.

Habilitation à Diriger les recherches

Spécialité: Mathématiques appliquées

présentée par

Virginie EHRLACHER

Mathematical and numerical analysis of some high-dimensional and multiscale problems in materials science

soutenue le 10 décembre 2020 devant le jury composé de:

Yvon Maday	*Président*
Yves Achdou	*Rapporteur*
Wolfgang Dahmen	*Rapporteur*
Ansgar Jüngel	*Rapporteur*
Eric Cancès	*Examinateur*
Guillaume Carlier	*Examinateur*
Marie Doumic	*Examinateur*
Laura Grigori	*Examinateur*

à Brigitte et Alain, mes parents,
à Charles, mon frère,
à Edouard, mon amour,
à Amélie et Mathieu, mes deux anges.

Je ne sais pas ce qui est beau,
mais je sais ce que j'aime
et je trouve ça amplement suffisant.
Boris Vian

Contents

Remerciements

Je souhaiterais adresser mes premiers remerciements aux rapporteurs de ce manuscrit, Yves Achdou, Wolfgang Dahmen et Ansgar Jüngel. Un grand merci à eux pour leur lecture attentive ainsi que leurs commentaires très encourageants sur ce travail, dans un contexte sanitaire et administratif particulier.

Un grand merci également aux autres membres du jury, Eric Cancès, Guillaume Carlier, Marie Doumic, Laura Grigori, Yvon Maday, qui ont accepté ma demande quand bien même certains me connaissent très peu. Je souhaiterais remercier tout particulièrement Guillaume Carlier d'avoir accepté d'être coordinateur de mon dossier d'habilitation auprès de l'Université Paris-Dauphine, et à Yvon Maday d'avoir accepté de présider le jury.

Je regrette profondément de ne pas avoir l'occasion de discuter science autour d'un tableau (et d'une coupe de champagne ;)) avec chacun d'entre vous à l'occasion de la soutenance, à cause du contexte sanitaire particulier que nous vivons en ce moment, mais j'espère de tout coeur avoir d'autres occasions de pouvoir me rattrapper dans le futur.

Le temps passe vite, très vite, et l'écriture de ce manuscrit pour l'Habilitation à Diriger les Recherches a été l'occasion pour moi de mesurer une fois de plus l'immense chance que j'ai d'avoir pu embrasser une carrière de recherche en mathématiques appliquées, tout particulièrement au CERMICS. Je souhaiterais profiter de cette occasion pour adresser mes plus profonds remerciements aux différentes personnes qui m'ont permis d'avoir cette chance.

Je souhaiterais remercier en tout premier lieu mes deux anciens directeurs de thèse, Eric Cancès et Tony Lelièvre. Je me rappelle nettement le grand plaisir que j'avais à suivre les cours qu'Eric dispensait à l'Ecole Polytechnique, à l'Ecole des Ponts et à l'Université Pierre et Marie Curie. Ce don extraordinaire qu'il a de transmettre des notions difficiles ou subtiles de manière parfaitement limpide a été déterminant dans l'origine de ma vocation. Je me considère comme extrêmement chanceuse d'avoir croisé sa route et voudrais ici lui exprimer toute ma gratitude.

Je suis également très reconnaissante vis-à-vis de Tony pour ses encouragements continuels depuis mes tout premiers pas dans le monde de la recherche, chose si importante en début de carrière. Son soutien m'a été extrêmement précieux, tout particulièrement dans les moments de doute (il y en a toujours!), et je voudrais sincèrement l'en remercier.

Je souhaiterais également ici exprimer une reconnaissance toute particulière à Yvon Maday, pour sa bienveillance constante, ses encouragements continuels et son optimisme sans limite! Merci à lui pour sa confiance dans de nombreuses occasions, en particulier en ce qui concerne l'organisation du CEMRACS 2021, à laquelle je suis très heureuse de contribuer (d'autant plus que j'ai un très joli souvenir en tant qu'étudiante de l'édition 2013, qui avait été à l'époque co-organisée par Tony!). Je suis tout particulièrement heureuse de pouvoir participer aux discussions scientifiques passionantes qui ont annuellement lieu à Roscoff, dans un cadre absolument exceptionnel. J'ai d'ailleurs une pensée toute spéciale pour Annick!

A tous trois, encore une fois, un immense merci. C'est une grande chance de pouvoir côtoyer quotidiennement des chercheurs aussi exceptionnels, tant par leurs qualités scientifiques et humaines que par leur enthousiasme pour la recherche. Nombreux sont les jeunes chercheurs qui rêveraient d'avoir de tels mentors sur qui prendre exemple.

Un grand merci également à Gabriel Stoltz pour sa gestion exemplaire du pôle MAS, pour son engagement total au service du CERMICS et de ses membres, ainsi que de m'avoir fait découvrir l'enseignement en classe inversée.

Je me rappelle d'une très très très forte intervention de Claude Le Bris au cours de ma thèse, qui s'est révélée déterminante dans la poursuite de mon parcours et voudrais lui en témoigner ici de ma gratitude. Un grand merci à lui également en tant que chef de l'équipe INRIA MATHERIALS.

Je souhaiterais également remercier vivement Aurélien Alfonsi, Alexandre Ern, Laura Grigori, Arthur Lebée, Frédéric Legoll et Damiano Lombardi de m'avoir offert l'opportunité de co-encadrer des étudiants et post-doctorants avec eux. Un grand merci à eux pour les passionantes discussions scientifiques que nous avons eues, ainsi qu'à tous mes autres collègues avec qui j'ai eu la chance de collaborer: Martin Burger, Clément Cancès, Laurent Monasse, Pierre Monmarché, Olga Mula, Anthony Nouy, Christoph Ortner, Jan-Frederik Pietschmann, Alexander Shapeev, Benjamin Stamm, François-Xavier Vialard. Je voudrais adresser des remerciements tout particuliers à la "Roscoff team": Geneviève Dusson, David Gontier et Antoine Levitt. Enfin, c'est une grande chance pour moi d'avoir eu l'occasion d'échanger avec de jeunes collègues: un grand merci à Athmane Bakhta, Thomas Boiveau, Amina Benaceur, Jean Cauvin-Vila, Rafaël Coyaud, Adrien Lesage, Idrissa Niakh, Shuyang Xiang.

Enfin, je souhaiterais remercier ici tous mes collègues du CERMICS et d'INRIA avec qui les échanges sont toujours un plaisir: la liste est beaucoup trop longue pour vous citer tous! Une mention toute spéciale pour Isabelle Simunic, secrétaire générale du CERMICS, qui ne ménage jamais ses efforts lorsqu'il s'agit de défendre ses troupes, et le moins que l'on puisse dire, c'est que les occasions de le faire ne manquent pas! Un grand merci également à Stéphanie Bonnel, gestionnaire administrative du CERMICS, et à Julien Guieu, assistant de l'équipe-projet MATHERIALS, pour leur aide si précieuse.

Embrasser une carrière de recherche n'aurait tout simplement pas pu être possible sans l'aide permanente des membres de ma famille, sur qui j'ai toujours pu compter sans faille. Je voudrais dire à mes parents à quel point je leur suis reconnaissante de l'importance et l'attention constante qu'ils ont portées à mon éducation et pour les nombreux sacrifices qu'ils ont réalisés pour que mon frère et moi puissions suivre des études dans les meilleures conditions possibles. Merci à toi Charles pour tous tes encouragements. Edouard, merci à toi pour ton soutien en toutes circonstances et d'être toujours aussi présent pour moi et les enfants. Je voudrais enfin remercier mes deux petits loulous, Amélie et Mathieu, grâce à qui je suis une maman comblée.

Introduction (version française)

Ce mémoire comporte quatre parties.

La première partie porte sur l'étude et l'analyse de méthodes numériques pour l'approximation de problèmes en grande dimension, en particulier de méthodes de tenseur couplées à des algorithmes gloutons pour l'approximation de différentes familles d'équations, tels que les problèmes de minimisation de fonctionnelle convexe [VE3], les problèmes linéaires non symmétriques [VE22] et tout particulièrement les équations paraboliques [VE14], ainsi que des problèmes aux valeurs propres linéaires [VE24]. L'efficacité et les limites de ce type de méthodes sont ensuite illustrées pour plusieurs problèmes en grande dimension, provenant de différentes applications en sciences des matériaux: les équations cinétiques [VE11, VE26], la dynamique moléculaire [VE27] et le calcul de structure électronique [VE25]. Mes co-auteurs sur ce thème sont A. Alfonsi, T. Boiveau, E. Cancès, R. Coyaud, A. Ern, L. Grigori, T. Lelièvre, D. Lombardi, P. Monmarché et A. Nouy.

La deuxième partie du mémoire présente le développement de méthodes de réduction de modèles pour des équations différentielles paramétrées, tout particulièrement de méthodes de Bases Réduites. Le travail [VE21] porte sur le développement d'une nouvelle méthode de réduction de modèles pour des problèmes de transport paramétrés, dont les solutions peuvent être vues comme des mesures de probabilité, en utilisant des barycentres pour la distance de Wasserstein. De nouvelles techniques de réduction de modèles ont été développées dans le cadre d'une collaboration avec Electricité de France (EDF) sur des problèmes d'évolution non-linéaires [VE13] et des inégalités variationnelles avec contraintes non-linéaires [VE17]. Mes co-auteurs sur ce thème sont A. Benaceur, A. Ern, D. Lombardi, S. Meunier, O. Mula, A. Nouy et F.-X. Vialard.

La troisième partie du mémoire traite de nouvelles méthodes numériques pour l'homogénéisation stochastique. Une matrice effective pour un problème de diffusion multi-échelle est calculée en utilisant un problème du correcteur approché, défini sur tout l'espace, où une matrice de diffusion constante est imposée dans la zone d'espace qui se trouve en dehors d'une boule dont le rayon a vocation à tendre vers l'infini [VE18]. Une méthode numérique très efficace basée sur une représentation intégrale de ce problème, une discrétisation en harmoniques sphériques et l'utilisation d'une méthode de *Fast Multipole* est développée dans [VE19], pour des matériaux constitués d'inclusions sphériques de matériau isotrope insérées dans une matrice également constituée d'un matériau isotrope. Mes co-auteurs sur ce thème sont E. Cancès, F. Legoll, B. Stamm et S. Xiang.

La dernière partie du mémoire porte sur l'étude de systèmes de diffusion croisée. Dans [VE16], l'existence et l'unicité de solutions fortes est prouvée pour un système de diffusion croisée particulier, défini sur un domaine fixe avec des conditions de flux nuls, sous l'hypothèse que les coefficients modélisant les propriétés de diffusion entre chaque paire d'espèces ne soient pas trop éloignés les uns des autres. Dans [VE12], un modèle uni-dimensionnel pour la simulation du processus de fabrication de cellules photovoltaïques à couches minces est étudié. Celui-ci s'écrit comme un système de diffusion croisée défini sur un domaine qui évolue au cours du temps. Mes co-auteurs sur ce thème sont A. Bakhta, J. Berendsen, M. Burger et J.-F. Pietschmann.

Je termine cette introduction en signalant des travaux qui ne sont pas résumés dans ce mémoire, car portant sur d'autres thématiques: [VE10] (analyse de conditions aux bords pour la simulation atomistique de défauts dans les cristaux), [VE15, VE20] (calcul de structure de bandes électroniques pour des solides cristallins), [VE23] (méthodes statistiques pour la détection de scénarios critiques en aéronautique). Ces travaux ont été réalisés en collaboration avec H. Alrachid, A. Bakhta, E. Cancès, D. Gontier, A. Levitt, D. Lombardi, C. Ortner, A. Shapeev and K. Tekkal.

Introduction

This manuscript is composed of four parts.

The first part deals with the analysis of numerical methods for the approximation of high-dimensional problems. Some theoretical and numerical results about the use of tensor methods together with greedy algorithms for the approximation of different kinds of problems are presented, in particular for convex minimisation problems [VE3], non-symmetric linear problems [VE22] with a specific focus on parabolic equations [VE14], and linear eigenvalue problems [VE24]. The efficiency and limits of these methods are illustrated on several types of high-dimensional problems, arising in different fields of materials science: kinetic equations [VE11, VE26], molecular dynamics [VE27] and electronic structure calculations [VE25]. My co-authors on this thematic are A. Alfonsi, T. Boiveau, E. Cancès, R. Coyaud, A. Ern, L. Grigori, T. Lelièvre, D. Lombardi and A. Nouy.

The second part focuses on the development of model-order reduction techniques for parametrized partial differential equations, in particular Reduced Basis methods. The work [VE21] aims at developping new model-order reduction methods for parametrized transport-dominated problems, the solutions of which can be seen as probability measures using barycenters for the Wasserstein metric. New reduced basis techniques have been developped within a collaboration with the Electricité de France (EDF) company, on nonlinear evolution problems [VE13] and variational inequalities with nonlinear constraints [VE17]. My co-authors on this topic are A. Benaceur, A. Ern, D. Lombardi, O. Mula and F.-X. Vialard.

The third part of the manuscript deals with new numerical methods for stochastic homogenization. Effective matrices for multiscale diffusion problems are computed using an approximate corrector problem, defined over the whole space, where a constant diffusion matrix is imposed in a region of the space which lies outside of a ball with radius going to infinity [VE18]. A very efficient numerical method based on an integral representation of the problem, a spherical harmonics discretization and the use of a *Fast Multipole Method* is developped in [VE19], for heterogeneous media made of spherical isotropic inclusions embedded in an isotropic matrix. My co-authors on this subject are E. Cancès, F. Legoll, B. Stamm and S. Xiang.

The last part of the manuscript is concerned with the analysis of cross-diffusion systems. In [VE16], the existence and uniqueness of strong solutions is proved for a particular cross-diffusion system, defined on a fixed domain with no-flux boundary conditions, under the assumption that the coefficients encoding the diffusion properties of each pair of species are close. In [VE12], a one-dimensional model for the simulation of the fabrication process of thin film solar cells is proposed and analyzed. This model reads as a cross-diffusion system defined over a time-dependent domain. My co-authors on this thematic are A. Bakhta, J. Berendsen, M. Burger and J.-F. Pietschmann.

Let me end this introduction by mentioning some works which are not summarized in this manuscript, because they are corresponding to topics not covered in the three parts mentioned above: [VE10] (analysis of boundary conditions for the atomistic simulations of crystal defects), [VE15, VE20] (band electronic structure

calculations for crystalline solids), [VE23] (statistical methods for critical scenarios in aeronautics). My co-authors on these works are H. Alrachid, A. Bakhta, E. Cancès, D. Gontier, A. Levitt, D. Lombardi, C. Ortner, A. Shapeev and K. Tekkal.

Publication list

I list here my publications, decomposing them into two categories, depending on whether the material has been produced or substantially initiated during my PhD, or afterwards.

Publications from works completed during the PhD or before

[VE1] *Data mined ionic substitutions for the discovery of new compounds*, Inorganic Chemistry, 50 (2), 2011, pp 656-663 (with G. Hautier, C. Fischer, A. Jain and G. Ceder)

[VE2] *Local defects are always neutral in the Thomas-Fermi-von Weiszäcker theory of crystals*, Archive for Rational Mechanics and Analysis, 202, 2011, pp 933-973 (with E. Cancès)

[VE3] *Convergence of a greedy algorithm for high-dimensional convex problems*, Mathematical Models and Methods in Applied Sciences, 21(12), 2011, pp 2433-2467 (with E. Cancès and T. Lelièvre)

[VE4] *Periodic Schrödinger operators with local defects and spectral pollution*, SIAM Journal of Numerical Analysis, 50(6), 2012, pp 3016-3035 (with E. Cancès and Y. Maday)

[VE5] *Non-consistent approximations of self-adjoint eigenproblems: application to the supercell method*, Numerische Mathematik, 128, 2014, pp 663-706 (with E. Cancès and Y. Maday)

Proceedings from works completed during the PhD

[VE6] *Convergence of a greedy algorithm on nonlinear convex problems and application to uncertainty quantification on obstacle problems*, ASME Proceedings, 3rd Joint US-European Fluids Engineering Summer Meeting, 2010, pp 2905-2912.

[VE7] *Investigation of solar cell properties by absolute measurement of spatially and spectrally resolved luminescence*, in Proceedings of the 27th European Photovoltaics Solar Energy Conference, 2012, pp 497-499 (with A. Delamarre, L. Lombez, J.-F. Guillemoles, T. Lelièvre and E. Cancès)

Publications from works completed after the PhD

[VE8] *Greedy algorithms for high-dimensional eigenvalue problems*, Constructive Approximation, 40, 2014, pp 387-423 (with E. Cancès and T. Lelièvre)

[VE9] *An embedded corrector problem to approximate the homogenized coefficients of an elliptic equation*, Comptes-Rendus Mathématiques, 353(9), 2015, pp 801-806 (with E. Cancès, F. Legoll and B. Stamm)

[VE10] *Analysis of boundary conditions for crystal defect atomistic simulations*, Archive for Rational Mechanics and Analysis, 222(3), 2016, pp 1217-1268, (with C. Ortner and A. Shapeev)

[VE11] *A dynamical adpative tensor method for the resolution of the Vlasov-Poisson system*, Journal of Computational Physics, 339, 2017, pp 285-306 (with D. Lombardi)

[VE12] *Cross-diffusion systems with non-zero flux and moving boundary conditions*, ESAIM: Mathematical Modelling and Numerical Analysis, 52(4), 2018, pp 1385-1415 (with A. Bakhta)

[VE13] *A progressive reduced basis/empirical interpolation method for nonlinear parabolic problems*, SIAM Journal of Scientific Computing, 40(5), 2018, pp A2930-A2955 (with A. Benaceur, A. Ern and S. Meunier)

[VE14] *Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods*, ESAIM: Mathematical Modelling and Numerical Analysis, 53(2), 2019, pp 635-658 (with T. Boiveau, A. Ern and A. Nouy)

[VE15] *Numerical reconstruction of the first band(s) in an inverse Hill's problem*, to appear in ESAIM: Control, Optimisation and Calculus of Variations, 2019 (with A. Bakhta and D. Gontier)

[VE16] *Uniqueness of strong solutions and weak-strong stability in a system of cross-diffusion equations*, Journal of Evolution Equations, 2019, pp 1-25 (with J. Berendsen, M. Burger and J.-F. Pietschmann)

[VE17] *A reduced basis method for parametrized variational inequalities applied to contact mechanics*, to appear in International Journal for Numerical Methods in Engineering, 2019, (with A. Benaceur and A. Ern)

[VE18] *An embedded corrector problem for homogenization. Part I: Theory*, to appear in Multiscale Modeling and Simulation, 20, (with E. Cancès, F. Legoll, B. Stamm and S. Xiang)

[VE19] *An embedded corrector problem for homogenization. Part II: Algorithms and discretization*, Journal of Computational Physics, 2020, pp 109254 (with E. Cancès, F. Legoll, B. Stamm and S. Xiang)

[VE20] *Numerical quadrature in the Brillouin zone for periodic Schrödinger operators*, Numerische Mathematik, 2020, pp 1-48 (with E. Cancès, D. Gontier, A. Levitt and D. Lombardi)

[VE21] *Nonlinear model reduction on metric spaces. Application to one-dimensional conservative PDEs in Wasserstein spaces*, accepted in ESAIM: Mathematical Modelling and Numerical Analysis, 2020, (with D. Lombardi, O. Mula and F.-X. Vialard)

Proceedings from works completed after the PhD

[VE22] *Greedy algorithms for high-dimensional linear non-symmetric problems*, ESAIM: Proceedings, 41, 2013, pp 95-131 (with E. Cancès and T. Lelièvre)

[VE23] *Statistical methods for critical scenarios in aeronautics*, ESAIM: Proceedings, 41, 2014, pp 95-131 (with H. Alrachid, K. Tekkal and T. Lelièvre)

[VE24] *Convergence results on greedy algorithms for high-dimensional eigenvalue problems*, ESAIM: Proceedings and Surveys, 45, 2014, pp 148-157.

Preprints

[VE25] *Approximation of optimal transport problems with marginal moment constraints*, accepted for publication in Mathematics of Computation, 2020 (with A. Alfonsi, R. Coyaud and D. Lombardi), https://hal.archives-ouvertes.fr/hal-02128374

[VE26] *Adaptive hierarchical subtensor partitioning for tensor compression*, 2019 (with L. Grigori, D. Lombardi and H. Song), https://hal.inria.fr/hal-02284456v1

[VE27] *Adaptive force biasing algorithms: new convergence results and tensor approximations of the bias*, 2019 (with T. Lelièvre and P. Monmarché), https://hal.archives-ouvertes.fr/hal-02314426

[VE28] *Existence of weak solutions to a cross-diffusion Cahn-Hilliard type system*, 2020 (with G. Marino and J.F. Pietschmann), https://hal.archives-ouvertes.fr/hal-02888479

Advising activities

Within the CERMICS lab, I had the opportunity to be part of the advising team of several PhD students and post-doctoral fellows:

- Athmane Bakhta, Université Paris-Est PhD entitled *Mathematical Models and Numerical Simulation of Photovoltaic Devices*, defended on the 19th December 2017 (advisors: E. Cancès and T. Lelièvre)
- Amina Benaceur, Université Paris-Est PhD entitled *Reduced Order Modeling in Thermo-mechanics*, defended on the 21st December 2018 (advisor: A. Ern)
- Rafaël Coyaud, Université Paris-Est PhD entitled *Deterministic and Stochastic Numerical Methods for Optimal Transport Problems*, started in September 2017 (advisor: A. Alfonsi)
- Adrien Lesage, Université Paris-Est PhD entitled *Multiscale Numerical Methods for Heterogeneous Plates*, started in September 2017 (advisors: F. Legoll, A. Lebée)

- Mohammed Raed Blel, Université Paris-Est PhD entitled *Model-order Reduction Methods for Stochastic Problems*, started in September 2018 (advisor: T. Lelièvre)
- Idrissa Niakh, Université Paris-Est PhD entitled *Model-order Reduction Methods for Variational Inequalities*, started in September 2019 (advisor: A. Ern)
- Thomas Boiveau, post-doctoral fellow, May 2016-September 2017 (advisors: A. Ern, A. Nouy)
- Shuyang Xiang, post-doctoral fellow, June 2017-June 2019 (advisors: E. Cancès, F. Legoll and B. Stamm)

Chapter 1

Greedy algorithms for high-dimensional problems

High-dimensional problems are ubiquitous in a large variety of applications: molecular dynamics, electronic structure calculation, finance, kinetic models, uncertainty quantification, etc. The aim of this chapter is to present some contributions related to the resolution of such problems.

In this chapter are summarized the contributions [VE22, VE8, VE11, VE25, VE26, VE27, VE14] and research perspectives on tensor methods, particularly on the use of greedy algorithms together with appropriate tensor formats for high-dimensional problems arising in materials science.

Section 1.1 summarizes some of my theoretical contributions to the mathematical analysis of greedy algorithms for high-dimensional problems, namely the works [VE3, VE6, VE8, VE14, VE22]. Three particular fields of interest arising from materials science where high-dimensional problems have to be tackled are considered in [VE11, VE25, VE26, VE27], i.e. kinetic equations [47, 86], molecular dynamics [105] and electronic structure calculations [30]. My contributions to the development of numerical methods for these problems are presented in Section 1.2.

1.1 Analysis of greedy algorithms for high-dimensional problems

A significant part of my research activity was devoted to the mathematical analysis of so-called greedy algorithms together with appropriate tensor formats for the resolution of such high-dimensional equations. The aim of this section is to summarize the contributions [VE3, VE6, VE8, VE14, VE22] on this topic.

A general introduction to the so-called curse of dimensionality and tensor formats is given in Section 1.1.1. Existing theoretical convergence results on greedy algorithms are presented in Section 1.1.2, in particular the results proved in [VE3, VE8, VE22, VE14].

1.1.1 High-dimensional problems and tensor formats

Curse of dimensionality

Standard algorithms cannot be carried out in practice for the resolution of problems involving a large number of variables because of the so-called curse of dimensionality [14].

A way to understand this phenomenon is the following: consider the domain $[0,1]^d$ and a function $u : [0,1]^d \to \mathbb{R}$ with regularity $\mathcal{C}^m$ for some $m \in \mathbb{N}^*$. Assume that one would like to reconstruct the function u from an ensemble of N values $\{u(y_i)\}_{1\leq i\leq N}$ where $y_1, \cdots, y_N \in [0,1]^d$. In this case, it is well-known [49] that if $(y_i)_{1\leq i\leq N}$ are the nodes of a uniform grid of $[0,1]^d$ with mesh size $h > 0$, and if a polynomial reconstruction scheme is used, then

$$\|u - R(u)\|_{L^\infty(\Omega)} \leq Ch^m,$$

where $C > 0$ is a constant independent on h, and $R(u)$ denotes the reconstructed function. Since the number of sample points N scales like h^{-d}, the approximation error reads

$$\|u - R(u)\|_{L^\infty(\Omega)} \leq CN^{-m/d}.$$

Thus, the higher the dimension, the slower the decay rate of the reconstruction error with respect to the number of sample points N. Actually, it is proved in [49] that it is impossible to design reconstruction schemes which would achieve better results. This can be explained in terms of nonlinear width. Let L be a normed space with associated norm $\|\cdot\|_L$ and $K \subset L$. Let us consider continuous maps $E : K \to \mathbb{R}^N$ (encoding) and $R : \mathbb{R}^N \to L$ (reconstruction). The distortion of the pair (E, R) over K is defined as

$$\sup_{u\in K} \|u - R(E(u))\|_L ,$$

i.e., it is the largest error made for all functions $u \in K$ by the encoding-reconstruction scheme. The nonlinear N-width of K is defined as the infimum of the distortion of all pairs of continuous maps (E, R):

$$d_N(K) := \inf_{\left\{\begin{array}{l} E : K \to \mathbb{R}^N \\ R : \mathbb{R}^N \to L \end{array}\right. \text{ continuous}} \sup_{u\in K} \|u - R(E(u))\|_L .$$

Then it is known [49] that in the case when $L = L^\infty([0,1]^d)$ and

$$K = \left\{u \in \mathcal{C}^m([0,1]^d) \mid \forall \alpha \in \mathbb{N}^d,\ |\alpha| \leq m, \|\partial^\alpha u\|_{L^\infty([0,1]^d)} \leq 1\right\}$$

is the unit ball of $\mathcal{C}^m([0,1]^d)$ in a suitable norm, then there exists $c, C > 0$ independent on d such that for all $N \in \mathbb{N}^*$,

$$cN^{-m/d} \leq d_N(K) \leq CN^{-m/d}.$$

In other words, if one wants to approximate a function $u \in \mathcal{C}^m([0,1]^d)$ so that the relative error is lower than a given error threshold, the number N of samples will necessarily scale exponentially with respect to the dimension d.

Prototypical examples of high-dimensional equations

Let $d \in \mathbb{N}^*$ and let $\Omega_1, \dots, \Omega_d$ be open bounded subsets of $\mathbb{R}^{p_1}, \dots, \mathbb{R}^{p_d}$ respectively for some $p_1, \dots, p_d \in \mathbb{N}^*$. Let $\Omega := \Omega_1 \times \cdots \times \Omega_d$ and let V be a Hilbert space of real-valued multivariate functions defined on Ω. For all $1 \leq i \leq d$, let V_i be a Hilbert space of real-valued functions defined on Ω_i.

For all $r_1 \in V_1, \dots, r_d \in V_d$, we denote by

$$r_1 \otimes \cdots \otimes r_d : \left\{ \begin{array}{ccc} \Omega = \Omega_1 \times \cdots \times \Omega_d & \to & \mathbb{R} \\ (x_1, \dots, x_d) & \mapsto & r_1(x_1) \cdots r_d(x_d) \end{array} \right.$$

A function of the form $r_1 \otimes \cdots \otimes r_d$ for some $r_1 \in V_1$, ..., $r_d \in V_d$ is called a *pure tensor product* function. We denote by $V_1 \otimes \cdots \otimes V_d$ the tensor product space of $V_1, \dots, V_d$. More precisley, $V_1 \otimes \cdots \otimes V_d$ is the closure of the set of finite linear combinations of pure tensor products for the canonical tensor norm $\| \cdot \|_\otimes$ defined by

$$\forall r_1 \in V_1, \dots, \forall r_d \in V_d, \quad \|r_1 \otimes \cdots \otimes r_d\|_\otimes := \|r_1\|_{V_1} \cdots \|r_d\|_{V_d},$$

where $\| \cdot \|_{V_i}$ denotes the norm of V_i for all $1 \leq i \leq d$.

We assume in the remainder of the section that $V_1 \otimes \cdots \otimes V_d \subset V$, that this embedding is dense, and consider a function $u \in V$ which will typically be the solution of some high-dimensional differential equation. We give below two prototypical examples of such situations:

Example 1: Let $V = H^1_0(\Omega)$, $V_i = H^1_0(\Omega_i)$ for all $1 \leq i \leq d$, and u be the unique solution in V to the high-dimensional Laplace problem

$$\left\{ \begin{array}{ll} -\Delta u = f, & \text{on } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{array} \right. \tag{1.1}$$

for some $f \in L^2(\Omega)$. Let us point out that u is equivalently the unique solution to the minimization problem

$$u = \operatorname*{argmin}_{v \in V} \mathcal{E}(v)$$

where

$$\forall v \in V, \quad \mathcal{E}(v) := \frac{1}{2} \int_\Omega |\nabla v|^2 - \int_\Omega f v = \frac{1}{2} \|u - v\|_V^2 - \frac{1}{2} \|u\|_V^2. \tag{1.2}$$

Example 2: Let $V = H^1_0(\Omega)$, $V_i = H^1_0(\Omega_i)$ for all $1 \leq i \leq d$, and $(\lambda, u) \in \mathbb{R} \times V$ be a solution to the high-dimensional eigenvalue problem

$$\left\{ \begin{array}{ll} -\Delta u + W u = \lambda u, & \text{on } \Omega, \\ u = 0, & \text{on } \partial\Omega, \end{array} \right. \tag{1.3}$$

for some $W \in L^\infty(\Omega)$.

Tensor formats

A manifestation of the curse of dimensionality is the complexity of the representation of such a function u in the particular case when the spaces V_i are finite-dimensional for all $1 \leq i \leq d$. Then, necessarily, it holds that $V = V_1 \otimes \cdots \otimes V_d$. Let us assume

for the sake of simplicity that there exists $N \in \mathbb{N}^*$ such that dim $V_i = N$ for all $1 \leq i \leq d$, and let $(\psi_1^i, \ldots, \psi_N^i)$ be a basis of V_i. Then, the set of tensorized functions $\left(\psi_{i_1}^1 \otimes \cdots \otimes \psi_{i_d}^d\right)_{1 \leq i_1, \ldots, i_d \leq N}$ forms a basis of V and any element $u \in V$ can then be decomposed in the following form

$$u = \sum_{(i_1, \ldots, i_d) \in \{1, \ldots, N\}^d} \lambda_{i_1, \ldots, i_d} \psi_{i_1}^1 \otimes \cdots \otimes \psi_{i_d}^d$$

for some $(\lambda_{i_1, \ldots, i_d})_{1 \leq i_1, \ldots, i_d \leq N} \in \mathbb{R}^{N^d}$. The complexity of this representation of the function u on the full tensorized basis then scales like N^d, which grows exponentially with the number of variables.

Tensor methods have been a very active field of research for the resolution of high-dimensional equations in the past few years [76, 71, 77, 38]. A significant research effort has been done by the mathematical community in order to develop *tensor formats* whose complexity increases slowly with the number of variables d.

We summarize below the most classical tensor formats used in the litterature and motivate their interest for the approximation of functions depending on a large number of variables.

