Theoretical and numerical study of nonlinear models in quantum mechanics

Antoine Levitt

▶ **To cite this version:**

Antoine Levitt. Theoretical and numerical study of nonlinear models in quantum mechanics. Other [q-bio.OT]. Université Paris Dauphine - Paris IX, 2013. English. NNT : 2013PA090011 . tel-00881031

HAL Id: tel-00881031

https://theses.hal.science/tel-00881031

Submitted on 7 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DAUPHINE UNIVERSITÉ PARIS | CEREMADE UMR CNRS 7534

Université Paris-Dauphine, École doctorale de Dauphine

Thèse de doctorat en sciences
Spécialité Mathématiques appliquées

Présentée par
Antoine Levitt

Étude théorique et numérique de modèles non linéaires en mécanique quantique

soutenue le 4 juillet 2013 devant le jury composé de

Rapporteurs :
Éric Cancès
Professeur, École des Ponts ParisTech
Laurent Di Menza
Professeur, Université de Reims

Examinateurs :
Xavier Blanc
Professeur, Université Paris Diderot
Guillaume Legendre
Maître de conférences, Université Paris-Dauphine
Mathieu Lewin
Chargé de recherche, Université de Cergy-Pontoise
Heinz Siedentop
Professeur, Ludwig-Maximilians-Universität, Münich

Directeur de thèse :
Éric Séré
Professeur, Université Paris-Dauphine

Remerciements

Je souhaite tout d'abord remercier Éric Séré pour sa disponibilité, ses conseils et ses encouragements tout au long de ces trois ans. Par sa patience et sa rigueur, il a su me permettre d'avancer quand je croyais la situation bloquée.

Je remercie Éric Cancès et Laurent Di Menza d'avoir accepté d'être rapporteurs de cette thèse, ainsi que de leur intérêt pour mes travaux.

Merci à Xavier Blanc pour sa participation à ce jury. Thanks to Heinz Siedentop for agreeing to be part of the jury, despite the fact that part of this thesis is in French.

Guillaume Legendre m'a permis de participer au développement du code de calcul ACCQUAREL, ce qui m'a été extrêmement utile notamment en début de thèse pour me familiariser avec la mécanique quantique. Merci à lui pour ses explications et son soutien au cours de cette thèse.

Je remercie Mathieu Lewin de son intérêt pour mes travaux, et de m'avoir suggéré l'étude des inégalités de Lieb-Thirring. Merci à lui ainsi qu'à Maria Esteban pour m'avoir permis de participer au projet ANR NoNAP et au trimestre thématique de l'Institut Henri Poincaré "Méthodes variationnelles et spectrales en mécanique quantique".

Je suis reconnaissant envers Julien Salomon, qui m'a montré comment utiliser l'inégalité de Łojasiewicz, et Jean Dolbeault, qui m'a permis de mieux comprendre les inégalités de Lieb-Thirring.

Je souhaite remercier les thésards du Ceremade pour les nombreuses discussions, mathématiques ou non, que j'ai pu avoir avec eux. Merci également aux thésards et postdocs du groupe de travail en mécanique quantique, avec qui j'ai pu élargir ma culture en physique mathématique. Merci en particulier à Loïc Le Treust et Julien Sabin, qui m'ont offert leurs remarques sur ce manuscrit.

Merci enfin à ma famille et à mes amis pour leur soutien pendant ces trois ans.

Table des matières

Chapitre 1

Introduction

Cette thèse s'inscrit dans le contexte scientifique de la chimie quantique *ab initio*, c'est-à-dire de l'étude de la matière en n'utilisant que les principes fondamentaux de la mécanique quantique et non des modèles établis empiriquement. Les avantages de cette approche sont sa précision, sa simplicité conceptuelle et sa flexibilité. En contrepartie, on obtient des modèles complexes qui nécessitent des solutions numériques souvent coûteuses en temps de calcul. Le succès en chimie de cette approche intermédiaire entre prédictions analytiques et expériences est attesté par le prix Nobel décerné à Kohn et Pople en 1998. Aujourd'hui, les méthodes *ab initio* sont utilisées pour prédire le comportement des réactions chimiques ou les propriétés des matériaux.

Bien que la mécanique quantique soit une théorie fondamentalement linéaire, la complexité d'un système quantique croît exponentiellement avec sa taille : on a besoin d'une fonction d'onde à $3N$ variables pour représenter l'état d'un système de N particules. Les physiciens sont donc amenés à introduire des approximations non-linéaires pour réduire la taille des problèmes et permettre des calculs numériques. Malgré leur importance pratique, les aspects mathématiques de ces modèles sont mal connus, et de nombreux problèmes théoriques comme numériques sont encore ouverts.

Cette thèse se veut une contribution à l'étude de ces questions importantes. Elle est divisée en deux parties indépendantes. Dans la première partie, on étudie le modèle de Hartree-Fock, sa généralisation au cas multiconfiguration, et la prise en compte des effets relativistes. Pour le modèle de Hartree-Fock, on prouve la convergence d'algorithmes de calcul de solutions (chapitre 2). Pour le modèle multiconfiguration relativiste, on prouve l'existence de solutions près de la limite non-relativiste (chapitre 3). Dans la seconde partie, on s'intéresse aux inégalités de Lieb-Thirring, qui lient les valeurs propres négatives de l'opérateur $-\Delta + V$ à la taille du potentiel V. Par des calculs d'éléments finis, on trouve des exemples de potentiels V qui fournissent des bornes inférieures sur les constantes optimales intervenant dans ces inégalités (chapitre 4).

Dans un premier temps, on présente rapidement les notions de mécanique quantique nécessaires à la compréhension de la thèse. On introduit ensuite les modèles étudiés, et les résultats obtenus. Enfin, les chapitres 2, 3 et 4, tirés d'articles publiés ou soumis, sont dédiés à l'exposition précise des résultats.

1.1 Le formalisme quantique

En mécanique classique, on modélise l'état d'une particule à un instant donné par un couple de vecteurs position et quantité de mouvement $(x, p) \in \mathbb{R}^3 \times \mathbb{R}^3$. Étant donné un champ de potentiel extérieur $V(x)$, l'évolution du système est donnée par les équations

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{p}{m}, \\ \frac{\mathrm{d}p}{\mathrm{d}t} &= -\nabla V(x). \end{aligned}$$

La première de ces équations est simplement la définition de la quantité de mouvement, $p = m\frac{\mathrm{d}x}{\mathrm{d}t}$. La seconde exprime la seconde loi de Newton, $ma = F$, où la force F découle du potentiel V par la relation $F = -\nabla V$. Lors de l'évolution du système, la quantité

$$H(x,p) = \frac{1}{2m}p^2 + V(x) \tag{1.1}$$

est conservée. Cette quantité, l'Hamiltonien, s'interprète comme l'énergie totale du système. L'Hamiltonien suffit à spécifier l'évolution, qui se réécrit sous la forme des équations de Hamilton

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= +\frac{\partial H}{\partial p},\\ \frac{\mathrm{d}p}{\mathrm{d}t} &= -\frac{\partial H}{\partial x}.\end{aligned}$$

On obtient un système d'équations différentielles non-linéaires du premier ordre. La donnée de la condition initiale (x_0, p_0) au temps $t = 0$ et du potentiel $V(x)$ suffit à déterminer la trajectoire du système, via le théorème de Cauchy-Lipschitz (sous des hypothèses de régularité sur V).

La mécanique classique s'est révélée une théorie extrêmement féconde, qui permet de décrire la matière sur plus de quinze ordres de grandeur, de la bactérie à la planète. Elle souffre néanmoins de deux limitations majeures. Premièrement, quand les vitesses approchent la vitesse de la lumière (300 000 kilomètres par seconde), on observe des contractions des longueurs et du temps qui imposent le recours à la théorie relativiste. Deuxièmement, quand les échelles impliquées deviennent trop petites (de l'ordre du nanomètre), on ne peut plus considérer la matière comme particulaire, et on doit adopter une description ondulatoire : c'est la mécanique quantique, dont nous allons brièvement exposer les principes.

Cette introduction n'est pas un cours de mécanique quantique, mais une présentation du formalisme utilisé pour la suite. On pourra se reporter à l'annexe de [CLBM06] pour une introduction rapide du point de vue mathématique, et à [Fey65; CTDL73] entre autres références pour une approche plus physique.

Alors que la mécanique classique décrit une particule par un vecteur de $\mathbb{R}^6$ et son évolution par une équation différentielle ordinaire non-linéaire, la mécanique quantique décrit une particule par sa fonction d'onde $\psi \in L^2(\mathbb{R}^3)$, et son évolution par une équation aux dérivées partielles linéaire, l'équation de Schrödinger. Au prix d'une complexité mathématique accrue, la théorie quantique fournit des prédictions extrêmement précises, et permet en particulier de rendre compte de la structure des atomes et des molécules.

Les postulats de la mécanique quantique dictent que l'état d'un système est représenté par un élément ψ de la sphère unité d'un espace de Hilbert complexe $\mathcal{H}$. Une quantité observable scalaire de ce système se modélise par un opérateur auto-adjoint A sur $\mathcal{H}$. La mesure de cet observable est un processus probabiliste, dont la loi est donnée par la mesure spectrale de A appliquée à l'état ψ. En particulier, $\langle \psi, A\psi \rangle$ est la valeur moyenne de l'observable A dans l'état ψ.

En mécanique classique, la fonctionnelle d'énergie d'un système dicte son évolution via les équations de Hamilton. En mécanique quantique, c'est l'observable

d'énergie H, aussi appelée Hamiltonien, qui joue ce rôle, via l'équation de Schrödinger[1]

$$i\partial_t \psi = H\psi. \tag{1.2}$$

Considérons l'exemple d'une particule sans spin. Le formalisme quantique non-relativiste représente son état par une fonction d'onde $\psi \in L^2(\mathbb{R}^3, \mathbb{C})$, avec $\|\psi\| = 1$. La composante x_i de la position de la particule est représentée par l'opérateur de multiplication par x_i. Il résulte alors des propriétés spectrales de cet opérateur de multiplication que $|\psi(x)|^2$ représente la densité de probabilité de présence de la particule au point x, ce qui permet une interprétation plus évidente de la fonction d'onde.

Pour construire l'Hamiltonien d'un système et obtenir sa dynamique, on se base sur l'Hamiltonien classique, et on utilise les règles de correspondance pour la position et la quantité de mouvement

$$V(x) \leftrightarrow \hat{V}, \tag{1.3}$$
$$p \leftrightarrow -i\nabla, \tag{1.4}$$

où on note $\hat{V}$ l'opérateur de multiplication $(\hat{V}\psi)(x) = V(x)\psi(x)$. Par la suite, on notera cet opérateur simplement V. Comme $\nabla(x\psi) = x\nabla\psi + \psi$, on a $[x, p] = -i$: l'opérateur de position x ne commute pas avec l'opérateur de quantité de mouvement p. Ces deux opérateurs ne peuvent donc pas être diagonalisés simultanément, ce qui implique qu'un état ψ ne peut pas avoir une position et une quantité de mouvement simultanément bien définies (c'est le principe d'incertitude de Heisenberg).

Un exemple fondamental de système quantique est celui de l'atome d'hydrogène, composé d'un électron soumis à l'attraction Coulombienne d'un proton fixe. Son énergie classique est

$$H(x, p) = \frac{1}{2}p^2 - \frac{1}{|x|}. \tag{1.5}$$

Par application des règles de correspondance, l'Hamiltonien H associé s'écrit

$$(H\psi)(x) = -\frac{1}{2}(\Delta\psi)(x) - \frac{1}{|x|}\psi(x). \tag{1.6}$$

Son spectre est composé de valeurs propres négatives $\lambda_n = -\frac{1}{2n^2}$, et d'un spectre essentiel $[0, +\infty[$. Les états propres associés aux valeurs propres négatives sont localisés, et sont appelés états liés. L'état propre associé à la première valeur propre λ_1 joue un rôle particulier : c'est l'état fondamental, celui où on trouvera l'électron au zéro absolu. Les états propres associés aux valeurs propres suivantes sont les états excités. Par un processus que nous ne décrivons pas ici, un électron peut passer de l'état fondamental à un état excité, en absorbant un photon de fréquence $\lambda_n - \lambda_1$. Le caractère discret des fréquences possibles explique le spectre de raies des éléments chimiques.

[1] Pour simplifier les notations, on utilise dans toute cette thèse le système d'unités atomiques, dans lequel les constantes fondamentales qui interviennent (la masse et charge de l'électron, les constantes de Planck et de Coulomb) sont égales à 1.

1.2 Le problème à N corps en mécanique quantique

On s'intéresse maintenant à l'équation de Schrödinger vérifiée par le système quantique d'une molécule ou d'un atome formé de N électrons autour de M noyaux, le noyau i comportant Z_i protons. L'espace de Hilbert du système composé est obtenu comme produit tensoriel des espaces de Hilbert individuels. Cela revient à considérer une fonction d'onde $\psi \in L^2(\mathbb{R}^{3(N+M)})$. Dans ce formalisme, la quantité $|\psi(x_1, \dots, x_N, z_1, \dots, z_M)|^2$ représente la probabilité de trouver les électrons aux coordonnées $x_1, \dots, x_N$ et les noyaux aux coordonnées $z_1, \dots, z_M$. Les noyaux étant plus lourds que les électrons de trois ordres de grandeur, leurs fonctions d'onde sont bien plus localisées que celles des électrons. On peut donc négliger leur caractère quantique et considérer qu'ils sont vus par les électrons comme des particules ponctuelles, placés en $z_i \in \mathbb{R}^3$. La fonction d'onde ne dépend alors plus que des variables électroniques : c'est l'approximation de Born-Oppenheimer. Une fois le calcul effectué pour une position particulière des noyaux, on peut faire varier la configuration pour optimiser l'énergie fondamentale, et ainsi obtenir la géométrie des molécules au repos. On s'intéresse dans la suite uniquement au problème électronique, pour une configuration des noyaux donnée.

L'état des N électrons est modélisé par une fonction d'onde ψ qui en toute généralité est un élément de l'espace $L^2(\mathbb{R}^{3N}, \mathbb{C}^2)$, où l'espace $\mathbb{C}^2$ représente les deux états de spin. Pour notre étude, les effets de spin ne sont pas fondamentaux et alourdissent simplement les notations : on travaille donc avec des fonctions d'onde $\psi \in L^2(\mathbb{R}^{3N}, \mathbb{R})$.

Une complication supplémentaire provient d'un autre postulat de la mécanique quantique : on ne peut pas distinguer deux particules de même nature. Ce postulat a des conséquences sur l'espace de Hilbert $\mathcal{H} \subset L^2(\mathbb{R}^{3N}, \mathbb{R})$ dans lequel vit ψ. Soit $P_{i,j}$ l'opérateur d'échange des électrons i et j :

$$(P_{i,j}\psi)(\dots, x_i, \dots, x_j, \dots) = \psi(\dots, x_j, \dots, x_i, \dots).$$

Alors les états ψ et $P_{i,j}\psi$ doivent être indistinguables, c'est-à-dire qu'ils doivent donner le même résultat pour la mesure de tout observable. On montre (voir par exemple [CLBM06]) que la seule possibilité est que $P_{i,j}\Big|_{\mathcal{H}}$ soit un opérateur de changement de phase, $P_{i,j}\Big|_{\mathcal{H}} = I$ ou $P_{i,j}\Big|_{\mathcal{H}} = -I$. Les deux types de situations sont observées pour différents types de particules. Dans le premier cas, les particules sont appelées bosons, dans le second, fermions.

Il a été montré expérimentalement que les électrons sont des fermions. Une fonction d'onde électronique doit donc être antisymétrique vis-à-vis de l'échange de deux particules :

$$\forall 1 \leq i < j \leq N, \;\; \psi(\dots, x_i, \dots, x_j, \dots) = -\psi(\dots, x_j, \dots, x_i, \dots). \tag{1.7}$$

On appelle $L^2_a(\mathbb{R}^{3N}, \mathbb{R})$ l'ensemble des fonctions de $L^2(\mathbb{R}^{3N}, \mathbb{R})$ vérifiant cette propriété, et on cherchera les fonctions d'onde électroniques dans cet espace. Cette condition signifie en particulier que $\psi(\dots, x, \dots, x, \dots) = 0$. C'est une forme du

principe d'exclusion de Pauli : deux électrons ne peuvent pas occuper la même position.

Pour une configuration donnée, les noyaux exercent un potentiel électrostatique

$$V(x) = -\sum_{i=1}^{M} \frac{Z_i}{|x - z_i|}. \tag{1.8}$$

L'Hamiltonien associé au système des N électrons, obtenu par application du principe de correspondance, est

$$H^{(N)} = \sum_{i=1}^{N} \left(-\frac{1}{2}\Delta_{x_i} + V(x_i) \right) + \sum_{1 \leq i < j \leq N} \frac{1}{\left| x_i - x_j \right|} + \sum_{1 \leq i < j \leq M} \frac{Z_i Z_j}{\left| z_i - z_j \right|}, \tag{1.9}$$

où la notation indicée Δ_{x_i} indique que le Laplacien agit sur la coordonnée x_i. L'Hamiltonien contient, dans l'ordre, les termes d'énergie cinétique des électrons, d'attraction Coulombienne électrons-noyaux, et de répulsion Coulombienne électrons-électrons et noyaux-noyaux. Pour une configuration des noyaux donnée, ce dernier terme est constant, et on l'ignorera dans la suite pour s'intéresser uniquement au problème électronique.

L'opérateur $H^{(N)}$ est auto-adjoint sur $L^2_a(\mathbb{R}^{3N}, \mathbb{R})$. Son spectre est constitué d'un spectre essentiel de la forme $[\Sigma, +\infty[$, et de valeurs propres. Les valeurs propres strictement inférieures à Σ sont isolées, de multiplicité finie, et peuvent être obtenues par les formules min-max classiques.

Le problème qui nous occupe est celui de la détermination de l'état fondamental, c'est-à-dire de la plus petite valeur propre λ_1 et de son ou ses états propres associés[2]. Cette valeur propre satisfait le principe variationnel

$$\lambda_1 = \inf_{\psi \in L^2_a(\mathbb{R}^{3N}, \mathbb{R}), \|\psi\| = 1} \langle \psi, H\psi \rangle_{L^2(\mathbb{R}^{3N})}. \tag{1.10}$$

La méthode la plus élémentaire pour le calcul numérique de λ_1 consiste à introduire un sous-espace $V_n \subset L^2_a(\mathbb{R}^{3N}, \mathbb{R})$ de dimension finie, et à résoudre le problème aux valeurs propres restreint à cet espace. On se heurte ici à un problème de dimensionalité : si on utilise une base construite sur une grille grossière de 10 points par dimension, on obtient un problème de valeurs propres en dimension 10^{3N}, ce qui surpasse les capacités de calcul des plus gros supercalculateurs même pour de faibles valeurs de N.

Deux solutions sont principalement utilisées pour remédier à ce problème. La première est de résumer l'information contenue dans ψ par la seule densité électronique ρ, une fonction de $\mathbb{R}^3$ dans $\mathbb{R}^+$. Il faut alors trouver une expression approchée de l'énergie de ρ : c'est la théorie de la fonctionnelle de la densité (DFT, pour *Density Functional Theory*), très utilisée en physique du solide mais considérée trop imprécise pour les utilisations en chimie quantique. Une deuxième possibilité est de restreindre l'espace de recherche dans (1.10) en postulant une forme particulière pour ψ. C'est dans cette catégorie que se placent les méthodes de Hartree-Fock.

[2]Ce problème n'est pas le seul intéressant en chimie quantique. Cependant, il est souvent une étape obligatoire pour une analyse plus poussée, comme par exemple l'optimisation de la géométrie des noyaux, ou le calcul de propriétés chimiques ou thermodynamiques.

1.3 Première partie : le modèle de Hartree-Fock

1.3.1 L'approximation de Hartree-Fock

Il est bien connu que le produit tensoriel de $L^2(\mathbb{R})$ par lui-même s'identifie naturellement à $L^2(\mathbb{R}^2)$. Plus généralement, toute fonction $\psi \in L^2(\mathbb{R}^{3N})$ peut être approchée par une combinaison linéaire de produits tensoriels :

$$\psi(x_1, \dots, x_N) \approx \sum_i \phi_i^1(x_1)\phi_i^2(x_2)\dots\phi_i^N(x_N),$$

où les ϕ_i^k sont des éléments de $L^2(\mathbb{R}^3)$.

Cette technique permet de séparer les variables x_i, et, en gardant un nombre fini de termes, de réduire la dimensionalité du problème. Cette approximation n'est pas utilisable telle quelle pour notre problème, car elle ne respecte pas la condition d'antisymétrie (1.7). On utilise donc une décomposition adaptée à $L^2_a(\mathbb{R}^{3N})$,

$$\psi = \sum_{i_1, i_2, \dots, i_N} \left|\phi_{i_1}, \phi_{i_2}, \dots, \phi_{i_N}\right\rangle, \tag{1.11}$$

où on décompose ψ sur les déterminants de Slater

$$\left|\phi_{i_1}, \phi_{i_2}, \dots, \phi_{i_N}\right\rangle = \frac{1}{\sqrt{N!}} \det\left((\phi_i(x_j))_{i,j}\right). \tag{1.12}$$

Ces déterminants de Slater peuvent être vus comme la généralisation naturelle des produits tensoriels élémentaires $\phi_i^1(x_1)\phi_i^2(x_2)\dots\phi_i^N(x_N)$ sous la contrainte d'antisymétrie (1.7). Par exemple, pour $N = 2$, on a

$$|\phi_1, \phi_2\rangle (x_1, x_2) = \frac{1}{\sqrt{2}} \left(\phi_1(x_1)\phi_2(x_2) - \phi_2(x_1)\phi_1(x_2)\right).$$

Cette décomposition est complète en norme L^2 : on peut approcher à une précision arbitraire toute fonction $\psi \in L^2_a(\mathbb{R}^{3N})$ par une combinaison linéaire finie de déterminants de Slater. En d'autres termes,

$$L^2_a(\mathbb{R}^{3N}) = \bigwedge_1^N L^2(\mathbb{R}^3),$$

où $\bigwedge$ est le produit tensoriel antisymétrique.

L'approximation de Hartree-Fock consiste à ne garder qu'un déterminant dans ce développement, c'est-à-dire à utiliser l'ansatz

$$\psi = |\phi_1, \phi_2, \dots, \phi_N\rangle, \tag{1.13}$$

où les $\phi_i \in L^2(\mathbb{R}^3)$ sont des fonctions d'ondes mono-électroniques, appelées *orbitales*. Cet ansatz revient essentiellement à considérer que les N électrons sont indépendants, et que, modulo les corrections imposées par l'antisymétrie de la fonction d'onde, on a $\psi(x_1, \dots, x_N) \approx \phi_1(x_1)\dots\phi_N(x_N)$. Cette indépendance implique une perte de corrélation entre les différents électrons.

L'approximation de Hartree-Fock, qui peut sembler grossière, est néanmoins étonnamment efficace. Mathématiquement, elle a l'avantage d'être variationnelle : l'énergie fondamentale obtenue par minimisation sur les ϕ_i est automatiquement supérieure à la première valeur propre de l'opérateur de Schrödinger à N corps. On obtient ainsi une borne supérieure de l'énergie, et non pas une approximation dont on ne contrôle pas le signe comme c'est le cas avec la théorie de la fonctionnelle de la densité par exemple.

En pratique, la méthode permet de prédire les énergies de repos et les géométries de la plupart des atomes et petites molécules à quelques pourcents près [SO89]. Cependant, ces petites erreurs peuvent souvent cacher des phénomènes physiques d'intérêt. Par exemple, l'enthalpie de formation du dioxyde de carbone (la différence entre l'énergie fondamentale du système et la somme des énergies fondamentales de ses constituants) est de l'ordre de 0.1% de son énergie totale [Can98]. Un calcul naïf peut donc conduire à des erreurs importantes dans l'estimation de cette enthalpie de formation (bien qu'en pratique, les erreurs se compensent souvent). Pour des résultats plus précis, on utilise des techniques dites post-Hartre-Fock, comme l'interaction de configuration (CI pour *Configuration Interaction*), la théorie de la perturbation de Møller-Plesset, ou les méthodes multiconfiguration.

On s'intéresse maintenant à la formulation mathématique et numérique du problème de la détermination de l'état fondamental dans la théorie Hartree-Fock. En utilisant l'ansatz (1.13) dans l'expression de l'énergie $\left\langle \psi, H^{(n)}\psi \right\rangle$, on obtient facilement l'énergie en fonction des orbitales $\Phi = (\phi_i)_{i=1,\dots,N}$:

$$\mathcal{E}^{\mathrm{HF}}(\Phi) = \frac{1}{2}\sum_{i=1}^{N}\int_{\mathbb{R}^3}|\nabla\phi_i|^2 + \int_{\mathbb{R}^3}\rho_\Phi V + \frac{1}{2}\int_{\mathbb{R}^6}\frac{\rho_\Phi(x)\rho_\Phi(y) - |\tau_\Phi(x,y)|^2}{|x-y|}\mathrm{d}x\mathrm{d}y, \quad (1.14)$$

où

$$\tau_\Phi(x,y) = \sum_{i=1}^{N}\phi_i(x)\phi_i(y), \quad (1.15)$$

$$\rho_\Phi(x) = \tau(x,x) \quad (1.16)$$

sont respectivement l'opérateur densité et la densité électronique.

On peut considérer sans perte de généralité que les orbitales forment une famille orthonormale, c'est-à-dire que $\mathrm{Gram}\,\Phi = 1$. C'est encore une expression du principe d'exclusion de Pauli : les orbitales électroniques doivent être orthogonales, et en particulier deux orbitales ne peuvent pas être identiques.

Le problème de détermination de l'état fondamental dans la théorie Hartree-Fock est donc de calculer un minimiseur du problème

$$I^{\mathrm{HF}} = \inf\left\{\mathcal{E}^{\mathrm{HF}}(\Phi), \Phi \in (H^1(\mathbb{R}^3,\mathbb{R}))^N, \mathrm{Gram}\,\Phi = 1\right\}. \quad (1.17)$$

L'énergie (1.14) comprend les termes d'énergie cinétique, d'attraction électrons-noyaux, et de répulsion électronique. Ce terme de répulsion, qui est le seul à coupler les orbitales, est à l'origine des difficultés mathématiques du modèle. Il comprend le terme direct, qui est l'énergie potentielle classique de répulsion d'un nuage électronique de densité ρ, et le terme d'échange, qui provient de l'antisymétrie de la fonction d'onde, et a donc une origine purement quantique.

L'équation d'Euler-Lagrange associée au problème (1.17) est

$$F_\Phi \phi_i = \sum_{j=1}^{N} \lambda_{i,j} \phi_j, \tag{1.18}$$

où l'opérateur de Fock F_Φ est défini par

$$(F_\Phi \psi)(x) = -\frac{1}{2}(\Delta\psi)(x) + V(x)\psi(x) + \left(\int_{\mathbb{R}^3} \frac{\rho_\Phi(y)}{|x-y|}\mathrm{d}y\right)\psi(x) - \int_{\mathbb{R}^3} \frac{\tau_\Phi(x,y)}{|x-y|}\psi(y)\mathrm{d}y. \tag{1.19}$$

et $\lambda_{i,j}$ est la matrice symétrique des multiplicateurs de Lagrange associés aux contraintes $\int \phi_i \phi_j = \delta_{i,j}$.

On peut simplifier ces équations en remarquant que le déterminant de Slater $|\phi_1, \phi_2, \ldots, \phi_N\rangle$ est invariant par transformation orthogonale sur les ϕ_i. En conséquence, pour toute matrice orthogonale $U \in \mathcal{O}(N)$,

$$\mathcal{E}^{\mathrm{HF}}(U\Phi) = \mathcal{E}^{\mathrm{HF}}(\Phi).$$

Si Φ vérifie (1.18) avec $\Lambda = (\lambda_{i,j})_{i,j}$, alors $U\phi$ vérifie

$$F_{U\Phi} U\phi_i = \sum_{j=1}^{N} \lambda'_{i,j} \phi_j,$$

où

$$(\lambda'_{i,j})_{i,j} = U\Lambda U^T.$$

À une transformation orthogonale près, on peut donc supposer que la matrice λ_{ij} est diagonale, et on obtient les équations de Hartree-Fock

$$F_\Phi \phi_i = \lambda_i \phi_i. \tag{1.20}$$

Commençons par étudier le cas où F ne dépend pas de Φ, c'est-à-dire où la répulsion électron-électron est négligée. Alors ϕ_i est un vecteur propre de $-\frac{1}{2}\Delta + V$ associé à la valeur propre λ_i, et l'énergie s'écrit

$$E(\Phi) = \sum_{i=1}^{N} \lambda_i.$$

Sans autre contrainte, le minimum serait atteint en prenant pour tous les ϕ_i le premier état propre de $-\frac{1}{2}\Delta + V$. Cependant, comme les électrons sont des fermions, on a nécessairement $\operatorname{Gram} \Phi = 1$, ce qui empêche cette condensation. À la place, le minimum est atteint en prenant pour ϕ_1 le premier état propre, pour ϕ_2 le second, etc. C'est le principe Aufbau ("construction" en allemand), qui justifie les constructions empiriques des chimistes (les règles de Klechkowski ou de Madelung, qui expliquent partiellement la classification périodique des éléments).

L'interaction entre les électrons fait que F_Φ dépend de Φ. On peut alors interpréter les équations de Hartree-Fock comme un problème aux valeurs propres non-linéaire : les ϕ_i sont des vecteurs propres de l'opérateur de champ moyen F_Φ, qui

représente l'opérateur de Schrödinger dans un champ de potentiel prenant en compte une répulsion générée par les électrons dans l'état Φ. Cet effet de rétroaction est la source des difficultés mathématiques et numériques du modèle de Hartree-Fock.

Une propriété des équations permet de reformuler le problème sous une forme plus utile pour les calculs numériques. On a vu que l'énergie $\mathcal{E}^{\mathrm{HF}}(\Phi)$ est invariante par transformation orthogonale : elle dépend donc uniquement du sous-espace engendré par les ϕ_i. On peut utiliser cette propriété pour réécrire l'énergie en fonction de l'opérateur densité τ_Φ. En effet,

$$\tau_\Phi(x,y) = \sum_{i=1}^{N} \phi_i(x)\phi_i(y)$$

est le noyau intégral de l'opérateur de projection sur le sous-espace engendré par les ϕ_i, également noté τ_Φ. On peut donc écrire

$$\mathcal{E}(\Phi) = \mathrm{tr}\left(\left(h + \frac{1}{2}G(\tau_\Phi)\right)\tau_\phi\right),$$

avec

$$\begin{aligned} h &= -\frac{1}{2}\Delta + V, \\ (G(\tau_\Phi)\psi)(x) &= \left(\rho_{\tau_\Phi} \star \frac{1}{|x|}\right)(x)\psi(x) - \int_{\mathbb{R}^3} \frac{\tau(x,y)}{|x-y|}\psi(y)\mathrm{d}y. \end{aligned}$$

On peut également caractériser l'ensemble des τ_Φ issus de $\Phi \in H^1(\mathbb{R}^3)^N$: c'est l'ensemble des projecteurs orthogonaux de rang N à image dans $H^1(\mathbb{R}^3)$:

$$\mathcal{P}_N = \Big\{\tau \in \mathcal{L}^1, \mathrm{Im}(\tau) \subset H^1(\mathbb{R}^3), \tau^2 = \tau^* = \tau, \mathrm{tr}\,\tau = N\Big\},$$

où $\mathcal{L}^1$ est l'ensemble des opérateurs à trace sur $L^2(\mathbb{R}^3)$.

On a donc

$$I^{\mathrm{HF}} = \inf\Big\{\mathcal{E}^{\mathrm{HF}}(\tau), \tau \in \mathcal{P}_N\Big\}.$$

Cette forme, a l'avantage d'être posée directement sur l'espace des fonctions d'onde quotienté par l'invariance orthogonale, alors que le minimiseur éventuel de $\mathcal{E}^{\mathrm{HF}}(\Phi)$ n'est défini qu'à une transformation orthogonale près. Elle permet également une formulation plus aisée des algorithmes.

1.3.2 Algorithmes pour Hartree-Fock

Pour résoudre numériquement le problème d'optimisation (1.17), on discrétise en développant les fonctions d'onde sur une base de Galerkin $(\chi_\alpha)_{\alpha=1\ldots N_b}$, qu'on suppose orthonormale pour simplifier les calculs :

$$\phi_i = \sum_{\alpha=1}^{N_b} C_{\alpha,i}\chi_\alpha.$$

Comme dans le cas continu où l'énergie de Φ ne dépendait que de l'opérateur densité τ_Φ, on peut ici réécrire l'énergie et les contraintes en termes de la matrice densité $D = CC^T$, qui est l'analogue discret de la matrice densité τ_Φ :

$$\mathcal{E}^{\mathrm{HF}}(D) = \mathrm{tr}\left(\left(h + \frac{1}{2}G(D)\right)D\right), \tag{1.21}$$

avec

$$h_{\alpha,\beta} = \left\langle \chi_\alpha, \left(-\frac{1}{2}\Delta + V\right)\chi_\beta \right\rangle,$$
$$G(D)_{\alpha,\beta} = \sum_{\mu,\nu=1}^{N_b} ((\alpha\beta|\mu\nu) - (\alpha\nu|\mu\beta))D_{\mu,\nu},$$

où on note $(\alpha\beta|\mu\nu)$ les intégrales biélectroniques[3]

$$(\alpha\beta|\mu\nu) = \int_{\mathbb{R}^6} \frac{\chi_\alpha(x)\chi_\beta(x)\chi_\mu(y)\chi_\nu(y)}{|x-y|}\mathrm{d}x\mathrm{d}y. \tag{1.22}$$

On a alors le problème de minimisation

$$I^{\mathrm{HF}}_{N_b} = \inf\left\{\mathcal{E}^{\mathrm{HF}}(D), D \in M_{N_b\times N_b}, D^2 = D^T = D, \mathrm{tr}\, D = N\right\}. \tag{1.23}$$

Plusieurs stratégies sont possibles pour résoudre le problème (1.23). Les premières tentatives de résolution utilisaient l'algorithme de Roothaan, parfois appelé "champ auto-cohérent simple" (*simple SCF* pour *Self-Consistent Field*), qui se base sur le principe Aufbau décrit précédemment. Les équations d'Euler-Lagrange, une fois diagonalisées, s'écrivent

$$F(D)C_i = \lambda_i C_i, \tag{1.24}$$

où C_i est la i-ième colonne de la matrice $C_{i,\alpha}$, et $F(D)$ est la matrice de Fock

$$F(D) = h + G(D). \tag{1.25}$$

L'algorithme de Roothaan consiste, pour D_n donné, à calculer $F(D_n)$ et à prendre $D_{n+1} = C_{n+1}C_{n+1}^T$, où C_{n+1} est la matrice formée à partir des N premiers états propres de $F(D_n)$. Cet algorithme a une interprétation physique simple : à chaque étape, on fixe les orbitales dans une configuration particulière, ce qui mène à un problème linéaire, qu'on résoud et qu'on utilise pour mettre à jour les orbitales. Ce procédé est bien défini à condition que la N-ième plus petite valeur propre λ_N de $F(D_n)$ soit séparée de la $N+1$-ième λ_{N+1}. On appelle *gap* l'écart $\lambda_{N+1} - \lambda_N$.

La question mathématique naturelle qui se pose est celle de la convergence de cet algorithme. Il a été observé numériquement par les chimistes que, même pour des systèmes simples et des bases de calcul petites, l'algorithme ne converge pas mais

[3]Le calcul des intégrales biélectroniques (1.22) est le goulot d'étranglement de la plupart des codes de calcul. On utilise souvent des bases de type Gaussienne fois polynôme, pour lequel le calcul de ces intégrales peut se simplifier algébriquement.

oscille entre deux états [SO89]. Ce phénomène a été étudié mathématiquement par Cancès et Le Bris en 2000 [CLB00b]. Ils ont expliqué ces oscillations en étudiant la fonctionnelle

$$E(D, D') = \mathrm{tr}(h(D + D')) + \mathrm{tr}(G(D)D').$$

On peut prouver que, si D_n est la suite de matrices densité générées par l'algorithme de Roothaan et si il existe un gap $\lambda_{N+1} - \lambda_N > 0$ uniforme en n, alors

$$E(D_{n+1}, D_{n+2}) \leq E(D_n, D_{n+1}) - \beta \|D_{n+2} - D_n\|^2$$

pour un certain $\beta > 0$, où $\|D\|$ est la norme de Hilbert-Schmidt (aussi appelée norme de Frobenius) de la matrice D.

Ce résultat implique que $\|D_{n+2} - D_n\|^2$ est sommable, ce qui est une forme faible de convergence de la suite (D_n, D_{n+1}). Cela laisse la possibilité d'obtenir une convergence de (D_n, D_{n+1}) vers (D, D'), avec $D \neq D'$, ce qui explique les oscillations constatées en pratique.

Pour remédier à ces oscillations, les chimistes utilisent parfois un algorithme de Level-Shifting, qui consiste à augmenter artificiellement le gap $\lambda_{N+1} - \lambda_N$ en utilisant à la place de $F(D_n)$ la matrice décalée $F(D_n) - bD_n$. Cancès et Le Bris ont montré que, si le paramètre de *shift* b est suffisamment important, $D_{n+1} - D_n \to 0$, mais que la limite éventuelle de D_n n'est en général pas un minimiseur. Pour obtenir un algorithme stable sans les défauts du Level-Shifting, ces auteurs ont proposé l'algorithme ODA (pour *Optimal Damping Algorithm* [CLB00a]), qui combine la bonne vitesse de convergence locale de l'algorithme de Roothaan avec la stabilisation des algorithmes amortis. Dans ce cas, on peut également prouver que $\|D_{n+1} - D_n\| \to 0$.

Même si ces résultats donnent une bonne idée de ce qui peut se produire en pratique et montrent l'intérêt d'une stabilisation de l'algorithme, ils ne fournissent pas de preuve de convergence. En effet, $D_{n+1} - D_n \to 0$ n'implique pas que D_n converge, comme le montre l'exemple de la divergence de la série harmonique $x_n = \sum_{i=1}^{n} 1/i$. Les divergences lentes de D_n ne sont donc pas exclues. De plus, il est intéressant en pratique de connaître la vitesse de convergence des algorithmes, qui est hors de portée de ces méthodes.

1.3.3 Résultats du chapitre 2

Dans le chapitre 2 de cette thèse, on étudie la convergence des algorithmes de gradient, de Roothaan et de Level-Shifting pour le modèle Hartree-Fock. Les résultats présentés ont fait l'objet d'un article publié dans M2AN [Lev12b].

