A Numerical Perspective on Hartree-Fock-Bogoliubov Theory

Mathieu Lewin, Séverine Paul

A Numerical Perspective on Hartree-Fock-Bogoliubov Theory

Mathieu LEWIN
CNRS and Laboratoire de Mathématiques (CNRS UMR 8088)
Université de Cergy-Pontoise, 95 000 Cergy-Pontoise - France.
Email: mathieu.lewin@math.cnrs.fr

Séverine PAUL
Laboratoire de Mathématiques (CNRS UMR 8088)
Université de Cergy-Pontoise, 95 000 Cergy-Pontoise - France.
Email: severine.paul@u-cergy.fr

June 26, 2012

The method of choice for describing attractive quantum systems is Hartree-Fock-Bogoliubov (HFB) theory. This is a nonlinear model which allows for the description of *pairing effects*, the main explanation for the superconductivity of certain materials at very low temperature.

This paper is the first study of Hartree-Fock-Bogoliubov theory from the point of view of numerical analysis. We start by discussing its proper discretization and then analyze the convergence of the simple fixed point (Roothaan) algorithm. Following works by Cancès, Le Bris and Levitt for electrons in atoms and molecules, we show that this algorithm either converges to a solution of the equation, or oscillates between two states, none of them being a solution to the HFB equations. We also adapt the Optimal Damping Algorithm of Cancès and Le Bris to the HFB setting and we analyze it.

The last part of the paper is devoted to numerical experiments. We consider a purely gravitational system and numerically discover that pairing always occurs. We then examine a simplified model for nucleons, with an effective interaction similar to what is often used in nuclear physics. In both cases we discuss the importance of using a damping algorithm.

Contents

1. Introduction

Hartree-Fock-Bogoliubov (HFB) theory is the method of choice for describing some special features of attractive fermionic quantum systems [41]. It is a generalization of the famous Hartree-Fock (HF) method [35] used in quantum chemistry. It also generalizes the Bardeen-Cooper-Schrieffer (BCS) theory of superconductivity [4] which was invented in 1957 to explain the complete loss of resistivity of certain materials at very low temperature. In 1958 Bogoliubov realized in [9] that the BCS theory was actually very similar to his previous works [6, 8, 7] on the superfluidity of certain bosonic systems. He adapted it to fermions, leading to a model that is now called Hartree-Fock-Bogoliubov, and which is the main interest of this article.

In Hartree-Fock-Bogoliubov theory the state of the system is completely determined by two operators [3]. The *one-particle density matrix* γ is the same as in Hartree-Fock theory [34], whereas the *pairing density matrix* α describes the Cooper pairing effect. This effect can only hold in attractive systems. When the interaction potential is positive, the energy is decreased by replacing α by 0.

One of the most interesting questions in HFB theory is precisely the existence of pairing, that is, the non-vanishing of the matrix α for a minimizer. This question has been settled in the simpler translation-invariant BCS theory [5, 48, 49, 39, 18, 23, 26] and in some cases for the translation-invariant Hubbard model [3]. But it remains completely open for general attractive systems with a few particles, like those encountered in nuclear physics. In [30], the first author of this paper has shown with Lenzmann the existence of HFB minimizers for a purely Newtonian system of N fermions, but it is not yet known if $\alpha \neq 0$. One of the purpose of this work is to answer this question numerically.

The HFB energy is a nonlinear function of γ and α. Because of nonlinearity, minimizing this functional on a computer is not an easy task. The simpler Hartree-Fock model in which $\alpha = 0$ is now well understood from the point of view of numerical analysis [29, 11], even

though most authors have concentrated their attention to the special case of electrons in an atom or a molecule. Cancès and Le Bris have studied in [13, 12] the simpler fixed point algorithm called the Roothaan algorithm [42] and they have shown that this algorithm either converges or oscillates between two points, none of them being the solution of the HF equation. This result was recently improved by Levitt [31]. Cancès and Le Bris have also proposed a new algorithm called *Optimal Damping*, which is now used in several chemistry programs. It is based on the fact that one can freely minimize the energy over mixed HF states instead of pure HF states, by Lieb's variational principle [33].

Because the HFB model is an extension of HF theory, it is natural to believe that these ideas can be applied to the HFB case as well. In particular, under appropriate assumptions on the interaction potential, Bach, Fröhlich and Jonsson have recently shown in [2] an HFB equivalent of Lieb's variational principle. Up to some difficulties that will be explained later, we will show in this paper that the previously mentioned results can indeed be transposed to the Hartree-Fock-Bogoliubov model.

The paper is organized as follows. In the next section we quickly recall the basic formulation of Hartree-Fock-Bogoliubov theory. Then, in Section 3, we derive the discretized HFB equations and we prove that, in the limit of a large Galerkin basis set, the discretized solution converges to the true solution. We also discuss at length the possible symmetries of the system and we formulate the theory when these symmetries are taken into account.

In Section 4 we study the HFB Roothaan algorithm and we prove that it either converges to a solution of the HFB equation, or oscillate between two points, none of them being a solution of the equations. We then introduce an equivalent of the Optimal Damping Algorithm of Cancès and Le Bris, which is based on an optimization in the set of mixed HFB states.

Section 5 is devoted to the presentation of some numerical results. We first consider the purely gravitational model studied by Lenzmann and Lewin [30] and we numerically discover that there is always pairing. Then, we introduce a simplified model for nucleons, for which we as well present some preliminary numerical results. We particularly discuss the importance of using a damped algorithm instead of a simple fixed point method, a fact which has already been noticed in nuclear physics [16]. Our approach could help in improving the existing numerical techniques.

Acknowledgement. The authors would like to thank Laurent Bruneau and Julien Sabin for useful discussions. They acknowledge financial support from the French Ministry of Research (ANR-10-BLAN-0101) and from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement MNIQS 258023).

2. A quick review of Hartree-Fock-Bogoliubov theory

2.1. *Hartree-Fock-Bogoliubov states and their energy*

We consider a system composed of N identical fermions, described by the many-body Hamiltonian

$$H(N) = \sum_{j=1}^{N} T_j + \sum_{1 \leqslant k < \ell \leqslant N} W_{k\ell}, \tag{2.1}$$

acting on the fermionic N-body space $\mathfrak{H}^N = \bigwedge_1^N \mathfrak{H}$, where $\mathfrak{H}$ is the space for one particle. Here, $T : \mathfrak{H} \to \mathfrak{H}$ is a one-body operator and $W : \mathfrak{H}^2 \to \mathfrak{H}^2$ accounts for the interactions between the particles. We use the notation T_j for the operator T which acts on the jth component of the tensor product $\mathfrak{H}^N = \bigwedge_1^N \mathfrak{H}$, that is $T_j = 1 \otimes \cdots \otimes T \otimes \cdots \otimes 1$, and a similar convention for $W_{k\ell}$.

Most of what follows is valid in an abstract setting. However, for the sake of simplicity, in the whole paper we will restrict ourselves to the special case of nonrelativistic fermions with q internal degrees of freedom, moving in $\mathbb{R}^3$ ($q = 2$ for spin-1/2 particles like electrons). We also assume that no external force is applied, and that their interaction is translation-invariant. Then, in units where $m = 1/2$ and $\hbar = 1$, we have

$$\mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C}^q), \qquad T = -\Delta, \qquad W_{k\ell} = W(x_k - x_\ell).$$

The N-body space $\mathfrak{H}^N = \bigwedge_1^N L^2(\mathbb{R}^3, \mathbb{C}^q)$ consists of wave functions $\Psi(x_1, \sigma_1, ..., x_N, \sigma_N)$ which are antisymmetric with respect to exchanges of the variables (x_i, σ_i). In principle $W(x_k - x_\ell)$ is also a function of the two internal variables $\sigma_k, \sigma_\ell \in \{1, ..., q\}$ of the particles k and ℓ. Again for simplicity, we will assume that W only depends on the space variable $x_k - x_\ell$. Finally, we make the assumption that W is smooth and decays fast enough at infinity to ensure that H is bounded from below. To make this more explicit, we assume in the whole paper that

$$W = W_1 + W_2 \in L^p(\mathbb{R}^3) + L^q(\mathbb{R}^3) \quad \text{for some } 2 \leqslant p \leqslant q < \infty. \tag{2.2}$$

Sometimes we will make more precise assumptions on W.

We are interested in the case where W is attractive ($W \leqslant 0$), or at least partially attractive ($W \leqslant 0$ on a set of measure non zero). By translation invariance, the Hamiltonian $H(N)$ has no ground state (that is, the bottom of its spectrum cannot be an eigenvalue). But it may have one once the center of mass is removed, if W is sufficiently negative.

In a nonlinear model approximating the many-body problem above, there could be a ground state, even if the system is translation-invariant. Of course, translation invariance is not lost and there are then infinitely many ground states, obtained by translating the system arbitrarily. In Hartree-Fock theory [35, 3], such breaking of symmetry is known to occur for instance when $W(x) = -1/|x|$ is a purely gravitational interaction and $T = -\Delta$ (nonrelativistic), or $T = \sqrt{1 - \Delta} - 1$ (pseudo-relativistic), see [30, 32] and Theorem 2.3 below.

For attractive systems, it is often convenient to allow for another symmetry breaking, namely that of *particle number*. This means that the fixed particle number N is replaced by an operator $\mathcal{N}$ whose eigenvalues are $0, 1, 2, ...$ Only the *average particle number* is well defined for a quantum state. The classical way to define $\mathcal{N}$ is to introduce the fermionic Fock space

$$\mathcal{F} = \mathbb{C} \oplus \bigoplus_{n \geqslant 1} \mathfrak{H}^n,$$

which gathers all the possible n-particle subspaces in a direct sum. A (pure) quantum state in $\mathcal{F}$ is a vector $\Psi = \psi_0 \oplus \psi_1 \oplus \cdots$ which is normalized in the sense that

$$\|\Psi\|_{\mathcal{F}}^2 = |\psi_0|^2 + \sum_{n \geqslant 1} \|\psi_n\|_{\mathfrak{H}^n}^2 = 1.$$

The *average particle number* is the diagonal operator

$$\mathcal{N} := 0 \oplus \bigoplus_{n \geqslant 1} n,$$

such that the average number of particles in a state Ψ is given by the formula

$$\langle \Psi, \mathcal{N}\Psi \rangle = \sum_{n \geqslant 1} n \left\| \psi_n \right\|^2_{\mathfrak{H}^n}.$$

Instead of imposing that $\Psi \in \mathfrak{H}^N$ which is equivalent to Ψ being an eigenvector of $\mathcal{N}$, $\mathcal{N}\Psi = N\Psi$, we will only fix the average particle number of Ψ:

$$\langle \Psi, \mathcal{N}\Psi \rangle = N.$$

Allowing to have $\psi_n \neq 0$ for $n \neq N$ is useful to describe some physical properties of attractive systems. In most practical cases it is expected that the variance $\sum_{n \geqslant 0} (n-N)^2 \left\| \psi_n \right\|^2_{\mathfrak{H}^n}$ will be quite small, i.e. that Ψ will live in a neighborhood of $\mathfrak{H}^N$.

