$ uniquely associates the vector containing its expansion in the basis $\Phi_l^{(i)}$; i.e.,

$$\xi = \sum_{k=1}^{l} \left(\vec{\xi}\right)_k \varphi_k^{(i)} \qquad \forall\, \xi \in \hat{\mathrm{H}}_l^{(i)}. \tag{4.73}$$

Of course, the mapping $\vec{\cdot}$ depends on the spring index i and on the dimension l_i; however, there is no risk of confusion, as long as it is always clear that its argument belongs to an identifiable $\hat{\mathrm{H}}_{l_i}^{(i)}$.

With this in mind, let us now assume $\boldsymbol{l}$ fixed, so we can avoid dimension-indicating subscripts in the matrices and vectors we will define. We introduce, for $i \in [N]$, the Gram matrices $\mathcal{I}^{(i)}$, $\mathcal{K}^{(i)}$, $\mathcal{T}^{(i)}$ and $\tilde{\mathcal{T}}^{(i)}$ in $\mathbb{R}^{l_i \times l_i}$, corresponding, respectively, to the I_i, K_i, T_i and $\tilde{T}_i$ bilinear forms defined in (4.67) with respect to the basis $\Phi_l^{(i)}$ of $\mathrm{H}(D_i; M_i)$; that is, for j and k in $[l_i]$,

$$\begin{aligned} \left(\mathcal{I}^{(i)}\right)_{j,k} &= I_i(\varphi_k^{(i)}, \varphi_j^{(i)}), & \left(\mathcal{K}^{(i)}\right)_{j,k} &= K_i(\varphi_k^{(i)}, \varphi_j^{(i)}), \\ \left(\mathcal{T}^{(i)}\right)_{j,k} &= T_i(\varphi_k^{(i)}, \varphi_j^{(i)}), \qquad \text{and} & \left(\tilde{\mathcal{T}}^{(i)}\right)_{j,k} &= \tilde{T}_i(\varphi_k^{(i)}, \varphi_j^{(i)}). \end{aligned} \tag{4.74}$$

Thus, for example,

$$I^{(i)}(\eta, \xi) = \vec{\xi}^{\mathrm{T}} \mathcal{I}^{(i)} \vec{\eta} \qquad \forall\, (\eta, \xi) \in \hat{\mathrm{H}}_{l_i}^{(i)}.$$

Similar identities hold for the other matrices defined in (4.74). We also define, for $(i, k) \in [N] \times [N_g]$, the load vector $\vec{g}_k^{(i)} \in \mathbb{R}^{l_i}$ by

$$\left(\vec{g}_k^{(i)}\right)_j := I^{(i)}(g_k^{(i)}, \varphi_j^{(i)}) = \langle g_k^{(i)}, \varphi \rangle_{\mathrm{L}^2_{1/M_i}(D_i)} \qquad \forall\, j \in [l_i]; \tag{4.75}$$

whence $I^{(i)}(g_k^{(i)},\xi) = \vec{\xi}^{\,\mathrm{T}}\vec{g}_k^{(i)}$ for all $\xi \in \hat{\mathrm{H}}_{l_i}^{(i)}$.

With all the introduced notation, the step `2.1.2` of the inner iteration (Algorithm VII) takes the form: $\vec{r}^{(i)}$ being fixed for $i \in [N]\setminus\{m\}$, solve for $\vec{r}^{(m)}$ in

$$\sum_{k,l=1}^{N} \frac{A_{k,l}}{4\mathrm{Wi}} \begin{cases} B_1 & \text{if } k=l, m\notin\{k,l\} \\ B_2 & \text{if } k\neq l, m\notin\{k,l\} \\ B_3 & \text{if } k=l=m \\ B_4 & \text{if } k\neq l, m=k \\ B_5 & \text{if } k\neq l, m=l \end{cases} + c\left[\prod_{i\in[N]\setminus\{m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{r}^{(i)}\right]\mathcal{I}^{(m)}\vec{r}^{(m)}$$

$$= \sum_{k=1}^{N_g}\left[\prod_{i\in[N]\setminus\{m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{g}_k^{(i)}\right]\mathcal{I}^{(m)}\vec{g}_k^{(m)}, \quad (4.76a)$$

where

$$B_1 = \left[\prod_{i\in[N]\setminus\{k,m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{r}^{(i)}\right]\left(\vec{r}^{(k)\,\mathrm{T}}\mathcal{K}^{(k)}\vec{r}^{(k)}\right)\mathcal{I}^{(m)}\vec{r}^{(m)}, \quad (4.76b)$$

$$B_2 = \left[\prod_{i\in[N]\setminus\{k,l,m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{r}^{(i)}\right]\left(\vec{r}^{(l)\,\mathrm{T}}\tilde{\mathcal{T}}^{(l)}\vec{r}^{(l)}\right)\left(\vec{r}^{(k)\,\mathrm{T}}\mathcal{T}^{(k)}\vec{r}^{(k)}\right)\mathcal{I}^{(m)}\vec{r}^{(m)}, \quad (4.76c)$$

$$B_3 = \left[\prod_{i\in[N]\setminus\{m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{r}^{(i)}\right]\mathcal{K}^{(m)}\vec{r}^{(m)}, \quad (4.76d)$$

$$B_4 = \left[\prod_{i\in[N]\setminus\{l,m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{r}^{(i)}\right]\left(\vec{r}^{(l)\,\mathrm{T}}\tilde{\mathcal{T}}^{(l)}\vec{r}^{(l)}\right)\mathcal{T}^{(m)}\vec{r}^{(m)}, \quad (4.76e)$$

$$B_5 = \left[\prod_{i\in[N]\setminus\{k,m\}} \vec{r}^{(i)\,\mathrm{T}}\mathcal{I}^{(i)}\vec{r}^{(i)}\right]\left(\vec{r}^{(k)\,\mathrm{T}}\mathcal{T}^{(k)}\vec{r}^{(k)}\right)\tilde{\mathcal{T}}^{(m)}\vec{r}^{(m)}. \quad (4.76f)$$

This is linear system of order l_m.

Before presenting an example, there are two numerical issues we want to comment on.

The first is related to the fact that the tensor-product operation is not injective. Indeed, given an ensemble $(s^{(1)},\dots,s^{(N)}) \in \times_{i\in[N]}\mathrm{H}(D_i;M_i)$ and, for $i\in[N]$, scalars $\beta_i\in\mathbb{R}$ such that $\prod_{i=1}^N \beta_i = 1$,

$$\bigotimes_{i=1}^{N} s^{(i)} = \bigotimes_{i=1}^{N}\left(\beta_i s^{(i)}\right).$$

Thus, there is a risk of having the inner iteration Algorithm VII sliding from one ensemble to another with a very similar tensor-product. This can hamper the convergence of the inner iteration; particularly so in the light of the termination criterion `2.2` of Algorithm VII. Therefore, we prefer to work with $\mathrm{L}^2_{1/M_i}(D_i)$-normalized factors $r^{(i)}$ together with a scale

factor in $\mathbb{R}$. This involves enriching (4.76) with a Lagrange multiplier that takes into account the restriction on the $\mathrm{L}^2_{1/M_i}(D_i)$ norms.