Dans le cas classique où on cherche à minimiser une fonctionnelle non contrainte $f(x)$, un algorithme simple est la descente de gradient à pas fixe t

$$x_{n+1} = x_n - t\nabla f(x_n). \tag{1.26}$$

Pour étendre cet algorithme au cas Hartree-Fock, on travaille sur la variété Riemannienne

$$\mathcal{P} = \{D \in M_{N_b \times N_b}, D^2 = D^T = D, \mathrm{tr}\, D = N\}.$$

On peut calculer le gradient de $\mathcal{E}^{\mathrm{HF}}$ projeté sur l'espace tangent à $\mathcal{P}$

$$\begin{aligned}\nabla_{\mathcal{P}}\,\mathcal{E}^{\mathrm{HF}}(D) &= [D,[D,\nabla\,\mathcal{E}^{\mathrm{HF}}(D)]],\\ &= [D,[D,F(D)]]\end{aligned} \tag{1.27}$$

où $[A,B] = AB - BA$ est le commutateur. Pour obtenir une généralisation de l'algorithme de descente (1.26) qui reste sur la variété $\mathcal{P}$ d'une itération à l'autre, on utilise le changement de base

$$D_{n+1} = U_n D_n U_n^T, \tag{1.28}$$

où U_n est une matrice orthogonale. On peut toujours écrire U_n comme $\exp(tA)$, où A est une matrice antisymétrique, et t représente le pas. Si t est petit, on peut développer $U_n = 1 + tA + o(t^2)$, ce qui donne

$$D_{n+1} = D_n + t[D_n, A] + o(t^2).$$

On est donc conduit à choisir $A = [D_n, F(D_n)]$, et on obtient

$$D_{n+1} = D_n + t\nabla_{\mathcal{P}}\,\mathcal{E}^{\mathrm{HF}}(D_n) + o(t^2). \tag{1.29}$$

On prouve alors que, pour t suffisamment petit,

$$\mathcal{E}^{\mathrm{HF}}(D_{n+1}) \leq \mathcal{E}^{\mathrm{HF}}(D_n) - \beta \left\|\nabla_{\mathcal{P}}\,\mathcal{E}^{\mathrm{HF}}(D_n)\right\|^2,$$

où $\beta > 0$. On obtient donc par (1.29) que $\|D_{n+1} - D_n\|^2$ est sommable, exactement comme dans la preuve de Cancès et Le Bris.

Comment aller plus loin et prouver la convergence de la suite D_n ? Revenons au cas classique où on minimise une fonction $f(x)$ par un algorithme de descente de gradient à pas fixe. Pour prouver la convergence de cet algorithme, on utilise souvent une information de non-dégénérescence du second ordre. Par exemple, on suppose que la Hessienne $\nabla^2 f(x_0)$ est définie positive pour un certain minimum local x_0, et on prouve la convergence quand l'initialisation de la méthode est suffisamment proche de x_0.

Ici, cette information est très difficile à obtenir (cette difficulté est à relier au manque de preuves d'unicité de l'état fondamental de Hartree-Fock). On peut cependant la remplacer par une information plus faible, obtenue grâce à l'analyticité de la fonctionnelle $\mathcal{E}^{\mathrm{HF}}$. En effet, pour une fonctionnelle analytique, toutes les dérivées ne peuvent pas être nulles simultanément, et on a donc non-dégénérescence à un certain ordre n_0. L'inégalité entre la fonctionnelle et sa dérivée qui en résulte est connue sous le nom d'inégalité de Łojasiewicz [Łoj65]. Elle s'écrit de la façon suivante : si f est analytique sur $\mathbb{R}^n$, alors pour tout $x_0 \in \mathbb{R}^n$, il existe un voisinage U de x_0, un entier $n_0 \geq 2$ et un réel $\kappa > 0$ tels que, quand $x \in U$,

$$|f(x) - f(x_0)|^{1-1/n_0} \leq \kappa \|\nabla f(x)\|.$$

En appliquant cette inégalité à notre problème, on obtient

Théorème 1.1 (Convergence de l'algorithme de gradient)**.** *Il existe $\alpha > 0$ tel que, si $D_0 \in \mathcal{P}$ et D_n est la suite des itérés donnés par* (1.28) *avec $t < \alpha$, alors D_n converge vers une solution des équations de Hartree-Fock.*

Cette méthode de preuve se généralise aux algorithmes de Roothaan et Level-Shifting. Comme expliqué précédemment, Cancès et Le Bris ont prouvé que, pour l'algorithme de Roothaan, la fonctionnelle

$$E(D, D') = \mathrm{tr}(h(D + D')) + \mathrm{tr}(G(D)D')$$

satisfait l'inégalité

$$E(D_{n+1}, D_{n+2}) \leq E(D_n, D_{n+1}) - \beta \|D_{n+2} - D_n\|^2 ,$$

sous une condition de gap uniforme.

On prouve la convergence de (D_n, D_{n+1}) par une méthode similaire à celle utilisée pour l'algorithme de gradient (on applique l'inégalité de Łojasiewicz à $E(D, D')$ sur $\mathcal{P} \times \mathcal{P}$), à condition de pouvoir relier $\|D_{n+2} - D_n\|$ à $\nabla_{\mathcal{P}\times\mathcal{P}} E(D_n, D_{n+1})$. Ce lien, qui peut être réalisé par des manipulations de commutateurs, montre le rapport étroit entre l'algorithme de Roothaan et la descente de gradient. Il est ensuite aisé de généraliser ce résultat à l'algorithme de Level-Shifting.

Dans les trois cas (algorithme de gradient, de Roothaan et de Level-Shifting), on obtient des estimations de vitesse de convergence. Bien que ces estimations dépendent des constantes de l'inégalité de Łojasiewicz et soient donc implicites, elles permettent dans une certaine mesure de comparer les méthodes entre elles. Par exemple, le taux de convergence de l'algorithme de Roothaan dépend du gap γ, alors que l'algorithme du gradient n'en dépend pas. On peut également extraire de ces taux de convergence des informations asymptotiques, notamment le comportement de l'algorithme de Level-Shifting quand le paramètre de shift tend vers l'infini, ou le comportement dans le pire cas des algorithmes quand on augmente la taille de la base.

Ces trois algorithmes ont été implémentés dans le code ACCQUAREL, développé par Guillaume Legendre et l'auteur. Les résultats que nous avons obtenus (voir par exemple la figure 1.1) montrent que la convergence est linéaire, ce qui suggère que le point critique vers lequel ces algorithmes convergent est non-dégénéré. L'algorithme de gradient converge bien plus lentement que l'algorithme de Roothaan, un fait que les estimations (grossières) de vitesse de convergence que nous avons montrées ne permet pas d'expliquer.

Ces résultats sont une première étape vers l'étude systématique de la convergence d'algorithmes pour Hartree-Fock. De nombreux points restent à clarifier, comme la structure du voisinage des minimiseurs, qui influe directement sur la vitesse de convergence, ou le comportement d'algorithmes plus sophistiqués comme ODA ou DIIS (pour *Direct Inversion in the Iterative Subspace* [Pul82]), l'algorithme le plus utilisé en chimie quantique, qui a été récemment lié à une méthode de quasi-Newton [RS11].

1.3.4 Existence de solutions pour Hartree-Fock

On s'intéresse maintenant au problème théorique de l'existence de solutions des équations de Hartree-Fock, et de ses généralisations au cas de plusieurs déterminants

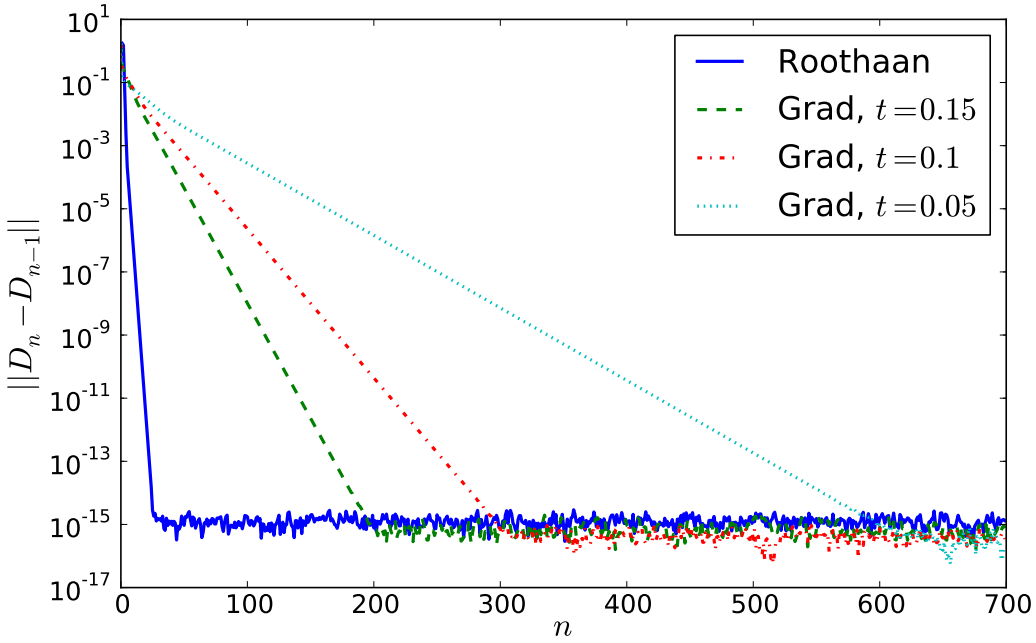

FIGURE 1.1: Calcul sur l'atome de carbone ($N = Z = 6$) avec les algorithmes de Roothaan et de gradient.

de Slater (modèle multiconfiguration), relativiste (modèle Dirac-Fock), et relativiste à plusieurs déterminants (modèle Dirac-Fock multiconfiguration), objet du chapitre 3 de cette thèse.

Lieb et Simon [LS77] ont prouvé en 1977 l'existence de minimiseurs de la fonctionnelle de Hartree-Fock (1.17) pour les systèmes pour lesquels $N < Z + 1$. Lions [Lio87] a généralisé cette preuve en 1987, et montré l'existence d'une infinité de points critiques (c'est-à-dire de solutions des équations de Hartree-Fock). Le seul résultat d'unicité disponible [GH12] considère la limite $Z \to \infty$, et la question reste ouverte en général.

On rappelle ici très brièvement la stratégie de preuve utilisée par Lieb, Simon [LS77] et Lions [Lio87]. Pour des présentations générales de ce résultat, on peut consulter [LBL05; CLBM06].

Cette preuve utilise de façon cruciale l'inégalité de Hardy, qui permet de contrôler l'énergie potentielle Coulombienne par l'énergie cinétique. Cette inégalité s'écrit

$$\forall \phi \in H^1(\mathbb{R}^3), \left\| \frac{\phi}{|x|} \right\|_{L^2} \leq 2 \|\nabla u\|_{L^2} . \tag{1.30}$$

Elle permet par exemple de montrer, par l'inégalité de Cauchy-Schwarz, que

$$\left| \int V \phi^2 \right| \leq 2Z \|\phi\|_{L^2} \|\nabla u\|_{L^2} .$$

Ainsi, on peut contrôler l'énergie potentielle de ϕ dans le champ V par (la racine de) son énergie cinétique.

À l'aide de cette inégalité, on montre qu'une suite minimisante $(\Phi_n)_{n\in\mathbb{N}}$ de (1.17) est bornée en norme H^1. Sans perte de généralité, on peut supposer qu'elle converge faiblement dans H^1, fortement dans L^p_{loc}, $2 \leq p < 6$, et presque partout vers Φ.

Par semi-continuité inférieure faible de $\mathcal{E}^{\text{HF}}$ sur H^1, on montre que $\mathcal{E}^{\text{HF}}(\Phi) \leq I^{\text{HF}}$. Toute la difficulté réside en le fait que la convergence faible implique uniquement que $\operatorname{Gram}\Phi \leq 1$. Pour prouver que $\operatorname{Gram}\Phi = 1$, et donc que Φ est un minimiseur du problème, on utilise des informations supplémentaires sur la suite minimisante : on choisit une suite qui vérifie la condition de Palais-Smale, c'est-à-dire telle que

$$F_{\Phi_n}\phi_n^i = \lambda_i\phi_n^i + \varepsilon_n^i,$$

avec $\varepsilon_n^i \to 0$ en norme H^{-1}. L'existence d'une telle suite est garantie par le principe variationnel d'Ekeland [Eke74]. Ensuite, on peut calculer

$$\begin{aligned}\lambda_i \langle\phi_i, \phi_i\rangle &= \langle\phi_i, F_\Phi\phi_i\rangle \\ &\leq \liminf \left\langle\phi_n^i, F_{\Phi_n}\phi_n^i\right\rangle \\ &= \liminf \lambda_i \left\langle\phi_n^i, \phi_n^i\right\rangle \\ &= \lambda_i.\end{aligned}$$

Si $\lambda_i < 0$, on peut donc en déduire que $\|\phi_i\|_{L^2} = 1$, ce qui montre que Φ_n converge fortement, et donc le résultat.

Nous allons prouver la condition $\lambda_i < 0$ sur les multiplicateurs en utilisant la notion d'indice de Morse, qui mesure le nombre de directions de descente d'une fonctionnelle autour d'un point critique. Le principe variationnel de Borwein-Preiss [FG92; BP87] permet d'obtenir une suite minimisante avec un indice de Morse nul, c'est-à-dire dont la Hessienne est (asymptotiquement) positive. On calcule alors que, en prenant une fonction test ψ orthogonale à tous les ϕ_i, on a

$$\left\langle\psi, \left(-\frac{1}{2}\Delta + V + \rho_n^i \star \frac{1}{|x|} - \lambda_i\right)\psi\right\rangle \geq -\delta_n\|\psi\|^2, \tag{1.31}$$

où $\int \rho_n^i = N - 1$, et $\delta_n \to 0$ (on a oublié à bon droit l'énergie d'échange, qui est négative).

Maintenant, choisissons pour ψ une fonction radiale, et dilatons-la par la transformation $\psi_\mu = \mu^{-3/2}\psi(\frac{x}{\mu})$, tout en préservant l'orthogonalité avec les ϕ_i. La fonction ψ_μ est soumise à un champ moyen $V + \rho_n^i \star \frac{1}{|x|}$ qui, si μ est assez grand, peut être approché par $\frac{(N-1)-Z}{|x|}$. Or on sait par la théorie de l'atome d'hydrogène que, pour $Z < N - 1$, l'opérateur $-\frac{1}{2}\Delta + \frac{(N-1)-Z}{|x|}$ admet une infinité de valeurs propres négatives. C'est incompatible avec (1.31) sauf si $\lambda_i < 0$, ce qui finit la preuve.

Par le même principe, on peut montrer la convergence de suites de Palais-Smale à indice de Morse fini, c'est-à-dire dont la Hessienne admet (asymptotiquement) un espace négatif de dimension finie. Par un argument de min-max, on peut trouver pour tout j un point critique d'indice de Morse au plus j.

1.3.5 Modèle multiconfiguration

Le modèle multiconfiguration est une généralisation du modèle Hartree-Fock. En considérant non plus un unique déterminant de Slater mais une combinaison linéaire

finie de déterminants, la méthode multiconfiguration permet en théorie une précision arbitraire, au prix d'un temps de calcul considérablement plus important.

L'ansatz multiconfiguration consiste à se donner un entier $K > N$, et à poser

$$\psi = \sum_{1\leq i_1<\cdots<i_N\leq K} a_{i_1,\ldots,i_N} \left|\phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N}\right\rangle, \tag{1.32}$$

où les $(\phi_i)_{i=1,\ldots,K}$ sont les orbitales qui constituent ψ.

En suivant [Lew04], on peut écrire l'énergie en fonction des orbitales Φ et des coefficients a sous la forme matricielle

$$\mathcal{E}^{\mathrm{MCHF}}(a,\Phi) = \left\langle \Phi, \left(\left(-\frac{1}{2}\Delta + V\right)\Gamma_a + W_{a,\Phi}\right)\Phi\right\rangle_{(L^2(\mathbb{R}^3))^K}, \tag{1.33}$$

où Γ_a et $W_{a,\Phi}$ sont des matrices symétriques $K \times K$ dont l'expression exacte est donnée au chapitre 3. Les valeurs propres de Γ_a sont appelées nombres d'occupation, et représentent le "poids" total donné à l'orbitale correspondante dans la fonction d'onde. Ces nombres d'occupation jouent un rôle important pour la suite.

La condition de normalisation s'écrit $\operatorname{Gram}\Phi = 1, \sum_I a_I^2 = 1$. Les équations d'Euler-Lagrange correspondantes sont

$$\begin{cases} \left(\left(-\frac{1}{2}\Delta + V\right)\Gamma_a + 2W_{a,\Phi}\right)\Phi = \Lambda\Phi, \\ \mathcal{H}_\Phi a = Ea, \end{cases} \tag{1.34}$$

où

$$(\mathcal{H}_\Phi)_{I,J} = \left\langle \psi_I, H^N \psi_J \right\rangle \tag{1.35}$$

est la matrice $\binom{K}{N} \times \binom{K}{N}$ du Hamiltonien à N corps H^N dans la base formée par les déterminants de Slater

$$\psi_I = \left|\phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N}\right\rangle.$$

Le problème mathématique majeur de la méthode multiconfiguration, outre la complexité algébrique des équations, est qu'on ne peut plus utiliser l'invariance de l'énergie par transformation orthogonale des ϕ_i pour découpler le problème : on peut diagonaliser soit la matrice des multiplicateurs de Lagrange Λ, soit la matrice des nombres d'occupation Γ_a, mais pas les deux simultanément.

Notons

$$I^K = \inf\left\{\mathcal{E}^{\mathrm{MCHF}}(a,\Phi), a \in \mathbb{R}^{\binom{K}{N}}, \Phi \in H^1(\mathbb{R}^3), \|a\| = 1, \operatorname{Gram}\Phi = 1\right\}.$$

l'énergie fondamentale du problème. Pour $K = N$, on retrouve le modèle Hartree-Fock, et dans la limite $K \to \infty$, on obtient le problème complet à N corps (1.10). Pour $K = N + 1$, les propriétés algébriques des déterminants de Slater font que $I^{N+1} = I^N$. Le premier cas intéressant est donc $K = N + 2$. Ce cas a été étudié mathématiquement par Le Bris [LB94], qui a montré l'existence de minimiseurs ainsi que l'inégalité stricte $I^{N+2} < I^N$ (qui justifie l'intérêt du modèle multiconfiguration).

Cette inégalité a plus tard été généralisée à $I^{K+2} < I^K$ pour tout $K \geq N$ par Friesecke [Fri03a]. La preuve utilise le fait que la singularité Coulombienne produit sur la fonction d'onde à N corps une singularité aux points de coalescence des électrons : la fonction d'onde ψ n'est pas dérivable en les points pour lesquels $x_i = x_j, i \neq j$. Cette caractéristique ne peut pas être reproduite par les solutions des équations de Hartree-Fock multiconfiguration, qui sont nécessairement analytiques en dehors des noyaux.

Une étude systématique de l'existence de solutions de cette équation pour tout K a été réalisée par Friesecke [Fri03b] et Lewin [Lew04], qui fournissent deux preuves différentes de l'existence de minimiseurs pour ce problème. La preuve de Friesecke utilise un principe de concentration-compacité sur la fonction d'onde à N corps, alors que la preuve de Lewin utilise des méthodes basées sur l'équation d'Euler-Lagrange, dans l'esprit de la méthode de Lions [Lio87]. On détaille ici la preuve de Lewin, sur laquelle l'étude du chapitre 3 est basée.

De même que dans la preuve de Lions [Lio87], on prend une suite minimisante (a_n, Φ_n) de Palais-Smale avec indice de Morse nul. En extrayant, on peut imposer que Γ_n converge vers une matrice Γ semi-définie positive. En utilisant l'invariance orthogonale déjà exploitée dans le cas Hartree-Fock, on peut diagonaliser $\Gamma_n = \mathrm{diag}(\gamma_n^i)_{i=1,\dots,N}$. Comme l'énergie est bornée, on obtient par l'inégalité de Hardy que $\sqrt{\gamma_n^i}\phi_n^i$ est borné en norme H^1.

La difficulté ici est que si $\gamma_n^i \to 0$, alors ϕ_n^i peut exploser en norme H^1. On prouve, grâce à l'équation d'Euler-Lagrange en a, que si $\gamma_n^i \to 0$, alors $\sqrt{\gamma_n^i}\phi_n^i \to 0$ en norme H^1. Les orbitales dont le nombre d'occupation γ_n^i tend vers 0 ne contribuent donc pas à l'énergie ni aux équations d'Euler-Lagrange, et on peut donc se restreindre à une suite minimisante à K' orbitales, avec $K' \leq K$.

Cette nouvelle suite minimisante vérifie $\Gamma_n \geq \gamma$ au sens des matrices symétriques, où $\gamma > 0$. À partir de là, on peut procéder comme dans la preuve de Lions, et obtenir alors la convergence vers un minimiseur. Par l'inégalité $I^{K+2} < I^K$, on sait alors que $K' \geq K - 1$, mais on ne peut pas éliminer la possibilité que K' soit égal à $K - 1$.

1.3.6 L'opérateur de Dirac

Nous avons jusqu'à présent négligé les effets relativistes : la théorie quantique est indépendante de la vitesse de la lumière c. On sait qu'en mécanique classique, les effets relativistes deviennent importants pour des vitesses proches de c. Or, en unités atomiques, la vitesse des électrons de coeur (proches du noyau) d'un atome est de l'ordre de Z, à comparer avec $c \approx 137$. Pour les atomes lourds, ces électrons subissent donc une contraction non-négligeable des distances. Cette contraction produit un écrantage plus fort du potentiel d'attraction du noyau, ce qui a un impact sur les électrons de valence (éloignés du noyau), qui déterminent les propriétés chimiques des éléments. Par exemple, ce sont les effets relativistes qui expliquent les différences de couleur entre l'argent et l'or [PD79]. Ces effets sont prédits par la théorie de Dirac, que nous décrivons brièvement.

Dirac a introduit en 1928 [Dir28] l'opérateur qui porte maintenant son nom pour remédier à certains défauts de la théorie quantique de Schrödinger. Premièrement, la notion de spin, introduite par Pauli pour expliquer le dédoublement des états

quantiques, était purement phénoménologique. Deuxièmement, l'équation de Schrödinger, d'ordre deux en espace et un en temps, ne respectait pas le principe de la relativité restreinte qui veut que le temps et l'espace occupent des places symétriques dans les équations. La théorie de Dirac a permis de résoudre ces deux obstacles théoriques, d'affiner les prédictions quantitatives de la mécanique quantique, et de poser les bases de l'électrodynamique quantique (QED, pour *Quantum Electrodynamics*) [Tha92].

Pour obtenir une équation qui vérifie l'invariance relativiste, Dirac est partir de l'équation relativiste pour l'énergie d'une particule libre

$$E^2 = c^4 + c^2 p^2, \tag{1.36}$$

où p est la quantité de mouvement, et c la vitesse de la lumière.

Une première possibilité pour quantifier cette équation est d'utiliser les règles de substitution $E \leftrightarrow i\partial_t, p \leftrightarrow -i\nabla$, pour obtenir l'équation de Klein-Gordon

$$\partial_t^2 \psi = c^2 \Delta \psi - c^4 \psi. \tag{1.37}$$

Cette équation n'est pas satisfaisante pour modéliser une particule, notamment parce qu'il est impossible d'obtenir une densité de probabilité positive satisfaisant une loi de conservation. Une autre possibilité est de prendre la racine de (1.36) pour obtenir une équation du premier ordre en temps

$$i\partial_t \psi = \sqrt{-\Delta c^2 + c^4}\psi, \tag{1.38}$$

mais cette équation n'est alors plus locale en espace. L'idée de Dirac est de considérer ψ comme un vecteur et non un scalaire. Il postule un opérateur D_c du premier ordre de la forme

$$D_c \psi = -ic\alpha \cdot \nabla + c^2 \beta, \tag{1.39}$$

et essaie ensuite d'obtenir $D_c^2 = c^4 - c^2\Delta$.

Le plus petit espace pour lequel cela est possible est $\mathbb{C}^4$ [Tha92], et la représentation choisie habituellement est donnée par

$$\alpha_k = \begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix}, \beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \tag{1.40}$$

avec les matrices de Pauli

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.41}$$

La difficulté mathématique majeure provient du spectre de cet opérateur. D_c peut se décomposer selon deux espaces propres E_c^+ et E_c^- tels que

$$D_c = P_c^+ \sqrt{c^4 - \Delta c^2} P_c^+ - P_c^- \sqrt{c^4 - \Delta c^2} P_c^-, \tag{1.42}$$

où $P_c^\pm$ est le projecteur sur $E_c^\pm$.

Ceci montre que le spectre de D_c est $]-\infty, -c^2] \cup [c^2, \infty[$, avec d'importantes conséquences en pratique : comment, par exemple, donner un sens à l'état fondamental de l'opérateur $D_c + V$, non borné inférieurement ? Comment interpréter les états d'énergie négatives ? Peut-on avoir une caractérisation variationnelle des solutions ? Le problème est encore pire dans le problème à plusieurs corps : dans ce cas, l'opérateur $H^{(n)}$ a pour spectre $\mathbb{R}$ tout entier, et c'est un problème ouvert de savoir si il possède des valeurs propres [Der12] (qui seraient de toute façon dynamiquement instables).

Ces difficultés d'interprétation font que la théorie de Dirac ne peut pas être considérée comme un modèle complet. Une réponse partielle à ces problèmes est apportée par l'électrodynamique quantique, théorie plus complexe que nous n'aborderons pas ici. Malgré ses faiblesses, l'opérateur de Dirac fait apparaître naturellement la notion de spin, et explique la structure fine des raies d'émission de l'atome d'hydrogène. Combiné avec des corrections perturbatives issues de la QED, il permet des calculs de chimie quantique extrêmement précis. Comme pour l'Hamiltonien non-relativiste, on utilise des approximations pour pouvoir traiter le problème à N corps.

1.3.7 Le modèle Dirac-Fock

Le modèle de Dirac-Fock est obtenu en remplaçant l'énergie cinétique $-\frac{1}{2}\Delta$ par l'opérateur de Dirac D_c dans l'expression de l'énergie de Hartree-Fock. Les orbitales ψ_i sont maintenant à valeurs dans $\mathbb{C}^4$, et l'énergie s'écrit

$$\mathcal{E}^{\mathrm{DF}}(\Psi) = \sum_{i=1}^{N} \langle \psi_i, D_c \psi_i \rangle + \int_{\mathbb{R}^3} \rho_\Psi V + \frac{1}{2} \int_{\mathbb{R}^6} \frac{\rho_\Psi(x)\rho_\Psi(y) - \mathrm{tr}_{\mathbb{C}^4}(\tau_\Psi(x,y)\tau_\Psi(y,x))}{|x-y|} \mathrm{d}x\mathrm{d}y, \tag{1.43}$$

où

$$\tau_\Psi(x,y) = \sum_{i=1}^{N} \psi_i(x)\psi_i(y)^*, \tag{1.44}$$

$$\rho_\Psi(x) = \mathrm{tr}_{\mathbb{C}^4}\, \tau(x,x) \tag{1.45}$$

sont respectivement la matrice densité et la densité électronique sommée sur les spins. De même que dans le cas Hartee-Fock, les équations d'Euler-Lagrange prennent la forme d'un problème aux valeurs propres non-linéaire, les équations de Dirac-Fock.

Cette fonctionnelle a été étudiée par Esteban et Séré [ES99]. Le résultat principal est un théorème analogue à celui de Lions pour Hartree-Fock : il existe une infinité de solutions aux équations de Dirac-Fock.

Une première difficulté pour réaliser ce programme est que la stratégie utilisée par Lions ne donne qu'un contrôle par le dessus des multiplicateurs de Lagrange $\lambda_i < c^2$. Pour assurer $\lambda_i > 0$, et ainsi un gap à l'opérateur de Fock, Esteban et Séré remplacent la contrainte $\mathrm{Gram}\, \Psi = 1$ par une fonctionnelle pénalisée

$$\mathcal{F}_p(\Psi) = \mathcal{E}^{\mathrm{DF}}(\Psi) - \pi_p(\Psi), \tag{1.46}$$

avec

$$\begin{aligned}\pi_p(\Psi) &= \operatorname{tr}\left((\operatorname{Gram}\Psi)^p(1-\operatorname{Gram}\Psi)^{-1}\right) \\ &= \sum_{i=1}^{n}\frac{\sigma_i^p}{1-\sigma_i} \\ &= \sum_{i=1}^{n} e_p(\sigma_i),\end{aligned} \tag{1.47}$$

où les σ_i sont les valeurs propres de $\operatorname{Gram}\Psi$.

De la sorte, un point critique de $\mathcal{F}_p$ vérifie, à transformation orthogonale près,

$$H_\Psi\psi_i = \lambda_i\psi_i, \tag{1.48}$$

avec

$$\lambda_i = e_p'(\sigma_i) > 0.$$

Pour chercher des points critiques de $\mathcal{F}_p$, on se heurte à la difficulté que la fonctionnelle a un indice de Morse infini en tout point, à cause du spectre négatif de l'opérateur de Dirac. Une stratégie de minimisation comme utilisée précédemment ne peut pas fonctionner, et on doit donc chercher un point critique. On peut néanmoins utiliser la concavité (pour c suffisamment grand) de la fonctionnelle $\mathcal{E}^{\mathrm{DF}}$ dans les directions du sous-espace d'énergie négative E_c^- de D_c pour définir la fonctionnelle réduite

$$I_p(\Psi^+) = \sup_{\Psi^-\in(E^-)^N,\operatorname{Gram}(\Psi^++\Psi^-)<1} \mathcal{F}_p(\Psi^+ + \Psi^-) \tag{1.49}$$

Des arguments de min-max similaires à ceux utilisés par Lions permettent de conclure à l'existence de points critiques de cette fonctionnelle. On passe ensuite à la limite $p \to \infty$. Comme e_p' converge uniformément vers 0 et qu'on obtient par ailleurs un contrôle $e_p'(\sigma_i) = \lambda_{i,p} \geq h_0 > 0$ uniforme en p, on doit nécessairement avoir $\sigma_{i,p} \to 1$, et donc, à la limite, $\operatorname{Gram}\Psi = 1$. On obtient ainsi une infinité de solutions Ψ_j, sous la condition

$$c > \frac{1}{2}\left(\frac{\pi}{2}+\frac{2}{\pi}\right)\max(Z, 3N-1). \tag{1.50}$$

La première de ces solutions Ψ_0 est obtenue par un principe de min-max qui en fait un bon candidat pour un "état fondamental" de la fonctionnelle de Dirac-Fock (voir le chapitre 3 pour plus de détails). En utilisant cette propriété, on peut montrer leur convergence quand $c \to \infty$ vers un minimiseur de la fonctionnelle de Hartree-Fock [ES01].

1.3.8 Dirac-Fock multiconfiguration

Le modèle de Dirac-Fock multiconfiguration est simplement le modèle multiconfiguration, où l'on remplace l'opérateur $-\frac{1}{2}\Delta$ par l'opérateur de Dirac D_c. L'objectif

de la partie 3 de cette thèse est de prouver l'existence de solutions aux équations de Dirac-Fock multiconfiguration

$$\left((D_c + V)\,\Gamma_a + 2W_{a,\Psi}\right)\Psi = \Lambda\Psi, \tag{1.51}$$

$$\mathcal{H}_\Psi a = Ea. \tag{1.52}$$

Pour prouver l'existence de solutions aux équations de Dirac-Fock, on utilisait de façon centrale la possibilité de "séparer" les parties positives et négatives de l'opérateur de Fock. Plus précisément, l'opérateur

$$D_c - \frac{Z}{|x|}$$

possède un trou spectral autour de 0 tant que

$$c > \frac{1}{2}\left(\frac{\pi}{2} + \frac{2}{\pi}\right)Z.$$

Dans le cadre multiconfiguration, l'analyse de l'opérateur de Fock $(D_c + V)\Gamma_a + 2W_{a,\Psi}$ est compliquée par le fait que Γ_a ne commute pas avec $W_{a,\Psi}$. On ne peut prouver un trou spectral sur cet opérateur que pour des valeurs de c dépendant du plus petit nombre d'occupation de Γ_a. Ainsi, on ne peut pas montrer la compacité d'une suite de Palais-Smale pour laquelle Γ_{a_n} n'est pas bornée inférieurement. On ne peut même pas montrer, comme dans le cadre non-relativiste, que les orbitales dont le nombre d'occupation tend vers 0 ne contribuent pas aux équations d'Euler-Lagrange, et on doit trouver une autre stratégie de preuve.

1.3.9 Résultats du chapitre 3

Dans le chapitre 3 de cette thèse, on prouve l'existence de solutions des équations de Dirac-Fock multiconfiguration, pour c suffisamment grand. Les résultats présentés ont fait l'objet d'un article soumis pour publication [Lev13].

Pour contourner les difficulté mentionnées précédemment, on s'inspire de la limite non-relativiste de ces équations. Dans ce cas, on sait par les travaux de Friesecke [Fri03a] que $I^{K+2} < I^K$. On se place à partir de maintenant dans le cas où $I^K < I^{K-1}$ (ce qui est au moins le cas pour un K sur deux). Alors, chaque minimiseur vérifie $\Gamma > 0$. Par la compacité de l'ensemble de ces minimiseurs, prouvée par Lewin [Lew04], il existe une borne inférieure uniforme $\gamma_0 > 0$ telle que $\Gamma_a \geq \gamma_0$.

On cherche maintenant à exploiter cette information pour prouver l'existence de solutions aux équations de Dirac-Fock multiconfiguration. Pour ce faire, on se donne un $\gamma < \gamma_0$, et on formule un principe variationnel dans le domaine réduit où $\Gamma \geq \gamma$. En passant ensuite à la limite non-relativiste, on vérifie que les solutions du principe variationnel ne saturent pas la contrainte, et on obtient des solutions des équations pour c suffisamment grand.

Posons les ensembles

$$\begin{aligned} S_\gamma &= \{a \in \mathbb{C}^{\binom{K}{N}}, \|a\| = 1, \Gamma_a \geq \gamma\}, \\ \Sigma &= \{\Psi \in E^K, \operatorname{Gram}\Psi = 1\}, \\ \Sigma^+ &= \Sigma \cap (E_c^+)^K, \end{aligned}$$

et, pour tout $\Psi \in E^K$ tel que $\operatorname{Gram} \Psi > 0$, la normalisation

$$g(\Psi) = (\operatorname{Gram} \Psi)^{-1/2} \Psi.$$

Alors, notre principe variationnel s'écrit

$$I_{c,\gamma} = \inf_{a \in S_\gamma, \Psi^+ \in \Sigma^+} \sup_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)). \tag{1.53}$$

On commence par prouver que ce principe variationnel est bien posé :

Théorème 1.2. *Soit $N < Z + 1$, $0 < \gamma < \gamma_0$. Pour c suffisamment grand, il existe un triplet $a_* \in S_\gamma, \Psi_*^+ \in \Sigma^+, \Psi_*^- \in (E^-)^K$ solution du principe variationnel* (1.53)*. En notant $\Psi_* = g(\Psi_*^+ + \Psi_*^-)$, Ψ_* vérifie*

$$\left((D_c + V)\,\Gamma_{a_*} + 2W_{a_*,\Psi_*}\right)\Psi_* = \Lambda_* \Psi_*.$$

Si de plus $\Gamma_{a_*} > \gamma$, alors a_* vérifie $\mathcal{H}_{\Psi_*}\, a_* = I_{c,\gamma}\, a_*$.

Pour prouver ce théorème, on procède en plusieurs étapes. On commence par montrer, comme dans [ES99], la compacité des suites de Palais-Smale dont les multiplicateurs de Lagrange sont en dehors du spectre essentiel de l'opérateur $D_c\Gamma$. Puis, à (a, Ψ^+) fixé, on étudie le principe variationnel

$$\sup_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)),$$

et on montre que celui-ci est concave, en utilisant des arguments notamment inspirés de [ES01]. Finalement, on montre que des conditions du second ordre de type Borwein-Preiss pour pour la fonctionnelle

$$\mathcal{F}(a, \Psi^+) = \sup_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-))$$

impliquent des bornes sur les multiplicateurs de Lagrange qui permettent de conclure à la convergence.

On étudie ensuite la limite non-relativiste de ces solutions :

Théorème 1.3. *Soit $N < Z + 1$, $0 < \gamma < \gamma_0$, et c_n une suite qui tend vers l'infini. En notant (a_n, Ψ_n) les solutions trouvées dans le théorème 1.2,*

$$a_n \to a,$$
$$\Psi_n \to \begin{pmatrix} \Phi \\ 0 \end{pmatrix}$$

en norme H^1, où (a, Φ) est un minimiseur du problème

$$I^K = \inf \left\{ \mathcal{E}^{MCHF}(a, \Phi), \|a\| = 1, \Phi \in (H^1(\mathbb{R}^3, \mathbb{C}^2))^K, \operatorname{Gram} \Phi = 1 \right\}.$$

Comme la limite vérifie $\Gamma_a \geq \gamma_0 > \gamma$, on obtient immédiatement le

Corollaire 1.1. *Si $I^K < I^{K-1}$, $N < Z + 1$, pour c suffisamment grand, il existe des solutions aux équations de Dirac-Fock multiconfiguration.*

La condition $N < Z + 1$ est classique, même dans le modèle non-relativiste mono-configuration, et correspond au fait qu'un électron qui s'échappe à l'infini voit un potentiel effectif $\frac{(N-1)-Z}{|x|}$. La condition $I^K < I^{K-1}$ est naturelle, et était déjà utilisée par Lewin [Lew04] dans le cadre non-relativiste. Puisque $I^{K+2} < I^K$ pour tout $K \geq N$, elle est vérifiée pour au moins un K sur deux.

Le gros défaut de cette preuve est qu'elle est implicite : on sait seulement qu'il existe des solutions pour des valeurs de c suffisamment grandes. Pour aller plus loin et obtenir des constantes explicites, il faudrait obtenir une borne inférieure sur les nombres d'occupation des minimiseurs de Hartree-Fock multiconfiguration, c'est-à-dire une version quantitative des travaux de Friesecke [Fri03a]. On se heurte ici encore au manque d'informations sur la structure locale de la fonctionnelle de Hartree-Fock au voisinage de ses minimiseurs.

1.4 Deuxième partie : inégalités de Lieb-Thirring

1.4.1 L'opérateur $-\Delta + V$

L'opérateur $H = -\Delta + V$ est omniprésent en mécanique quantique, où il représente l'Hamiltonien d'une particule non-relativiste dans un champ de potentiel V. Le cas qui nous intéresse ici est celui d'un potentiel V négatif (donc attractif), et localisé (au contraire de par exemple un potentiel V périodique, utilisé dans la modélisation des cristaux).

Par analogie avec les potentiels connus en mécanique quantique (potentiel Coulombien, barrière de potentiel ...), on s'attend à ce que le spectre d'un tel opérateur soit composé d'un spectre essentiel $[0, +\infty[$, et de valeurs propres négatives, isolées et de multiplicité finie. Les valeurs propres négatives correspondent aux états liés, et le spectre continu positif correspond aux états de diffusion.

Mathématiquement, la situation est loin d'être aussi claire, et dépend fortement des caractéristiques de régularité et de décroissance à l'infini de V. On peut cependant définir pour une classe assez large de potentiels V les valeurs propres négatives λ_i par un principe de min-max (voir par exemple le chapitre 12 de [LL01]).