The set of tensors in canonical format of rank lower than $R \in \mathbb{N}^*$ is defined as the set

$$\mathcal{T}_R^{\text{can}} := \left\{ z = \sum_{k=1}^{R} r_k^1 \otimes \cdots \otimes r_k^d,\ r_k^i \in V_i,\ \forall 1 \leq k \leq R,\ \forall 1 \leq i \leq d \right\}. \tag{1.4}$$

The number of terms R in the expression above is called the canonical rank of the function z. This decomposition can also be found in the literature under the names CANDECOMP or PARAFAC [92]. In the case when $\dim(V_i) = N$ for all $1 \leq i \leq d$, the complexity of the canonical format (1.6) is equal to RdN, which makes the canonical format a very popular choice for the treatment of high-dimensional problems [19]. However, the set $\mathcal{T}_R^{\text{can}}$ is not a weakly closed subset of V as soon as $d \geq 3$ and $R \geq 2$ [46]. This implies that there may not exist a best approximation, i.e. there may not exist a minimizer to the problem

$$\inf_{z \in \mathcal{T}_R^{\text{can}}} \|u - z\|_V.$$

Moreover, $\mathcal{T}_R^{\text{can}}$ is not an embedded manifold. This makes difficult the identification of a tangent space, which is needed in practice for the resolution of high-dimensional partial differential equations.

Other tensor formats exist in the literature to avoid the shortcomings of the canonical format. Among them, the set of tensors in Tucker format with rank $\boldsymbol{R} := (R_1, \cdots, R_d) \in (\mathbb{N}^*)^d$ is defined as follows

$$\mathcal{T}_{\boldsymbol{R}}^{\text{Tucker}} := \left\{ \begin{array}{c} z = \sum_{k_1=1}^{R_1} \cdots \sum_{k_d=1}^{R_d} c_{k_1, \ldots, k_d} r_{k_1}^1 \otimes \cdots \otimes r_{k_d}^d,\ r_{k_i}^i \in V_i,\ \forall 1 \leq k_i \leq R_i,\ \forall 1 \leq i \leq d \\ (c_{k_1, \ldots, k_d})_{1 \leq k_1 \leq R_1, \ldots, 1 \leq k_d \leq R_d} \in \mathbb{R}^{R_1 \times \cdots \times R_d} \end{array} \right\}. \tag{1.5}$$

The set $\mathcal{T}_{\boldsymbol{R}}^{\text{Tucker}}$ is weakly closed in V, which ensures the existence of best approximations, is an embedded manifold [117] with a well-defined tangent space. Thus,

it possesses nice mathematical properties but, unfortunately, its complexity scales exponentially with d. Actually, if $\boldsymbol{R} = (R, R, \cdots, R)$ and if $\dim V_i = N$ for all $1 \leq i \leq d$, the complexity scales as $\mathcal{O}(R^d + NRd)$, which limits the applicability of the Tucker format for very large values of d.

The Tensor Train (TT) format [129] enables one to get rid of this exponential complexity. The set of tensors in Tensor Train format with rank $\boldsymbol{R} := (R_{1,2}, R_{2,3}, \cdots, R_{d-1,d}) \in (\mathbb{N}^*)^{d-1}$ is defined as follows

$$\mathcal{T}_{\boldsymbol{R}}^{\mathrm{TT}} := \left\{ \begin{array}{c} z(x_1, \ldots, x_d) = S_1(x_1)^T M_2(x_2) \cdots M_{d-1}(x_{d-1}) S_d(x_d), \ \forall (x_1, \ldots, x_d) \in \Omega, \\ S_1 \in (V_1)^{R_{1,2}}, \ S_d \in (V_d)^{R_{d-1,d}}, M_i \in (V_i)^{R_{i-1,i} \times R_{i,i+1}}, \ \forall 2 \leq i \leq d-1 \end{array} \right\}. \tag{1.6}$$

Thus, a function z belonging to the set $\mathcal{T}_{\boldsymbol{R}}^{\mathrm{TT}}$ can be seen as a product of vector-valued or matrix-valued univariate functions, and hence is also called in the litterature a *matrix product state*. Again, the set $\mathcal{T}_{\boldsymbol{R}}^{\mathrm{TT}}$ is a weakly closed subset of V and an embedded manifold which possesses a stable local parametrization of its tangent space [81]. Besides, its storage complexity scales as $\mathcal{O}(R^2 N d)$ if $\boldsymbol{R} = (R, R, \cdots, R)$ and if $\dim V_i = N$ for all $1 \leq i \leq d$. This enables one to get rid of the exponential dependence in the dimension of the Tucker format. That is why the TT format is a very popular way for treating high-dimensional problems.

The hierarchical Tucker (HT) format introduced in [78] is a generalization of the TT format, which uses a hierarchical splitting, described by a dimension partition tree. We refer the reader to [76, 71] for an exhaustive review of the different tensor formats and their mathematical properties.

1.1.2 Greedy (or Proper Generalized Decomposition) algorithms

In the continuation of some of my PhD work, some of my contributions are concerned with the mathematical analysis of a class of algorithms to compute tensor approximations of solutions of high-dimensional PDEs. These methods are called greedy algorithms [145] in the field of nonlinear approximation or Progressive Generalized Decomposition (PGD) in the computational mechanics community. They were introduced for the resolution of high-dimensional PDEs in different contexts by Pierre Ladevèze [96], Francesco Chinesta [7] and Anthony Nouy [126].

Dictionnary

Before going further, let us first introduce the definition of *dictionnary* which is used in the sequel.

Definition 1.1.1. *A set $\Sigma \subset V$ is called a dictionary of V if and only if it satisfies the three following conditions:*

(D1) The set Span Σ *is dense in V.*

(D2) For all $\lambda \in \mathbb{R}$ and $z \in \Sigma$, $\lambda z \in \Sigma$.

(D3) Σ is weakly closed in V.

The sets $\mathcal{T}_1^{\text{can}}$, $\mathcal{T}_{\boldsymbol{R}}^{\text{Tucker}}$ and $\mathcal{T}_{\boldsymbol{R}}^{\text{TT}}$ introduced in the preceding section are examples of dictionaries of V, for instance when $V = H_0^1(\Omega)$ and $V_i = H_0^1(\Omega_i)$ for all $1 \leq i \leq d$.

Convex minimization problems

During my PhD, in a joint work with Eric Cancès and Tony Lelièvre, we considered greedy (or PGD) algorithms in the case when u is the unique solution to a minimization problem of the form

$$u = \operatorname*{argmin}_{v \in V} \mathcal{E}(v), \tag{1.7}$$

for some functional $\mathcal{E} : V \to \mathbb{R}$ which is assumed to satisfy the following conditions:

(E1) $\mathcal{E}$ is differentiable and its gradient is Lipschitz on bounded sets of V, i.e. for all $K \subset V$ bounded, there exists $L_K > 0$ such that

$$\forall v, w \in K, \quad \|\nabla \mathcal{E}(v) - \nabla \mathcal{E}(w)\|_V \leq L_K \|v - w\|_V.$$

(E2) the functional $\mathcal{E}$ is strictly convex in V and there exist $\alpha > 0$ and $s > 1$ such that

$$\forall v, w \in V, \quad \mathcal{E}(v) \geq \mathcal{E}(w) + \langle \nabla \mathcal{E}(w), v - w \rangle_V + \frac{\alpha}{2} \|v - w\|_V^s.$$

The functional $\mathcal{E}$ defined in (1.2) in the example of the high-dimensional Laplace equation clearly satisfies assumptions (E1)-(E2). We refer the reader to [VE3] for other examples of PDEs whose solution can be written as the unique solution of a minimization problem of the form (1.7) with (potentially non-quadratic) functionals $\mathcal{E}$ satisfying these conditions.

The *Pure Greedy* algorithm for the approximation of the solution u of (1.7) then reads as follows:

Pure Greedy Algorithm:

- **Initialization:** Set $u_0 := 0$ and $n := 1$.
- **Iteration** $n \geq 1$**:**
  1. Find $z_n \in \Sigma$ solution to
  $$z_n \in \operatorname*{argmin}_{z \in \Sigma} \mathcal{E}(u_{n-1} + z). \tag{1.8}$$
  2. Define $u_n := u_{n-1} + z_n$.
  3. Set $n := n + 1$ and return to Step 1.

Then, the following result holds [101, VE3, 59]:

Theorem 1.1.1. *Let us assume that Σ is a dictionary of V, in the sense of Definition 1.1.1, and that $\mathcal{E} : V \to \mathbb{R}$ satisfies assumptions (E1)-(E2). Then, each iteration of the Pure Greedy algorithm is well-defined, in the sense that for all $n \in \mathbb{N}^*$, there*

always exists at least one solution $z_n \in \Sigma$ to (1.8). Besides, a solution $z_n \in \Sigma$ to (1.8) is nonzero if and only if $u_{n-1} \neq u$. Moreover, the sequence $(u_n)_{n\in\mathbb{N}^*}$ strongly converges in V to u.

Let us point out that Theorem 1.1.1 was proved in the case of the Laplace equation (Example 1) when $d = 2$ and when $\Sigma = \mathcal{T}_1^{\text{can}}$ in [101]. It was proved for general non-quadratic functionals $\mathcal{E}$ satisfying (E1) and (E2) with $s = 2$, $d = 2$ and $\Sigma = \mathcal{T}_1^{\text{can}}$ in [VE3]. The present statement of Theorem 1.1.1 was finally proved in [59]. It is to be noted that it was also proved in [59] that similar convergence results also hold in the case when V is a Banach space.

It was also proved in [VE17] that, in the case when $d = 2$ and when V is a finite-dimensional space, the sequence $(\|u - u_n\|_V)_{n\in\mathbb{N}}$ decays exponentially fast with n.

Similar convergence results can also be obtained for a different version of the greedy algorithm, namely the *Orthogonal Greedy Algorithm*. At the n^{th} iteration of the algorithm, step 2 is modified as follows: instead of defining $u_n := u_{n-1} + z_n$, u_n is defined as the unique minimizer of

$$u_n \in \underset{w\in \text{Span}\{z_k,\, 1\leq k\leq n\}}{\text{argmin}} \mathcal{E}(w).$$

In other words, u_n is defined as the linear combination of all elements $(z_k)_{1\leq k\leq n}$ which minimizes the functional $\mathcal{E}$.

Let us consider the particular case when $\mathcal{E}$ is a quadratic functional. More precisely, let $a : V \times V \to \mathbb{R}$ be a symmetric coercive continuous bilinear form on $V \times V$, $l : V \to \mathbb{R}$ a continuous linear form on V and define

$$\forall v \in V, \quad \mathcal{E}(v) := \frac{1}{2}a(v, v) - l(v). \tag{1.9}$$

The unique solution $u \in V$ to (1.7) is then well-known to be the unique solution to the variational problem

$$a(u, v) = l(v), \quad \forall v \in V,$$

by the Lax-Milgram theorem. Then, $\mathcal{E}$ naturally satisfies (E1) and (E2). Besides, in the case when Σ is an embedded manifold, denoting by $T_z(\Sigma)$ the tangent space to Σ at a point $z \in \Sigma$, the Euler equation associated to (1.8) reads as:

$$a(z_n, \delta z_n) = l(\delta z_n) - a(u_{n-1}, \delta z_n), \quad \forall \delta z_n \in T_{z_n}(\Sigma). \tag{1.10}$$

Non-symmetric linear and parabolic problems

The resolution of non-symmetric linear problems by means of such PGD algorithms is an intricate question. Indeed, consider $u \in V$ the unique solution of

$$a(u, v) = l(v), \quad \forall v \in V, \tag{1.11}$$

for some $a : V \times V \to \mathbb{R}$ continuous coercive bilinear form (not symmetric) and $l : V \to \mathbb{R}$ a continuous linear form on V. Then, u cannot be interpreted in general as the minimizer of the functional $\mathcal{E}$ defined by (1.9) on V. Nevertheless, in

analogy with (1.10) in the case when a is symmetric, the so-called Galerkin-PGD algorithm [126] consists in using a greedy algorithm for the approximation of u where Step 1 of iteration $n \in \mathbb{N}^*$ is replaced by: find $z_n \in \Sigma$ solution to

$$a(z_n, \delta z_n) = l(\delta z_n) - a(u_{n-1}, \delta z_n), \quad \forall \delta z_n \in T_{z_n}(\Sigma). \tag{1.12}$$

This Galerkin-PGD algorithm is very common in the computational mechanics litterature. However, in a joint work with Eric Cancès and Tony Lelièvre [VE22], we were able to identify some counter-examples of problems of the form (1.11) where for some $n \in \mathbb{N}^*$, $u_{n-1} \neq u$ and the only solution to (1.12) is $z_n = 0$. The Galerkin-PGD method thus does not appear as a reliable and mathematically sound numerical method for the resolution of problems of the form (1.11). Let us point out that, in [VE22], were proved some convergence results of greedy-type algorithms in the case where the antisymmetric part of the bilinear form a is assumed to be small in some sense compared to its symmetric part.

However, it is possible to circumvent the difficulty inherent to the non-symmety of linear parabolic problems, as shown in a joint work with Thomas Boiveau, Alexandre Ern and Anthony Nouy [VE14]. Indeed, let

$$V \hookrightarrow L = L' \hookrightarrow V'$$

be a Gelfand triple where V and L are separable real Hilbert spaces respectively equipped with inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_L$, with associated norms $\| \cdot \|_V$ and $\| \cdot \|_L$. The symbol $\hookrightarrow$ represents a dense and continuous embedding. Let $T > 0$ be the time horizon and let $I := (0, T)$ be the time interval. Let $A : I \to \mathcal{L}(V, V')$ be a strongly measurable time-function with values in the Hilbert space of bounded linear operators from V to V'. We assume that the following boundedness and coercivity properties hold true: there exist $0 < \alpha \leq M < +\infty$ such that for almost all $t \in I$,

$$\forall v \in V, \quad \|A(t)v\|_{V'} \leq M\|v\|_V \quad \text{and} \quad \langle A(t)v, v\rangle_{V',V} \geq \alpha\|v\|_V^2.$$

Note that $A(t)$ is not required to be selfadjoint. Let us define the Hilbert–Bochner spaces

$$X := L^2(I; V) \cap H^1(I; V') \quad \text{and} \quad Y := L^2(I; V).$$

Let $f \in Y' = L^2(I; V')$ and $u_0 \in L$. We consider the following parabolic problem [113, 44, 154]: find $u \in X$ such that

$$\begin{cases} \partial_t u(t) + A(t)u(t) = f(t), & \text{in } V', \text{ a.e. } t \in I, \\ u(0) = u_0, & \text{in } L. \end{cases} \tag{1.13}$$

Using

Then, it holds that u can be equivalently characterized as the unique minimizer of

$$u = \operatorname*{argmin}_{v \in X} \mathcal{E}(v), \tag{1.14}$$

where

$$\forall v \in V, \quad \mathcal{E}(v) := \int_0^T \|\partial_t v(t) + A(t)v(t) - f\|_{V'}^2 \, dt + \alpha\|v(0) - u_0\|_L^2$$

is a strongly convex functional on X. The proof of this result follows from arguments similar to the ones presented in [144] and generalizes the results of [148, 58] for the specific case of the heat equation. In addition, it is proved in [VE14] that the set

$$\Sigma := \left\{ z(t) := r(t)s, \text{ for almost every } t \in I,\ s \in V,\ r \in H^1(I) \right\}$$

is a dictionary of X. Thus, the Pure Greedy algorithm introduced in the preceding section can be used to approximate the solution u and is provably convergent using Theorem 1.1.1. We refer the reader to [VE14] for details on numerical tests and discretization spaces used for the approximation of the solution of various types of parabolic equations using formulation (4.1) together with greedy algorithms.

Linear symmetric eigenvalue problems

The work [VE8], done in collaboration with Eric Cancès and Tony Lelièvre, was concerned with the analysis of greedy algorithms for the resolution of symmetric linear eigenvalue problems. More precisely, let us consider two Hilbert spaces V and H, endowed respectively with the scalar products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_H$, such that

(HV) the embedding $V \hookrightarrow H$ is dense and compact.

The associated norms are denoted respectively by $\|\cdot\|_V$ and $\|\cdot\|_H$. Let $a : V \times V \to \mathbb{R}$ be a symmetric continuous bilinear form on $V \times V$ such that

(HA) there exist $\gamma, \nu > 0$, such that

$$\forall v \in V,\ a(v,v) \geq \gamma \|v\|_V^2 - \nu \|v\|_H^2.$$

The bilinear form $\langle \cdot, \cdot \rangle_a$, defined by

$$\forall v, w \in V,\ \langle v, w \rangle_a := a(v,w) + \nu \langle v, w \rangle_H,$$

is a scalar product on V, whose associated norm, denoted by $\|\cdot\|_a$, is equivalent to the norm $\|\cdot\|_V$. It is well-known (see e.g. [136]) that, under the assumptions (HA) and (HV), there exists a sequence $(\psi_p, \mu_p)_{p \in \mathbb{N}^*}$ of solutions to the elliptic eigenvalue problem: find $(\psi, \mu) \in V \times \mathbb{R}$ such that $\|\psi\|_H = 1$ and

$$\forall v \in V, \quad a(\psi, v) = \mu \langle \psi, v \rangle_H, \tag{1.15}$$

such that $(\mu_p)_{p \in \mathbb{N}^*}$ forms a non-decreasing sequence of real numbers going to infinity and $(\psi_p)_{p \in \mathbb{N}^*}$ is an orthonormal basis of H. The work [VE8] focuses on the computation of μ_1, the lowest eigenvalue of a, and of an associated H-normalized eigenvector. Let Σ be a dictionary of V in the sense of Definition 1.1.1. In this work, we propose two greedy algorithms inspired from the Pure Greedy Algorithm presented in Section 1.1.2.

The first algorithm exploits the fact that μ_1 can be seen as the infimum of the Rayleigh quotient associated to a, more precisely that

$$\mu_1 = \inf_{v \in V} \mathcal{J}(v),$$

where

$$\forall v \in V, \quad \mathcal{J}(v) := \begin{cases} \dfrac{a(v,v)}{\|v\|_H^2} & \text{if } v \neq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

For later use, we define $\lambda_\Sigma := \inf_{z\in\Sigma} \mathcal{J}(z)$.

The *Pure Rayleigh Greedy Algorithm (PRaGA)* reads as follows:

Pure Rayleigh Greedy Algorithm:

- **Initialization:** Choose an initial guess $u_0 \in V$ such that $\|u_0\|_H = 1$ and $\lambda_0 := a(u_0, u_0) < \lambda_\Sigma$ and set $n := 1$.
- **Iteration** $n \geq 1$**:**
  1. Find $z_n \in \Sigma$ solution to
  $$z_n \in \operatorname*{argmin}_{z\in\Sigma} \mathcal{J}(u_{n-1} + z). \tag{1.16}$$
  2. Define $u_n := \dfrac{u_{n-1} + z_n}{\|u_{n-1} + z_n\|_H}$ and $\lambda_n := a(u_n, u_n)$.
  3. Set $n := n + 1$ and return to Step 1.

Let us point out here that $\mathcal{J}$ is not a convex functional, so that the analysis of the PRaGA algorithm does not fall into the scope of Theorem 1.1.1.

The second algorithm, called the *Pure Residual Greedy Algorithm (PReGA)*, is based on the use of a residual for problem (1.15) and reads as follows:

Pure Residual Greedy Algorithm:

- **Initialization:** Choose an initial guess $u_0 \in V$ such that $\|u_0\|_H = 1$, let $\lambda_0 := a(u_0, u_0)$ and set $n := 1$.
- **Iteration** $n \geq 1$**:**
  1. Find $z_n \in \Sigma$ solution to
  $$z_n \in \operatorname*{argmin}_{z\in\Sigma} \frac{1}{2}\|u_{n-1} + z\|_a^2 - (\lambda_{n-1} + \nu)\langle u_{n-1}, z\rangle_H. \tag{1.17}$$
  2. Define $u_n := \dfrac{u_{n-1} + z_n}{\|u_{n-1} + z_n\|_H}$ and $\lambda_n := a(u_n, u_n)$.
  3. Set $n := n + 1$ and return to Step 1.

The denomination *Residual* can be justified as follows. It is easy to check that for all $n \in \mathbb{N}^*$, the minimization problem (1.17) is equivalent to the minimization problem: find $z_n \in \Sigma$ such that

$$z_n \in \operatorname*{argmin}_{z\in\Sigma} \|R_{n-1} - z\|_a^2,$$

where $R_{n-1} \in V$ is the Riesz representant in V of the linear form $l_{n-1} : V \ni v \mapsto \lambda_{n-1}\langle u_{n-1}, v\rangle - a(u_{n-1}, v)$.

Then, the following convergence result, proved in [VE8], holds:

Theorem 1.1.2. *Let V and H be separable Hilbert spaces satisfying (HV), Σ a dictionary of V and $a : V \times V \to \mathbb{R}$ a symmetric continuous bilinear form satisfying (HA). The following properties hold for the PRaGA and PReGA:*

1. *All the iterations of the algorithms are well-defined, in the sense that there always exists at least one solution to (1.16) or (1.17).*

2. *The sequence $(\lambda_n)_{n\in\mathbb{N}}$ is non-increasing and converges towards a limit λ which is an eigenvalue of a for the scalar product $\langle\cdot,\cdot\rangle_H$.*

3. *The sequence $(u_n)_{n\in\mathbb{N}}$ is bounded in V and any subsequence of $(u_n)_{n\in\mathbb{N}}$ which weakly converges in V also strongly converges in V towards an H-normalized eigenvector associated with λ. This implies in particular that*

$$d(u_n, F_\lambda) := \inf_{w\in F_\lambda} \|w - u_n\|_a \xrightarrow[n\to+\infty]{} 0,$$

   where F_λ denotes the set of the H-normalized eigenvectors of a associated with λ.

4. *If λ is a simple eigenvalue, then there exists an H-normalized eigenvector w_λ associated with λ such that the whole sequence $(u_n)_{n\in\mathbb{N}}$ converges to w_λ strongly in V.*

Let us point out that orthogonal versions of these algorithms can easily be defined and similar convergence results for them can be proved as well. These convergence results can also be generalized to cases where the injection $V \hookrightarrow H$ is not compact, and in particular when the self-adjoint operator A on H may have some essential spectrum, provided that the initial guess u_0 used in these algorithms satisfies $\lambda_0 := a(u_0, u_0) < \inf \sigma_{\text{ess}}(A)$.

Rates of convergence can also be obtained in the case when V is finite-dimensional using the so-called Łojasiewicz inequality [115].

From a theoretical point of view, the greedy algorithms presented above may not converge towards the lowest eigenvalue μ_1 associated with the bilinear form a. Of course, if the initial guess u_0 is chosen so that $\lambda_0 = a(u_0, u_0) < \mu_2 = \inf_{j\in\mathbb{N}^*}\{\mu_j \mid \mu_j > \mu_1\}$, then the sequences $(\lambda_n)_{n\in\mathbb{N}}$ generated by the greedy algorithms are ensured to converge to μ_1. However, the construction of such an initial guess u_0 in the general case is not obvious.

Research perspectives on the mathematical analysis of tensor methods

Let us mention here some research perspectives related to the mathematical analysis of tensor numerical methods for high-dimensional PDEs.

- The approximation of high-dimensional non-symmetric eigenvalue problems is a very interesting (and challenging!) issue. This is typically of practical interest for criticity calculations in neutronics. The operators of interest (of transport-type) are acting on high-dimensional functions (typically defined on some phase-space domain), and satisfy the assumptions of the Krein-Rutman theorem. Even if there is no Rayleigh quotient for such eigenvalue problems, one could hope that a residual-based algorithm, using residuals of the forward and adjoint eigenvalue problems, might yield some provably convergent scheme.

- Tensor approximation using separation between the space and time variable for time-dependent problems, as studied in [VE14], may be of particular interest for the approximation of the solution of mean field games problems [1]. The adaptation of such methods in this context gives rise to very interesting mathematical problems.

- A more fundamental (and difficult!) theoretical question is the following: can one a priori predict if the solution u of some high-dimensional PDE can be efficiently approximated by some low-rank tensor formats? Very few results of this kind exist in the litterature. Some results can be obtained by making some assumptions on the regularity of the solution u [137]. However, low-rank approximation properties are not directly related to smoothness. Some results exist for the solution of the Laplace problem [43]: in this work, the authors prove (in a nutshell) that, if f can be accurately approximated by low-rank tensor formats, then so can the function u solution to (1.3). Generalizing such kinds of results to more general types of equations is a challenging open problem.

- A related issue related to the problem mentioned above is the quantification of the error of an approximation of the solution of some PDE in a tensorized form. The development of *a posteriori error estimators*, the computation of which should not be too expensive, is a possible way to be able to assess the quality of a tensorized approximation, and I intend to work on this problematic in the future.

1.2 Applications of tensor methods in some materials science problems

The aim of this section is to present the works [VE11, VE25, VE26, VE27] which illutrate the advantages and some limitations of tensor methods for the numerical resolution of some high-dimensional problems arising in materials science.

Three particular fields of interest are considered here: kinetic equations, molecular dynamics and electronic structure calculations. My contributions on these three topics are summarized respectively in Section 1.2.1, Section 1.2.2 and Section 1.2.3.

1.2.1 Kinetic equations

The purpose of kinetic equations [47, 86] is to describe the evolution of dilute particle gases at an intermediate scale between the microscopic scale and the hydrodynamical scale occupying the physical domain $\Omega_x \subset \mathbb{R}^3$. A dilute gase is a system with a large number of particles, for which a description of the position and of the velocity of each particle is irrelevant, but for which the description cannot be reduced to the computation of an average velocity at any time $t \in \mathbb{R}^+$ and any position $x \in \Omega_x$.

For such particle systems, one wants to take into account more than one possible velocity at each point. In these kinetic equations, the state of the particle system is then described (at the statistical level) by a distribution function $f(t, x, v)$ which encodes the probability at time $t \in \mathbb{R}_+$ of finding a particle at the position $x \in \Omega_x$ with velocity $v \in \mathbb{R}^3$. The normalized particle density at time $t \in \mathbb{R}_+$ and point $x \in \Omega_x$ is then defined as

$$\rho(t, x) := \int_{\mathbb{R}^3} f(t, x, v)\, dv. \tag{1.18}$$

The interpretation of $f(t, x, v)$ as a probability density implies that f has to be non-negative and has to satisfy the following normalization condition:

$$\int_{\Omega_x \times \mathbb{R}^3} f(t, x, v)\, dx\, dv = 1, \quad \forall t \in \mathbb{R}_+.$$

Let us denote by $F(t, x) \in \mathbb{R}^3$ the force field acting on the particle system at time $t \in \mathbb{R}_+$ and point $x \in \Omega_x$. Since the function f has to describe the statistical evolution of the system of particles, it has to satisfy the so-called *transport equation*:

$$\partial_t f(t, x, v) + v \cdot \nabla_x f(t, x, v) + F(t, x) \cdot \nabla_v f(t, x, v) = 0 \quad \text{for } (t, x, v) \in \mathbb{R}_+ \times \Omega_x \times \mathbb{R}^3. \tag{1.19}$$

The *Vlasov-Poisson system* is a particular example of kinetic equations where the particles composing the system are assumed to be electrically charged and where the force field F derives from the electrostatic field generated by the particles themselves. Such a system is used in particular for the description of plasmas or in semi-classical models of electron transport in semiconductors and reads:

$$\begin{cases} \partial_t f(t, x, v) + v \cdot \nabla_x f(t, x, v) - \nabla_x U(t, x) \cdot \nabla_v f(t, x, v) = 0, & \text{in } \mathbb{R}_+ \times \Omega_x \times \mathbb{R}^3, \\ -\Delta_x U(t, x) = \rho(t, x) - 1, & \text{in } \mathbb{R}_+ \times \Omega_x, \end{cases} \tag{1.20}$$

where ρ is defined by (1.18) together with appropriate initial and boundary conditions.

The numerical resolution of a system of the form (1.20) is a very challenging task due to the high-dimensionality of the space $\mathbb{R}_+ \times \Omega_x \times \mathbb{R}^3 \subset \mathbb{R}^7$ on which the function f is defined. Three classes of approaches are used in the litterature to tackle this problem from a numerical point of view: particle methods (Particle-In-Cell [66, 24] and Particle-In-Cloud [150]), semi-Lagrangian approaches [42, 33, 93, 28] and full-deterministic Eulerian methods [62]. While Eulerian approaches are appealing to describe the evolution of the unknown quantities of interest, the high dimensionality

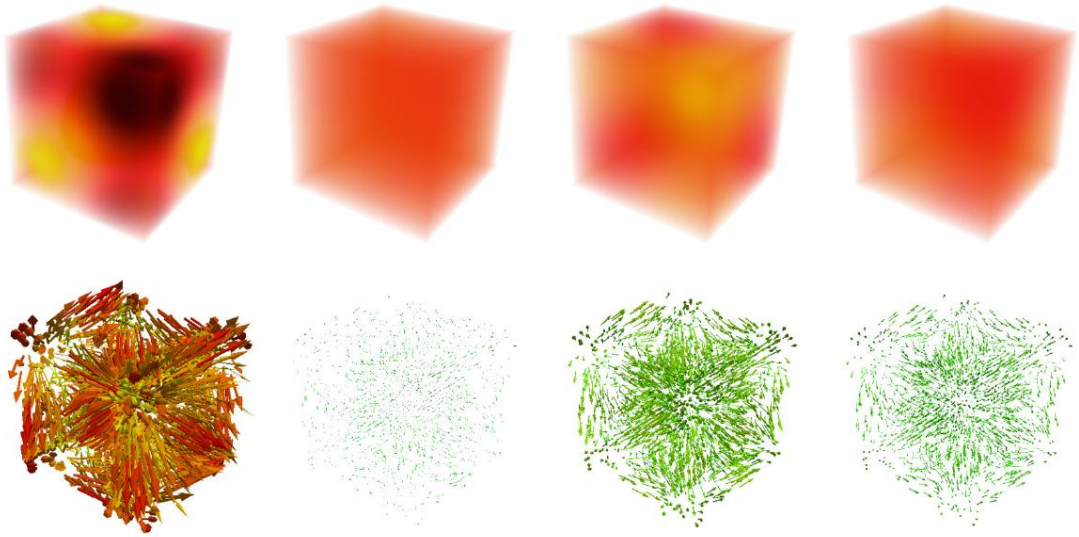

Figure 1.1: Numerical results for a 3D-3D Landau damping test at times $t = 0, 0.33, 0.66, 1.0$ from left to right. Above: Density $1 - \rho(t, x)$. Below: Electric field $E(t, x) = -\nabla_x U(t, x)$.

of the phase space domain makes them often prohibitive in terms of memory and computational cost, especially when 2D-2D and 3D-3D problems are considered.

In [VE11], Damiano Lombardi and I proposed a numerical method to solve the Vlasov-Poisson system (1.20) using a tensorized approximation of the function $f(t, x, v)$ under the form

$$f(t, x, v) \approx \sum_{k=1}^{n_t} r_k(t, x) s_k(t, v). \tag{1.21}$$

Actually, we show in [VE11] that the use of such a tensorised representation of the solution f induces a natural splitting of the system which respects the Hamiltonian nature of the Vlasov-Poisson equations. This leads to the definition of a symplectic time discretization scheme, the different steps of which are solved by tensor methods. More precisely, a modified PGD method based on a well-chosen fixed-point algorithm is proposed to solve the resulting (non-symmetric) equations using tensorised functions at each time step. The convergence of the scheme is proved under restrictions on the size of the time step, which are close to CFL conditions. The proposed method dynamically adapts through time the rank n_t of the decomposition (either increasing or decreasing it). This is an important feature, as was noted in [36, 93], since the number of tensorised terms needed to approximate the solution at a certain time with a given error tolerance is not known a priori.

For the sake of brievity, we do not give all the details of the algorithm proposed in [VE11]. This method enabled us to obtain encouraging results, helped in tremendously accelerating the computations, in particular in situations where the initial condition for f is a low-rank tensor, especially on 3D-3D test cases (see Figure 1.1).