Similarly to the particle number operator $\mathcal{N}$, the many-body Hamiltonian $H(N)$ is now replaced by a many-body Hamiltonian $\mathbb{H}$ on Fock space

$$\mathbb{H} := 0 \oplus \bigoplus_{n \geqslant 1} H(n) \tag{2.3}$$

which is nothing else but the diagonal operator which coincides with $H(n)$ on each n-particle subspace. We will not discuss here the problem of defining $\mathbb{H}$ as a self-adjoint operator on $\mathcal{F}$.

The Hartree-Fock-Bogoliubov (HFB) model generalizes the well-known Hartree-Fock (HF) method and it allows for breaking of particle number in a very simple fashion. The method consists in restricting the many-body Hamiltonian $\mathbb{H}$ on $\mathcal{F}$ to a special class of states called *Hartree-Fock-Bogoliubov states* (or *quasi-free states*), which are completely characterized by their one-particle density matices [3].

Let us recall that a state in Fock space has two one-particle density matrices, instead of one in usual HF theory. These are two operators $\gamma : \mathfrak{H} \to \mathfrak{H}$ and $\alpha : \mathfrak{H} \to \mathfrak{H}$, which are defined by means of creation and annihilation operators by the relations [3]

$$\left\langle \Psi, a^\dagger(f) a(g) \Psi \right\rangle_{\mathcal{F}} = \langle g, \gamma f \rangle_{\mathfrak{H}}, \qquad \langle \Psi, a(f) a(g) \Psi \rangle_{\mathcal{F}} = \left\langle g, \alpha \overline{f} \right\rangle_{\mathfrak{H}}.$$

When Ψ lives in a particular N-particle subspace $\mathfrak{H}^N$, then $a(f)a(g)\Psi \in \mathfrak{H}^{N-2}$ hence $\langle \Psi, a(f)a(g)\Psi \rangle_{\mathcal{F}} = 0$ for all $f, g \in \mathfrak{H}$ and the matrix α vanishes. However for a general state $\Psi \in \mathcal{F}$, one can have $\alpha \neq 0$.

The two operators γ and α satisfy several constraints. First, we have $\gamma^* = \gamma$, $0 \leqslant \gamma \leqslant 1$ (in the sense of operators) and $\operatorname{Tr} \gamma = \langle \Psi, \mathcal{N}\Psi \rangle = N$, for the one-particle matrix γ. On the other hand, the so-called *pairing matrix* α satisfies $\alpha^T = -\alpha$. Its kernel $\alpha(x, \sigma, x', \sigma') = -\alpha(x', \sigma', x, \sigma)$ can thus be seen as an antisymmetric two-body wavefunction in $\mathfrak{H}^2$. It is interpreted as describing pairs of virtual particles, called *Cooper pairs*.

It is well known that a quantum state $\Psi \in \mathfrak{H}^N$ such that $(\gamma_\Psi)^2 = \gamma_\Psi$ is necessarily a Slater determinant (that is, a Hartree-Fock state). The same is true for states in Fock space. Consider a pair (γ, α) which is such that

$$\Gamma^2 = \Gamma, \quad \text{with} \quad \Gamma := \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1 - \overline{\gamma} \end{pmatrix} \tag{2.4}$$

on $\mathfrak{H} \oplus \mathfrak{H}$. Hence we have for instance $\alpha\alpha^* = \gamma - \gamma^2$. Then there exists a unique state Ψ in $\mathcal{F}$ which has γ and α as density matrices. This state has the property that any observable can be computed using only γ and α, by Wick's formula (see Thm 2.3 in [3]). The quantum states obtained by considering projections Γ are called Hartree-Fock-Bogoliubov states and they generalize usual Hartree-Fock states. When $\alpha \equiv 0$, then $\gamma = \sum_{j=1}^N |\varphi_j\rangle\langle\varphi_j|$ is a rank-N projection and the corresponding state is the usual Slater determinant

$$\Psi = \varphi_1 \wedge \cdots \wedge \varphi_N = \frac{1}{\sqrt{N!}} \det\big(\varphi_i(x_j, \sigma_j)\big).$$

When $\alpha \neq 0$, Ψ can be obtained by applying a Bogoliubov rotation to the vacuum but we will not explain this further. For the present work, we will only need the formula of the total energy, in terms of γ and α:

$$\begin{aligned}\langle \Psi, \mathbb{H}\Psi\rangle_{\mathcal{F}} &= \sum_{n\geqslant 0} \langle \psi_n, H(n)\psi_n\rangle_{\mathfrak{H}^n} \\ &= \mathrm{Tr}(-\Delta)\gamma + \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} W(x-y)\Big(\rho_\gamma(x)\rho_\gamma(y) - |\gamma(x,y)|^2 + |\alpha(x,y)|^2\Big)dx\,dy \\ &:= \mathcal{E}(\gamma,\alpha) \end{aligned} \tag{2.5}$$

where $\rho_\gamma(x) = \mathrm{Tr}_{\mathbb{C}^q}(\gamma(x,x))$ is the density of particles in the system. The terms in the double integral are respectively called the *direct*, *exchange* and *pairing* terms. Taking $\alpha \equiv 0$ one recovers the usual Hartree-Fock energy which has been studied by many authors [35, 38, 1, 3]. Our main goal in this paper is to investigate the minimization of the more complicated nonlinear functional $\mathcal{E}(\gamma,\alpha)$, when γ and α are submitted to the above constraints, and its numerical implementation. We will show below that the energy $\mathcal{E}$ is well defined in an appropriate function space, under our assumption (2.2) on W.

Note that the variance of the particle number for a HFB state Ψ in Fock space can be expressed only in terms of α by

$$\Big\langle \Psi, \big(\mathcal{N} - \langle\Psi,\mathcal{N}\Psi\rangle_{\mathcal{F}}\big)^2\Psi\Big\rangle_{\mathcal{F}} = \sum_{n\geqslant 0}(n-N)^2\,\|\psi_n\|^2_{\mathfrak{H}^n} = 2\,\mathrm{Tr}_{\mathfrak{H}}(\alpha\alpha^*),$$

see Lemma 2.7 in [3]. The spreading of the HFB state over the different spaces $\mathfrak{H}^n$ is therefore determined by the Hilbert-Schmidt norm of the pairing matrix α. We recover the fact that an HFB state has a given particle number if and only if its pairing matrix α vanishes.

2.2. *Pure vs mixed states*

In our previous description, we have only considered pure states in Fock space, that is states given by a normalized vector $\Psi \in \mathcal{F}$. For practical purposes, it is very convenient to extend the model to mixed states, which are nothing else but convex combinations of pure states, given by a (many-body) density matrix in $\mathcal{F}$

$$D = \sum_j \lambda_j |\Psi_j\rangle\langle\Psi_j| \quad \text{with} \quad \lambda_j \geqslant 0, \quad \sum_j \lambda_j = 1, \quad \langle\Psi_i,\Psi_j\rangle_{\mathcal{F}} = \delta_{ij}.$$

The average particle number and energy are then given by the formulas

$$\mathrm{Tr}_{\mathcal{F}}\big(\mathcal{N}D\big) = \sum_j \lambda_j\langle\Psi_j,\mathcal{N}\Psi_j\rangle, \qquad \mathrm{Tr}_{\mathcal{F}}\big(\mathbb{H}D\big) = \sum_j \lambda_j\langle\Psi_j,\mathbb{H}\Psi_j\rangle.$$

Resorting to mixed states is mandatory at positive temperature, when the equilibrium state of the system will actually always be a mixed state. But it is also very useful at zero temperature, even if the true ground state is a pure state. We will recall later the practical advantages of using mixed HFB states.

Similarly to what we have explained in the previous section, there is a class of *mixed* Hartree-Fock-Bogoliubov states, which are completely characterized by their one-particle density matrices γ and α. The latter now satisfy the constraint

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \leqslant \Gamma := \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1-\overline{\gamma} \end{pmatrix} \leqslant \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.6}$$

on $\mathfrak{H} \oplus \mathfrak{H}$, which is nothing else but the relaxation of the constraint $\Gamma^2 = \Gamma$ of pure states. The set of one-particle density matrices of mixed HFB states is therefore a convex set, whose extremal points are the density matrices of pure HFB states. One should remember that the set of mixed HFB states is *not* the convex hull of HFB pure states, however. The relation between the density matrices (γ, α) and the corresponding HFB states in $\mathcal{F}$ is highly nonlinear.

The energy of a mixed HFB state described by the density matrices (γ, α) is given by the same formula (2.5) as for pure states. Hence, minimizing this energy under the relaxed constraint (2.6) is equivalent to minimizing the full quantum energy over all mixed HFB states. The natural question arises whether a minimizer, when it exists, is automatically a pure state. The answer to this question is positive in many situations, as we will see below.

Before turning to the comparison between the minimization among pure and mixed states, we first introduce the variational sets on which the energy is well defined. The sets of all pure and mixed HFB states with finite kinetic energy are respectively given by

$$\mathcal{P} := \left\{(\gamma,\alpha) \in \mathfrak{S}_1(\mathfrak{H}) \times \mathfrak{S}_2(\mathfrak{H}) \ : \ \alpha^T = -\alpha,\ \Gamma = \Gamma^* = \Gamma^2,\ \mathrm{Tr}(-\Delta)\gamma < \infty\right\} \tag{2.7}$$

and

$$\mathcal{K} := \left\{(\gamma,\alpha) \in \mathfrak{S}_1(\mathfrak{H}) \times \mathfrak{S}_2(\mathfrak{H}) \ : \ \alpha^T = -\alpha,\ 0 \leqslant \Gamma = \Gamma^* \leqslant 1_{\mathfrak{H}\oplus\mathfrak{H}},\ \mathrm{Tr}(-\Delta)\gamma < \infty\right\}, \tag{2.8}$$

(the matrix Γ is the one appearing in (2.4)). Here $\mathfrak{S}_1(\mathfrak{H})$ and $\mathfrak{S}_2(\mathfrak{H})$ denote the spaces of trace-class and Hilbert-Schmidt operators [44]. The expression $\mathrm{Tr}(-\Delta)\gamma$ is to be understood in the sense of quadratic forms, that is

$$\mathrm{Tr}(-\Delta)\gamma = \sum_{k=1}^{3} \mathrm{Tr}(p_k \gamma p_k) \in [0,+\infty], \qquad \text{with} \quad p_k = -i\partial_{x_k}.$$

In practice we want to fix the total average number of particles. For this reason we also define the constrained sets

$$\mathcal{P}(N) := \{(\gamma,\alpha) \in \mathcal{P} \ : \ \mathrm{Tr}\,\gamma = N\} \tag{2.9}$$

and

$$\mathcal{K}(N) := \{(\gamma,\alpha) \in \mathcal{K} \ : \ \mathrm{Tr}\,\gamma = N\}, \tag{2.10}$$

of pure and mixed states with average particle number N. In practice N is an integer but it is convenient to allow any non-negative real number.