Le nombre et la magnitude de ces valeurs propres négatives dépend de la capacité des fonctions d'ondes normalisées ψ à rendre l'énergie

$$E(\psi) = \int_{\mathbb{R}^d} |\nabla\psi|^2 + V\psi^2$$

négative. En d'autres termes, les fonctions ψ doivent se localiser (pour rendre $\int V\psi^2$ le plus négatif possible), sans trop augmenter leur énergie cinétique $\int|\nabla\psi|^2$. Plus le potentiel V sera négatif (pour diminuer $\int V\psi^2$ à ψ fixé) et étendu (pour pouvoir supporter le maximum de fonctions d'ondes orthogonales ψ), plus les valeurs propres seront négatives et nombreuses.

Une façon de rendre ce raisonnement rigoureux mathématiquement est d'essayer d'estimer en fonction de V les moments des valeurs propres négatives, c'est-à-dire

$$\sum_{i,\lambda_i<0} |\lambda_i|^\gamma = \mathrm{tr}((-\Delta + V)_-^\gamma), \tag{1.54}$$

où l'opérateur $(-\Delta + V)_-^\gamma$ est défini par le calcul fonctionnel dès que $-\Delta + V$ est auto-adjoint [RS80].

Il existe un régime où cette estimation peut être faite facilement : c'est la limite semi-classique.

1.4.2 La limite semi-classique

Pour simplifier les calculs, on se restreint au cas où $d = 1$. L'approximation semi-classique d'un système quantique revient à supposer que tous les états ψ sont localisés dans l'espace des phases (x, p), où ils occupent le volume minimal permis par le principe d'incertitude d'Heisenberg, c'est-à-dire $1/(2\pi)$. Ainsi, on passe d'un modèle continu à un modèle discret où les états possibles du système quantique sont

les $\psi_{m,n}$, de position et de quantité de mouvement

$$x_m \in \left[\frac{1}{\sqrt{2\pi}}m, \frac{1}{\sqrt{2\pi}}(m+1)\right],$$
$$p_n \in \left[\frac{1}{\sqrt{2\pi}}n, \frac{1}{\sqrt{2\pi}}(n+1)\right].$$

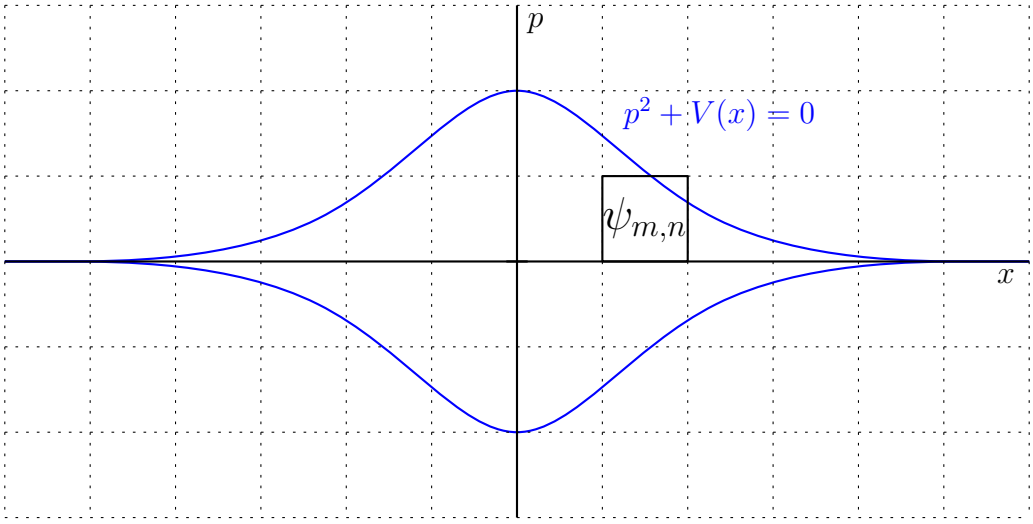

FIGURE 1.2: L'approximation semi-classique

Bien que cette localisation exacte soit impossible (une fonction à support compact ne peut pas avoir une transformée de Fourier à support compact), elle est utile pour calculer approximativement les états propres associés à une valeur propre négative de l'opérateur $-\Delta + V$: ce sont ceux pour lesquels $p_n^2 + V(x_m) < 0$, ce qui définit une région dans l'espace des phases (x, p) (voir Figure 1.2). Les erreurs sur la position et la quantité de mouvement, liés au principe d'incertitude de Heisenberg, empêchent de savoir précisément si un état $\psi_{m,n}$ appartient à cette région. Cependant, dans la limite où V est très grand (ou très étendu ; par scaling, c'est en fait la même chose), la discrétisation devient négligeable, et on peut considérer heuristiquement que

$$\begin{aligned}\sum_i |\lambda_i|^\gamma &\approx \sum_{m,n} (p_n^2 + V(x_m))_-^\gamma \\ &\approx \frac{1}{2\pi}\int_x \int_p (p^2 + V(x))_-^\gamma \\ &= \frac{1}{2\pi}\int_x \int_{|p|\leq\sqrt{-V(x)}} \left(-p^2 - V(x)\right)^\gamma \\ &= \frac{1}{2\pi}\int_{|p|\leq 1}(1-p^2)^\gamma \int_x |V|^{\gamma+1/2}.\end{aligned}$$

Le calcul se mène de même en dimension d quelconque, et donne

$$\sum_i |\lambda_i|^\gamma \approx L^{\mathrm{sc}}_{\gamma,d}\int |V|^{\gamma+d/2}, \tag{1.55}$$

où

$$L^{\mathrm{sc}}_{\gamma,d} = \frac{1}{(2\pi)^d}\int_{|p|\leq 1}(1-p^2)^\gamma = \frac{\Gamma(\gamma+1)}{(4\pi)^{d/2}\Gamma(\gamma+1+d/2)}.$$

Cette heuristique peut être formalisée dans la limite où V est infiniment grand. On peut alors prouver grâce à la notion d'état cohérent (voir par exemple [LL01]) que, pour V suffisamment régulier,

$$\lim_{\mu\to\infty}\frac{\mathrm{tr}((-\Delta+\mu V)^\gamma_-)}{\int(\mu V)^{\gamma+\frac{d}{2}}} = L^{\mathrm{sc}}_{\gamma,d}. \tag{1.56}$$

1.4.3 Inégalités de Lieb-Thirring

Dans le régime semi-classique, on a vu que le moment des valeurs propres peut être directement calculé en fonction de V. En dehors de ce régime, le calcul n'est plus valable. Néanmoins, les inégalités de Lieb-Thirring montrent qu'une forme de cette heuristique survit sous la forme d'une majoration du moment des valeurs propres.

L'inégalité de Lieb-Thirring en dimension d pour l'exposant $\gamma \geq 0$ s'écrit : il existe une constante $L_{\gamma,d} > 0$ telle que, pour tout potentiel $V \in L^{\gamma+d/2}(\mathbb{R}^d,\mathbb{R})$,

$$\mathrm{tr}((-\Delta+V)^\gamma_-) \leq L_{\gamma,d}\int V_-^{\gamma+d/2}. \tag{1.57}$$

En dehors de leur intérêt propre, l'application principale des inégalités de Lieb-Thirring concerne le cas $\gamma = 1$, où elles permettent de lier l'énergie cinétique d'une fonction d'onde fermionique à sa densité. Ce lien est une étape importante dans la preuve de la stabilité de la matière en mécanique quantique, c'est-à-dire l'asymptotique à N grand de l'état fondamental d'un système de N fermions (voir [Lie90] pour une présentation générale). C'est d'ailleurs dans ce but que les inégalités ont été introduites en 1975 par Lieb et Thirring [LT75]. Détaillons à titre d'illustration cette application.

Soit $\psi \in L^2_a(\mathbb{R}^d)$ une fonction d'onde fermionique à N corps normalisée, et V un potentiel négatif quelconque. Posons

$$H^{(N)} = \sum_{i=1}^N -\Delta_{xi} + V(x_i)$$

l'Hamiltonien à N corps formé à partir de V. Alors, en notant E_0 l'état fondamental fermionique de H^N,

$$E_0 \leq \left\langle \psi, H^N\psi\right\rangle = K_\psi + \int V\rho,$$

où $\rho(x) = N\int_{x_2,\dots,x_N}\psi^2(x,x_2,\dots,x_N)$ est la densité associée à ψ, et

$$K_\psi = \sum_{i=1}^N \int_{\mathbb{R}^{3N}}\left|\nabla_{x_i}\psi\right|^2 \tag{1.58}$$

son énergie cinétique.

Comme $H^{(n)}$ ne couple pas les variables x_i entre elles, E_0 se calcule facilement : en notant λ_i les valeurs propres négatives de $-\Delta + V$ rangées par ordre croissant (éventuellement complétées par 0 si il n'y en a pas N),

$$\begin{aligned} E_0 &= \sum_{i=1}^{N} \lambda_i \\ &\geq \sum_{i=1}^{\infty} \lambda_i \\ &\geq -L_{1,d} \int (-V)^{1+d/2}. \end{aligned}$$

On a donc

$$K_\psi \geq -L_{1,d} \int (-V)^{1+d/2} - \int V\rho.$$

Cette inégalité est valide pour tout V. Pour obtenir un résultat qui ne dépend que de ρ, on peut choisir $V = -C\rho^{2/d}$, et obtenir

$$K_\psi \geq (C - L_{1,d}\, C^{1+d/2}) \int \rho^{1+2/d}.$$

On maximise maintenant cette inégalité par rapport à C et on obtient, pour $C = (L(1+d/2))^{-2/d}$,

$$K_\psi \geq L_d^* \int \rho^{1+2/d}, \tag{1.59}$$

avec la constante

$$L_d^* = \frac{d}{2} L_{1,d}^{-2/d} (1+d/2)^{-2/d-1}.$$

Cette inégalité a un sens physique immédiat : une densité ρ ne peut provenir que d'une fonction d'onde ψ qui a une énergie cinétique au moins égale à $L_d^* \int \rho^{1+2/d}$. C'est une forme de principe d'incertitude qui permet de mieux comprendre physiquement les modèles qui utilisent la densité électronique comme variable principale, comme la théorie de la fonctionnelle de la densité. Elle permet également, par réduction au modèle de Thomas-Fermi, de prouver la stabilité de la matière, ce qui était la motivation originale des inégalités de Lieb-Thirring.

1.4.4 Constantes optimales

L'inégalité de Lieb-Thirring (1.57) n'est pas valable pour tout (γ, d). Elle ne peut évidemment pas être valable pour $\gamma < 0$. En une dimension, l'opérateur $-\Delta + \mu V$, avec μ petit, peut posséder une valeur propre de l'ordre de μ^2 : l'inégalité ne peut donc être valide que si $2\gamma \geq \gamma + 1/2$, c'est-à-dire $\gamma \geq 1/2$. En deux dimensions, il est bien connu en mécanique quantique qu'un potentiel quelconque, arbitrairement petit, a au moins une valeur propre négative, ce qui montre que l'inégalité ne peut être valide que pour $\gamma > 0$.

L'article original de Lieb et Thirring [LT76] en 1976 prouvait l'inégalité dans le cas $\gamma > 1/2$ en une dimension, et $\gamma > 0$ pour $d \geq 2$. Le cas $\gamma = 0, d \geq 3$ nécessite des méthodes complètement différentes, et il a été prouvé indépendamment par Cwikel, Lieb et Rozenblum [Cwi77; Lie76; Roz72]. Il a ensuite fallu attendre les travaux de Weidl en 1996 [Wei96] pour obtenir le cas limite $\gamma = 1/2, d = 1$.

L'inégalité est donc valide pour $\gamma \geq 1/2$ en une dimension, $\gamma > 0$ en deux dimensions, et $\gamma \geq 0$ en dimension supérieure. Cependant, les constantes $L_{\gamma,d}$ obtenues dans les preuves sont substantiellement supérieures aux constantes semi-classiques $L^{\mathrm{sc}}_{\gamma,d}$. La question se pose alors de la valeur des constantes optimales, c'est-à-dire les plus petites constantes $L_{\gamma,d}$ pour lesquelles l'inégalité de Lieb-Thirring est valide.

Notons

$$R_{\gamma,d} = \frac{L_{\gamma,d}}{L^{\mathrm{sc}}_{\gamma,d}}. \tag{1.60}$$

Par le résultat dans la limite semi-classique (1.56), $R_{\gamma,d} \geq 1$. Il est intéressant de savoir pour quels γ, d on a $R_{\gamma,d} = 1$ (l'asymptotique semi-classique est une borne exacte) ou $R_{\gamma,d} > 1$ (l'asymptotique semi-classique doit être modifiée pour donner une borne). Un résultat de scaling d'Aizenman et Lieb [AL78] montre que $R_{\gamma,d}$ est décroissant avec γ. Ainsi, si $R_{\gamma_0,d} = 1$ pour un certain γ_0, on sait que $R_{\gamma,d} = 1$ pour tout $\gamma \geq \gamma_0$.

Beaucoup d'études théoriques essayent d'obtenir des constantes $L_{\gamma,d}$ les plus proches possibles des constantes semi-classiques. Néanmoins, les seuls résultats optimaux ($R_{\gamma,d} = 1$) sont obtenus pour $\gamma \geq 3/2$, d quelconque [LW00b].

Pour comprendre les valeurs des constantes optimales, une autre approche est d'obtenir des bornes inférieures sur $L_{\gamma,d}$ en exhibant des potentiels V particuliers. Dans une annexe de l'article original de Lieb et Thirring [LT76], Barnes a étudié numériquement le problème restreint aux potentiels tels que $-\Delta + V$ n'a qu'une seule valeur propre négative. Dans ce cas, si ϕ est un vecteur propre associé à la valeur propre λ qui réalise l'égalité $(-\lambda)^\gamma = L_{\gamma,d} \int V^{\gamma+d/2}$, avec $L_{\gamma,d}$ la constante optimale, alors la condition d'optimalité du premier ordre donne une relation entre V et ϕ qui conduit à l'équation différentielle

$$-\Delta\phi + K\phi^{2/(\gamma+d/2-1)}\phi = \lambda\phi.$$

Barnes a résolu numériquement cette équation dans le cas radial par une méthode de tir. On note $R^1_{\gamma,d}$ les bornes inférieures des constantes optimales obtenues par cette méthode. Elles sont plus grandes que 1 pour $\gamma < 3/2$ en dimension 1, $\gamma \lesssim 1.165$ en dimension 2, et $\gamma \lesssim 0.863$ en dimension 3.

Helffer et Robert [HR90] ont étudié le cas particulier où $V = \mu(1 - |x|^2)$. Dans le développement asymptotique de $L_{\gamma,d}$ pour μ grand, le premier terme est la limite semi-classique. En étudiant le signe du second terme, ils ont prouvé que $R_{\gamma,d} > 1$ quand $\gamma < 1$, indépendamment de la dimension.

Tous ces résultats montrent qu'il existe un γ^d_c critique tel que $R_{\gamma,d} = 1$ pour $\gamma \geq \gamma^d_c$, et $R_{\gamma,d} > 1$ pour $\gamma < \gamma^d_c$. Ce γ^d_c est compris entre 1 et 3/2. On ne connaît sa valeur exacte qu'en une dimension, où il vaut 3/2.

Pour résumer :

- En une dimension, $R_{\gamma,1}$ est infini pour $\gamma \leq 1/2$. Pour $\gamma > 1/2$, il est fini et supérieur à $R^1_{\gamma,1}$ qui croise 1 en $\gamma = 3/2$. Pour $\gamma \geq 3/2$, $R_{\gamma,1} = 1$.

- En deux dimensions, $R_{\gamma,2}$ est infini pour $\gamma = 0$. Pour $\gamma > 0$, il est fini et supérieur à $R^1_{\gamma,2}$, qui croise 1 en $\gamma \approx 1.165$. Il vaut 1 pour $\gamma \geq 3/2$.

- En trois dimensions, pour $\gamma \geq 0$, $R_{\gamma,3}$ est fini et supérieur à $R^1_{\gamma,3}$, qui croise 1 en $\gamma \approx 0.863$. Il reste strictement supérieur à 1 jusqu'à $\gamma = 1$, et vaut 1 pour $\gamma \geq 3/2$.

- Pour les dimensions supérieures, la situation est semblable à la dimension 3, mis à part que le point d'intersection de $R^1_{\gamma,d}$ avec 1 intervient pour des γ de plus en plus faibles.

1.4.5 Résultats du chapitre 4

Dans le chapitre 4 de cette thèse, on étudie numériquement les valeurs des meilleures constantes pour l'inégalité de Lieb-Thirring. Les résultats présentés ont fait l'objet d'un article accepté pour publication dans le Journal of Spectral Theory [Lev12a].

Notre but est d'obtenir des potentiels V qui maximisent le rapport

$$\frac{\operatorname{tr}((-\Delta + V)_-^\gamma)}{\int V_-^{\gamma+d/2}}.$$

Sans perte de généralité, on peut choisir V négatif et normalisé :

$$\begin{aligned} L_{\gamma,d} &= \sup_{V \in L^{\gamma+d/2}} \frac{\operatorname{tr}((-\Delta + V)_-^\gamma)}{\int V_-^{\gamma+d/2}} \\ &= \sup \Big\{ \operatorname{tr}((-\Delta + V)_-^\gamma) \;/\; V \in L^{\gamma+d/2}, \int V^{\gamma+d/2} = 1, V \leq 0 \Big\}. \end{aligned} \tag{1.61}$$

On se propose de résoudre ce problème de maximisation numériquement. Une maximisation globale est évidemment hors de portée d'une application numérique : on se contente de maximiser localement cette quantité, à l'aide d'un algorithme qui produit une suite V_n de potentiels à partir d'un point initial V_0. Ensuite, on varie la condition initiale V_0 pour parcourir différentes régions du paysage d'énergie.

Pour décrire l'algorithme, on utilise une généralisation de la notion de matrice densité. Soit $\mathcal{S}_p$ l'ensemble des opérateurs τ sur $L^2(\mathbb{R}^d)$ symétriques bornés de norme de Schatten

$$\|\tau\|_p = \operatorname{tr}(|\tau|^p)^{1/p} \tag{1.62}$$

finie. $\mathcal{S}_1$ est la classe des opérateurs symétriques à trace, $\mathcal{S}_2$ la classe des opérateurs symétriques de Hilbert-Schmidt, et $\mathcal{S}_\infty$ la classe des opérateurs symétriques bornés.

Pour $\gamma = 1$,

$$\operatorname{tr}((-\Delta + V)_-) = -\operatorname{tr}((-\Delta + V)\,\chi_{\mathbb{R}^-}(-\Delta + V)), \tag{1.63}$$

où $\chi_{\mathbb{R}^-}(-\Delta+V)$ est le projecteur sur le spectre négatif de $-\Delta+V$. Il est facile de voir que ce projecteur est même l'élément de $\{\tau \in \mathcal{S}_\infty, \tau \geq 0, \|\tau\|_\infty = 1\}$ qui minimise la quantité $\mathrm{tr}((-\Delta+V)\tau)$. Par l'inégalité de Hölder sur $\mathcal{S}_p$, on peut généraliser cette propriété à tout $\gamma \geq 1$:

Proposition 1.1. *Pour tout $V \in L^{\gamma+\frac{d}{2}}(\mathbb{R}^d, \mathbb{R})$ et $\gamma \geq 1$,*

$$[\mathrm{tr}((-\Delta+V)_-^\gamma)]^{1/\gamma} = \max\Big\{ -\mathrm{tr}((-\Delta+V)\tau) \;/\; \tau \in \mathcal{S}_{\gamma'}, \tau \geq 0, \|\tau\|_{\gamma'} = 1\Big\}, \quad (1.64)$$

où γ' est le conjugué de Hölder de γ, tel que $\frac{1}{\gamma}+\frac{1}{\gamma'}=1$.

On a donc

$$\begin{aligned} L^{1/\gamma}_{\gamma,d} = \sup\Big\{ -\mathrm{tr}((-\Delta+V)\tau) \;/\; \tau \in \mathcal{S}_{\gamma'}, \|\tau\|_{\gamma'} = 1, \tau \geq 0, \\ V \in L^{\gamma+d/2}, \int V^{\gamma+d/2} = 1, V \leq 0\Big\}. \end{aligned} \quad (1.65)$$

L'intérêt de cette formulation est qu'on peut maximiser explicitement la fonctionnelle par rapport à V et τ indépendamment. En généralisant (1.63), on voit qu'à V fixé, l'optimum est atteint pour

$$\tau = K_\tau(-\Delta+V)_-^{\gamma-1}, \quad (1.66)$$

où la constante K_τ est choisie pour assurer la normalisation $\|\tau\|_{\gamma'} = 1$. De plus, pour τ fixé, le problème se ramène à minimiser $\int V(x)\tau(x,x)\mathrm{d}x$ pour $\|V\|_{\gamma+d/2} = 1$. La solution est un cas d'égalité dans l'inégalité de Hölder, et donc

$$V(x) = -K_V \tau(x,x)^{\frac{1}{\gamma+\frac{d}{2}-1}}, \quad (1.67)$$

où K_V est une constante choisie pour assurer la normalisation $\|V\|_{\gamma+d/2} = 1$.

On utilise donc numériquement l'algorithme suivant : pour un V_n fixé, on choisit τ_n par (1.66), puis V_{n+1} par (1.67). C'est un algorithme de champ auto-cohérent similaire à l'algorithme de Roothaan pour Hartree-Fock. Cependant, comme il revient à maximiser (1.65) alternativement par rapport à chaque variable, on a la certitude que $-\mathrm{tr}((\Delta+V_n)\tau_n)$ est croissant avec n : on évite ainsi les oscillations de l'algorithme de Roothaan.

L'algorithme est formulé avec $\gamma \geq 1$, car c'est dans ce cas qu'on a une théorie de dualité Hölderienne. Cependant, les formules (1.66) et (1.67) peuvent être appliqués pour tout $\gamma > 0$. Il n'y a par contre plus de garanties que le schéma soit monotone. Néanmoins, on a observé numériquement une convergence satisfaisante, et on utilise ce schéma pour tout γ.

Reste à implémenter ce schéma en pratique. Pour simplifier, on considère principalement le cas radial. Si V_0 est radial, les itérés suivants le sont également, ce qui permet de formuler tous les calculs en une dimension. Pour résoudre le problème radial, on choisit un espace d'éléments finis pour les potentiels V et les fonctions propres de $-\Delta+V$. Cela permet de réduire l'algorithme à un calcul de valeurs propres de matrice. On se rapportera au chapitre 4 pour les détails.

Ce procédé de discrétisation par éléments finis a un grand avantage. Étant donné que les valeurs propres de $-\Delta+V$ sont calculées sur un sous-espace, elles sont systématiquement surévaluées. Ainsi, pour un V donné, on peut calculer, à la précision

machine près, une borne inférieure de $-\mathrm{tr}(-\Delta + V)$. Cela nous assure de pouvoir obtenir des bornes inférieures sur $L_{\gamma,d}$.

Les résultats principaux, détaillés dans le chapitre 4, sont les suivants :

Dimension 1 En initialisant l'algorithme avec une gaussienne de largeur appropriée, on retrouve les potentiels à un seul état propre V^1 calculés par Barnes. Si on initialise l'algorithme avec des conditions initiales sensiblement plus étendues que V^1, ou présentant plusieurs minimums, le potentiel se sépare au fur et à mesure des itérations en plusieurs bosses correspondant à V^1, comme on peut le voir figure 1.3. Cela semble indiquer que V^1 est bien le potentiel maximisant, et que $R_{\gamma,d} = R^1_{\gamma,d}$ pour $\gamma \leq 3/2$.

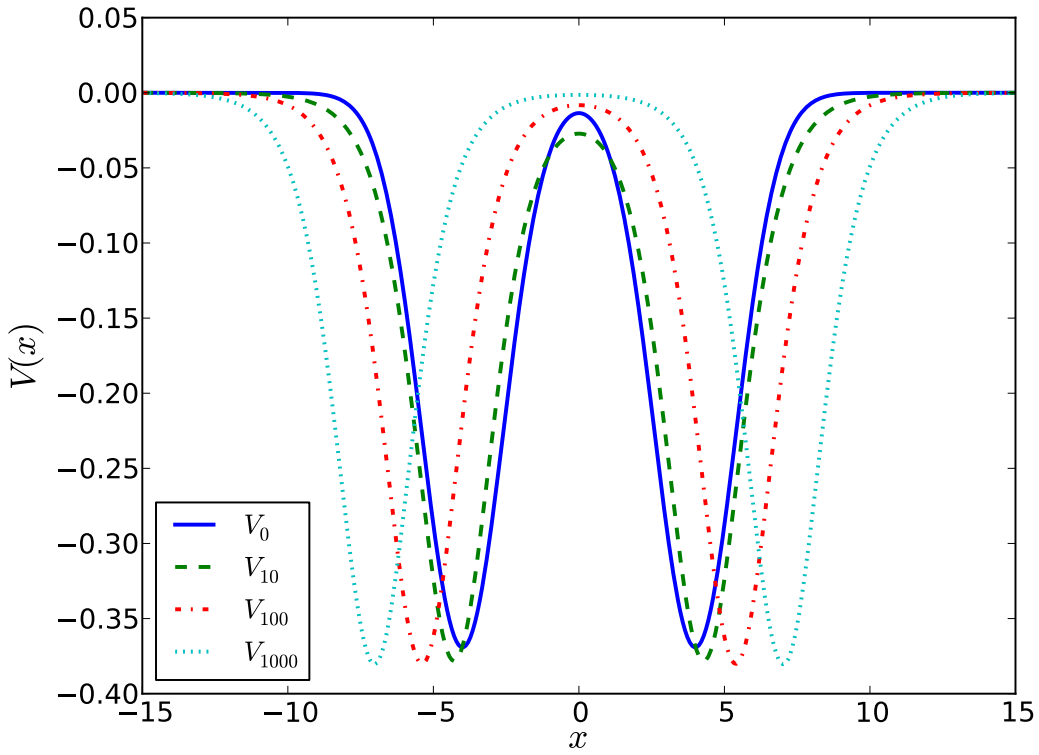

FIGURE 1.3: En une dimension, un potentiel se sépare au cours des itérations en plusieurs bosses qui s'éloignent

Dimension 2 De même qu'en dimension 1, on retrouve les potentiels à un état propre V^1. En plus de ceux-ci, on trouve également des points stationnaires de l'algorithme possédant plus d'un état propre. Cependant, ces potentiels ont une énergie plus faible que V^1, de sorte qu'il semble que $R_{\gamma,d} = R^1_{\gamma,d}$ jusqu'à $\gamma \approx 1.165$, où $R_{\gamma,d} = 1$. Il est à noter que les points stationnaires trouvés, qui semblent être des minimums locaux dans le cadre radial, sont instables par des perturbations non-radiales, comme on peut le voir figure 1.4.

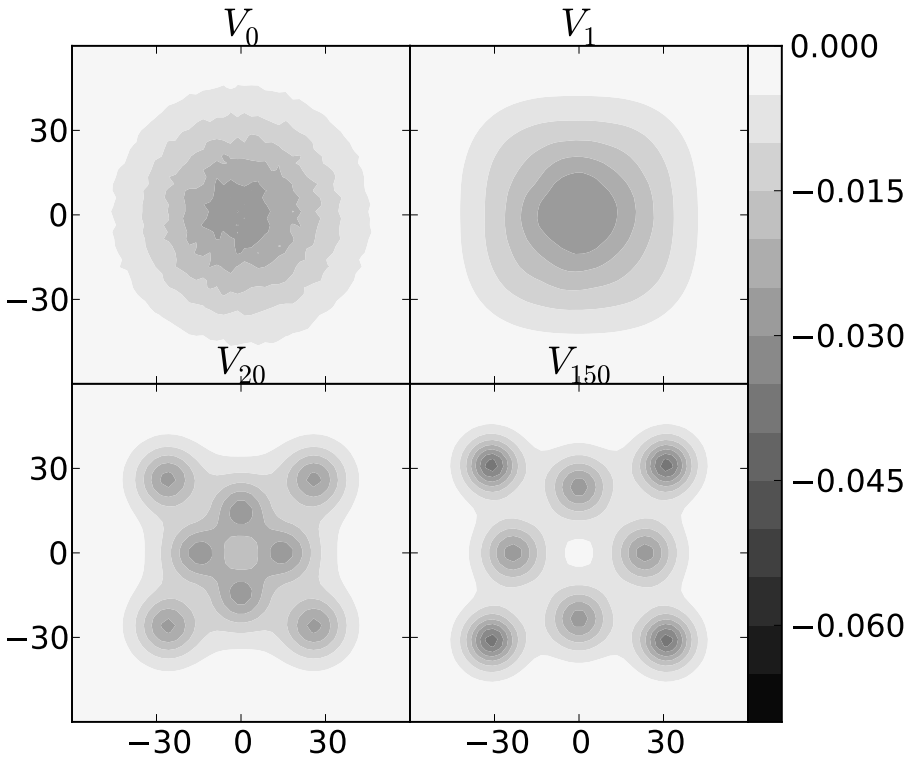

FIGURE 1.4: En deux dimensions, on initialise l'algorithme avec un point critique trouvé dans le cadre radial, qu'on perturbe aléatoirement. Il se sépare en huit copies du potentiel V^1 calculé par Barnes.

Dimension 3 En dimension 3, comme en dimension 2, on trouve des potentiels ayant plus d'un état propre. Pour certaines valeurs des paramètres, ces potentiels ont une énergie supérieure à $R^1_{\gamma,d}$. Par exemple, à partir de $\gamma \approx 0.85$, on trouve un potentiel à 5 états propres qui a une énergie supérieure à V^1. Puis, vers $\gamma \approx 0.87$, c'est un potentiel à 14 états propres qui dépasse le potentiel à cinq états propres, et ainsi de suite (voir figure 1.5).

Ces résultats vont dans le sens du résultat analytique de Helffer et Robert, qui prouve que pour $\gamma < 1$, $R_{\gamma,d} > 1$. Ils semblent confirmer la conjecture que $R_{1,3} = 1$.

Dimension $d \geq 4$ Pour les dimensions plus grandes, le problème aux valeurs propres devient de plus en plus mal conditionné, nécessitant plus de calculs. Les résultats partiels obtenus indiquent que la situation est semblable à la dimension 3, avec une suite de potentiels de plus en plus étendus qui deviennent successivement maximisants.

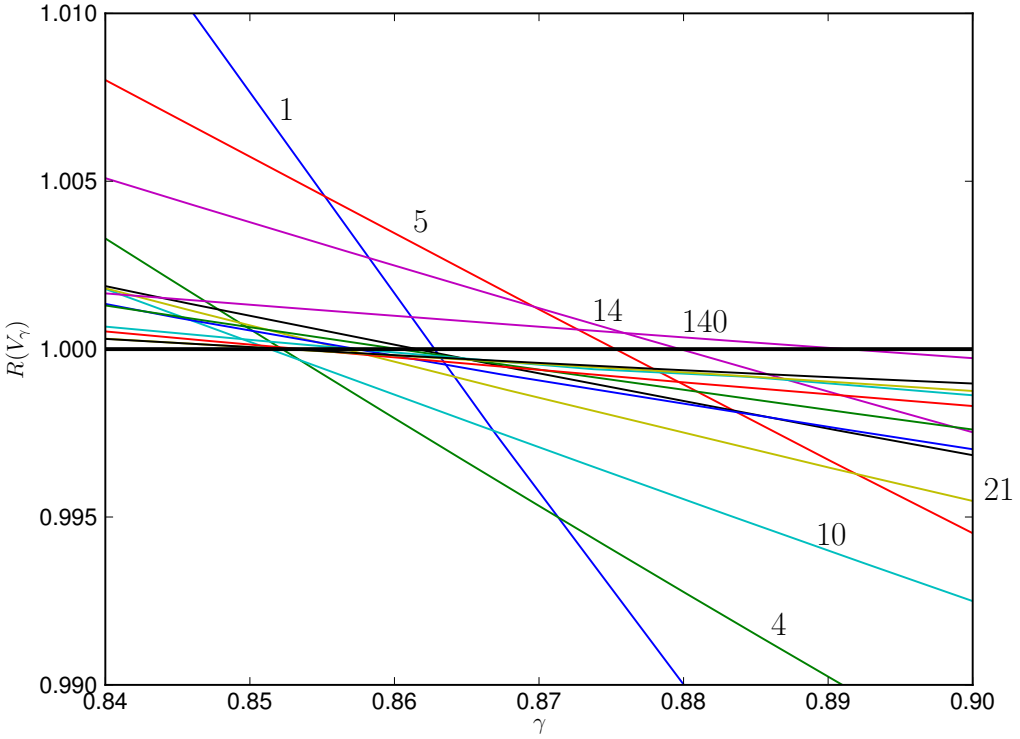

FIGURE 1.5: En trois dimensions, valeurs de R pour certains potentiels radiaux calculés numériquement. On numérote les potentiels par le nombre de valeurs propres négatives de $-\Delta + V$. Le graphique est agrandi autour de $\gamma \approx 0.86$, le point où le potentiel V^1 passe en-dessous du seuil critique $R = 1$.

Bibliographie

[AL78] M. Aizenman and E.H. Lieb. On semi-classical bounds for eigenvalues of Schrödinger operators. *Phys. Lett. A* 66.6 (1978), pp. 427–429.

[BP87] J.M. Borwein and D. Preiss. A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions. *Trans. Amer. Math. Soc* 303.51 (1987), pp. 7–527.

[CLB00a] E. Cancès and C. Le Bris. Can we outperform the DIIS approach for electronic structure calculations ? *International Journal of Quantum Chemistry* 79.2 (2000), pp. 82–90.

[CLB00b] E. Cancès and C. Le Bris. On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 34.4 (2000), pp. 749–774.

[CLBM06] E. Cancès, C. Le Bris, and Y. Maday. *Méthodes mathématiques en chimie quantique. Une introduction*. Vol. 53. Springer, 2006.

[Can98] E Cancès. "Molecular Simulation and Environmental Effects : A Mathematical and Numerical Perspective". PhD thesis. Ecole des Ponts ParisTech, 1998.

[CTDL73] C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Mécanique quantique*. 1973.

[Cwi77] M. Cwikel. Weak type estimates for singular values and the number of bound states of Schrodinger operators. *Ann. of Math.* (1977), pp. 93–100.

[Der12] J. Derezinski. Open problems about many-body Dirac operators. *Bulletin of International Association of Mathematical Physics* (2012).

[Dir28] P.A.M. Dirac. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 117.778 (1928), pp. 610–624.

[Eke74] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications* 47.2 (1974), pp. 324–353.

[ES99] M.J. Esteban and E. Séré. Solutions of the Dirac-Fock equations for atoms and molecules. *Communications in Mathematical Physics* 203.3 (1999), pp. 499–530.

[ES01] M.J. Esteban and E. Séré. Nonrelativistic limit of the Dirac-Fock equations. *Annales Henri Poincare* 2 (5 2001), pp. 941–961.

[FG92] G. Fang and N. Ghoussoub. Second order information on Palais-Smale sequences in the mountain pass theorem. *Manuscripta mathematica* 75.1 (1992), pp. 81–95.

[Fey65] R.P. Feynman. *Feynman lectures on physics. Volume 3 : Quantum mechanics*. 1965.

[Fri03a] G. Friesecke. On the infinitude of non-zero eigenvalues of the single-electron density matrix for atoms and molecules. *Proceedings of the Royal Society of London. Series A : Mathematical, Physical and Engineering Sciences* 459.2029 (2003), pp. 47–52.

[Fri03b] G. Friesecke. The multiconfiguration equations for atoms and molecules : charge quantization and existence of solutions. *Archive for Rational Mechanics and Analysis* 169.1 (2003), pp. 35–71.

[GH12] M. Griesemer and F. Hantsch. Unique Solutions to Hartree–Fock equations for closed-shell atoms. *Archive for Rational Mechanics and Analysis* 203 (3 2012), pp. 883–900.

[HR90] B. Helffer and D. Robert. Riesz means of bounded states and semi-classical limit connected with a Lieb-Thirring conjecture II. *Ann. Inst. Henri Poincaré (A)* 53.2 (1990), pp. 139–147.

[LW00b] A. Laptev and T. Weidl. Sharp Lieb-Thirring inequalities in high dimensions. *Acta Math.* 184.1 (2000), pp. 87–111.

[LB94] C. Le Bris. A general approach for multiconfiguration methods in quantum molecular chemistry. *Annales de l'Institut Henri Poincaré. Analyse non linéaire* 11.4 (1994), pp. 441–484.

[LBL05] C. Le Bris and P.L. Lions. From atoms to crystals : a mathematical journey. *Bull. Amer. Math. Soc* 42 (2005), pp. 291–363.

[Lev12a] A. Levitt. Best constants in Lieb-Thirring inequalities : a numerical investigation. Accepted for publication in Journal of Spectral Theory. 2012.

[Lev12b] A. Levitt. Convergence of gradient-based algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 46.06 (2012), pp. 1321–1336.

[Lev13] A. Levitt. Solutions of the multiconfiguration Dirac-Fock equations. Submitted. 2013.

[Lew04] M. Lewin. Solutions of the multiconfiguration equations in quantum chemistry. *Archive for Rational Mechanics and Analysis* 171.1 (2004), pp. 83–114.

[Lie76] E.H. Lieb. The number of bound states of one-body Schrödinger operators and the Weyl problem. *Bull. Amer. Math. Soc.* 82 (1976), pp. 751–753.

[Lie90] E.H. Lieb. The stability of matter : from atoms to stars. *Bull. Amer. Math. Soc.* 22.1 (1990).

[LL01] E.H. Lieb and M. Loss. Analysis. *American Mathematical Society, Providence, RI,* 4 (2001).

[LS77] E.H. Lieb and B. Simon. The Hartree-Fock theory for Coulomb systems. *Comm. Math. Phys.* 53.3 (1977), pp. 185–194.

[LT75] E.H. Lieb and W.E. Thirring. Bound for the kinetic energy of fermions which proves the stability of matter. *Phys. Rev. Lett.* 35 (1975), pp. 687–689.

[LT76] E.H. Lieb and W.E. Thirring. Inequalities for the moments of the eigenvalues of the Schrodinger Hamiltonian and their relation to Sobolev inequalities. *Studies in Math. Phys., Essays in Honor of Valentine Bargmann* (1976).

[Lio87] P.L. Lions. Solutions of Hartree-Fock equations for Coulomb systems. *Communications in Mathematical Physics* 109.1 (1987), pp. 33–97.

[Łoj65] S. Łojasiewicz. *Ensembles semi-analytiques*. Institut des Hautes Etudes Scientifiques, 1965.

[Pul82] P. Pulay. Improved SCF convergence acceleration. *Journal of Computational Chemistry* 3.4 (1982), pp. 556–560.

[PD79] P. Pyykko and J.P. Desclaux. Relativity and the periodic system of elements. *Accounts of Chemical Research* 12.8 (1979), pp. 276–281.

[RS80] M. Reed and B. Simon. *Methods of modern mathematical physics : Functional analysis*. Vol. 1. 1980.

[RS11] T. Rohwedder and R. Schneider. An analysis for the DIIS acceleration method used in quantum chemistry calculations. *Journal of mathematical chemistry* 49.9 (2011), pp. 1889–1914.