The second is that we have observed that the inner iteration benefits from using some form of relaxation, wherein, at least for a few of the first inner iterations, each new iteration is averaged with the previous one.

4.2.4. Numerical example. We implemented the Discrete Orthogonal Greedy Algorithm (Algorithm VI) as discussed in the previous two subsections, including the use of the inner iteration (Algorithm VII). The problem we approximate is given by (2.5) with parameters

$$N = 3, \qquad \frac{1}{4\mathrm{Wi}} = 1, \qquad c = 1 \tag{4.77a}$$

and, for i and j in $[N]$,

$$D_i = (-1, 1), \qquad \alpha_i = 10, \qquad M_i = \left(1 - \bullet^2\right)^{\alpha_i}, \qquad A_{i,j} = \delta_{i,j}. \tag{4.77b}$$

and its solution $\psi\colon \mathsf{D} = (-1,1)^3 \to \mathbb{R}$ is given by

$$\psi(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3) = (\cos(3\pi \boldsymbol{q}_1)\boldsymbol{q}_2 + \exp(\boldsymbol{q}_2 - \boldsymbol{q}_3)) \prod_{i=1}^{3} \rho(\boldsymbol{q}_i)^{\alpha_i}. \tag{4.77c}$$

The right hand side $f \in \mathrm{H}(\mathsf{D};\mathsf{M})'$ has the form (4.64) exactly. As the number of terms in this expansion, N_g, is 36 in this case, we do not write it down.

As, most of all, the alteration of the matrix A (cf. (1.6)) implies, this example is only useful for numerical testing purposes and has no rheological interpretation. Note, however, that the weights M_i have the right kind of singular behavior, and that the alteration of A has not altered the elliptical nature of the system nor the size of each inner iteration system (4.76).

For $i \in [N]$, we decided to set the basis functions which span $\hat{\mathrm{H}}^{(i)}_{l_i}$ according to

$$\varphi^{(i)}_k := M_i \, J^{(\alpha_i)}_k \qquad \forall\, k \in \{0, \dots, l_i\}.$$

Here, we have switched to a 0-based indexing of the basis functions. Thus,

$$\hat{\mathrm{H}}^{(i)}_{l_i} = M_i \, \mathbb{P}_{l_i}.$$

This was the choice studied Subsection 4.1.6 (cf. (4.30)). The use Jacobi-polynomial-based basis functions with parameter α_i greatly simplifies the generation of the Gram matrices defined in (4.74) and the load vectors defined in (4.75). At last, we will use the same number of degrees of freedom in each direction; that is, for some $l \in \mathbb{N}$, $l_i = l$ for all $i \in [N]$.

Theorem 4.29 applies in this case, so we should expect a convergence rate of the order $\mathcal{O}(n^{-1/2})$, where n is the iteration number, plus a projection error. As the approximated function is smooth, the projection error decays very rapidly.

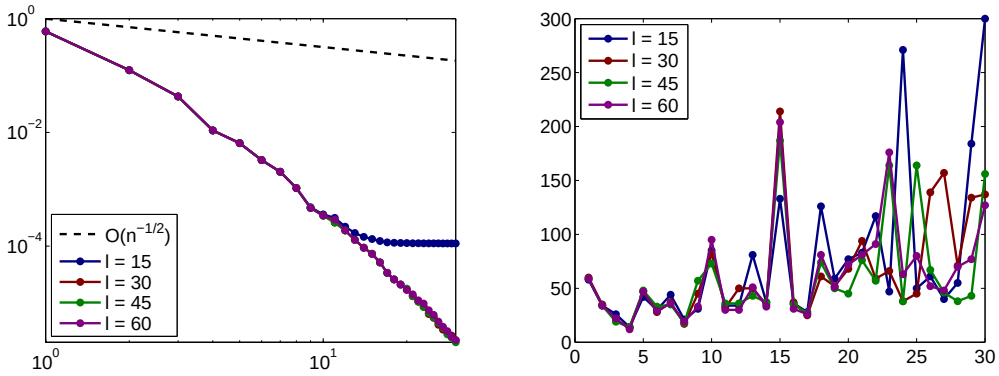

FIGURE 4.2. *Left*: Log-log plot of iteration number against $\mathrm{H}(\mathsf{D};\mathsf{M})$ error for discretization parameter l in $\{15, 30, 45, 60\}$. *Right*: Number of inner iterations used in each case.

The results are shown in Figure 4.2 and Figure 4.3. From the left plot of Figure 4.2 it is apparent that the error decays faster than the expected $\mathcal{O}(n^{-1/2})$ rate, more like $\mathcal{O}(n^{-2})$. In the case of the lowest discretization parameter used, $l = 15$, stagnation is observed after roughly ten iterations, which is compatible with the fact that Algorithm VI was proved to converge to the projection of the solution onto $\hat{\mathrm{H}}_{\boldsymbol{l}}$, which, for low discretization parameters, can be quite different to the solution itself.

In Figure 4.3 we compare the approximation at the end of the last iteration when using the highest number of degrees of freedom in each direction (top row) with the true solution ψ (bottom row). We chose to plot ψ/M, because ψ itself decays very fast, making its features difficult to distinguish. A close look at the plots will reveal that the quality of the approximation of ψ/M is worse near the boundary of the domain, which is to be expected; indeed, as the $\mathrm{H}(\mathsf{D};\mathsf{M})$ norm of ψ equals the $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ norm of ψ/M (cf. (2.4)), point evaluations of the latter near the boundary of the domain contribute very little to the overall error.

This suggests two main observations. One is that the convergence rate given by our theory falls short of what can be seen in practice; some theoretical reformulation might be needed to explain this phenomenon, for the $\mathcal{O}(n^{-1/2})$ rate seems to be an intrinsic part of the approach followed in this work. The other is that the inner iteration given in Algorithm VII—with, in our case, normalization of the separated factors—might indeed be a reasonable way of approximating the minimizers required by the Discrete Orthogonal Greedy Algorithm (Algorithm VI) in its step `2.1`.

We have shown that the discrete variants of the greedy algorithms do converge to a finite-dimensional projection of the sought after solution for fixed discretization parameter. We have also shown that stability and approximation estimates can be combined to give composite convergence rates of the discrete greedy algorithms with respect to the true solution. We