The following lemma says that the energy is a well-defined functional on the largest of the above sets $\mathcal{K}$, and that it is bounded from below on $\mathcal{K}(N)$ for any $N \geqslant 0$.

Lemma 2.1 (The HFB energy is bounded-below on $\mathcal{K}(N)$). *When $W = W_1 + W_2 \in L^p(\mathbb{R}^3) + L^q(\mathbb{R}^3)$ with $2 \leqslant p \leqslant q < \infty$, then $\mathcal{E}(\gamma,\alpha)$ is well defined for any $(\gamma,\alpha) \in \mathcal{K}$. It also satisfies a bound of the form*

$$\forall (\gamma,\alpha) \in \mathcal{K}, \qquad \mathcal{E}(\gamma,\alpha) \geqslant \frac{1}{2}\operatorname{Tr}(-\Delta)\gamma - C(N) \tag{2.11}$$

for some constant $C(N)$ depending only on $N = \operatorname{Tr}(\gamma)$.

Proof. The assumption that $W = W_1 + W_2 \in L^p(\mathbb{R}^3) + L^q(\mathbb{R}^3)$ with $2 \leqslant p \leqslant q < \infty$ implies that W is relatively form-bounded with respect to the Laplacian, with relative bound as small as we want [15]. This means $|W| \leqslant \epsilon(-\Delta) + C_\epsilon$ in the sense of quadratic forms, for all $\epsilon > 0$ and for some constant C_ϵ. This can now be used to verify that the energy is well defined under the assumption that $\operatorname{Tr}(-\Delta)\gamma < \infty$. First, we have for the direct term

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} |W(x-y)|\,\rho_\gamma(x)\rho_\gamma(y)\,dx\,dy \leqslant \epsilon N \int_{\mathbb{R}^3} |\nabla\sqrt{\rho_\gamma}|^2 + C_\epsilon\, N^2 \leqslant \epsilon N \operatorname{Tr}(-\Delta)\gamma + C_\epsilon\, N^2,$$

where in the last line we have used the Hoffmann-Ostenhof inequality [27],

$$\int_{\mathbb{R}^3} |\nabla\sqrt{\rho_\gamma}|^2 \leqslant \operatorname{Tr}(-\Delta)\gamma. \tag{2.12}$$

The exchange term is bounded similarly by applying the inequality $|W| \leqslant \epsilon(-\Delta) + C_\epsilon$ in x with y fixed:

$$\begin{aligned}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} |W(x-y)|\,|\gamma(x,y)|^2\,dx\,dy &\leqslant \epsilon \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} |\nabla_x\gamma(x,y)|^2\,dx\,dy + C_\epsilon \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} |\gamma(x,y)|^2\,dx\,dy \\ &= \epsilon \operatorname{Tr}(-\Delta)\gamma^2 + C_\epsilon \operatorname{Tr}\gamma^2 \leqslant \epsilon \operatorname{Tr}(-\Delta)\gamma + C_\epsilon N,\end{aligned}$$

since $\gamma^2 \leqslant \gamma$. Similarly we have, since $\alpha\alpha^* \leqslant \gamma - \gamma^2 \leqslant \gamma$,

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} |W(x-y)|\,|\alpha(x,y)|^2\,dx\,dy \leqslant \operatorname{Tr}\big(\epsilon(-\Delta) + C_\epsilon\big)\alpha\alpha^* \leqslant \epsilon \operatorname{Tr}(-\Delta)\gamma + C_\epsilon N.$$

All this shows that all the terms in the energy are well defined when $(\gamma,\alpha) \in \mathcal{K}(N)$. Also, we have

$$\mathcal{E}(\gamma,\alpha) \geqslant \big(1 - \epsilon - \epsilon N/2\big)\operatorname{Tr}(-\Delta)\gamma - C_\epsilon N - C_\epsilon N^2/2. \tag{2.13}$$

Taking $\epsilon = 1/(2+N)$ finishes the proof. □

Lemma 2.1 allows us to define the minimization problems for pure and mixed states as follows:

$$I(N) := \inf_{(\gamma,\alpha)\in\mathcal{K}(N)} \mathcal{E}(\gamma,\alpha), \tag{2.14}$$

$$J(N) := \inf_{(\gamma,\alpha)\in\mathcal{P}(N)} \mathcal{E}(\gamma,\alpha). \tag{2.15}$$

Of course we have $J(N) \geqslant I(N)$ since $\mathcal{P}(N) \subset \mathcal{K}(N)$. In many cases, we have that $I(N) = J(N)$ and that any minimizer, when it exists, is automatically a pure HFB state. We give two results in the literature going in this direction. The first deals with *purely repulsive interactions* and it is Lieb's famous variational principle [33] (see also Thm. 2.11 in [3]).

Theorem 2.1 (Lieb's HF Variational Principle [33]). *Assume that*

$$W \geqslant 0$$

and let N be an integer. Then for any $(\gamma,\alpha)\in\mathcal{K}(N)$, there exists $(\gamma',0)\in\mathcal{P}(N)$ such that

$$\mathcal{E}(\gamma,\alpha)\geqslant\mathcal{E}(\gamma',0).$$

In particular, we have $I(N)=J(N)$.

If $W>0$ a.e., then any minimizer for $I(N)$, when it exists, is necessarily of the form $(\gamma',0)$ with $(\gamma')^2=\gamma'$.

We see that for repulsive interactions, $W>0$, there is never pairing ($\alpha\equiv 0$) and the ground state is always a pure HF state, that is, a Slater determinant. The fact that, in HF theory, one can minimize over mixed states and get the same ground state energy is very important from a numerical point of view. This was used by Cancès and Le Bris [13, 12] to derive well-behaved numerical strategies, to which we will come back later in Section 4.2.

For *purely attractive interactions*, the following recent result of Bach, Fröhlich and Jonsson [2] is relevant.

Theorem 2.2 (HFB Constrained Variational Principle [2]). *Assume that the number of spin states is $q=2$ (spin-1/2 fermions), and that W can be decomposed in the form*

$$W(x-y)=-\int_\Omega d\mu(\omega)\mathbb{1}_{A_\omega}(x)\,\mathbb{1}_{A_\omega}(y) \tag{2.16}$$

on a given measure space (Ω,μ), with A_ω a measurable family of bounded domains in $\mathbb{R}^3$. Let N be any positive real number. Then for any $(\gamma,\alpha)\in\mathcal{K}(N)$, we have

$$\mathcal{E}(\gamma,\alpha)\geqslant\mathcal{E}(\gamma',\alpha'), \tag{2.17}$$

with

$$\gamma'=g\otimes\begin{pmatrix}1&0\\0&1\end{pmatrix},\qquad \alpha'=\sqrt{g(1-g)}\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix} \tag{2.18}$$

(the matrices act on the spin variables), and

$$g=g^T=\overline{g}=\frac{\gamma_{\uparrow\uparrow}+\gamma_{\downarrow\downarrow}+\overline{\gamma_{\uparrow\uparrow}}+\overline{\gamma_{\downarrow\downarrow}}}{4}.$$

This HFB state is pure: $(\gamma',\alpha')\in\mathcal{P}(N)$. In particular, we have $I(N)=J(N)$.

Furthermore, if $W<0$ a.e., then any minimizer for $I(N)$, when it exists, is necessarily of the previous form.

Remark 2.1. *Note that N does not have to be an even integer in this result. Since $\mathrm{Tr}_{L^2(\mathbb{R}^3)}(g)=N/2$, the operator g must have one eigenvalue different from 1 when N is odd, and it follows that $\alpha\neq 0$ in this special case.*

In Theorem 2.2, we have decomposed the operator γ acting on $L^2(\mathbb{R}^3\times\{\uparrow,\downarrow\},\mathbb{C})$ according to the spin variables as follows:

$$\gamma=\begin{pmatrix}\gamma_{\uparrow\uparrow}&\gamma_{\downarrow\uparrow}\\\gamma_{\uparrow\downarrow}&\gamma_{\downarrow\downarrow}\end{pmatrix}.$$

Theorem 2.2 says that when W satisfies (2.16), one can minimize over states which are pure, real, and have a simple spin symmetry. The antisymmetry of α is only contained in the spin

variables, hence the Cooper pairs are automatically in a singlet state. Of course, one can express the total energy only in terms of the real operator g, as follows

$$\mathcal{E}(\gamma',\alpha') = 2\ \mathrm{Tr}_{L^2(\mathbb{R}^3)}(-\Delta)g + \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} W(x-y)\Big(2\rho_g(x)\rho_g(y) - |g(x,y)|^2 + |\sqrt{g(1-g)}(x,y)|^2\Big).$$

In practice it will be more convenient to keep a pairing term $a(x,y)$ not *a priori* related to g and to optimize over both g and a, that is, to consider mixed states. When W satisfies the assumptions of the theorem, any ground state will automatically lead to $a = \pm\sqrt{g(1-g)}$.

Let us conclude our comments on Theorem 2.2, by noticing that several simple attractive potentials W can be written in the form (2.16). For instance the Fefferman-de la Llave formula [17]

$$\frac{1}{|x-y|} = \frac{1}{\pi}\int_0^\infty \frac{dr}{r^5}\int_{\mathbb{R}^3} dz\ \mathbb{1}_{B(z,r)}(x)\,\mathbb{1}_{B(z,r)}(y)$$

shows that a simple Newtonian interaction $W(x-y) = -|x-y|^{-1}$ is covered (here $\mathbb{1}_{B(z,r)}$ is the characteristic function of the ball centered at z, of radius r). Hainzl and Seiringer showed in [25] that any smooth enough radial function W can be written in the form

$$W(x-y) = \int_0^\infty dr\, g(r)\int_{\mathbb{R}^3} dz\ \mathbb{1}_{B(z,r)}(x)\,\mathbb{1}_{B(z,r)}(y)$$

for some explicit function g, whose sign can easily be studied.

2.3. *Existence results and properties of minimizers*

Before turning to the discretization and the numerical study of the HFB minimization problem, we make some comments on the existence of minimizers. In addition to the infinite dimension and the nonlinearity of the model, an important difficulty is the invariance under translations. For instance, there are always minimizing sequences which do not converge strongly (assuming there exists a minimizer, one can simply translate it far away).

Consider the electrons in an atom or in a molecule, with fixed classical nuclei (Born-Oppenheimer approximation). From the point of view of the electrons, the problem is no more translation-invariant, once the nuclei have been given a fixed position. Since the Coulomb interaction between the electrons is repulsive, $W(x-y) = |x-y|^{-1}$, Theorem 2.1 tells us that there is never pairing, $\alpha \equiv 0$. In this case there are several existence results, starting with the fundamental works of Lieb and Simon [35] and continuing with works by Lions [38], Bach [1], Solovej [46, 47].