[Roz72] G.V. Rozenblum. Distribution of the discrete spectrum of singular differential operators. *Soviet Math. Doki.* 202 (1972), pp. 1012–1015.

[SO89] A. Szabo and N.S. Ostlund. *Modern quantum chemistry*. McGraw-Hill New York, 1989.

[Tha92] B. Thaller. *The Dirac equation*. Springer-Verlag, 1992.

[Wei96] T. Weidl. On the Lieb-Thirring constants $L_{\gamma,1}$ for $\gamma \geq 1/2$. *Comm. Math. Phys.* 178.1 (1996), pp. 135–146.

Chapitre 2

Convergence d'algorithmes pour Hartree-Fock

Ce chapitre reprend le texte intégral de l'article "Convergence of gradient-based algorithms for the Hartree-Fock equations", paru dans ESAIM : Mathematical Modelling and Numerical Analysis, volume 46, numéro 6, pages 1321-1336, 2012.

Convergence of gradient-based algorithms for the Hartree-Fock equations

Abstract

The numerical solution of the Hartree-Fock equations is a central problem in quantum chemistry for which numerous algorithms exist. Attempts to justify these algorithms mathematically have been made, notably in [CLB00b], but, to our knowledge, no complete convergence proof has been published, except for the large-Z result of [GH12]. In this paper, we prove the convergence of a natural gradient algorithm, using a gradient inequality for analytic functionals due to Łojasiewicz [Łoj65]. Then, expanding upon the analysis of [CLB00b], we prove convergence results for the Roothaan and Level-Shifting algorithms. In each case, our method of proof provides estimates on the convergence rate. We compare these with numerical results for the algorithms studied.

2.1 Introduction

In quantum chemistry, the Hartree-Fock method is one of the simplest approximations of the electronic structure of a molecule. By assuming minimal correlation between the N electrons, it reduces Schrödinger's equation, a linear partial differential equation on $\mathbb{R}^{3N}$, to the Hartree-Fock equations, a system of N coupled nonlinear equations on $\mathbb{R}^3$. This approximation makes it much more tractable numerically. It is used both as a standalone description of the molecule and as a starting point for more advanced methods, such as the Møller-Plesset perturbation theory, or multi-configuration methods. Mathematically, the Hartree-Fock method leads to a coupled system of nonlinear integro-differential equations, which are discretized by expanding the solution on a finite Galerkin basis. The resulting nonlinear algebraic equations are then solved iteratively, using a variety of algorithms, the convergence of which is the subject of this work.

The mathematical structure of the Hartree-Fock equations was investigated in the 70's, culminating in the proof of the existence of solutions by Lieb and Simon [LS77], later generalized by Lions [Lio87]. On the other hand, despite their ubiquitous use in computational chemistry, the convergence of the various algorithms used to solve them is still poorly understood. A major step forward in this direction is the recent work of Cancès and Le Bris [CLB00b]. Using the density matrix formulation,

they provided a mathematical explanation for the oscillatory behavior observed in the simplest algorithm, the Roothaan method, and proposed the Optimal Damping Algorithm (ODA), a new algorithm inspired directly by the mathematical structure of the constraint set [CLB00a]. This algorithm was designed to decrease the energy at each step, and linking the energy decrease to the difference of iterates allowed the authors to prove that this algorithm "numerically converges" in the weak sense that $\|D_k - D_{k-1}\| \to 0$. In addition, the algorithm numerically converges towards an Aufbau solution [Can00]. This, to our knowledge, is the strongest convergence result available for an algorithm to solve the Hartree-Fock equations.

However, this is still mathematically unsatisfactory, as it does not guarantee convergence, and merely prohibits fast divergence. The difficulty in proving convergence of the algorithms used to solve the Hartree-Fock equations lies in the lack of understanding of the second-order properties of the energy functional. To our knowledge, the only uniqueness proof is the large-Z limit of [GH12], which proves as a by-product the convergence of the Roothaan algorithm. However, this perturbative proof only works for very negatively charged ions (for $N = 2$, the condition is $Z \geq 35$).

In other domains, the convergence of gradient-based methods has been established using the Łojasiewicz inequality for analytic functionals [Łoj65] (see for instance [Sal07; HJK03]). This method of proof has the advantage of not requiring any second-order information.

In this paper, we use a gradient descent algorithm to solve the Hartree-Fock equations. This algorithm builds upon ideas from differential geometry [EAS98] and the various projected gradient algorithms used in the context of quantum chemistry [CP08; McW56; AA09]. To our knowledge, this particular algorithm has never been applied to the Hartree-Fock equations. Although it lacks the sophistication of modern minimization methods (for instance, see the trust region methods of [FMM04] and [Høs+08]), it is the most natural generalization of the classical gradient descent, and, as such, the simplest one to analyze mathematically. For this algorithm, following the method of [Sal07], we prove convergence, and obtain explicit estimates on the convergence rate. We also apply the method to the widely used Roothaan and Level-Shifting algorithms, effectively linking these fixed-point algorithms to gradient methods.

The structure of this paper is as follows. We first introduce the Hartree-Fock problem in the mathematical setting of density matrices and prove a Łojasiewicz inequality on the constrained parameter space. We then introduce the gradient algorithm, and prove some estimates. We show the convergence and obtain convergence rates for this algorithm, then extend our method to the Roothaan and Level-Shifting algorithm, using an auxiliary energy functional following [CLB00b]. We finally test all these results numerically and compare the convergence of the algorithms.

2.2 Setting

We are concerned with the numerical solution of the Hartree-Fock equations. We will consider for simplicity of notation the spinless Hartree-Fock equations, where

each orbital ϕ_i is a function in $L^2(\mathbb{R}^3, \mathbb{R})$, although our results are easily transposed to other variants such as General Hartree-Fock (GHF) and Restricted Hartree-Fock (RHF).

In this paper, we consider a Galerkin discretization space with finite orthonormal basis $(\chi_i)_{i=1\dots N_b}$. In this setting, the orbitals ϕ_i are expanded on this basis, and the operators we consider are $N_b \times N_b$ matrices. This finite dimension hypothesis is crucial for our results, and we comment on it in the conclusion.

The Hartree-Fock problem consists in minimizing the total energy of a N-body system. We describe the mathematical structure of the energy functional and the minimization set, and introduce a natural gradient descent to solve this problem numerically.

2.2.1 The energy

We consider the quantum N-body problem of N electrons in a potential V (in most applications, V is the Coulombic potential created by a molecule or atom). In the spinless Hartree-Fock model, this problem is simplified by assuming that the N-body wavefunction ψ is a single Slater determinant of N L^2-orthonormal orbitals ϕ_i. A simple calculation then shows that the energy of the wavefunction ψ can be expressed as a function of the orbitals ϕ_i,

$$\mathcal{E}(\psi) = \sum_{i=1}^{N} \int_{\mathbb{R}^3} \frac{1}{2}(\nabla\phi_i)^2 + \int_{\mathbb{R}^3} V\rho + \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho(x)\rho(y) - \tau(x,y)^2}{|x-y|}\mathrm{d}x\mathrm{d}y,$$

where $\tau(x,y) = \sum_{i=1}^{N} \phi_i(x)\phi_i(y)$ and $\rho(x) = \tau(x,x)$.

The energy is then to be minimized over all sets of orthonormal orbitals ϕ_i. An alternative way of looking at this problem is to reformulate it using the density operator D. This operator, defined by its integral kernel $D(x,y) = \tau(x,y)$, can be seen to be the orthogonal projection on the space spanned by the ϕ_i's. The energy can be written as a function of D only:

$$E(D) = \mathrm{Tr}((h + \frac{1}{2}G(D))D), \tag{2.1}$$

where

$$h = -\frac{1}{2}\Delta + V,$$
$$(G(D)\phi)(x) = \left(\rho \star \frac{1}{|\cdot|}\right)(x)\phi(x) - \int_y \frac{\phi(y)\tau(x,y)}{|x-y|}.$$

This time, the energy is to be minimized over all orthogonal projection operators of rank N. In the discrete setting, the orbitals ϕ_j are discretized as $\phi_j = \sum_{i=1}^{N_b} c_{ij}\chi_i$, and the operators D, h, and $G(D)$ become $N_b \times N_b$ matrices.

2.2.2 The manifold $\mathcal{P}$

The Hartree-Fock energy is to be minimized over the set of density matrices

$$D \in \mathcal{P} = \{D \in M_{N_b}(\mathbb{R}), D^T = D, D^2 = D, \mathrm{Tr}\, D = N\}.$$

The manifold $\mathcal{P}$ is equipped with the canonical inner product in $M_{N_b}(\mathbb{R})$

$$\langle A, B\rangle = \mathrm{Tr}(A^T B).$$

We denote by $\|A\| = \sqrt{\langle A, A\rangle}$ the Frobenius (or Hilbert-Schmidt) norm of A, which is the most natural here, and by $\|A\|_{\mathrm{op}} = \max_{\|x\|=1}\|Ax\|$ the operator norm of A.

The Riemannian structure of $\mathcal{P}$ allows us to compute the gradient of E. The tangent space $T_D\mathcal{P}$ at a point D is the set of Δ such that $D+\Delta$ verifies the constraints up to first order in Δ, that is, such that $\Delta^T = \Delta, D\Delta + \Delta D = \Delta, \mathrm{Tr}\,\Delta = 0$. Block-decomposing Δ on the two orthogonal spaces range(D) and ker(D), this implies that the tangent space $T_D\mathcal{P}$ is the set of matrices Δ of the form

$$\Delta = \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix},$$

where A is an arbitrary $(N_b - N) \times N$ matrix.

Hence, the projection on the tangent space of an arbitrary symmetric matrix M is given by

$$\begin{aligned} P_D(M) &= DM(1-D) + (1-D)MD \\ &= [D, [D, M]]. \end{aligned}$$

Note that if M has decomposition $\begin{pmatrix} B & A^T \\ A & C \end{pmatrix}$, then $[D, M] = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}$ and $[D, [D, M]] = \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix}$, which shows that $\|[D[D, M]]\| = \|[D, M]\|$, a property that will be useful in the sequel.

We can now compute the gradient of E. First, the unconstrained gradient in $M_{N_b}(\mathbb{R})$ is

$$\nabla E(D) = F_D = h + G(D),$$

the Fock operator describing the mean field generated by the electrons of D. We obtain the constrained gradient $\nabla_{\mathcal{P}} E$ by projecting onto the tangent space:

$$\begin{aligned} \nabla_{\mathcal{P}} E(D) &= P_D(\nabla E(D)) \\ &= [D, [D, F_D]]. \end{aligned}$$

2.2.3 Łojasiewicz inequality

The Łojasiewicz inequality for a functional f around a critical point x_0 is a local inequality that provides a lower bound on ∇f. Its only hypothesis is analyticity. In particular, no second order information is needed, and the inequality accommodates degenerate critical points.

The classical Łojasiewicz inequality

Theorem 2.1 (Łojasiewicz inequality in $\mathbb{R}^n$)**.** *Let f be an analytic functional from $\mathbb{R}^n$ to $\mathbb{R}$. Then, for each $x_0 \in \mathbb{R}^n$, there is a neighborhood U of x_0 and two constants $\kappa > 0$, $\theta \in (0, 1/2]$ such that when $x \in U$,*

$$|f(x) - f(x_0)|^{1-\theta} \leq \kappa \|\nabla f(x)\| .$$

This inequality is trivial when x_0 is not a critical point. When x_0 is a critical point, a simple Taylor expansion shows that, if the Hessian $\nabla^2 f(x_0)$ is invertible, we can choose $\theta = \frac{1}{2}$ and $\kappa > \frac{1}{\sqrt{2|\lambda_1|}}$, where λ_1 is the eigenvalue of lowest magnitude $\nabla^2 f(x_0)$. When $\nabla^2 f(x_0)$ is singular (meaning that x_0 is a degenerate critical point), the analyticity hypothesis ensures that the derivatives cannot all vanish simultaneously, and that there exists a differentiation order n such that the inequality holds with $\theta = \frac{1}{n}$.

Łojasiewicz inequality on $\mathcal{P}$

We now extend this inequality to functionals defined on the manifold $\mathcal{P}$.

Theorem 2.2 (Łojasiewicz inequality on $\mathcal{P}$)**.** *Let f be an analytic functional from $\mathcal{P}$ to $\mathbb{R}$. Then, for each $D_0 \in \mathcal{P}$, there is a neighborhood U of D_0 and two constants $\kappa > 0$, $\theta \in (0, 1/2]$ such that when $D \in U$,*

$$|f(D) - f(D_0)|^{1-\theta} \leq \kappa \|\nabla_{\mathcal{P}} f(D)\| .$$

Proof. Let $D_0 \in \mathcal{P}$. Define the map R_{D_0} from $T_{D_0}\mathcal{P}$ to $\mathcal{P}$ by

$$\begin{aligned} R_{D_0}(\Delta) &= U D_0 U^T, \\ U &= \exp(-[D_0, \Delta]). \end{aligned}$$

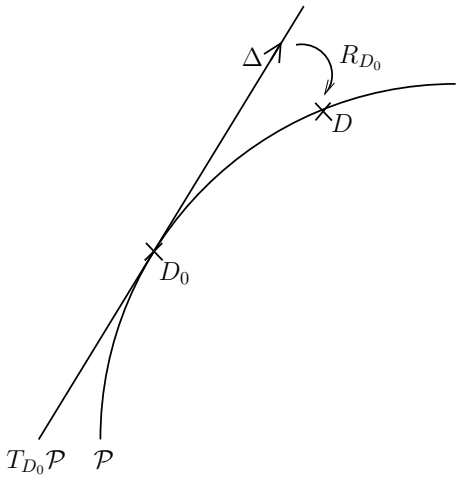

Figure 2.1: The map R_{D_0}

This map is analytic and verifies $R_{D_0}(0) = D_0$, $\mathrm{d}R_{D_0}(0) = \mathrm{id}_{T_{D_0}\mathcal{P}}$. Therefore, by the inverse function theorem, the image of a neighborhood of zero contains a

neighborhood of D_0. We now compute the gradient of $\widetilde{f} = f \circ R_{D_0}$ at a point Δ, with $D = R_{D_0}(\Delta)$.

$$\begin{aligned}\widetilde{f}(\Delta+\delta) &= f(D) + \langle \nabla_{\mathcal{P}} f(D), \mathrm{d}R_{D_0}(\Delta)\delta\rangle + O(\delta^2)\\ &= f(D) + \langle \mathrm{d}R_{D_0}(\Delta)^* \nabla_{\mathcal{P}} f(D), \delta\rangle + O(\delta^2)\\ &= f(D) + \langle P_{D_0}\mathrm{d}R_{D_0}(\Delta)^* \nabla_{\mathcal{P}} f(D), \delta\rangle + O(\delta^2).\end{aligned}$$

We deduce

$$\nabla_{T_{D_0}\mathcal{P}} \widetilde{f}(\Delta) = P_{D_0}\mathrm{d}R_{D_0}(\Delta)^* \nabla_{\mathcal{P}} f(D).$$

We can now apply the Łojasiewicz inequality to $\widetilde{f}$, which is an analytic functional defined on the Euclidean space $T_{D_0}\mathcal{M}$. We obtain a neighborhood of zero in $T_{D_0}\mathcal{P}$, and therefore a neighborhood U of D_0 on which

$$|f(D) - f(D_0)|^{1-\theta} \leq \kappa \| P_{D_0}\mathrm{d}R_{D_0}(\Delta)^* \nabla_{\mathcal{P}} f(D)\|.$$

As $\mathrm{d}R^*_{D_0}$ is continuous in Δ, up to a change in the neighborhood U and the constant κ,

$$|f(D) - f(D_0)|^{1-\theta} \leq \kappa \| \nabla_{\mathcal{P}} f(D)\|.$$

□

2.3 The gradient method

2.3.1 Description of the method

The gradient flow

$$\begin{aligned}\frac{\mathrm{d}D}{\mathrm{d}t} &= -\nabla_{\mathcal{P}} E(D)\\ &= -[D,[D,F_D]]\end{aligned} \tag{2.2}$$

is a way of minimizing the energy over the manifold $\mathcal{P}$. This continuous flow was already used to solve the Hartree-Fock equations in [AA09] (although the authors used a formulation in terms of orbitals, whereas we use density matrices).

The naive Euler discretization

$$D_{k+1} = D_k - t[D_k,[D_k,F_k]]$$

is not suitable because it does not stay on $\mathcal{P}$. A variety of approaches deal with this problem. One of the first algorithms to solve the Hartree-Fock equations [McW56] used a purification method to project D_{k+1} back onto $\mathcal{P}$. More recently, an orthogonal projection on the convex hull of $\mathcal{P}$ was used for that purpose [CP08]. Although we focus in this paper on a different gradient method, such projection methods have the same behavior for small stepsizes and can be treated in the same framework, provided one can prove results similar to Lemmas 2.1 and 2.2 below.

We look for D_{k+1} on a curve on $\mathcal{P}$ that is tangent to $\nabla_{\mathcal{P}} E(D_k)$. A natural curve on $\mathcal{P}$ is the change of basis

$$D'(t) = U_t D U_t^T,$$

where U_t is a smooth family of orthogonal matrices. If we take

$$U_t = \exp(t[D, F_D]),$$

we get

$$\left.\frac{\mathrm{d}D'}{\mathrm{d}t}\right|_{t=0} = -[D,[D,F_D]],$$

so $D'(t)$ is a smooth curve on $\mathcal{P}$, tangent to the gradient flow at $t = 0$.

Our gradient method with a fixed step t is then

$$D_{k+1} = U_k D_k U_k^T, \tag{2.3}$$

with

$$U_k = \exp(t[D_k, F_k]). \tag{2.4}$$

This method, as well as various generalizations, is described in [EAS98].

We now prove a number of lemmas which are the main ingredients of the convergence proof. First, we bound the second derivative of the energy to obtain quantitative estimates on the energy decrease, then we link the difference of iterates $D_{k+1} - D_k$ to the gradient $\nabla_{\mathcal{P}} E(D_k)$, and finally we use the Łojasiewicz inequality to establish convergence.

2.3.2 Derivatives

We start from a point D_0 and compute the derivatives of E along the curve $D_t = U_t D_0 U_{-t}$. For ease of notation we will write $\epsilon(t) = E(D_t)$, $F_t = F(D_t)$ and $C_t = [D_t, F_t]$.

$$\begin{aligned}
\frac{\mathrm{d}D_t}{\mathrm{d}t} &= \frac{\mathrm{d}U_t}{\mathrm{d}t} D_0 U_{-t} + U_t D_0 \frac{\mathrm{d}U_{-t}}{\mathrm{d}t} \\
&= [C_0, D_t], \\
\frac{\mathrm{d}^n D_t}{\mathrm{d}t^n} &= \frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} [C_0, D_t] \\
&= \underbrace{[C_0, [C_0, \dots [C_0, D_t] \dots]]}_{n \text{ times } C_0}, \\
\frac{\mathrm{d}\epsilon}{\mathrm{d}t} &= \mathrm{Tr}(F_t [C_0, D_t]), \\
\left.\frac{\mathrm{d}\epsilon}{\mathrm{d}t}\right|_{t=0} &= -\|C_0\|^2, \\
\frac{\mathrm{d}^2\epsilon}{\mathrm{d}t^2} &= \mathrm{Tr}(F_t [C_0, [C_0, D_t]]) + \mathrm{Tr}(G([C_0, D_t])[C_0, D_t]).
\end{aligned}$$

2.3.3 Control on the curvature

Lemma 2.1. *There exists $\alpha > 0$ such that for every D_0 and t,*

$$\left|\frac{\mathrm{d}^2\epsilon}{\mathrm{d}t^2}\right|(t) \leq \alpha \|C_0\|^2 .$$

Proof.

$$\frac{\mathrm{d}^2\epsilon}{\mathrm{d}t^2} = \mathrm{Tr}(F_t[C_0,[C_0,D_t]]) + \mathrm{Tr}(G([C_0,D_t])[C_0,D_t]). \tag{2.5}$$

First, since the Laplacian in $F(D)$ acts on a finite dimensional space, we can bound $F(D)$:

$$\begin{aligned}\|F(D)\|_{\mathrm{op}} &\leq \frac{1}{2}\|-\Delta\|_{\mathrm{op}} + \|V\|_{\mathrm{op}} + \|G(D)\|_{\mathrm{op}} \\ &\leq \frac{1}{2}\|-\Delta\|_{\mathrm{op}} + 2(2N+Z)\sqrt{\|-\Delta\|_{\mathrm{op}}}\end{aligned} \tag{2.6}$$

by the Hardy inequality. Next, making use of the operator inequality $\mathrm{Tr}(AB) \leq \|A\|_{\mathrm{op}}\|B\|$, we show that

$$\mathrm{Tr}(F_t[C_0,[C_0,D_t]]) \leq 2\left(\frac{1}{2}\|-\Delta\|_{\mathrm{op}} + 2(2N+Z)\sqrt{\|-\Delta\|_{\mathrm{op}}}\right)\|C_0\|^2 .$$

For the second term of (2.5),

$$\begin{aligned}\mathrm{Tr}(G([C_0,D_t])[C_0,D_t]) &\leq \|G([C_0,D_t])\|_{\mathrm{op}}\,\mathrm{Tr}\left(|[C_0,D_t]|\right) \\ &\leq 4\sqrt{\|-\Delta\|_{\mathrm{op}}}\,\mathrm{Tr}\left(|[C_0,D_t]|\right)^2 \\ &\leq 16N\sqrt{\|-\Delta\|_{\mathrm{op}}}\|C_0\|^2 .\end{aligned}$$

The result is now proved with

$$\alpha = \|-\Delta\|_{\mathrm{op}} + 4(6N+Z)\sqrt{\|-\Delta\|_{\mathrm{op}}}.$$

□

2.3.4 Choice of the stepsize

We can now expand the energy:

$$\epsilon(t) \leq \epsilon(0) - t\|C_0\|^2 + \frac{t^2}{2}\alpha\|C_0\|^2 .$$

If we choose

$$t < \frac{2}{\alpha}, \tag{2.7}$$

we obtain a decrease of the energy

$$\epsilon(t) \leq \epsilon(0) - \beta\|C_0\|^2 \tag{2.8}$$

with $\beta = t - \frac{t^2}{2}\alpha > 0$.

The optimal choice of t with this bound on the curvature is $t = \frac{1}{\alpha}$, with $\beta = \frac{1}{2\alpha}$. Of course it could be that the actual optimal t is different, and could vary wildly, which is why we will not consider optimal stepsizes.

2.3.5 Study of $D_{k+1} - D_k$

We now prove that our iteration $D_{k+1} = U_k D_k U_k^T$ coincides with an unconstrained gradient method up to first order in t.

We say that $M \in o(\|N\|)$ when for all $\varepsilon > 0$, there is a neighborhood U of zero such that when $N \in U, \|M\| \leq \varepsilon \|N\|$. Note that this neighborhood U is not allowed to depend on N, meaning that the resulting bound is uniform, which will allow us to manipulate the remainders more easily.

Lemma 2.2. *For any k,*

$$D_{k+1} = D_k + t[C_k, D_k] + o(t\|C_k\|).$$

Proof.

$$D_{k+1} - D_k - t[C_k, D_k] = \sum_{n=2}^{\infty} \frac{t^n}{n!} \underbrace{[C_k, [C_k, \dots [C_k, D_k] \dots]]}_{n \text{ times } C_k}$$

$$\begin{aligned} \|D_{k+1} - D_k - t[C_k, D_k]\| &\leq t\|[C_k, D_k]\| \sum_{n=2}^{\infty} t^{n-1} \|C_k\|^{n-1} \\ &\leq t\|[C_k, D_k]\| \frac{t\|C_k\|}{1 - t\|C_k\|} \end{aligned}$$

□

2.4 Convergence of the gradient algorithm

Theorem 2.3 (Convergence of the gradient algorithm)**.** *Let $D_0 \in \mathcal{P}$ be any density matrix and D_k be the sequences of iterates generated from D_0 by $D_{k+1} = U_k D_k U_k^T$, with stepsize $t < \frac{2}{\alpha}$. Then D_k converges towards a solution of the Hartree-Fock equations.*

Proof. The energy $E(D)$ is bounded from below on $\mathcal{P}$, and therefore E_k converges to a limit E_∞. In the sequel we will work for convenience with $\tilde{E}(D) = E(D) - E_\infty$ and drop the tildes. Immediately, summing (2.8) implies that C_k converges to 0, and therefore so does $D_k - D_{k-1}$ (this is what Cancès and Le Bris call "numerical convergence" in [CLB00b]). Note that we only get that $\|D_k - D_{k-1}\|^2$ is summable, which alone is not enough to guarantee convergence (the harmonic series $x_k = \sum_{l=1}^{k} 1/l$ is a simple counter-example). To obtain convergence, we shall use the Łojasiewicz inequality.

Let us denote by Γ the level set $\Gamma = \{D \in \mathcal{P}, E(D) = 0\}$. It is non-empty and compact. We apply the Łojasiewicz inequality to every point of Γ to obtain a cover $(U_i)_{i \in \mathcal{I}}$ of Γ in which the Łojasiewicz inequality holds with constants κ_i, θ_i.

By compactness, we extract a finite subcover from the U_i, from which we deduce $\delta > 0$, $\kappa > 0$ and $\theta \in (0, 1/2]$ such that whenever $d(D, \Gamma) < \delta$,

$$E(D)^{1-\theta} \leq \kappa \|\nabla_{\mathcal{P}} E(D)\| = \kappa \|[D, C_D]\| = \kappa \|C_D\|. \tag{2.9}$$

(recall from Section 2.2.2 that $\|[D, [D, M]]\| = \|[D, M]\|$ for M symmetric)

To apply the Łojasiewicz inequality to our iteration, it remains to show that $d(D_k, \Gamma)$ tends to zero. Suppose this is not the case. Then we can extract a subsequence, still denoted by D_k, such that $d(D_k, \Gamma) \geq \varepsilon$ for some $\varepsilon > 0$. By compactness of $\mathcal{P}$ we extract a subsequence that converges to a D such that $d(D, \Gamma) \geq \varepsilon$ and (by continuity) $E(D) = 0$, a contradiction. Therefore $d(D_k, \Gamma) \to 0$, and for k larger than some k_0,

$$E(D_k)^{1-\theta} \leq \kappa \|C_k\| . \tag{2.10}$$

For $k \geq k_0$,

$$\begin{aligned} E(D_k)^\theta - E(D_{k+1})^\theta &\geq \frac{\theta}{E(D_k)^{1-\theta}}(E(D_k) - E(D_{k+1})) \\ &\geq \frac{\theta}{\kappa\|C_k\|}(E(D_k) - E(D_{k+1})) \\ &\geq \frac{\theta\beta}{\kappa}\|C_k\| \\ &\geq \frac{\theta\beta}{\kappa t}\|D_{k+1} - D_k\| + o(\|D_{k+1} - D_k\|) \end{aligned}$$

hence

$$\frac{\theta\beta}{\kappa t}\|D_{k+1} - D_k\| + o(\|D_{k+1} - D_k\|) \leq E(D_k)^\theta - E(D_{k+1})^\theta. \tag{2.11}$$

As the right-hand side is summable, so is the left-hand side, which implies that $\sum\|D_{k+1} - D_k\| < \infty$. D_k is therefore Cauchy and converges to some limit D_∞. $C_k \to 0$, so D_∞ is a critical point.

Note that now that we know that D_k converges to D_∞, we can replace the θ and κ in this inequality by the ones obtained from the Łojasiewicz inequality around D_∞ only. □

Let

$$e_k = \sum_{l=k}^{\infty} \|D_{l+1} - D_l\| .$$

This is a (crude) measure of the error at iteration number k. In particular, $\|D_k - D_\infty\| \leq e_k$.

Theorem 2.4 (Convergence rate of the gradient algorithm)**.**

1. *If $\theta = 1/2$ (non-degenerate case), then for any $\nu' < \frac{\beta}{2\kappa^2}$, there exists $c > 0$ such that*

$$e_k \leq c(1 - \nu')^k. \tag{2.12}$$

2. *If $\theta \neq 1/2$ (degenerate case), then there exists $c > 0$ such that*

$$e_k \leq ck^{-\frac{\theta}{1-2\theta}}. \tag{2.13}$$

Proof. Summing (2.11) from $l = k$ to ∞ with $k \geq k_0$, we obtain

$$\begin{aligned} e_k + o(e_k) &\leq \frac{\kappa t}{\theta\beta} E(D_k)^\theta \\ \left(\frac{\theta\beta}{\kappa t} e_k + o(e_k)\right)^{\frac{1-\theta}{\theta}} &\leq E(D_k)^{1-\theta} \\ &\leq \kappa \|C_k\| \\ &\leq \frac{\kappa}{t}(e_k - e_{k+1}) + o(e_k - e_{k+1}) \end{aligned}$$

Hence,

$$\begin{aligned} e_{k+1} \leq e_k - \nu e_k^{\frac{1-\theta}{\theta}} + o(e_k{}^{\frac{1-\theta}{\theta}}), \text{ with} \\ \nu = \frac{t}{\kappa}\left(\frac{\theta\beta}{\kappa t}\right)^{\frac{1-\theta}{\theta}} \end{aligned}$$

Two cases are to be distinguished. If $\theta = \frac{1}{2}$, the above inequality reduces to

$$e_{k+1} \leq (1 - \nu + o(1)) e_k$$

with $\nu = \frac{\beta}{2\kappa^2}$ and the result follows.

The case $\theta \neq 1/2$ is more involved. We define

$$y_k = ck^{-p},$$

which verifies

$$\begin{aligned} y_{k+1} &= c(k+1)^{-p} \\ &= ck^{-p}(1 + 1/k)^{-p} \\ &\geq ck^{-p}(1 - \frac{p}{k}) \\ &\geq y_k(1 - pc^{-\frac{1}{p}} y_k^{\frac{1}{p}}) \end{aligned}$$

We set $p = \frac{\theta}{1-2\theta}$ and c large enough so that $c > (\frac{\nu}{p})^{-p}$ and $y_{k_0} \geq e_{k_0}$. We then prove by induction $e_k \leq y_k$ for $k \geq k_0$. The result follows by increasing c to ensure that $e_k \leq y_k$, for $k < k_0$. □

In the non-degenerate case $\theta = 1/2$ (which was the case for the numerical simulations we performed, see Section 2.7), the convergence is asymptotically geometric with rate $1 - \nu$, where

$$\nu = \frac{\beta}{2\kappa^2}.$$

With the choice $t = \frac{1}{\alpha}$ suggested by our bounds, the convergence rate is

$$\nu = \frac{1}{4\kappa^2\alpha}.$$

2.5 Convergence of the Roothaan algorithm

The Roothaan algorithm (also called simple SCF in the literature) is based on the observation that a minimizer D of the energy satisfies the *Aufbau* principle: D is the projector onto the space spanned by the eigenvectors associated with the first N eigenvalues of $F(D)$. This suggests a simple fixed-point algorithm: take for D_{k+1} the projector onto the space spanned by the eigenvectors associated with the first N eigenvalues of $F(D_k)$, and iterate. Unfortunately, this procedure does not always work: in some circumstances, oscillations between two states occur, and the algorithm never converges. This behavior was explained mathematically in [CLB00b], where the authors notice that the Roothaan algorithm minimizes the bilinear functional

$$E(D, D') = \mathrm{Tr}(h(D + D')) + \mathrm{Tr}(G(D)D')$$

with respect to its first and second argument alternatively. Thanks to the Łojasiewicz inequality, we can improve on their result and prove the convergence or oscillation of the method.

The bilinear functional verifies $E(D, D') = E(D', D)$, $E(D, D) = 2E(D)$. In fact, $\frac{1}{2}E(\cdot, \cdot)$ is the symmetric bilinear form associated with the quadratic form $E(\cdot)$. In the following, we assume the uniform well-posedness hypothesis of [CLB00b], *i.e.*that there is a uniform gap of at least $\gamma > 0$ between the eigenvalues number N and $N + 1$ of $F(D_k)$. Under this assumption, it can be proven [Can+03] that there is a decrease of the bilinear functional at each iteration

$$\begin{aligned} E(D_{k+1}, D_{k+2}) &= E(D_{k+2}, D_{k+1}) \\ &= \min_{D \in \mathcal{P}} E(D, D_{k+1}) \\ &\leq E(D_k, D_{k+1}) - \gamma \|D_{k+2} - D_k\|^2 \end{aligned}$$

Since $E(\cdot, \cdot)$ is bounded from below on $\mathcal{P} \times \mathcal{P}$, this immediately shows that $D_k - D_{k+2} \to 0$, which shows that D_{2k} and D_{2k+1} converge up to extraction, which was noted in [CLB00b]. We now prove convergence of these two subsequences, again using the Łojasiewicz inequality.

$E(\cdot, \cdot)$ is defined on $\mathcal{P} \times \mathcal{P}$, which inherits the Riemannian structure of $\mathcal{P}$ by the natural inner product $\langle (D_1, D_1'), (D_2, D_2') \rangle = \langle D_1, D_2 \rangle + \langle D_1', D_2' \rangle$. In this setting, the gradient is

$$\nabla_{\mathcal{P} \times \mathcal{P}} E(D, D') = \begin{pmatrix} [D, F(D')] \\ [D', F(D)] \end{pmatrix}.$$

and therefore, using the fact that D_{k+1} (resp. D_{k+2}) and $F(D_k)$ (resp $F(D_{k+1})$) commute,

$$\begin{aligned} \|\nabla_{\mathcal{P} \times \mathcal{P}} E(D_k, D_{k+1})\| &= \sqrt{\|[D_k, F(D_{k+1})]\|^2 + \|[D_{k+1}, F(D_k)]\|^2} \\ &= \|[D_k, F(D_{k+1})]\| \\ &= \|[D_k - D_{k+2}, F(D_{k+1})]\| \\ &\leq 2\|F(D_{k+1})\|_{\mathrm{op}} \|D_{k+2} - D_k\| \end{aligned}$$

A trivial extension of Theorem 2.2 to the case of a functional defined on $\mathcal{P} \times \mathcal{P}$ shows that we can apply the Łojasiewicz inequality to $E(\cdot,\cdot)$. By the same compactness argument as before, the inequality

$$\begin{aligned} E(D_k, D_{k+1})^{1-\theta'} &\leq \kappa' \|\nabla_{\mathcal{P}\times\mathcal{P}} E(D_k, D_{k+1})\| \\ &\leq 2\kappa' \|F(D_{k+1})\|_{\mathrm{op}} \|D_{k+2} - D_k\| \end{aligned}$$

holds for k large enough, with constants $\kappa' > 0$ and $\theta' \in (0, \frac{1}{2}]$.

The exact same reasoning as for the gradient algorithm proves the following theorems

Theorem 2.5 (Convergence/oscillation of the Roothaan algorithm)**.** *Let $D_0 \in \mathcal{P}$ such that the sequence D_k of iterates generated by the Roothaan algorithms verifies the uniform well-posedness hypothesis with uniform gap $\gamma > 0$. Then the two subsequences D_{2k} and D_{2k+1} are convergent. If both have the same limit, then this limit is a solution of the Hartree-Fock equations.*

Theorem 2.6 (Convergence rate of the Roothaan algorithm)**.** *Let D_k be the sequence of iterates generated by a uniformly well-posed Roothaan algorithm, and let*

$$e_k = \sum_{l=k}^{\infty} \|D_{l+2} - D_l\| .$$

Then,

1. *If $\theta' = 1/2$ (non-degenerate case), then for any $\nu' < \frac{\gamma}{8\kappa'^2 \|F\|_{op}^2}$, where $\|F\|_{op}$ is the uniform bound on F proved in* (2.6)*, there exists $c > 0$ such that*

$$e_k \leq c(1-\nu')^k. \tag{2.14}$$

2. *If $\theta' \neq 1/2$ (degenerate case), then there exists $c > 0$ such that*

$$e_k \leq c k^{-\frac{\theta'}{1-2\theta'}}. \tag{2.15}$$

2.6 Level-shifting

The Level-Shifting algorithm was introduced in [SH73] as a way to avoid oscillation in self-consistent iterations. By shifting the N lowest energy levels (eigenvalues of F), one can force convergence, although denaturing the equations in the process. We now prove the convergence of this algorithm.

The same arguments as before apply to the functional

$$\begin{aligned} E^b(D, D') &= \mathrm{Tr}(h(D+D')) + \mathrm{Tr}(G(D)D') + \frac{b}{2}\left\|D - D'\right\|^2 \\ &= \mathrm{Tr}(h(D+D')) + \mathrm{Tr}(G(D)D') - b\,\mathrm{Tr}(DD') + bN \end{aligned}$$

with associated Fock matrix $F^b(D) = F(D) - bD$. The difference with the Roothaan algorithm is that for b large enough, there is a uniform gap $\gamma^b > 0$, and $D_k - D_{k+1}$ converges to 0 [CLB00b]. Therefore, we have the following theorems

Theorem 2.7 (Convergence of the Level-Shifting algorithm)**.** *Let $D_0 \in \mathcal{P}$. Then there exists $b_0 > 0$ such that for every $b > b_0$, the sequence D_k of iterates generated by the Level-Shifting algorithm with shift parameter b verifies the uniform well-posedness hypothesis with uniform gap $\gamma > 0$ and converges.*

Theorem 2.8 (Convergence rate of the Level-Shifting algorithm)**.** *Let D_k be the sequence of iterates generated by the Level-Shifting with shift parameter $b > b_0$, and let*

$$e_k = \sum_{l=k}^{\infty} \|D_{l+2} - D_l\| .$$

Then,

1. *If $\theta' = 1/2$ (non-degenerate case), then for any $\nu' < \frac{\gamma^b}{8\kappa'^2 \left\|F^b\right\|_{op}^2}$, there exists $c > 0$ such that*

$$e_k \leq c(1-\nu')^k. \tag{2.16}$$

2. *If $\theta' \neq 1/2$ (degenerate case), then there exists $c > 0$ such that*

$$e_k \leq c k^{-\frac{\theta'}{1-2\theta'}}. \tag{2.17}$$

We can use this result to heuristically predict the behavior of the algorithm when b is large. γ^b and $\left\|F^b\right\|_{\text{op}}$ both scale as b for large values of b. Assuming non-degeneracy, we can take $\kappa' > \frac{1}{\sqrt{2|\lambda_1|}}$, where λ_1 is the eigenvalue of smallest magnitude of the Hessian $H_1 + \frac{b}{2}H_2$, where $H_1 = H_{\mathcal{P}\times\mathcal{P}}E(D^\infty, D^\infty)$ and $H_2 = H_{\mathcal{P}\times\mathcal{P}}\|D - D'\|^2 (D^\infty, D^\infty)$. But H_2 admits zero as an eigenvalue (for instance, note that $\left\|D - D'\right\|^2$ is constant along the curve $(D_t, D'_t) = (U_t D U_t^T, U_t D' U_t^T)$, where U_t is a family of orthogonal matrices), so that, when b goes to infinity, λ_1 tends to the eigenvalue of smallest magnitude of H_1 restricted to the nullspace of H_2, and therefore stays bounded. Therefore, ν' scales as $\frac{1}{b}$, which suggests that b should not be too large for the algorithm to converge quickly.