However, in particular for the two-stream instability test case, we observe numerically that the rank n_t of the approximation of the solution f under the form

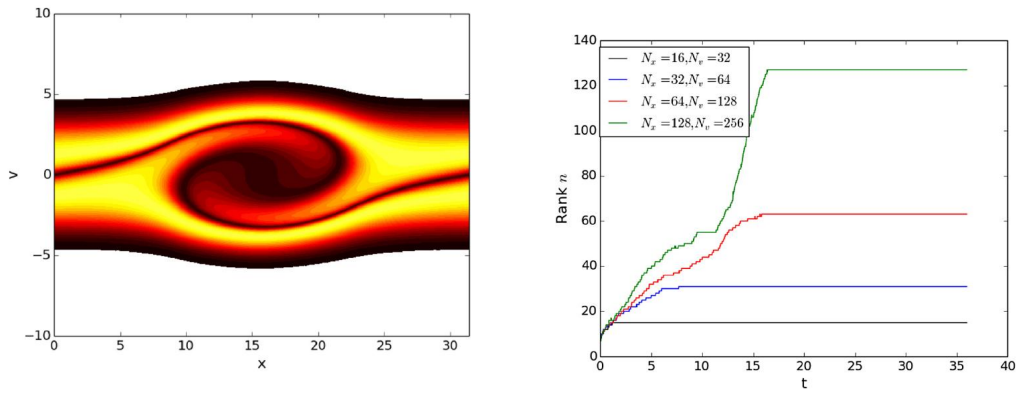

Figure 1.2: 1D-1D double stream instability test case: Left: Contour plots of the reference solution at time $t = 36$. Right: Evolution of the rank n_t as a function of time for different sizes of discretisation grids.

(1.21) given by the algorithm grows with the time t. This is not surprising: indeed the rank of a prescribed accuracy POD approximation of the reference full solution increases also at a similar rate as time increases. Figure 1.2 represents the contour plot of the solution of a standard 1D-1D double stream instability test case and the evolution of the rank n_t with time for different values of the size of the discretization grid used for the space or velocity variable.

Of course, this is not good news for tensor methods with respect to long time simulations of the Vlasov-Poisson system, and motivates the development of new numerical methods to circumvent this difficulty. A potential promising track of research is to find a technique to approximate the solution f of (1.20) using jointly domain decomposition and tensor methods. More precisely, the idea is to partition the phase-space domain into different subdomains, and to introduce different tensor approximations of $f(x, v)$ on each subdomain. Indeed, Figure 1.2 illustrates the fact that this idea is natural: it can be seen that the solution cannot be accurately approximated by a global low-rank tensor, whereas restrictions of the solution to particular subdomains of the phase-space domain can. Let us point out that close ideas were suggested in [89]. Of course, such an approach raises the following question: how should the partition and tensor approximations be chosen (and adapted through the time evolution) so that a prescribed error tolerance is satisfied with the lowest possible computational and memory cost?

The development of a solver for (1.20) relying on these ideas is work in progress. In a joint work with Damiano Lombardi, Laura Grigori and Hao Song [VE26], we first considered the following simpler problem, which can be seen as a first step in this direction. Given some error threshold $\epsilon > 0$ and a particular function $f(x, v)$, we developped a numerical strategy to find a partition of the phase-space domain such that, if the function f is approximated under a tensorized form on each subdomain of this partition, the global error remains lower than ϵ and the memory needed to store the tensor approximations in all the subdomains is as low as possible. It appeared that the efficiency of such an approximation, in terms of memory savings given a prescribed error tolerance, is very sensitive to the way the error is distributed among the different subdomains.

Let us also mention the related works [55, 56]. In these articles, the authors develop elegant numerical schemes in order to approximate in a quasi-optimal way the solution $f(t,x,v)$ of the Vlasov-Poisson system under a tensorized form $\sum_{k=1}^{n} r_k(t,x)s_k(t,v)$ where the rank n has to be chosen a priori and kept fixed during the whole evolution. Their method is related to the so-called Dynamical Orthogonal Decomposition method [91]. The numerical strategy proposed in these works is very elegant in the sense that it enables to capture a quasi-optimal approximation of rank n of the solution through time. However, as mentioned above, keeping a fixed rank approximation can be problematic if one wishes to guarantee some prescribed level of acuracy of the approximation for long-time simulations. A second track of research could be the design of a numerical scheme which should be a combination of both ideas: at each time step of a scheme, a first guess for the new approximation of the solution could be obtained using the strategy of [55, 56] by keeping the rank constant. In a second step, corrections to this guess should be computed to guarantee that the final approximation reaches a given error threshold with respect to the full solution, in a similar way than in [VE11] for instance, which would lead to an adaptation of the rank through time.

A last track of research concerning the approximation of kinetic equations is the design of (potentially tensor-based) numerical schemes for problems with realistic boundary conditions, namely *secular boundary conditions*. Indeed, in [VE11, 55, 56], a tensorized representation of the solution based on the separation of the velocity v and space x variables is possible because the authors consider the approximation of the Vlasov-Poisson system (1.20) on finite-size space and velocity with periodic boundary conditions. However, secular boundary conditions are more realistic boundary conditions which are difficult to treat with tensor methods. Other types of separations have to be thought of, and this could be possibly done on simple toric geometries which would correspond to simplified models of plasma evolutions inside Tokamaks.

1.2.2 Molecular dynamics

In this section are presented some high-dimensional problems arising from molecular dynamics simulations, in particular for the computation of free energies [105, 104].

Let $\mathbb{T} := \mathbb{R}/\mathbb{Z}$. Let us consider a system composed of N (classical) particles, whose positions in space are denoted by $(x_1, \cdots, x_N) =: x \in \mathbb{T}^{3N}$ (assuming periodic boundary conditions) and which interact through a potential function $V : \mathbb{T}^{3N} \to \mathbb{R}$. In the NVT canonical ensemble, the positions of the particles are distributed in space according to the Boltzmann-Gibbs probability measure:

$$d\mu(x) := \frac{1}{Z} e^{-\beta V(x)}\, dx$$

where $Z := \int_{\mathbb{T}^{3N}} e^{-\beta V(x)}\, dx$ and $\beta = \frac{1}{k_B T}$ with k_B the Boltzmann constant and T the temperature. One of the main objectives of molecular dynamics simulations is to compute averages of the form

$$\int_{\mathbb{T}^{3N}} \phi(x)\, d\mu(x), \tag{1.22}$$

for some functions $\phi : \mathbb{T}^{3N} \to \mathbb{R}$. Indeed, *macroscopic quantities* such as pressure fields or likelihoods of molecular configurations can be expressed as averages of the form (1.22) for particular functions ϕ.

The compution of averages of the form (1.22) is a very high-dimensional problem which is currently tackled by Monte-Carlo methods, based on the use of Markov chains. One possible method to sample the measure μ is to consider the solution X_t of the stochastic differential equation

$$dX_t = -\nabla V(X_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t. \tag{1.23}$$

The dynamics (1.23) is called the *overdamped Langevin* dynamics. Under appropriate assumptions on the potential V, the following ergodicity property holds: for μ-almost all initial conditions $X_0 \in \mathbb{R}^{3N}$,

$$\lim_{t\to+\infty} \frac{1}{t} \int_0^t \phi(X_s)\, ds = \int_{\mathbb{T}^{3N}} \phi(x)\, d\mu(x). \tag{1.24}$$

Unfortunately, the convergence (1.24) as t goes to infinity is extremely slow due to the *metastability* of the process X_t. Indeed, in practice, the process X_t remains trapped during very long times in a local minimum of the potential function V, and thus cannot explore all the sets of local minima of the potential function, especially in high dimension, in a reasonable amount of simulation time.

Several methods have been proposed in the litterature to overcome this metastability problem. In the sequel, one of these methods, namely the Adaptive Biasing Force (ABF) method, is explained in more details. This method relies on the assumption that a few *slow* coordinates of the system are known and can be used to bias the dynamics (1.23) in an efficient way. More precisely, let us assume that such slow variates, called hereafter *reaction coordinates*, are encoded by a map $\xi : \mathbb{T}^{3N} \to \mathbb{T}^p$ with $p \ll N$, which is assumed to be smooth in the sequel. In practice, ξ is a collection of geometrical quantities such as the distance between two (groups of) atoms of the molecular system, or angles between pairs of bonds. Of course, the choice of an appropriate map ξ is delicate and requires some a priori knowledge on the physical properties of the molecular system. In the following, we assume that the map ξ is given and fixed.

To these reaction coordinates ξ is associated the corresponding free energy $A : \mathbb{T}^p \to \mathbb{R}$, given by

$$\forall z \in \mathbb{T}^p, \quad A(z) := -\frac{1}{\beta} \ln \left(\int_{\Sigma(z)} e^{-\beta V(x)} \delta_{\xi(x)-z}(\, dx) \right), \tag{1.25}$$

where $\delta_{\xi(x)-z}(\, dx)$ is the so-called delta measure, which can be defined from the Lebesgue measure on the submanifold

$$\Sigma(z) := \{x \in \mathbb{T}^{3N}, \quad \xi(x) = z\}$$

through the co-area formula, see for example [105]. This definition ensures that, if X is a random variable with law μ on $\mathbb{T}^{3N}$, then $\xi(X)$ is a random variable with law

$\frac{1}{Z_A} e^{-\beta A(z)}\, dz$ on $\mathbb{T}^p$ with $Z_A := \int_{\mathbb{T}^p} e^{-\beta A(z)}\, dz$. The heuristic of the ABF algorithm is the following. If we were to sample from the process

$$dY_t = -\nabla (V - A \circ \xi)(Y_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t,$$

the equilibrium measure would be $\frac{1}{\tilde{Z}} e^{-\beta(V(x) - A\circ\xi(x)}\, dx$ with $\tilde{Z} = \int_{\mathbb{T}^{3N}} e^{-\beta(V(x) - A\circ\xi(x)}\, dx$. The image of this measure through ξ, by definition of A, is the uniform measure on $\mathbb{T}^p$. This means that there would be no more metastability along ξ, since all the regions of $\mathbb{T}^p$ would be equally visited by $\xi(Y_t)$. Unfortunately, it is not possible to use directly this free-energy biased dynamics in practice, since it would require the knowledge of A and thus the computation of expectations of the form (1.25) in large dimension. The idea of the ABF method is to learn A on the fly, i.e. to run a process $\tilde{X}_t$ solving

$$d\tilde{X}_t = -\nabla \left(V - A_{\mathrm{bias},t} \circ \xi\right)(\tilde{X}_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t, \tag{1.26}$$

with a biased free energy $A_{\mathrm{bias},t} : \mathbb{T}^p \to \mathbb{R}$ constructed from $\left(\tilde{X}_s\right)_{0\leq s<t}$ and designed to target A in the longtime limit.

In practice, the choice of good reaction coordinates is a difficult problem. Up to recently, their definition has been based on the knowledge and intuition of experts. The question of the automatic learning of suitable reaction coordinates is currently a vivid research area, see for instance [125] for a review on the latest works in this direction. Moreover, some techniques like the orthogonal space random walk [116] provide a general way to construct new reaction coordinates from previous ones. Due to these recent progresses, one would like to consider a relatively large value of p. From a numerical point of view, since $A_{\mathrm{bias},t}$ is adaptively learned on the fly, the values of the latter function have to be kept in memory, which requires a grid whose size typically scales exponentially with p. This limits the application of ABF to small dimensional reaction coordinates. The aim of the work [VE27], done in collaboration with Tony Lelièvre and Pierre Monmarché, is to lift this limitation by approximating $A_{\mathrm{bias},t}$ using a tensor product of decomposition, which significantly reduces the size of the memory needed to store this approximation.

The obtained method, named *Tensorized Adaptive Biasing Force* (TABF), gives very satisfactory results and numerical tests show that this algorithm is able to recover non-trivial correlations between reaction coordinates. Let me present some numerical results obtained on a particular test case. The system considered here is constituted of two types of particles, solvent particles and polymer particles (see Figure 1.3).

In a two-dimensional periodic box, we consider 100 particles among which $d = 5$ form a polymer and the others are solvent particles. Each pair of particles that involves at least one solvent particle interacts through the purely repulsive WCA pair potential, which is the Lennard-Jones potential truncated at its minimum. The polymer particles interact through a potential to form a ring. More precisely, each pair of consecutive particles in the polymer ring interacts through a double well potential. The minimum of this potential is attained at two values of the distance between two polymer particles $r_1 < r_2$. Denoting by r the distance between two consecutive polymer particles, if $r = r_1$, the two particles are said to be in a so-called

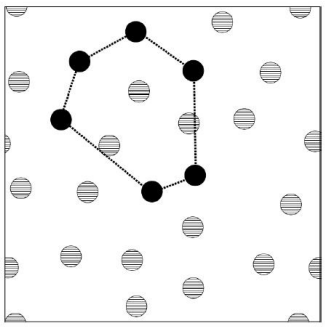

Figure 1.3: The slow motions of the system are the transitions of each bond between two consecutive particles of the polymer ring from its compact state to its stretched state.

compact state, and if $r = r_2$, the particles are said to be in a *stretched* state. Finally, each triplet of consecutive particles in the polymer also interacts through the angle they form with the potential in order to favor the value of one particular equilibrium angle that ensures that the total angular potential is minimized when the polymer particles form a regular pentagon. We refer the reader to [VE27] for the precise expression of these interaction potentials. We chose d reaction coordinates, which are the distances between two consecutive polymer particles.

Figure 1.4 represents the cumulative one-dimensional histograms of the five reaction coordinates at $t = 50$ for the TABF algorithm and for a non-biased process. We clearly see from Figure 1.4 that the values of the different reaction coordinates are much more efficiently sampled by the process obtained by the TABF algorithm than with a non-biased process.

Let me end this section on the interest of tensor approximations for molecular dynamics simulation by mentioning one perspective of research we are currently working on with Tony Lelièvre and Raed Blel(PhD student at CERMICS). In practice, the potential which describes the interactions between the atoms of the molecular system depends on a set of parameters $\mu \in \mathbb{R}^n$ for some $n \in \mathbb{N}^*$, whose precise values are not known a priori. Let us denote by $V_\mu : \mathbb{T}^{3N} \to \mathbb{R}$ the interaction potential corresponding to the set of parameters μ. It is then particularly important with respect to practical applications to find numerical methods in order to efficiently sample $\frac{1}{Z_\mu} e^{-\beta V_\mu(x)}\, dx$ with $Z_\mu := \int_{\mathbb{T}^{3N}} e^{-\beta V_\mu(x)}\, dx$. A promising idea would be to approximate the solution of the SDE

$$dX_t^\mu = -\nabla V_\mu(X_t^\mu)\, dt + 2\beta^{-1}\, dW_t, \tag{1.27}$$

using a decomposition of the form

$$X_t^\mu \approx \sum_{k=1}^{K} r_k(t;\mu) Z_t^k \tag{1.28}$$

for some low value of the rank $K \in \mathbb{N}^*$, and for all $1 \leq k \leq K$, some parameter-independent stochastic process $\left(Z_t^k\right)_{t\geq 0}$ and time-dependent real-valued functions $r_k(t;\cdot) : \mathbb{R}^n \to \mathbb{R}$ defined on the parameter domain. The Dynamical Orthogonal method [91, 32] could be a potential way to obtain an approximation of the

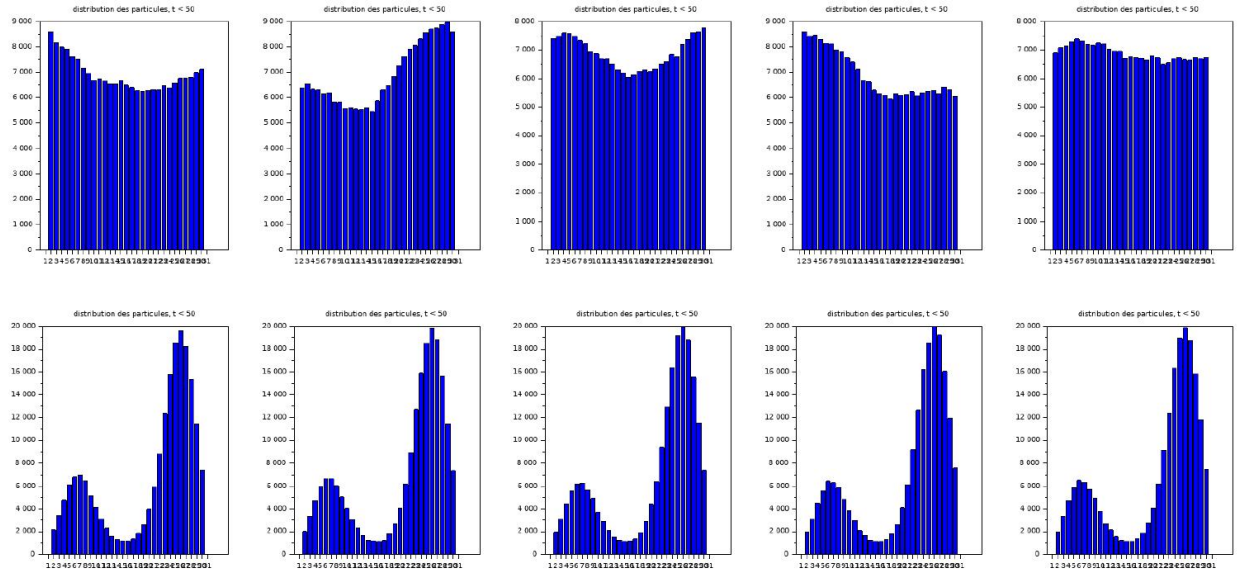

Figure 1.4: Cumulative one-dimensional histograms of the five reaction coordinates at $t = 50$ for the TABF algorithm (up) and for a non-biased process (down).

form (1.28). The adaptation of this method to this context leads to interesting mathematical issues.

1.2.3 Electronic structure calculations

Tensor methods for the electronic Schrödinger problem

The electronic Schrödinger problem for the computation of the electronic structure of molecules is another example of high-dimensional problem arising in materials science. In the sequel, atomic units are used, for which

$$\hbar = 1, \quad e = 1, \quad m_e = 1, \quad 4\pi\epsilon_0 = 1,$$

where $\hbar$ is the reduced Planck constant, e the elementary charge, m_e the mass of the electron, and ϵ_0 the dielectric permittivity of the void. For the simplicity of exposition, the spin variable is omitted here.

In the Born-Oppenheimer approximation, a molecule is a system composed of

- $M \in \mathbb{N}^*$ nuclei, which are considered as classical point-like particles, whose positions are denoted by $R_1, \cdots, R_M \in \mathbb{R}^3$ and electrical charges by $Z_1, \cdots, Z_M \in \mathbb{N}^*$;
- N electrons, which are modeled as quantum particles, and whose state is described by a function

$$\psi : \left\{ \begin{array}{ccc} \mathbb{R}^{3N} & \to & \mathbb{C} \\ (x_1, \ldots, x_N) & \mapsto & \psi(x_1, \ldots, x_N), \end{array} \right.$$

called the *wavefunction* of the system of electrons.

The physical interpretation of a wavefunction ψ is the following: given $A \subset \mathbb{R}^{3N}$, $\int_A |\psi|^2$ represents the probability that the positions of the N electrons belong to the set A. In particular, this implies that $\|\psi\|^2_{L^2(\mathbb{R}^{3N})} = 1$. In addition, the wavefunction ψ is *antisymmetric* with respect to its variables. This is a consequence of the fact that the electrons are fermionic particles. More precisely, denoting by $\mathcal{S}_N$ the set of permutations of the set $\{1, \dots, N\}$, it holds that for all $p \in \mathcal{S}_N$ and all $(x_1, \dots, x_N) \in \mathbb{R}^{3N}$,

$$\psi(x_{p(1)}, \dots, x_{p(N)}) = \epsilon(p)\psi(x_1, \dots, x_N),$$

where $\epsilon(p)$ denotes the signature of p.

The energy $E[\psi]$ of a system of N electrons whose state is described by a wavefunction ψ in the molecule described above is the sum of three contributions:

- the kinetic energy:
$$T[\psi] := \frac{1}{2} \int_{\mathbb{R}^{3N}} |\nabla \psi|^2;$$

- the Coulomb energy associated to the interactions between the electrons and the nuclei:
$$C_{\text{nuc}}[\psi] := \int_{\mathbb{R}^{3N}} \left(\sum_{i=1}^{N} V_{\text{nuc}}(x_i) \right) |\psi(x_1, \dots, x_N)|^2 \, dx_1 \cdots dx_N,$$
where, for all $x \in \mathbb{R}^3$,
$$V_{\text{nuc}}(x) := -\sum_{k=1}^{M} \frac{Z_k}{|x - R_k|};$$

- the Coulomb energy associated to the interactions between the electrons:
$$C_{\text{elec}}[\psi] := \int_{\mathbb{R}^{3N}} c(x_1, \dots, x_N) |\psi(x_1, \dots, x_N)|^2 \, dx_1 \dots dx_N,$$
where for almost all $(x_1, \dots, x_N) \in \mathbb{R}^{3N}$,
$$c(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}.$$

Computing a ground state of the electrons in the molecule amounts to computing a wavefunction ψ_0 among all admissible wavefunctions which minimize the energy of the system. More precisely, let us denote by

$$\mathcal{A} := \left\{ \psi \in L^2(\mathbb{R}^{3N}),\ \nabla \psi \in L^2(\mathbb{R}^{3N})^{3N},\ \psi \text{ antisymmetric},\ \|\psi\|_{L^2(\mathbb{R}^{3N})} = 1 \right\}$$

the set of wavefunctions associated to a system of N electrons with finite kinetic energy. Then, it holds that

$$E_0 = \min_{\psi \in \mathcal{A}} T[\psi] + C_{\text{nuc}}[\psi] + C_{\text{elec}}[\psi]. \tag{1.29}$$

Let $H := -\frac{1}{2}\Delta + \sum_{i=1}^N V_{\rm nuc}(x_i) + c(x_1, \cdots, x_N)$ be the so-called *many-body Schrödinger operator*. The operator H is a self-adjoint, bounded from below, operator on

$$L^2_{\rm antisym}(\mathbb{R}^{3N}) := \{\psi \in L^2(\mathbb{R}^{3N}),\ \psi \text{ antisymmetric}\}$$

with domain

$$H^2_{\rm antisym}(\mathbb{R}^{3N}) := \{\psi \in H^2(\mathbb{R}^{3N}),\ \psi \text{ antisymmetric}\}.$$

We also denote by

$$H^1_{\rm antisym}(\mathbb{R}^{3N}) := \{\psi \in H^1(\mathbb{R}^{3N}),\ \psi \text{ antisymmetric}\}.$$

In the case when $E_0 := \inf \sigma(H)$ is a discrete eigenvalue of H (which occurs for instance when the molecule is neutral or positively charged from Zhislin's theorem [158]), there exists at least one minimizer ψ_0 to (1.29), and any minimizer is necessarily an eigenvector of H associated to the eigenvalue E_0. Thus, solving the electronic Schrödinger problem amounts to solving a linear *high-dimensional* eigenvalue problem of the form

$$H\psi_0 = E_0\psi_0.$$

In view of Section 1.1.2, it is natural to consider tensor approximations of the function ψ_0 using greedy algorithms for linear eigenvalue problems. Indeed, problem (1.29) can be rewritten equivalently as

$$\psi_0 \in \operatorname*{argmin}_{\psi \in H^1_{\rm antisym}(\mathbb{R}^{3N})} \frac{a(\psi,\psi)}{\|\psi\|^2_{L^2(\mathbb{R}^{3N})}}, \tag{1.30}$$

where

$$\forall \psi, \zeta \in H^1_{\rm antisym}(\mathbb{R}^{3N}), \quad a(\psi,\zeta) := \frac{1}{2}\int_{\mathbb{R}^{3N}} \nabla\overline{\psi}\cdot\nabla\zeta + \int_{\mathbb{R}^{3N}} (W_{\rm nuc} + c)\,\overline{\psi}\zeta,$$

where $W_{\rm nuc}(x_1, \ldots, x_N) := \sum_{i=1}^N V_{\rm nuc}(x_i)$ for all $(x_1, \ldots, x_N) \in \mathbb{R}^{3N}$.

The antisymmetry of the wavefunction has to be taken into account in the definition of appropriate dictionaries for this problem.

The set of Slater determinants is the antisymmetric version of the set of rank-1 canonical tensors (or pure tensor products) $\mathcal{T}_1^{\rm can}$. The set of Slater determinant functions is defined more precisely as

$$\mathcal{S} := \left\{S_\Phi, \quad \Phi := (\phi_1, \cdots, \phi_N) \in H^1(\mathbb{R}^3)^N\right\},$$

where for all $\Phi := (\phi_1, \cdots, \phi_N) \in H^1(\mathbb{R}^3)^N$, the Slater determinant $S_\Phi \in H^1_{\rm antisym}(\mathbb{R}^{3N})$ is defined by

$$S_\Phi(x_1, \cdots, x_N) := \frac{1}{\sqrt{N!}}\det\begin{pmatrix} \phi_1(x_1) & \phi_1(x_2) & \cdots & \phi_1(x_N) \\ \phi_2(x_1) & \phi_2(x_2) & \cdots & \phi_N 2(x_N) \\ \vdots & \ddots & \ddots & \vdots \\ \phi_N(x_1) & \phi_N(x_2) & \cdots & \phi_N(x_N) \end{pmatrix}.$$

The set $\mathcal{S}$ then defines a dictionary for $V := H^1_{\text{antisym}}(\mathbb{R}^{3N})$.

The analysis of greedy algorithms using one of these formats for problem (1.30) then falls under the scope of the extension of Theorem 1.1.1 in the case when the injection $V := H^1_{\text{antisym}}(\mathbb{R}^{3N}) \hookrightarrow H := L^2_{\text{antisym}}(\mathbb{R}^{3N})$ is not compact, which was proved in [VE8]. It appears that such algorithms were earlier proposed in the chemistry litterature in [90, 146, 70] without any mathematical analysis.

Unfortunately, for strongly correlated systems, where the influence of the electron-electron Coulomb interactions is more significant than the other contributions to the energy of the system, the convergence of such greedy algorithms is very slow. Such algorithms are thus not competitive with other methods that are widely used in quantum chemistry, such as *Coupled-Cluster* [11] methods for instance.

Let us mention that an alternative formulation of problem (1.30), which is called the *second quantization* formulation, which we do not describe in full details here for the sake of brievity, has been used in several works in conjunction with tensor methods [143]. Such techniques yield very interesting results but this second quantization formulation suffers from the fact that its definition requires the use of a particular orthonormal basis of $L^2(\mathbb{R}^3)$ and depends on this particular choice. The tensor approximation suggested in [143], and thus its accuracy, then depends on the choice of this particular basis. How to choose such a basis in an optimal way remains an open question, at least up to my knowledge.

Semi-classical limit of the Lévy-Lieb functional

The poor quality of the approximation of ψ_0 given by greedy algorithms, and the dependence of the second quantization formulation of problem (1.30) on the choice of a particular orthonormal basis of $L^2(\mathbb{R}^3)$, motivated me to consider a different approach to tackle the approximation of (1.30) for strongly correlated systems, which relies on the so-called *Density Functional Theory (DFT)*. For any $\psi \in \mathcal{A}$, we denote by ρ_ψ the electronic density of the system of electrons characterized by the wavefunction ψ, which is defined by

$$\rho_\psi(x) := N \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, \ldots, x_n)|^2 \, dx_2 \cdots dx_N.$$

The principle of DFT, and of all the models which are derived from it, is the reformulation of problem (1.29) with the density (and no longer the wavefunction) as the main variable. The key advantage of this method is that optimization problems are then formulated with functions defined over the domain $\mathbb{R}^3$ instead of $\mathbb{R}^{3N}$. The theoretical justification of this approach was first provided by Hohenberg and Kohn [80] and was later complemented by Levy [109] and Lieb [111]. Indeed, the Density Functional Theory states that the energy E_0 and the associated electronic density of the ground state of the electronic problem can be found by solving a problem of the form

$$E_0 = \inf \left\{ F_{LL}(\rho) + \int_{\mathbb{R}^3} \rho V, \quad \rho \in \mathcal{I}_{\mathcal{N}} \right\},$$

where $\mathcal{I}_N := \left\{\rho \in L^1(\mathbb{R}^3),\ \rho \geq 0,\ \sqrt{\rho} \in H^1(\mathbb{R}^3),\ \int_{\mathbb{R}^3} \rho = N\right\}$ and where for all $\rho \in \mathcal{I}_N$, the so-called *Lévy-Lieb functional* is defined as

$$F_{LL}(\rho) := \left\{T[\psi] + C_{\text{elec}}[\psi],\ \psi \in \mathcal{A},\ \rho_\psi = \rho\right\}.$$

This is a very appealing theory, but unfortunately, the exact computation of $F_{LL}(\rho)$ is out-of-reach since it requires the resolution of a problem almost as complex as the original electronic Schrödinger problem.

In practice then, approximations of the functional F_{LL} are used, which gives rise to a wide zoology of DFT models. One of this approximation, which was suggested by theoretical chemists in [139, 140], consists in considering the *semi-classical* limit of the Lévy-Lieb functional, with a view to use it in order to design approximate DFT models for strongly correlated systems. This semi-classical limit is the limit as α goes to 0 to the functional F^α_{LL} defined as follows for $\rho \in \mathcal{I}_N$ and $0 < \alpha \leq 1$:

$$F^\alpha_{LL}(\rho) := \left\{\alpha T[\psi] + C_{\text{elec}}[\psi],\ \psi \in \mathcal{A},\ \rho_\psi = \rho\right\}.$$

In this semi-classical limit, the influence of the kinetic term $T[\psi]$ is then neglected in front of the contributions due to the electron-electron Coulombic interaction term $C_{\text{elec}}[\psi]$. It has been rigorously proven in the series of works [41, 40, 110] that the limit as α goes to 0 of the functional $F^\alpha_{LL}(\rho)$ reads as a *symmetric multi-marginal optimal transport problem with Coulomb cost*. More precisely, for all $\rho \in \mathcal{I}_N$, let us denote by ν_ρ the probability measure on $\mathbb{R}^3$ defined by $d\nu_\rho(x) := \frac{\rho(x)}{N}\,dx$ and by $\mathcal{P}_{\text{sym}}(\mathbb{R}^{3N})$ the set of symmetric probability measures on $\mathbb{R}^{3N}$. For all $\gamma \in \mathcal{P}_{\text{sym}}(\mathbb{R}^{3N})$, we denote by μ_γ the probability measure on $\mathbb{R}^3$ defined as the marginal of γ, i.e.

$$d\mu_\gamma(x) := \int_{(x_2,\dots,x_N)\in\mathbb{R}^{3(N-1)}} d\gamma(x, x_2, \dots, x_N).$$

Then, it holds that [41, 40, 110],

$$\lim_{\alpha\to 0} F^\alpha_{LL}(\rho) = I(\nu_\rho),$$

where for all probability measure ν on $\mathbb{R}^3$,

$$I(\nu) := \inf_{\substack{\gamma \in \mathcal{P}_{\text{sym}}(\mathbb{R}^{3N}),\\ \mu_\gamma = \nu}} \int_{\mathbb{R}^{3N}} c\,d\gamma. \tag{1.31}$$

A classical way to approximate the problem (1.31) is to use a (fixed) discrete state space $\{y_1, \dots, y_M\} \subset \mathbb{R}^3$ for some $M \in \mathbb{N}^*$ and compute an approximation of a solution γ to (1.31) under the form

$$\gamma \approx \sum_{1\leq m_1,\dots,m_N\leq M} \lambda_{m_1,\dots,m_N}\delta_{(y_{m_1},\dots,y_{m_N})}$$

where the M^N real coefficients $(\lambda_{m_1,\dots,m_N})_{1\leq m_1,\dots,m_N\leq M}$ have to be determined. This leads to a very high-dimensional linear optimization problem.