For interactions W which have no particular sign, the pure HF problem was studied by Friesecke in [19] and by the first author of this paper in [32]. There is an HVZ-type theorem for HF wavefunctions which states that some binding conditions imply the existence of minimizers (Theorem 22 in [32]).

After the fundamental paper of Bach, Lieb and Solovej [3] with its study of the Hubbard model, to our knowledge the existence of ground states for the HFB model with pairing was only studied recently by Lenzmann and the first author of this paper in [30]. Some caricatures of HFB in nuclear physics had been previously considered by Gogny and Lions in [22].

We give here an existence result for the variational problem $I(N)$. In some cases we have $I(N) = J(N)$ (see the previous section) and for this reason, we only concentrate on $I(N)$. The next theorem can be proved by following the method of [30], which dealt with the more complicated case of a pseudo-relativistic kinetic operator $\sqrt{1-\Delta}-1$.

Theorem 2.3 (Existence of minimizers and compactness of minimizing sequences). *We assume as before that $W = W_1+W_2 \in L^p(\mathbb{R}^3)+L^q(\mathbb{R}^3)$ with $2 \leqslant p \leqslant q < \infty$. Let $\lambda > 0$. Then the following assertions are equivalent:*

(1) All the minimizing sequences $(\gamma_n, \alpha_n) \subset \mathcal{K}(\lambda)$ for $I(\lambda)$ are precompact up to translations, that is there exists a sequence $(x_k) \subset \mathbb{R}^3$ and $(\gamma, \alpha) \in \mathcal{K}(\lambda)$ such that, for a subsequence,

$$\lim_{k\to\infty} \left\|(1-\Delta)^{1/2}(\tau_{x_k}\gamma_{n_k}\tau_{-x_k}-\gamma)(1-\Delta)^{1/2}\right\|_{\mathfrak{S}_1} = \lim_{k\to\infty} \left\|(1-\Delta)^{1/2}(\tau_{x_k}\alpha_{n_k}\tau_{-x_k}-\alpha)\right\|_{\mathfrak{S}_2} = 0.$$

In particular (γ, α) is a minimizer for $I(\lambda)$.

(2) The binding inequalities

$$I(\lambda) < I(\lambda-\mu)+I(\mu) \qquad \text{for all } 0<\mu<\lambda \tag{2.19}$$

are satisfied.

Furthermore, if W is Newtonian at infinity, that is

$$W(x) \leqslant -\frac{a}{|x|} \quad \text{for } a>0 \text{ and } |x| \geqslant R, \tag{2.20}$$

then the previous two equivalent conditions are verified.

The assumption that the interaction is Newtonian at infinity is a big simplification, as it means that two subsystems receiding from each other always attract at large distances. One can expect that minimizers exist even if the potential is not attractive at infinity, as soon as it has a sufficiently large negative component. A typical effective potential $W(x)$ used in nuclear physics is nonnegative for small $|x|$, and has a negative well at intermediate distances [41]. At infinity it typically decays like $+\kappa|x|^{-1}$ for two protons, and exponentially fast when one of the two particles is a neutron. Even in HF theory, we are not aware of any existence result dealing with such potentials, however. We will numerically investigate a model of this form in Section 5.3.

The form of the nonlinear equation solved by minimizers is well-known in the physics literature, and it was re-explained in [3]. The following result summarizes some known properties.

Theorem 2.4 (HFB equation and properties of minimizers [3, 30]). *A HFB minimizer on $\mathcal{K}(N)$ solves the nonlinear equation*

$$\Gamma = \mathbb{1}_{(-\infty,0)}\big(F_\Gamma - \mu\mathcal{N}\big) + \delta \tag{2.21}$$

where $0 \leqslant \delta \leqslant \mathbb{1}_{\{0\}}(F_\Gamma - \mu\mathcal{N})$ has the same form as Γ, and where

$$\mathcal{N} := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad F_\Gamma = \begin{pmatrix} h_\gamma & \pi \\ \pi^* & -\overline{h_\gamma} \end{pmatrix} \tag{2.22}$$

with $h_\gamma = -\Delta + \rho_\gamma * W - W(x-y)\gamma(x,y)$ and $\pi(x,y) = \alpha(x,y)W(x-y)$.

If $W(x-y) = -\kappa|x-y|^{-1}$ (Newtonian interaction) and N is an integer, then all the minimizers are of the special form (2.18)*. In this case, we have either $\alpha \equiv 0$ and γ is a projector of rank N, or $\alpha \neq 0$ and γ has an infinite rank.*

The nonlinear equation (2.21) is in principle similar to the usual equation obtained in HF theory,

$$\gamma = \mathbb{1}_{(-\infty,\mu)}(h_\gamma) + \delta \tag{2.23}$$

Indeed, (2.21) reduces to (2.23) when $\alpha \equiv 0$. Let us however emphasize that the mean-field operator F_Γ has a spectrum which is symmetric with respect to 0. Hence F_Γ is usually not even semi-bounded, on the contrary to h_γ which is always bounded from below. Furthermore, the operator $\mathcal{N}$ does *not* commute with F_Γ (except when $\alpha \equiv 0$) and the equation cannot be written in a simple form as in HF theory. This will cause several difficulties to which we will come back at length later.

The fact that γ has an infinite rank when there is pairing, $\alpha \neq 0$, is a dramatic change of behavior compared to the simple HF case. However, no information on the decay of the eigenvalues of γ seems to be known.

An important open question is to show that minimizers actually exhibit non-vanishing pairing $\alpha \neq 0$, at least for a sufficiently strong attractive potential W. On heuristic grounds, one expects such a phenomenon of *"Cooper pair formation"* to be energetically favorable due to the attractive interaction among particles. However, it seems to be a formidable task to find mathematical proof for this claim. The existence of pairing is known in some particular situations (when N is odd and W is Newtonian, see Remark 2.1, for the Hubbard model [3], or in translation-invariant BCS theory [5, 48, 49, 39, 18, 23, 26]), but for the model presented here, we are not aware of any result of this sort. One of the purposes of this paper is to investigate this question numerically.

3. Discretized Hartree-Fock-Bogoliubov theory

In this section, we write and study the Hartree-Fock-Bogoliubov model in a discrete basis.

3.1. *Convergence analysis*

Here we show that the ground state HFB energy in a finite basis converges to the true HFB ground state energy when the size of the basis grows. We consider a sequence of finite-dimensional spaces $V_h \subset H^1(\mathbb{R}^3, \mathbb{C}^q)$ for $h \to 0$. We assume that any function $f \in H^1(\mathbb{R}^3, \mathbb{C}^q)$ can be approximated by functions from V_h:

$$\forall f \in H^1(\mathbb{R}^3, \mathbb{C}^q), \quad \exists f_h \in V_h \quad \text{such that} \quad \|f - f_h\|_{H^1} \underset{h\to 0}{\longrightarrow} 0. \tag{3.1}$$

We typically think of a sequence V_h given by the Finite Elements Method. Let π_h denote the orthogonal projection on V_h in $L^2(\mathbb{R}^3, \mathbb{C}^q)$. We define the set of density matrices living on V_h (with average particle number N) as follows

$$\mathcal{K}_h(N) = \big\{(\gamma,\alpha) \in \mathcal{K}(N) \ : \ \pi_h \gamma \pi_h = \gamma, \ \pi_h \alpha \overline{\pi_h} = \alpha\big\}. \tag{3.2}$$

The corresponding minimization problem is now

$$I_h(N) = \inf_{(\gamma,\alpha)\in\mathcal{K}_h(N)} \mathcal{E}(\gamma,\alpha). \tag{3.3}$$

Since $\mathcal{K}_h(N) \subset \mathcal{K}(N)$ by definition, it is obvious that $I_h(N) \geqslant I(N)$. The following result is a consequence of Theorem 2.3.

Theorem 3.1 (Convergence of the approximate HFB problem). *When $W = W_1 + W_2 \in L^p(\mathbb{R}^3) + L^q(\mathbb{R}^3)$ with $2 \leqslant p \leqslant q < \infty$ and under Assumption* (3.1) *on the sequence (V_h), we have*

$$\lim_{h\to 0} I_h(N) = I(N). \tag{3.4}$$

If the binding inequality (2.19) *is satisfied, then any sequence of minimizers $(\gamma_h, \alpha_h) \in \mathcal{K}_h(N)$ for $I_h(N)$ converges, up to a subsequence and up to a translation, to a minimizer $(\gamma, \alpha) \in \mathcal{K}(N)$ of $I(N)$, in the sense that*

$$\lim_{h_k\to 0} \left\| (1-\Delta)^{1/2}(\tau_{x_k}\gamma_{h_k}\tau_{-x_k} - \gamma)(1-\Delta)^{1/2} \right\|_{\mathfrak{S}_1} = \lim_{h_k\to 0} \left\| (1-\Delta)^{1/2}(\tau_{x_k}\alpha_{h_k}\tau_{-x_k} - \alpha) \right\|_{\mathfrak{S}_2} = 0. \tag{3.5}$$

Proof. We only have to show that $I_h(N) \to I(N)$ as $h \to 0$. Then, any sequence of exact minimizers (γ_h, α_h) for $I_h(N)$ is also a minimizing sequence for $I(N)$. Applying Theorem 2.3 concludes the proof.

We know that finite-rank operators are dense in $\mathcal{K}(N)$. Let $(\gamma, \alpha) \in \mathcal{K}(N)$ be any such finite rank operator. Let $(f_i)_{i=1}^K$ be an orthonormal basis of the range of γ. By Löwdin's theorem (Lemma 13 in [32]), we know that the two-body wavefunction α can be expanded in the same basis $(f_1, ..., f_K)$. Now, for every $i = 1, .., K$, we apply (3.1) and take a sequence $f_i^h \in V_h$ be such that $f_i^h \to f_i$ in $H^1(\mathbb{R}^3)$. The system $(f_i^h)_{i=1}^k$ is not necessarily orthonormal but we have $\langle f_i^h, f_j^h \rangle \to \langle f_i, f_j \rangle = \delta_{ij}$. Applying the Gram-Schmidt procedure, we can therefore replace the $(f_i^h)_{i=1}^K$ by an orthonormal set $(g_i^h)_{i=1}^K \subset V_h$ having the same properties. An equivalent procedure is to take $g_i^h = \sum_{j=1}^K (S_h^{-1/2})_{ji} f_j^h$ where S_h is the Gram matrix $(\langle f_i^h, f_j^h \rangle)$. Let now U_h be any unitary operator on $L^2(\mathbb{R}^3)$ which is such that $U_h f_i = g_i^h$ for all $i = 1, ..., K$. We then take $\gamma_h := U_h \gamma U_h^*$ and $\alpha_h := U_h \alpha U_h^T$. In words, we just replace the f_i by g_i^h in the decomposition of γ and α. To see that $(\gamma_h, \alpha_h) \in \mathcal{K}(N)$, we just notice that

$$\begin{pmatrix} \gamma_h & \alpha_h \\ \alpha_h^* & 1-\overline{\gamma_h} \end{pmatrix} = \begin{pmatrix} U_h & 0 \\ 0 & \overline{U_h} \end{pmatrix} \begin{pmatrix} \gamma & \alpha \\ \alpha^* & 1-\overline{\gamma} \end{pmatrix} \begin{pmatrix} U_h^* & 0 \\ 0 & \overline{U_h}^* \end{pmatrix}.$$