2.7 Numerical results

We illustrate our results on atomic systems, using gaussian basis functions. The gradient method was implemented using the software Expokit [Sid98] to compute matrix exponentials. In our computations, the cost of a gradient step is not much higher than a step of the Roothaan algorithm, since the limiting step is computing the Fock matrix, not the exponential. However, the situation may change if the Fock matrix is computed using linear scaling techniques. In this case, one can use more efficient ways of computing geodesics, as described in [EAS98].

First, the Łojasiewicz inequality with exponent $\frac{1}{2}$ was checked to hold in the molecular systems and basis sets we encountered, suggesting that the minimizers

are non-degenerate. Consequently, we never encountered sublinear convergence of any algorithm.

For a given molecular system and basis, we checked that the Level-Shifting algorithm converged as $(1-\nu)^k$, where ν is asymptotically proportional to $\frac{1}{b}$, which we predicted theoretically in Section 2.6 (see Figure 2.2). This means that the estimates we used have at least the correct scaling behavior.

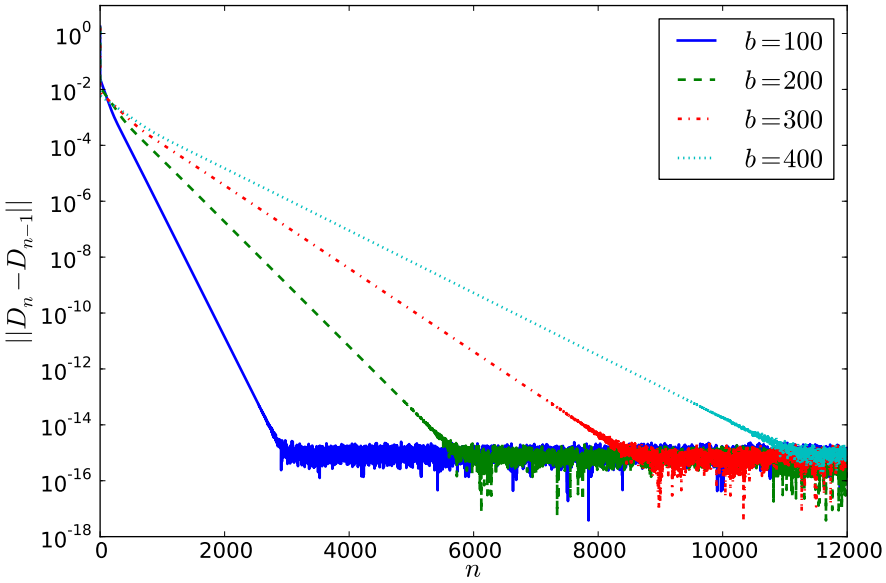

Figure 2.2: Convergence of the Level-Shifting algorithm. The convergence is linear until machine precision. The horizontal spacing of the curves reveals the asymptotic relationship $\nu \propto \frac{1}{b}$. The system considered is the carbon atom ($N = Z = 6$), under the RHF formalism, using the 3-21G gaussian basis functions.

Next, we compared the efficiency of the Roothaan algorithm and of the gradient algorithm, in the case where the Roothaan algorithm converges. Our analysis leads to the estimate $\nu = \frac{\gamma}{8\kappa'^2\|F\|_{\text{op}}^2}$ for the Roothaan algorithm, and $\nu = \frac{1}{4\kappa^2\alpha}$ for the gradient algorithm with stepsize $t = \frac{1}{\alpha}$.

It is immediate to see that, up to a constant multiplicative factor, $\kappa' > \kappa$, $\gamma \leq \|F\|_{\text{op}}$ and for the cases of interest $\alpha \approx \|F\|_{\text{op}}$, so from our estimates we would expect the gradient algorithm to be faster than the Roothaan algorithm. However, in our tests the Roothaan algorithm was considerably faster than the gradient algorithm (see Figure 2.3). This conclusion holds even when the stepsize is adjusted at each iteration with a line search.

The reason that the Roothaan algorithm performs better than expected is that the inequality

$$\|[D_{k+2} - D_k, F(D_{k+1})]\| \leq 2\|F(D_{k+1})\|\|D_{k+2} - D_k\|$$

is very far from optimal. Whether an improved bound (in particular, one that does not depend on the dimension) can be derived is an interesting open question.

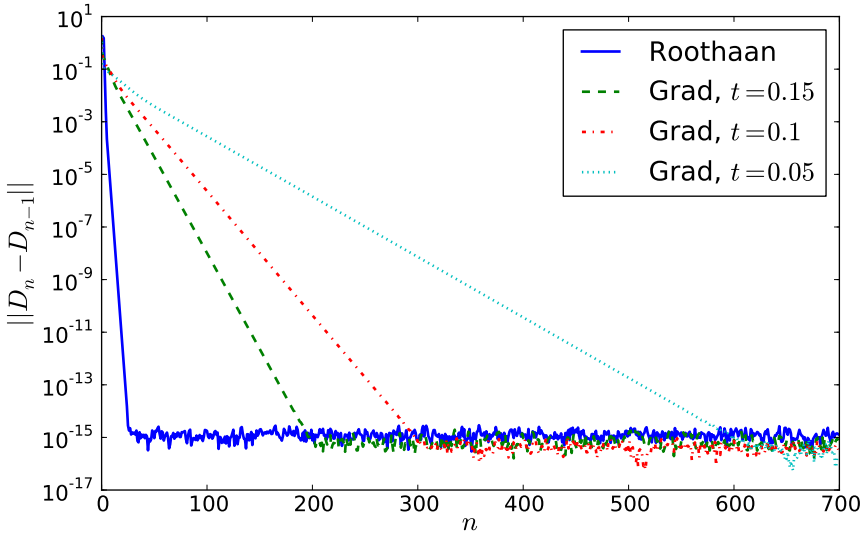

Figure 2.3: Comparison of Roothaan and gradient algorithm. The system considered is the carbon atom ($N = Z = 6$), under the RHF formalism, using the 3-21G gaussian basis functions.

The outcome of these tests seems to be that the gradient algorithm is slower. It might prove to be faster in situations where the gap is small, or whenever κ' is much larger than κ. We have been unable to find concrete examples of such cases.

2.8 Conclusion, perspectives

In this paper, we introduced an algorithm based on the idea of gradient descent. By using the analyticity of the objective function and of the constraint manifold, we were able to derive a Łojasiewicz inequality, and use that to prove the convergence of the gradient method, under the assumption of a small enough stepsize. Next, expanding on the analysis of [CLB00b], we extended the Łojasiewicz inequality to a Lyapunov function for the Roothaan algorithm. By linking the gradient of this Lyapunov function to the difference in the iterates of the algorithm, we proved convergence (or oscillation), an improvement over previous results which only prove a weaker version of this. In this framework, the Level-Shifting algorithm can be seen as a simple modification of the Roothaan algorithm, and as such can be treated by the same methods. In each case, we were also able to derive explicit bounds on the convergence rates.

The strength of the Łojasiewicz inequality is that no higher-order hypothesis are needed for its use. As a consequence, the rates of convergence we obtain weaken considerably if the algorithm converges to a degenerate critical point. A more precise study of the local structure of critical points is necessary to understand why the algorithms usually exhibit geometric convergence. This is related to the problem of local uniqueness and is likely to be hard in a non-perturbative regime (see [GH12] for a large-Z result).

Even though our results hide the complexity of the local structure in the constants of the Łojasiewicz inequality, they still provide insight as to the influence of the basis on the speed of convergence, and can be used to compare algorithms. All of our results use crucially the hypothesis of a finite-dimensional Galerkin space. For the gradient algorithm, we need it to ensure the existence of a stepsize that decreases the energy. This is analogous to a CFL condition for the discretization of the equation $\frac{\mathrm{d}D}{\mathrm{d}t} = -[D,[D,F_D]]$, and can only be lifted with a more implicit discretization of this equation. For the Roothaan and Level-Shifting algorithms, we use the finite dimension hypothesis to bound $F(D)$. As noted in Section 2.7, the inequality is not sharp, so it could be that the infinite-dimensional version of the Roothaan and Level-Shifting algorithms still converge. More research is needed to examine this.

The gradient algorithm we examined only converges towards a stationary point of the energy, that may not be a local minimizer, or even an Aufbau solution. However, it will generically converge towards a local minimizer, unlike the Level-Shifting algorithm with large b. Therefore, it is the most robust of the algorithms considered. Although it was found to be slower than other algorithms on the numerical tests we performed, it has the advantage that its convergence rate does not depend on the gap $\lambda_{N+1} - \lambda_N$, and might therefore prove useful in extreme situations.

An algorithm that could achieve the speed of the fixed-point algorithms with the robustness granted by the energy monotonicity seems to be the ODA algorithm of Cancès and Le Bris [CLB00b], along with variants such as EDIIS, or combinations of EDIIS and DIIS algorithms [KSC02]. We were not able to examine these algorithms in this paper. At first glance, the ODA algorithm should fit into our framework (indeed, the ODA algorithm was built to satisfy an energy decrease inequality similar to (2.8)). However, it works in a relaxed parameter space $\widetilde{P}$, and using the commutator to control the differences of iterates as we did only makes sense on $\mathcal{P}$. Therefore, other arguments have to be used.

A variant on the gradient algorithm used here is to modify the local geometry of the manifold $\mathcal{P}$ by using a different inner product, leading to a variety of methods, including conjugate gradient algorithms [EAS98]. These methods fit into our framework, as long as one can prove that they are "gradient-like", in the sense that one can control the gradient by the difference $D_{k+1} - D_k$. However, precise estimates of convergence rates might be hard to obtain.

Also missing from the present contribution is the study of other commonly used algorithms, such as DIIS [Pul82], and variants of (quasi)-Newton algorithms [Bac81; Høs+08]. DIIS numerically exhibits a complicated behavior that is probably hard to explain analytically, and the (quasi)-Newton algorithms require a study of the second-order structure of the critical points, which we are unable to do.

Acknowledgments

The author would like to thank Eric Séré for his extensive help, Guillaume Legendre for the code used in the numerical simulations and Julien Salomon for introducing him to the Łojasiewicz inequality. He also thanks the anonymous referees for many constructive remarks.

Bibliography

[AA09] F. Alouges and C. Audouze. Preconditioned gradient flows for nonlinear eigenvalue problems and application to the Hartree-Fock functional. *Numerical Methods for Partial Differential Equations* 25.2 (2009), pp. 380–400.

[Bac81] G.B. Bacskay. A quadratically convergent Hartree–Fock (QC-SCF) method. Application to closed shell systems. *Chemical Physics* 61.3 (1981), pp. 385–404.

[CLB00a] E. Cancès and C. Le Bris. Can we outperform the DIIS approach for electronic structure calculations? *International Journal of Quantum Chemistry* 79.2 (2000), pp. 82–90.

[CLB00b] E. Cancès and C. Le Bris. On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 34.4 (2000), pp. 749–774.

[CP08] E. Cancès and K. Pernal. Projected gradient algorithms for Hartree-Fock and density matrix functional theory calculations. *The Journal of chemical physics* 128 (2008), p. 134108.

[Can+03] E. Cancès et al. Computational quantum chemistry: a primer. *Handbook of numerical analysis* 10 (2003), pp. 3–270.

[Can00] E. Cancès. SCF algorithms for Hartree-Fock electronic calculations. In: *Mathematical models and methods for ab initio quantum chemistry, Lecture Notes in Chemistry*. Vol. 74. 2000.

[EAS98] A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* 20 (1998), p. 303.

[FMM04] J.B. Francisco, J.M. Martínez, and L. Martínez. Globally convergent trust-region methods for self-consistent field electronic structure calculations. *The Journal of chemical physics* 121 (2004), p. 10863.

[GH12] M. Griesemer and F. Hantsch. Unique Solutions to Hartree–Fock equations for closed-shell atoms. *Archive for Rational Mechanics and Analysis* 203 (3 2012), pp. 883–900.

[HJK03] A. Haraux, M.A. Jendoubi, and O. Kavian. Rate of decay to equilibrium in some semilinear parabolic equations. *Journal of Evolution equations* 3.3 (2003), pp. 463–484.

[Høs+08] S. Høst et al. The augmented Roothaan-Hall method for optimizing Hartree-Fock and Kohn-Sham density matrices. *The Journal of chemical physics* 129 (2008), p. 124106.

[KSC02] K.N. Kudin, G.E. Scuseria, and E. Cancès. A black-box self-consistent field convergence algorithm: One step closer. *The Journal of chemical physics* 116 (2002), p. 8255.

[LS77] E.H. Lieb and B. Simon. The Hartree-Fock theory for Coulomb systems. *Comm. Math. Phys.* 53.3 (1977), pp. 185–194.

[Lio87] P.L. Lions. Solutions of Hartree-Fock equations for Coulomb systems. *Communications in Mathematical Physics* 109.1 (1987), pp. 33–97.

[Łoj65] S. Łojasiewicz. *Ensembles semi-analytiques*. Institut des Hautes Etudes Scientifiques, 1965.

[McW56] R. McWeeny. The density matrix in self-consistent field theory. I. Iterative construction of the density matrix. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 235.1203 (1956), p. 496.

[Pul82] P. Pulay. Improved SCF convergence acceleration. *Journal of Computational Chemistry* 3.4 (1982), pp. 556–560.

[Sal07] J. Salomon. Convergence of the time-discretized monotonic schemes. *ESAIM: Mathematical Modelling and Numerical Analysis* 41.01 (2007), pp. 77–93.

[SH73] V.R. Saunders and I.H. Hillier. A Level-Shifting method for converging closed shell Hartree-Fock wave functions. *International Journal of Quantum Chemistry* 7.4 (1973), pp. 699–705.

[Sid98] R.B. Sidje. Expokit: a software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)* 24.1 (1998), pp. 130–156.

Chapitre 3

Le modèle de Dirac-Fock multiconfiguration

Ce chapitre reprend le texte intégral de l'article "Solutions of the multiconfiguration Dirac-Fock equations", soumis pour publication.

Solutions of the multiconfiguration Dirac-Fock equations

Antoine Levitt

Abstract

The multiconfiguration Dirac-Fock (MCDF) model uses a linear combination of Slater determinants to approximate the electronic N-body wave function of a relativistic molecular system, resulting in a coupled system of nonlinear eigenvalue equations, the MCDF equations. In this paper, we prove the existence of solutions of these equations in the weakly relativistic regime. First, using a new variational principle as well as results of Lewin on the multiconfiguration nonrelativistic model, and Esteban and Séré on the single-configuration relativistic model, we prove the existence of critical points for the associated energy functional, under the constraint that the occupation numbers are not too small. Then, this constraint can be removed in the weakly relativistic regime, and we obtain non-constrained critical points, i.e. solutions of the multiconfiguration Dirac-Fock equations.

3.1 Introduction

Consider an atom or molecule with N electrons. Nonrelativistic quantum mechanics dictates that, under the Born-Oppenheimer approximation, the electronic rest energy is given by the lowest fermionic eigenvalue of the N-body Hamiltonian. The complexity of this problem grows exponentially with N, and approximations are used to keep the problem tractable. Hartree-Fock theory uses the variational ansatz that the N-body wavefunction is a single Slater determinant. The optimization of the resulting energy over the orbitals gives rise to a nonlinear eigenvalue problem, which is solved iteratively.

It is well-known that this method overestimates the true ground state energy by a quantity known as the correlation energy, whose size can be significant in many cases of chemical interest [SO89]. This can be remedied by considering several Slater determinants, a technique known as multiconfiguration Hartree-Fock (MCHF) theory. This brings the model closer to the full N-body problem, and, in the limit of an infinite number of determinants, one recovers the true ground state energy.

Another source of errors is that the Hamiltonian used is non-relativistic. Indeed, in large atoms, the core electrons reach relativistic speeds (in atomic units, of the order of Z, compared with the speed of light $c \approx 137$). This causes a length contraction which affects the screening by the core electrons of the attractive potential of the nucleus. This has important consequences for the valence electrons and the chemistry of elements. Neglecting these effects leads to incorrect conclusions, and for instance fails to account for the difference in color between silver and gold [PD79].

For a fully relativistic treatment of the electrons, one should use quantum electrodynamics (QED). But this very precise theory is also extremely complex for all but the simplest systems. Therefore, physicists and chemists use approximate Hamiltonians to avoid working in the full Fock space of QED. The multiconfiguration Dirac-Fock (MCDF) model is obtained by using the Dirac operator in the multiconfiguration Hartree-Fock model. It incorporates relativistic effects into the multiconfiguration Hartree-Fock model, and has been used successfully in a number of applications [DFJ07; Gra07].

Although these models, and more complicated ones, are used routinely by physicists, many problems still remain in their mathematical analysis. The first rigorous proof of existence of ground states of the Hartree-Fock equations was given by Lieb and Simon [LS77] and later generalized to excited states by Lions [Lio87]. The multiconfiguration equations were studied by Le Bris [LB94], who proved existence in the particular case of doubly excited states. Friesecke later proved the existence of minimizers for an arbitrary number of determinants [Fri03b], and Lewin generalized his proof to excited states, in the spirit of the method of Lions [Lew04]. For relativistic models, Esteban and Séré proved existence of single-configuration solutions to the Dirac-Fock equations [ES99], and studied their non-relativistic limit [ES01]. To our knowledge, the present work is the first mathematical study of a relativistic multiconfiguration model.

The main mathematical difficulty of the multiconfiguration equations, apart from the increased algebraic complexity, is that one cannot simultaneously diagonalize the Fock operator and the matrix of Lagrange multipliers. Lewin rewrote the Euler-Lagrange equations in a vector formalism and used the same arguments as in the Hartree-Fock case [LS77; Lio87] to prove the existence of solutions.

The Dirac-Fock equations are considerably more difficult to handle than the Hartree-Fock equations. The main difficulty is that the Dirac operator is not bounded from below. This fact, which causes important problems already in the linear theory, complicates the search for solutions of the equations, because every critical point has an infinite Morse index. One can therefore no longer minimize the energy functional, or even use standard critical point theory. Esteban and Séré [ES99], later generalized by Buffoni, Esteban and Séré [BES06], used the concavity of the energy with respect to the negative directions of the free Dirac operator to reduce the problem to one whose critical points have a finite Morse index.

The MCDF model combines the two mathematical problems and adds the difficulty that, for the theory to make sense, the speed of light has to be above a constant that depends on a lower bound on the occupation numbers. Note that this difficulty with small occupation numbers is also encountered in numerical computations [ID93], and theoretical studies of the nonrelativistic evolution problem [Bar+10].

In this paper, we prove the existence of solutions, when the speed of light is large enough (weakly relativistic regime). We now describe our formalism.

3.2 Definitions

In atomic units, the Dirac operator is given by

$$D_c = -ic(\alpha \cdot \nabla) + c^2\beta. \tag{3.1}$$

In standard representation, α and β are 4×4 matrices given by

$$\alpha_k = \begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix}, \beta_k = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix},$$

where the σ_k are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The speed of light c has the physical value $c \approx 137$.

The operator D_c is self-adjoint on $L^2(\mathbb{R}^3, \mathbb{C}^4)$ with domain $H^1(\mathbb{R}^3, \mathbb{C}^4)$ and form domain $H^{1/2}(\mathbb{R}^3, \mathbb{C}^4)$. It verifies the relativistic identity $D_c^2 = c^4 - c^2\Delta$. More precisely, it admits the spectral decomposition

$$D_c = P^+\sqrt{c^4 - c^2\Delta}\, P^+ - P^-\sqrt{c^4 - c^2\Delta}\, P^-, \tag{3.2}$$

where the projectors $P^\pm$ are given in the Fourier domain by

$$P^\pm(\xi) = \frac{1}{2}\left(1_{\mathbb{C}^4} + \pm\frac{c\alpha \cdot \xi + c^2\beta}{\sqrt{c^4 + c^2\xi^2}}\right). \tag{3.3}$$

We denote by

$$E = H^{1/2}(\mathbb{R}^3, \mathbb{C}^4) \tag{3.4}$$

the form-domain of D_c, and $E^\pm = P^\pm E$ the two positive and negative spectral subspaces.

We will use three scalar products in this paper:

$$\begin{aligned} \langle \psi, \phi \rangle_{L^2} &= \int_{\mathbb{R}^3} \psi^* \phi, \\ \langle \psi, \phi \rangle_E &= \left\langle \psi, \sqrt{1 - \Delta}\phi \right\rangle_{L^2}, \\ \langle \psi, \phi \rangle_c &= \left\langle \psi, \sqrt{1 - \frac{\Delta}{c^2}}\phi \right\rangle_{L^2}, \end{aligned}$$

with associated norms $\|\psi\|_{L^2}, \|\psi\|_E, \|\psi\|_c$. The purpose of this last norm is to simplify several estimates. It is related to the change of variables $d_c(\psi)(x) = c^{-3/2}\psi(\frac{x}{c})$ used in [ES01] in the sense that

$$\langle \Psi, \Phi \rangle_c = \langle d_c(\Psi), d_c(\Phi) \rangle_E$$

A molecule made of M nuclei with positions z_i and charges Z_i creates an attractive potential

$$V(x) = -\sum_{i=1}^{M} \frac{Z_i}{|x - z_i|}.$$

More generally, we consider a charge distribution $\mu \geq 0$ with $\mu(\mathbb{R}^3) = Z$, which creates a potential

$$V = -\mu \star \frac{1}{|x|}. \tag{3.5}$$

In the sequel, we shall always assume that $N < Z + 1$, which is the only case where we can prove existence of solutions to our equations. This assumption is made in existence proofs for the Hartree-Fock model to ensure that an electron cannot "escape to infinity", because it will then feel the effective attractive potential $\frac{(N-1)-Z}{|x|}$ [Lio87; LS77]. Mathematically, it is used to prove that second order information on Palais-Smale sequences implies that the Lagrange multipliers are not in the essential spectrum.

The Hamiltonian $D_c + V$ has a spectral gap around zero as long as

$$Z < \frac{2}{\pi/2 + 2/\pi} c.$$

This is related to the following Hardy-type inequality (see [Tix98; Her77; Kat66]) :

$$|\langle \psi, V\psi \rangle| \leq \frac{Z}{2}(\pi/2 + 2/\pi) \left\langle \psi, \sqrt{1 - \Delta}\psi \right\rangle \tag{3.6}$$

for all $\psi \in E^{\pm}$, a refinement of the Kato inequality

$$|\langle \psi, V\psi \rangle| \leq \frac{Z\pi}{2} \left\langle \psi, \sqrt{-\Delta}\psi \right\rangle \tag{3.7}$$

for all $\psi \in E$, which we will use extensively in this paper. We also recall the standard Hardy inequality:

$$\|V\phi\|_{L^2} \leq 2Z \|\nabla\phi\|_{L^2} \tag{3.8}$$

for all $\phi \in H^1$.

The N-body relativistic Hamiltonian is given by

$$H^N = \sum_{i=1}^{N} (D_{c,x_i} + V(x_i)) + \sum_{1 \leq i < j \leq N} \frac{1}{\left|x_i - x_j\right|}.$$

This Hamiltonian acts on $\bigwedge^N L^2_a(\mathbb{R}^3, \mathbb{C}^4)$, the fermionic N-body space. Its interpretation is problematic : in particular, its essential spectrum is all of $\mathbb{R}$, and it is not even known whether eigenvectors exists [Der12].

For a given $K \geq N$, the multiconfiguration ansatz is

$$\psi = \sum_{1 \leq i_1 < \cdots < i_N \leq K} a_{i_1,\ldots,i_N} \left| \psi_{i_1} \ \ldots \ \psi_{i_N} \right\rangle, \tag{3.9}$$

where

$$\left| \psi_{i_1} \ \ldots \ \psi_{i_N} \right\rangle (X_1, \ldots, X_N) = \frac{1}{\sqrt{N!}} \det(\psi_{i_k}(X_l))_{k,l}$$

with $X_l = (x_l, s_l), x_l \in \mathbb{R}^3, s_l \in \{1,2,3,4\}$ are Slater determinants, and $a \in S, \Psi \in \Sigma$, where

$$S = \{a \in \mathbb{C}^{\binom{K}{N}}, \|a\|^2 = \sum_{1 \leq i_1 < \cdots < i_N \leq K} \left| a_{i_1,\ldots,i_N} \right|^2 = 1\}, \tag{3.10}$$

$$\begin{aligned} \Sigma &= \{\Psi \in E^K, \operatorname{Gram} \Psi = 1\}, \\ &= \{\Psi \in E^K, \left\langle \psi_i, \psi_j \right\rangle_{L^2} = \delta_{ij}\}. \end{aligned} \tag{3.11}$$

Our convention here and in the rest of this paper is to use lower case greek letters for orbitals $\psi \in E$, and upper case greek letters for vectors of orbitals $\Psi \in E^K$. We extend in a straightforward way the scalar products $\langle \cdot, \cdot \rangle_{L^2}, \langle \cdot, \cdot \rangle_E$ and $\langle \cdot, \cdot \rangle_c$ to the space E^K:

$$\langle \Psi, \Phi \rangle_* = \sum_{k=1}^{K} \langle \psi_k, \phi_k \rangle_* .$$

Following [Lew04], we define

$$\alpha_{i_1 \ldots i_N} = \begin{cases} 0 & \text{if } \#(i_1 \ldots i_N) < N, \\ \frac{\epsilon(\sigma)}{\sqrt{N!}} a_{i_{\sigma(1)},\ldots,i_{\sigma(N)}} & \text{otherwise,} \end{cases}$$

where, for all $i_1, \ldots, i_N$ with $\#(i_1 \ldots i_N) = N$, σ is the unique permutation such that $i_{\sigma(1)} < \cdots < i_{\sigma(N)}$.

With this definition,

$$\psi(X_1, \ldots, X_N) = \sum_{1 \leq i_1 \leq N, \ldots, 1 \leq i_N \leq N,} \alpha_{i_1,\ldots,i_N} \psi_{i_1}(X_1) \ldots \psi_{i_N}(X_N).$$

Then, substituting into the relativistic energy $\left\langle \psi, H^N \psi \right\rangle$, we obtain [Lew04]

$$\mathcal{E}(a, \Psi) = \left\langle \Psi, \left(D_c \Gamma_a + V \Gamma_a + W_{a,\Psi} \right) \Psi \right\rangle_{(L^2(\mathbb{R}^3, \mathbb{C}^4))^K}, \tag{3.12}$$

with the $K \times K$ Hermitian matrices

$$\begin{aligned} (\Gamma_a)_{i,j} &= N \sum_{k_2 \ldots k_N} \alpha^*_{i,k_2 \ldots k_N} \alpha_{j,k_2 \ldots k_N}, \\ (W_{a,\Psi})_{i,j} &= \frac{N(N-1)}{2} \sum_{k_3 \ldots k_N} \sum_{k,l} \alpha^*_{i,k,k_3 \ldots k_N} \alpha_{j,l,k_3 \ldots k_N} \left(\psi^*_k \psi_l \star \frac{1}{|x|} \right). \end{aligned}$$

The eigenvalues γ_i of Γ_a, for $a \in S$, satisfy $0 \leq \gamma_i \leq 1$, and are called occupation numbers. They measure the total weight of the corresponding orbital in the N-body wave function.

For reference, we define similarly the multiconfiguration Hartree-Fock energy

$$\mathcal{E}^{\mathrm{HF}}(a, \Phi) = \left\langle \Phi, \left(-\frac{1}{2}\Delta\, \Gamma_a + V\Gamma_a + W_{a,\Phi} \right) \Phi \right\rangle_{(L^2(\mathbb{R}^3,\mathbb{C}^2))^K}, \tag{3.13}$$

on $S \times \{\Phi \in (H^1(\mathbb{R}^3, \mathbb{C}^2))^K, \mathrm{Gram}\, \Phi = 1\}$.

One can define a group action on $S \times \Sigma$ that leaves $\mathcal{E}$ invariant : for any unitary matrix $U \in \mathcal{U}(K)$,

$$U \cdot (a, \Psi) = (a', U\Psi), \tag{3.14}$$

where a' is defined via the equivalent variables α' :

$$\alpha'_{i_1,\dots,i_N} = \sum_{j_1,\dots,j_N} (U^*)_{i_1,j_1} \dots (U^*)_{i_N,j_N} \alpha_{j_1,\dots,j_N},$$

where U^* is the adjoint of U. This group action is the multiconfiguration analogue of the well-known unitary invariance of the Hartree-Fock equations.

The MCDF equations, obtained as the Euler-Lagrange equations of $\mathcal{E}$ under the constraints $a \in S$ and $\Psi \in \Sigma$, are, for Ψ and a respectively,

$$H_{a,\Psi}\Psi = \Lambda\Psi, \tag{3.15}$$

$$\mathcal{H}_\Psi a = Ea, \tag{3.16}$$

where

$$H_{a,\Psi} = D_c\Gamma_a + V\Gamma_a + 2W_{a,\Psi} \tag{3.17}$$

is the Fock operator, and

$$(\mathcal{H}_\Psi)_{I,J} = \left\langle \psi_{i_1} \dots \psi_{i_N} \left| H^N \right| \psi_{j_1} \dots \psi_{j_N} \right\rangle \tag{3.18}$$

are the coefficients of the $\binom{K}{N} \times \binom{K}{N}$ matrix of the N-body Hamiltonian H^N in the basis of the Slater determinants. Our goal in this paper is to prove the existence of solutions to (3.15) and (3.16) by finding critical points of $\mathcal{E}$ on $S \times \Sigma$.

3.3 Strategy of proof

There are several major mathematical difficulties in the study of the MCDF model. Unlike in the single-configuration case, one can use the group action (3.14) to diagonalize Γ or Λ, but not both at the same time. Worse, because $W_{a,\Psi}$ does not in general commute with Γ, one can only prove that the Fock operator $H_{a,\Psi}$ has a spectral gap around 0 for values of c that depend on a lower bound on the eigenvalues of Γ. This gap is used centrally to prove the convergence of Palais-Smale sequences. Therefore, one needs a lower bound on Γ.

To obtain this lower bound, we consider the (formal) nonrelativistic limit of the multiconfiguration Dirac-Fock model, the multiconfiguration Hartree-Fock model. Let

$$I^K = \inf \left\{ \mathcal{E}^{\mathrm{HF}}(a, \Phi), a \in S, \Phi \in (H^1(\mathbb{R}^3, \mathbb{C}^2))^K, \operatorname{Gram} \Phi = 1 \right\} \tag{3.19}$$

be the ground-state energy of the nonrelativistic multiconfiguration method of rank $K \geq N$. I^N is the Hartree-Fock energy. I^K is non-increasing, and converges to I^∞, the Schrödinger energy. The behavior of I^K is not precisely known, but a result by Friesecke [Fri03a] shows that $I^{K+2} < I^K$. Therefore, $I^K < I^{K-1}$ at least for one every two K. When this strict inequality holds, every minimizer satisfies $\Gamma > 0$ in the sense of Hermitian matrices. Because of the compactness of these minimizers (implicitly proved in [Lew04]), there is a uniform bound $\gamma_0 > 0$ such that for every minimizer, $\Gamma_a \geq \gamma_0$ in the sense of Hermitian matrices.

Because there is no well-defined "ground state energy" in the relativistic case, we cannot use information of this type directly. Instead, we fix $\gamma < \gamma_0$, and use a min-max principle to look for solutions in the domain

$$S_\gamma = \{a \in S, \Gamma_a \geq \gamma\}.$$

By arguments inspired by [ES99; ES01; Lew04], we prove that the min-max principle yields solutions of $H_{a,\Psi}\Psi = \Lambda\Psi$, for c large enough. But these are only solutions of the equation $\mathcal{H}_\Psi a = Ea$ if the constraint is not saturated, *i.e.* if $\Gamma_a > \gamma$.

To prove that this is the case, we take the nonrelativistic ($c \to \infty$) limit of the critical points found in the first step. By arguments similar to the ones in [ES01], we prove that these critical points converge, up to a subsequence, to a minimizer of the Hartree-Fock functional. Therefore, for c large, the constraint $\Gamma_a \geq \gamma$ is not saturated, and we obtain solutions of the MCDF equations.

In the rest of this paper, we will always assume that $I^K < I^{K-1}$, so that $\Gamma \geq \gamma_0$ on the nonrelativistic minimizers. $\gamma > 0$ is a fixed constant, taken to be less than γ_0. We also assume $N < Z + 1$.

First, for all $\Psi \in (L^2)^K$ such that $\operatorname{Gram} \Psi > 0$, we define the normalization

$$g(\Psi) = (\operatorname{Gram} \Psi)^{-1/2} \Psi. \tag{3.20}$$

This normalization was used in [ES01] to prove another variational principle for the relativistic "ground state", which we shall not use here.

Define

$$\begin{aligned} \Sigma^+ &= \Sigma \cap (E^+)^K, \\ &= \left\{ \Psi \in (E^+)^K, \operatorname{Gram} \Psi = 1 \right\}. \end{aligned}$$

We will find solutions to our equations as a result of the following variational principle:

$$I_{c,\gamma} = \inf_{a \in S_\gamma, \Psi^+ \in \Sigma^+} \sup_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)). \tag{3.21}$$

3.4 Results

Our first result is the well-posedness of our variational principle:

Theorem 3.1. *There are constants $K_1, K_2 > 0$ such that, for c large enough, there is a triplet $a_* \in S_\gamma, \Psi_*^+ \in \Sigma^+, \Psi_*^- \in (E^-)^K$ solution of the variational principle* (3.21)*:*

$$\begin{aligned}\mathcal{E}(a_*, g(\Psi_*^+ + \Psi_*^-)) &= \max_{\Psi^- \in (E^-)^K} \mathcal{E}(a_*, g(\Psi_*^+ + \Psi^-)),\\ &= \min_{a \in S_\gamma, \Psi^+ \in \Sigma^+} \max_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)).\end{aligned}$$

Denoting $\Psi_* = g(\Psi_*^+ + \Psi_*^-)$, Ψ_* is a solution of the equation $H_{a_*,\Psi_*}\Psi_* = \Lambda_*\Psi_*$ in Σ.

The Hermitian matrix of Lagrange multipliers Λ_* satisfies the estimates

$$(c^2 - K_1)\Gamma_* \leq \Lambda_* \leq (c^2 - K_2)\Gamma_*. \tag{3.22}$$

Furthermore, if $\Gamma_* > \gamma$, then a_* is a solution of $\mathcal{H}_{\Psi_*} a_* = I_{c,\gamma} a_*$.

We now study the nonrelativistic limit of these solutions, thanks to the control (3.22) on the Lagrange multipliers:

Theorem 3.2. *Let $c_n \to \infty$, and let (a_n, Ψ_n) be the solution of* (3.21) *obtained by Theorem 3.1 with $c = c_n$. Then, up to a subsequence,*

$$\begin{aligned}a_n &\to a,\\ \Psi_n &\to \begin{pmatrix}\Phi\\ 0\end{pmatrix}\end{aligned}$$

in H^1 norm, where $(a, \Phi) \in S_\gamma \times (H^1(\mathbb{R}^3, \mathbb{C}^2))^K$ is a minimizer of

$$I^K = \inf\left\{\mathcal{E}^{HF}(a, \Phi), a \in S, \Phi \in (H^1(\mathbb{R}^3, \mathbb{C}^2))^K, \operatorname{Gram}\Phi = 1\right\}. \tag{3.23}$$

The min-max level $I_{c,\gamma}$ satisfies the asymptotics

$$I_{c,\gamma} = Nc^2 + I^K + o_{c\to\infty}(1).$$

Since any minimizer of (3.23) must satisfy $\Gamma \geq \gamma_0 > \gamma$, we immediately obtain

Corollary 1. *For c large enough, there are solutions of the multiconfiguration Dirac-Fock equations* (3.15)-(3.16).

The remainder of this paper is dedicated to the proof of Theorems 3.1 and 3.2.

For Theorem 3.1, we first begin with Lemma 3.1, a convergence result for Palais-Smale sequences of the functional $\mathcal{E}$ with Lagrange multipliers bounded away from the essential spectrum of $D_c\Gamma$. Then, at $(a, \Psi^+) \in S_\gamma \times \Sigma^+$ fixed, we study the variational principle

$$\sup_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-))$$

in Lemma 3.2, under the condition that $\mathcal{E}(a, \Psi^+) < Nc^2$. We prove in Lemma 3.6 an upper bound on the asymptotic behavior of $I_{c,\gamma}$ which will enable us to restrict to this domain, and finally, we prove in Lemma 3.7 that Palais-Smale sequences with Morse-type information for the functional

$$\mathcal{F}_a(\Psi^+) = \sup_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-))$$

satisfy the hypotheses of Lemma 3.1, and therefore are precompact. Their limit up to extraction is a solution of our min-max problem (3.21).

To prove Theorem 3.2„ we use the estimates (3.22) on the Lagrange multipliers to prove the compactness of the sequence (a_n, Ψ_n), and the asymptotic behavior from Lemma 3.7 to show that the limit is a minimizer.

3.5 Proof of Theorem 3.1

Our first result is the convergence of Palais-Smale sequences with bounds on the Lagrange multipliers. The proof proceeds as in Lemma 2.1 of [ES99] for the single-configuration case.

3.5.1 Palais-Smale sequences for the energy functional

Lemma 3.1. *For c large enough, if $(a_n, \Psi_n) \in S_\gamma \times \Sigma$ satisfies:*

(i) $H_{a_n,\Psi_n}\Psi_n - \Lambda_n\Psi_n = \Delta_n \to 0$ *in* $H^{-1/2}$ *with* Λ_n *Hermitian matrices,*

(ii) $\liminf \Lambda_n > 0$,

(iii) $\limsup c^2\Gamma_n - \Lambda_n > 0$,

then, up to extraction, $(a_n, \Psi_n) \to (a, \Psi)$ in $S_\gamma \times \Sigma$, where (a, Ψ) is a solution of $H_{a,\Psi}\Psi = \Lambda\Psi$.

Proof.

Step 1 : convergence in $H^{1/2}_{\mathbf{loc}}$ Let $\Psi \in E^K$, and $\Psi^\pm = P^\pm\Psi$. Using the inequality (3.7),

$$\begin{aligned}\left\langle \Psi^+, H_{a_n,\Psi_n}\Psi^+ \right\rangle &\geq \left\langle \Psi^+, \Gamma_n\sqrt{c^4 - c^2\Delta}\Psi^+ \right\rangle + \left\langle \Psi^+, \Gamma_n V\Psi^+ \right\rangle, \\ &\geq \left\langle \Psi^+, \Gamma_n\sqrt{c^4 - c^2\Delta}\Psi^+ \right\rangle - C_1\left\|\Psi^+\right\|_E^2, \\ &\geq (\gamma c^2 - C_1 c)\left\|\Psi^+\right\|_c^2,\end{aligned}$$

where $C_1 > 0$. Similarly,

$$\left\langle \Psi^-, H_{a_n,\Psi_n}\Psi^- \right\rangle \leq -(\gamma c^2 - C_2 c)\left\|\Psi^-\right\|_c^2,$$

with $C_2 > 0$.