In a joint work with Aurélien Alfonsi, the PhD student Rafaël Coyaud and Damiano Lombardi [VE25], we considered an alternative way to approximate the

symmetric optimal transport problem (1.31). In this approach, we still consider a continuous state space $\mathbb{R}^3$, but the marginal contraint appearing in (1.31) is relaxed into a finite number of moment constraints. For the sake of simplicity, let us present our results here in the case when the support of the measure ν is included in a compact set $Y \subset \mathbb{R}^3$. Let $(f_m)_{m\in\mathbb{N}^*} \subset \mathcal{C}(Y)$, satisfying the following natural density assumption

$$\forall f \in \mathcal{C}(Y), \qquad \inf_{g_M \in \mathrm{Span}\{f_1,\dots,f_M\}} \|f - g_M\|_{L^\infty} \underset{M\to+\infty}{\longrightarrow} 0,$$

and consider the approximate moment constrained optimal transport problem

$$I^M(\nu) := \inf_{\substack{\gamma \in \mathcal{P}_{\mathrm{sym}}(\mathbb{R}^{3N}),\\ \forall 1 \le m \le M,\\ \int_{\mathbb{R}^{3N}} \left(\frac{1}{N}\sum_{i=1}^N f_m(x_i)\right) d\gamma(x_1,\dots,x_N) = \int_{\mathbb{R}^3} f_m \, d\nu}} \int_{\mathbb{R}^{3N}} c \, d\gamma. \tag{1.32}$$

Then, the result is proved in [VE25], where $\mathcal{P}(\mathbb{R}^{3N})$ denotes the set of (not necessarily symmetric) probability measures on $\mathbb{R}^{3N}$.

Theorem 1.2.1. *Under the preceding assumptions, it holds that*

$$I^M(\nu) \underset{M\to+\infty}{\longrightarrow} I(\nu).$$

Besides, it holds that

$$I^M(\nu) = \inf_{\substack{\gamma \in \mathcal{P}(\mathbb{R}^{3N}),\\ \forall 1 \le m \le M,\\ \int_{\mathbb{R}^{3N}} \left(\frac{1}{N}\sum_{i=1}^N f_m(x_i)\right) d\gamma(x_1,\dots,x_N) = \int_{\mathbb{R}^3} f_m \, d\nu}} \int_{\mathbb{R}^{3N}} c \, d\gamma, \tag{1.33}$$

and there exists at least one minimizer $\gamma^M \in \mathcal{P}(\mathbb{R}^{3N})$ to (1.33) which reads as

$$\gamma^M = \sum_{k=1}^K w_k \delta_{(x_1^k,\dots,x_N^k)}$$

for some $1 \le K \le M+2$, and for some $w_k \ge 0$ and $(x_1^k,\dots,x_N^k) \in Y^N$ for all $1 \le k \le K$. Besides,

$$\gamma^M_{\mathrm{sym}} = \frac{1}{N!}\sum_{p\in\mathcal{S}_N}\sum_{k=1}^K w_k \delta_{(x^k_{p(1)},\dots,x^k_{p(N)})},$$

the symmetrized version of γ^M, is a minimizer to (1.32).

Theorem 1.2.1 states two things: (i) it is possible to drop the symmetry constraint of the measure γ in problem (1.32) to compute $I^M(\nu)$; (ii) there exists a minimizer of (1.33) which reads as a discrete measure which charges a low number of points (less than $M+2$), and a minimizer to (1.32) can be obtained as the symmetrized version

of this discrete measure. In particular, this means that it is sufficient to identify at most $\mathcal{O}(NM)$ scalars to compute γ^M. This suggests considering the following optimization problem for the computation of $I^M(\nu)$, since

$$I^M(\nu) = \min_{\substack{(w_k)_{1\leq k\leq M+2} \in \mathbb{R}_+^{M+2}, \\ \sum_{k=1}^{M+2} w_k = 1, \\ \left(x_1^k, \dots, x_N^k\right) \in Y^N,\ \forall 1 \leq k \leq M+2, \\ \sum_{k=1}^{M+2} w_k \left(\frac{1}{N}\sum_{i=1}^N f_m(x_i^k)\right) = \int_{\mathbb{R}^3} f_m \, d\nu}} \sum_{k=1}^{M+2} w_k c(x_1^k, \dots, x_N^k).$$

The use of this sparse structure for the design of efficient numerical methods for the resolution of (1.32) is currently work in progress with Aurélien Alfonsi and Rafaël Coyaud. Another nice research perspective lies in the analysis of a stochastic version of such an algorithm, which reads as a manifold-constrained Langevin process.

Let us mention that a similar sparsity result was obtained for a discrete state space approximation of (1.31) in [63], using different mathematical arguments. However, the latter result seems more delicate to exploit from a numerical point of view, since it requires the resolution of an optimization problem defined on a discrete state space.

Chapter 2

Model-order reduction methods for parametrized Partial Differential Equations

In this chapter are summarized some contributions [VE13, VE17, VE21] to the development of numerical methods for the construction of reduced-order models for parameter-dependent Partial Differential Equations (PDEs).

The objective of a reduced-order model reduction method is the following: it may sometimes be very expensive from a computational point of view to simulate the properties of a complex system described by a complicated model, typically a set of PDEs. This cost may become prohibitive in situations where the solution of the model has to be computed for a very large number of values of the parameters involved in the model. Such a parametric study is nevertheless necessary in several contexts, for instance when the value of these parameters has to be calibrated so that numerical simulations give approximations of the solutions that are as close as possible to some measured data. A reduced-order model method then consists in constructing, from a few complex simulations which were performed for a small number of well-chosen values of the parameters, a so-called *reduced model*, much cheaper and quicker to solve from a numerical point of view, and which enables to get an accurate approximation of the solution of the model for any other values of the parameters.

Section 2.1 presents the main features and issues of model-order reduction techniques, with a particular emphasis on the so-called Reduced Basis (RB) method.

The work [VE21] concerns the development of new RB techniques for parameter-dependent transport-dominated problems and is summarized in Section 2.2.

The methods proposed in [VE13, VE17] are mostly application-driven and were motivated by a still on-going collaboration with the Electricité de France (EDF) company. This series of works were developped during the PhD thesis of Amina Benaceur [15], which was done under the joint supervision of Alexandre Ern and myself and was defended in 2018. They are presented in Section 2.3.

2.1 Reduced Basis method and greedy algorithms

Let V be a Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle_V$ with associated norm $\| \cdot \|_V$. As mentioned above, the main goal of model reduction is to approximate as accurately and quickly as possible the solution $u(\mu) \in V$ of a problem of the form

$$\mathcal{A}(u(\mu), \mu) = 0 \tag{2.1}$$

for many different values of a vector $\mu = (\mu_1, \cdots, \mu_p)$ in a certain range $\mathcal{P} \subset \mathbb{R}^p$. In the above formula, $\mathcal{A}$ is a differential or integro-differential operator parametrized by μ, and we assume that for each $\mu \in \mathcal{P}$ there exists a unique solution $u(\mu) \in V$ to the problem (2.1). The set of all solutions is defined as

$$\mathcal{M} := \{u(\mu) : \mu \in \mathcal{P}\} \subset V,$$

and is often referred to as the solution manifold with some abuse of terminology.

The reduced basis (RB) method [25] aims at constructing efficient reduced-order models for the approximation of the solution of parameter-dependent partial differential equations of the form (2.1). It relies on the computation of the exact solution $u(\mu)$ for a small number (say $n \in \mathbb{N}^*$) of well-chosen values of parameters $\mu_1, \cdots, \mu_n \in \mathcal{P}$ in a preliminary *off-line* stage. These functions then form the Galerkin basis of a discretization space used to solve the differential equation for any other value of the parameter $\mu \in \mathcal{P}$ in an *online* stage.

The Kolmogorov n-width gives a good indication on how well a compact subset $\mathcal{M} \subset V$ can be approximated by a n-dimensional linear subspace. It is defined as

$$d_n^V(\mathcal{M}) := \inf_{\substack{V_n \subset V \text{ vector subspace} \\ \dim V_n = n}} \sup_{u \in \mathcal{M}} \inf_{v_n \in V_n} \|u - v_n\|_V.$$

In the case when $d_n^V(\mathcal{M})$ decays rapidly with increasing n, the reduced basis method is likely to provide a good approximation of the solution $u(\mu)$ for any $\mu \in \mathcal{P}$. The difficulty now relies on finding an appropriate set of parameters $(\mu_i)_{1 \leq i \leq n}$ such that $\sup_{u \in \mathcal{M}} \inf_{v_n \in V_n} \|u - v_n\|_V$, with $V_n := \text{Span}\{u(\mu_i),\ 1 \leq i \leq n\}$, is close to the Kolmogorov n-width of the set $\mathcal{M}$. Greedy algorithms stand as the state-of-the-art technique to find such a subset in practice.

The ideal version of a greedy algorithm in this context reads as follows:

RB-Greedy Algorithm:

- **Initialization:** Find $\mu_1 \in \mathcal{P}$ such that

$$\mu_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \|u(\mu)\|_V.$$

  Set $V_1 := \mathrm{Span}\{u(\mu_1)\}$ and $n = 2$.

- **Iteration** $n \geq 2$**:** Find $\mu_n \in \mathcal{P}$ such that

$$\mu_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \|u(\mu) - \Pi_{V_{n-1}} u(\mu)\|_V,$$

  where $\Pi_{V_{n-1}}$ denotes the orthogonal projection of V onto V_{n-1}. Set $V_n := V_{n-1} + \mathrm{Span}\{u(\mu_n)\} = \mathrm{Span}\{u(\mu_1), \cdots, u(\mu_n)\}$.

For all $n \in \mathbb{N}^*$, let us denote by $\sigma_{n-1}(\mathcal{M}) := \max_{\mu \in \mathcal{P}} \|u(\mu) - \Pi_{V_{n-1}} u(\mu)\|_V$ where V_{n-1} is the $n-1$-dimensional subspace of V given by the greedy algorithm defined above. Then, the decay rate of the sequence $(\sigma_n(\mathcal{M}))_{n \in \mathbb{N}^*}$ is related to the decay rate of $\left(d_n^V(\mathcal{M})\right)_{n \in \mathbb{N}^*}$ as was pointed out in the series of works [25, 20, 50]. Indeed, it is proved in [50] that

$$\forall n \in \mathbb{N}^*, \quad \sigma_{2n}(\mathcal{M}) \leq \sqrt{2} d_n^V(\mathcal{M}). \tag{2.2}$$

Inequality (2.2) states that the greedy algorithm presented above yields a sequence of finite-dimensional subspaces $(V_n)_{n \in \mathbb{N}^*}$ that are quasi-optimal with respect to the approximation of the elements of the set $\mathcal{M}$.

In practice, when $\mathcal{M}$ is given as the set of solutions of a parametrized PDE of the form (2.1), it is not easy to compute quantities of the form $\|u(\mu) - \Pi_{V_{n-1}} u(\mu)\|_V$, for $\mu \in \mathcal{P}$ and V_{n-1} some finite-dimensional subspace of V. Instead, a posteriori error estimators are used in order to estimate these quantities.

2.2 Model-order reduction for transport-dominated problems

This section summarizes the contributions of [VE21], which is a joint work with Damiano Lombardi, Olga Mula and François-Xavier Vialard. In this work, the situation of parameter-dependent conservative transport-dominated partial differential equations is considered. More precisely, let $\Omega \subset \mathbb{R}$ be an open interval, $T > 0$ and $Y \subset \mathbb{R}^p$ a compact set. For a given $y \in Y$, let us consider $\rho_y : [0, T] \times \Omega \to \mathbb{R}$ the solution to

$$\partial_t \rho_y(t, x) - \partial_x F(\rho_y(t, x); y, t) = 0, \quad \forall (t, x) \in [0, T] \times \Omega, \tag{2.3}$$

with appropriate initial and boundary conditions. We assume that $F(\rho; y, t)$ is a real-valued mapping defined on a set of functions $\rho : \Omega \to \mathbb{R}$ so that the solution to (2.3) is well-defined and unique.

Let $\mathcal{P} \subset Y \times [0,T]$ and for all $\mu := (y,t) \in \mathcal{P}$, let $\rho(\mu)$ denote the function $\rho_y(t,\cdot)$. We then introduce the *solution set*

$$\mathcal{M} := \{\rho(\mu),\ \mu \in \mathcal{P}\},$$

and assume that $\mathcal{M}$ is included in some Hilbert space V of real-valued functions defined on Ω, for instance $V = L^2(\Omega)$.

Let us give here two prototypical examples of such problems.

- **Example 1: Pure transport equation:** Let $\Omega = (-1,1)$, $Y := [0,1]$, $T = 1$, $\mathcal{P} := Y \times \{1\}$, and consider for all $y \in Y$, the solution ρ_y to

  $$\partial_t \rho_y(t,x) + y\partial_x \rho_y(t,x) = 0, \quad x \in \Omega, t \geq 0,$$

  with initial condition

  $$\rho_y(t=0,x) := \begin{cases} 1 & \text{if } x \in (-1,0), \\ 0 & \text{otherwise.} \end{cases}$$

- **Example 2: Inviscid Burger's equation:** Let $\Omega = (-1,4)$, $Y := [1/2,3]$, $T = 5$, $\mathcal{P} = Y \times [0,T]$ and consider for all $y \in Y$, the solution ρ_y to the inviscid Burger's equation

  $$\partial_t \rho_y(t,x) + \partial_x \left(\rho_y^2\right)(t,x) = 0, \quad x \in \Omega, t \geq 0,$$

  with periodic boundary conditions on Ω and initial condition

  $$\rho_y(0,x) := \begin{cases} 0 & \text{if } -1 \leq x < 0, \\ y & \text{if } 0 \leq x < \frac{1}{y}, \\ 0 & \text{if } \frac{1}{y} \leq x \leq 4. \end{cases}$$

The motivation for [VE21] is the following. For transport-dominated problems of the form (2.3), the sequence $\left(d_n^V(\mathcal{M})\right)_{n\in\mathbb{N}^*}$ decays very slowly with increasing n. For instance, in the case of the pure transport equation introduced above (Example 1), it is proved in [51, 127] that there exists a constant $c > 0$ such that

$$\forall n \in \mathbb{N}^*, \quad d_n^{L^2(\Omega)}(\mathcal{M}) \geq cn^{-1/2}.$$

Thus, standard reduced basis techniques as the method described in 2.1 are doomed to perform poorly on this type of problems. Several numerical methods, based on *nonlinear* approximation techniques, have recently been proposed in the litterature [27, 151, 102] to overcome this difficulty.

The purpose of [VE21] is to propose one particular approach which is based on the following remark. In th case of conservative partial differential equations, it holds that for all $\mu \in \mathcal{P}$, the measure $u(\mu)$ defined by

$$u(\mu)(\,dx) := \rho(\mu)(x)\,dx$$

belongs to $\mathcal{P}_2(\Omega)$ where $\mathcal{P}_2(\Omega)$ denotes the set of probability measures on Ω with finite second-order moments. From now on, let us assume that this is the case and that

$$\widetilde{\mathcal{M}} := \{u(\mu),\ \mu \in \mathcal{P}\} \subset \mathcal{P}_2(\Omega).$$

This is the case in particular for Example 1 and Example 2 introduced above.

For all $v \in \mathcal{P}_2(\Omega)$, let us denote by $\mathrm{cdf}_v : \Omega \to [0,1]$ the *cumulative distribution function* associated to v, defined by

$$\forall x \in \Omega, \quad \mathrm{cdf}_v(x) := \int_{(\inf \Omega, x]} dv(y),$$

and by $\mathrm{icdf}_v : [0,1] \to \Omega$ the *(generalized) inverse cumulative distribution* function associated to v, defined by

$$\forall s \in [0,1], \quad \mathrm{icdf}_v(s) := \inf \{x \in \Omega,\ \mathrm{cdf}_v(x) > s\}.$$

It then holds that for all $v \in \mathcal{P}_2(\Omega)$, icdf_v belongs to $L^2(0,1)$ and that

$$\mathcal{I} := \{\mathrm{icdf}_v,\ v \in \mathcal{P}_2(\Omega)\}$$

is a closed convex set of $L^2(0,1)$.

Let us now denote by $\mathcal{T} := \left\{\mathrm{icdf}_{u(\mu)},\ \mu \in \mathcal{P}\right\}$. The set $\mathcal{T}$ is called hereafter the *transformed solution set*.

From a reduction point of view, it is then more convenient for certain types of transport-dominated conservative equations of the form (2.3) to approximate the transformed solution set $\mathcal{T}$ rather than directly approximate the solution set $\mathcal{M}$. Actually, the interest of such an approach can be easily seen on the particular example of the pure-transport equation. Indeed, in this case, it holds that $d_n^{L^2(0,1)}(\mathcal{T}) = 0$ for all $n \geq 2$.

In the case of Example 2, the following proposition is proved in [VE21].

Proposition 2.2.1 (VE, Lombardi, Mula, Vialard, 2020)**.** *In the case of Example 2, there exists $C > 0$ such that*

$$\forall n \in \mathbb{N}^*, \quad d_n^{L^2(0,1)}(\mathcal{T}) \leq C n^{-21/10}.$$

No lower bounds on the decay rate of $\left(d_n^{L^2(0,1)}(\mathcal{M})\right)_{n\in\mathbb{N}^*}$ has been proved for Example 2. However, numerical observations indicate that $\left(d_n^{L^2(0,1)}(\mathcal{M})\right)_{n\in\mathbb{N}^*}$ may decay much slower than $\left(d_n^{L^2(0,1)}(\mathcal{T})\right)_{n\in\mathbb{N}^*}$ as n increases. Indeed, Figure 2.1 represents the decay of the singular values of the family of functions $(\rho(\mu))_{\mu\in\mathcal{P}}$ in $L^2(\Omega)$ (blue curve) and $\left(\mathrm{icdf}_{u(\mu)}\right)_{\mu\in\mathcal{P}}$ in $L^2(0,1)$ (red curve) for the case of the Burger's equation. Even if the decay of these singular values is not directly linked to the decay of the Kolomogorov n-widths, we can clearly see in Figure 2.1 the potential interest of approximating the set $\mathcal{T}$ rather than the set $\mathcal{M}$.

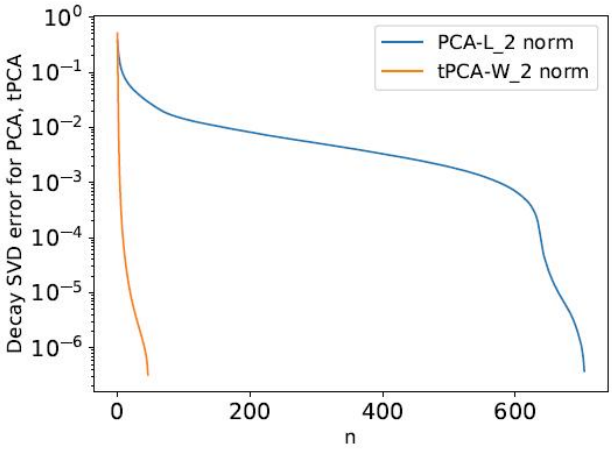

Figure 2.1: Decay of the singular values of the set of functions $(\rho(\mu))_{\mu\in\mathcal{P}}$ (blue curve) and $\left(\text{icdf}_{u(\mu)}\right)_{\mu\in\mathcal{P}}$ (red curve) for the case of the Burger's equation.

The two previous examples illustrate the potential interest of approximating the set $\mathcal{T}$ rather than the set $\mathcal{M}$ for transport-dominated problems. Relying on this observation, we developped two different numerical schemes in order to build reduced-order models for conservative transport-dominated parametrized PDEs. One of these methods is based on the observation that the set $\{\text{icdf}_v,\ v\in\mathcal{P}_2(\Omega)\}$ is a convex subset of $L^2(0,1)$. It then consists in constructing an approximation of $\text{icdf}_{u(\mu)}$ for all values of $\mu\in\mathcal{P}$ as a convex combination of $\text{icdf}_{u(\mu_1)},\cdots,\text{icdf}_{u(\mu_n)}$ for some values $\mu_1,\cdots,\mu_n\in\mathcal{P}$. This amounts to approximating the measure $u(\mu)$ as a barycenter for the 2-Wasserstein metric [2] of $u(\mu_1),\cdots,u(\mu_n)$. The values of the parameters $\mu_1,\cdots,\mu_n\in\mathcal{P}$ are selected by a *barycentric* greedy algorithm which reads as follows, using the notation Conv to denote the convex hull of a set.

Barycentric-Greedy Algorithm:

- **Initialization:** Find $\mu_1,\mu_2\in\mathcal{P}$ such that

$$(\mu_1,\mu_2)\in\underset{\mu,\mu'\in\mathcal{P}}{\text{argmax}}\,\|\text{icdf}_{u(\mu)}-\text{icdf}_{u(\mu')}\|_{L^2(0,1)}.$$

  Set $C_2:=\text{Conv}\{\text{icdf}_{u(\mu_1)},\text{icdf}_u(\mu_2)\}$ and $n=3$.

- **Iteration** $n\geq 3$**:** Find $\mu_n\in\mathcal{P}$ such that

$$\mu_n\in\underset{\mu\in\mathcal{P}}{\text{argmax}}\,\|u(\mu)-\Pi_{C_{n-1}}u(\mu)\|_{L^2(0,1)},$$

  where $\Pi_{C_{n-1}}$ denotes the orthogonal projection of $L^2(0,1)$ onto the closed convex set C_{n-1}. Set $C_n:=\text{Conv}\{C_{n-1},u(\mu_n)\}=\text{Conv}\{u(\mu_1),\cdots,u(\mu_n)\}$ and $n:=n+1$.

This approach yields interesting numerical results for the reduction of differ-

ent types of conservative one-dimensional transport problems, including Burger's, Camassa-Holm and Korteveg-de Vries equations. We refer the reader to [VE21] for more details.

2.3 Collaboration with EDF

In the context of Amina Benaceur's PhD thesis, the EDF company was interested in the development of reduced-order models for the simulation of the thermo-mechanical behaviour of regulation valves used in nuclear reactor operation. Indeed, the computation of the evolution of the temperature and displacement fields in these components (usually done using finite element methods) are usually very expensive from a computational point of view. Indeed, they require the use of very fine meshes (with a very high number of degrees of freedom), a significant number of time steps, complex nonlinear thermal and mechanical behaviour laws, together with nonlinear constraints due to mechanical contact between the different parts of the valve.

The two contributions [VE13, VE17] are motivated by EDF's issues and done in collaboration with Amina Benaceur and Alexandre Ern.

In [VE13], a reduced-basis approach is developped for the reduction of parameter-dependent nonlinear parabolic problems. For such problems, other reduced basis methods were proposed in earlier contributions [72, 73], but the latter works do not take into account the possibly prohibitively large computational cost of the offline phase. In particular for nonlinear problems, the efficiency of RB methods relies on the use of a special interpolation technique to approximate the nonlinearity of the model, called the Empirical Integration Method [121, 120]. This EIM approximation is also usually computed in the offline phase, independently of the reduced basis. Recently, the reduction of the cost of the offline phase in the construction of a reduced-order model has become an important issue which attracted a lot of attention from mathematicians. Among these, in [45], the authors introduce the idea of progressively and jointly enrich the reduced basis and improve on the EIM approximation of the non-linearity for stationary nonlinear PDEs. In [VE13], this idea is adapted to nonlinear evolution problems: the EIM approximation of the nonlinearity and the reduced basis are jointly enriched jointly but according to a different criterion than the one used in [45]. This methodology gave very satisfactory results on the industrial test case of the thermal behaviour of a valve component.

Few works deal with the adaptation of RB methods to variational inequalities [75, 157, 10, 61, 69], and all of them concern the reduction of problems with *linear* constraints. In [VE17], a new RB methodology was developped for the reduction of variational inequalities with nonlinear constraints, which makes use of an EIM interpolation method to deal with the nonlinearity of the constraint. In this context, a *primal* reduced basis is constructed fo the approximation of the primal solution and a *dual* reduced basis is used for the approximation of the Lagrange multipliers. The latter is constructed using a new hierarchical algorithm which is guarantees the non-negativity of the Lagrange multipliers of the obtained reduced model. This reduction strategy is then applied to some contact problems of two elastic bodies with non coincident meshes. The nonlinearity of the constraints stems from the fact that the meshes do not coincide, which is a very common situation in engineering

applications, and had not been previously studied in the litterature, at least up to our knowledge.

2.4 Research perspectives on model-order reduction of parametrized PDEs

Let us mention in this section some research perspectives that I am currently investigating with different collaborators, linked to the development and analysis of model-order reduction methods in different contexts.

Let us begin by mentioning some perspectives and questions related to the work [VE21], which was summarized in Section 2.2.

- Of course, natural extensions of the methodology proposed in [VE21] would be to extend the proposed methodologies to problems defined in two or more dimensions and non-conservative problems. Using the so-called Sinkhorn algorithm [16, 141], it is possible to compute efficiently Wasserstein barycenters in dimension 2 and 3, and we wish to exploit this numerical tool to extend our approach to higher dimensional cases.

- Is it possible to prove results on the rate of decay of Kolmogorov widths of solution sets or transformed solution sets for more general equations?

- Another interesting open question is the following: does the barycentric greedy algorithm enjoy, in some sense, the same quasi-optimality properties than the standard greedy algorithm presented in Section 2.1? These are research tracks I would like to explore in the future.

Let us now outline some other research perspectives on model reduction of parameter-dependent equations.

- Amina Benaceur defended her PhD in 2018. The aim of the PhD thesis of Idrissa Niakh, which started in November 2019 under the supervision of Alexandre Ern and myself, is to further develop Reduced Basis methods for problems of interest for EDF, together with appropriate a posteriori error estimators. Problems that will be considered in this context are parameter-dependent contact problems with friction and geo-mechanical problems, a prototypical example of which being the Drücker-Prager model [119].

- In the context of the PhD thesis of Raed Blel, which started in 2018 and is co-supervised by Tony Lelièvre and myself, the analysis of model order reduction techniques which are used to reduce the simulation time of complex high dimensional sampling problems is currently under study. More precisely, we are currently analyzing some extensions of the RB method which have been proposed by S. Boyaval and T. Lelièvre for parameterized stochastic problems [23].

- Another field of research, currently investigated in collaboration with G. Dusson, T. Lelièvre and F. Madiot, concerns the development of an appropriate a

posteriori error estimator for RB non-symmetric eigenvalue problems. The motivation for considering such problems stems from a collaboration with the Commissariat aux Energies Alternatives (CEA) on the construction of reduced-order models for criticality calculations in neutronics, in particular for nuclear reactors [123].

- Another track of research concerns the reduction of parametrized cross-diffusion systems defined on moving domains. Such systems arise in the modeling of the fabrication process of solar cells and yield challenging difficulties for model-order reduction techniques. We refer the reader to Chapter 4 of this manuscript for more details on this topic.

Chapter 3

Numerical methods for multiscale problems

The contributions presented in this chapter are motivated by the development of numerical methods for multiscale problems, and summarize the results of [VE9, VE18, VE19].

This chapter is organized as follows. In Section 3.1, we recall some basic elements on the theory of stochastic homogenization, and review the standard associated numerical methods used for the computation of the homogenized matrix of a stochastic ergodic heterogeneous diffusion problem.

In the joint works with E. Cancès, F. Legoll, B. Stamm and S. Xiang [VE9, VE18], some alternative methods to approximate the homogenized matrix are proposed. These are based on the use of a so-called *embedded corrector problem*. The embedded corrector problem and three associated different approaches for the computation of homogenized matrices are presented in Section 3.2, together with the associated convergence results.

Our motivation for considering such a family of embedded corrector problems is the following. A very efficient numerical method has been proposed and developed in the series of works [31, 114] in order to solve Poisson problems arising in implicit solvation models. The adaptation of this algorithm, which is based on a boundary integral formulation of the problem, has enabled us to solve these embedded corrector problems in a very efficient way in situations when the considered heterogeneous medium is composed of (possibly polydisperse) spherical inclusions embedded into a homogeneous material. This algorithm was proposed in [VE19] and is summarized in Section 3.3.

Research perspectives on numerical methods for multiscale problems are given in Section 3.4.

3.1 Motivation: numerical stochastic homogenization

In the sequel, the following notation is used. Let $d \in \mathbb{N}^\star$, $0 < \alpha \leq \beta < +\infty$ and

$$\mathcal{M} := \left\{ A \in \mathbb{R}^{d\times d},\ A^T = A \text{ and, for any } \xi \in \mathbb{R}^d,\ \alpha|\xi|^2 \leq \xi^T A \xi \leq \beta |\xi|^2 \right\}.$$

Let $(e_i)_{1\leq i\leq d}$ be the canonical basis of $\mathbb{R}^d$. Taking $\xi = e_i$ and next $\xi = e_i + e_j$ in the above definition, we see that any $A := (A_{ij})_{1\leq i,j\leq d} \in \mathcal{M}$ satisfies $|A_{ij}| \leq \beta$ for any $1 \leq i, j \leq d$. We further denote by $\mathcal{D}(\mathbb{R}^d)$ the set of C^∞ functions with compact supports in $\mathbb{R}^d$.

In this section, we briefly recall the well-known homogenization theory in the stationary ergodic setting, as well as standard strategies to approximate the homogenized coefficients. We refer to [94, 130, 57, 17, 39, 83, 8] for some seminal contributions, books and review articles on this topic. The stationary ergodic setting can be viewed as a prototypical example of contexts in which the alternative method we propose here for approximating the homogenized matrix can be used.

3.1.1 Theoretical setting

Let us recall the definition of G-convergence introduced by F. Murat and L. Tartar in [124]:

Definition 3.1.1 (G-convergence)**.** *Let D be a smooth bounded domain of $\mathbb{R}^d$. A family of matrix-valued functions $\left(\mathbb{A}^R\right)_{R>0} \subset L^\infty(D,\mathcal{M})$ is said to converge in the sense of homogenization (or to G-converge) in D to a matrix-valued function $\mathbb{A}^\star \in L^\infty(D,\mathcal{M})$ if, for all $f \in H^{-1}(D)$, the sequence $(u^R)_{R>0}$ of solutions to*

$$u^R \in H^1_0(D), \quad -div\left(\mathbb{A}^R\nabla u^R\right) = f \text{ in } \mathcal{D}'(D)$$

satisfies

$$\begin{cases} u^R \underset{R\to+\infty}{\rightharpoonup} u^\star \text{ weakly in } H^1_0(D), \\ \mathbb{A}^R\nabla u^R \underset{R\to+\infty}{\rightharpoonup} \mathbb{A}^\star\nabla u^\star \text{ weakly in } L^2(D), \end{cases}$$

where $u^\star$ is the unique solution to the homogenized equation

$$u^\star \in H^1_0(D), \quad -\mathrm{div}\left(\mathbb{A}^\star\nabla u^\star\right) = f \text{ in } \mathcal{D}'(D).$$

The stationary ergodic setting is a prototypical example of family of matrix-valued functions $\left(\mathbb{A}^R\right)_{R>0} \subset L^\infty(D,\mathcal{M})$ which G-converges to a *constant* matrix $\mathbb{A}^\star = A^\star$ for some matrix $A^\star \in \mathcal{M}$.