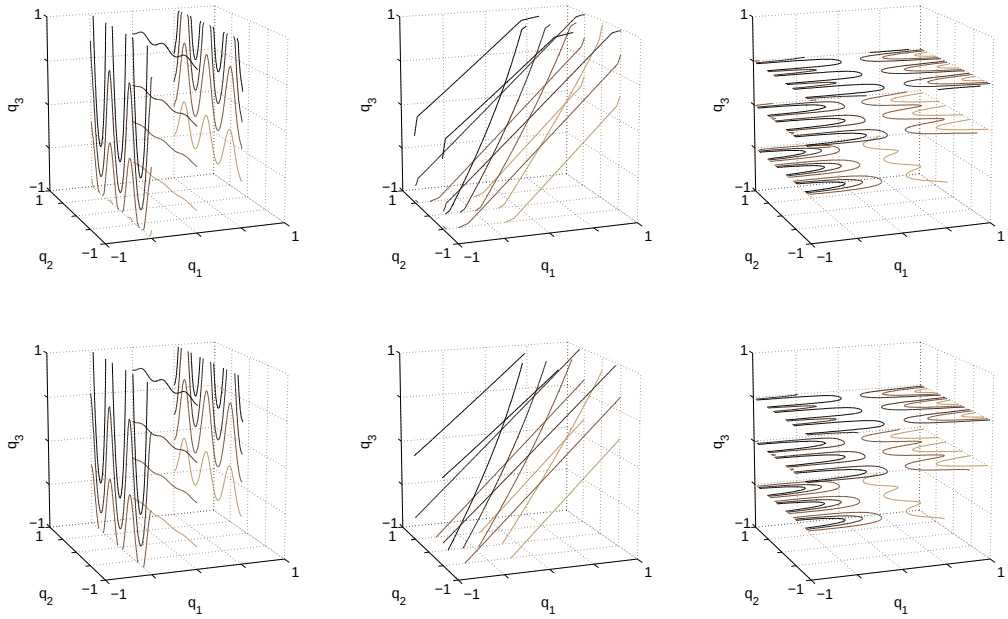

FIGURE 4.3. *Top row*: Some plane-restricted contour lines (generated using the command `contourslice` of MATLAB) of the ratio of one of the computed approximations and the Maxwellian. *Bottom row*: Contour lines drawn at the same values and at the same planes as those in the top row, but of the ratio of the true solution ψ and the Maxwellian.

then gave particular examples of finite dimensional subspaces and gave conditions on the true solution so that the composite convergence rates are valid. We ended the chapter on a cautionary note on the gaps existing between the discrete greedy algorithms and algorithms fit for computational purposes, and illustrated the behavior of one of the latter with a simple numerical example.

CHAPTER 5

Conclusions

The main focus of this work was the study of the convergence properties of the Separated Representation strategy using tools from the theory of greedy algorithms.

The degeneracy of the Fokker–Planck operator compelled us to give careful consideration to the functional-analytical setting we would give to our arguments. We studied the properties of the Maxwellian-weighted Sobolev spaces and tensor-product-weighted Sobolev spaces that arose naturally from our analysis, and expounded on the geometric properties of the domains they are defined upon.

We performed the basic analysis on the well-posedness and convergence of the continuous greedy algorithms Algorithm I and Algorithm II and then proceeded to characterize subspaces of the space of solutions with rapid convergence $\mathcal{B}_1$. We defined families of subspaces in terms of weighted summability of squared Fourier coefficients, which motivated the study of the spectral properties of the single-domain (low-dimensional) elliptic Fokker–Planck operators. Then, we defined subspaces of $\mathcal{B}_1$ in terms of weighted Sobolev regularity of their members. We found that spaces which control derivatives of high overall order but still relatively low component-by-component order are especially well suited to this task. We used elliptic regularity results to show either (in the single-domain case) equivalence between certain spaces of weighted summability of squared Fourier coefficients and spaces defined in terms of regularity or (in the tensor-product domain case) inclusion relations between these two classes of families of subspaces of $\mathcal{B}_1$.

We also analyzed the discrete greedy algorithms Algorithm V and Algorithm VI, where the approximation terms are sought in the tensor product of finite dimensional subspaces. In the case of Algorithm VI we found a way to produce, in this nonlinear approximation setting, composite convergence rate estimates involving both the iteration number of the greedy algorithm and a measure of the discretization error. For certain families of finite-dimensional spaces—in particular, for spaces defined using polynomials—we gave regularity conditions on the true solution that guarantee the validity of the composite rate just mentioned. We discussed what stands between the algorithms in their current form and algorithms suitable for computational work, and illustrated an implementation with a simple numerical example.

Among the possible avenues of further work, of various importance, we mention:

- Approximating convection terms implicitly, using least-squares-like formulations to maintain the characterization of the solution to the Fokker–Planck equation as a minimum of a functional.
- Proving that Maxwellians associated with Inverse Langevin force law are also, near the boundary of their domain, the product of a power of the distance-to-the-boundary function and a smooth function.
- Simplifying the definition of the $\tilde{\mathrm{H}}^{m,\mathrm{mix}}_{\mathsf{M}(\mathsf{D})}$ spaces, $m \in \{2,4\}$, defined in (3.78) and (3.79) and, possibly, of their $(1/2,2)$ (real) interpolation, tentatively named $\tilde{\mathrm{H}}^{3,\mathrm{mix}}_{\mathsf{M}(\mathsf{D})}$ after (3.83).
- Obtaining full equivalence between the subspaces of $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ defined by the weighted summability of their squared Fourier coefficients and the subspaces of $\mathrm{H}^1_{\mathsf{M}}(\mathsf{D})$ defined by weighted integrability of their derivatives.
- Understanding why we had to use a non-standard spherical coordinate system in the proof of the iterated elliptic regularity result Lemma 3.26.
- Extending the results obtained in Subsection 4.1.6 for polynomial-based search subspaces from the tensor product of intervals to the tensor product of discs and spheres.
- Obtaining $\mathcal{A}_{1,\mathcal{U}}$-stability results from approximation results or vice versa, as they seem to be related.
- Analyzing the minimization procedure used in practice to obtain the iterates of the discrete greedy algorithms.
- Applying the various auxiliary results to the analysis of other approximation schemes.

Finally, we believe that, in the process of producing this work, we have not only advanced the understanding of the Separated Representation strategy, but also obtained a number of results that are of wider interest, which, to the best of our knowledge, were not previously available in the literature.

APPENDIX A

Auxiliary results

A.1. Some results on distributions

Throughout this section Ω will denote an open subset of $\mathbb{R}^d$.

Lemma A.1. *Let $\alpha \in \mathbb{N}_0^d$ and $f, g \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$. Then,*

$$A_{f,g}(\varphi) := \int_\Omega \Big(f\partial_\alpha \varphi - (-1)^{|\alpha|} g\varphi \Big) = 0 \quad \forall\, \varphi \in \mathrm{C}_0^\infty(\Omega) \iff A_{f,g}(\varphi) = 0 \quad \forall\, \varphi \in \mathrm{C}_0^{|\alpha|}(\Omega).$$

In other words, a weak derivative of order $|\alpha|$ can be defined by using $\mathrm{C}_0^{|\alpha|}(\Omega)$ test functions instead of the usual $\mathrm{C}_0^\infty(\Omega)$ test functions.

Proof. The right-to-left implication is immediate. The left-to-right implication is a consequence of Theorem 2.1.6 in [**Hör83**]. □

Lemma A.2. *Let $\alpha \in \mathbb{N}_0^d$ and let $f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ and $g \in \mathrm{C}(\Omega)$ be such that $\partial_\beta f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ and $\partial_\beta g \in \mathrm{C}(\Omega)$ for all $\beta \le \alpha$. Then, $\partial_\alpha(fg) \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ and*

$$\partial_\alpha(fg) = \sum_{\beta \le \alpha} \frac{\alpha!}{\beta!(\alpha-\beta)!} \partial_\beta f\, \partial_{\alpha-\beta} g, \tag{A.1}$$

where the convention $\gamma! = \prod_{i\in[d]} \gamma_i!$ for all $\gamma \in \mathbb{N}_0^d$ has been followed.