Note also that $\mathrm{Tr}(\gamma_h) = \mathrm{Tr}(\gamma) = N$ since U_h is unitary. Now γ_h and α_h belong to $\mathcal{K}_h(N)$ by definition, hence we have that $\mathcal{E}(\gamma_h, \alpha_h) \geqslant I_h(N)$. On the other hand, by the convergence of f_i^h (hence of g_i^h) towards f_i in $H^1(\mathbb{R}^3)$, we easily see that

$$\lim_{h\to 0} \mathcal{E}(\gamma_h, \alpha_h) = \mathcal{E}(\gamma, \alpha)$$

by continuity of $\mathcal{E}$. Therefore we have proved that

$$\limsup_{h\to 0} I_h(N) \leqslant \mathcal{E}(\gamma, \alpha).$$

This is valid for all finite-rank $(\gamma, \alpha) \in \mathcal{K}(N)$, hence we deduce that

$$\limsup_{h\to 0} I_h(N) \leqslant \inf_{(\gamma,\alpha)\in\mathcal{K}(N)} \mathcal{E}(\gamma, \alpha) = I(N).$$

On the other hand the inequality $I_h(N) \geqslant I(N)$ is always satisfied, hence we have proved the claimed convergence $I_h(N) \to I(N)$. □

3.2. *Discretization*

In this section we compute the HFB energy $\mathcal{E}(\gamma,\alpha)$ of a discretized state $(\gamma,\alpha)\in\mathcal{K}_h(N)$ and we write the corresponding self-consistent equation. We fix once and for all the approximation space V_h and we consider a basis set $(\chi_i)_{i=1}^{N_b}$ of V_h, which is not necessarily orthonormal. We will assume that V_h is stable under complex conjugation, which means that $f\in V_h\Rightarrow\overline{f}\in V_h$ (this amounts to replacing V_h by $\mathrm{Span}(V_h,\overline{V_h})$). Then we can choose a basis $(\chi_i)_{i=1}^{N_b}$ which is real, that is $\overline{\chi_i}=\chi_i$ for all $i=1,...,N_b$. This will dramatically simplify the calculation below.

Since $\pi_h\gamma\pi_h=\gamma$ and $\pi_h\alpha\overline{\pi_h}=\alpha$, we can write the kernels of γ and α as follows

$$\gamma(x,y)_{\sigma,\sigma'}=\sum_{i,j=1}^{N_b}G_{ij}\,\chi_i(x)_\sigma\overline{\chi_j(y)_{\sigma'}},\qquad \alpha(x,y)_{\sigma,\sigma'}=\sum_{i,j=1}^{N_b}A_{ij}\,\chi_i(x)_\sigma\chi_j(y)_{\sigma'}. \tag{3.6}$$

The complex conjugation on $\overline{\chi_j}$ is superfluous but we keep it to emphasize the difference between γ and α. The matrices G and A (defined by the previous relation) satisfy the constraints $G^*=G$ and $A^T=-A$. Note that since A is antisymmetric, we can also write

$$\begin{aligned}\alpha(x,y)_{\sigma,\sigma'}&=\sum_{1\leqslant i<j\leqslant N_b}A_{ij}\left(\chi_i(x)_\sigma\chi_j(y)_{\sigma'}-\chi_j(x)_\sigma\chi_i(y)_{\sigma'}\right)\\&=\sqrt{2}\sum_{1\leqslant i<j\leqslant N_b}A_{ij}\,\chi_i\wedge\chi_j(x,\sigma,y,\sigma')\end{aligned}$$

where $(\chi_i\wedge\chi_j)(x,\sigma;y,\sigma'):=\left(\chi_i(x)_\sigma\chi_j(y)_{\sigma'}-\chi_i(y)_{\sigma'}\chi_j(x)_\sigma\right)/\sqrt{2}$ is a two-body Slater determinant. Let us also remark that (3.6) can be written in the operator form

$$\gamma=\sum_{i,j=1}^{N_b}G_{ij}\,|\chi_i\rangle\langle\chi_j|,\qquad \alpha=\sum_{i,j=1}^{N_b}A_{ij}\,|\chi_i\rangle\langle\overline{\chi_j}|.$$

Again the complex conjugation on $\overline{\chi_j}$ is superfluous.

Let us define the so-called *overlap matrix* $S=S^*=\overline{S}$ by

$$S_{ij}=\langle\chi_i,\chi_j\rangle_{\mathfrak{H}}=\sum_{\sigma=1}^{q}\int_{\mathbb{R}^3}\overline{\chi_i(x)_\sigma}\chi_j(x)_\sigma\,dx=\sum_{\sigma=1}^{q}\int_{\mathbb{R}^3}\chi_i(x)_\sigma\chi_j(x)_\sigma\,dx. \tag{3.7}$$

It is tedious but straightforward to verify that the constraint

$$\begin{pmatrix}0&0\\0&0\end{pmatrix}\leqslant\Gamma:=\begin{pmatrix}\gamma&\alpha\\\alpha^*&1-\overline{\gamma}\end{pmatrix}\leqslant\begin{pmatrix}1&0\\0&1\end{pmatrix}$$

can be written for the matrices G and A in the form

$$\begin{pmatrix}0&0\\0&0\end{pmatrix}\leqslant\begin{pmatrix}SGS&SAS\\SA^*S&S-S\overline{G}S\end{pmatrix}\leqslant\begin{pmatrix}S&0\\0&S\end{pmatrix} \tag{3.8}$$

or, equivalently,

$$0\leqslant\Upsilon\mathbf{S}\Upsilon\leqslant\Upsilon \tag{3.9}$$

where Υ and $\mathbf{S}$ are defined by

$$\Upsilon:=\begin{pmatrix}G&A\\A^*&S^{-1}-\overline{G}\end{pmatrix}\quad\text{and}\quad\mathbf{S}:=\begin{pmatrix}S&0\\0&S\end{pmatrix}. \tag{3.10}$$

Another way to write the constraint is $0 \leqslant \mathbf{S}^{1/2}\Upsilon\mathbf{S}^{1/2} \leqslant 1$. Let us notice that we have used everywhere the fact that $S = S^* = \overline{S} = S^T$. The formulas are much more complicated when S is not real.

The energy can be expressed in terms of the matrices G and A, as well. A calculation shows that

$$\mathcal{E}(\gamma, \alpha) = \mathrm{Tr}(hG) + \frac{1}{2}\,\mathrm{Tr}(G\,J(G)) - \frac{1}{2}\,\mathrm{Tr}(G\,K(G)) + \frac{1}{2}\,\mathrm{Tr}(A^*\,X(A)). \tag{3.11}$$

The trace here is the usual one for $N_b \times N_b$ matrices. As we think that there is no possible confusion with $\mathcal{E}(\gamma, \alpha)$, we will also denote by $\mathcal{E}(G, A)$ this discretized energy functional. In formula (3.11),

$$h_{ij} = \langle \chi_i, (-\Delta)\chi_j \rangle = \sum_{\sigma=1}^{q} \int_{\mathbb{R}^3} \nabla\chi_i(x)_\sigma \cdot \nabla\chi_j(x)_\sigma \, dx,$$

$$J(G)_{ij} = \sum_{k,\ell=1}^{N_b} (ij|\ell k)\, G_{k\ell}, \qquad K(G)_{ij} = \sum_{k,\ell=1}^{N_b} (ik|\ell j)\, G_{k\ell}, \qquad X(A)_{ij} = \sum_{k,\ell=1}^{N_b} (ik|j\ell)\, A_{k\ell}, \tag{3.12}$$

and

$$(ij|k\ell) := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} W(x-y)\,\chi_i(x)^*\;\chi_j(x)\;\chi_k(y)^*\;\chi_\ell(y)\,dx\,dy. \tag{3.13}$$

We use here the notation $a^*b = \sum_{\sigma=1}^{q} \overline{a}_\sigma\, b_\sigma$ (but the complex conjugation is superfluous for our real basis, hence $K = X$). Similarly, we have

$$\mathrm{Tr}(\gamma) = \mathrm{Tr}(SG). \tag{3.14}$$

We define the discretized number operator as

$$\mathbf{N} = \begin{pmatrix} S & 0 \\ 0 & -S \end{pmatrix} \tag{3.15}$$

The constraint $\mathrm{Tr}(\gamma) = N$ can be written equivalently as

$$\mathrm{Tr}(\mathbf{N}\Upsilon) = 2N - N_b$$

We deduce from this calculation that the variational problem $I_h(N)$ can be written in finite dimension as

$$I_h(N) = \min\Big\{\mathcal{E}(G, A)\ :\ 0 \leqslant \Upsilon\mathbf{S}\Upsilon \leqslant \Upsilon,\ \mathrm{Tr}(\mathbf{N}\Upsilon) = 2N - N_b\Big\} \tag{3.16}$$

where we recall that Υ and $\mathbf{S}$ have been defined in (3.10). Here the infimum is always attained because the problem is finite dimensional.

In this form, the discretized problem is very similar to the usual discretized Hartree-Fock problem [11, 29], in dimension $2N_b$ instead of N_b. There is a big difference, however. In HF theory the constraint involves the matrix $\mathbf{S}$ instead of $\mathbf{N}$. This difference will cause several difficulties. To understand the problem, let us introduce a new variable $\Upsilon' = \mathbf{S}^{1/2}\Upsilon\mathbf{S}^{1/2}$ (which is the same as orthonormalize the basis (χ_i)). Then the constraint $0 \leqslant \Upsilon\mathbf{S}\Upsilon \leqslant \Upsilon$ is transformed into $0 \leqslant \Upsilon' \leqslant 1$. However, the constraint on the number of particles becomes $\mathrm{Tr}(\mathbf{S}^{-1/2}\mathbf{N}\mathbf{S}^{-1/2}\Upsilon') = 2N - N_b$. In usual Hartree-Fock theory, the matrix $\mathbf{S}^{-1/2}\mathbf{N}\mathbf{S}^{-1/2}$ is

replaced by the identity. The fact that this matrix then commutes with the Fock Hamiltonian (defined below) simplifies dramatically the self-consistent equations. Here we get

$$\mathbf{S}^{-1/2}\mathbf{N}\mathbf{S}^{-1/2} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which commutes with the Fock Hamiltonian if and only if $A \equiv 0$.

The self-consistent equation is obtained like in [3]. The result is as follows.