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $Q := \left(-\frac{1}{2},\frac{1}{2}\right)^d$. For a random variable $X \in L^1(\Omega, d\mathbb{P})$, we denote by $\mathbb{E}[X] := \displaystyle\int_\Omega X(\omega)\, d\mathbb{P}(\omega)$ its expectation value. For the sake of convenience, we restrict the presentation to the case of discrete stationarity, even though the ideas presented here can be readily extended to the case of continuous stationarity. We assume that the group $(\mathbb{Z}^d,+)$ acts on Ω. We denote by $(\tau_k)_{k\in\mathbb{Z}^d}$ this action, and assume that it preserves the measure $\mathbb{P}$, i.e.

$$\forall k \in \mathbb{Z}^d, \quad \forall F \in \mathcal{F}, \quad \mathbb{P}(\tau_k(F)) = \mathbb{P}(F).$$

We also assume that τ is ergodic, that is,

$$\forall F \in \mathcal{F}, \quad \left(\forall k \in \mathbb{Z}^d,\ \tau_k F = F\right) \implies \left(\mathbb{P}(F) = 0 \text{ or } 1\right).$$

A funtion $\mathcal{S} \in L^1_{\text{loc}}\left(\mathbb{R}^d, L^1(\Omega)\right)$ is said to be stationary if

$$\forall k \in \mathbb{Z}^d, \quad \mathcal{S}(x+k,\omega) = \mathcal{S}(x,\tau_k\omega) \text{ for almost all } x \in \mathbb{R}^d \text{ and almost surely.} \quad (3.1)$$

The following theorem is a classical result of stochastic homogenization theory (see e.g. [83]):

Theorem 3.1.1. *Let $\mathbb{A} \in L^\infty(\mathbb{R}^d, L^1(\Omega))$ be such that $\mathbb{A}(x,\omega) \in \mathcal{M}$ almost surely and for almost all $x \in \mathbb{R}^d$. We assume that $\mathbb{A}$ is stationary in the sense of* (3.1)*. For any $R > 0$ and $\omega \in \Omega$, we set $\mathbb{A}^R(\cdot,\omega) := \mathbb{A}(R\cdot,\omega)$. Then, almost surely, for any arbitrary smooth bounded domain $D \subset \mathbb{R}^d$, the sequence $\left(\mathbb{A}^R(\cdot,\omega)\right)_{R>0} \subset L^\infty(D;\mathcal{M})$ G-converges to a* constant and deterministic *matrix $A^\star \in \mathcal{M}$, which is given by*

$$\forall p \in \mathbb{R}^d, \quad A^\star p = \mathbb{E}\left[\frac{1}{|Q|}\int_Q \mathbb{A}(x,\cdot)\left(p + \nabla w_p(x,\cdot)\right)\,dx\right],$$

where w_p is the unique solution (up to an additive constant) in

$$\left\{v \in L^2_{\text{loc}}(\mathbb{R}^d, L^2(\Omega)), \quad \nabla v \in \left(L^2_{\text{unif}}(\mathbb{R}^d, L^2(\Omega))\right)^d\right\}$$

to the so-called corrector problem

$$\begin{cases} -div\left(\mathbb{A}(\cdot,\omega)(p + \nabla w_p(\cdot,\omega))\right) = 0 \text{ almost surely in } \mathcal{D}'(\mathbb{R}^d), \\ \nabla w_p \text{ is stationary in the sense of } (3.1), \\ \mathbb{E}\left[\displaystyle\int_Q \nabla w_p(x,\cdot)\,dx\right] = 0. \end{cases} \quad (3.2)$$

In Theorem 3.1.1, the notation L^2_{unif} refers to the uniform L^2 space:

$$L^2_{\text{unif}}(\mathbb{R}^d, L^2(\Omega)) := \left\{u \in L^2_{\text{loc}}(\mathbb{R}^d; L^2(\Omega)), \quad \sup_{x\in\mathbb{R}^d}\int_{x+(0,1)^d} \|u(y,\cdot)\|^2_{L^2(\Omega)}\,dy < \infty\right\}.$$

The major difficulty to compute the homogenized matrix $A^\star$ is the fact that the corrector problem (3.2) is set over the whole space $\mathbb{R}^d$ and cannot be reduced to a problem posed over a bounded domain (in contrast e.g. to periodic homogenization). This is the reason why approximation strategies yielding practical approximations of $A^\star$ are necessary.

3.1.2 Standard numerical practice

A common approach to approximate $A^\star$ consists in introducing a truncated version of (3.2), see e.g. [22].

Let $Q := \left(-\frac{1}{2}, \frac{1}{2}\right)$ and let $\left(\mathbb{A}^R\right)_{R>0} \subset L^\infty(Q;\mathcal{M})$ a general family of matrix-valued fields which G-converges in the sense of Definition 3.1.1 to a *constant* matrix $A^\star$ in Q. Recall that the family $\left(\mathbb{A}^R(\cdot,\omega)\right)_{R>0}$ introduced above in the ergodic stationary setting provides one prototypical example of such a family.

Let us introduce

$$H^1_{\text{per}}(Q) := \left\{w \in H^1_{\text{loc}}(\mathbb{R}^d), \quad w \text{ is } \mathbb{Z}^d\text{-periodic}\right\}.$$

For all $R > 0$ and for any $p \in \mathbb{R}^d$, let w_p^R is the unique solution in $H^1_{\rm per}(Q)/\mathbb{R}$ to

$$-\mathrm{div}\left(\mathbb{A}^R\left(p + \nabla w_p^R\right)\right) = 0 \text{ almost surely in } \mathcal{D}'(\mathbb{R}^d), \tag{3.3}$$

and define the matrix $A^{\star,R} \in \mathcal{M}$ such that

$$\forall p \in \mathbb{R}^d, \quad A^{\star,R}\, p = \frac{1}{|Q|}\int_Q \mathbb{A}^R\left(p + \nabla w_p^R\right). \tag{3.4}$$

A. Bourgeat and A. Piatniski proved in [22] that the sequence of matrices $\left(A^{\star,R}\right)_{R>0}$ converges almost surely to $A^\star$ as R goes to infinity.

Solving (3.3) by means of standard finite element methods requires the use of very fine discretization meshes, which may lead to prohibitive computational costs. This motivates our work and the alternative definitions of effective matrices that were proposed in [VE9, VE18] and are presented in the next section.

3.2 Embedded corrector problem for homogenization: theoretical analysis

Let $B = B(0,1)$ be the unit open ball of $\mathbb{R}^d$, $\Gamma = \partial B$ and $n(x)$ be the outward pointing unit normal vector at point $x \in \Gamma$. Let $\left(\mathbb{A}^R\right)_{R>0} \subset L^\infty(B;\mathcal{M})$ a family of matrix-valued fields which G-converges in the sense of Definition 3.1.1 to a *constant* matrix $A^\star$ in B.

3.2.1 Embedded corrector problem

In this section, we introduce an *embedded corrector* problem, which is used in the sequel to define new approximations of the homogenized coefficient $A^\star$.

We introduce the vector spaces

$$V := \left\{v \in L^2_{\rm loc}(\mathbb{R}^d),\ \nabla v \in \left(L^2(\mathbb{R}^d)\right)^d\right\} \quad \text{and} \quad V_0 := \left\{v \in V,\ \int_B v = 0\right\}. \tag{3.5}$$

The space V_0, endowed with the scalar product $\langle\cdot,\cdot\rangle$ defined by

$$\forall v, w \in V_0, \quad \langle v, w\rangle := \int_{\mathbb{R}^d} \nabla v \cdot \nabla w,$$

is a Hilbert space.

For any matrix-valued field $\mathbb{A} \in L^\infty(B,\mathcal{M})$, any constant matrix $A \in \mathcal{M}$, and any vector $p \in \mathbb{R}^d$, we denote by $w_p^{\mathbb{A},A}$ the unique solution in V_0 to

$$-\mathrm{div}\Big(\mathcal{A}^{\mathbb{A},A}\left(p + \nabla w_p^{\mathbb{A},A}\right)\Big) = 0 \text{ in } \mathcal{D}'(\mathbb{R}^d), \tag{3.6}$$

where

$$\mathcal{A}^{\mathbb{A},A}(x) := \left| \begin{array}{l} \mathbb{A}(x) \text{ if } x \in B, \\ A \text{ if } x \in \mathbb{R}^d \setminus B. \end{array} \right.$$

The variational formulation of (3.6) reads as follows: find $w_p^{\mathbb{A},A} \in V_0$ such that

$$\forall v \in V_0, \quad \int_B (\nabla v)^T \mathbb{A}(p + \nabla w_p^{\mathbb{A},A}) + \int_{\mathbb{R}^d \setminus B} (\nabla v)^T A \nabla w_p^{\mathbb{A},A} - \int_\Gamma (Ap \cdot n)\, v = 0. \quad (3.7)$$

Problem (3.6) is linear and the above bilinear form is coercive in V_0. This problem is thus equivalent to a minimization problem (recall that $\mathbb{A}$ and A are symmetric). The solution $w_p^{\mathbb{A},A}$ to (3.6) is equivalently the unique solution to the minimization problem

$$w_p^{\mathbb{A},A} = \operatorname*{argmin}_{v \in V_0} J_p^{\mathbb{A},A}(v), \quad (3.8)$$

where

$$J_p^{\mathbb{A},A}(v) := \frac{1}{|B|} \int_B (p+\nabla v)^T \mathbb{A}(p+\nabla v) + \frac{1}{|B|} \int_{\mathbb{R}^d \setminus B} (\nabla v)^T A \nabla v - \frac{2}{|B|} \int_\Gamma (Ap \cdot n) v. \quad (3.9)$$

We define the map $\mathcal{J}_p^{\mathbb{A}} : \mathcal{M} \to \mathbb{R}$ by

$$\forall A \in \mathcal{M}, \quad \mathcal{J}_p^{\mathbb{A}}(A) := J_p^{\mathbb{A},A}\left(w_p^{\mathbb{A},A}\right) = \min_{v \in V_0} J_p^{\mathbb{A},A}(v). \quad (3.10)$$

The linearity of the map $\mathbb{R}^d \ni p \mapsto w_p^{\mathbb{A},A} \in V_0$ yields that, for any $A \in \mathcal{M}$, the map $\mathbb{R}^d \ni p \mapsto \mathcal{J}_p^{\mathbb{A}}(A)$ is quadratic. As a consequence, for all $A \in \mathcal{M}$, there exists a unique symmetric matrix $G^{\mathbb{A}}(A) \in \mathbb{R}^{d \times d}$ such that

$$\forall p \in \mathbb{R}^d, \quad \mathcal{J}_p^{\mathbb{A}}(A) = p^T G^{\mathbb{A}}(A) p. \quad (3.11)$$

The motivation for considering problems of the form (3.6) is twofold. First, we show below that the solution $w_p^{\mathbb{A}^R,A}$ to (3.6) can be used to define consistent approximations of $A^\star$. Second, problem (3.6) can be efficiently solved in some cases [VE19], and we refer to Section 3.3 for the presentation of the main ingredients of the numerical method.

The rest of the section is devoted to the presentation of different methods for constructing approximate effective matrices, using corrector problems of the form (3.6).

3.2.2 Three definitions of approximate homogenized matrices using embedded corrector problems

Two definitions of approximate homogenized matrices rely on the following result, proved in [VE18].

Lemma 3.2.1. *For any $\mathbb{A} \in L^\infty(B, \mathcal{M})$, the function $\mathcal{J}^{\mathbb{A}} : \mathcal{M} \ni A \mapsto \sum_{i=1}^d \mathcal{J}_{e_i}^{\mathbb{A}}(A) = \operatorname{Tr}\left(G^{\mathbb{A}}(A)\right)$ is concave. Moreover, when $d \leq 3$, $\mathcal{J}^{\mathbb{A}}$ is strictly concave.*

We infer from Lemma 3.2.1 that, for any $R > 0$, there exists a matrix $A_1^R \in \mathcal{M}$ such that

$$A_1^R \in \operatorname*{argmax}_{A \in \mathcal{M}} \sum_{i=1}^d \mathcal{J}_{e_i}^{\mathbb{A}^R}(A) = \operatorname*{argmax}_{A \in \mathcal{M}} \operatorname{Tr}\left(G^{\mathbb{A}^R}(A)\right). \quad (3.12)$$

Moreover, in dimension $d \leq 3$, this matrix is unique. A matrix A_1^R satisfying (3.12) provides a first definition of approximate homogenized matrix.

For all $R > 0$, let $A_2^R \in \mathbb{R}^{d\times d}$ be matrix such that

$$A_2^R = G^{\mathbb{A}^R}(A_1^R), \tag{3.13}$$

where A_1^R is a solution to (3.12). Then, a matrix A_2^R satisfying (3.13) provides a second definition of approximate homogenized matrix.

The following convergence result is proved in [VE18].

Proposition 3.2.1. *Let $(\mathbb{A}^R)_{R>0} \subset L^\infty(B, \mathcal{M})$ be a family of matrix-valued fields which G-converges in B to a constant matrix $A^\star \in \mathcal{M}$ as R goes to infinity.*

Then, for any sequence of matrices $\left(A_1^R\right)_{R>0}$ and $\left(A_2^R\right)_{R>0}$ respectively satisfying (3.12) *and* (3.13)*, it holds that*

$$A_1^R \underset{R\to+\infty}{\longrightarrow} A^\star \quad \text{and} \quad A_2^R \underset{R\to+\infty}{\longrightarrow} A^\star.$$

We eventually introduce a third definition, inspired by [37]. Let us assume that, for any $R > 0$, there exists a matrix $A_3^R \in \mathcal{M}$ such that

$$A_3^R = G^{\mathbb{A}^R}(A_3^R). \tag{3.14}$$

This third definition also yields a converging approximation of $A^\star$, as stated in the following proposition which is proved in [VE18]:

Proposition 3.2.2. *Let $(\mathbb{A}^R)_{R>0} \subset L^\infty(B, \mathcal{M})$ be a family of matrix-valued fields which G-converges in B to a constant matrix $A^\star \in \mathcal{M}$ as R goes to infinity.*

Let us assume that, for any $R > 0$, there exists a matrix $A_3^R \in \mathcal{M}$ satisfying (3.14)*. Then,*

$$A_3^R \underset{R\to+\infty}{\longrightarrow} A^\star.$$

In general, we are not able to prove the existence of a matrix A_3^R satisfying (3.14). However, the following weaker existence result holds in the case of an isotropic homogenized medium.

Proposition 3.2.3. *Let $(\mathbb{A}^R)_{R>0} \subset L^\infty(B, \mathcal{M})$ be a family of matrix-valued fields which G-converges in B to a constant matrix $A^\star \in \mathcal{M}$ as R goes to infinity. In addition, assume that $A^\star = \mathsf{a}^\star \mathsf{I}$, where I is the identity matrix of $\mathbb{R}^{d\times d}$.*

Then, for any $R > 0$, there exists a positive number $\mathsf{a}_3^R \in [\alpha, \beta]$ (which is unique at least in the case when $d \leq 3$) such that

$$\mathsf{a}_3^R = \frac{1}{d}\mathrm{Tr}\left(G^{\mathbb{A}^R}\left(\mathsf{a}_3^R\,\mathsf{I}\right)\right). \tag{3.15}$$

In addition,

$$\mathsf{a}_3^R \underset{R\to+\infty}{\longrightarrow} \mathsf{a}^\star. \tag{3.16}$$

Note that, since $A^\star = \mathsf{a}^\star \mathsf{I} \in \mathcal{M}$, we have that $\mathsf{a}^\star \in [\alpha, \beta]$. Note also that (3.15) is weaker than (3.14), which would read $\mathsf{a}_3^R\,\mathsf{I} = G^{\mathbb{A}^R}(\mathsf{a}_3^R\,\mathsf{I})$. However, this weaker result is sufficient to prove that a_3^R is a converging approximation of $\mathsf{a}^\star$.

3.3 Embedded corrector problem for homogenization: numerical method

3.3.1 Isotropic materials with spherical inclusions

As pointed out above, the embedded corrector problem can, in some cases, be very efficiently solved. We describe this situation here. In the sequel, we assume that $d = 3$ and for any $x \in \mathbb{R}^3$ and $r > 0$, we denote by $B_r(x)$ the ball of $\mathbb{R}^3$ of radius r centered at x. For all $R > 0$, we also denote by B_R the ball $B_R(0)$.

In this section, we focus on the particular case where for all $R > 0$, $\mathbb{A}^R := \mathbb{A}(R\cdot)$ for some $\mathbb{A} \in L^\infty(\mathbb{R}^3; \mathcal{M})$. Recall that this is indeed the case in the stochastic ergodic setting. Let us assume in addition that $\mathbb{A}$ satisfies the following assumptions: there exist $\eta > 0$, $(x_n)_{n\in\mathbb{N}^\star} \subset \mathbb{R}^3$, $(r_n)_{n\in\mathbb{N}^\star} \subset \mathbb{R}_+^\star$, $(\mathsf{a}_n)_{n\in\mathbb{N}^\star} \subset [\alpha, \beta]$ and $\mathsf{a}_0, \mathsf{a}^\star \in [\alpha, \beta]$ such that

(A1) for all $n \neq m \in \mathbb{N}^\star$, $\mathrm{dist}(B_{r_n}(x_n), B_{r_m}(x_m)) \geq \eta$;

(A2) for all $n \in \mathbb{N}^\star$, $\mathbb{A}(x) = \mathsf{a}_n\,\mathsf{I}$ when $x \in B_{r_n}(x_n)$;

(A3) $\mathbb{A}(x) = \mathsf{a}_0\,\mathsf{I}$ on $\mathbb{R}^3 \setminus \bigcup_{n\in\mathbb{N}^\star} B_{r_n}(x_n)$;

(A4) the sequence $\left(\mathbb{A}^R\right)_{R>0}$ G-converges in B to $A^\star = \mathsf{a}^\star\,\mathsf{I}$.

In other words, we focus here on the case when the matrix-valued field $\mathbb{A}$ models a material composed only of isotropic phases ($\mathbb{A}$ is everywhere proportional to the identity matrix I), with spherical inclusions embedded into a homogeneous material, and such that the associated homogenized material is also isotropic. Let us point out that the algorithm presented below enables to compute very efficiently the effective thermal properties of polydisperse materials.

Assume now that we are interested in computing the homogenized coefficient $\mathsf{a}^\star$ associated to a matrix-valued field $\mathbb{A}$ satisfying Assumptions (A1)-(A4). Then, following the results of [VE18], for each value of $R > 0$, one can define three approximate effective coefficients a_R^1, a_R^2 and a_R^3, which are scalar versions of (3.12), (3.13) and (3.14), as follows:

$$\mathsf{a}_R^1 = \operatorname*{arg\,max}_{\mathsf{a}_\infty \in [\alpha,\beta]} \mathcal{J}^{\mathbb{A}^R}(\mathsf{a}_\infty\,\mathsf{I}), \tag{3.17}$$

$$\mathsf{a}_R^2 = \mathcal{J}^{\mathbb{A}^R}(\mathsf{a}_R^1\,\mathsf{I}), \tag{3.18}$$

$$\mathsf{a}_R^3 \in [\alpha, \beta] \ \text{ such that } \ \mathsf{a}_R^3 = \mathcal{J}^{\mathbb{A}^R}(\mathsf{a}_R^3\,\mathsf{I}), \tag{3.19}$$

where for all $A \in \mathcal{M}$, $\mathcal{J}^{\mathbb{A}^R}(A) = \frac{1}{3}\sum_{i=1}^3 \mathcal{J}_{e_i}^{\mathbb{A}^R}(A)$. Since $\mathcal{J}^{\mathbb{A}^R}$ is strictly concave, and using an easy adaptation of Proposition 3.2.3, we obtain that the above three approximations are well-defined. It then holds that, for any $1 \leq i \leq 3$,

$$\lim_{R\to+\infty} \mathsf{a}_R^i = \mathsf{a}^\star.$$

3.3.2 Main ingredients of the numerical method

The computation of a_R^1, a_R^2 and a_R^3 requires the resolution of embedded corrector problems of the form

$$-\mathrm{div}\Big(\mathcal{A}^{\mathbb{A}^R,A^R}\left(p+\nabla w_p^{\mathbb{A}^R,A^R}\right)\Big)=0 \text{ in } \mathcal{D}'(\mathbb{R}^d),$$

where $\mathcal{A}^{\mathbb{A}^R,A^R}$ is defined by

$$\mathcal{A}^{\mathbb{A}^R,A^R}(x):=\left|\begin{array}{l}\mathbb{A}^R(x) \text{ if } x\in B,\\ A^R \text{ if } x\in\mathbb{R}^d\setminus B,\end{array}\right.$$

with $A^R=\mathsf{a}_\infty^R\,\mathsf{I}$ for some $\mathsf{a}_\infty^R>0$.

Let us point out that, using a standard rescaling argument similar to the one explained in Section 3.1.2, computing $w_p^{\mathbb{A}^R,A^R}$ for some $p\in\mathbb{R}^3$ is equivalent to computing $\widetilde{w}_p^{\mathbb{A}^R,A^R}:=Rw_p^{\mathbb{A}^R,A^R}\left(\frac{\cdot}{R}\right)$ and that $\widetilde{w}_p^{\mathbb{A}^R,A^R}$ is the unique solution in $\widetilde{V}_0^R:=\left\{v\in V,\ \int_{\mathsf{B}_R}v=0\right\}$ to

$$-\mathrm{div}\left(\widetilde{\mathcal{A}}^{\mathbb{A},A^R}\left(p+\nabla\widetilde{w}_p^{\mathbb{A},A^R}\right)\right)=0 \text{ in } \mathcal{D}'(\mathbb{R}^d), \tag{3.20}$$

where

$$\widetilde{\mathcal{A}}^{\mathbb{A},A^R}(x):=\left\{\begin{array}{l}\mathbb{A}(x) \text{ if } x\in\mathsf{B}_R,\\ A^R \text{ otherwise.}\end{array}\right.$$

Computing the solution of the embedded corrector problems (3.20) with $A^R=\mathsf{a}_\infty^R\,\mathsf{I}$ for some $\mathsf{a}_\infty^R>0$ can be done very efficiently, using the algorithm developped in [VE19], provided that the sphere $\partial\mathsf{B}_R$ does not intersect any of the spherical inclusions $B_{r_n}(x_n)$ for $n\in\mathbb{N}^\star$.

We do not present the algorithm in full details for the sake of brievity, but summarize its main ingredients:

- Since the sphere $\partial\mathsf{B}_R$ does not intersect any of the spherical inclusions $B_{r_n}(x_n)$ for $n\in\mathbb{N}^\star$, one can derive an integral equation formulation of Problem (3.20). The unknown functionof this integral equation formulation is λ, the trace of $\widetilde{w}_p^{\mathbb{A},A^R}$ on the sphere $\partial\mathsf{B}_R$ and on the surface of the spherical inclusions $\partial B_{r_n}(x_n)$ for all $n\in\mathbb{N}^*$ such that $B_{r_n}(x_n)\subset\mathsf{B}_R$.

- A Galerkin approximation with truncated series of real spherical harmonics is used to approximate λ. More precisely, for some fixed value of $N\in\mathbb{N}^*$, for all $n\in\mathbb{N}^*$ such that $B_{r_n}(x_n)\subset\mathsf{B}_R$, the trace $\lambda|_{\partial B_{r_n}(x_n)}$ is approximated by an element of

$$V_N^n:=\mathrm{Span}\left\{\mathcal{Y}_{lm}\left(\frac{\cdot-x_n}{r_n}\right),\ 0\le l\le N,\ -l\le m\le l\right\},$$

  where $(\mathcal{Y}_{lm})_{l\in\mathbb{N}^*,-l\le m\le l}$ denotes the set of real spherical harmonics for the unit sphere $\mathbb{S}^2$ of $\mathbb{R}^3$. Similarly, the trace $\lambda|_{\partial\mathsf{B}_R}$ is approximated by an element of

$$V_N^\infty:=\mathrm{Span}\left\{\mathcal{Y}_{lm}\left(\frac{\cdot}{R}\right),\ 0\le l\le N,\ -l\le m\le l\right\}.$$

This approximation is obtained as the Galerkin approximation of λ associated to the variational formulation of the integral equation formulation of Problem (3.20).

- A numerical integration scheme is used to compute approximations of the discretized matrices involved in the resulting discrete problem using a sufficiently high number N_g of Lebedev integration points.
- A significant speed up of the computations of the discrete matrices can be obtained using the Fast Multipole Method (FMM) [13].

Of course, if the matrix-valued field $\mathbb{A}$ satisfies Assumptions (A1)-(A4), it is not always possible to find arbitrarily large values of $R > 0$ such that $\partial\mathsf{B}_R$ does not intersect any of the spherical inclusions of the material (see the left side of Figure 3.1 for an illustration). Adapting the algorithm presented in [VE19] to the case when spherical inclusions can intersect with each other and/or when spherical inclusions can intersect with the outer sphere $\partial\mathsf{B}_R$ will be the subject of a future work.

Here, we consider a heuristic procedure which consists, for a given value of $R > 0$, in replacing the value of the material coefficient field $\mathbb{A}$ inside the ball B_R by a modified coefficient field $\overline{\mathbb{A}}$. The aim of this procedure is to ensure that the volume of the inclusions in $\overline{\mathbb{A}}|_{\mathsf{B}_R}$ is equal to the volume of the inclusions in $\mathbb{A}|_{\mathsf{B}_R}$, but ensuring that no inclusion associated to the field $\overline{\mathbb{A}}$ intersect the sphere $\partial\mathsf{B}_R$. More precisely, the field $\overline{\mathbb{A}}$ is defined as follows.

Let us assume that $\mathrm{Card}\{n \in \mathbb{N}^\star, \quad B_{r_n}(x_n) \subset \mathsf{B}_R\} = M$ for some $M \in \mathbb{N}^\star$: there are exactly M spherical inclusions that are contained in the ball B_R. Up to reordering the elements of the sequence $(x_n, r_n, \mathsf{a}_n)_{n\in\mathbb{N}^\star}$, we can assume that $\{n \in \mathbb{N}^\star, \quad B_{r_n}(x_n) \subset \mathsf{B}_R\} = \{1, \ldots, M\}$ without loss of generality. Let us assume in addition that there are $\widehat{M} \in \mathbb{N}^\star$ balls that intersect with B_R but do not lie entirely in B_R:

$$\mathrm{Card}\Big\{n \in \mathbb{N}^\star, \quad B_{r_n}(x_n) \cap \mathsf{B}_R \neq \emptyset \ \text{ and } \ B_{r_n}(x_n) \not\subset \mathsf{B}_R\Big\} = \widehat{M}.$$

We denote by $\widehat{x}_1, \ldots, \widehat{x}_{\widehat{M}}$ (respectively $\widehat{r}_1, \ldots, \widehat{r}_{\widehat{M}}$ and $\widehat{a}_1, \ldots, \widehat{a}_{\widehat{M}}$) their centers (respectively their radii and diffusion coefficients).

We define

$$\gamma := \frac{\sum_{i=1}^{M} |B_{r_i}(x_i)| + \sum_{j=1}^{\widehat{M}} |\mathsf{B}_R \cap B_{\widehat{r}_j}(\widehat{x}_j)|}{\sum_{i=1}^{M} |B_{r_i}(x_i)|}. \tag{3.21}$$

The material coefficient $\mathbb{A}$ inside the ball B_R is then replaced by the modified material coefficient

$$\overline{\mathbb{A}}(x) := \left| \begin{array}{ll} \mathsf{a}_i\,\mathsf{I} & \text{if } x \in B_{\gamma r_i}(x_i) \text{ for some } 1 \leq i \leq M, \\ \mathsf{a}_0\,\mathsf{I} & \text{if } x \in \mathsf{B}_R \setminus \bigcup_{i=1}^{M} B_{\gamma r_i}(x_i). \end{array} \right. \tag{3.22}$$

In other words, in the proposed procedure, the spherical inclusions which intersect with $\partial\mathsf{B}_R$ are deleted and the ones included in B_R are rescaled by the factor $\gamma \geq 1$. A sketch of this procedure is shown on the right side of Figure 3.1.

Since the spherical inclusions included in B_R grow in the rescaling process, some of them may no longer be included in B_R after rescaling, or may intersect with another inclusion. However, since the scaling factor γ converges to one as R goes

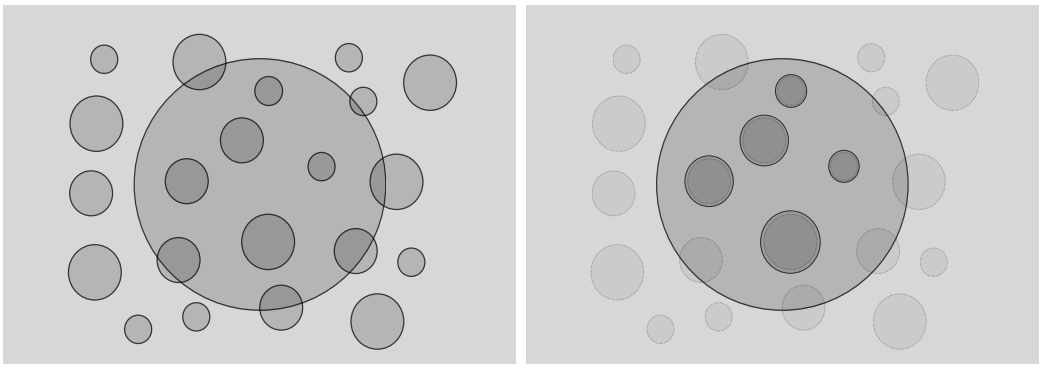

Figure 3.1: Process of restricting and scaling all spherical inclusions inside B_R.

to infinity with rate $\mathcal{O}\left(R^{-2/3}\right)$, we do not observe this problem in practice in our numerical tests, for large values of R. Effective coefficients a_R^1, a_R^2 and a_R^3 are then computed using formulas (3.17), (3.18) and (3.19), using $\overline{\mathbb{A}}$ instead of $\mathbb{A}$.

3.3.3 Numerical results

We present here some numerical tests in the deterministic and stochastic homogenization frameworks to illustrate the performance of the proposed method.

Several parameters influence the accuracy of the approximation of the true homogenized coefficient $\mathsf{a}^\star$:

- First, the truncation of the material with the ball of radius R introduces a model error.
- Second, as mentioned in Section 3.3.2, the solution to (3.20) for fixed R is approximated by a Galerkin scheme based on real spherical harmonics of maximum degree N and using numerical quadrature with N_g points. These approximations create a discretization error. In practice, in what follows, we choose N_g such that the product of two spherical harmonics of degree N is exactly integrated (which is possible when using Lebedev points).
- Third, an iterative solver (with a stopping criterion based on some error tolerance on the residual) is used to solve the obtained linear system.
- Finally, the optimization or fixed-point algorithm involved in the computation of the approximate homogenized coefficients (3.17)-(3.19) also requires an error tolerance.

In all computations, unless otherwise stated, the following convergence criteria are used:

- the iterations of the linear solver are stopped as soon as the relative l^2 norm of the residual is smaller than $\eta_{\text{ls}} = 10^{-7}$;
- the iterations of the optimization or fixed-point algorithm are stopped when the absolute value of the difference between two consecutive values of the diffusion parameter a_∞ (which also corresponds to the *relative error*, since $\mathsf{a}^\star$ is of the order of one in the numerical tests below) is smaller than $\eta_{\text{opt}} = 10^{-5}$.

We recall the reader that, unless otherwise stated, we use the heuristic procedure described in Section 3.3.1 to ensure that (for any fixed value of $R > 0$) the inclusions Ω_i, $1 \leq i \leq M$, do not intersect the sphere $\partial \mathsf{B}_R$.

Test case 1: periodic inclusions

We first consider a case where spherical inclusions of radius $r_n = 0.25$ are periodically arranged on the cubic lattice $\mathbb{Z}^3$. All the spherical inclusions share the same diffusion coefficient $\mathsf{a}_n = 10$, while the diffusion constant of the matrix is fixed to $\mathsf{a}_0 = 1$. Due to the symmetries of the geometrical setting, the value $\mathcal{J}_{e_i}^{\mathbb{A}^R,N}(\mathsf{a}_\infty)$ does not depend on i.