Proof. To begin with, the result is obviously true for $\alpha = (0, \ldots, 0)$. Once we have shown that the result is valid for $|\alpha| = 1$, the final result will follow from standard combinatorial arguments and an induction procedure.

Let, then, $|\alpha| = 1$ and let $\varphi \in \mathrm{C}_0^\infty(\Omega)$. From Lemma A.1 and the fact that $g\varphi \in \mathrm{C}_0^1(\Omega)$ we have that

$$0 = \int_\Omega \left(f\partial_\alpha(g\varphi) + \partial_\alpha f g\varphi \right) = \int_\Omega \left((f\partial_\alpha g + \partial_\alpha f g)\varphi + fg\partial_\alpha\varphi \right).$$

Then, as the products fg, $\partial_\alpha f g$ and $f\partial_\alpha g$ are $\mathrm{L}^1_{\mathrm{loc}}(\Omega)$ functions, $\partial_\alpha(fg) = f\partial_\alpha g + \partial_\alpha f g$ is a distributional identity. □

The purpose of the following lemma is to formulate a result analogous to Theorem 3.41 of [**AF03**] for weighted Sobolev spaces without resorting to density arguments, which may be unavailable for one or both of the weighted Sobolev spaces being connected.

Lemma A.3. *Let T be an invertible $\mathrm{C}^\infty(\overline{\Omega})$ transformation with codomain $\tilde{\Omega}$ and let $f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ be such that its distributional derivatives are in $\mathrm{L}^1_{\mathrm{loc}}(\Omega)$ up to the order $\alpha \in \mathbb{N}^d$. Then,*

$$\partial_\alpha(f \circ T^{-1}) = \sum_{1 \le |\beta| \le |\alpha|} M_{\alpha,\beta}(\partial_\beta f \circ T^{-1}) \in \mathrm{L}^1_{\mathrm{loc}}(\tilde{\Omega}), \tag{A.2}$$

where $M_{\alpha,\beta}$ is a polynomial of degree not exceeding $|\beta|$ in derivatives of orders not exceeding $|\alpha|$ of the various components of T^{-1}.

Proof. Let S denote the inverse of T and let S_k denote the its k-th component. From Theorem 6.1.2 in [**Hör83**] and the remark that follows it we know, first, that there exists a unique continuous linear map $S^*\colon \mathcal{D}'(\Omega) \to \mathcal{D}'(\tilde{\Omega})$ whose restriction to $\mathrm{C}(\Omega)$ is $u \mapsto u \circ S$ and, second, that the chain rule,

$$\partial_j S^* u = \sum_{k=1}^d \partial_j S_k \, S^* \partial_k u$$

holds in $\mathcal{D}'(\tilde{\Omega})$. It is easy to see (either directly or from the proof of Theorem 6.1.2 of [**Hör83**]) that S^*u has the explicit form

$$S^*u(\varphi) = u\big((\varphi \circ T)\,|\det(\nabla T)|\,\big) \qquad \forall\, \varphi \in \mathrm{C}^\infty_0(\tilde{\Omega}).$$

For a regular distribution such as f the above characterization and the change of variable formula for integrable functions (see, e.g., [**Bog07**, Theorem 3.7.1]) makes S^*f precisely the regular distribution associated with the $\mathrm{L}^1_{\mathrm{loc}}(\tilde{\Omega})$ function $f \circ S$. Similarly, $S^*\partial_k f$ will be the regular distribution associated with the $\mathrm{L}^1_{\mathrm{loc}}(\tilde{\Omega})$ function $\partial_k f \circ S$. Hence, $\partial_j(f \circ T^{-1}) = \sum_{k=1}^d \partial_j S_k \partial_k f \circ S$ and (A.2) is proved for $|\alpha| = 1$. An induction argument then establishes (A.2) in the general case. □

Lemma A.4. *Let $\tilde{\Omega}$ and T be as in Lemma A.3 and let w be a weight function defined on Ω. Then, $f \in \mathrm{H}^m_w(\Omega)$ if, and only if, $f \circ T^{-1} \in \mathrm{H}^m_{\tilde{w}}(\tilde{\Omega})$ and there exist positive constants $c_1(m)$ and $c_2(m)$ such that*

$$c_1 \|f \circ T^{-1}\|_{\mathrm{H}^m_{\tilde{w}}(\tilde{\Omega})} \le \|f\|_{\mathrm{H}^m_w(\Omega)} \le c_2 \|f \circ T^{-1}\|_{\mathrm{H}^m_{\tilde{w}}(\tilde{\Omega})},$$

where $\tilde{w} = w \circ T^{-1}$.

Proof. We use Lemma A.3 to replace the first part of the proof of Theorem 3.41 of [**AF03**]. Then, the rest of that proof, *mutatis mutandis*, carries over to our case. □

A.2. Variational eigenvalue problems

The following abstract lemma states standard results (essentially, the Hilbert–Schmidt theorem and some of its corollaries). As we could not find these results in the literature in the precise form stated here, we provide a brief proof.

Lemma A.5. *Let H and V be separable infinite-dimensional Hilbert spaces, with $V \Subset H$ and $\overline{V} = H$ in the norm of H. Let $a\colon V \times V \to \mathbb{R}$ be a nonzero, symmetric, bounded and elliptic bilinear form. Then, there exist sequences of real numbers, denoted by $(\lambda_n\colon n \in \mathbb{N})$, and unit H-norm members of V, denoted by $(e_n\colon n \in \mathbb{N})$, which solve the following problem:* Find $\lambda \in \mathbb{R}$ and $e \in H \setminus \{0\}$ such that

$$a(e, v) = \lambda \langle e, v\rangle_H \quad \forall\, v \in V. \tag{A.3}$$

The λ_n, which can be assumed to be in increasing order with respect to n, are positive, bounded from below away from 0, and $\lim_{n\to\infty} \lambda_n = \infty$.

Additionally, the e_n form an H-orthonormal system whose H-closed span is H and the rescaling $e_n/\sqrt{\lambda_n}$ gives rise to an a-orthonormal system whose a-closed span is V, so we have

$$h = \sum_{n=1}^{\infty} \langle h, e_n\rangle_H e_n \quad \textit{and} \quad \|h\|_H^2 = \sum_{n=1}^{\infty} \langle h, e_n\rangle_H^2 \qquad \forall\, h \in H \tag{A.4}$$

and

$$v = \sum_{n=1}^{\infty} a\left(v, \frac{e_n}{\sqrt{\lambda_n}}\right) \frac{e_n}{\sqrt{\lambda_n}} \quad \textit{and} \quad \|v\|_a^2 = \sum_{n=1}^{\infty} a\left(v, \frac{e_n}{\sqrt{\lambda_n}}\right)^2 \qquad \forall\, v \in V; \tag{A.5}$$

further,

$$h \in H \quad \textit{and} \quad \sum_{n=1}^{\infty} \lambda_n \langle h, e_n\rangle_H^2 < \infty \iff h \in V. \tag{A.6}$$

Proof. This proof is an adaptation of the proof of Theorem IX.31 in [**Bre83**]. The Lax–Milgram lemma implies the existence of an operator $\tilde{T}\colon H \to V$ where, given $h \in H$, $\tilde{T}(h)$ is defined as the unique solution in V to the variational problem

$$a(\tilde{T}(h), v) = \langle h, v\rangle_H \quad \forall\, v \in V. \tag{A.7}$$