Lemma 3.1 (Discretized HFB equation). *Let Υ be a minimizer for the variational problem $I_h(N)$. Then there exists $\mu \in \mathbb{R}$ such that Υ solves the linear problem*

$$\min \Big\{ \operatorname{Tr}(\mathbf{F}_\Upsilon - \mu\mathbf{N})\tilde{\Upsilon} \ : \ 0 \leqslant \tilde{\Upsilon}\mathbf{S}\tilde{\Upsilon} \leqslant \tilde{\Upsilon} \Big\} \tag{3.17}$$

where

$$\mathbf{F}_\Upsilon := \begin{pmatrix} h_G & X(A) \\ X(A)^* & -\overline{h_G} \end{pmatrix}, \qquad h_G := h + J(G) - K(G). \tag{3.18}$$

The solution Υ can be written in the form

$$\Upsilon = \mathbf{S}^{-1/2}\Big(\mathbb{1}_{(-\infty,0)}\Big(\mathbf{S}^{-1/2}(\mathbf{F}_\Upsilon - \mu\mathbf{N})\mathbf{S}^{-1/2}\Big) + \delta \Big)\mathbf{S}^{-1/2} \tag{3.19}$$

where $0 \leqslant \delta \leqslant 1$ lives only in the kernel of $\mathbf{S}^{-1/2}(\mathbf{F}_\Upsilon - \mu\mathbf{N})\mathbf{S}^{-1/2}$.

The solution Υ of the self-consistent equation (3.19) may be equivalently written by considering the generalized eigenvalue problem

$$\big(\mathbf{F}_\Upsilon - \mu\mathbf{N}\big) f_i = \epsilon_i \, \mathbf{S} \, f_i, \qquad \langle f_i, \mathbf{S} f_j \rangle = \delta_{ij}. \tag{3.20}$$

Then we have simply (assuming $\epsilon_i \neq 0$ for all $i = 1, ..., 2N_b$)

$$\Upsilon = \sum_{\epsilon_i < 0} f_i \, f_i^*.$$

Again, this is similar to the Hartree-Fock solution [11, 29] except that μ is unknown and $\mathbf{N}$ does not always commute with $\mathbf{F}_\Upsilon$.

Remark that although the basis functions χ_i are real, the density matrix Υ is not necessarily real. In the next section, we will restrict ourselves to real-valued density matrices and impose some spin symmetry.

3.3. *Using symmetries*

The HFB energy $\mathcal{E}(\Gamma)$ has some natural symmetry invariances which we describe in detail in this section. Recall that since $\mathcal{E}$ is a *nonlinear* functional, it cannot be guaranteed that the HFB minimizers will all have the same symmetries as the HFB energy. The set of all minimizers will be invariant under the action of the symmetry group, but each minimizer alone does not have to be invariant.

We have already allowed for the breaking of particle-number symmetry and we hope to find an HFB ground state. It will then automatically break the translational invariance of the system. There are three other symmetries (spin, complex conjugation and rotations) which are of interest to us. We have the choice of imposing these symmetries by adding appropriate constraints, or not. Because this reduces the computational cost, it will be convenient to impose them.

3.3.1. *Time-reversal symmetry*

Let us now assume that $q = 2$, which means that our fermions are spin-1/2 particles. Since the Laplacian and the interaction function W do not act on the spin variable, the HFB energy has some *spin symmetry*, which can be written for $q = 2$ as

$$\forall k = 1, 2, 3, \qquad \mathcal{E}(\Sigma_k \Gamma \Sigma_k^*) = \mathcal{E}(\Gamma)$$

where

$$\Sigma_k := \begin{pmatrix} i\sigma_k & 0 \\ 0 & i\sigma_k \end{pmatrix},$$

with $\sigma_1, \sigma_2, \sigma_3$ being the usual Pauli matrices. Note that Σ_k has the form of a Bogoliubov transformation, hence $\Sigma_k \Gamma \Sigma_k^*$ is also an HFB state. The number operator is also invariant which means that

$$\Sigma_k \mathcal{N} \Sigma_k^* = \mathcal{N}$$

for all $k = 1, 2, 3$. Thus, the contraint $\mathrm{Tr}(\gamma) = N$ is conserved and we have $\Sigma_k \Gamma \Sigma_k^* \in \mathcal{K}(N)$ when $\Gamma \in \mathcal{K}(N)$.

Another important symmetry is that of *complex conjugation* which means this time that

$$\mathcal{E}(\overline{\Gamma}) = \mathcal{E}(\Gamma)$$

and which is based on the fact that the Laplacian and W are real operators. Again we have $\mathrm{Tr}(\overline{\gamma}) = \mathrm{Tr}(\gamma)$ hence $\mathcal{K}(N)$ is invariant under complex conjugation.

As was explained in [24] (see Remark 5 page 1032), the density matrices (γ, α) can be written in the special form

$$\gamma' = g \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \alpha' = a \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \text{with } g = g^T = \overline{g} \text{ and } a = a^T = \overline{a} \tag{3.21}$$

(the 2×2 matrix refers to the spin variables), if and only if

$$\begin{cases} \Sigma_k \Gamma \Sigma_k^* = \Gamma, & \text{for } k = 1, 2, \text{ and} \\ \overline{\Gamma} = \Gamma. \end{cases} \tag{3.22}$$

In other words, Γ is invariant under the action of the group generated by Σ_1, Σ_2 and $\mathcal{C}$. This invariance is sometimes called the *time-reversal symmetry*. As remarked in [24], imposing $\Sigma_k \Gamma \Sigma_k^* = \Gamma$ for all $k = 1, 2, 3$ implies $\alpha \equiv 0$ which is not interesting for us.

When W can be written in the form (2.16), the Theorem 2.2 of Bach, Fröhlich and Jonsson [2], tells us that there is no breaking of the time-reversal symmetry. That is, we can always minimize over such special states. For other interactions W this is not necessarily true but it is often convenient to impose this symmetry anyhow.

Because it holds

$$F_{\Sigma_k \Gamma \Sigma_k^*} = \Sigma_k F_\Gamma \Sigma_k^*, \quad F_{\overline{\Gamma}} = \overline{F_\Gamma},$$

it can then be verified that minimizers under the additional symmetry constraint, satisfy the same self-consistent equation as when no constraint is imposed.

When we discretize the problem by imposing time-reversal symmetry, we use two real symmetric matrices G and A, related through the constraint that

$$0 \leqslant \Upsilon \mathbf{S} \Upsilon \leqslant \Upsilon, \quad \text{with} \quad \Upsilon := \begin{pmatrix} G & A \\ A & S^{-1} - G \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}. \tag{3.23}$$

The energy becomes

$$\mathcal{E}(G,A) = 2\,\mathrm{Tr}(hG) + 2\,\mathrm{Tr}(G\,J(G)) - \mathrm{Tr}(G\,K(G)) + \mathrm{Tr}(A\,K(A)). \tag{3.24}$$

and the associated particle number constraint is $\mathrm{Tr}(SG) = N/2$. In practice we always assume that N is even for simplicity. The basis (χ_i) is now composed of (real-valued) functions in $H^1(\mathbb{R}^3,\mathbb{R})$, instead of functions in $H^1(\mathbb{R}^3,\mathbb{C}^q)$ as before, and the formulas for S, h, J, K and X are the same as before.

3.3.2. *Rotational symmetry*

The group $SO(3)$ of rotations in $\mathbb{R}^3$ also acts on HFB states and it leaves the energy invariant when the interaction W is a radial function. In this section we always assume that the spin variable has already been removed according to the previous section and we denote by g and a the corresponding (real symmetric) density matrices. If the spin were still present, rotations would act on it as well.

To any rotation $R \in SO(3)$ we can associate a unitary operator on $L^2(\mathbb{R}^3)$, denoted also by R, defined by $(R\varphi)(x) = \varphi\big(R^{-1}x\big)$. Now we say that an HFB state Γ with density matrices (γ,α) is invariant under rotations when it satisfies

$$\mathscr{R}\,\Gamma\,\mathscr{R}^* = \Gamma, \qquad \text{where} \quad \mathscr{R} = \begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix}.$$

Note that $\mathscr{R}$ is a Bogoliubov rotation since $\overline{R} = R$. The density matrices of an invariant state satisfy

$$g(Rx,Ry) = g(x,y), \qquad a(Rx,Ry) = a(x,y)$$

for all $x,y \in \mathbb{R}^3$ and any rotation $R \in SO(3)$.

As the angular momentum $L = x \times (-i\nabla)$ generates the group of rotations, a (smooth enough) HFB state is invariant under rotations if and only if

$$\mathscr{L}\Gamma = \Gamma\mathscr{L}, \qquad \text{where} \quad \mathscr{L} = \begin{pmatrix} L & 0 \\ 0 & L \end{pmatrix}.$$

The density matrices g and a can then be written in the special form

$$g(x,y) = \frac{1}{4\pi}\sum_{\ell\geqslant 0} g^\ell(|x|,|y|)\,(2\ell{+}1)\,P_\ell\big(\omega_x{\cdot}\omega_y\big), \;\; a(x,y) = \frac{1}{4\pi}\sum_{\ell\geqslant 0} a^\ell(|x|,|y|)\,(2\ell{+}1)\,P_\ell\big(\omega_x{\cdot}\omega_y\big) \tag{3.25}$$

where P_ℓ is the Legendre polynomial of degree ℓ, which is such that $P_\ell(1) = 1$. The constraint on g and a are transfered in each angular momentum sector (labelled by $\ell \geqslant 0$), leading to

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \leqslant \begin{pmatrix} g^\ell & a^\ell \\ a^\ell & 1-g^\ell \end{pmatrix} \leqslant \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

on $L^2([0,\infty), r^2\,dr) \oplus L^2([0,\infty), r^2\,dr)$. However, there is no such constraint between different ℓ's. The total average particle number is given by

$$\mathrm{Tr}(g) = \sum_{\ell\geqslant 0}(2\ell+1)\,\mathrm{Tr}(g^\ell) = N/2.$$

This is the only constraint which mixes the different angular momentum density matrices.