We observe numerically that the dependence in N for $\mathsf{a}^2_{R,N}$ and $\mathsf{a}^3_{R,N}$ is negligible with respect to the error in R, in the sense that the error due to the truncation in N is dominated by the error introduced by the embedded corrector method on the one hand and by neglecting the spherical inclusions that intersect $\partial \mathsf{B}_R$ on the other hand. The situation is slightly different for $\mathsf{a}^1_{R,N}$ as the dependence in N is more pronounced. The approximations $\mathsf{a}^2_{R,N}$ and $\mathsf{a}^3_{R,N}$ of the exact homogenized coefficient are almost identical, and depend only slightly on N. In the following tests, the value $N = 1$ has been chosen.

Figure 3.2 illustrates how the scaling procedure (3.21)–(3.22), which is used to better account for the inclusions intersecting $\partial \mathsf{B}_R$, modifies the values of the approximate coefficients $\mathsf{a}^1_{R,N}$, $\mathsf{a}^2_{R,N}$ and $\mathsf{a}^3_{R,N}$. We monitor the homogenized coefficient on the interval $R \in [2, 20]$, again with $N = 1$. The dotted line refered to as "without scaling" illustrates the value of the effective coefficients computed when the material coefficient inside the ball B_R is replaced by a coefficient of the form (3.22) with $\gamma = 1$ (i.e. when the inclusions intersecting with $\partial \mathsf{B}_R$ have been deleted, but the ones inside B_R have *not* been enlarged). We also report the extrapolated value obtained from the FEM computations as reference value. We observe that the scaling procedure prevents a systematic and very slowly convergent bias of the approximation introduced by discarding the inclusions intersecting $\partial \mathsf{B}_R$. This motivates the scaling procedure (3.21)–(3.22), which we use in all the other computations.

To study the convergence with respect to R, we have computed a reference solution which is obtained as the average of the results for $R = 40$, 40.25, 40.5 and 40.75 with the tighter convergence criteria $\eta_{\text{ls}} = 10^{-8}$ and $\eta_{\text{opt}} = 10^{-6}$. This geometrical set-up contains up to 278,370 spherical inclusions (for $R = 40.75$). The errors on the coefficients $\mathsf{a}^1_{R,N}$ and $\mathsf{a}^2_{R,N}$ are shown on Figure 3.3. We can observe decays that are proportional at least to $1/R$ for $\mathsf{a}^1_{R,N}$, and approximately to $1/R^2$ for $\mathsf{a}^2_{R,N}$.

Test case 2: random inclusions

We now consider the case of a random material with polydisperse spherical inclusions. The radii of the inclusions are uniformly distributed between 0.1 and 0.25, their centers are uniformly distributed under the constraint that the distance between two spheres is not smaller than 0.4, and the diffusion coefficient in each inclusion is uniformly distributed between 10 and 50. On average, there is one inclusion

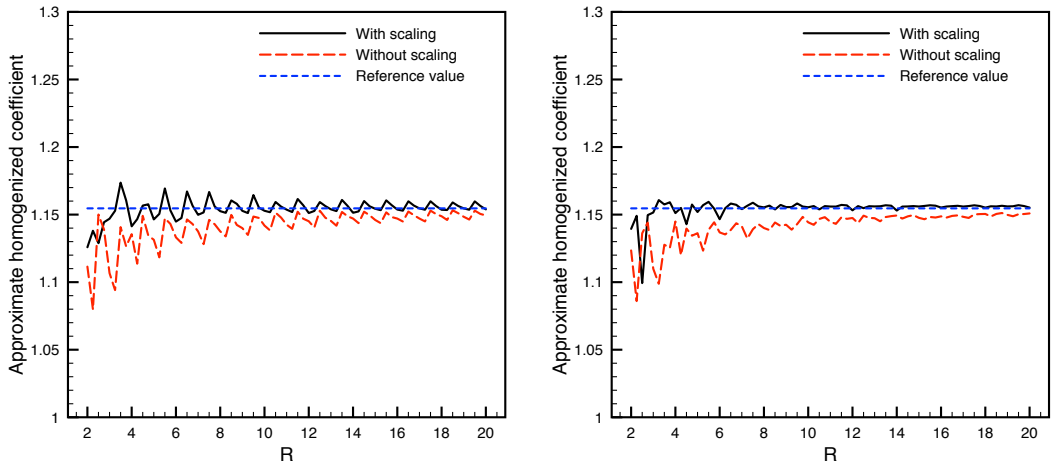

Figure 3.2: [Test case 1] Plots of the functions $R \mapsto \mathsf{a}^1_{R,N}$ (left) and $R \mapsto \mathsf{a}^2_{R,N}$ (right) for $N = 1$, with and without scaling of the inclusions inside B_R (see text). The coefficients $\mathsf{a}^2_{R,N}$ and $\mathsf{a}^3_{R,N}$ are identical at the scale of the figures.

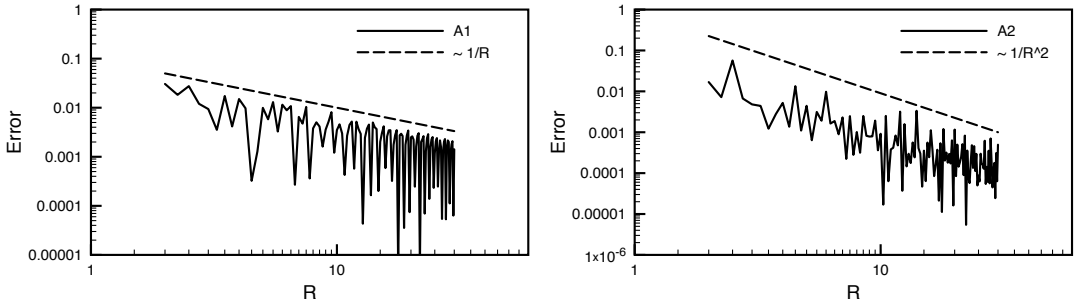

Figure 3.3: [Test case 1] Errors $|\mathsf{a}^1_{R,N} - \mathsf{a}^\star|$ (left) and $|\mathsf{a}^2_{R,N} - \mathsf{a}^\star|$ (right) as functions of R for $N = 1$ (log-log scale).

per cube of unit size. The three random variables (radius, position and diffusion coefficient) are independent. We consider R in the range $[2, 20]$. The largest simulated configuration consists of 32,442 inclusions.

On the left side of Figure 3.4, we plot the three approximate homogenized coefficients $\mathsf{a}^1_{R,N}$, $\mathsf{a}^2_{R,N}$ and $\mathsf{a}^3_{R,N}$ as functions of R (we have set $N = 1$). We have run two simulations, which are based on the same geometric configuration, that is on the *same realization* of the random material. In the first one, we have simply discarded the inclusions intersecting with the boundary $\partial\mathsf{B}_R$, while in the second one the scaling procedure (3.21)-(3.22) is used. The results with and without using the scaling procedure are significantly different, and the approximations converge much faster with respect to R when the scaling procedure is used.

On the right side of Figure 3.4, we present some timings. All simulations have been run on a 4 GHz Intel Core i7 processor, without any parallelization (all computations have been run on a single processor). The method is implemented in Matlab and calls the ScalFMM-library through the MEX-interface. We show the wall-clock timings to compute $\mathsf{a}^1_{R,N}$, $\mathsf{a}^2_{R,N}$ and $\mathsf{a}^3_{R,N}$ for different values of the convergence threshold η_{ls} and for different numbers of inclusions (the largest system,

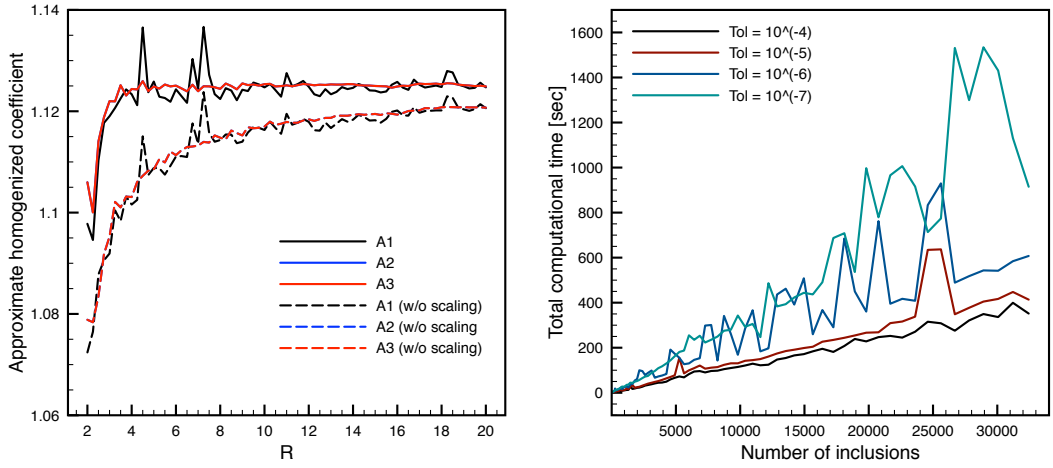

Figure 3.4: [Test case 2] Left: Plots of the functions $R \mapsto \mathsf{a}^1_{R,N}$ (marked as A1), $R \mapsto \mathsf{a}^2_{R,N}$ (marked as A2) and $R \mapsto \mathsf{a}^3_{R,N}$ (marked as A3) for $N = 1$. Right: Total computational time to determine the effective diffusion constant.

consisting of 32,442 inclusions, corresponds to $R = 20$). We have set $N = 1$ for these tests. The threshold η_{opt} for the Armijo line seach is chosen 100 times as large, i.e. $\eta_{\text{opt}} = 100\,\eta_{\text{ls}}$. We observe that the cost increases only linearly with respect to the number M of inclusions, a direct consequence of the use of FMM (without FMM, the cost scales quadratically with respect to M).

3.4 Research perspectives on numerical methods for multiscale problems

The work contained in [VE9, VE18, VE19] concerns a numerical method to approximate the homogenized diffusion matrix associated to, for instance, a stochastic ergodic heterogeneous diffusion problem. The fact that the homogenized matrix field is constant makes the resolution of the associated homogenized problem much more convenient to solve from a computational point of view than the original multiscale problem. However, it is relevant in some applications, especially when the typical size of the heterogeneities is not very small, to obtain better approximations of the solution of the actual multiscale problem, rather than the solution of the homogenized problem. Several numerical methods have been developped in the last decades to address this kind of issues; the Multiscale Finite Element method [54] is one of them. In an ongoing work with Arthur Lebée, Frédéric Legoll and the PhD student Adrien Lesage, we are currently developping certified MsFEM method for heterogeneous plate and shell structures, for diffusion and elasticity problems. The mathematical difficulty in this context is that one of the direction of the domain is assumed to go to 0 at a speed comparable with the typical size of the heterogeneities in the plate or the shell.

More generally, I intend to go on working with the development and analysis of multiscale finite element methods in different contexts in the future.

Chapter 4

Cross-diffusion systems

This chapter summarizes some of my contributions concerning the analysis of cross-diffusion systems. My interest in these problems stems from a collaboration with researchers from the Institut Photovoltaïque de France (IPVF), the aim of which is to propose and analyze a model for the simulation of the fabrication process of thin film solar cells, which is usually done via a Physical Vapor Deposition (PVD) process. This fabrication process and the motivation for considering cross-diffusion systems for its modeling are presented in Section 4.1. Such a model then reads as a system of partial differential equations defined over a time-dependent domain.

My contributions are two-fold. First, some theoretical results were proved for cross-diffusion systems defined on fixed domains with no-flux boundary conditions. In [VE16], the existence and uniqueness of strong solutions is proved for a particular cross-diffusion system, defined on a fixed domain with no-flux boundary conditions, under the assumption that the coefficients encoding the diffusion properties of each pair of species are close. This contribution is summarized in Section 4.2.

Second, a one-dimensional model for the simulation of the fabrication process of thin film solar cells, defined on a time-dependent domain, is proposed and analyzed in [VE12]. This contribution is summarized in Section 4.3.

4.1 Motivation: modeling of a Physical Vapor Deposition process

The contributions presented in this chapter are motivated by the modeling, simulation and control of a Physical Vapor Deposition (PVD) process, the different steps of which are described in details for instance in [122]. Such a technique is used in several contexts, for instance for the fabrication of thin film crystalline solar cells. The procedure works as follows: a substrate wafer is introduced in a hot chamber where the different chemical species composing the film are injected under a gaseous form. Molecules deposit on the substrate surface, so that a solid thin film layer grows. In addition, due to the high temperatures imposed inside the hot chamber, the different components diffuse on the surface and inside the bulk of the film, so that the local volumic fractions of each chemical species evolve through time. The temperature in the chamber and the rates at which the different chemical entities are injected can be modified during the process. Once the wafer is taken out of

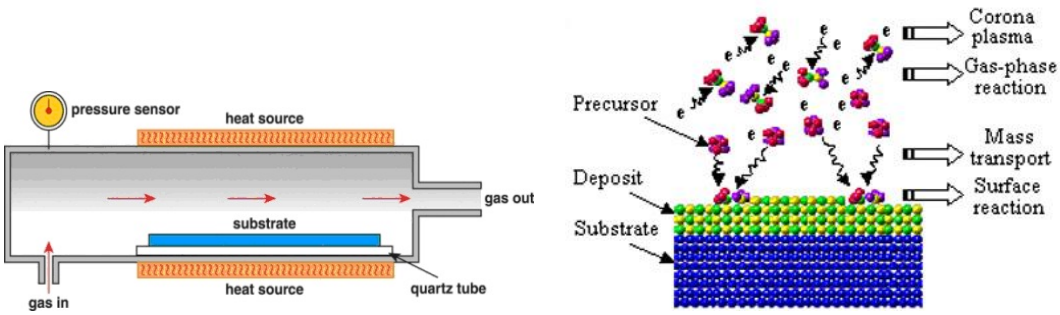

Figure 4.1: Illustration of the PVD process: hot chamber (left) and molecule deposition (right).

the chamber and brought to room temperature, the bulk and surface cross-diffusion phenomena mentioned above are stopped and the final chemical composition of the film is thus *frozen*. The efficiency of the obtained solar cell crucially depends on its chemical composition, which is characterized by the final profile of the volumic fractions of the various chemical compounds inside the solid phase. The chemical composition and geometry of the surface of the film also play a very important role because they determine the quality of the mechanical adherence with the coating layer which will be deposited on top of the semiconducting film in a later stage of the fabrication process. Optical and energetical properties such as the amount of light absorbed by the solar cell also strongly depend on this surface state.

A major challenge consists in optimizing the gazeous fluxes of the various atomic species injected and the temperature of the hot chamber during the process for the final volumic fractions in the bulk and surface state of the layer to be as close as possible to some desired targets. To this aim, it is essential to dispose of a trustworthy model to account for the evolution of the chemical composition and geometry of the film during the PVD process, as well as accurate mathematical methods to approximate the solutions of this model via numerical simulations. Two main phenomena have to be taken into account: the first is naturally the evolution of the surface of the film; the second is the diffusion of the various species in the bulk, due to the high temperature conditions.

The simulation of heteroepitaxial growth processes has been a very active field of research lately. Several numerical methods already exist to simulate this kind of processes, but all of them suffer from serious limitations in the context described above. They can be roughly divided into two main categories, stochastic and continuous, which we describe in more details hereafter.

In stochastic methods, the evolution of the position of each individual atom (or molecule) involved in the heteroepitaxial growth is simulated via a stochastic process. Each atom can move according to various rules, and the rate of occurence of displacement events depends strongly on its chemical environment (typically the positions and chemical natures of the neighbouring atoms). Molecular Dynamics (MD) [106, 103] and Kinetic Monte-Carlo (KMC) [138, 82, 12] algorithms are the most widely used stochastic methods for the simulation of epitaxial growth processes. Their popularity stems from the fact that they allow to reproduce deposition and

diffusion phenomena at an atomistic level, which can provide very useful insights on the growth mechanisms occuring during a heteroepitaxial process such as PVD. However, the richness of the information offered by such simulations comes at a high price: describing the evolution of the state of all the atoms involved in the process can be afforded only for very small systems on very small time scales. This huge limitation makes this family of methods unsuitable for the simulation of the growth of semiconducting thin films in realistic fabrication conditions. Indeed, the typical thickness of a thin film layer in a photovoltaic cell is of the order of 100 μm (thus approximately 10^6 atomistic layers), and tracking the positions of all the atoms composing the layer is thus unfeasible. Besides, such simulations can only be carried out to describe the evolution of an atomistic system during times of the order of the nanosecond. In comparison, the PVD process used for the fabrication of thin film solar cells described in the previous section lasts several hours.

Continuous models of matter offer interesting alternatives to the simulation of heteroepitaxial growth processes [18, 153, 149]. They are appealing in our context since they enable to account for the surface evolution of thin films for realistically-sized systems. More precisely, let $d \in \mathbb{N}^*$ denote the dimension of the physical space (typically $d \leq 3$). Let us assume that at a time $t \geq 0$, the solid layer is composed of $n+1$ different chemical species and occupies a domain $\Omega(t) \subset \mathbb{R}^d$. At time $t > 0$ and point $x \in \Omega(t)$, the local volumic fractions of the different species are denoted respectively by $u_0(t,x), \cdots, u_n(t,x)$ and are the unknown quantities of interest which encode the composition of the film. The evolution of the domain $\Omega(t)$ and of the local volumic fractions $u_0(t,x), \cdots, u_n(t,x)$ has to be determined and depends in particular on the fluxes of atoms that are absorbed at the surface of the layer.

Hydrodynamic limits of MD or KMC models (i.e. limits of the models when the number of atoms in the system goes to infinity) are usually given by multi-species cross-diffusion systems (see [134] for instance). From a modeling point of view, it is thus natural to model the evolution of the local volumic fractions of each chemical species inside the bulk of the film by a system of equations of the form:

$$\partial_t U - \mathrm{div}_x \left(\mathcal{A}(U) \nabla_x U \right) = 0, \qquad \text{for } t > 0,\ x \in \Omega(t), \tag{4.1}$$

where $U = (u_0, \cdots, u_n)$ and $\mathcal{A} : \mathbb{R}^{n+1} \to \mathbb{R}^{(n+1)\times(n+1)}$ is a matrix-valued function encoding the cross-diffusion properties of the different species. This set of equations is defined on the time-dependent domain occupied by the film $\Omega(t)$, whose evolution possibly depends on (i) the fluxes of atoms that are absorbed at the surface of the layer (ii) the values of $u_0, \cdots, u_n$ on $\partial\Omega(t)$.

From a physical point of view, since for all $0 \leq i \leq n$, $u_i(t,x)$ represents the local volumic fraction of the i^{th} chemical species at time t and point x, $u_i(t,x)$ has to be non-negative and to satisfy the so-called *volumic constraint* :

$$\forall 0 \leq i \leq n, \quad u_i(t,x) \geq 0 \text{ and } \sum_{i=0}^{n} u_i(t,x) = 1, \quad \forall t \in \mathbb{R}_+,\ x \in \Omega(t). \tag{4.2}$$

Systems such as (4.1)-(4.2) have received much attention from the mathematical community in the case when no-flux boundary conditions are imposed on a fixed domain. However, very few works have considered the case of time-dependent domains.

We present in Section 4.2.1 the mathematical difficulties raised by the analysis of such systems, together with some mathematical methods used in the case of fixed domains with no-flux boundary conditions, with a particular emphasis on the so-called *boundedness by entropy* method, which was introduced and developped in particular in [26, 87]. For the sake of simplicity, we restrict our presentation here to the case when solutions are expected to satisfy a volumic constraint of the form (4.2), which is called in the litterature the *volume-filling case*.

4.2 Cross-diffusion systems on fixed domains

The aim of this section is to present some contributions related to the analysis of cross-diffusion systems defined on fixed domains with no-flux boundary conditions contained in [VE12, VE16]. A general introduction to the challenges raised by the mathematical analysis of such sytems is given in Section 4.2.1.

A particular focus on the so-called *boundedness by entropy* method introduced and developped in [26, 87] to prove the existence of weak solutions to systems which exhibit a formal gradient flow structure is made in Section 4.2.3. A particular example of cross-diffusion system which exhibits such a formal gradient flow structure is presented in Section 4.2.2. For this particular system, under some appropriate assumptions on the value of some cross-diffusion coefficients, the existence and uniqueness of strong solutions has been proved in [VE16] and this contribution is summarized in Section 4.2.4. Research perspectives are mentioned in Section 4.2.5.

4.2.1 Mathematical analysis: challenges

We assume in Section 4.2 that the domain is time-independent, namely for all $t \geq 0$, $\Omega(t) = \Omega \subset \mathbb{R}^d$ where Ω is a fixed bounded domain with smooth boundary. We denote by $\mathbf{n}$ the outward unit normal vector to $\partial\Omega$ and consider the cross-diffusion system with no-flux boundary conditions:

$$\left\{ \begin{array}{rl} \partial_t U - \mathrm{div}_x \left(\mathcal{A}(U) \nabla_x U \right) = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \Omega, \\ \mathcal{A}(U) \nabla_x U \cdot \mathbf{n} = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \partial\Omega, \end{array} \right. \tag{4.3}$$

together with the initial condition $U^0 := (u_0^0, \cdots, u_n^0) \in (L^1(\Omega))^{n+1}$. This initial condition is assumed to satisfy:

$$\forall 0 \leq i \leq n, \quad u_i^0(x) \geq 0, \quad \sum_{i=0}^{n} u_i^0(x) = 1 \text{ and } u_i(0,x) = u_i^0(x) \quad \text{a.e. in } \Omega. \tag{4.4}$$

Denoting by $(u_0, \cdots, u_n)$ the $n+1$ components of U, the volumic constraint (4.2) reads in this case as

$$\forall 0 \leq i \leq n, \quad u_i(t,x) \geq 0 \text{ and } \sum_{i=0}^{n} u_i(t,x) = 1, \quad \forall t \in \mathbb{R}_+, \ x \in \Omega. \tag{4.5}$$

In other words, the condition (4.5) formulates that a solution U to (4.3) is expected to take values in $\overline{\mathcal{P}} \subset [0,1]^{n+1}$ where

$$\mathcal{P} := \left\{ z \in (\mathbb{R}_+^*)^{n+1}, \quad \sum_{i=1}^{n+1} z_i = 1 \right\}. \tag{4.6}$$

Bearing condition (4.5) in mind, it is natural to consider an equivalent reformulation of the system (4.3) using the fact that u_0 has to be equal to $1 - \sum_{i=1}^n u_i$, so that $u := (u_1, \cdots, u_n)$ is solution to

$$\left\{ \begin{array}{rl} \partial_t u - \mathrm{div}_x \left(A(u) \nabla_x u \right) = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \Omega, \\ A(u) \nabla_x u \cdot \mathbf{n} = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \partial\Omega, \end{array} \right. \tag{4.7}$$

with initial condition $u^0 = (u_1^0, \cdots, u_n^0)$, and where $A : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ is given by:

$$\forall u = (u_1, \cdots, u_n) \in \mathbb{R}^n, \ \forall 1 \leq i, j \leq n, \quad A_{ij}(u) = \mathcal{A}_{ij} \left(1 - \sum_{i=1}^n u_i, u \right) - \mathcal{A}_{i0} \left(1 - \sum_{i=1}^n u_i, u \right).$$

A solution u to (4.7) is then expected to take values in $\overline{\mathcal{D}} \subset [0,1]^n$ where

$$\mathcal{D} := \left\{ z \in (\mathbb{R}_+^*)^n, \quad \sum_{i=1}^n z_i < 1 \right\}. \tag{4.8}$$

The analysis of cross-diffusion systems of the form (4.3) (or equivalently of the form (4.7) is a challenging task from a mathematical point of view [107, 3, 95, 135, 34, 35, 52, 87, 155, 74, 131, 88, 108] for the following reasons:

- The equations are strongly nonlinearly coupled. As a consequence, standard tools such as the maximum/minimum principle do not apply in general. Besides, there is no regularity theory as in the scalar case. Nice counterexamples are given in [142]: there exist Hölder continuous solutions to certain cross-diffusion systems which blows up in finite time, and there exist bounded weak solutions which develop singularities in finite time.

- The diffusion matrix A or $\mathcal{A}$ is in general not elliptic and may be degenerate. Thus, even the local-in-time existence of solutions is not guaranteed.

- Solutions to (4.3) or (4.7) are not guaranteed to satisfy the volumic constraints (4.5) in general. Thus, these upper and lower bounds must be shown to be satisfied. But, as already mentioned, the standard tools as maximum/minimum principle do not apply in general.

- In addition, proving uniqueness of weak or strong solutions is in general out of reach for these systems.

Several attempts have been proposed in the mathematical litterature to overcome some of these difficulties.

Existence of weak (or strong) solutions

Amann developed a theory of parabolic systems in [4, 5] where the diffusion matrices $\mathcal{A}$ or A are assumed to be normal parabolic. We recall the definition of normal parabolicity below.

Definition 4.2.1. *A system of the form (4.7) is said to be parabolic if the diffusion matrix $A(u)$ is elliptic for all $u \in \overline{\mathcal{D}}$, i.e. if*

$$\forall u \in \overline{\mathcal{D}}, \quad \det\left(\frac{1}{2}(A(u) + A(u)^T)\right) > 0,$$

and said to be normal parabolic if the diffusion matrix $A(u)$ is normally elliptic for all $u \in \overline{\mathcal{D}}$, i.e. if

$$\forall u \in \overline{\mathcal{D}}, \quad \sigma(A(u)) \subset \{z \in \mathbb{C},\ \mathrm{Re}(z) > 0\},$$

where Re *denotes the real part of a complex number.*

This enables to prove the existence of local-in-time classical solutions for initial conditions in $W^{1,p}$. He also showed that the existence of global-in-time solutions is reduced to deriving suitable $W^{1,p}$ bounds for the local solutions. In particular, the following alternative holds: either the $W^{1,p}$ norm of the local-in-time solutions explodes in finite time, or the global-in-time solutions exist.

As mentioned above, the question of regularity of the solutions is a difficult problem. As remarked in [142, 53] and unlike in the scalar case, one cannot expect in general that bounded weak solutions to cross-diffusion systems are Hölder continuous everywhere. For some particular systems with smooth diffusion matrices, partial regularity results were established in [67]. The everywhere Hölder continuity was investigated in [84] only for low dimensional systems $d \leq 2$ and in [152] for an arbitrary space dimension $d \in \mathbb{N}^*$ but with rather restrictive structural conditions. The everywhere regularity of the weak solutions to possibly degenerate systems of the form was investigated in [99]. Sufficient conditions for the everywhere Hölder continuity of the solutions are given for arbitrary space dimension under several structural assumptions of the diffusion matrix.

The mathematical understanding of multi-species cross-diffusion systems defined on fixed domains with no-flux boundary conditions greatly improved in the last years [97, 6, 100, 74]. It was in particular understood that the decay of some entropy and the control of its dissipation is of paramount interest [87, 26, 108, 48].

Indeed, it appears that some of these cross-diffusion systems have a formal gradient flow structure. Recently, an elegant idea, which consists in introducing an entropy density that appears to be a Lyapunov functional for these systems, has been introduced and developped in [26, 87]. This analysis strategy, which was later extended by Jüngel in [87] and named *boundedness by entropy* technique, enables to obtain the existence of global in time weak solutions satisfying (4.2) under suitable assumptions on the diffusion matrix A. It was successfully applied in several contexts (see for instance [88, 85, 155, 156]). We present this entropy structure and illustrate the results proved in [26, 87] on a particular example in Section 4.2.2.

Let us also mention another method for the analysis of such cross-diffusion systems, which was developped by Desvillettes, Lepoutre, Moussa and collaborators, called *duality method*. The main idea of this method is to adapt the a priori duality estimates proved in [132] in order to obtain uniform L^2 bounds in addition to the entropy bounds stemming from the entropy dissipation property. This method was developped mainly for the analysis of generalized SKT systems where solutions are not expected to satisfy volumic constraints of the form (4.5).

Uniqueness of solutions

For systems which cannot be written as a set of fully decoupled equations, there exist few methods which enable to prove uniqueness of solutions, like for instance the H^{-1} method or the Gajewski method [64, 65, 155]. All these methods heavily rely on the fact that the cross-diffusion system has to satisfy some very specific structure. In the so-called H^{-1} method for instance, $A(u)\nabla u$ has to be written as $\nabla\Psi(u)$ for some function $\Psi : \mathbb{R}^n \to \mathbb{R}$ which satisfies the monotonicity proprety

$$\forall u, v \in \mathcal{D}, \quad \langle \nabla\Psi(u) - \nabla\Psi(v), u - v\rangle \geq 0.$$

Let us finally mention that other non-general uniqueness results can be obtained in some particular cases. We mention for example [21, 79] where the uniqueness of local-in-time solutions to the Stefan-Maxwell system were proved. In [68, 26], the uniqueness of the global-in-time weak solutions is obtained for an initial condition that is sufficiently close to the constant steady states.

One contribution of this manuscript, based on the joint work [VE16] with Judith Berendsen, Martin Burger and Jan-Frederik Pietschmann, is the proof of the uniqueness of strong solutions to the particular example presented in Section 4.2.2, which does not have the specific structure mentioned above, at the price of making assumptions on the value of the cross-diffusion coefficients appearing in the system. This contribution is detailed in Section 4.2.4.

4.2.2 Formal gradient flow structure of a particular cross-diffusion system

In this section, we illustrate the formal gradient flow structure of one particular example of system of cross-diffusion equations, which is of particular interest for the simulation of PVD processes mentioned above. This system, with no-flux boundary conditions, is studied in [VE12, VE16] and reads as follows : for any $0 \leq i \leq n$,

$$\begin{cases} \partial_t u_i - \mathrm{div}_x \left(\displaystyle\sum_{0\leq j\neq i\leq n} K_{ij}(u_j\nabla_x u_i - u_i\nabla_x u_j) \right) = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \Omega, \\ \left(\displaystyle\sum_{0\leq j\neq i\leq n} K_{ij}(u_j\nabla_x u_i - u_i\nabla_x u_j) \right) \cdot \mathbf{n} = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \partial\Omega, \end{cases} \tag{4.9}$$

where the positive real numbers K_{ij} satisfy

$$\forall 0 \leq i \neq j \leq n, \quad K_{ij} = K_{ji} > 0.$$

These coefficients represent the cross-diffusion coefficients of the chemical species of type i with the chemical species of type j. This set of equations can be formally derived from a discrete stochastic lattice hopping model, which is detailed in the Appendix.