It also follows, via the elliptic stability estimate of the Lax–Milgram lemma and the continuity of the embedding $V \hookrightarrow H$, that $\tilde{T}$ is bounded. Let $i\colon V \to H$ denote the embedding operator that maps V into H, i.e., $v \in V \mapsto i(v) = v \in H$. Then, $T := i \circ \tilde{T}$ is a bounded operator defined on H with values in H; as $i\colon V \to H$ is a compact linear operator, it follows that $T\colon H \to H$ is a compact linear operator. Further, for all $(h, h') \in H \times H$,

$$\begin{aligned}\langle T(h), h'\rangle_H = \langle \tilde{T}(h), h'\rangle_H = \langle h', \tilde{T}(h)\rangle_H = a(\tilde{T}(h'), \tilde{T}(h)) \\ = a(\tilde{T}(h), \tilde{T}(h')) = \langle h, \tilde{T}(h')\rangle_H = \langle h, T(h')\rangle_H,\end{aligned}$$

whence T is self-adjoint. Thus, thanks to Theorem VI.11 in [**Bre83**] (the spectral theorem for compact and self-adjoint operators in Hilbert spaces), there exists an H-orthonormal system $(e_n\colon n \geq 1)$ of eigenvectors of T such that

$$h = \sum_{n=1}^{\infty} \langle h, e_n\rangle_H e_n \quad \text{and} \quad \|h\|_H^2 = \sum_{n=1}^{\infty} \langle h, e_n\rangle_H^2 \qquad \forall\, h \in H. \tag{A.8}$$

As, for all $h \in H$, $\langle T(h), h\rangle_H = a(\tilde{T}(h), \tilde{T}(h))$ and a is V-elliptic, all the eigenvalues of T are nonnegative. Also, as T is bounded, the set of its eigenvalues is also bounded. Now, by Theorem VI.8 in [**Bre83**], the set of nonzero eigenvalues of T is either empty, or finite, or countable with 0 as its only accumulation point. However, on account of (A.8), the latter alternative is then the one that holds.

If 0 were an eigenvalue of T, there would exist $e \in H \setminus \{0\}$ such that $T(e) = 0$; i.e., $e \in \mathrm{Ker}(T)$. However, from (A.7) we then have that $e \in V^{\perp_H}$. As $H = \overline{V} \oplus V^{\perp_H}$ in the norm of H and V is dense in H, $V^{\perp_H} = \{0\}$, which contradicts $e \neq 0$. Therefore, 0 is not an eigenvalue of T.

From the above, we can take the eigenvectors e_n of (A.8) as associated to positive eigenvalues μ_n bounded from above, arranged in decreasing order ($\mu_{n+1} \leq \mu_n$ for $n \geq 1$) with $\lim_{n\to\infty} \mu_n = 0$. A consequence of the absence of 0 from the spectrum of T is that all the eigenvectors of T have to be members of the smaller space V.

Assuming that $\mu \neq 0$ and $e \in V \setminus \{0\}$, we have that $T(e) = \mu e$ if, and only if, $a(e, w) = \mu^{-1}\langle e, w\rangle_H$ for all $w \in V$. Then, all the eigenvalues of the eigenvalue problem (A.3) are reciprocals of eigenvalues of T with the possible exception of 0. However, from the V-ellipticity of a, 0 cannot be an eigenvalue of the problem (A.3). On defining $\lambda_n := \mu_n^{-1}$ and setting the e_n to be the same as in (A.8) we obtain the desired existence and distribution statements about of the eigenvalues of (A.3).

We observe from $a(e_n, e_m) = \lambda_n\langle e_n, e_m\rangle_H$, $n \geq 1$, that the sequence $\big(e_n/\sqrt{\lambda_n} \colon n \geq 1\big)$ is an a-orthonormal system in V. Let us denote the a-closure of its span by $\hat{V}$. Then, $v \in \hat{V}^{\perp_a}$ if, and only if, $a(v, e_n) = 0$ for all $n \geq 1$. As each e_n is an eigenfunction of the problem (A.3) associated to a nonzero eigenvalue, it follows from (A.8) that $v = 0$ and therefore $\hat{V}^{\perp_a} = \{0\}$. Thus, $V = \hat{V} \oplus \hat{V}^{\perp_a} = \hat{V}$ and we have the desired V-spanning property of the e_n (the desired H-spanning property was already given in (A.8)).

Using Theorem VI.9 of [**Bre83**] on the properties of Hilbert sums, we can turn the H- and V-spanning properties of the e_n into the expressions (A.4) and (A.5), respectively.

As (λ_n, e_n) is an eigenpair of (A.3), $a(v, e_n/\sqrt{\lambda_n}) = \sqrt{\lambda}\langle v, e_n\rangle_H$ for all $v \in V$; this and the second expression of (A.5) give the right-to-left implication in (A.6). Let us now consider an $h \in H$ that satisfies the left-hand side of (A.6). As the e_n are members of V, the partial sums

$$h_k := \sum_{n=1}^{k} \langle h, e_n\rangle_H e_n$$

also belong to V. The a-orthonormality of the $e_n/\sqrt{\lambda_n}$ leads to the equality, for $1 \leq k < l$,

$$\|h_l - h_k\|_a^2 = \sum_{n=k+1}^{l} \lambda_n \langle h, e_n\rangle_H^2.$$

As the real series $\sum_{n=1}^{\infty} \lambda_n \langle h, e_n \rangle_H^2$ is assumed to converge, the above expression tends to 0 as k and l tend to ∞. Hence, $(h_k \colon k \geq 1)$ is a Cauchy sequence in V and thus converges to some $\hat{h} \in V$. As V is continuously embedded in H (a consequence of being compactly embedded), the limit $\hat{h}$ has to be the same limit the h_k have in H. That is, $h = \hat{h} \in V$. This completes the proof of (A.6). □

Bibliography

[AF03] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, second ed., Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003, MR2424078 (2009e:46025). 27, 39, 135, 136

[AMCK06] A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings, *A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids*, J. Non-Newton. Fluid Mech. **139** (2006), no. 3, 153–176, doi:10.1016/j.jnnfm.2006.07.007. 19

[AMCK07] ———, *A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modelling of complex fluids: Part II: Transient simulation using space-time separated representations*, J. Non-Newton. Fluid Mech. **144** (2007), no. 2–3, 98–121, doi:10.1016/j.jnnfm.2007.03.009. 19

[AND+10] A. Ammar, M. Normandin, F. Daim, D. González, E. Cueto, and F. Chinesta, *Non incremental strategies based on separated representations: applications in computational rheology*, Commun. Math. Sci. **8** (2010), no. 3, 671–695, MR2730326. 19

[AR10] R. A. Askey and R. Roy, *Gamma function*, `http://dlmf.nist.gov/5`, 5 2010, Digital Library of Mathematical Functions. National Institute of Standards and Technology. 75

[Bat67] G. K. Batchelor, *An introduction to fluid dynamics*, Cambridge University Press, 1967. 3

[BCAH87] R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of polymeric liquids, volume 2, kinetic theory*, second ed., John Wiley and Sons, New York, 1987. 2, 3, 5, 6, 17, 18