Now we can discretize the radial HFB problem. We choose a finite-dimensional subspace $V_{\rm rad}$ in $L^2([0,\infty), r^2 dr)$ with basis $(\chi_1, ..., \chi_{N_b})$, which we use to expand the density matrices g^ℓ and a^ℓ. Then we fix a maximal angular momentum $\ell_{\max}$ and we truncate the series in (3.25). This is the same as taking as discretization space

$$V = \Big\{ f(|x|)\, Y_\ell^m(\omega_x) \ : \ f \in V_{\rm rad},\ 0 \leqslant \ell \leqslant \ell_{\max},\ -\ell \leqslant m \leqslant \ell \Big\} \subset L^2(\mathbb{R}^3, \mathbb{R})$$

where Y_ℓ^m is the spherical harmonics of total angular momentum ℓ and azimuthal angular momentum m. We then assume that g and a are radial and live in this space. The matrices G^ℓ and A^ℓ of g^ℓ and a^ℓ in the basis (χ_i) are defined similarly as before by

$$g^\ell(r,r') = \sum_{i,j=1}^{N_b} G_{ij}^\ell\, \chi_i(r)\chi_j(r'), \qquad a^\ell(r,r') = \sum_{i,j=1}^{N_b} A_{ij}^\ell\, \chi_i(r)\chi_j(r'). \tag{3.26}$$

The constraints on the matrices G^ℓ and A^ℓ are

$$0 \leqslant \Upsilon^\ell \mathbf{S} \Upsilon^\ell \leqslant \Upsilon^\ell, \quad \text{with} \quad \Upsilon^\ell := \begin{pmatrix} G^\ell & A^\ell \\ A^\ell & S^{-1} - G^\ell \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix} \tag{3.27}$$

with

$$S_{ij} = \int_0^\infty \chi_i(r)\, \chi_j(r)\, r^2\, dr$$

and

$$\sum_{\ell=0}^{\ell_{\max}} (2\ell+1)\, {\rm Tr}(S G^\ell) = N/2. \tag{3.28}$$

The total HFB energy is now

$$\begin{aligned} \mathcal{E}(G^0, ..., G^{\ell_{\max}}, A^0, ..., A^{\ell_{\max}}) &= 2 \sum_{\ell=0}^{\ell_{\max}} (2\ell+1)\, {\rm Tr}(h^\ell G^\ell) \\ &+ \sum_{\ell,\ell'=0}^{\ell_{\max}} (2\ell+1)(2\ell'+1) \Big(2\, {\rm Tr}(G^\ell\, J(G^{\ell'})) - {\rm Tr}(G^\ell\, K^{\ell\ell'}(G^{\ell'})) + {\rm Tr}(A^\ell\, K^{\ell\ell'}(A^{\ell'})) \Big), \end{aligned} \tag{3.29}$$

where

$$h_{ij}^\ell := \int_0^\infty \chi_i'(r)\, \chi_j'(r)\, r^2\, dr + \ell(\ell+1) \int_0^\infty \chi_i(r)\, \chi_j(r)\, dr,$$

$$J(G^{\ell'})_{ij} := \sum_{m,n=0}^{N_b} (ij|nm)_{0,0}\, G_{mn}^{\ell'}, \qquad K^{\ell\ell'}(G^{\ell'})_{ij} := \sum_{m,n=0}^{N_b} (im|jn)_{\ell,\ell'}\, G_{mn}^{\ell'},$$

$$(ij|mn)_{\ell,\ell'} := \int_0^\infty r^2\, dr \int_0^\infty s^2\, ds\ \chi_i(r)\, \chi_j(r)\, \chi_m(s)\, \chi_n(s)\, w_{\ell,\ell'}(r,s)$$

and

$$w_{\ell,\ell'}(r,s) := \frac{1}{2} \int_{-1}^1 W\left(\sqrt{r^2 + s^2 - 2rst} \right)\, P_\ell(t)\, P_{\ell'}(t)\, dt.$$

Any minimizer $(G^0, ..., G^{\ell_{\max}}, A^0, ..., A^{\ell_{\max}})$ of $\mathcal{E}$ under the constraints (3.27) and (3.28) will be of the form

$$\Upsilon^\ell = \sum_{\epsilon_i^\ell < 0} f_i^\ell \big(f_i^\ell\big)^T, \qquad 0 \leqslant \ell \leqslant \ell_{\max}, \tag{3.30}$$

where the vectors f_i^ℓ's solve the generalized eigenvalue problem

$$\Big(\mathbf{F}^\ell - \mu(2\ell+1)\mathbf{N}\Big) f_i^\ell = \epsilon_i^\ell \, \mathbf{S} \, f_i^\ell, \qquad \langle f_i, \mathbf{S} f_j \rangle = \delta_{ij} \tag{3.31}$$

with

$$\mathbf{F}^\ell := \begin{pmatrix} h^\ell & 0 \\ 0 & -h^\ell \end{pmatrix} + \sum_{\ell'=0}^{\ell_{\max}} \begin{pmatrix} 2J\big(G^{\ell'}\big) - K^{\ell\ell'}\big(G^{\ell'}\big) & K^{\ell\ell'}\big(A^{\ell'}\big) \\ K^{\ell\ell'}\big(A^{\ell'}\big) & -2J\big(G^{\ell'}\big) + K^{\ell\ell'}\big(G^{\ell'}\big) \end{pmatrix}.$$

The Euler-Lagrange multipler μ appearing in (3.31) is common to all the different angular momentum sectors and it is chosen to ensure the validity of the constraint (3.28).

4. Algorithmic strategies and convergence analysis

In this section we study the convergence of two algorithms which can be used in practice to solve the HFB minimization problem (2.15). In order to simplify our presentation, we restrict ourselves to the finite-dimensional case, that is, to the discretized problem (3.16). We also assume that the discretization basis (χ_j) is orthonormal, such that $\mathbf{S} = I_{2N_b}$, the $(2N_b) \times (2N_b)$ identity matrix. Finally, we only consider states which are invariant under time-reversal symmetry like in Section 3.3.1. This means that the HFB state is described by real and symmetric matrices G and A such that

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \leqslant \Upsilon := \begin{pmatrix} G & A \\ A & 1-G \end{pmatrix} \leqslant \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{4.1}$$

The energy is given by (3.24),

$$\mathcal{E}(\Upsilon) = 2\,\mathrm{Tr}(hG) + 2\,\mathrm{Tr}(G\,J(G)) - \mathrm{Tr}(G\,K(G)) + \mathrm{Tr}(A\,K(A)). \tag{4.2}$$

The extension to more general situations is straightforward.

The energy $\mathcal{E}$ is continuous (it is indeed real-analytic) with respect to Υ. Also the set $\mathcal{K}$ of density matrices Υ of the form (4.1) is compact in finite dimension. Hence minimizers always exist and, as we have seen, they solve the nonlinear equation

$$\Upsilon = \mathbb{1}_{(-\infty,0)}\Big(\mathbf{F}_\Upsilon - \mu\mathbf{N}\Big) + \delta, \tag{4.3}$$

where μ is a Lagrange multiplier chosen to ensure the constraint that $\mathrm{Tr}(G) = N/2$. Of course we must have $N/2 \leqslant N_b$, the dimension of the (no-spin) discretization space V_h, otherwise the minimization set is always empty.

4.1. *Roothaan Algorithm*

The most natural technique used in practice to solve the equation (4.3) is a simple fixed point method [40, 16]. This is usually refered to as the *Roothaan algorithm* in the chemistry literature [42]. The iteration scheme is the following

$$\Upsilon_{n+1} = \mathbb{1}_{(-\infty,0)}\Big(\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}\Big) + \delta_{n+1}. \tag{4.4}$$

At each step, one has to determine μ_{n+1} such as to satisfy the constraint $\mathrm{Tr}(G_{n+1}) = N/2$. If the operator $\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}$ has a trivial kernel, then $\delta_{n+1} \equiv 0$. This is the usual situation encountered in practice. In the iteration (4.4), the state is assumed to be pure at each step (Υ is an orthogonal projection). Recall that by Theorem 2.2 we know that minimizers of $\mathcal{E}$ under a particle number constraint are always pure, under suitable assumptions on the interaction potential W. The algorithm is stopped when the commutator

$$\left\| \left[\Upsilon_n, F_{\Upsilon_n} \right] \right\|$$

and/or when the variation of the HFB state

$$\| \Upsilon_{n+1} - \Upsilon_n \|$$

are smaller than a prescribed ε.

Our purpose in this section is to study the behavior of the Roothaan algorithm (4.4). In the Hartree-Fock case, it was shown in a fundamental work of Cancès and Le Bris [12, 13], that the algorithm converges or oscillate between two points, none of them being a solution to the equation (4.3). This result was recently improved by Levitt in [31]. We will explain that the results of Cancès-Le Bris and Levitt can be generalized to the HFB model. Actually, in a discretization space of dimension N_b, HFB is equivalent to a Hartree-Fock-like minimization problem in dimension $2N_b$, with additional constraints. The adaptation of the previously cited works in the HF case reduces to handling these contraints.

In order to avoid the convergence problems of the Roothaan algorithm, Cancès and Le Bris have proposed the *Optimal Damping Algorithm* (ODA). We will study the equivalent of this algorithm in HFB theory in Section 4.2.

To start with, we show that the Roothaan algorithm is well defined, in the sense that for any HFB state Υ_n, there exists $(\Upsilon_{n+1}, \mu_{n+1}, \delta_{n+1})$ solving (4.4). To this end, we follow [12, 13] and introduce the auxillary functional

$$\tilde{\mathcal{E}}(\Upsilon, \Upsilon') := \mathrm{Tr}(hG) + \mathrm{Tr}(hG') + 2\,\mathrm{Tr}(G\,J(G')) - \mathrm{Tr}(G\,K(G')) + \mathrm{Tr}(A\,K(A')) \tag{4.5}$$

as well as the variational problem

$$I_\Upsilon(\lambda) := \min_{\Upsilon'} \left\{ \tilde{\mathcal{E}}(\Upsilon, \Upsilon') \ : \ \mathrm{Tr}(G') = \lambda \right\} \tag{4.6}$$

which consists in minimizing over Υ' with Υ fixed. The matrix Υ' must be an admissible HFB state which, in our context, means that

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \leqslant \Upsilon' := \begin{pmatrix} G' & A' \\ A' & 1 - G' \end{pmatrix} \leqslant \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (G')^T = \overline{G'} = G', \quad (A')^T = \overline{A'} = A'. \tag{4.7}$$

Recall that we have chosen an orthonormal basis and that the spin has been eliminated. It is clear that (4.6) admits at least one solution Υ', as soon as $0 \leqslant \lambda \leqslant N_b$, where we recall that N_b is the dimension of the discretization space V_h. The following says that these solutions are exactly those solving the equation of the Roothaan method.