Now, Ψ^+ and Ψ^- are orthogonal for the c scalar product, so

$$\|\Psi\|_c^2 = \left\|\Psi^+\right\|_c^2 + \left\|\Psi^-\right\|_c^2 = \left\|\Psi^+ - \Psi^-\right\|_c^2.$$

Denoting by $\|\cdot\|_c^*$ the dual norm of $\|\cdot\|_c$,

$$\begin{aligned}
\left\|H_{a_n,\Psi_n}\Psi\right\|_c^* &\geq \frac{1}{\|\Psi\|_c}\left\langle \Psi^+ - \Psi^-, H_{a_n,\Psi_n}\Psi\right\rangle, \\
&= \frac{1}{\|\Psi\|_c}\left(\left\langle \Psi^+, H_{a_n,\Psi_n}\Psi^+\right\rangle - \left\langle \Psi^-, H_{a_n,\Psi_n}\Psi^-\right\rangle\right), \\
&\geq \frac{1}{\|\Psi\|_c}\left(c^2\gamma - c\max(C_1,C_2)\right)\left(\left\|\Psi^+\right\|_c^2 + \left\|\Psi^-\right\|_c^2\right), \\
&\geq h_0\|\Psi\|_c, \qquad (3.24)
\end{aligned}$$

with $h_0 > 0$ when c is large enough.

We then have

$$\begin{aligned}
\limsup_{n\to\infty}\|\Psi_n\|_c &\leq \limsup_{n\to\infty}\frac{1}{h_0}\left\|H_{a_n,\Psi_n}\Psi_n\right\|_c^*, \\
&\leq \limsup_{n\to\infty}\frac{1}{h_0}\left(\|\Delta_n\|_c^* + \|\Lambda_n\Psi_n\|_{L^2}\right).
\end{aligned}$$

Therefore, Ψ_n is bounded in c norm, *i.e.* in $H^{1/2}$. Extracting a subsequence, again denoted by (a_n,Ψ_n), we may assume that $a_n \to a$, $\Gamma_n \to \Gamma$, $\Lambda_n \to \Lambda$, and $\Psi_n \to \Psi$ weakly in $H^{1/2}$, strongly in $L^p_{\rm loc}$, $2 \leq p < 3$.

Since H_{a,Ψ_n} is self-adjoint from E^K to $(E^K)^*$ and bounded away from zero, it is invertible. Define Ψ'_n by

$$H_{a,\Psi_n}\Psi'_n = \Lambda\Psi_n.$$

Ψ'_n is bounded in $H^{1/2}$, and therefore precompact in $L^p_{\rm loc}$, $2 \leq p < 3$.

We partially invert

$$\Psi'_n = (D_c\Gamma + V\Gamma)^{-1}(\Lambda\Psi_n - 2W_{a,\Psi_n}\Psi'_n).$$

From Young's inequality, $W_{a,\Psi_n}\Psi'_n$ is precompact in $L^p_{\rm loc}$, $1 \leq p < 3$, so $\Lambda\Psi_n - 2W_{a,\Psi_n}\Psi'_n$ is precompact in $L^2_{\rm loc}$. Therefore, Ψ'_n is precompact in $H^{1/2}_{\rm loc}$. We extract again and impose $\Psi'_n \to \Psi$ in $H^{1/2}_{\rm loc}$. But since

$$H_{a,\Psi_n}(\Psi_n - \Psi'_n) \to 0$$

in $H^{-1/2}$, from (3.24), $\Psi_n \to \Psi$ in $H^{1/2}_{\rm loc}$.

Step 2 : convergence in $H^{1/2}$ We now have convergence of Ψ_n to Ψ in $H^{1/2}_{\rm loc}$. Ψ satisfies

$$H_{a,\Psi}\Psi = \Lambda\Psi.$$

We now look at the convergence in $H^{1/2}$ by obtaining an approximate Euler-Lagrange equation satisfied by the error $\varepsilon_n = \Psi_n - \Psi$. We have the Euler-Lagrange equations satisfied by Ψ_n and Ψ:

$$\begin{aligned}(D_c\Gamma + V\Gamma + 2W_{a,\Psi_n})\Psi_n - \Lambda\Psi_n = \Delta'_n,\\ (D_c\Gamma + V\Gamma + 2W_{a,\Psi})\Psi - \Lambda\Psi = 0.\end{aligned}$$

with $\Delta'_n \to 0$ in $H^{-1/2}$. Subtracting and using the fact that $\varepsilon^n \to 0$ weakly in $H^{1/2}$ and strongly in $H^{1/2}_{\mathrm{loc}}$, we get

$$L_n\varepsilon_n \to 0 \tag{3.25}$$

in $H^{-1/2}$, where

$$L_n = D_c\Gamma + 2W_{a,\Psi_n} - \Lambda \tag{3.26}$$

is the Hamiltonian "at infinity" seen by ε_n.

We now use a concavity argument to extract information on the positive and negative components $\varepsilon_n^{\pm} = P^{\pm}\varepsilon_n$ of ε_n separately.

Define the quadratic functional Q_n on $(E^-)^K$ by

$$Q_n(\delta^-) = \left\langle \varepsilon_n + \delta^-, L_n(\varepsilon_n + \delta^-)\right\rangle.$$

The second order terms are

$$\begin{aligned}\left\langle \delta^-, L_n\delta^-\right\rangle &= \left\langle \delta^-, (D_c\Gamma + 2W_{a,\Psi_n} - \Lambda)\delta^-\right\rangle,\\ &\le -(c^2\gamma - C_2c)\left\|\delta^-\right\|_c^2 - \langle \delta_n, \Lambda\delta_n\rangle.\end{aligned} \tag{3.27}$$

Since $\Lambda > 0$, we obtain that Q_n is strictly concave for n large.

The concavity allows us to write

$$\begin{aligned}\left\langle \varepsilon_n^+, L_n\varepsilon_n^+\right\rangle &= Q_n(-\varepsilon_n^-)\\ &\le Q_n(0) - \nabla Q_n(0)[\varepsilon_n^-],\\ &= \langle \varepsilon_n, L_n\varepsilon_n\rangle - 2\left\langle \varepsilon_n^-, L_n\varepsilon_n\right\rangle,\\ &\le 3\|\varepsilon_n\|_E\|L_n\varepsilon_n\|_{E^*}.\end{aligned}$$

Hence

$$\limsup_{n\to\infty}\left\langle \varepsilon_n^+, L_n\varepsilon_n^+\right\rangle \le 0.$$

But

$$\left\langle \varepsilon_n^+, L_n\varepsilon_n^+\right\rangle \ge \left\langle \varepsilon_n^+, (c^2\Gamma - \Lambda)\varepsilon_n^+\right\rangle$$

Since $\Lambda < c^2\Gamma$, this implies convergence to 0 of ε_n^+ in L^2 and then in $H^{1/2}$. But, by (3.25), this implies that $L_n\varepsilon_n^- \to 0$ in $H^{-1/2}$ and therefore that $\left\langle \varepsilon_n^-, L_n\varepsilon_n^-\right\rangle \to 0$. By (3.27), we deduce $\varepsilon_n^- \to 0$ in $H^{1/2}$, which proves that $\Psi_n \to \Psi$ strongly in Σ. □

3.5.2 The reduced functional

For $(a, \Psi^+) \in S_\gamma \times \Sigma^+$, define the functional

$$F_{a,\Psi^+}(\Psi^-) = \mathcal{E}(a, g(\Psi^+ + \Psi^-))$$

on $(E^-)^K$. Our goal in this section is to prove

Lemma 3.2. *There is a constant $M_- > 0$ such that, for c large enough, for all $(a, \Psi^+) \in S_\gamma \times \Sigma^+$ with $\mathcal{E}(a, \Psi^+) \leq Nc^2$, the functional F_{a,Ψ^+} has a unique maximizer $h(a, \Psi^+)$ in $(E^-)^K$. The map h is smooth, and satisfies*

$$\left\| h(a, \Psi^+) \right\|_c \leq \frac{M_-}{c}. \tag{3.28}$$

We first begin with estimates on Ψ^+, for which we use the property $\mathcal{E}(a, \Psi^+) \leq Nc^2$.

Lemma 3.3 (A priori bounds on Ψ^+)**.** *There are $M_+, M_D > 0$ such that, for c large enough, if $(a, \Psi^+) \in S_\gamma \times \Sigma^+$ verifies $\mathcal{E}(a, \Psi^+) \leq Nc^2$, then*

$$\left\| \Psi^+ \right\|_E \leq M_+, \tag{3.29}$$

$$D_c|_{\mathrm{Span}(\{\psi_i^+\})} \leq c^2 + M_D. \tag{3.30}$$

Proof. From Kato's inequality (3.7),

$$\mathcal{E}(a, \Psi^+) \geq \left\langle \Psi^+, D_c \Gamma \Psi^+ \right\rangle - C \left\langle \Psi^+, \sqrt{-\Delta} \Psi^+ \right\rangle. \tag{3.31}$$

Here and in the rest of this paper, C denotes various positive constants independent of c. Since $\mathcal{E}(a, \Psi^+) < Nc^2$ and $\left\langle \Psi^+, \Gamma \Psi^+ \right\rangle = N$,

$$\left\langle \Psi^+, \left(\sqrt{c^4 - c^2\Delta} - c^2 - \frac{C}{\gamma} \sqrt{-\Delta} \right) \Gamma \Psi^+ \right\rangle \leq 0.$$

In the Fourier domain, we can write for all $0 < \alpha < c^2$ by the Cauchy-Schwarz inequality

$$\begin{aligned} \sqrt{c^4 + c^2|\xi|^2} &\geq c^2 \left(1 - \frac{\alpha}{c^2}\right) + c|\xi| \sqrt{1 - \left(1 - \frac{\alpha}{c^2}\right)^2}, \\ &= c^2 - \alpha + |\xi| \sqrt{2\alpha - \frac{\alpha^2}{c^2}}. \end{aligned}$$

Therefore, we obtain

$$\left\langle \Psi^+, \left(-\alpha + \left(\sqrt{\alpha - \frac{\alpha^2}{c^2}} - \frac{C}{\gamma} \right) \sqrt{-\Delta} \right) \Gamma \Psi^+ \right\rangle \leq 0,$$

so

$$\left\langle \Psi^+, \sqrt{-\Delta} \Gamma \Psi^+ \right\rangle \leq \frac{N\alpha}{\sqrt{\alpha - \frac{\alpha^2}{c^2}} - \frac{C}{\gamma}}.$$

Taking $\alpha > \sqrt{C/\gamma}$ and c large, $\left\langle \Psi^+, \sqrt{-\Delta}\Gamma\Psi^+ \right\rangle$ is bounded independently of c. Since $\Gamma \geq \gamma$, so is $\left\|\Psi^+\right\|_E$, and (3.29) is proved.

Now, using (3.31) again along with our new estimate (3.29), we have

$$\begin{aligned} \left\langle \Psi^+, D_c\Gamma\Psi^+ \right\rangle &\leq Nc^2 + CM_+^2, \\ \left\langle \Psi^+, (D_c - c^2)\Gamma\Psi^+ \right\rangle &\leq CM_+^2, \\ \operatorname{tr} A &\leq CM_+^2, \end{aligned}$$

where A is the $K \times K$ Hermitian matrix

$$A_{ij} = \Gamma_{ij} \left\langle \psi_i^+, (D_c - c^2)\psi_j^+ \right\rangle.$$

A is positive semi-definite and its trace is bounded by CM_+^2, so $A \leq CM_+^2$. Since $\Gamma \geq \gamma$, we conclude that

$$\left(\left\langle \psi_i^+, (D_c - c^2)\psi_j^+ \right\rangle \right)_{1 \leq i,j \leq K} \leq \frac{CM_+^2}{\gamma},$$

and therefore (3.30) is proved. □

We now restrict our search for a maximizer to a neighborhood of zero.

Lemma 3.4 (A priori bounds on Ψ^-)**.** *There is a constant $M_- > 0$ such that, for c large enough, for all $(a, \Psi^+) \in S_\gamma \times \Sigma^+$ with $\mathcal{E}(a, \Psi^+) \leq Nc^2$,*

$$\sup_{\Psi^- \in (E^-)^K} F_{a,\Psi^+}(\Psi^-)$$

cannot be achieved outside a neighborhood of zero of size $\frac{M_-}{c}$ in the c norm.

Proof. Let $\Psi^- \in (E^-)^K$, $G = \operatorname{Gram}(\Psi^+ + \Psi^-)$, $\Psi = G^{-1/2}(\Psi^+ + \Psi^-)$. Using (3.30),

$$\begin{aligned} \langle \Psi, D_c\Gamma\Psi \rangle &= (c^2 + M_D) \langle \Psi, \Gamma\Psi \rangle + \left\langle \Psi, (D_c - c^2 - M_D)\Gamma\Psi \right\rangle, \\ &\leq N(c^2 + M_D) + \left\langle G^{-1/2}\Psi^-, (D_c - c^2 - M_D)\Gamma G^{-1/2}\Psi^- \right\rangle, \\ &\leq N(c^2 + M_D) - \gamma c^2 \left\| G^{-1/2}\Psi^- \right\|_c^2. \end{aligned}$$

On the other hand,

$$\begin{aligned} \left\langle \Psi, (V\Gamma + 2W_{a,\Psi})\Psi \right\rangle &\leq C \left\| G^{-1/2}\Psi^+ \right\|_E^2 + C \left\| G^{-1/2}\Psi^- \right\|_E^2, \\ &\leq CM_+ + Cc \left\| G^{-1/2}\Psi^- \right\|_c^2. \end{aligned}$$

All together,

$$F_{a,\Psi^+}(\Psi^-) \leq Nc^2 + NM_D + CM_+ - \left(\gamma c^2 - Cc\right) \left\| G^{-1/2}\Psi^- \right\|_c^2.$$

But we also have

$$\begin{aligned} F_{a,\Psi^+}(0) &= \mathcal{E}(a,\Psi^+) \\ &\geq \left\langle \Psi^+, (D_c\Gamma + V\Gamma)\Psi^+ \right\rangle, \\ &\geq Nc^2 - C\left\|\Psi^+\right\|_E^2, \\ &\geq Nc^2 - CM_+^2. \end{aligned}$$

Therefore,

$$F_{a,\Psi^+}(\Psi^-) \leq F_{a,\Psi^+}(0) + NM_D + 2CM_+^2 - \left(\gamma c^2 - Cc\right)\left\|G^{-1/2}\Psi^-\right\|_c^2.$$

So, in order to have $F_{a,\Psi^+}(\Psi^-) \leq F_{a,\Psi^+}(0)$, Ψ^- must satisfy

$$\begin{aligned} \left\|G^{-1/2}\Psi^-\right\|_c^2 &\leq O(1/c^2), \\ \left\|\Psi^-\right\|_c^2 &\leq O(1/c^2)(1+\|\Psi_-\|_{L^2}^2), \\ &\leq O(1/c^2)(1+\|\Psi_-\|_c^2), \end{aligned}$$

and therefore

$$\left\|\Psi^-\right\|_c^2 = O(1/c^2).$$

□

Restricting now to this domain, we prove that F_{a,Ψ^+} is strictly concave:

Lemma 3.5. *For c large, for all $(a,\Psi^+) \in S_\gamma \times \Sigma$ such that $\mathcal{E}(a,\Psi^+) \leq Nc^2$, for all Ψ^- in the region $\left\|\Psi^-\right\|_c \leq \frac{M_-}{c}$, for all $\Phi^- \in (E^-)^K$,*

$$F''_{a,\Psi^+}(\Psi^-)[\Phi^-,\Phi^-] \leq -\frac{\gamma c^2}{2}\left\|\Phi^-\right\|_c^2.$$

Proof. We have $g(\Psi^+ + \Psi^-) = (1 + \frac{1}{c}B(\Psi^-))(\Psi^+ + \Psi^-)$, where B is a matrix-valued function. In the region $\left\|\Psi^-\right\|_c \leq \frac{M_-}{c}$, B and its derivatives are bounded independently of c. Let Ψ^- be such that $\left\|\Psi^-\right\|_c \leq \frac{M_-}{c}$, and define $G = \mathrm{Gram}(\Psi^+ + \Psi^-)$. Then, for all $\Phi^- \in (E^-)^K$,

$$\begin{aligned} \frac{1}{2}F''_{a,\Psi^+}(\Psi^-)[\Phi^-,\Phi^-] &= \partial_\Psi \mathcal{E}(a,\Psi)\left[\frac{1}{c}B'(\Psi^-)[\Phi^-]\Phi^- + \frac{1}{2c}B''(\Psi^-)[\Phi^-,\Phi^-](\Psi^+ + \Psi^-)\right] \\ &\quad + \frac{1}{2}\partial_\Psi^2 \mathcal{E}(a,\Psi)\left[G^{-1/2}\Phi^- + \frac{1}{c}B'(\Psi^-)[\Phi^-](\Psi^+ + \Psi^-)\right]^2, \\ &= \frac{1}{2}\partial_\Psi^2 \mathcal{E}(a,\Psi)[\Phi^-,\Phi^-] + O\left(c\left\|\Phi^-\right\|_c^2\right). \end{aligned}$$

But we also have

$$\begin{aligned} \frac{1}{2}\partial_\Psi^2 \mathcal{E}(a,\Psi)[\Phi^-,\Phi^-] &= -\left\langle \Phi^-, \sqrt{c^4 - c^2\Delta}\Gamma\Phi^- \right\rangle + O\left(\left\|\Phi^-\right\|_E^2\right), \\ &\leq \left(-\gamma c^2 + O(c)\right)\left\|\Phi^-\right\|_c^2. \end{aligned}$$

and so the result follows for c large.

□

Lemma 3.2 is now proved as a direct consequence of Lemmas 3.3, 3.4 and 3.5.

3.5.3 Asymptotic behavior of $I_{c,\gamma}$

In order to restrict to the domain $\mathcal{E}(a,\Psi^+) \leq Nc^2$, we prove that solutions of our min-max principle have to be in this domain for c large:

Lemma 3.6.

$$I_{c,\gamma} \leq Nc^2 + I^K + o_{c\to\infty}(1).$$

In particular, for c large enough,

$$\begin{aligned} I_{c,\gamma} &= \inf_{a\in S_\gamma, \Psi^+\in\Sigma^+} \sup_{\Psi^-\in(E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)), \\ &= \inf_{\substack{a\in S_\gamma, \Psi^+\in\Sigma^+, \\ \mathcal{E}(a,\Psi^+)<Nc^2}} \sup_{\Psi^-\in(E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)). \end{aligned} \tag{3.32}$$

Proof. Let $(a_*, \Phi_*) \in S_\gamma \times H^1(\mathbb{R}^3, \mathbb{C}^2)$ be a minimizer of the nonrelativistic multi-configuration Hartree-Fock functional, and $\Psi_* = \begin{pmatrix} \Phi_* \\ 0 \end{pmatrix}$. Set

$$\Psi_*^+ = g(P^+\Psi_*).$$

Ψ_*^+ belongs to Σ^+, and converges in H^1 to Ψ_* as $c \to \infty$. From the concavity inequality

$$\sqrt{1+|\xi|^2} \leq 1 + \frac{1}{2}|\xi|^2$$

in the Fourier domain, we get

$$\begin{aligned} \mathcal{E}(a_*, \Psi_*^+) &= \left\langle \Psi_*^+, (\sqrt{c^4 - c^2\Delta}\Gamma + V\Gamma + W_{a,\Psi_*^+})\Psi_*^+ \right\rangle, \\ &\leq \left\langle \Psi_*^+, \left(c^2\Gamma - \frac{1}{2}\Delta\Gamma + V\Gamma + W_{a,\Psi_*^+} \right) \Psi_*^+ \right\rangle, \\ &= Nc^2 + \mathcal{E}^{\mathrm{HF}}(a_*, \Psi_*) + o_{c\to\infty}(1), \\ &= Nc^2 + I^K + o_{c\to\infty}(1). \end{aligned}$$

From Lemmas 3.2 and 3.5, we now have

$$\begin{aligned} I_{c,\gamma} &\leq F_{a_*,\Psi_*^+}(h(\Psi_*^+)), \\ &\leq \mathcal{E}(a_*, \Psi_*^+) + F'_{a_*,\Psi_*^+}(0)[h(\Psi_*^+)] - \frac{\gamma c^2}{2}\left\|h(\Psi_*^+)\right\|_c^2, \\ &\leq \mathcal{E}(a_*, \Psi_*^+) + C\left\|h(\Psi_*^+)\right\|_{L^2}, \\ &\leq \mathcal{E}(a_*, \Psi_*^+) + O\left(\frac{1}{c}\right), \end{aligned}$$

so that

$$I_{c,\gamma} \leq Nc^2 + I^K + o_{c\to\infty}(1).$$

Since $I^K < 0$ and, for all $a \in S_\gamma$, $\Psi^+ \in \Sigma$,

$$\sup_{\Psi^-\in(E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)) \geq \mathcal{E}(a, \Psi^+),$$

(3.32) holds and the lemma is proved. □

3.5.4 Borwein-Preiss sequences for the reduced functional

Let

$$S'_\gamma = \{a \in S_\gamma, \inf_{\Psi^+ \in \Sigma^+} \mathcal{E}(a, \Psi^+) < Nc^2\}.$$

For $a \in S'_\gamma$ fixed, we minimize the functional

$$\mathcal{F}_a(\Psi^+) = \mathcal{E}(a, g(\Psi^+ + h(a, \Psi^+)))$$

on the manifold

$$\Sigma_a^+ = \{\Psi^+ \in \Sigma^+, \mathcal{E}(a, \Psi^+) < Nc^2\}.$$

For all $\Psi^+ \in \Sigma_a^+, \Psi \in \Sigma$, define the tangent spaces

$$T_{\Psi^+}\Sigma_a^+ = \{\Phi^+ \in (E^+)^K, \left\langle \phi_i^+, \psi_j^+ \right\rangle = 0 \text{ for all } i,j \in \{1,\dots,K\}\},$$
$$T_\Psi \Sigma = \{\Phi \in E^K, \left\langle \phi_i, \psi_j \right\rangle = 0 \text{ for all } i,j \in \{1,\dots,K\}\}.$$

Lemma 3.7. *There are constants $K_1 > 0, K_2 > 0$ such that, for all c large enough, $a \in S'_\gamma$, if $\Psi_n^+ \in \Sigma_a^+$ is a Borwein-Preiss sequence for $\mathcal{F}_a$ on Σ_a^+, i.e. satisfies*

(i) $\mathcal{F}_a(\Psi_n^+) \to \inf_{\Psi^+ \in \Sigma_a^+} \mathcal{F}_a(\Psi^+)$,

(ii) $\mathcal{F}'_a(\Psi_n^+)\Big|_{T_{\Psi_n^+}\Sigma_a^+} \to 0$ *in* $H^{-1/2}$,

(iii) There is a sequence $\beta_n \to 0$ such that the quadratic form $\Phi^+ \to \mathcal{F}''_a(\Psi_n^+)[\Phi^+,\Phi^+] + \beta_n \left\|\Phi^+\right\|_E^2$ is non-negative on $T_{\Psi_n^+}\Sigma_a^+$,

then, denoting $\Psi_n = g(\Psi_n^+ + h(a, \Psi_n^+))$,

1. *There is a sequence of Hermitian matrices Λ_n such that $H_{a,\Psi_n}\Psi_n - \Lambda_n\Psi_n = \Delta_n \to 0$ in $H^{-1/2}$,*
2. $\limsup \Lambda_n \le (c^2 - K_2)\Gamma$,
3. $\liminf \Lambda_n \ge (c^2 - K_1)\Gamma$.

Proof. Define

$$k(\Psi^+, \Psi^-) = g(\Psi^+ + \Psi^-)$$

on $\Sigma_a^+ \times (E^-)^K$. From hypothesis *(ii)* and the definition of h, $(\Psi_n^+, h(\Psi_n^+))$ is a Palais-Smale sequence for $\mathcal{E}(a, k(\cdot))$. But $k'(\Psi_n^+, h(\Psi_n^+))$ is an isomorphism from $T_{\Psi_n^+}\Sigma_a^+ \times (E^-)^K$ to $T_{\Psi_n}\Sigma$, so that $\Psi_n = k(\Psi_n^+, h(\Psi_n^+))$ is a Palais-Smale sequence for $\mathcal{E}$ on Σ, and *(1)* is proved.

Upper bound on the Lagrange multipliers

Let us now prove the upper bound *(2)*. First, note that, since $a \in S'_\gamma$ and $\Psi_n^+ \in \Sigma_a^+$, the a priori estimates of Lemmas 3.2 and 3.3 hold.

Let $\delta_n \in T_{\Psi_n^+}\Sigma^+$. Let $\Psi_n^- = h(\Psi_n^+)$, and $G_n = \mathrm{Gram}(\Psi_n^+ + \Psi_n^-) = 1 + \mathrm{Gram}\,\Psi_n^-$. For ε small enough, define the curve on Σ_a^+

$$\Psi_n^+(\varepsilon) = G_n^{1/2}\left(\frac{(G_n^{-1/2}\Psi_n^+)_1 + \varepsilon\delta_n}{\sqrt{1+\varepsilon^2}}, (G_n^{-1/2}\Psi_n^+)_2, \ldots, (G_n^{-1/2}\Psi_n^+)_K\right).$$

Define the associated

$$\begin{aligned}
\Psi_n^-(\varepsilon) &= h(\Psi_n^+(\varepsilon)),\\
G_n(\varepsilon) &= \mathrm{Gram}(\Psi_n^+(\varepsilon) + \Psi_n^-(\varepsilon)) = 1 + \mathrm{Gram}\,\Psi_n^-(\varepsilon),\\
\Psi_n(\varepsilon) &= G_n^{-1/2}(\varepsilon)(\Psi_n^+(\varepsilon) + \Psi_n^-(\varepsilon)),
\end{aligned}$$

and the infinitesimal increments

$$\begin{aligned}
\Phi_n^+ &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Psi_n^+(\varepsilon)\Big|_{\varepsilon=0},\\
&= G_n^{1/2}(\delta_n, 0, \ldots, 0),\\
\Phi_n^- &= h'(\Psi_n)[\Phi_n^+].
\end{aligned}$$

Step 1 Our first step is a control on Φ_n^-.

Define

$$\begin{aligned}
\mathcal{G}(\Psi^+, \Psi^-) &= F_{a,\Psi^+}(\Psi^-),\\
&= \mathcal{E}(a, g(\Psi^+ + \Psi^-)).
\end{aligned}$$

Now, for all $\Psi^+ \in \Sigma_a^+$, $\Phi^- \in (E^-)^K$,

$$\partial_{\Psi^-}\mathcal{G}(\Psi^+, h(\Psi^+))[\Phi^-] = 0.$$

Differentiating with respect to Ψ^+, we get, for all $\Phi^+ \in T_{\Psi^+}\Sigma$,

$$\partial^2_{\Psi^-\Psi^+}\mathcal{G}(\Psi^+, h(\Psi^+))[\Phi^-, \Phi^+] + \partial^2_{\Psi^-\Psi^-}\mathcal{G}(\Psi^+, h(\Psi^+))[\Phi^-, h'(\Psi^+)[\Phi^+]] = 0,$$

and therefore, from the definition of $\mathcal{G}$,

$$-F''_{a,\Psi^+}(\Psi^-)[\Phi^-, h'(\Psi^+)[\Phi^+]] = \partial^2_{\Psi^-\Psi^+}\mathcal{G}(\Psi^+, h(\Psi^+))[\Phi^-, \Phi^+].$$

We now apply this to $\Psi^+ = \Psi_n^+$, $\Psi^- = \Psi_n^-$, $\Phi^+ = \Phi_n^+$ and $\Phi^- = \Phi_n^-$, and get

$$-F''_{a,\Psi_n^+}(\Psi_n^-)[\Phi_n^-, \Phi_n^-] = \partial^2_{\Psi^-\Psi^-}\mathcal{G}(\Psi_n^+, \Psi_n^-)[\Phi_n^-, \Phi_n^-]. \tag{3.33}$$

Using estimates similar to but slightly more complicated than those in [ES99],

$$\partial^2_{\Psi^-\Psi^-}\mathcal{G}(\Psi_n^+, \Psi_n^-)[\Phi_n^-, \Phi_n^-] \le O\left(\left\|\nabla\Phi_n^+\right\|_{L^2}\left\|\Phi_n^-\right\|_{L^2}\right),$$

where the notation O is for both c and n large.

But, by Lemma 3.5,

$$F''_{a,\Psi_n^+}(\Psi_n^-)[\Phi_n^-,\Phi_n^-] \leq -\frac{\gamma c^2}{2}\left\|\Phi_n^-\right\|_c^2,$$

from which we conclude, from (3.33), that

$$\frac{\gamma c^2}{2}\left\|\Phi_n^-\right\|_c^2 \leq O\left(\left\|\nabla\Phi_n^+\right\|_{L^2}\left\|\Phi_n^-\right\|_{L^2}\right),$$

and therefore that

$$\left\|\Phi_n^-\right\|_c \leq \frac{1}{c^2}O\left(\left\|\nabla\Phi_n^+\right\|_{L^2}\right). \tag{3.34}$$

Step 2 We now write the Hessian of $\mathcal{E}$ along the curve $\Psi_n^+(\varepsilon)$.

First, we compute

$$\begin{aligned}\Phi_n &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Psi_n(\varepsilon)\bigg|_{\varepsilon=0}\\ &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}G_n^{-1/2}(\varepsilon)\left(\Psi_n^+(\varepsilon)+\Psi_n^-(\varepsilon)\right)\bigg|_{\varepsilon=0}\\ &= (\delta_n,0,\dots,0)+R_n^+ + R_n^-,\end{aligned} \tag{3.35}$$

where, using (3.34), we can estimate the remainder terms $R_n^\pm \in E^\pm$ as

$$\begin{aligned}\left\|R_n^+\right\|_c &= O\left(\frac{1}{c^4}\|\nabla\delta_n\|_{L^2}\right),\\ \left\|R_n^-\right\|_c &= O\left(\frac{1}{c^2}\|\nabla\delta_n\|_{L^2}\right).\end{aligned}$$

Using these estimates and the same arguments as in [Lew04], we can now compute

$$\begin{aligned}\mathcal{E}''(\Psi_n)[\Phi_n,\Phi_n] \leq \Gamma_{11}&\left(c^2\|\delta_n\|_c^2+\left\langle\delta_n,(V+\rho_n\star\frac{1}{|x|})\delta_n\right\rangle\right)\\ &+O\left(\frac{1}{c^2}\|\nabla\delta_n\|_{L^2}^2+\frac{1}{c^2}\|\nabla\delta_n\|_{L^2}\|\delta_n\|_{L^2}\right),\end{aligned}$$

with $\int\rho_n = N-1$.

Now, let U be an arbitrary vector subspace of $H^1(\mathbb{R}^3,\mathbb{C}^4)$ consisting of functions of the form

$$\begin{pmatrix} f(|x|)\\ 0\\ 0\\ 0\end{pmatrix}$$

with dimension at least $K+1$. Let U_λ^+ be the positive projection of the dilation of U of a factor λ, *i.e.*

$$U_\lambda = P^+ \left\{ \psi\left(\frac{x}{\lambda}\right), \psi \in U \right\}.$$

U_λ^+ is also of dimension $K+1$ for c large enough, so we can find a function $\delta_n \in U_\lambda^+$ normalized in L^2 which is orthogonal to Ψ_n^+. For such a function,

$$\mathcal{E}''(\Psi_n)[\Phi_n, \Phi_n] \leq \left(c^2 - \eta \frac{Z-N-1}{\lambda} + O\left(\frac{1}{\lambda^2} + \frac{1}{\lambda c^2}\right)\right) \Gamma_{11},$$

with $\eta > 0$, where the O notation is understood for n, c and λ large. So, taking λ large enough independently of n and c, we get

$$\mathcal{E}''(\Psi_n)[\Phi_n, \Phi_n] \leq \left(c^2 - \kappa\right) \Gamma_{11}, \tag{3.36}$$

with $\kappa > 0$ independent of n and c.

Step 3 Using (3.35) again, we estimate

$$\langle \Phi_n, \Lambda_n \Phi_n \rangle = (\Lambda_n)_{11} + O\left(\frac{1}{c^4}\Lambda_n\right).$$

But we can obtain a very crude control on Λ_n thanks to the estimates in Lemmas 3.3 and 3.4 :

$$\begin{aligned}(\Lambda_n)_{ij} &= \left\langle \Psi_n^i, H_{a,\Psi_n} \Psi_n^j \right\rangle + o_{n\to\infty}(1), \\ &= c^2 \Gamma_{ij} + O(\|\Psi_n\|_E^2) + o_{n\to\infty}(1),\end{aligned}$$

and therefore

$$\Lambda_n = c^2 \Gamma + O(1). \tag{3.37}$$

Step 4 We now use the second order condition to conclude.

We have $\langle \Psi_n(\varepsilon), \Lambda_n \Psi_n(\varepsilon) \rangle = \operatorname{tr} \Lambda_n$, so, defining the Lagrangian $L_n(\Psi) = \mathcal{E}(a, \Psi) - \langle \Psi, \Lambda_n \Psi \rangle$, we get from the second order information *(iii)* that

$$\beta_n \left\| \Psi_n^+ \right\|_E^2 \leq L_n''(\Psi_n)\left[\Phi_n, \Phi_n\right] + O\left(\left\| \frac{\mathrm{d}^2}{\mathrm{d}^2 \varepsilon} \Psi_n(\varepsilon)\Big|_{\varepsilon=0} \right\|_{H^{1/2}} \left\| \mathcal{F}'(\Psi_n^+) \right\|_{H^{-1/2}} \right).$$

Therefore, from the Palais-Smale condition *(ii)*

$$\mathcal{E}''(\Psi_n)[\Phi_n, \Phi_n] \geq \langle \Phi_n, \Lambda_n \Phi_n \rangle + o_{n\to\infty}(1).$$

Finally, from (3.36) and (3.37), we obtain, for c and n large enough,

$$(\Lambda_n)_{11} \leq (c^2 - K_2)\Gamma_{11},$$

with $K_2 > 0$.

Using the group action (3.14), we could apply the same procedure to $(\widetilde{a}, \widetilde{\Psi}_n^+) = U \cdot (a, \Psi_n^+)$ for any $U \in \mathcal{U}(K)$, and obtain

$$(U\Lambda_n U^*)_{11} \leq (c^2 - K_2)(U\Gamma U^*)_{11},$$

which proves our result

$$\Lambda_n \leq (c^2 - K_2)\Gamma.$$

Lower bound on the Lagrange multipliers

Let $A_n = (c^2 - K_2)\Gamma - \Lambda_n$. We know that, for n large enough, $A_n \geq 0$, and, from (3.37),

$$\begin{aligned} \operatorname{tr} A_n &= Nc^2 - NK_2 - \operatorname{tr} \Lambda_n, \\ &= O(1). \end{aligned}$$

So $A_n = O(1)$, and therefore $\Lambda_n \geq (c^2 - K_2)\Gamma - O(1)$. Because $\Gamma \geq \gamma > 0$, the result follows for c large. □

3.5.5 Proof of Theorem 3.1

For any $a \in S'_\gamma$, we can apply the Borwein-Preiss variational principle [BP87] to the functional $\mathcal{F}_a$ on Σ_a^+, and obtain a sequence Ψ_n^+ that satisfies the hypotheses of Lemma 3.7. The associated sequence Ψ_n satisfies the hypotheses of Lemma 3.1 so, the sequence (a, Ψ_n^+) converges up to extraction to a limit Ψ_a^+, solution of the min-max principle

$$\mathcal{F}_a(\Psi_a^+) = \min_{\Psi^+ \in \Sigma^+} \max_{\Psi^- \in (E^-)^K} \mathcal{E}(a, g(\Psi^+ + \Psi^-)).$$

We now take a minimizing sequence a_n for the continuous functional $F_a(\Psi_a^+)$ on S'_γ. The sequence $(a_n, \Psi_{a_n}^+)$ again verifies the hypotheses of Lemma 3.1, and therefore converges to (a_*, Ψ_*^+). The triplet $(a_*, \Psi_*^+, h(\Psi_*^+))$ is now a solution of the variational principle (3.21), and Theorem 3.1 is proved.

3.6 Proof of Theorem 3.2

3.6.1 Nonrelativistic limit

We begin with a lemma that is the multiconfiguration analogue of Theorem 3 of [ES01].

Lemma 3.8. *Let $c_n \to \infty, (a_n, \Psi_n) \in S_\gamma \times \Sigma$ solutions of*

$$H_{a_n, \Psi_n} \Psi_n = \Lambda_n \Psi_n \tag{3.38}$$

such that

$$({c_n}^2 - K_1)\Gamma_n \leq \Lambda_n \leq ({c_n}^2 - K_2)\Gamma_n$$

for constants $K_1, K_2 > 0$.

Then, up to a subsequence, $a_n \to a \in S_\gamma$, $\Psi_n \to \begin{pmatrix} \Phi \\ 0 \end{pmatrix}$ *in* H^1, *and* $\mathcal{E}(a_n, \Psi_n) - Nc^2 \to \mathcal{E}^{HF}(a, \Phi)$

Proof. First, we need a uniform bound on Ψ_n in H^1.

$$\begin{aligned} \|D_c\Gamma_n\Psi_n\|_{L^2}^2 &= \left\langle \Gamma_n\Psi_n, (c^4 - c^2\Delta)\Gamma_n\Psi_n \right\rangle, \\ &= {c_n}^4\|\Gamma_n\Psi_n\|_{L^2}^2 + {c_n}^2\|\Gamma_n\nabla\Psi_n\|_{L^2}^2 . \end{aligned}$$

On the other hand,

$$\begin{aligned} \|D_c\Gamma_n\Psi_n\|_{L^2}^2 &= \left\|(V\Gamma + 2W_{a_n,\Psi_n})\Psi_n - \Lambda_n\Psi_n\right\|_{L^2}^2 , \\ &\leq {c_n}^4\|\Gamma_n\Psi_n\|_{L^2}^2 + C\|\nabla\Psi_n\|_{L^2}^2 + C{c_n}^2\|\Gamma_n\nabla\Psi_n\|_{L^2} . \end{aligned}$$

by the classical Hardy inequality, with $C > 0$. Therefore, Ψ_n is bounded in H^1.

We now write $\Psi_n = \begin{pmatrix} \Phi_n \\ \mathcal{X}_n \end{pmatrix}$, where $\Phi_n, \mathcal{X}_n \in H^1(\mathbb{R}^3, \mathbb{C}^2)$. We rewrite the equations (3.38) as

$$c_n\Gamma_n L\mathcal{X}_n + (V\Gamma_n + 2W_{a_n,\Psi_n})\Phi_n = (\Lambda_n - {c_n}^2\Gamma_n)\Phi_n, \tag{3.39}$$

$$c_n\Gamma_n L\Phi_n + (V\Gamma_n + 2W_{a_n,\Psi_n})\mathcal{X}_n = (\Lambda_n + {c_n}^2\Gamma_n)\mathcal{X}_n, \tag{3.40}$$

with the operator

$$L = -i\nabla \cdot \sigma \tag{3.41}$$

such that $L^2 = -\Delta$.