System (4.9) is of the form (4.3) with

$$\mathcal{A}_{ij}(u_0, \cdots, u_n) = -K_{ij}u_i \text{ if } i \neq j \text{ and } \mathcal{A}_{ii}(u_0, \cdots, u_n) = -\sum_{j=0, j\neq i}^{n} K_{ij}u_j. \tag{4.10}$$

Under the assumption that $u_0 = 1 - \sum_{i=1}^n u_i$, it can be easily checked that the system can be rewritten under the form (4.7) with $A(u)$ defined by

$$\left\{ \begin{array}{l} \forall 1 \leq i \leq n, \quad A_{ii}(u) = \displaystyle\sum_{1 \leq j \neq i \leq n} (K_{ij} - K_{i0}) u_j + K_{i0}, \\ \forall 1 \leq i \neq j \leq n, \quad A_{ij}(u) = -(K_{ij} - K_{i0}) u_i. \end{array} \right. \tag{4.11}$$

We detail here the formal gradient flow structure of this particular system. To this aim, it is more convenient to consider the formulation (4.7) with A given by (4.11). Let us introduce the classical *logarithmic entropy density* h (see for instance [26, 87, 156, 112]) defined by

$$h : \left\{ \begin{array}{ccc} \overline{\mathcal{D}} & \longrightarrow & \mathbb{R} \\ u := (u_i)_{1 \leq i \leq n} & \longmapsto & h(u) := \displaystyle\sum_{i=1}^n u_i \log u_i + (1 - \rho_u) \log(1 - \rho_u), \end{array} \right. \tag{4.12}$$

where $\rho_u := \sum_{i=1}^n u_i$. Some properties of h can be easily checked:

(P1) the function h belongs to $\mathcal{C}^0(\overline{\mathcal{D}}) \cap \mathcal{C}^2(\mathcal{D})$; consequently, h is bounded on $\overline{\mathcal{D}}$;

(P2) the function h is strictly convex on $\mathcal{D}$;

(P3) its gradient

$$Dh : \left\{ \begin{array}{ccc} \mathcal{D} & \longrightarrow & \mathbb{R}^n \\ (u_i)_{1 \leq i \leq n} & \mapsto & \left(\log \left(\frac{u_i}{1 - \rho_u} \right) \right)_{1 \leq i \leq n}, \end{array} \right.$$

is invertible and its inverse is given by

$$(Dh)^{-1} : \left\{ \begin{array}{ccc} \mathbb{R}^n & \longrightarrow & \mathcal{D} \\ (w_i)_{1 \leq i \leq n} & \mapsto & \frac{e^{w_i}}{1 + \sum_{j=1}^n e^{w_j}}. \end{array} \right.$$

In the following, we denote by $D^2 h$ the Hessian of h. The *entropy functional* $\mathcal{E}$ associated to h is defined by

$$\mathcal{E} : \left\{ \begin{array}{ccc} L^\infty(\Omega; \overline{\mathcal{D}}) & \longrightarrow & \mathbb{R} \\ u & \longmapsto & \mathcal{E}(u) := \displaystyle\int_\Omega h(u(x))\, dx. \end{array} \right. \tag{4.13}$$

Throughout the chapter, for all $u \in L^\infty(\Omega; \mathcal{D})$, we shall denote by $D\mathcal{E}(u)$ the measurable vector-valued function defined by

$$D\mathcal{E}(u) : \left\{ \begin{array}{ccc} \Omega & \to & \mathbb{R}^n \\ x & \mapsto & Dh(u(x)). \end{array} \right.$$

System (4.7) can then be formally rewritten under the following gradient flow structure

$$\left\{ \begin{array}{ll} \partial_t u - \operatorname{div}_x \left(M(u) \nabla_x D\mathcal{E}(u) \right) = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \Omega, \\ (M(u) \nabla_x D\mathcal{E}(u)) \cdot \mathbf{n} = 0, & \text{for } (t,x) \in \mathbb{R}_+^* \times \partial\Omega, \end{array} \right. \tag{4.14}$$

where $M : \overline{\mathcal{D}} \to \mathbb{R}^{n\times n}$ is the so-called *mobility matrix* of the system. It holds that

$$\forall u \in \mathcal{D}, \quad M(u) := A(u)(D^2h(u))^{-1}.$$

Moreover, when A is given by (4.11), the components of $M(u)$ read, for all $u \in \overline{\mathcal{D}}$ and all $1 \leq i \neq j \leq n$,

$$M_{ii}(u) = K_{i0}(1-\rho_u)u_i + \sum_{1\leq j\neq i\leq n} K_{ij}u_iu_j \quad \text{and} \quad M_{ij}(u) = -K_{ij}u_iu_j. \tag{4.15}$$

4.2.3 Boundedness by entropy method for the existence of weak solutions

The formal gradient flow formulation of a system of cross-diffusion equations is a key point in the boundedness by entropy technique. In the example presented in Section 4.2.2, it implies in particular that $\mathcal{E}$ is a Lyapunov functional for the system (4.3) [26, 87]. The existence of a global weak solution to (4.9) can be obtained, using Theorem 2 of [87], whose proof heavily relies on the existence of such a formal gradient flow structure. We recall here a simplified version of the latter theorem which is adapted to our context.

Theorem 4.2.1 (Theorem 2 of [87])**.** *Let $\mathcal{D} \subset \mathbb{R}^n$ be the domain defined by (4.8). Let $A : u \in \overline{\mathcal{D}} \mapsto A(u) := (A_{ij}(u))_{1\leq i,j\leq n} \in \mathbb{R}^{n\times n}$ be a matrix-valued functional defined on $\overline{\mathcal{D}}$ satisfying $A \in \mathcal{C}^0(\overline{\mathcal{D}}; \mathbb{R}^{n\times n})$ and the following assumptions:*

(H1) There exists a bounded from below convex function $h \in \mathcal{C}^2(\mathcal{D}, \mathbb{R})$ such that its derivative $Dh : \mathcal{D} \to \mathbb{R}^n$ is invertible on $\mathbb{R}^n$;

(H2) There exists $\alpha > 0$, and for all $1 \leq i \leq n$, there exist $1 \geq m_i > 0$, such that for all $z = (z_1, \cdots, z_n)^T \in \mathbb{R}^n$ and $u = (u_1, \cdots, u_n)^T \in \mathcal{D}$,

$$z^T D^2h(u)A(u)z \geq \alpha \sum_{i=1}^{n} u_i^{2m_i-2} z_i^2.$$

Let $u^0 \in L^1(\Omega; \mathcal{D})$ so that $w^0 := Dh(u^0) \in L^\infty(\Omega; \mathbb{R}^n)$. Then, there exists a weak solution u with initial condition u^0 to (4.7) such that for almost all $(t, x) \in \mathbb{R}_+^* \times \Omega$, $u(t, x) \in \overline{\mathcal{D}}$ with

$$u \in L^2_{\text{loc}}(\mathbb{R}_+; H^1(\Omega, \mathbb{R}^n)) \text{ and } \partial_t u \in L^2_{\text{loc}}(\mathbb{R}_+; (H^1(\Omega; \mathbb{R}^n))').$$

Building on ideas from [156], it was remarked in [VE12] that the prototypical example presented in Section 4.2.2 falls into the framework of Theorem 4.2.1. This is a consequence of Lemma 4.2.1, the proof of which can be found in [VE12].

Lemma 4.2.1. *Let $\mathcal{D} \subset \mathbb{R}^n$ be the domain defined by (4.8) and A be the matrix-valued function defined by (4.11). Then, $A \in \mathcal{C}^0(\overline{\mathcal{D}}; \mathbb{R}^{n\times n})$ satisfies assumptions (H1)-(H2) of Theorem 4.2.1, with h given by (4.12), $\alpha = \min_{1\leq i\neq j\leq n} K_{ij} > 0$ and $m_i = \frac{1}{2}$ for all $1 \leq i \leq n$.*

The existence of global weak solutions to (4.9) satisfying volumic constraints (4.5) is then a direct consequence of Theorem 4.2.1 and Lemma 4.2.1. We equivalently formulate this result in terms of existence of a weak solution to (4.3) in the following proposition for later reference.

Proposition 4.2.1. *Let $\mathcal{P} \subset \mathbb{R}^{n+1}$ be the domain defined by (4.6) and $\mathcal{A}$ be the matrix-valued function defined by (4.10). Let $U^0 = (u_0^0, \cdots, u_n^0) \in L^1(\Omega; \mathcal{P})$ such that $u^0 := (u_1^0, \cdots u_n^0)$ satisfies the assumptions of Theorem 4.2.1. Then, there exists a weak solution U with initial condition U^0 to (4.3) such that for almost all $(t,x) \in \mathbb{R}_+^* \times \Omega$, $U(t,x) \in \overline{\mathcal{P}}$ with*

$$U \in L^2_{\text{loc}}(\mathbb{R}_+; H^1(\Omega, \mathbb{R}^{n+1})) \text{ and } \partial_t U \in L^2_{\text{loc}}(\mathbb{R}_+; (H^1(\Omega; \mathbb{R}^{n+1}))').$$

Uniqueness or existence of **strong** of solutions to general systems of the form (4.3) or (4.7) remains an open theoretical question, at least up to our knowledge. Some partial results, which were proved in [VE16] in a joint work with Judith Berendsen, Martin Burger and Jan-Frederik Pietschmann, can be obtained for the particular system (4.9) under additional assumptions on the value of the cross-diffusion coefficients $(K_{ij})_{0\leq i\neq j\leq n}$. These results are presented in the following section.

4.2.4 Existence and uniqueness of strong solutions for the particular example

The aim of the joint work [VE16] with Judith Berendsen, Martin Burger and Jan-Frederik Pietschmann, is to study the existence and uniqueness of *strong* solutions to system (4.9) satisfying (4.5).

To obtain such existence and uniqueness result, an additional assumption on the value of cross-diffusion coefficients $(K_{ij})_{0\leq i\neq j\leq n}$ is required. For all $0 \leq i \leq n$, let

$$K^+ := \max_{0\leq j\neq i\leq n} K_{ij}, \quad K^- := \min_{0\leq j\neq i\leq n} K_{ij}, \quad K := \frac{K^+ + K^-}{2} \quad \text{and } \kappa := \frac{K^+ - K^-}{2}. \tag{4.16}$$

Let us point out that definitions (4.16) implies that for all $0 \leq j \neq i \leq n$, $|K_{ij} - K| \leq \kappa$.

The additional assumption on the cross-diffusion coefficients reads as follows:

Assumption 1. *It holds that $2n\kappa < K$.*

In other words, Assumption 1 means that all the coefficients K_{ij} should be sufficiently close to one another. The motivation for considering such a situation stems from the following observation: if there exists a constant $K > 0$ such that for all $0 \leq i \neq j \leq n$, $K_{ij} = K$, then $\kappa = 0$ and system (4.9) boils down to a system of $n+1$ independent heat equations for which the existence and uniqueness of strong solutions satisfying (4.5) is obvious.

We conjecture here that Assumption (1) is technical for the existence and uniqueness of strong solutions of system (4.3) (or equivalently (4.7). Lifting this assumption will be the object of future research.

We are now in position to state the two main results of [VE16]. For this analysis, considering formulation (4.3) appeared to be more convenient than considering formulation (4.7). This is the reason why we state here the results in terms of existence and uniqueness of the solution $U = (u_0, \cdots, u_n)$ (rather than $u = (u_1, \cdots u_n)$).

Theorem 4.2.2 (Existence and uniqueness of strong solutions)**.** *Let $d \leq 3$, $T > 0$ and let us assume that Assumption 1 holds. Let $U^0 \in [H^1(\Omega)]^{n+1}$, with $U^0(x) \in \overline{\mathcal{P}}$ for almost all $x \in \Omega$. Then, there exists one and only one strong solution U to* (4.3) *with $\mathcal{A}$ defined by (4.10) and with initial condition U^0 such that*

(i) $U \in [L^2((0,T), H^2(\Omega)) \cap H^1((0,T), L^2(\Omega))]^{n+1}$,

(ii) $U(t,x) \in \overline{\mathcal{P}}$ *for almost all* $(t,x) \in (0,T) \times \Omega$.

For the particular case when $d = 1$, we can also prove a weak-strong stability result which implies that there exists a unique weak solution to the system (4.9) satisfying (4.5) and that this solution is strong.

Theorem 4.2.3. *(Weak-strong stability estimate in $d = 1$) Let us assume that $d = 1$ and that Assumption 1 holds. Let $\tilde{U}$ be a weak solution to* (4.3) *with $\mathcal{A}$ defined by (4.10) in the sense of Proposition 4.2.1, and let U be the strong solution in the sense of Theorem 4.2.2. Then, there exists a constant $C > 0$ such that the following stability estimate holds for all $0 < t \leq T$:*

$$\|U(t,\cdot) - \tilde{U}(t,\cdot)\|^2_{L^2(\Omega)} \leq e^{C\|\nabla U\|^2_{L^2(0,t;L^\infty(\Omega))}} \|U(0,\cdot) - \tilde{U}(0,\cdot)\|^2_{L^2(\Omega)}. \tag{4.17}$$

In particular, if the corresponding initial data $U(0,\cdot)$ and $\tilde{U}(0,\cdot)$ agree a.e. on Ω, we also have

$$U = \tilde{U} \text{ a.e. in } \Omega \times (0,T).$$

Theorem 4.2.3 is restricted to spatial dimension 1 since the proof relies on the embedding $H^2(\Omega) \hookrightarrow W^{1,\infty}(\Omega)$, so that a strong solution $U \in [L^2((0,T), H^2(\Omega)) \cap H^1((0,T), L^2(\Omega))]^{n+1}$ in the sense of Theorem 4.2.2 satisfies $\nabla U \in \left[L^2(0,T; L^\infty(\Omega))^d\right]^{n+1}$.

4.2.5 Research perspectives on cross-diffusion systems on fixed domains with no-flux boundary conditions

Let me present here some research perspectives I would like to address in the future, related to cross-diffusion systems on fixed boundary domains with no-flux boundary conditions.

Let me begin with some short-term objectives.

- We proved the existence and uniqueness of strong solutions of (4.9) under the assumption that value of the cross-diffusion coefficients should be close enough to one single positive constant. In my opinion, such results could be also proved for the Stefan-Maxwell system [88]. Indeed, both systems share common features, even though the formulation of the Stefan-Maxwell system is slightly more complicated than the formulation of (4.9). Since the Stefan-Maxwell system is widely used by materials scientists, such an extension would be very interesting from a practical point of view.

- Recently, a finite volume numerical scheme for the discretization of system (4.9) was proposed and studied in [29]. With Laurent Monasse and Clément Cancès, we are currently working on the generalization of such a scheme for more general cross-diffusion systems, including the Stefan-Maxwell system, with similar desirable mathematical properties.

Let us now present some long-term issues, which lead to potentially quite hard (but very interesting and challenging!) problems.

- System (4.9) (as a large number of cross-diffusion models) is a *phenomenological* model, in the sense that its formulation is derived from formal calculations or arguments. On the other hand, more realistic (but more complicated) cross-diffusion systems can be rigorously identified as hydrodynamic limits of stochastic particle systems (see [134] for instance). The analysis of such systems may be more intricate, due to the fact that the expression of the diffusion matrix $\mathcal{A}$ or A is not known explicitly. For instance, in the system identified in [134], in the expression of $\mathcal{A}$ appears the so-called *self-diffusion coefficient*, whose value depends on the long-time limit of the expectation of a random quantity, obtained using a Symmetric Simple Exclusion Process [98]. Whether such a system satisfies a formal gradient flow structure similar to the one of (4.9) is not clear. This makes the study of some mathematical properties of such a system like its long-time behaviour difficult and challenging. The study of dedicated numerical schemes for the resolution of such systems is also a very interesting field of study I would like to address in the future.

- Consider a (stochastic) particle system involving a very large number of particles. Prototypical examples of such situations may be encountered in Molecular Dynamics or Kinetic Monte-Carlo simulations, where the evolution of the particles is modeled either by a continuous or discrete state space Markov process. The simulation of the evolution of the positions of each particle is in general very expensive from a computational point of view, and may not be necessary except in some small parts of the physical domain where the most relevant phenomena occur. In such situations, it may be very interesting from a practical point of view to couple a stochastic microscopic particle model in a small region of interest with a continuous model, which may be seen as some kind of hydrodynamic limit for the particle system, in the remainder of the domain. For instance, in the case of the PVD process mentioned above, it may be of interest to keep track of the way the atoms deposit onto the surface of the substrate using a stochastic particle model, and to model the bulk diffusion phenomena of the different chemical species with a continuous cross-diffusion model. Such a coupling could, in principle, help in accelerating such simulations. However, how to couple both kinds of models with a sound mathematical ground is not clear at all. However, there may be hope in cases where both models share gradient flow structures that are connected to one another. For instance, it has been shown in [60], in the particular case of the Symmetric Simple Exclusion Process, that some gradient flow structures of the microscopic Markov process and its hydrodynamic limit (which leads to a simple heat equation) are strongly connected. Even for this simple system, the use of both gradient flow structures in order to define a consistent

mathematical scheme to couple the microscopic stochastic process together with a continuous hydrodynamic limit model has not been studied at least up to my knowledge. One desirable feature of such a scheme, in analogy with atomistic-to-continuum schemes used in *deterministic* micro-macro simulations [118, 128, VE10], should be its convergence, at least in some sense, to the full hydrodynamic limit model in the limit of a large number of particles.

4.3 One-dimensional cross-diffusion systems on moving domains

There are very few works which focus on the analysis of cross-diffusion systems with non zero-flux boundary conditions and time-dependent domains. To my knowledge, only systems containing at most two different species have been studied, so that $n = 1$ and the evolution of the concentrations inside the domain are decoupled and follow independent linear heat equations [133].

In this section, I present the contribution of [VE12], where a one-dimensional model for the PVD process presented in Section 4.1 was proposed and analyzed. The model is presented in Section 4.3.1 and the main theoretical results of [VE12] are summarized in Section 4.3.2. Research perspectives are outlined in Section 4.3.3.

4.3.1 Presentation of the model

In the sequel, the study is restricted to the case when $d = 1$. For the sake of simplicity, we assume that non-zero fluxes are only imposed on the right-hand side of the domain occupied by the solid. At some time $t > 0$, this domain is denoted by $\Omega_t := (0, e(t))$ where $e(t) > 0$ models the thickness of the layer. Initially, we assume that the domain Ω_0 occupied by the solid at time $t = 0$ is the interval $(0, e_0)$ for some initial thickness $e_0 > 0$.

The evolution of the thickness of the film $e(t)$ is determined by the external fluxes of the atomic species that are absorbed at its surface. More precisely, let us assume that there are $n + 1$ different chemical species composing the solid layer and let $(\phi_0, \cdots, \phi_n)$ belong to $L^\infty_{\rm loc}(\mathbb{R}_+; \mathbb{R}^{n+1}_+)$. For all $0 \leq i \leq n$, the function $\phi_i(t)$ represents the flux of the species i absorbed at the surface at time $t > 0$ and is assumed to be non-negative. In this one-dimensional model, the evolution of the thickness of the solid is assumed to be given by

$$e(t) := e_0 + \int_0^t \sum_{i=0}^{n} \phi_i(s)\, ds. \tag{4.18}$$

In the following, we will denote by $\varphi := (\phi_1, \cdots, \phi_n)^T$.

For all $t \geq 0$ and $0 \leq i \leq n$, the local concentration of species i at time t and point $x \in (0, e(t))$ is denoted by $u_i(t, x)$. The evolution of the vector $u := (u_1, \cdots, u_n)$ is given by the system of cross-diffusion equations

$$\partial_t u - \partial_x \left(A(u)\partial_x u\right) = 0, \text{ for } t \in \mathbb{R}^*_+,\ x \in (0, e(t)), \tag{4.19}$$

where $A : \overline{\mathcal{D}} \to \mathbb{R}^{n\times n}$ is a well-chosen diffusion matrix satisfying (H1)-(H2) of Theorem 4.2.1.

We consider that for every $t > 0$, the system satisfies the following conditions on the boundary $\partial\Omega_t$:

$$(A(u)\partial_x u)(t,0) = 0 \text{ and } (A(u)\partial_x u)(t,e(t)) + e'(t)u(t,e(t)) = \varphi(t). \tag{4.20}$$

An easy calculation shows that these boundary conditions, in addition to (4.18) and (4.19), ensure that, for all $0 \le i \le n$,

$$\frac{d}{dt}\left(\int_{\Omega_t} u_i(t,x)\,dx\right) = \phi_i(t).$$

Indeed, it holds that

$$\begin{aligned}
\frac{d}{dt}\left(\int_{\Omega_t} u(t,x)\,dx\right) &= \int_0^{e(t)} \partial_t u(t,x)\,dx + e'(t)u(t,e(t)) \\
&= \int_0^{e(t)} \partial_x\left(A(u)\partial_x u\right) + e'(t)u(t,e(t)) \\
&= (A(u)\partial_x u)(t,e(t)) + e'(t)u(t,e(t)) - (A(u)\partial_x u)(t,0) \\
&= \varphi(t).
\end{aligned}$$

The calculation for the 0^{th} species reads:

$$\begin{aligned}
\frac{d}{dt}\left(\int_{\Omega_t} u_0(t,x)\,dx\right) &= \frac{d}{dt}\left(|\Omega_t| - \sum_{i=1}^n \int_{\Omega_t} u_i(t,x)\,dx\right) \\
&= e'(t) - \sum_{i=1}^n \frac{d}{dt}\left(\int_{\Omega_t} u_i(t,x)\,dx\right) \\
&= \sum_{i=0}^n \phi_i(t) - \sum_{i=1}^n \phi_i(t) = \phi_0(t).
\end{aligned}$$

To sum up, the final system of interest reads:

$$\begin{cases}
e(t) = e_0 + \int_0^t \sum_{i=0}^n \phi_i(s)\,ds, & \text{for } t \in \mathbb{R}_+^*, \\
\partial_t u - \partial_x\left(A(u)\partial_x u\right) = 0, & \text{for } t \in \mathbb{R}_+^*,\ x \in (0,e(t)), \\
(A(u)\partial_x u)(t,0) = 0, & \text{for } t \in \mathbb{R}_+^*, \\
(A(u)\partial_x u)(t,e(t)) + e'(t)u(t,e(t)) = \varphi(t), & \text{for } t \in \mathbb{R}_+^*, \\
u(0,x) = u^0(x), & \text{for } x \in (0,e_0),
\end{cases} \tag{4.21}$$

where $u^0 \in L^1(0,e_0)$ is an initial condition satisfying $u^0(x) \in \mathcal{D}$ for almost all $x \in (0,e_0)$. We assume in addition that $w^0 := Dh(u^0)$ belongs to $L^\infty((0,e^0);\mathbb{R}^n)$.

Rescaled version of the model 4.21

We introduce here a rescaled version of system (4.21). For all $0 \leq i \leq n$, $t \geq 0$ and $y \in (0,1)$, let us denote by $v_i(t,y) := u_i(t, e(t)y)$. It holds that

$$\partial_t v(t,y) = \partial_t u(t,e(t)y) + e'(t)y\partial_x u(t,e(t)y) \text{ and } \partial_y v(t,y) = e(t)\partial_x u(t,e(t)y),$$

where $v := (v_1, \cdots, v_n)$. Thus, u is a solution of (4.21) if and only if v is a solution to the following system:

$$\begin{cases} e(t) = e_0 + \displaystyle\int_0^t \sum_{i=0}^{n} \phi_i(s)\, ds, & \text{for } t \in \mathbb{R}_+^*, \\ \partial_t v - \frac{1}{e(t)^2}\partial_y \left(A(v)\partial_y v\right) - \frac{e'(t)}{e(t)} y\partial_y v = 0, & \text{for } (t,y) \in \mathbb{R}_+^* \times (0,1), \\ \frac{1}{e(t)}(A(v)\partial_y v)(t,1) + e'(t)v(t,1) = \varphi(t), & \text{for } (t,y) \in \mathbb{R}_+^* \times (0,1), \\ \frac{1}{e(t)}(A(v)\partial_y v)(t,0) = 0, & \text{for } (t,y) \in \mathbb{R}_+^* \times (0,1) \\ v(0,y) = v^0(y), & \text{for } y \in (0,1), \end{cases} \tag{4.22}$$

where $v^0(y) := u^0(e_0 y)$.

Proving the existence of a global weak solution to (4.21) is equivalent to proving the existence of a global weak solution to (4.22).

Actually, it can be seen that the entropy of the system (4.22) satisfies a formal inequality at the continuous level, which is key in the proof of our existence result. Indeed, let us denote by

$$\mathcal{E}(t) := \int_0^1 h(v(t,y))\, dy,$$

where v is a solution to (4.22). Then, formal calculations yield that

$$\begin{aligned} \frac{d\mathcal{E}}{dt}(t) &= \int_0^1 \partial_t v(t,y) \cdot Dh(v(t,y))\, dy \\ &= \frac{1}{e(t)^2}\int_0^1 \partial_y \left(A(v(t,y))\partial_y v(t,y)\right) \cdot Dh(v(t,y))\, dy + \frac{e'(t)}{e(t)}\int_0^1 y\partial_y v(t,y) \cdot Dh(v(t,y))\, dy \\ &= -\frac{1}{e(t)^2}\int_0^1 \partial_y v(t,y) \cdot D^2 h(v(t,y))A(v(t,y))\partial_y v(t,y)\, dy \\ &+ \frac{1}{e(t)^2}\left(A(v(t,1))\partial_y v(t,1)\right) \cdot Dh(v(t,1)) + \frac{e'(t)}{e(t)}\int_0^1 y\partial_y(h(v(t,y)))\, dy \\ &= -\frac{1}{e(t)^2}\int_0^1 \partial_y v(t,y) \cdot D^2 h(v(t,y))A(v(t,y))\partial_y v(t,y)\, dy + \frac{1}{e(t)}\left(\varphi(t) - e'(t)v(t,1)\right) \cdot Dh(v(t,1)) \\ &+ \frac{e'(t)}{e(t)}h(v(t,1)) - \frac{e'(t)}{e(t)}\int_0^1 h(v(t,y))\, dy. \end{aligned}$$

Denoting by $\overline{f}(t) := \frac{\varphi(t)}{e'(t)}$, it holds that $\overline{f}(t) \in \overline{\mathcal{D}}$ for all $t > 0$. Besides, using assumption (H2) of Theorem 4.2.1, we obtain that

$$-\int_0^1 \partial_y v(t,y) \cdot D^2 h(v(t,y))A(v(t,y))\partial_y v(t,y)\, dy \leq 0,$$

which yields that

$$\frac{d\mathcal{E}}{dt}(t) \leq \frac{e'(t)}{e(t)} \left[h(v(t,1) + Dh(v(t,1)) \cdot \big(\overline{f}(t) - v(t,1)\big) - \int_0^1 h(v(t,y))\, dy \right].$$

Using the convexity of h, we obtain that $h(v(t,1) + Dh(v(t,1)) \cdot \big(\overline{f}(t) - v(t,1)\big) \leq h(\overline{f}(t))$, so that

$$\frac{d\mathcal{E}}{dt}(t) \leq \frac{e'(t)}{e(t)} \left[h(\overline{f}(t)) - \mathcal{E}(t) \right]. \tag{4.23}$$

Inequality (4.23) is not an entropy dissipation inequality in the sense that the quantity $\mathcal{E}(t)$ may increase with time. However, using the fact $e' \in L^\infty_{\rm loc}(\mathbb{R}_+;\mathbb{R}_+)$ and assumption (H3), it implies that the quantity $\mathcal{E}(t)$ cannot blow up in finite time, which is sufficient for our purpose.

4.3.2 Theoretical results

Global in time existence of weak solutions

Our first result deals with the global in time existence of bounded weak solutions to (4.22) (and thus to (4.21)).

Theorem 4.3.1. *Let $\mathcal{D} := \{(u_1,\cdots,u_n)^T \in (\mathbb{R}_+^*)^n,\ \sum_{i=1}^n u_i < 1\} \subset (0,1)^n$. Let $A : \overline{\mathcal{D}} \to \mathbb{R}^{n\times n}$ be a matrix-valued functional satisfying $A \in \mathcal{C}^0(\overline{\mathcal{D}};\mathbb{R}^{n\times n})$ and assumptions (H1)-(H2) of Theorem 4.2.1 for some well-chosen entropy density $h : \overline{\mathcal{D}} \to \mathbb{R}$. We assume in addition that*

(H3) $h \in \mathcal{C}^0(\overline{\mathcal{D}})$.

Let $e_0 > 0$, $u^0 \in L^1((0,e_0);\mathcal{D})$ so that $w^0 := (Dh)^{-1}(u^0) \in L^\infty((0,e_0);\mathbb{R}^n)$ and $(\phi_0,\cdots,\phi_n) \in L^\infty_{\rm loc}(\mathbb{R}_+;\mathbb{R}_+^{n+1})$. Let us define for almost all $y \in (0,1)$, $v^0(y) := u^0(e_0 y)$ and $\varphi := (\phi_1,\cdots,\phi_n)^T$. Then, there exists a weak solution v with initial condition v^0 to (4.22) such that for almost all $(t,y) \in \mathbb{R}_+^* \times (0,1)$, $v(t,y) \in \overline{\mathcal{D}}$. Besides,

$$v \in L^2_{\rm loc}(\mathbb{R}_+; H^1((0,1);\mathbb{R}^n)) \text{ and } \partial_t v \in L^2_{\rm loc}(\mathbb{R}_+; (H^1((0,1);\mathbb{R}^n))').$$

In particular, $v \in \mathcal{C}^0(\mathbb{R}_+; L^2((0,1);\mathbb{R}^n))$.

Let us point out that the example described in Section 4.2.2 satisfies all the assumptions of Theorem 4.3.1 since the entropy density h defined by (4.12) belongs to $\mathcal{C}^0(\overline{\mathcal{D}})$. Let us also point here that the form of (4.22) is different from the system considered in [87] through (i) the boundary conditions and (ii) the existence of the drift term $\frac{e'(t)}{e(t)} y \partial_y v$.

The proof of Theorem 4.3.1 relies on a careful adaptation of the boundedness by entropy method developped in [26, 87].

Long-time behaviour for constant fluxes

In the case when the fluxes are constant in time, we obtain long-time asymptotics for the functions v_i, provided that the entropy density h is given by (4.12). More precisely, the following result holds:

Proposition 4.3.1. *Suppose that the assumptions of Theorem 4.3.1 hold, as well as the following additional hypotheses:*

(T1) for all $0 \leq i \leq n$, there exists $\overline{\phi}_i > 0$ so that $\phi_i(t) = \overline{\phi}_i$, for all $t \in \mathbb{R}_+$;

(T2) for all $u \in \overline{D}$, the entropy density h reads as $h(u) = \sum_{i=1}^n u_i \log u_i + (1 - \rho_u)\log(1-\rho_u)$.

For all $0 \leq i \leq n$, let us define $\overline{f}_i := \frac{\overline{\phi}_i}{\sum_{j=0}^n \overline{\phi}_j}$ and $\overline{f} := (\overline{f}_i)_{1\leq i\leq n} \in \mathcal{D}$. Let us also denote by

$$\overline{h} : \left\{ \begin{array}{ccc} \overline{\mathcal{D}} & \mapsto & \mathbb{R} \\ u & \mapsto & h(u) - h(\overline{f}) - Dh(\overline{f})(u - \overline{f}) \end{array} \right.$$

the relative entropy associated with h and $\overline{f}$. Then, there exists a global weak solution v to (4.22) and a constant $C > 0$ such that

$$\int_0^1 \overline{h}\left(v(t,y)\right)\, dy \leq \frac{C}{t+1}, \tag{4.24}$$

and

$$\forall 1 \leq i \leq n, \quad \|v_i(t,\cdot) - \overline{f}_i\|_{L^1(0,1)} \leq \frac{C}{\sqrt{t+1}} \text{ and } \left\|\left(1 - \rho_{v(t,\cdot)}\right) - \overline{f}_0\right\|_{L^1(0,1)} \leq \frac{C}{\sqrt{t+1}}. \tag{4.25}$$

Let us comment here on assumption (T2). For the sake of simplicity, we chose to restrict ourselves to the case of logarithmic entropy density in Proposition 4.3.1. Actually, Proposition 4.3.1 can be easily generalized provided that the relative entropy density $\overline{h}$ satisfies a generalized Csizar-Kullback type inequality [147].