[BCDD08] A. R. Barron, A. Cohen, W. Dahmen, and R. A. DeVore, *Approximation and learning by greedy algorithms*, Ann. Statist. **36** (2008), no. 1, 64–94, doi:10.1214/009053607000000631, MR2387964 (2009c:62055). 106, 107

[BG04] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer. **13** (2004), 147–269, doi:10.1017/S0962492904000182, MR2249147 (2007e:65102). 18

[BM97] C. Bernardi and Y. Maday, *Spectral methods*, Handbook of numerical analysis, Vol. V (P. G. Ciarlet and J. L. Lions, eds.), North-Holland, Amsterdam, 1997, MR1470226, pp. 209–485. 112, 113

[Bog07] V. I. Bogachev, *Measure theory. Volumes I and II*, Springer-Verlag, Berlin, Heidelberg, 2007, doi:10.1007/978-3-540-34514-5, MR2267655 (2008g:28002). 136

[Boy01] J. Boyd, *Chebyshev and Fourier spectral methods*, 2nd ed., Dover books on mathematics, Dover Publications, 2001. 123

[Bre83] H. Brezis, *Analyse fonctionnelle: Théorie et applications*, Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris, 1983, MR697382 (85a:46001). 71, 137, 138

[Bro61] F. E. Browder, *On the spectral theory of elliptic differential operators. I*, Math. Ann. **142** (1961), 22–130, doi:10.1007/BF01343363, MR0209909 (35 #804). 34

[BS70] M. Š. Birman and M. Z. Solomjak, *The principal term of the spectral asymptotics for "non-smooth" elliptic problems*, Funkcional. Anal. i Prilożen. **4** (1970), no. 4, 1–13, MR0278126 (43 #3857), Translated in Funct. Anal. Appl., **4** (1970), 265–275. 35

[BS72] ———, *Spectral asymptotics of nonsmooth elliptic operators. I*, Trudy Moskov. Mat. Obšč. **27** (1972), 3–52, MR0364898 (51 #1152), Translated in Trans. Moscow Math. Soc. **27** (1972), 1–5. 38

[BS07] J. W. Barrett and E. Süli, *Existence of global weak solutions to some regularized kinetic models for dilute polymers*, Multiscale Model. Simul. **6** (2007), no. 2, 506–546 (electronic), doi:10.1137/060666810, MR2338493 (2009i:76042). 3, 10

[BS08] ———, *Existence of global weak solutions to dumbbell models for dilute polymers with microscopic cut-off*, Math. Models Methods Appl. Sci. **18** (2008), no. 6, 935–971, doi:10.1142/S0218202508002917, MR2419205 (2009b:35317). 10, 27

[BS09] ———, *Numerical approximation of corotational dumbbell models for dilute polymers*, IMA J. Numer. Anal. **29** (2009), no. 4, 937–959, doi:10.1093/imanum/drn022, MR2557051 (2010m:65213). 10, 18

[BS11a] ———, *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, doi:10.1142/S0218202511005313. 10, 18

[BS11b] ———, *Finite element approximation of kinetic dilute polymer models with microscopic cut-off*, M2AN Math. Model. Numer. Anal. **45** (2011), no. 1, 39–89, doi:10.1051/m2an/2010030. 10, 18

[CALK11] F. Chinesta, A. Ammar, A. Leygue, and R. Keunings, *An overview of the proper generalized decomposition with applications in computational rheology*, J. Non-Newton. Fluid Mech. **166** (2011), no. 11, 578–592, doi:10.1016/j.jnnfm.2010.12.012. 19

[CEK+07] S. R. Chinnamsetty, M. Espig, B. N. Khoromskij, W. Hackbusch, and H.-J. Flad, *Tensor product approximation with optimal rank in quantum chemistry*, J. Chem. Phys. **127** (2007), no. 8, 084110, doi:10.1063/1.2761871. 124

[CEL11] E. Cancès, V. Ehrlacher, and T. Lelièvre, *Convergence of a greedy algorithm for high-dimensional convex nonlinear problems*, Math. Models Methods Appl. Sci. **21** (2011), no. 12, 2433–2467, doi:10.1142/S0218202511005799. 19, 68, 98

[CH53] R. Courant and D. Hilbert, *Methods of mathematical physics. Vol. I*, Interscience Publishers, Inc., New York, N.Y., 1953, MR0065391 (16,426a). 34

[CL04a] C. Chauvière and A. Lozinski, *Simulation of complex viscoelastic flows using the fokker-planck equation: 3D FENE model*, J. Non-Newton. Fluid Mech. **122** (2004), no. 1–3, 201–214, doi:10.1016/j.jnnfm.2003.12.011. 17

[CL04b] ———, *Simulation of dilute polymer solutions using a Fokker–Planck equation*, Computers & Fluids **33** (2004), no. 5–6, 687–696, doi:10.1016/j.compfluid.2003.02.002. 17

[Cla67] C. Clark, *The asymptotic distribution of eigenvalues and eigenfunctions for elliptic boundary value problems*, SIAM Rev. **9** (1967), 627–646, doi:10.1137/1009105, MR0510064 (58 #23164). 34

[CO01] C. Chauvière and R. G. Owens, *A new spectral element method for the reliable computation of viscoelastic flow*, Comput. Methods Appl. Mech. Engrg. **190** (2001), no. 31, 3999–4018, doi:10.1016/S0045-7825(01)00177-3, MR1829215 (2002b:76075). 17

[Coh91] A. Cohen, *A Padé approximant to the inverse Langevin function*, Rheol. Acta **30** (1991), no. 3, 270–273, doi:10.1007/BF00366640. 9, 10

[CQ82] C. Canuto and A. Quarteroni, *Approximation results for orthogonal polynomials in Sobolev spaces*, Math. Comp. **38** (1982), no. 157, 67–86, doi:10.2307/2007465, MR637287 (82m:41003). 116

[Dav95] E. B. Davies, *Spectral theory and differential operators*, Cambridge Studies in Advanced Mathematics, vol. 42, Cambridge University Press, Cambridge, 1995, doi:10.1017/CBO9780511623721, MR1349825 (96h:47056). 33

[DiB02] E. DiBenedetto, *Real analysis*, Birkhäuser Advanced Texts: Basler Lehrbücher [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Boston Inc., Boston, MA, 2002, MR1897317 (2003d:00001). 79

[DLO07] P. Delaunay, A. Lozinski, and R. G. Owens, *Sparse tensor-product Fokker–Planck-based methods for nonlinear bead-spring chain models of dilute polymer solutions*, High-dimensional partial differential equations in science and engineering (A. Bandrauk, M. C. Delfour, and C. Le Bris, eds.), CRM Proc. Lecture Notes, vol. 41, Amer. Math. Soc., Providence, RI, 2007, MR2359669 (2008k:82153), pp. 73–89. 3, 18

[DLY05] Q. Du, C. Liu, and P. Yu, *FENE dumbbell model and its several linear and nonlinear closure approximations*, Multiscale Model. Simul. **4** (2005), no. 3, 709–731 (electronic), doi:10.1137/040612038, MR2203938 (2007b:76002). 17