Because $\Lambda_n < ({c_n}^2 - K_2)\Gamma_n$, using the Hardy inequality and the boundedness of Φ_n in H^1, the first equation (3.39) yields

$$\|\Gamma_n L\mathcal{X}_n\|_{L^2} = \|\Gamma_n\nabla\mathcal{X}_n\|_{L^2} = O(1/c_n). \tag{3.42}$$

The second equation (3.40) gives

$$\begin{aligned} \mathcal{X}_n &= \frac{1}{2c}\left(\frac{1}{2}\left(\Gamma_n + \Lambda_n/{c_n}^2\right)\right)^{-1}\Gamma_n L\Phi_n + \frac{1}{{c_n}^2}O(\|\mathcal{X}_n\|_{H^1}) \\ &= \mathrm{KB}(\Phi_n) + \frac{1}{{c_n}^2}O(\|\mathcal{X}_n\|_{H^1}) + O\left(\frac{1}{{c_n}^3}\right) \end{aligned} \tag{3.43}$$

in L^2 norm, where the "kinetic balance" operator KB is given by

$$\mathrm{KB}(\Phi) = \frac{1}{2c}L\Phi. \tag{3.44}$$

Equation (3.43) gives $\|\mathcal{X}_n\|_{L^2} = \frac{1}{2c_n}\|L\Phi_n\|_{L^2} + O(1/{c_n}^2) = O(1/c_n)$, and then

$$\mathcal{X}_n = \mathrm{KB}(\Phi_n) + O\left(\frac{1}{{c_n}^3}\right) \tag{3.45}$$

again in L^2 norm. Φ_n satisfies

$$\left(-\frac{1}{2}\Delta\Gamma_n + V\Gamma_n + 2W_{\Phi_n}\right)\Phi_n = (\Lambda_n - {c_n}^2\Gamma_n)\Phi_n + \Delta_n$$
$$\mathrm{Gram}\,\Phi_n = 1 + o(1)$$

with $\Delta_n \to 0$ in L^2 and therefore H^{-1} norm. $(a_n, g(\Phi_n))$ is a Palais-Smale sequence for the nonrelativistic functional, with control on the Lagrange multipliers $(\Lambda_n - {c_n}^2\Gamma_n) < 0$ and non-degeneracy information $\Gamma_n \geq \gamma$. By the arguments in the proof of Theorem 1 of [Lew04], (a_n, Φ_n) converges, up to a subsequence, to (a, Φ) in H^1 norm, and it is easy to compute from (3.45) that

$$\langle \Psi_n, D_{c_n}\Gamma_n\Psi_n\rangle = N{c_n}^2 + \frac{1}{2}\langle \Phi_n, (-\Delta)\Gamma_n\Phi_n\rangle + o(1),$$

and the result follows. □

We are now ready to prove Theorem 3.2.

3.6.2 Proof of Theorem 3.2

Proof. The sequence (a_n, Ψ_n) satisfies the hypotheses of Lemma 3.8 : up to a subsequence, it converges strongly in H^1 to $\left(a, \begin{pmatrix}\Phi\\0\end{pmatrix}\right)$, with $\lim \mathcal{E}(a_n, \Psi_n) - N{c_n}^2 = \mathcal{E}^{\mathrm{HF}}(a, \Phi)$. But since by Lemma 3.6 we have

$$\mathcal{E}(a_n, \Psi_n) = I_{c_n,\gamma} \leq Nc_n^2 + I^K + o_{c_n\to\infty}(1),$$

we obtain $\mathcal{E}^{\mathrm{HF}}(a, \Phi) = I^K$, hence the result. □

Acknowledgements

I would like to thank Éric Séré for his attention and helpful advice.

Bibliography

[Bar+10] C. Bardos et al. Setting and analysis of the multi-configuration time-dependent Hartree-Fock equations. *Archive for Rational Mechanics and Analysis* 198.1 (2010), pp. 273–330.

[BP87] J.M. Borwein and D. Preiss. A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions. *Trans. Amer. Math. Soc* 303.51 (1987), pp. 7–527.

[BES06] B. Buffoni, M.J. Esteban, and E. Séré. Normalized solutions to strongly indefinite semilinear equations. *Adv. Nonlinear Stud.* 6 (2 2006), pp. 323–347.

[Der12] J. Derezinski. Open problems about many-body Dirac operators. *Bulletin of International Association of Mathematical Physics* (2012).

[DFJ07] K.G. Dyall and K. Faegri Jr. *Introduction to relativistic quantum chemistry.* Oxford University Press, 2007.

[ES99] M.J. Esteban and E. Séré. Solutions of the Dirac-Fock equations for atoms and molecules. *Communications in Mathematical Physics* 203.3 (1999), pp. 499–530.

[ES01] M.J. Esteban and E. Séré. Nonrelativistic limit of the Dirac-Fock equations. *Annales Henri Poincare* 2 (5 2001), pp. 941–961.

[Fri03a] G. Friesecke. On the infinitude of non-zero eigenvalues of the single-electron density matrix for atoms and molecules. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 459.2029 (2003), pp. 47–52.

[Fri03b] G. Friesecke. The multiconfiguration equations for atoms and molecules: charge quantization and existence of solutions. *Archive for Rational Mechanics and Analysis* 169.1 (2003), pp. 35–71.

[Gra07] I. Grant. *Relativistic Quantum Theory of Atoms and Molecules.* Springer, 2007.

[Her77] I. Herbst. Spectral theory of the operator $(p^2 + m^2)^{-1/2} - Ze^2/r$. *Communications in Mathematical Physics* 53.3 (1977), pp. 285–294.

[ID93] P. Indelicato and J.P. Desclaux. Projection operator in the multiconfiguration Dirac-Fock method. *Physica Scripta* 1993.T46 (1993), p. 110.

[Kat66] T. Kato. *Perturbation theory for linear operators.* Springer Verlag, 1966.

[LB94] C. Le Bris. A general approach for multiconfiguration methods in quantum molecular chemistry. *Annales de l'Institut Henri Poincaré. Analyse non linéaire* 11.4 (1994), pp. 441–484.

[Lew04] M. Lewin. Solutions of the multiconfiguration equations in quantum chemistry. *Archive for Rational Mechanics and Analysis* 171.1 (2004), pp. 83–114.

[LS77] E.H. Lieb and B. Simon. The Hartree-Fock theory for Coulomb systems. *Comm. Math. Phys.* 53.3 (1977), pp. 185–194.

[Lio87] P.L. Lions. Solutions of Hartree-Fock equations for Coulomb systems. *Communications in Mathematical Physics* 109.1 (1987), pp. 33–97.

[PD79] P. Pyykko and J.P. Desclaux. Relativity and the periodic system of elements. *Accounts of Chemical Research* 12.8 (1979), pp. 276–281.

[SO89] A. Szabo and N.S. Ostlund. *Modern quantum chemistry*. McGraw-Hill New York, 1989.

[Tix98] C. Tix. Strict positivity of a relativistic Hamiltonian due to Brown and Ravenhall. *Bulletin of the London Mathematical Society* 30.3 (1998), pp. 283–290.

Chapitre 4

Constantes optimales pour Lieb-Thirring

Ce chapitre reprend le texte intégral de l'article "Best constants in Lieb-Thirring inequalities : a numerical investigation", accepté pour publication dans le Journal of Spectral Theory.

Best constants in Lieb-Thirring inequalities: a numerical investigation

Antoine Levitt

Abstract

We investigate numerically the optimal constants in Lieb-Thirring inequalities by studying the associated maximization problem. Using a monotonic fixed-point algorithm and a finite element discretization, we obtain radial trial potentials which provide lower bounds on the optimal constants. These results confirm existing conjectures, and provide insight into the behavior of the maximizers. Based on our numerical results, we formulate a complete conjecture about the best constants for all possible values of the parameters.

4.1 Introduction

In this paper we study the family of Lieb-Thirring inequalities [LT75; LT76], which state that for any potential $V \in L^{\gamma+d/2}(\mathbb{R}^d, \mathbb{R})$ and non-negative exponent γ,

$$\operatorname{tr}((-\Delta + V)_-^\gamma) \leq L_{\gamma,d} \int V_-^{\gamma+d/2}, \tag{4.1}$$

where $V_- = \max(-V, 0)$ is the negative part of V, and $(-\Delta + V)_-^\gamma$ is the γ-th power of the negative part of the operator $(-\Delta + V)$, as defined by the functional calculus. In other words, $\operatorname{tr}((-\Delta + V)_-^\gamma) = \sum_i |\lambda_i|^\gamma$ is the γ-th moment of the negative eigenvalues of $-\Delta + V$.

This inequality was originally used by Lieb and Thirring in the case $\gamma = 1, d = 3$ as an important tool to prove the stability of fermionic matter in [LT75] (see [Lie90; LS10] for an overview). The generalization (4.1) to any γ and d has since then attracted a great deal of attention (see for instance [LW00a] for a review). Of particular interest are the values of γ and d for which this inequality holds, and the value of the optimal constants $L_{\gamma,d}$.

Despite the physical significance of the Lieb-Thirring inequality and the amount of mathematical research on the subject, a number of questions are still open. This paper aims to investigate some of these numerically. To our knowledge, this has not been done since the work of Barnes in the appendix of the original paper by Lieb and Thirring [LT76], in 1976.

It is easy to see that the bound (4.1) cannot hold for $\gamma < 1/2$ when $d = 1$ by scaling, and $\gamma = 0$ when $d = 2$ (because any arbitrarily small potential in two dimension creates at least one bound state). The inequality was proved for $\gamma > 1/2$ in one dimension and $\gamma > 0$ in other dimensions in the original paper by Lieb and Thirring [LT76]. The proof in the case $\gamma = 1$ was recently greatly simplified by Rumin [Rum11]. The case $\gamma = 0, d \geq 3$ requires completely different methods and was proven independently by Cwikel, Lieb and Rozenblum [Cwi77; Lie76; Roz72]. The limit case $\gamma = 1/2, d = 1$ remained unsolved for twenty years until it was finally settled by Weidl [Wei96], and refined with a sharp constant by Hundertmark, Lieb and Thomas [HLT98].

The inequality (4.1) can be interpreted as a comparison of the quantum mechanical energy of a system with Hamiltonian $-\Delta+V$ to its semiclassical approximation. The semiclassical regime is obtained by letting the Planck constant $\hbar$ tend to zero in the Hamiltonian $-\hbar\Delta + V$. In this paper, we have scaled $\hbar$ away from inequality (4.1) for convenience, and the corresponding limit is a large (or extended) V. In this limit, the eigenfunctions become localized in the region of the phase space defined by $p^2 + V(x) \leq 0$ and explicit computations are possible. More precisely, using the Weyl asymptotics, one can prove

$$\lim_{\mu\to\infty} \frac{\mathrm{tr}((-\Delta + \mu V)_-^\gamma)}{\int (\mu V)^{\gamma+d/2}} = L_{\gamma,d,\mathrm{sc}},$$

where

$$L_{\gamma,d,\mathrm{sc}} = 2^{-d}\pi^{-d/2}\frac{\Gamma(\gamma+1)}{\Gamma(\gamma+d/2+1)}.$$

Therefore, denoting by $L_{\gamma,d}$ the optimal constant in (4.1), we have $L_{\gamma,d} \geq L_{\gamma,d,\mathrm{sc}}$. We set

$$R_{\gamma,d} = \frac{L_{\gamma,d}}{L_{\gamma,d,\mathrm{sc}}} \geq 1.$$

The value of $R_{\gamma,d}$ describes by how much the quantum mechanical energy can exceed its semiclassical counterpart. In this paper, we investigate numerically the problem

$$R_{\gamma,d} = \frac{1}{L_{\gamma,d,\mathrm{sc}}} \sup\Big\{ \mathrm{tr}((-\Delta + V)_-^\gamma)\ ,\ V \in L^{\gamma+d/2}, \int V^{\gamma+d/2} = 1\Big\}, \qquad (P_{\gamma,d})$$

where we impose the condition $\int V^{\gamma+d/2} = 1$ to eliminate the scaling invariance of the Lieb-Thirring inequality. As a global maximization of this non-concave functional is out of reach, we develop an algorithm to maximize it locally, and present numerical results for different starting points.

Several important facts about $R_{\gamma,d}$ are known. A scaling argument by Aizenman and Lieb [AL78] shows that $R_{\gamma,d}$ is decreasing with respect to γ. Laptev and Weidl [LW00b] showed that $R_{\gamma,d} = 1$ for $\gamma \geq 3/2$. Helffer and Robert [HR90] proved that $R_{\gamma,d} > 1$ for $\gamma < 1$ by expanding $\mathrm{tr}(-\Delta + V)_-^\gamma$ for the harmonic potential

$V(x) = 1 - |x|^2$ in the semiclassical limit. From these results, we can deduce that for each dimension d there exists a critical $\gamma_{c,d}$ such that

$$\begin{cases} R_{\gamma,d} > 1 & \text{when } \gamma < \gamma_{c,d}, \\ R_{\gamma,d} = 1 & \text{when } \gamma \geq \gamma_{c,d}, \end{cases}$$

with

$$1 \leq \gamma_{c,d} \leq \frac{3}{2}.$$

A trial potential of [LT76] provides a lower bound on $R_{\gamma,d}$ and therefore γ_c. In the appendix of this paper, Barnes solved numerically the restricted problem

$$R_{\gamma,d,1} = \frac{1}{L_{\gamma,d,\mathrm{sc}}} \sup_{\int V^{\gamma+d/2}=1} |\lambda_1|^\gamma, \tag{4.2}$$

where λ_1 is the lowest eigenvalue of $-\Delta + V$. This is equivalent to restricting $(P_{\gamma,d})$ to the potentials V such that $-\Delta+V$ only has one negative eigenvalue. The solution $V_{\gamma,d,1}$ of this problem is negative, radial and only has bound state, *i.e.* $-\Delta + V_{\gamma,d,1}$ only has one negative eigenvalue. The corresponding $R_{\gamma,d,1}$ is decreasing with γ, and intersects 1 at a critical $\gamma_{c,d,1}$. In low dimensions, $\gamma_{c,1,1} = \frac{3}{2}$, $\gamma_{c,2,1} \approx 1.165$, $\gamma_{c,3,1} \approx 0.863$ (our numerical results agree with these values). Therefore, $\gamma_{c,1} = \frac{3}{2}$ and $\gamma_{c,2} \geq 1.165$, but nothing can be said about $\gamma_{c,d}$ for $d \geq 3$. A famous conjecture of Lieb and Thirring states that $R_{1,3} = 1$, *i.e.* $\gamma_{c,3} = 1$. This would imply improved bounds on the energy of a system of fermions, and is of great importance to the Thomas-Fermi theory [Lie81].

Better upper bounds on $R_{\gamma,d}$ have also been derived recently. For instance, it is proven in [DLL08] that $R_{\gamma,d} \leq \frac{2\pi}{\sqrt{3}} \approx 3.63$ for $\gamma \geq \frac{1}{2}$, $R_{\gamma,d} \leq \frac{\pi}{\sqrt{3}} \approx 1.82$ for $\gamma \geq 1$.

Despite these advances on upper bounds, not much is known about lower bounds for $R_{\gamma,d}$, except for the one-bound state potential $V_{\gamma,d,1}$ of [LT76] and the asymptotic result in the semiclassical limit of [HR90]. In this paper, we attempt to bridge the gap between these two results by looking numerically for maximizers of the variational problem $(P_{\gamma,d})$. To our knowledge, this is the first numerical study of the Lieb-Thirring inequalities since the work of Barnes [LT76] (and unpublished recent work by Arnold and Dolbeault [Dol12] in 1D).

The method we use is a finite element discretization of a natural fixed point algorithm. We describe this algorithm, and specialize it to the case of radial potentials. Then we discretize it, and use it to obtain numerical results. The critical points we obtain are the natural generalization of the potential with one bound state obtained by Barnes. They serve as a partial bridge between this potential and asymptotic results, and yield new lower bounds on $R_{\gamma,d}$.

4.2 The optimization algorithm

4.2.1 Optimization scheme

Let us denote $E(V) = \mathrm{tr}((-\Delta + V)_-^\gamma)$, so that

$$L_{\gamma,d} = \sup \Big\{ E(V) \ , \ V \in L^{\gamma+d/2}, \int V^{\gamma+d/2} = 1 \Big\}.$$

Let $\mathcal{S}_p$ be the set of symmetric operators on $L^2(\mathbb{R}^d)$ with finite Schatten norm

$$\|\tau\|_p = \mathrm{tr}(|\tau|^p)^{1/p}.$$

The fixed point algorithm we use is based on the following property:

Proposition 4.1. *For any $V \in L^{\gamma+d/2}(\mathbb{R}^d, \mathbb{R})$ and any $\gamma \geq 1$,*

$$E(V)^{1/\gamma} = \max\Big\{ -\mathrm{tr}((-\Delta + V)\tau) \;/\; \tau \in \mathcal{S}_{\gamma'}, \tau \geq 0, \|\tau\|_{\gamma'} = 1\Big\},$$

where $\gamma' = \frac{\gamma}{\gamma-1}$ is the Hölder conjugate of γ.

Proof. For any $V \in L^{\gamma+d/2}$, $\tau \in \mathcal{S}_\gamma$ with $\tau \geq 0, \|\tau\|_{\gamma'} = 1$,

$$\begin{aligned} -\mathrm{tr}((-\Delta + V)\,\tau) &\leq \mathrm{tr}((-\Delta + V)_-\,\tau) \\ &\leq \|(-\Delta + V)_-\|_\gamma \|\tau\|_{\gamma'} \\ &= \mathrm{tr}((-\Delta + V)_-^\gamma)^{1/\gamma} \\ &= E(V)^{1/\gamma}, \end{aligned}$$

and the equality is achieved when

$$\tau = K_\tau (-\Delta + V)_-^{\gamma-1}, \tag{4.3}$$

where K_τ is a normalization constant chosen to ensure that $\|\tau\|_{\gamma'} = 1$.

□

From this, we see that when $\gamma \geq 1$,

$$\begin{aligned} L_{\gamma,d}^{1/\gamma} = \sup\Big\{ -\mathrm{tr}((-\Delta + V)\tau) \;/\; \tau \in \mathcal{S}_{\gamma'}, \|\tau\|_{\gamma'} = 1, \tau \geq 0, \\ V \in L^{\gamma+d/2}, \int V^{\gamma+d/2} = 1, V \leq 0\Big\}. \end{aligned}$$

What we gain from this formulation is that we can explicitely maximize $-\mathrm{tr}((-\Delta + V)\tau)$ with respect to V and τ separately. Indeed, for a fixed V, the maximum with respect to τ is given by (4.3), and for a fixed τ, the maximum with respect to V is given by

$$V(x) = -K_V \tau(x,x)^{\frac{1}{\gamma+d/2-1}},$$

where K_V is a normalization parameter chosen to ensure that $\int V^{\gamma+d/2} = 1$, and $\tau(x,y)$ is the integral kernel of τ.

This suggests the following maximization scheme: Given an approximate maximum V_n, we set $\tau_n = K_\tau(-\Delta + V_n)_-^{\gamma-1}$ and $V_{n+1}(x) = -K_V \tau_n(x,x)^{\frac{1}{\gamma+d/2-1}}$, and iterate. Explicitely:

Algorithm 1. Maximization algorithm

1. Compute the negative eigenvalues λ_i and corresponding eigenvectors ψ_i of $-\Delta + V_n$

2. Set $\rho_n = \sum_i (-\lambda_i)^{\gamma-1}\psi_i^2$
3. Compute $K_n = \left\| \rho_n^{\frac{1}{\gamma+d/2-1}} \right\|_{\gamma+d/2}^{-1} = \left(\int \rho_n^{\frac{\gamma+d/2}{\gamma+d/2-1}} \right)^{-\frac{1}{\gamma+d/2}}$
4. Set $V_{n+1} = -K_n \rho_n^{\frac{1}{\gamma+d/2-1}}$

By construction, this algorithm ensures that

$$\begin{aligned} E(V_{n+1})^{1/\gamma} &= -\operatorname{tr}((-\Delta + V_{n+1})\tau_{n+1}) \\ &\geq -\operatorname{tr}((-\Delta + V_n)\tau_{n+1}) \\ &\geq -\operatorname{tr}((-\Delta + V_n)\tau_n) \\ &= E(V_n)^{1/\gamma}. \end{aligned}$$

Therefore, the sequence $E(V_n)$ is non-decreasing, and since it is bounded from above, it converges. In general though, V_n may not converge. We give examples in Section 4.4 where, because of the translation invariance of the functional, V_n splits into two bumps that separate from each each other. Even when this behavior is forbidden, for instance by imposing a finite domain as we do in numerical computations, a rigorous convergence analysis of the algorithm is still an open problem.

Note that, even if Algorithm 1 was derived from Proposition 4.1 for $\gamma \geq 1$, it remains a reasonable algorithm when $\gamma < 1$. In this case, though, the monotonicity of $E(V_n)$ is not guaranteed.

This algorithm can be seen as a fixed-point scheme for the critical points of the maximization problem. Indeed, the Euler-Lagrange equations for $(P_{\gamma,d})$ are

$$V(x) = -K_V \left[(-\Delta + V)_-^{\gamma-1}(x,x) \right]^{\frac{1}{\gamma+d/2-1}}, \tag{4.4}$$

This is a self-consistent set of equations similar to systems such as the Hartree-Fock equations of quantum chemistry [LS77]. Our algorithm is similar in spirit to the Roothaan method [Roo51]. In our case, at least for $\gamma \geq 1$, the scheme monotonically increases the objective function. Therefore, the oscillatory behavior often seen in the Hartree-Fock model [CLB00b; Lev12b] cannot occur here. Even for $\gamma < 1$, we did not see any such oscillations numerically, and always observed linear convergence (*i.e.* $\|V_n - V_\infty\| \leq \nu^n$ for some ν, $0 < \nu < 1$), or slow separation of bumps, as will be discussed in Section 4.4.

We also note that this algorithm is used in the context of the Lieb-Thirring inequalities by Arnold and Dolbeault in 1D [Dol12].

4.2.2 Radial algorithm

Most of our multidimensional computations were done in a radial setting. To see why this is possible, consider the above iteration for $d \geq 2$, when V_n is radial. Then, the Laplacian splits into $\Delta_r + \frac{1}{r^{d-1}}\Delta_\theta$, where Δ_r is the radial Laplacian, and Δ_θ the Laplace-Beltrami operator on S^{d-1}. The eigenvectors of Δ_θ are explicitly known to be the spherical harmonics, labeled by the integers ℓ and m. Since V_n is

radial, $-\Delta_r + V_n$ commutes with Δ_θ and can therefore be diagonalized in the same basis (separation of variables). We write these eigenvectors of the form $\psi_{i,\ell,m}(x) = \phi_{i,\ell}(|x|)J_{\ell,m}(x/|x|)$, where $J_{\ell,m}$ is the spherical harmonic of degree ℓ and order m. The radial parts $\phi_{i,\ell}$ satisfy the equation

$$-\frac{1}{r^{d-1}}(r^{d-1}\phi'_{i,\ell})' + \frac{\ell(\ell+d-2)}{r^2}\phi_{i,\ell} + V\phi_{i,\ell} = \lambda_{i,\ell}\phi_{i,\ell}, \tag{4.5}$$

and each $\lambda_{i,\ell}$ has multiplicity

$$h(d,\ell) = \binom{d+\ell-1}{\ell} - \binom{d+\ell-3}{\ell-2}$$

(see [SW71].)

Therefore, we can obtain all the negative eigenvectors and eigenvalues by solving only (4.5). Furthermore, since the lowest eigenvalue increases as ℓ increases, we can iterate over ℓ and stop whenever the lowest eigenvalue becomes positive. Next, we compute

$$\begin{aligned}\rho_n(x) &= \sum_\ell \sum_i \sum_m (-\lambda_{i,\ell})^{\gamma-1}\psi_{i,\ell,m}(x)^2 \\ &= \sum_\ell h(d,\ell)\sum_i (-\lambda_{i,\ell})^{\gamma-1}\phi_{i,\ell}(r)^2\end{aligned}$$

where the sum only ranges over all negative eigenvalues. This is again radial, and therefore so is V_{n+1}. To summarize:

Algorithm 2. Radial maximization algorithm

1. Compute all the negative eigenvalues $\lambda_{i,\ell}$ and associated eigenvectors $\phi_{i,\ell}$ of of (4.5) by increasing ℓ until the lowest eigenvalue $\lambda_{0,\ell}$ becomes positive
2. Set $\rho_n = \sum_\ell \sum_i h(d,\ell)(-\lambda_{i,\ell})^{\gamma-1}\phi_{i,\ell}^2$
3. Compute $K_n = \left\| \rho_n^{\frac{1}{\gamma+d/2-1}} \right\|_{\gamma+d/2}^{-1} = \left(|\mathbb{S}_{d-1}| \int \rho_n^{\frac{\gamma+d/2}{\gamma+d/2-1}}(r)\mathrm{d}r \right)^{-\frac{1}{\gamma+d/2}}$
4. Set $V_{n+1} = -K_N \rho_n^{\frac{1}{\gamma+d/2-1}}$

Note that this algorithm is not an approximation: the iterates generated by this algorithm coincide with the ones from Algorithm 1 when the initial guess V_0 is radial.

4.3 Discretization

4.3.1 Galerkin basis and weak formulation

To discretize this algorithm, we use a Galerkin finite element basis on which we expand V and the eigenvectors. This variational discretization has the advantage that numerical computations can provide accurate lower bounds on the best constants $R_{\gamma,d}$. A disadvantage is that the algorithm we use involves taking powers of

functions. There is no exact way to do that in a Galerkin basis, and we must use an approximation, which causes a loss of accuracy in the fixed-point algorithm.

For the non-radial 1D and 2D cases, we simply use standard finite elements. The radial case is less standard, and we must derive a weak formulation of (4.5). There are several possibilities here. The simplest is to use a change of variable $\varphi(r) = r^{\frac{d-1}{2}}\phi(r)$ to transform the equation back to the more standard form

$$-\varphi'' + \frac{(l+\frac{d-1}{2})(l+\frac{d-3}{2})}{r^2}\varphi + V(r)\varphi = \lambda\varphi.$$

We obtain a weak formulation by multiplying by a test function u and integrating:

$$\int \left[\varphi' u' + \left(\frac{(l+\frac{d-1}{2})(l+\frac{d-3}{2})}{r^2} + V\right)\varphi u\right] = \lambda \int \varphi u. \tag{4.6}$$

Expanding the function φ on a Galerkin basis $\varphi(r) = \sum_i x_i \chi_i(r)$, this problem transforms into the generalized matrix eigenvalue problem

$$Ax = \lambda M x, \tag{4.7}$$

where

$$A_{ij} = \int \chi_i' \chi_j' + \left(\frac{(l-\frac{d-1}{2})(l+\frac{d-3}{2})}{r^2} + V\right)\chi_i\chi_j, \tag{4.8}$$

$$M_{ij} = \int \chi_i \chi_j, \tag{4.9}$$

are the stiffness and mass matrices.

When V is expanded on the same basis, we can compute the matrix elements, solve the eigenvalue problem (4.7), and transform back to ϕ. However, for $d = 2, l = 0$, φ behaves like $\sqrt{r}$ at 0, and the singularity of the derivative prevents an accurate discretization.

Another possibility is to obtain a weak formulation of (4.5) directly. We have to remember that since the ϕ functions are only radial parts of a d-dimensional function, the L^2 inner product between two functions ϕ_1 and ϕ_2 is $\int \phi_1\phi_2 r^{d-1}$. This has to be taken into account in the weak formulation in order to obtain a self-adjoint equation. Multiplying (4.5) by $r^{d-1}u$, where u is a test function, and integrating by parts, we obtain:

$$\int \left[r^{d-1}\phi' u' + \left(\frac{l(l+d-2)}{r^2} + V\right) r^{d-1}\phi u\right] = \lambda \int r^{d-1}\phi u. \tag{4.10}$$

Then, as with (4.6), we can transform this into a matrix equation and solve it. The disadvantage is that the integrations required for the assembly step are more involved, and therefore we only use it for the case $d = 2, l = 0$ where the other method fails.

4.3.2 Finite elements

We use a Finite Element basis of piecewise linear functions on a grid of $[0, L]$. The grid we use is a nonuniform grid of N points, with more points around 0, to accommodate for the singularity of (4.5). L is chosen large enough so that all the eigenvectors associated to negative eigenvalues of $-\Delta + V$ can be represented accurately in the basis. But if ψ is an eigenvector with negative eigenvalue λ, then, ψ decays as $\exp(-\sqrt{-\lambda}r)$. This shows that the discretization is problematic for eigenvalues close to zero, as we shall see in Section 4.4.

For our piecewise linear basis functions, it is easy to compute the matrix elements (4.8,4.9) when V_n is expanded on this same basis χ_i. We then solve the eigenproblem (4.7), and obtain the eigenvectors. To expand $\rho^{\frac{1}{\gamma+d/2-1}}$ on the basis, we simply chose the expansion that is exact on the grid points. Then the normalization can be performed exactly, and the iteration is carried out.

We present convergence results in Figure 4.1. As expected from piecewise linear basis functions, we obtain $O(1/N)$ convergence of the eigenfunctions ϕ in H^1 norm, and $O(1/N^2)$ convergence of the eigenvalues λ (and therefore of $E(V)$). For a given stepsize, the discretization error is exponential with respect to L, with a decay rate equal to the decay rate of the eigenfunctions, *i.e.*$\sqrt{-\lambda}$. Although we used a uniform grid in one dimension in Figure 4.1, similar results were checked to hold for our non-uniform grid in d dimensions, using the weak formulation (4.6), except for $d = 2, l = 0$, where the convergence was slower and the weak formulation (4.10) had to be used to get the same convergence rates.

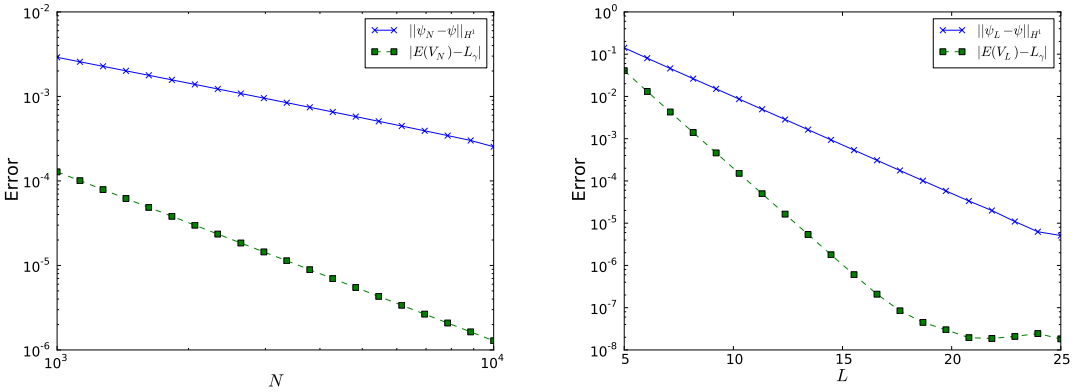

Figure 4.1: Convergence analysis with respect to N and L. The figure on the left illustrates the convergence of the eigenfunctions and eigenvalues with respect to the number of grid points N, with a fixed $L = 40$, large enough to cause a negligible error. The slopes, found by linear regression, are -1.043 for the eigenvectors and -1.999 for the eigenvalues, close to their theoretical values of -1 and -2. The figure on the right illustrates the convergence of the eigenfunctions and eigenvalues with respect to the domain size, with constant stepsize $h = 5 \times 10^{-4}$. The slope for the eigenvectors is $-0.5285 \approx \sqrt{-\lambda} = -0.5283$. When the domain size gets large enough, the error due to h dominates, and causes a plateau.

4.3.3 Diagonalization

The most computationally challenging step in our algorithm is the problem of computing all the negative eigenvalues and associated eigenvectors of a generalized eigenvalue problem $Ax = \lambda Mx$, where A and M are large sparse symmetric matrices. The spectrum of A consists of n negative eigenvalues, and a large number of positive eigenvalues, which can be seen as perturbations of the spectrum of the discrete Laplacian.

In order to compute the n negative eigenvalues, where n is unknown, we compute the k lowest eigenvalues, check if the k-th one is positive, and repeat the process with a larger k if not. The computation of the k first eigenvalues can be done by standard packages such as ARPACK [LSY98]. However, standard algorithms such as Arnoldi iteration are not suited for the computation of inner eigenvalues, and one has to solve for the largest eigenvalues of the shifted and inverted problem $(A-\sigma)^{-1}x = (\lambda-\sigma)^{-1}x$ instead to ensure reasonably fast convergence. This requires an adequate shift (one that is close to the bottom of the spectrum of A). Even with this shift-and-invert strategy, the group of eigenvalues one needs to locate is not well-separated from the rest of the spectrum, and becomes less and less so as L increases. This leads to slow convergence for large L.

4.3.4 Implementation

We implemented the algorithm in Python, using the Numpy/Scipy libraries [JOP+01], with the ARPACK eigenvalue solver [LSY98].

4.3.5 Error control

Our numerical methods introduce errors. For each V we construct, we can get a lower bound of $L_{\gamma,d}$, provided we can accurately compute $E(V)$ and the normalization constant $\int V^{\gamma+d/2}$.

Because the basis functions we use are piecewise linear, all the integrals involved in our computations, such as $\int V^{\gamma+d/2}$, can be reduced to integrals of piecewise polynomials, which can be computed explicitely, up to machine precision.

It now remains to examine the accuracy of $E(V)$, *i.e.* of the computation of eigenvalues of $-\Delta + V$. Because we use a Galerkin discretization, the min-max theorem guarantees that the eigenvalues of the matrix problem we solve are larger than the true eigenvalues. The matrix problem is solved using ARPACK, which yields a collection of orthogonal approximate eigenvectors. Again using the min-max theorem, the eigenvalues of the submatrix formed by restricting the eigenvalue problem to the subspace spanned by the approximate eigenvectors are upper bounds on the true eigenvalues of $-\Delta + V$. This submatrix is of modest size, and its eigenvalues can be computed accurately by ARPACK.

For the reasons mentioned above, we believe our method to yield lower bounds on $L_{\gamma,d}$ with an accuracy comparable to machine precision (about 10^{-16}). However, guaranteed lower bounds would require methods such as interval arithmetic, which we have not implemented.

4.4 Numerical results

We used the algorithm described above to compute critical points of the variational problem $(P_{\gamma,d})$, that is, a solution to the self-consistent equation (4.4). Our strategy was to find a critical point by running the algorithm on a suitable initial potential V_0 (for instance, a Gaussian of specified width). Once convergence is achieved for a specific γ, we can run the algorithm for $\gamma + \Delta\gamma$, using as initial potential the one found for γ. We have been unable to analytically prove the correctness of this method, for instance by checking the conditions of the implicit function theorem. However, we found it reliable enough for our purpose, as long as $\Delta\gamma$ was chosen small enough.

4.4.1 The 1D case

In one dimension we simply used a grid of size $[-L, L]$ with Dirichlet boundary conditions. We reproduced the potential with one bound state $V_{\gamma,1,1}$ of [LT76] by using for V_0 a Gaussian of relatively small variance (see Figure 4.2). In this case, the algorithm converges linearly (*i.e.* $\|V_{n+1} - V_n\| \approx \nu^n$, for some convergence rate $\nu < 1$), and very quickly (about twenty iterations to achieve machine precision).

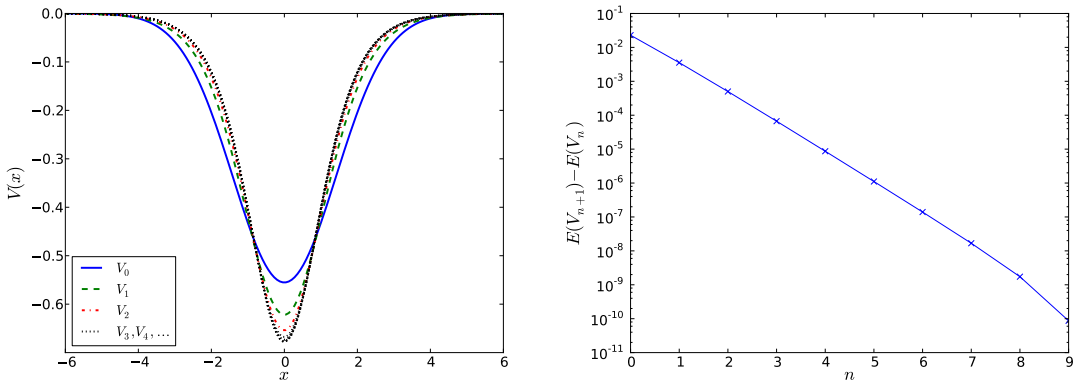

Figure 4.2: Linear convergence from a Gaussian initial data towards $V_{\gamma,d,1}$. This plot is for $\gamma = 1.2$.

Contrary to what we observe in higher dimensions, initializing the algorithm with a Gaussian of larger variance did not make the algorithm converge to other critical points. Instead, it leads to a slow divergence where "bumps" separate, each bump corresponding to the potential of $V_{\gamma,1,1}$ (see Figure 4.3).

This effect occured regardless of the value of γ in the range $\gamma \in (1/2, 3/2)$. The divergence appears to be logarithmic. More precisely, a numerical fit showed the asymptotic relationship for the distance L_n between the two bumps

$$L_n \approx \frac{1}{2\sqrt{-\lambda}} \log n, \tag{4.11}$$

where λ is the unique negative eigenvalue of $-\Delta + V_{\gamma,1,1}$.

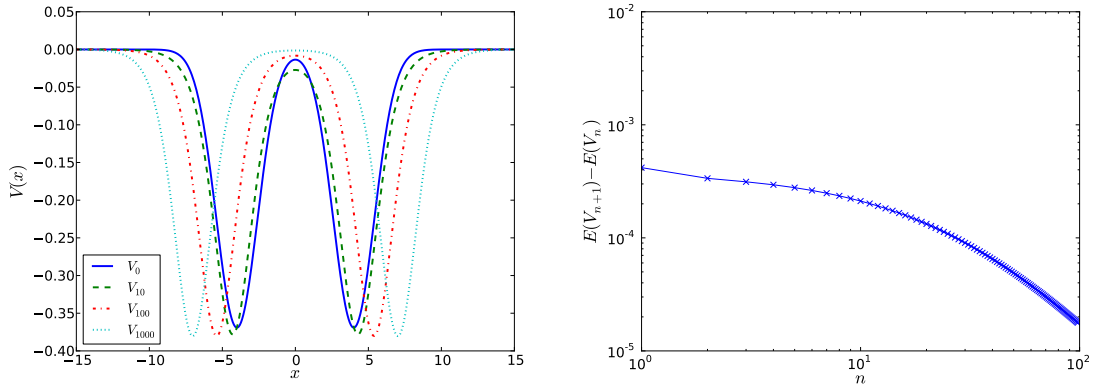

Figure 4.3: Divergence from a sum of Gaussian bumps at $\gamma = 1.2$. Initializing to a single Gaussian of large width yields similar results. The two bumps separate slowly, until the finiteness of L forces convergence. Although not displayed here, the asymptotic logarithmic divergence (4.11) can be checked graphically, e.g. the spacing between V_{100} and V_{1000} is the same than between V_{1000} and V_{10000}. The asymptotic slope of the log-log convergence plot is -1, which fits with the heuristic arguments given.