The central ingredient of the proof is the following formal entropy inequality. In the case when h is given by (4.12), it can be easily seen that $\overline{h}$ is also a valid entropy density for the diffusion coefficient A in the sense that $\overline{h}$ also satisfies assumptions (H1)-(H2)-(H3). Thus, inequality (4.23) holds with $\overline{h}$ instead of h so that

$$\frac{d\overline{\mathcal{E}}}{dt}(t) \leq \frac{e'(t)}{e(t)}\left[\overline{h}(\overline{f}) - \int_0^1 \overline{h}(v(t,y))\, dy\right] = \frac{e'(t)}{e(t)}\left[\overline{h}(\overline{f}) - \overline{\mathcal{E}}(t)\right],$$

where for all $t > 0$, $\overline{\mathcal{E}}(t) := \int_0^1 \overline{h}(v(t,y))\, dy$. Denoting by $V := \sum_{i=0}^n \overline{\phi}_i$, it holds that $e'(t) = V$ and $e(t) = e_0 + Vt$ for all $t \geq 0$. Finally, using the fact that $\overline{h} \geq 0$ and that $\overline{h}(\overline{f}) = 0$, we obtain that

$$\left(\frac{e_0}{V} + t\right)\frac{d\overline{\mathcal{E}}}{dt}(t) + \overline{\mathcal{E}}(t) = \frac{d}{dt}\left(\left(\frac{e_0}{V} + t\right)\overline{\mathcal{E}}(t)\right) \leq 0.$$

This inequality implies that there exists a constant $C > 0$ such that for all $t \geq 0$,

$$\overline{\mathcal{E}}(t) \leq \frac{C}{t+1}.$$

The rates on the L^1 norm of the solutions are then obtained using the Csizàr-Kullback inequality.

Let us finally point out that the quantity $\int_0^1 h(v(t,y))\,dy = \frac{1}{e(t)}\int_0^{e(t)} h(u(t,x))\,dx$ can be seen as an average entropy. In particular, the result of Propositon 4.3.1 does not imply in general the convergence of $u(t,x)$ to a constant vector $L^1_{\text{loc}}(\mathbb{R}_+)$ for instance. Whether such a convergence may hold true remains an open question.

4.3.3 Research perspectives for cross-diffusion systems on time-dependent domains

Comparisons between numerical simulations of the model presented in Section 4.3.1 and experimental measurements on actual solar cells yielded encouraging results on the relevance of this approach [9]. However, this one-dimensional model suffers from several limitations. Since it does not allow to study geometrical effects due to surface tension or surfacic cross-diffusion phenomena which occur at the surface of the film. These phenomena are nevertheless extremely important to take into account, in particular for the production of curved solar cells for building-integrated photovoltaics.

There is a crucial need for overcoming these limitations and proposing a multi-dimensional model for the PVD process along with accurate and efficient numerical schemes for the approximation of its solutions, which can be used in order to optimize the production process of such thin film solar cells. This represents a significant scientific advance with respect to the existing models and numerical methods which I wish to study in the future. I am the Principal Investigator of an ANR JCJC project (COMODO, short for CrOss-diffusion systems in MOving Domains) which started on the 1st of January 2020, and the objectives of which are to address these issues. The other members of the project are Martin Burger, Clément Cancès, Laurent Monasse and Jan-Frederik Pietschmann.

In this project, four main tasks are identified:

- a first task consists in identifying appropriate models for the evolution of the local volumic fractions of the various chemical species inside the film and of its surface. Such models read as cross-diffusion systems defined on a domain with moving boundary, taking into account surface cross-diffusion phenomena. We wish to justify such models by the means of an asymptotic analysis in the sharp interface limit of multi-species Cahn-Hilliard like models.

- the second task aims at developing numerical schemes for such models, which should respect the mathematical properties of the considered systems. Appropriate numerical methods such as Arbitrary Lagrangian Eulerian methods, surface finite elements or the Embedded boundary method should be considered as well to treat the fact that the domain also evolves through the PVD process;

- the third task concerns the parallelization of the obtained algorithms for the simulation of large-scale problems;
- the last task consists in calibrating the obtained models with experimental data given by researchers of the Institut Photovoltaïque d'Ile-de-France (IPVF) in order to select the values of the parameters involved (typically the value of some cross-diffusion coefficients for instance). To perform this task, the construction of an adapted reduced-order model shall be necessary. This raises several challenging mathematical difficulties, among which the treatment of the evolution of the domain. We then wish to use this calibrated simulation tool in order to optimize the fabrication process of thin film solar cells, so that the final geometry of the film and volumic fraction profiles of the different chemical components become as close as possible to well-chosen targets.
- Let us make here a last comment: model (4.3) is unable to model segragation phenomena, which do happen in the thin film layer during the PVD process. In order to reproduce such physical phenomena, it will be very interesting in the future to consider cross-diffusion models inside the layer that account for the segregatino of some chemical species during the fabrication process.

Appendix: Formal derivation of system (4.9)

We present in this section a simplified formal derivation of the cross-diffusion model (4.9) from a one-dimensional microscopic lattice hopping model with size exclusion, in the same spirit than the one proposed in [26].

We consider here a solid occupying the whole space $\mathbb{R}$ and discretize the domain using a uniform grid of step size $\Delta x > 0$. At any time $t \in [0,T]$, we denote by $u_i^{k,t}$ the number of atoms of type i ($0 \le i \le n$) in the k^{th} interval $[k\Delta x, (k+1)\Delta x)$ ($k \in \mathbb{Z}$). Let $\Delta t > 0$ denote a small enough time step. We assume that during the time interval Δt, an atom i located in the k^{th} interval can exchange its position with an atom of type j ($j \neq i$) located in one of the two neighbouring intervals with probability $p_{ij} = p_{ji} > 0$. On average, we obtain the following evolution equation for $u_i^{k,t}$:

$$\begin{aligned} u_i^{k,t+\Delta t} - u_i^{k,t} &= \sum_{0\le j\neq i\le n} p_{ij}\left(u_i^{k+1,t}u_j^{k,t} + u_i^{k-1,t}u_j^{k,t} - u_i^{k,t}u_j^{k+1,t} - u_i^{k,t}u_j^{k-1,t}\right) \\ &= \sum_{0\le j\neq i\le n} p_{ij}\left[u_j^{k,t}\left(u_i^{k+1,t} + u_i^{k-1,t} - 2u_i^{k,t}\right) - u_i^{k,t}\left(u_j^{k+1,t} + u_j^{k-1,t} - 2u_j^{k,t}\right)\right]. \end{aligned}$$

This yields that

$$\frac{u_i^{k,t+\Delta t} - u_i^{k,t}}{\Delta t} = \frac{2\Delta x^2}{\Delta t}\sum_{0\le j\neq i\le n} p_{ij}\left[u_j^{k,t}\frac{u_i^{k+1,t} + u_i^{k-1,t} - 2u_i^{k,t}}{2\Delta x^2} - u_i^{k,t}\frac{u_j^{k+1,t} + u_j^{k-1,t} - 2u_j^{k,t}}{2\Delta x^2}\right].$$

Choosing Δt and Δx so that these quantities satisfy a classical diffusion scaling $\frac{2\Delta x^2}{\Delta t} = \alpha > 0$, denoting by $K_{ij} := \alpha p_{ij}$ and letting the time step and grid size

go to 0, we formally obtain the following equation for the evolution of u_i on the continuous level:

$$\partial_t u_i = \sum_{0 \leq j \neq i \leq n} K_{ij} \left(u_j \Delta_x u_i - u_i \Delta_x u_j \right),$$

which is identical to the system of equations (4.9) introduced in the first section. Of course, this formal argument can be easily extended to any arbitrary dimension.

[1] Y. Achdou and I. Capuzzo-Dolcetta. Mean field games: numerical methods. *SIAM Journal on Numerical Analysis*, 48(3):1136–1162, 2010.

[2] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[3] N.D. Alikakos. L^p bounds of solutions of reaction-diffusion equations. *Communications in Partial Differential Equations*, 4(8):827–868, 1979.

[4] H. Amann. Parabolic evolution equations and nonlinear boundary conditions. *Journal of Differential Equations*, 72(2):201–269, 1988.

[5] H. Amann. Highly degenerate quasilinear parabolic systems. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 18(1):135–166, 1991.

[6] Herbert Amann et al. Dynamic theory of quasilinear parabolic equations. II. Reaction-diffusion systems. *Differential and Integral Equations*, 3(1):13–75, 1990.

[7] 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. *Journal of Non-Newtonian Fluid Mechanics*, 139(3):153–176, 2006.

[8] A. Anantharaman, R. Costaouec, C. Le Bris, F. Legoll, and F. Thomines. Introduction to numerical stochastic homogenization and the related computational challenges: some recent developments. In W. Bao and Q. Du, editors, *Multiscale modeling and analysis for materials simulation*, volume 22 of *Lect. Notes Series, Institute for Mathematical Sciences, National University of Singapore*, pages 197–272. World Sci. Publ., Hackensack, NJ, 2011.

[9] A. Bakhta. *Mathematical models and numerical simulation of photovoltaic devices*. PhD thesis, 2017.

[10] M. Balajewicz, D. Amsallem, and C. Farhat. Projection-based model reduction for contact problems. *International Journal for Numerical Methods in Engineering*, 106(8):644–663, 2016.

[11] R. Bartlett and M. Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291, 2007.

[12] A. Baskaran, J. Devita, and P. Smereka. Kinetic Monte Carlo simulation of strained heteroepitaxial growth with intermixing. *Continuum Mechanics and Thermodynamics*, 22(1):1–26, 2010.

[13] Rick Beatson and Leslie Greengard. A short course on fast multipole methods. *Wavelets, multilevel methods and elliptic PDEs*, 1:1–37, 1997.

[14] R. Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

[15] A. Benaceur. *Reduced order modeling in thermo-mechanics*. PhD thesis, 2018.

[16] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[17] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures*. American Mathematical Society, 1978.

[18] R. Bergamaschini, M. Salvalaglio, R. Backofen, A. Voigt, and F. Montalenti. Continuum modelling of semiconductor heteroepitaxy: an applied perspective. *Advances in Physics: X*, 1(3):331–367, 2016.

[19] G. Beylkin and M. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM Journal on Scientific Computing*, 26(6):2133–2159, 2005.

[20] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM journal on mathematical analysis*, 43(3):1457–1472, 2011.

[21] D. Bothe. On the maxwell-stefan approach to multicomponent diffusion. In *Parabolic problems*, pages 81–93. Springer, 2011.

[22] A. Bourgeat and A. Piatniski. An optimal error estimate in stochastic homogenization of discrete elliptic equations. *Ann. I. H. Poincaré*, 40:153–165, 2004.

[23] S. Boyaval and T. Lelièvre. A variance reduction method for parametrized stochastic differential equations using the reduced basis paradigm. *Communications in Mathematical Sciences*, 8(3):735–762, 2010.

[24] J. Brackbill. On energy and momentum conservation in particle-in-cell plasma simulation. *Journal of Computational Physics*, 317:405–427, 2016.

[25] A. Buffa, Y. Maday, A. Patera, C. Prud'homme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical modelling and numerical analysis*, 46(3):595–603, 2012.

[26] M. Burger, M. Di Francesco, and B. Pietschmann, J.-F.and Schlake. Nonlinear cross-diffusion with size exclusion. *SIAM Journal on Mathematical Analysis*, 42(6):2842–2871, 2010.

[27] N. Cagniart, Y. Maday, and B. Stamm. Model order reduction for problems with large convection effects. In *Contributions to Partial Differential Equations and Applications*, pages 131–150. Springer, 2019.

[28] M. Campos Pinto and M. Mehrenberger. Convergence of an adaptive semi-Lagrangian scheme for the Vlasov-Poisson system. *Numerische Mathematik*, 108(3):407–444, 2008.

[29] C. Cancès and B. Gaudeul. Entropy diminishing finite volume approximation of a cross-diffusion system. 2019.

[30] E. Cancès, C. Le Bris, and Y. Maday. Méthodes mathématiques en chimie quantique. Une introduction, Mathématiques & Applications (Berlin)[Mathematics & Applications], vol. 53, 2006.

[31] E. Cancès, Y. Maday, and B. Stamm. Domain decomposition for implicit solvation models. *Journal of Chemical Physics*, 139:054111, 2013.

[32] Y. Cao and J. Lu. Stochastic dynamical low-rank approximation method. *Journal of Computational Physics*, 372:564–586, 2018.

[33] F. Charles, B. Després, and M. Mehrenberger. Enhanced convergence estimates for semi-Lagrangian schemes: application to the Vlasov-Poisson equation. *SIAM Journal on Numerical Analysis*, 51(2):840–863, 2013.

[34] L. Chen and A. Jüngel. Analysis of a multidimensional parabolic population model with strong cross-diffusion. *SIAM journal on mathematical analysis*, 36(1):301–322, 2004.

[35] L. Chen and A. Jüngel. Analysis of a parabolic cross-diffusion population model without self-diffusion. *Journal of Differential Equations*, 224(1):39–59, 2006.

[36] H. Cho, D. Venturi, and G. Karniadakis. Numerical methods for high-dimensional probability density function equations. *Journal of Computational Physics*, 305:817–837, 2016.

[37] R.M. Christensen and K.H. Lo. Solutions for effective shear properties in three sphere and cylinder models. *J. Mech. Phys. Solids*, 27:315–330, 1979.

[38] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, and D. Mandic. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends in Machine Learning*, 9(4-5):249–429, 2016.

[39] D. Cioranescu and P. Donato. *An Introduction to Homogenization*. Oxford University Press, New York, 1999.

[40] C. Cotar, G. Friesecke, and C. Klüppelberg. Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the Hohenberg-Kohn functional. *Archive for Rational Mechanics and Analysis*, 228(3):891–922, 2018.

[41] Co. Cotar, G. Friesecke, and C. Klüppelberg. Density functional theory and optimal transportation with Coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013.

[42] N. Crouseilles, G. Latu, and E. Sonnendrücker. A parallel Vlasov solver based on local cubic spline interpolation on patches. *Journal of Computational Physics*, 228(5):1429–1446, 2009.

[43] W. Dahmen, R. Devore, L. Grasedyck, and E. Süli. Tensor-sparsity of solutions to high-dimensional elliptic partial differential equations. *Foundations of Computational Mathematics*, 16(4):813–874, 2016.

[44] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 6 Evolution Problems II*. Springer Science & Business Media, 2012.

[45] C. Daversin and C. Prud'Homme. Simultaneous empirical interpolation and reduced basis method for non-linear problems. *Comptes Rendus Mathematique*, 353(12):1105–1109, 2015.

[46] V. De Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, 2008.

[47] P. Degond, L. Pareschi, and G. Russo. *Modeling and computational methods for kinetic equations*. Springer Science & Business Media, 2004.

[48] L. Desvillettes, T. Lepoutre, A. Moussa, and A. Trescases. On the entropic structure of reaction-cross diffusion systems. *Communications in Partial Differential Equations*, 40(9):1705–1747, 2015.

[49] R. DeVore, R. Howard, and C. Micchelli. Optimal nonlinear approximation. *Manuscripta mathematica*, 63(4):469–478, 1989.

[50] R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy algorithms for reduced bases in Banach spaces. *Constructive Approximation*, 37(3):455–466, 2013.

[51] R.A. DeVore. The theoretical foundation of reduced basis methods. *Model Reduction and approximation: Theory and Algorithms*, pages 137–168, 2014.

[52] M. Di Francesco and J. Rosado. Fully parabolic Keller-Segel model for chemotaxis with prevention of overcrowding. *Nonlinearity*, 21(11):2715, 2008.

[53] L. Dung. Remarks on hölder continuity for parabolic equations and convergence to global attractors. *Nonlinear Analysis: Theory, Methods & Applications*, 41(7-8):921–941, 2000.

[54] Y. Efendiev and T. Hou. *Multiscale finite element methods: theory and applications*, volume 4. Springer Science & Business Media, 2009.

[55] L. Einkemmer and C. Lubich. A low-rank projector-splitting integrator for the Vlasov-Poisson equation. *SIAM Journal on Scientific Computing*, 40(5):B1330–B1360, 2018.

[56] L. Einkemmer and C. Lubich. A quasi-conservative dynamical low-rank algorithm for the Vlasov equation. *SIAM Journal on Scientific Computing*, 41(5):B1061–B1081, 2019.

[57] B. Engquist and P.E. Souganidis. Asymptotic and numerical homogenization. *Acta Numerica*, 17:147–190, 2008.

[58] A. Ern, I. Smears, and M. Vohralík. Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. *SIAM Journal on Numerical Analysis*, 55(6):2811–2834, 2017.

[59] A. Falcó and A. Nouy. Proper generalized decomposition for nonlinear convex problems in tensor Banach spaces. *Numerische Mathematik*, 121(3):503–530, 2012.

[60] M. Fathi and M. Simon. The gradient flow approach to hydrodynamic limits for the simple exclusion process. In *From Particle Systems to Partial Differential Equations III*, pages 167–184. Springer, 2016.

[61] J. Fauque, I. Ramière, and D. Ryckelynck. Hybrid hyper-reduced modeling for contact mechanics problems. *International Journal for Numerical Methods in Engineering*, 115(1):117–139, 2018.

[62] F. Filbet and E. Sonnendrücker. Comparison of Eulerian Vlasov solvers. *Comput. Phys. Commun.*, 150(IRMA-2001-035):247–266, 2001.

[63] G. Friesecke and D. Vögler. Breaking the curse of dimension in multi-marginal Kantorovich optimal transport on finite state spaces.

[64] H. Gajewski. On a variant of monotonicity and its application to differential equations. *Nonlinear Analysis: Theory, Methods & Applications*, 22(1):73–80, 1994.

[65] H. Gajewski. On the uniqueness of solutions to the drift-diffusion model of semiconductor devices. *Mathematical Models and Methods in Applied Sciences*, 4(01):121–133, 1994.

[66] K. Germaschewski, W. Fox, S. Abbott, N. Ahmadi, K. Maynard, L. Wang, H. Ruhl, and A. Bhattacharjee. The Plasma Simulation Code: A modern particle-in-cell code with patch-based load-balancing. *Journal of Computational Physics*, 318:305–326, 2016.

[67] M. Giaquinta and M. Struwe. On the partial regularity of weak solutions of nonlinear parabolic systems. *Mathematische Zeitschrift*, 179(4):437–451, 1982.

[68] V. Giovangigli. Multicomponent flow modeling. *Science China Mathematics*, 55(2):285–308, 2012.

[69] S. Glas and K. Urban. On noncoercive variational inequalities. *SIAM Journal on Numerical Analysis*, 52(5):2250–2271, 2014.

[70] M. Goto, H.and Kojo, A. Sasaki, and K. Hirose. Essentially exact ground-state calculations by superpositions of nonorthogonal Slater determinants. *Nanoscale research letters*, 8(1):1–7, 2013.

[71] L. Grasedyck, D. Kressner, and C. Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013.

[72] M. Grepl. Certified reduced basis methods for nonaffine linear time-varying and nonlinear parabolic partial differential equations. *Mathematical Models and Methods in Applied Sciences*, 22(03):1150015, 2012.

[73] M. Grepl, Y. Maday, N. Nguyen, and A. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(3):575–605, 2007.

[74] J. Griepentrog and L. Recke. Local existence, uniqueness and smooth dependence for nonsmooth quasilinear parabolic problems. *Journal of Evolution Equations*, 10(2):341–375, 2010.

[75] B. Haasdonk, J. Salomon, and B. Wohlmuth. A reduced basis method for parametrized variational inequalities. *SIAM Journal on Numerical Analysis*, 50(5):2656–2676, 2012.

[76] W. Hackbusch. *Tensor spaces and numerical tensor calculus*, volume 42. Springer, 2012.

[77] W. Hackbusch and B. Khoromskij. Tensor-product approximation to operators and functions in high dimensions. *Journal of Complexity*, 23(4-6):697–714, 2007.

[78] W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *Journal of Fourier analysis and applications*, 15(5):706–722, 2009.

[79] M. Herberg, M. Meyries, J. Prüss, and M. Wilke. Reaction-diffusion systems of maxwell-stefan type with reversible mass-action kinetics. *Nonlinear Analysis*, 159:264–284, 2017.

[80] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[81] S. Holtz, T. Rohwedder, and R. Schneider. On manifolds of tensors of fixed TT-rank. *Numerische Mathematik*, 120(4):701–731, 2012.

[82] A. Jansen. *An introduction to kinetic Monte Carlo simulations of surface reactions*, volume 856. Springer, 2012.

[83] V. Jikov, S. Kozlov, and O. Oleinik. *Homogenization of differential operators and integral functionals*. Springer, Berlin, 1995.

[84] O. John and J. Stará. On the regularity of weak solutions to parabolic systems in two spatial dimensions. *Communications in partial differential equations*, 23(7-8):437–451, 1998.

[85] A. Juengel and I. Stelzer. Entropy structure of a cross-diffusion tumor-growth model. *Mathematical Models and Methods in Applied Sciences*, 22(07):1250009, 2012.

[86] A. Jüngel. *Transport equations for semiconductors*, volume 773. Springer, 2009.

[87] A. Jüngel. The boundedness-by-entropy method for cross-diffusion systems. *Nonlinearity*, 28(6):1963, 2015.

[88] A. Jungel and I. Stelzer. Existence analysis of Maxwell-Stefan systems for multicomponent mixtures. *SIAM Journal on Mathematical Analysis*, 45(4):2421–2440, 2013.

[89] B. Khoromskij. Structured data-sparse approximation to high order tensors arising from the deterministic Boltzmann equation. *Mathematics of computation*, 76(259):1291–1315, 2007.

[90] H. Koch and E. Dalgaard. Linear superposition of optimized non-orthogonal Slater determinants for singlet states. *Chemical physics letters*, 212(1-2):193–200, 1993.

[91] O. Koch and C. Lubich. Dynamical low-rank approximation. *SIAM Journal on Matrix Analysis and Applications*, 29(2):434–454, 2007.

[92] T. Kolda and B. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[93] K. Kormann. A semi-Lagrangian Vlasov solver in tensor train format. *SIAM Journal on Scientific Computing*, 37(4):B613–B632, 2015.

[94] S.M. Kozlov. Averaging of random operators. *Matematicheskii Sbornik*, 151(2):188–202, 1979.

[95] K.H.W. Küfner. Invariant regions for quasilinear reaction-diffusion systems and applications to a two population model. *Nonlinear Differential Equations and Applications NoDEA*, 3(4):421–444, 1996.

[96] P. Ladevèze. *Nonlinear computational structural mechanics: new approaches and non-incremental methods of calculation*. Springer Science & Business Media, 2012.

[97] O. Ladyzenskaja and V. Solonnikov. Linear and quasilinear equations of parabolic type, Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23. *American Mathematical Society, Providence, RI*, 63:64, 1967.

[98] C. Landim, S. Olla, and S.R.S. Varadhan. Symmetric simple exclusion process: Regularity of the self-diffusion coefficient. *Communications in Mathematical Physics*, 224(1):307–321, 2001.

[99] D. Le and T. Nguyen. Everywhere regularity of solutions to a class of strongly coupled degenerate parabolic systems. *Communications in Partial Differential Equations*, 31(2):307–324, 2006.

[100] D. Le and T. T. Nguyen. Everywhere regularity of solutions to a class of strongly coupled degenerate parabolic systems. *Communications in Partial Differential Equations*, 31(2):307–324, 2006.

[101] C. Le Bris, T. Lelièvre, and Y. Maday. Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations. *Constructive Approximation*, 30(3):621, 2009.

[102] K. Lee and K.T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, 2020.

[103] B. Leimkuhler and C. Matthews. *Molecular Dynamics*. Springer, 2016.

[104] T. Lelièvre and G. Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[105] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[106] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[107] T. Lepoutre, M. Pierre, and G. Rolland. Global well-posedness of a conservative relaxed cross diffusion system. *SIAM Journal on Mathematical Analysis*, 44(3):1674–1693, 2012.

[108] T. Lepoutre, M. Pierre, and G. Rolland. Global well-posedness of a conservative relaxed cross diffusion system. *SIAM Journal on Mathematical Analysis*, 44(3):1674–1693, 2012.

[109] M. Levy. An energy-density equation for isoelectronic changes in atoms. *The Journal of Chemical Physics*, 68(11):5298–5299, 1978.

[110] M. Lewin. Semi-classical limit of the Levy-Lieb functional in density functional theory. *Comptes Rendus Mathematique*, 356(4):449–455, 2018.

[111] E. Lieb. Density functionals for Coulomb systems. In *Inequalities*, pages 269–303. Springer, 2002.

[112] M. Liero and A. Mielke. Gradient structures and geodesic convexity for reaction-diffusion systems. *Phil. Trans. R. Soc. A*, 371(2005):20120346, 2013.

[113] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications*, volume 1. Springer Science & Business Media, 2012.

[114] F. Lipparini, B. Stamm, E. Cancès, Y. Maday, and B. Mennucci. A fast domain decomposition algorithm for continuum solvation models: Energy and first derivatives. *J. Chem. Theory Comput.*, 9:3637–3648, 2013.

[115] S. Lojasiewicz. Ensembles semi-analytiques. *IHES notes*, 1965.

[116] C. Lu, X. Li, D. Wu, L. Zheng, and W. Yang. Predictive sampling of rare conformational events in aqueous solution: designing a generalized orthogonal space tempering method. *Journal of chemical theory and computation*, 12(1):41–52, 2016.

[117] C. Lubich and I. Oseledets. A projector-splitting integrator for dynamical low-rank approximation. *BIT Numerical Mathematics*, 54(1):171–188, 2014.

[118] M. Luskin and C. Ortner. Atomistic-to-continuum coupling. *Acta Numerica*, 22:397–508, 2013.

[119] C. Lusso, A. Ern, F. Bouchut, A. Mangeney, M. Farin, and O. Roche. Two-dimensional simulation by regularization of free surface viscoplastic flows with Drücker-Prager yield stress and application to granular collapse. *Journal of Computational Physics*, 333:387–408, 2017.

[120] Y. Maday, O. Mula, and G. Turinici. A priori convergence of the Generalized Empirical Interpolation Method. 2013.

[121] Y. Maday, N.C. Nguyen, A. Patera, and S.H. Pau. A general multipurpose interpolation procedure: the magic points. *Communications on Pure & Applied Analysis*, 8(1):383, 2009.

[122] D. Mattox. *Handbook of physical vapor deposition (PVD) processing*. William Andrew, 2010.

[123] O. Mula. *Some contributions towards the parallel simulation of time dependent neutron transport and the integration of observed data in real time*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2014.

[124] F. Murat and L. Tartar. H-convergence. *Séminaire d'Analyse Fonctionnelle et Numérique de l'Université d'Alger*, 1978.

[125] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi. Machine learning for molecular simulation. *arXiv preprint arXiv:1911.02792*, 2019.

[126] A. Nouy. Recent developments in spectral stochastic methods for the numerical solution of stochastic partial differential equations. *Archives of Computational Methods in Engineering*, 16(3):251–285, 2009.

[127] M. Ohlberger and S. Rave. Reduced basis methods: Success, limitations and future challenges. *arXiv preprint arXiv:1511.02021*, 2015.

[128] C. Ortner and L. Zhang. Energy-based atomistic-to-continuum coupling without ghost forces. *Computer Methods in Applied Mechanics and Engineering*, 279:29–45, 2014.

[129] I. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[130] G.C. Papanicolaou and S.R.S. Varadhan. Boundary value problems with rapidly oscillating random coefficients. In J. Fritz, J.L. Lebaritz, and D. Szasz, editors, *Proc. Colloq. on Random fields: Rigorous results in statistical mechanics and quantum field theory*, volume 10 of *Colloq. Math. Soc. János Bolyai*, pages 835–873. North-Holland, Amsterdam-New York, 1981.

[131] M. Pierre. Global existence in reaction-diffusion systems with control of mass: a survey. *Milan Journal of Mathematics*, 78(2):417–455, 2010.

[132] M. Pierre and D. Schmitt. Blowup in reaction-diffusion systems with dissipation of mass. *SIAM review*, 42(1):93–106, 2000.

[133] J. Portegies and M. Peletier. Well-posedness of a parabolic moving-boundary problem in the setting of Wasserstein gradient flows. *arXiv preprint arXiv:0812.1269*, 2008.

[134] J. Quastel. Diffusion of color in the simple exclusion process. *Communications on Pure and Applied Mathematics*, 45(6):623–679, 1992.

[135] R. Redlinger. Invariant sets for strongly coupled reaction-diffusion systems under general boundary conditions. *Archive for Rational Mechanics and Analysis*, 108(4):281–291, 1989.

[136] M. Reed and B. Simon. *Methods of Modern Mathematical Physics IV: Analysis of Operators*. Academic Press, 1978.

[137] R. Schneider and A. Uschmajew. Approximation rates for the hierarchical tensor format in periodic Sobolev spaces. *Journal of Complexity*, 30(2):56–71, 2014.

[138] T.P. Schulze and P. Smereka. Kinetic Monte Carlo simulation of heteroepitaxial growth: Wetting layers, quantum dots, capping, and nanorings. *Physical Review B*, 96(23):235313.

[139] M. Seidl. Strong-interaction limit of density-functional theory. *Physical Review A*, 60(6):4387, 1999.

[140] M. Seidl, P. Gori-Giorgi, and A. Savin. Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities. *Physical Review A*, 75(4):042511, 2007.

[141] M. Sharify, S. Gaubert, and L. Grigori. Solution of the optimal assignment problem by diagonal scaling algorithms. *arXiv preprint arXiv:1104.3830*, 2011.

[142] J. Stará and O. John. Some (new) counterexamples of parabolic systems. *Commentationes Mathematicae Universitatis Carolinae*, 36(3):503–510, 1995.

[143] S. Szalay, M. Pfeffer, V. Murg, G. Barcza, F. Verstraete, R. Schneider, and O. Legeza. Tensor product methods and entanglement optimization for ab initio quantum chemistry. *International Journal of Quantum Chemistry*, 115(19):1342–1391, 2015.

[144] F. Tantardini and A. Veeser. The L^2-projection and quasi-optimality of galerkin methods for parabolic equations. *SIAM Journal on Numerical Analysis*, 54(1):317–340, 2016.

[145] V. Temlyakov. Greedy approximation. *Acta Numerica*, 17:235–409, 2008.

[146] S. Ten-no. Superposition of nonorthogonal Slater determinants towards electron correlation problems. *Theoretical Chemistry Accounts*, 98(4):182–191, 1998.

[147] A. Unterreiter, A. Arnold, P. Markowich, and G. Toscani. On Generalized Csiszár-Kullback Inequalities. *Monatshefte für Mathematik*, 131(3):235–253, 2000.

[148] K. Urban and A. Patera. A new error bound for reduced basis approximation of parabolic partial differential equations. *Comptes Rendus Mathematique*, 350(3-4):203–207, 2012.

[149] A. Voigt. *Multiscale Modeling in Epitaxial Growth*, volume 149. Springer Science & Business Media, 2006.

[150] X. Wang, R. Samulyak, X. Jiao, and K. Yu. AP-Cloud: Adaptive Particle-in-Cloud method for optimal solutions to Vlasov-Poisson equation. *Journal of Computational Physics*, 316:682–699, 2016.

[151] G. Welper. Interpolation of functions with parameter dependent jumps by transformed snapshots. *SIAM Journal on Scientific Computing*, 39(4):A1225–A1250, 2017.

[152] M. Wiegner. Global solutions to a class of strongly coupled parabolic systems. *Mathematische Annalen*, 292(1):711–727, 1992.

[153] S.M. Wise, J.S. Lowengrub, J.S. Kim, and W.C. Johnson. Efficient phase-field simulation of quantum dot formation in a strained heteroepitaxial film. *Superlattices and Microstructures*, 36(1-3):293–304, 2004.

[154] J. Wloka. *Partial Differential Equations*. Cambridge University Press, New York, 1982.

[155] N. Zamponi and A. Jüngel. Analysis of degenerate cross-diffusion population models with volume filling. In *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*. Elsevier, 2015.

[156] N. Zamponi and A. Jüngel. Analysis of degenerate cross-diffusion population models with volume filling. In *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*. Elsevier, 2015.

[157] Z. Zhang, E. Bader, and K. Veroy. A slack approach to reduced-basis approximation and error estimation for variational inequalities. *Comptes Rendus Mathematique*, 354(3):283–289, 2016.

[158] G. M. Zhislin. Finiteness of the discrete spectrum in the quantum N-particle problem. *Teoreticheskaya i Matematicheskaya Fizika*, 21(1):60–73, 1974.