[DPL04] G. Da Prato and A. Lunardi, *On a class of elliptic operators with unbounded coefficients in convex domains*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. **15** (2004), no. 3-4, 315–326, MR2148888 (2006a:35044). 78

[DT96] R. A. DeVore and V. N. Temlyakov, *Some remarks on greedy algorithms*, Adv. Comput. Math. **5** (1996), no. 2-3, 173–187, doi:10.1007/BF02124742, MR1399379 (97g:41029). 19, 65, 68, 104, 105

[Eve05] W. N. Everitt, *A catalogue of Sturm-Liouville differential equations*, Sturm-Liouville theory (W. O. Amrein, A. M. Hinz, and D. B. Pearson, eds.), Birkhäuser, Basel, 2005, doi:10.1007/3-7643-7359-8_12, MR2145086, pp. 271–331. 35

[FLÖ95] K. Feigl, M. Laso, and H. C. Öttinger, *CONNFFESSIT approach for solving a two-dimensional viscoelastic fluid problem*, Macromolecules **28** (1995), no. 9, 3261–3274, doi:10.1021/ma00113a031. 2

[Fre87] D. A. French, *The finite element method for a degenerate elliptic equation*, SIAM J. Numer. Anal. **24** (1987), no. 4, 788–815, doi:10.1137/0724051, MR899704 (88k:65110). 80, 82

[GACC10] D. González, A. Ammar, F. Chinesta, and E. Cueto, *Recent advances on the use of separated representations*, Internat. J. Numer. Methods Engrg. **81** (2010), no. 5, 637–659, MR2640987. 19

[GL10] S. R. Ghorpade and B. V. Limaye, *A course in multivariable calculus and analysis*, Undergraduate Texts in Mathematics, Springer, New York, 2010, doi:10.1007/978-1-4419-1621-1, MR2583676. 72, 76

[GP02] S. C. Glotzer and W. Paul, *Molecular and mesoscale simulation methods for polymer materials*, Annu. Rev. Mater. Res. **32** (2002), 401–436, doi:10.1146/annurev.matsci.32.010802.112213. 1

[Gri85] P. Grisvard, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics, vol. 24, Pitman (Advanced Publishing Program), Boston, MA, 1985, MR775683 (86m:35044). 45, 49

[GT01] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001, MR1814364 (2001k:35004), Reprint of the 1998 edition. 88

[Guo00] B.-y. Guo, *Jacobi approximations in certain Hilbert spaces and their applications to singular differential equations*, J. Math. Anal. Appl. **243** (2000), no. 2, 373–408, doi:10.1006/jmaa.1999.6677, MR1741531 (2001b:65082). 116

[GW04] B.-y. Guo and L.-l. Wang, *Jacobi approximations in non-uniformly Jacobi-weighted Sobolev spaces*, J. Approx. Theory **128** (2004), no. 1, 1–41, doi:10.1016/j.jat.2004.03.008, MR2063010 (2005h:41010). 123

[HK07] W. Hackbusch and B. N. Khoromskij, *Tensor-product approximation to operators and functions in high dimensions*, J. Complexity **23** (2007), no. 4-6, 697–714, doi:10.1016/j.jco.2007.03.007, MR2372023 (2008k:65042). 124

[HO06] C. Helzel and F. Otto, *Multiscale simulations for suspensions of rod-like molecules*, J. Comput. Phys. **216** (2006), no. 1, 52–75, doi:10.1016/j.jcp.2005.11.028, MR2223436 (2006k:76008). 18

[Hör83] L. Hörmander, *The analysis of linear partial differential operators. I*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], no. 256, Springer-Verlag, Berlin, 1983, MR717035 (85g:35002a). 135, 136

[HS02] C. Horgan and G. Saccomandi, *A molecular-statistical basis for the gent constitutive model of rubber elasticity*, Journal of Elasticity **68** (2002), no. 1, 167–176, doi:10.1023/A:1026029111723. 10

[HUL01] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001, MR1865628 (2002i:90002), Abridged version of *Convex analysis and minimization algorithms. I* [Springer, Berlin, 1993; MR1261420 (95m:90001)] and *II* [ibid.; MR1295240 (95m:90002)]. 49

[HvHvdB97] M. A. Hulsen, A. P. G. v. Heel, and B. H. A. A. v. d. Brule, *Simulation of viscoelastic flows using brownian configuration fields*, J. Non-Newtonian Fluid Mech. **70** (1997), no. 1–2, 79–101, doi:10.1016/S0377-0257(96)01503-0. 16

[JBL04] B. Jourdain, C. L. Bris, and T. Lelièvre, *On a variance reduction technique for micro-macro simulations of polymeric fluids*, J. Non-Newtonian Fluid Mech. **122** (2004), no. 1–3, 91–106, doi:10.1016/j.jnnfm.2003.09.006. 16

[Jos90] D. D. Joseph, *Fluid dynamics of viscoelastic liquids*, Applied Mathematical Sciences, no. 84, Springer-Verlag, New York, 1990, 1051193 (91d:76003). 1

[KB09] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev. **51** (2009), no. 3, 455–500, doi:10.1137/07070111X, MR2535056 (2010j:15027). 124

[Keu00] R. Keunings, *A survey of computational rheology*, XIIIth International Congress on Rheology (Cambridge, UK) (D. M. Binding, N. E. Hudson, J. Mewis, J.-M. Piau, C. J. S. Petrie, T. P., W. H. Wagner, and K. Walters, eds.), 8 2000, Available at `http://www.mate.tue.nl/~hulsen`. 1, 2

[Keu04] ———, *Micro-macro methods for the multiscale simulation of viscoelastic flow using molecular models of kinetic theory*, Rheology Reviews (2004), 67–98. 2, 16

[KG42] W. Kuhn and F. Grün, *Beziehungen zwischen elastischen Konstanten und Dehnungsdoppelbrechung hochelastischer Stoffe*, Kolloid Z. **101** (1942), no. 3, 248–271, doi:10.1007/BF01793684. 6, 9

[Kne08] D. J. Knezevic, *Analysis and implementation of numerical methods for simulating dilute polymeric fluids*, Ph.D. thesis, University of Oxford, 10 2008, Available at `http://www.cs.ox.ac.uk/publications/publication2740-abstract.html`. 3, 5, 6, 15, 17

[KO84] A. Kufner and B. Opic, *How to define reasonably weighted Sobolev spaces*, Comment. Math. Univ. Carolin. **25** (1984), no. 3, 537–554, MR775568 (86i:46036). 27, 28

[KS09a] D. J. Knezevic and E. Süli, *A heterogeneous alternating-direction method for a micro-macro dilute polymeric fluid model*, M2AN Math. Model. Numer. Anal. **43** (2009), no. 6, 1117–1156, doi:10.1051/m2an/2009034, MR2588435. 17

[KS09b] ———, *Spectral Galerkin approximation of Fokker-Planck equations with unbounded drift*, M2AN Math. Model. Numer. Anal. **43** (2009), no. 3, 445–485, doi:10.1051/m2an:2008051, MR2536245. 17, 123

[Kuf85] A. Kufner, *Weighted Sobolev spaces*, John Wiley & Sons Inc., New York, 1985, MR802206 (86m:46033), Translated from the Czech. 78, 79