This strikingly simple relationship can heuristically be understood by the fact that, because the eigenfunctions ϕ_1 and ϕ_2 corresponding to the two bumps have exponential decay with decay constant $\sqrt{-\lambda}$, their interaction is of order $\exp(-\sqrt{-\lambda}L_n)$. Due to cancellations, this interaction leads to a correction of order $\exp(-2\sqrt{-\lambda}L_n)$ in V_{n+1}, and we have the approximate relationship $L_{n+1} \approx L_n + K\exp(-2\sqrt{-\lambda}L_n)$, which yields (4.11). A rigorous explanation of this is an interesting question.

Based on these results, we conjecture that $V_{\gamma,d,1}$ is, up to translation, the only maximizer of the functional. This would imply that

$$R_{\gamma,1} = 2\left(\frac{\gamma - 1/2}{\gamma + 1/2}\right)^{\gamma - 1/2}$$

for $\gamma \leq \frac{3}{2}$, in agreement with the original conjecture of Lieb and Thirring [LT76].

4.4.2 The 2D case

Due to the high cost of accurately solving the maximization problem in more than one dimension, we only obtained partial results in the non-radial 2D case. The results indicate that the algorithm either converges to radial critical points, or form slowly separating bumps, as in one dimension. We have been unable to find any example of non-radial critical points, although this does not mean that they do not exist. An example of separation in bumps is provided in Figure 4.4.

In the radial case, we followed branches of critical points of $(P_{\gamma,d})$ using the continuation method on γ we described in Section 4.4. We found this branch continuation procedure robust. By varying the shape of the initial data, and in particular its width, we were able to obtain different branches of critical points. Generally

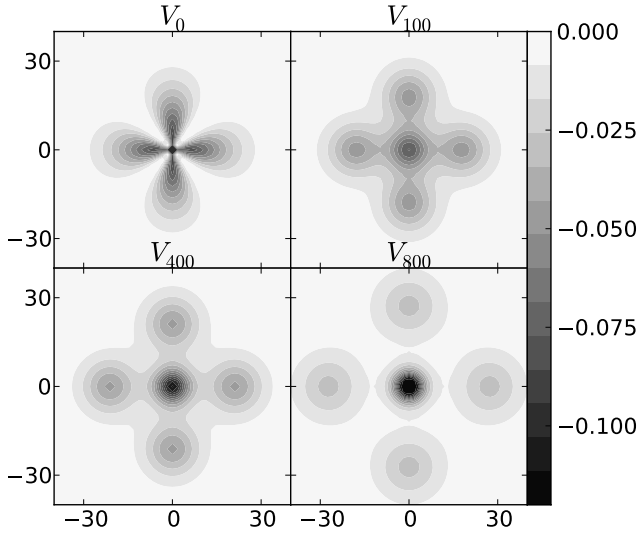

Figure 4.4: Separation in bumps of 2D non-radial initial data at $\gamma = 1.2$, computed from Algorithm 1 with standard FEM. The initial data is taken to be a Gaussian in r multiplied by an angular factor $(1 + \cos(4\theta))$.

speaking, as the width of the initial potential increases, the number of negative eigenvalues of $-\Delta + V$ increases.

We display the energy of these branches as a function of γ in Figure 4.5.

k	1	4	6	8	11	13	17	28	54	81	139
$\gamma_{c,2,k}$	1.165	1.150	1.141	1.135	1.126	1.124	1.124	1.119	1.110	1.105	1.100

Table 4.1: Values $\gamma_{c,2,k}$ of γ at which some branches with k bound states cross the threshold $R = 1$.

First, we see that as the number of eigenvalues increases, $R(V_\gamma)$ tends to 1, the semiclassical limit. For a given γ, the Lieb-Thirring constant will be given by the supremum of $R(V_\gamma)$ over such curves. From the branches depicted here, we see that the supremum is always given by either $V_{\gamma,d,1}$ for $\gamma < \gamma_c^1 \approx 1.16$, and by the semiclassical limit for $\gamma > \gamma_c^1$. All the other curves are below these two (although not depicted in this zoomed plot, this holds true for $0.5 < \gamma < 1.5$).

It is important to keep in mind that these branches only represent critical points of the functional. They are generically local maxima with respect to radial perturbations, but might not be stable with respect to non-radial ones. For instance, Figure 4.6 depicts what happens when the radial potential with eight bound states is perturbed non-radially by a Gaussian multiplicative noise, with $\gamma = 1.1$. Because the potential with one bound state has a higher energy, the potential splits into eight bumps. Performing the same experiment γ larger than about 1.17, where the potential with one bound state has a lower energy, when the potential with eight bound states has a larger energy, we found that no splitting occurs, which suggests that the potentials are stable beyond their crossing with the one with one bound state.

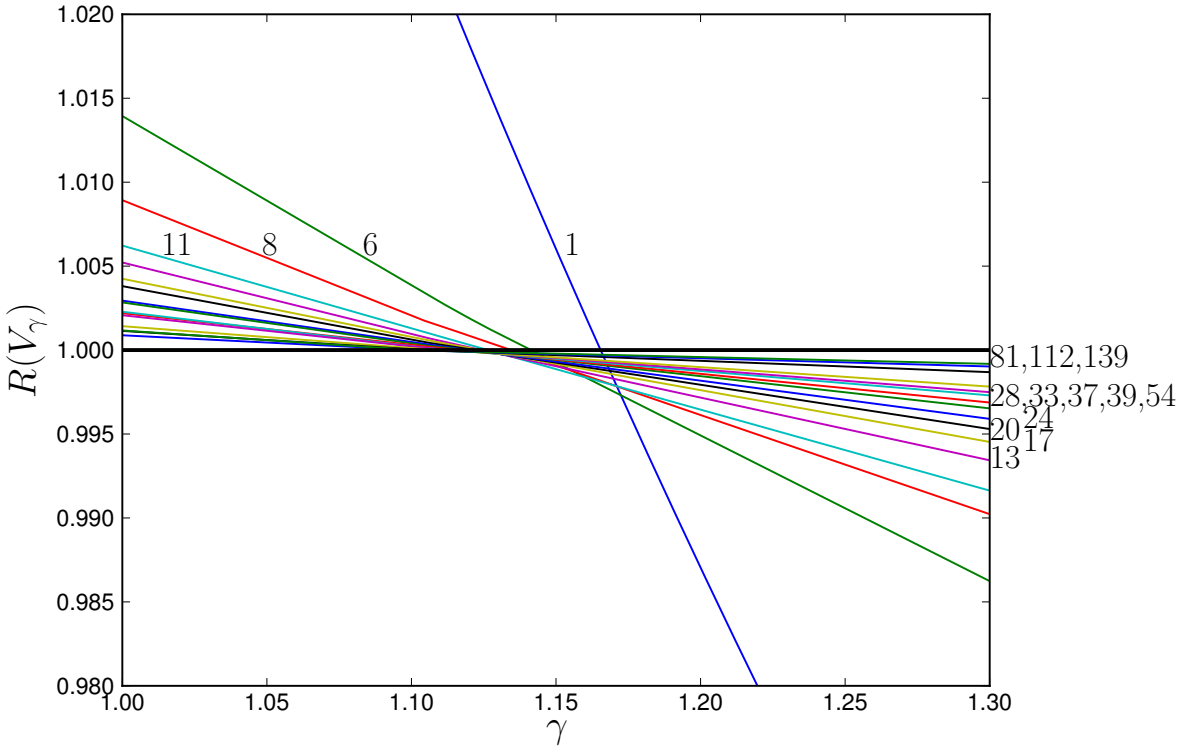

Figure 4.5: $R(V)$ as a function of γ for $d = 2$ using the branch continuation procedure in the radial setting (Algorithm 2). The branches are labeled by the number of bound states they have. Based on these results, the maximizer seems to be $V_{\gamma,d,1}$, up to $\gamma \approx 1.16$. Above this, no branches were above the semiclassical regime $R = 1$.

Based on our numerical results, we conjecture that $\gamma_c = \gamma_c^1 \approx 1.16$. The only way this conjecture could be false is if some other curve is above the one with one bound state. We have not been able to find such a curve.

4.4.3 The 3D case

Due to the high cost of computation, we were unable to get meaningful results in the non-radial setting, and only present our findings in the radial case. Some of the radial potentials we found are presented in Figure 4.7, with numerical data about the crossings of the curves in Table 4.2.

k	1	4	5	10	14	21	30	55	111	140	341
$\gamma_{c,3,k}$	0.863	0.852	0.875	0.851	0.880	0.857	0.862	0.860	0.854	0.891	0.853

Table 4.2: Values $\gamma_{c,3,k}$ of γ at which some branches with k bound states cross the threshold $R = 1$.

Contrary to the dimension 2, here some potentials with a higher number of bound states have a higher R than $V_{\gamma,d,1}$. The corresponding curves in the $\gamma - R$

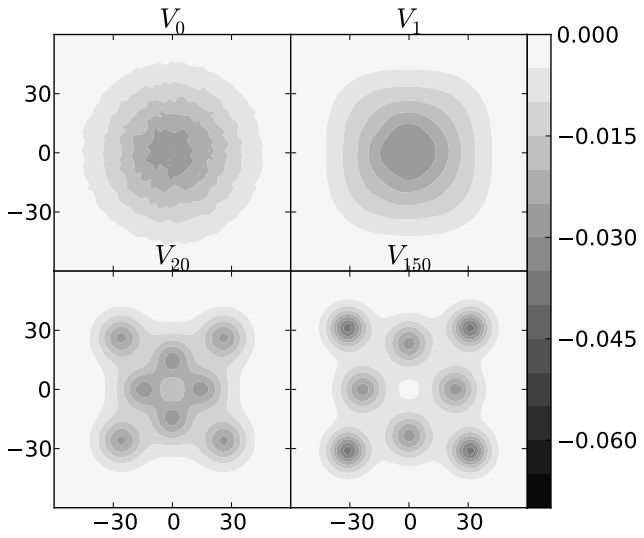

Figure 4.6: Separation in bumps of the randomly perturbed radial potential with eight bound states, $\gamma = 1.1$.

plane are flatter and flatter, and intersect at a sequence of increasing γ_c^k. This sequence seems to accumulate at 1, in accordance to Helffer and Robert's result [HR90], which predicts the existence of similar potentials with a R larger than 1 for every $\gamma < 1$ in the nearly semiclassical regime. However, their study of an harmonic potential of varying width showed a highly oscillatory behavior of $E(V)$ with respect to the width of the potential. Although our numerical methods do not permit us to investigate this regime (it would require a very large domain size, with a correspondingly large number of points, and a very poor conditioning of the eigenvalue problem), we expect the same behavior to occur. This means that the computation of the maximizing branches (here, those with 1, 5, 14 and 140 bound states) becomes harder and harder.

The relative energies of the branches display a greater variety than in the case $d = 2$, reflecting a more complicated energy landscape. Figure 4.8 shows the profiles of some critical potentials. Note that the potentials in dashed lines can be regarded as "anomalous" versions of their similarly extended counterparts, and have a lower energy. As can be expected, the potential branches which are maximizers for some γ have a profile that is decreasing in r.

Table 4.3 displays the eigenvalue repartitions of a particular potential, the one with 140 bound states, at $\gamma = 0.88$. Several patterns can be noted. First, for low values of k and l, the approximate relationship $\lambda_{k+1,l} = \lambda_{k,l+2}$ holds. This is because the associated eigenfunctions are localized close to 0. In this region, the potential can be approximated by a harmonic potential $V(0) + \frac{1}{2}V''(0)x^2$, which leads to the approximation $\lambda_{k,l} = V(0) + \sqrt{V''(0)}(2k + l + \frac{3}{2})$, explaining the relationship $\lambda_{k+1,l} \approx \lambda_{k,l+2}$. This is only valid for small k and l, where the harmonic approximation is valid. When the eigenvalues are close to zero, another pattern emerges: the last negative eigenvalues have the same value of $k + l$, leading to a triangular pattern in Table 4.3. This seems to be true for the maximizing branches (1, 5, 14, 140), but

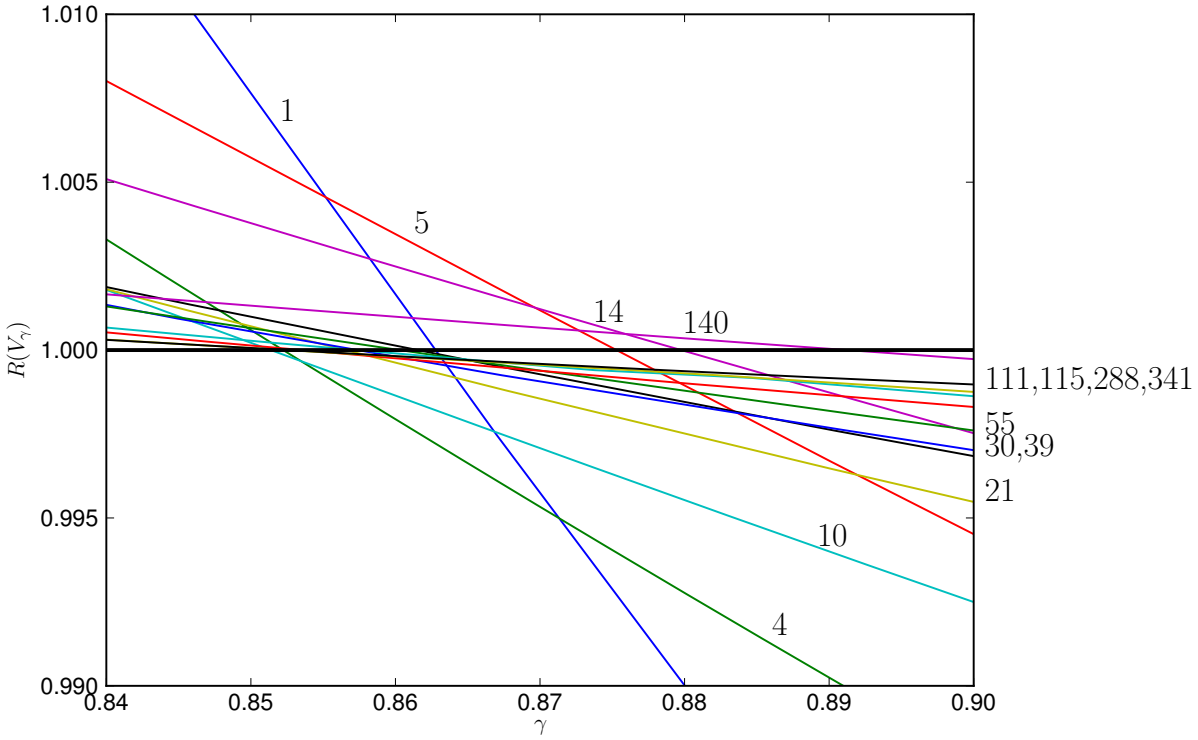

Figure 4.7: $R(V)$ as a function of γ for $d = 3$, with the same methodology as in Figure 4.5. As γ increases, different branches become maximizers, until $\gamma = 1$. The fact that only some branches are above the threshold $R = 1$ when the number of bound states increases results from the highly oscillatory behavior near $\gamma = 1$, as predicted in [HR90].

not for the others (which tend to have a different ordering of the eigenvalues close to zero). We do not have an explanation for this.

$\lambda_{k,l}$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$l=0$	-85	-54	-30	-13	-4	-0.7	-0.02	+0.25
$l=1$	-69	-41	-20	-7.6	-1.8	-0.1	+0.02	
$l=2$	-54	-29	-12	-3.4	-0.3	+0.03		
$l=3$	-39	-18	-5.8	-0.7	+0.04			
$l=5$	-26	-9	-1.3	+0.06				
$l=6$	-14	-2.3	+0.08					
$l=7$	-3.8	+0.1						

Table 4.3: Eigenvalue repartition for the branch with 140 bound states, with $\gamma = 0.88$, in units of 10^{-6} for readability. As in Section 4.2.2, $\lambda_{k,l}$ is the k-th eigenvalue of equation (4.5). The positive eigenvalues are finite size effects and have no physical meaning, therefore we do not write them beyond the first one.

Every potential we were able to compute was below the $R = 1$ threshold for

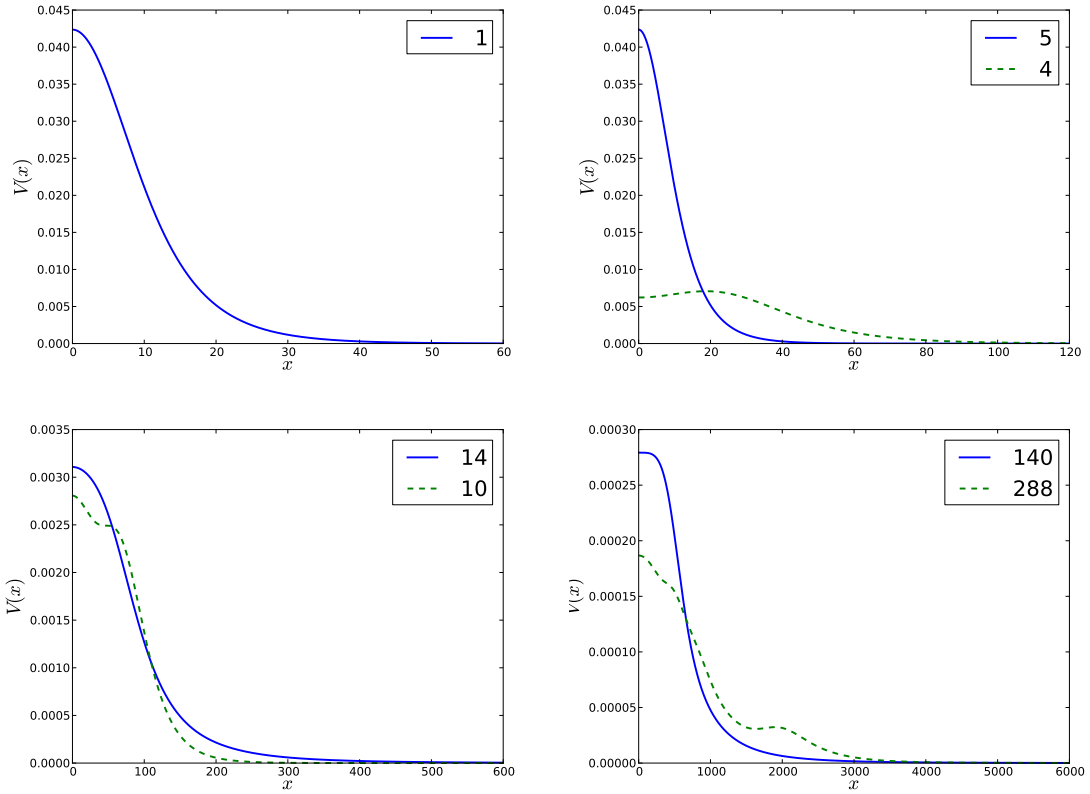

Figure 4.8: Shape of some critical points of $E(V)$, at $\gamma = 1$. The qualitative shape of the curves does not change much when varying γ.

$\gamma > 1$. These results confirm the Lieb-Thirring conjecture that $\gamma_{c,3} = 1$. However, the subtle behavior near $\gamma = 1$ means that there might exist maximizing potentials for $\gamma > 1$ we were unable to find.

4.4.4 The case $d \geq 4$

The results we obtained in the case $d \geq 4$ are similar to the case $d = 3$, with $V_{\gamma,d,1}$ as the maximizer for small γ, until it is outperformed by branches with a larger number of bound states, which all fall under the semiclassical limit before $\gamma = 1$. However, the high cost of computation (the higher the dimension, the more ill-conditioned the eigenvalue problem is) prevents us from performing a systematic study.

4.5 Conclusion

In this paper, we used a maximization algorithm to investigate the best constants of the Lieb-Thirring inequalities, an important open problem of quantum mechanics. Discretizing the problem using finite elements, we were able to numerically compute critical points of the functional. Our main findings are the following:

- In one dimension, the only critical point we found is the potential with exactly one bound-state $V_{\gamma,1,1}$. For all initial data, the algorithm seems to either converge to this potential, or to split in separating bumps. This supports the conjecture that these potentials are the only maximizers of the functional for $\frac{1}{2} \leq \gamma < \frac{3}{2}$.

- In two dimensions, the only critical points we found were radial. For all initial data, the algorithm seems to either converge to a radial potential, or to split in diverging bumps. Among the radial potentials we were able to compute, $V_{\gamma,2,1}$ was the maximizer for $\gamma < 1.16$. After this point, the maximum corresponds to the semiclassical regime. The natural conjecture is that $V_{\gamma,d,1}$ is the maximizer for $\gamma < 1.16$, and therefore that $\gamma_{c,2} \approx 1.16$.

- In three dimensions, branches of radial potentials with more than one bound state outperformed $V_{\gamma,d,1}$. This confirms the theoretical result from [HR90], and provides new lower bounds (see Figure 4.7). The natural conjecture is that there is a non-decreasing sequence of integers n_γ with $n_\gamma \to \infty$ as $\gamma \to 1$ such that the maximizer of the functional $(P_{\gamma,d})$ has n_γ bound states.

- In dimension $d \geq 4$, the results are similar to the three-dimensional case, and we conjecture the same behavior.

Beyond these results, the study hinted at a very rich and highly nonlinear behavior of the maximizers of Lieb-Thirring inequalities. This study is based on a simple numerical method (finite element discretization). More involved computations could use a more appropriate Galerkin basis. This could allow for a more accurate computation of potentials with a larger number of bound states, and a more detailed exploration of the energy landscape.

On the theoretical side, the properties of the functional $(P_{\gamma,d})$ remain unexplored. An open question is whether one can prove the existence of maximizers. The behavior of the maximization algorithm is also an interesting question, with the separation in bumps particularly interesting. Finally, we believe that our method could be adapted to generalizations such as models with positive temperatures or positive density [Fra+12], negative values of γ [Dol+06] or polyharmonic operators $(-\Delta)^l + V$ (see [NW96]).

Acknowledgments

I would like to thank Mathieu Lewin and Éric Séré for their help and advice.

Bibliography

[AL78] M. Aizenman and E.H. Lieb. On semi-classical bounds for eigenvalues of Schrödinger operators. *Phys. Lett. A* 66.6 (1978), pp. 427–429.

[CLB00b] E. Cancès and C. Le Bris. On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 34.4 (2000), pp. 749–774.

[Cwi77] M. Cwikel. Weak type estimates for singular values and the number of bound states of Schrodinger operators. *Ann. of Math.* (1977), pp. 93–100.

[Dol12] J. Dolbeault. Personal communication. 2012.

[DLL08] J. Dolbeault, A. Laptev, and M. Loss. Lieb-Thirring inequalities with improved constants. *J. Eur. Math. Soc. (JEMS)* 10 (2008), pp. 1121–1126.

[Dol+06] J. Dolbeault et al. Lieb–Thirring type inequalities and Gagliardo–Nirenberg inequalities for systems. *J. Funct. Anal.* 238.1 (2006), pp. 193–220.

[Fra+12] R.L. Frank et al. A positive density analogue of the Lieb-Thirring inequality. *Duke Math. J., to appear* (2012).

[HR90] B. Helffer and D. Robert. Riesz means of bounded states and semi-classical limit connected with a Lieb-Thirring conjecture II. *Ann. Inst. Henri Poincaré (A)* 53.2 (1990), pp. 139–147.

[HLT98] D. Hundertmark, E.H. Lieb, and L.E. Thomas. A sharp bound for an eigenvalue moment of the one-dimensional Schrodinger operator. *Adv. Theor. Appl. Math.* 2 (1998), pp. 719–732.

[JOP+01] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy: Open source scientific tools for Python*. 2001–.

[LW00a] A. Laptev and T. Weidl. Recent results on Lieb-Thirring inequalities. *Journées Équations aux dérivées partielles* (2000), pp. 1–14.

[LW00b] A. Laptev and T. Weidl. Sharp Lieb-Thirring inequalities in high dimensions. *Acta Math.* 184.1 (2000), pp. 87–111.

[LSY98] R.B. Lehoucq, D.C. Sorensen, and C. Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. Vol. 6. Siam, 1998.

[Lev12b] A. Levitt. Convergence of gradient-based algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 46.06 (2012), pp. 1321–1336.

[Lie76] E.H. Lieb. The number of bound states of one-body Schrödinger operators and the Weyl problem. *Bull. Amer. Math. Soc.* 82 (1976), pp. 751–753.

[Lie81] E.H. Lieb. Thomas-Fermi and related theories of atoms and molecules. *Rev. Mod. Phys.* 53.4 (1981), p. 603.

[Lie90] E.H. Lieb. The stability of matter: from atoms to stars. *Bull. Amer. Math. Soc.* 22.1 (1990).

[LS10] E.H. Lieb and R. Seiringer. *The stability of matter in quantum mechanics*. Cambridge Univ. Press, 2010.

[LS77] E.H. Lieb and B. Simon. The Hartree-Fock theory for Coulomb systems. *Comm. Math. Phys.* 53.3 (1977), pp. 185–194.

[LT75] E.H. Lieb and W.E. Thirring. Bound for the kinetic energy of fermions which proves the stability of matter. *Phys. Rev. Lett.* 35 (1975), pp. 687–689.

[LT76] E.H. Lieb and W.E. Thirring. Inequalities for the moments of the eigenvalues of the Schrodinger Hamiltonian and their relation to Sobolev inequalities. *Studies in Math. Phys., Essays in Honor of Valentine Bargmann* (1976).

[NW96] Y. Netrusov and T. Weidl. On Lieb-Thirring inequalities for higher order operators with critical and subcritical powers. *Comm. Math. Phys.* 182.2 (1996), pp. 355–370.

[Roo51] C.C.J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.* 23.2 (1951), pp. 69–89.

[Roz72] G.V. Rozenblum. Distribution of the discrete spectrum of singular differential operators. *Soviet Math. Doki.* 202 (1972), pp. 1012–1015.

[Rum11] M. Rumin. Balanced distribution-energy inequalities and related entropy bounds. *Duke Math. J.* 160.3 (2011), pp. 567–597.

[SW71] E.M. Stein and G.L. Weiss. *Introduction to Fourier analysis on Euclidean spaces*. Princeton Univ. Press, 1971.

[Wei96] T. Weidl. On the Lieb-Thirring constants $L_{\gamma,1}$ for $\gamma \geq 1/2$. *Comm. Math. Phys.* 178.1 (1996), pp. 135–146.

Bibliographie générale

[AL78] M. Aizenman and E.H. Lieb. On semi-classical bounds for eigenvalues of Schrödinger operators. *Phys. Lett. A* 66.6 (1978), pp. 427–429.

[AA09] F. Alouges and C. Audouze. Preconditioned gradient flows for nonlinear eigenvalue problems and application to the Hartree-Fock functional. *Numerical Methods for Partial Differential Equations* 25.2 (2009), pp. 380–400.

[Bac81] G.B. Bacskay. A quadratically convergent Hartree–Fock (QC-SCF) method. Application to closed shell systems. *Chemical Physics* 61.3 (1981), pp. 385–404.

[Bar+10] C. Bardos et al. Setting and analysis of the multi-configuration time-dependent Hartree-Fock equations. *Archive for Rational Mechanics and Analysis* 198.1 (2010), pp. 273–330.

[BP87] J.M. Borwein and D. Preiss. A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions. *Trans. Amer. Math. Soc* 303.51 (1987), pp. 7–527.

[BES06] B. Buffoni, M.J. Esteban, and E. Séré. Normalized solutions to strongly indefinite semilinear equations. *Adv. Nonlinear Stud.* 6 (2 2006), pp. 323–347.

[CLB00a] E. Cancès and C. Le Bris. Can we outperform the DIIS approach for electronic structure calculations ? *International Journal of Quantum Chemistry* 79.2 (2000), pp. 82–90.

[CLB00b] E. Cancès and C. Le Bris. On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 34.4 (2000), pp. 749–774.

[CLBM06] E. Cancès, C. Le Bris, and Y. Maday. *Méthodes mathématiques en chimie quantique. Une introduction.* Vol. 53. Springer, 2006.

[CP08] E. Cancès and K. Pernal. Projected gradient algorithms for Hartree-Fock and density matrix functional theory calculations. *The Journal of chemical physics* 128 (2008), p. 134108.

[Can+03] E. Cancès et al. Computational quantum chemistry : a primer. *Handbook of numerical analysis* 10 (2003), pp. 3–270.

[Can98] E Cancès. “Molecular Simulation and Environmental Effects : A Mathematical and Numerical Perspective”. PhD thesis. Ecole des Ponts ParisTech, 1998.

[Can00] E. Cancès. SCF algorithms for Hartree-Fock electronic calculations. In: *Mathematical models and methods for ab initio quantum chemistry, Lecture Notes in Chemistry*. Vol. 74. 2000.

[CTDL73] C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Mécanique quantique*. 1973.

[Cwi77] M. Cwikel. Weak type estimates for singular values and the number of bound states of Schrodinger operators. *Ann. of Math.* (1977), pp. 93–100.

[Der12] J. Derezinski. Open problems about many-body Dirac operators. *Bulletin of International Association of Mathematical Physics* (2012).

[Dir28] P.A.M. Dirac. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 117.778 (1928), pp. 610–624.

[Dol12] J. Dolbeault. Personal communication. 2012.

[DLL08] J. Dolbeault, A. Laptev, and M. Loss. Lieb-Thirring inequalities with improved constants. *J. Eur. Math. Soc. (JEMS)* 10 (2008), pp. 1121–1126.

[Dol+06] J. Dolbeault et al. Lieb–Thirring type inequalities and Gagliardo–Nirenberg inequalities for systems. *J. Funct. Anal.* 238.1 (2006), pp. 193–220.

[DFJ07] K.G. Dyall and K. Faegri Jr. *Introduction to relativistic quantum chemistry*. Oxford University Press, 2007.

[EAS98] A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* 20 (1998), p. 303.

[Eke74] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications* 47.2 (1974), pp. 324–353.

[ES99] M.J. Esteban and E. Séré. Solutions of the Dirac-Fock equations for atoms and molecules. *Communications in Mathematical Physics* 203.3 (1999), pp. 499–530.

[ES01] M.J. Esteban and E. Séré. Nonrelativistic limit of the Dirac-Fock equations. *Annales Henri Poincare* 2 (5 2001), pp. 941–961.

[FG92] G. Fang and N. Ghoussoub. Second order information on Palais-Smale sequences in the mountain pass theorem. *Manuscripta mathematica* 75.1 (1992), pp. 81–95.

[Fey65] R.P. Feynman. *Feynman lectures on physics. Volume 3 : Quantum mechanics*. 1965.

[FMM04] J.B. Francisco, J.M. Martínez, and L. Martínez. Globally convergent trust-region methods for self-consistent field electronic structure calculations. *The Journal of chemical physics* 121 (2004), p. 10863.

[Fra+12] R.L. Frank et al. A positive density analogue of the Lieb-Thirring inequality. *Duke Math. J., to appear* (2012).

[Fri03a] G. Friesecke. On the infinitude of non-zero eigenvalues of the single-electron density matrix for atoms and molecules. *Proceedings of the Royal Society of London. Series A : Mathematical, Physical and Engineering Sciences* 459.2029 (2003), pp. 47–52.

[Fri03b] G. Friesecke. The multiconfiguration equations for atoms and molecules : charge quantization and existence of solutions. *Archive for Rational Mechanics and Analysis* 169.1 (2003), pp. 35–71.

[Gra07] I. Grant. *Relativistic Quantum Theory of Atoms and Molecules*. Springer, 2007.

[GH12] M. Griesemer and F. Hantsch. Unique Solutions to Hartree–Fock equations for closed-shell atoms. *Archive for Rational Mechanics and Analysis* 203 (3 2012), pp. 883–900.

[HJK03] A. Haraux, M.A. Jendoubi, and O. Kavian. Rate of decay to equilibrium in some semilinear parabolic equations. *Journal of Evolution equations* 3.3 (2003), pp. 463–484.

[HR90] B. Helffer and D. Robert. Riesz means of bounded states and semi-classical limit connected with a Lieb-Thirring conjecture II. *Ann. Inst. Henri Poincaré (A)* 53.2 (1990), pp. 139–147.

[Her77] I. Herbst. Spectral theory of the operator $(p^2 + m^2)^{-1/2} - Ze^2/r$. *Communications in Mathematical Physics* 53.3 (1977), pp. 285–294.

[Høs+08] S. Høst et al. The augmented Roothaan-Hall method for optimizing Hartree-Fock and Kohn-Sham density matrices. *The Journal of chemical physics* 129 (2008), p. 124106.

[HLT98] D. Hundertmark, E.H. Lieb, and L.E. Thomas. A sharp bound for an eigenvalue moment of the one-dimensional Schrodinger operator. *Adv. Theor. Appl. Math.* 2 (1998), pp. 719–732.

[ID93] P. Indelicato and J.P. Desclaux. Projection operator in the multiconfiguration Dirac-Fock method. *Physica Scripta* 1993.T46 (1993), p. 110.

[JOP+01] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy : Open source scientific tools for Python*. 2001–.

[Kat66] T. Kato. *Perturbation theory for linear operators*. Springer Verlag, 1966.

[KSC02] K.N. Kudin, G.E. Scuseria, and E. Cancès. A black-box self-consistent field convergence algorithm : One step closer. *The Journal of chemical physics* 116 (2002), p. 8255.

[LW00a] A. Laptev and T. Weidl. Recent results on Lieb-Thirring inequalities. *Journées Équations aux dérivées partielles* (2000), pp. 1–14.

[LW00b] A. Laptev and T. Weidl. Sharp Lieb-Thirring inequalities in high dimensions. *Acta Math.* 184.1 (2000), pp. 87–111.

[LB94] C. Le Bris. A general approach for multiconfiguration methods in quantum molecular chemistry. *Annales de l'Institut Henri Poincaré. Analyse non linéaire* 11.4 (1994), pp. 441–484.

[LBL05] C. Le Bris and P.L. Lions. From atoms to crystals : a mathematical journey. *Bull. Amer. Math. Soc* 42 (2005), pp. 291–363.

[LSY98] R.B. Lehoucq, D.C. Sorensen, and C. Yang. *ARPACK users' guide : solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. Vol. 6. Siam, 1998.

[Lev12a] A. Levitt. Best constants in Lieb-Thirring inequalities : a numerical investigation. Accepted for publication in Journal of Spectral Theory. 2012.

[Lev12b] A. Levitt. Convergence of gradient-based algorithms for the Hartree-Fock equations. *ESAIM Math. Model. Numer. Anal.* 46.06 (2012), pp. 1321–1336.

[Lev13] A. Levitt. Solutions of the multiconfiguration Dirac-Fock equations. Submitted. 2013.

[Lew04] M. Lewin. Solutions of the multiconfiguration equations in quantum chemistry. *Archive for Rational Mechanics and Analysis* 171.1 (2004), pp. 83–114.

[Lie76] E.H. Lieb. The number of bound states of one-body Schrödinger operators and the Weyl problem. *Bull. Amer. Math. Soc.* 82 (1976), pp. 751–753.

[Lie81] E.H. Lieb. Thomas-Fermi and related theories of atoms and molecules. *Rev. Mod. Phys.* 53.4 (1981), p. 603.

[Lie90] E.H. Lieb. The stability of matter : from atoms to stars. *Bull. Amer. Math. Soc.* 22.1 (1990).

[LL01] E.H. Lieb and M. Loss. Analysis. *American Mathematical Society, Providence, RI,* 4 (2001).

[LS10] E.H. Lieb and R. Seiringer. *The stability of matter in quantum mechanics*. Cambridge Univ. Press, 2010.

[LS77] E.H. Lieb and B. Simon. The Hartree-Fock theory for Coulomb systems. *Comm. Math. Phys.* 53.3 (1977), pp. 185–194.

[LT75] E.H. Lieb and W.E. Thirring. Bound for the kinetic energy of fermions which proves the stability of matter. *Phys. Rev. Lett.* 35 (1975), pp. 687–689.

[LT76] E.H. Lieb and W.E. Thirring. Inequalities for the moments of the eigenvalues of the Schrodinger Hamiltonian and their relation to Sobolev inequalities. *Studies in Math. Phys., Essays in Honor of Valentine Bargmann* (1976).

[Lio87] P.L. Lions. Solutions of Hartree-Fock equations for Coulomb systems. *Communications in Mathematical Physics* 109.1 (1987), pp. 33–97.

[Łoj65] S. Łojasiewicz. *Ensembles semi-analytiques*. Institut des Hautes Etudes Scientifiques, 1965.

[McW56] R. McWeeny. The density matrix in self-consistent field theory. I. Iterative construction of the density matrix. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 235.1203 (1956), p. 496.

[NW96] Y. Netrusov and T. Weidl. On Lieb-Thirring inequalities for higher order operators with critical and subcritical powers. *Comm. Math. Phys.* 182.2 (1996), pp. 355–370.

[Pul82] P. Pulay. Improved SCF convergence acceleration. *Journal of Computational Chemistry* 3.4 (1982), pp. 556–560.

[PD79] P. Pyykko and J.P. Desclaux. Relativity and the periodic system of elements. *Accounts of Chemical Research* 12.8 (1979), pp. 276–281.

[RS80] M. Reed and B. Simon. *Methods of modern mathematical physics : Functional analysis*. Vol. 1. 1980.

[RS11] T. Rohwedder and R. Schneider. An analysis for the DIIS acceleration method used in quantum chemistry calculations. *Journal of mathematical chemistry* 49.9 (2011), pp. 1889–1914.

[Roo51] C.C.J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.* 23.2 (1951), pp. 69–89.

[Roz72] G.V. Rozenblum. Distribution of the discrete spectrum of singular differential operators. *Soviet Math. Doki.* 202 (1972), pp. 1012–1015.

[Rum11] M. Rumin. Balanced distribution-energy inequalities and related entropy bounds. *Duke Math. J.* 160.3 (2011), pp. 567–597.

[Sal07] J. Salomon. Convergence of the time-discretized monotonic schemes. *ESAIM : Mathematical Modelling and Numerical Analysis* 41.01 (2007), pp. 77–93.

[SH73] V.R. Saunders and I.H. Hillier. A Level-Shifting method for converging closed shell Hartree-Fock wave functions. *International Journal of Quantum Chemistry* 7.4 (1973), pp. 699–705.

[Sid98] R.B. Sidje. Expokit : a software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)* 24.1 (1998), pp. 130–156.

[SW71] E.M. Stein and G.L. Weiss. *Introduction to Fourier analysis on Euclidean spaces*. Princeton Univ. Press, 1971.

[SO89] A. Szabo and N.S. Ostlund. *Modern quantum chemistry*. McGraw-Hill New York, 1989.

[Tha92] B. Thaller. *The Dirac equation*. Springer-Verlag, 1992.

[Tix98] C. Tix. Strict positivity of a relativistic Hamiltonian due to Brown and Ravenhall. *Bulletin of the London Mathematical Society* 30.3 (1998), pp. 283–290.

[Wei96] T. Weidl. On the Lieb-Thirring constants $L_{\gamma,1}$ for $\gamma \geq 1/2$. *Comm. Math. Phys.* 178.1 (1996), pp. 135–146.