[LBLM09] C. Le Bris, T. Lelièvre, and Y. Maday, *Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations*, Constr. Approx. **30** (2009), no. 3, 621–651, doi:10.1007/s00365-009-9071-1, MR2558695. 19, 20, 21, 54, 56, 59, 60, 61, 64, 70, 93, 94

[LC03] A. Lozinski and C. Chauvière, *A fast solver for Fokker–Planck equation applied to viscoelastic flows calculations: 2D FENE model*, J. Comput. Phys. **189** (2003), no. 2, 607–625, doi:10.1016/S0021-9991(03)00248-1, MR1996059. 17

[LÖ93] M. Laso and H. C. Öttinger, *Calculation of viscoelatic flow using molecular models: the CONNFFESSIT approach*, J. Non-Newtonian Fluid Mech. **47** (1993), 1–20, doi:10.1016/0377-0257(93)80042-A. 2, 16

[Loz03] A. Lozinski, *Spectral methods for kinetic theory models of viscoelastic fluids*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, 2003, doi:10.5075/epfl-thesis-2860. 2, 16

[LP09] G. M. Leonenko and T. N. Phillips, *On the solution of the Fokker-Planck equation using a high-order reduced basis approximation*, Comput. Methods Appl. Mech. Engrg. **199** (2009), no. 1-4, 158–168, doi:10.1016/j.cma.2009.09.028, MR2566221 (2010j:76006). 19

[Mas08] N. Masmoudi, *Well-posedness for the FENE dumbbell model of polymeric flows*, Comm. Pure Appl. Math. **61** (2008), no. 12, 1685–1714, doi:10.1002/cpa.20252, MR2456183. 63

[MMP98] M. Marcus, V. J. Mizel, and Y. Pinchover, *On the best constant for Hardy's inequality in* $\mathbf{R}^n$, Trans. Amer. Math. Soc. **350** (1998), no. 8, 3237–3255, doi:10.1090/S0002-9947-98-02122-9, MR1458330 (98k:26035). 33

[MÖ96] M. Melchior and H. C. Öttinger, *Variance reduced simulations of polymer dynamics*, J. Chem. Phys. **105** (1996), no. 8, 3316–3331, doi:10.1063/1.472186. 16

[Nay98] R. Nayak, *Molecular simulation of liquid crystal polymer flow: a wavelet-finite element analysis*, Ph.D. thesis, Massachusetts Institute of Technology, 1998, Available at `http://hdl.handle.net/1721.1/9609`. 18

[Neč62] J. Nečas, *Sur une méthode pour résoudre les équations aux dérivées partielles du type elliptique, voisine de la variationnelle*, Ann. Scuola Norm. Sup. Pisa (3) **16** (1962), 305–326, MR0163054 (29 #357). 50

[NLM09] A. Nouy and O. P. Le Maître, *Generalized spectral decomposition for stochastic nonlinear problems*, J. Comput. Phys. **228** (2009), no. 1, 202–235, doi:10.1016/j.jcp.2008.09.010, MR2464076 (2010d:60154). 19

[Nou07] A. Nouy, *A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations*, Comput. Methods Appl. Mech. Engrg. **196** (2007), no. 45-48, 4521–4537, doi:10.1016/j.cma.2007.05.016, MR2354451 (2008g:65012). 19

[Nou08] ———, *Generalized spectral decomposition method for solving stochastic finite element equations: invariant subspace problem and dedicated algorithms*, Comput. Methods Appl. Mech. Engrg. **197** (2008), no. 51-52, 4718–4736, doi:10.1016/j.cma.2008.06.012, MR2464512 (2009m:60151). 19

[OK90] B. Opic and A. Kufner, *Hardy-type inequalities*, Pitman Research Notes in Mathematics Series, vol. 219, Longman Scientific & Technical, Harlow, 1990, MR1069756 (92b:26028). 49, 79, 80

[OP02] R. G. Owens and T. N. Phillips, *Computational rheology*, Imperial College Press, London, 2002, doi:10.1142/9781860949425, MR1906885 (2003h:76007). 16

[ÖvdBH97] H. C. Öttinger, B. H. A. A. v. d. Brule, and M. A. Hulsen, *Brownian configuration fields and variance reduced connffessit*, J. Non-Newton. Fluid Mech. **70** (1997), no. 3, 255–261, doi:10.1016/S0377-0257(96)01547-9. 16

[RS80] M. Reed and B. Simon, *Methods of modern mathematical physics. I*, second ed., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, MR751959 (85e:46002), Functional analysis. 23

[SC88] C. E. Seymour and R. B. Carraher, *Polymer chemistry*, second ed., Dekker, New York, 1988. 1

[Sch07] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen*, Math. Ann. **63** (1907), no. 4, 433–476, doi:10.1007/BF01449770. 19

[Tar07] L. Tartar, *An introduction to Sobolev spaces and interpolation spaces*, Lecture Notes of the Unione Matematica Italiana, vol. 3, Springer, Berlin, 2007, MR2328004 (2008g:46055). 95

[Taš75] G. M. Taščijan, *The spectral asymptotic behavior of elliptic boundary value problems with weak degeneracy*, Proceedings of the Sixth Winter School on Mathematical Programming and Related Questions (Drogobych, 1973), Functional analysis and its applications (Russian), Akad. Nauk SSSR Central. Èkonom.-Mat. Inst., Moscow, 1975, MR0481633 (58 #1739), pp. 277–293. 38

[Taš81] ———, *The classical formula of the asymptotic behavior of the spectrum of elliptic equations that are degenerate on the boundary of the domain*, Mat. Zametki **30** (1981), no. 6, 871–880, 959, MR641661 (83b:35128), Translated in Mathematical Notes **30** (1981), no. 6, 937–942, doi:10.1007/BF01145775. 38

[Tem08] V. N. Temlyakov, *Greedy approximation*, Acta Numer. **17** (2008), 235–409, doi:10.1017/S0962492906380014, MR2436013 (2009g:41066). 68

[Tre54] L. R. G. Treloar, *The photoelastic properties of short-chain molecular networks*, T. Faraday Soc. **50** (1954), 881–896, doi:10.1039/TF9545000881. 9

[TS71] R. I. Tanner and W. Stehrenberger, *Stresses in dilute solutions of bead?nonlinear?spring macromolecules. i. steady potential and plane flows*, J. Chem. Phys. **55** (1971), no. 4, 1958–1964, doi:10.1063/1.1676334. 7

[Vla02] V. S. Vladimirov, *Methods of the theory of generalized functions*, Analytical Methods and Special Functions, vol. 6, Taylor & Francis, London, 2002, MR2012831 (2005b:46077). 47

[VS74] I. L. Vulis and M. Z. Solomjak, *Spectral asymptotic analysis for degenerate second order elliptic operators*, Izv. Akad. Nauk SSSR Ser. Mat. **38** (1974), 1362–1392, MR0358081 (50 #10546), Translated in Mathematics of the USSR-Izvestiya **8** (1974), no. 6, 1343–1371. 37, 38

[War72] H. R. Warner, *Kinetic theory and rheology of dilute suspensions of finitely extendible dumbbells*, Ind. Eng. Chem. Fundamentals **11** (1972), no. 3, 379–387, doi:10.1021/i160043a017. 9