Lemma 4.1 (The Roothaan algorithm is well defined). *The function $\lambda \in [0, N_b] \mapsto I_\Upsilon(\lambda)$ is convex, hence left and right differentiable. For any $\lambda \in [0, N_b]$, the minimizers Υ' of $I_\Upsilon(\lambda)$ are exactly the states of the form*

$$\Upsilon' = \mathbb{1}_{(-\infty, 0)}\big(\mathbf{F}_\Upsilon - \mu'\mathbf{N}\big) + \delta' \tag{4.8}$$

where $\mu' \in [\partial_- I_\Upsilon(\lambda), \partial_+ I_\Upsilon(\lambda)]$ *and* $0 \leqslant \delta' \leqslant \mathbb{1}_{\{0\}}(\mathbf{F}_\Upsilon - \mu'\mathbf{N})$*. If* $0 \notin \sigma(\mathbf{F}_\Upsilon - \mu'\mathbf{N})$*, then* $\delta' \equiv 0$ *and* Υ' *is unique, for any such* $\mu' \in [\partial_- I_\Upsilon(\lambda), \partial_+ I_\Upsilon(\lambda)]$*.*

Proof. To see that $I_\Upsilon(\lambda)$ is convex, let $0 \leqslant \lambda_1 < \lambda_2 \leqslant N_b$ and let Υ'_i be a minimizer for $I_\Upsilon(\lambda_i)$ with $i = 1, 2$. Then $t\Upsilon'_1 + (1-t)\Upsilon'_2$ is a test state for the problem $I_\Upsilon(t\lambda_1 + (1-t)\lambda_2)$. Therefore it holds

$$\begin{aligned} I_\Upsilon(t\lambda_1 + (1-t)\lambda_2) \leqslant \tilde{\mathcal{E}}(\Upsilon, t\Upsilon'_1 + (1-t)\Upsilon'_2) &= t\tilde{\mathcal{E}}(\Upsilon, \Upsilon'_1) + (1-t)\tilde{\mathcal{E}}(\Upsilon, \Upsilon'_2) \\ &= tI_\Upsilon(\lambda_1) + (1-t)I_\Upsilon(\lambda_2). \end{aligned}$$

Then, by convexity we get that $I_\Upsilon(\lambda') \geqslant I_\Upsilon(\lambda) + \mu(\lambda' - \lambda)$ for any $\lambda' \in [0, N_b]$ and any $\mu \in [\partial_- I_\Upsilon(\lambda), \partial_+ I_\Upsilon(\lambda)]$. Thus

$$\begin{aligned} I_\Upsilon(\lambda) - \mu\lambda &= \min\{I_\Upsilon(\lambda') - \mu\lambda' \ : \ 0 \leqslant \lambda' \leqslant N_b\} \\ &= \min_{\Upsilon'} \left\{\tilde{\mathcal{E}}(\Upsilon, \Upsilon') - \mu \operatorname{Tr}(G')\right\} \\ &= \operatorname{Tr}(hG) + \frac{1}{2} \min_{\Upsilon'} \operatorname{Tr}(\mathbf{F}_\Upsilon - \mu\mathbf{N})\Upsilon'. \end{aligned}$$

In the previous two mins, Υ' is varied over all possible HFB states, without any particle number constraint. It is well known that the minimizers of the problem on the right side are exactly the solutions of the equation (4.8). □

Lemma 4.1 tells us that for any given Υ_n, there always exists at least one solution $(\Upsilon_{n+1}, \mu_{n+1}, \delta_{n+1})$ of the equation (4.4). It is obtained by solving the minimization problem $I_{\Upsilon_n}(N/2)$, and one has to take $\mu_{n+1} \in [\partial_- I_{\Upsilon_n}(N/2), \partial_+ I_{\Upsilon_n}(N/2)]$. We can always take by convention

$$\mu_{n+1} := \frac{\partial_- I_{\Upsilon_n}(N/2) + \partial_+ I_{\Upsilon_n}(N/2)}{2}.$$

However, Υ_{n+1} is not uniquely defined yet because of the possibility of having $\delta_{n+1} \neq 0$. As we have seen it is unique when

$$0 \notin \sigma\big(\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}\big). \tag{4.9}$$

In this section we always assume that it is possible to find $(\Upsilon_{n+1}, \mu_{n+1}, \delta_{n+1})$ exactly. In practice, we will only know $(\Upsilon_{n+1}, \mu_{n+1}, \delta_{n+1})$ approximately. Later in Section 4.3 we explain how to do this numerically. We will also see that the condition (4.9) is "very often" satisfied. This vague statement is made precise in Lemma 4.4 below. Following Cancès and Le Bris, we now introduce the concept of uniform well-posedness.

Definition 4.1 (Uniform well-posedness). We say that for a given initial HFB state Υ_0, the sequence (Υ_n) generated by the Roothaan algorithm is *uniformly well posed* when there exists $\eta > 0$ such that

$$|\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}| \geqslant \eta \tag{4.10}$$

for all $n \geqslant 0$, where $\mu_{n+1} = \big(\partial_- I_{\Upsilon_n}(N/2) + \partial_+ I_{\Upsilon_n}(N/2)\big)/2$.

Note that the condition $|\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}| \geqslant \eta$ is equivalent to $(-\eta, \eta) \cap \sigma(\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}) = \emptyset$. Later in Section 4.3 we will make several comments concerning Assumption (4.10).

We have seen that the sequence generated by the Roothaan algorithm can be obtained by solving the minimization problem

$$I_{\Upsilon_n} = \min_{\Upsilon'} \tilde{\mathcal{E}}(\Upsilon_n, \Upsilon').$$

Since $\tilde{\mathcal{E}}(\Upsilon, \Upsilon') = \tilde{\mathcal{E}}(\Upsilon', \Upsilon)$, we conclude that the Roothaan algorithm is the same as minimizing $\tilde{\mathcal{E}}$ with respect to its first and second variables one after another, inductively. This fact allows to prove the following result, which is the HFB equivalent of Theorem 7 in [13] and Theorem 5.1 in [31] in the HF case.

Theorem 4.1 (Convergence of the Roothaan algorithm). *Assume that* $0 < N/2 < N_b$. *Let* Υ_0 *be an initial HFB state such that the sequence* (Υ_n) *generated by the Roothaan algorithm is* uniformly well posed. *Then*

- *The sequence* $\tilde{\mathcal{E}}(\Upsilon_{2n}, \Upsilon_{2n+1})$ *decreases towards a critical value of* $\tilde{\mathcal{E}}$*;*
- *The sequence* $(\Upsilon_{2n}, \Upsilon_{2n+1})$ *converges towards a critical point* (Υ, Υ') *of* $\tilde{\mathcal{E}}$*;*
- *If* $\Upsilon = \Upsilon'$*, then this state is a solution of the original HFB equation* (4.3)*, but if* $\Upsilon \neq \Upsilon'$*, then none of these two states is a solution to* (4.3).

Theorem 4.1 says that (provided it is uniformly well posed) the sequence Υ_n will either converge to a solution of the self-consistent Equation (4.3), or oscillate between two points Υ and Υ', none of them being a solution to the desired equation.

Proof. We split the proof into several steps.

Step 1: μ_n ***is uniformly bounded.*** It will be very useful to know that the sequence μ_n is uniformly bounded. The following says that, as soon as we fix $\mathrm{Tr}(G) = N/2$ with $0 < N/2 < N_b$, the chemical potential μ cannot be too negative and too positive.

Lemma 4.2 (Bounds on the multiplier μ). *Let* Υ' *be any fixed HFB state and*

$$\Upsilon_\mu := \mathbb{1}_{(-\infty,0)}\big(\mathbf{F}_{\Upsilon'} - \mu\mathbf{N}\big) = \begin{pmatrix} G_\mu & A_\mu \\ A_\mu & 1 - G_\mu \end{pmatrix}$$

the corresponding HFB ground state at chemical potential μ. *There exists a constant* C *which is independent of* Υ' *and* μ*, such that*

$$\forall \mu \leqslant -C, \quad \mathrm{Tr}\, G_\mu \leqslant \frac{C}{|\mu|} \qquad \text{and} \qquad \forall \mu \geqslant C, \quad \mathrm{Tr}\, G_\mu \geqslant N_b - \frac{C}{\mu}. \tag{4.11}$$

The lemma says that the average number of particles in $\mathbb{1}_{(-\infty,0)}\big(\mathbf{F}_\Upsilon - \mu\mathbf{N}\big)$ tends to N_b when $\mu \to \infty$ whereas it tends to 0 when $\mu \to -\infty$, this *uniformly with respect to the state* Υ' used to build the mean-field operator $\mathbf{F}_{\Upsilon'}$.

Proof. We first remark that there exists a constant C such that

$$\|F_{\Upsilon'}\| \leqslant C \tag{4.12}$$

for any HFB state Υ'. This follows from the fact that $F_{\Upsilon'}$ is continuous with respect to Υ' and that the latter lives in a compact set since we always have $0 \leqslant \Upsilon \leqslant 1$. The chosen norm

for $\|F_{\Upsilon'}\|$ does not matter since we are in finite dimension. Now, for μ large enough we can use regular perturbation theory and obtain that

$$\begin{aligned}\big\|\mathbb{1}_{(-\infty,0)}\big(F_{\Upsilon'}-\mu\mathbf{N}\big)-\mathbb{1}_{(-\infty,0)}\big(-\mu\mathbf{N}\big)\big\| \\ = \left\|\mathbb{1}_{(-\infty,0)}\left(\frac{F_{\Upsilon'}}{|\mu|}-\frac{\mu}{|\mu|}\,\mathbf{N}\right)-\mathbb{1}_{(-\infty,0)}\left(-\frac{\mu}{|\mu|}\,\mathbf{N}\right)\right\| \leqslant \frac{C}{|\mu|}.\end{aligned}$$

Note that

$$\mathbb{1}_{(-\infty,0)}\left(-\frac{\mu}{|\mu|}\,\mathbf{N}\right)=\begin{cases}\begin{pmatrix}1&0\\0&0\end{pmatrix} & \text{for } \mu>0,\\ \begin{pmatrix}0&0\\0&1\end{pmatrix} & \text{for } \mu<0.\end{cases}$$

Taking the trace against $\mathbf{N}$ gives the result. □

From Lemma 4.2 we deduce that our sequence μ_n is bounded. Indeed, since we have by construction $\mathrm{Tr}(G_{n+1}) = N/2$ with $0 < N/2 < N_b$, we must have

$$-\max\left(C\,,\,\frac{2C}{N}\right)\leqslant\mu_{n+1}\leqslant\max\left(C\,,\,\frac{C}{N_b-N/2}\right) \tag{4.13}$$

otherwise $\mathrm{Tr}(G_{n+1})$ would be too small or too large.

Step 2: *convergence of* $\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon_{n+1})$. We now follow [12, 13, 10]. At each step, we know from Lemma 4.1 that Υ_{n+1} is a solution of the minimization problem $\min_{\Upsilon'}\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon')$. In particular, we deduce that

$$\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon_{n+1})\leqslant\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon_{n-1})=\tilde{\mathcal{E}}(\Upsilon_{n-1},\Upsilon_n). \tag{4.14}$$

Thus the sequence of real numbers $\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon_{n+1})$ is non-increasing. It is also bounded from below, hence it converges to a limit ℓ. We now use the uniform well-posedness to prove an inequality which is more precise than (4.14). We remark that

$$\begin{aligned}\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon_{n-1})-\tilde{\mathcal{E}}(\Upsilon_n,\Upsilon_{n+1}) &= \frac{1}{2}\,\mathrm{Tr}\,\mathbf{F}_{\Upsilon_n}\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)\\ &= \frac{1}{2}\,\mathrm{Tr}\,\big(\mathbf{F}_{\Upsilon_n}-\mu_{n+1}\mathbf{N}\big)\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)\\ &\geqslant \frac{1}{2}\,\mathrm{Tr}\,\big|\mathbf{F}_{\Upsilon_n}-\mu_{n+1}\mathbf{N}\big|\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)^2\\ &\geqslant \frac{\eta}{2}\,\mathrm{Tr}\,\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)^2=\frac{\eta}{2}\,\|\Upsilon_{n-1}-\Upsilon_{n+1}\|^2.\end{aligned} \tag{4.15}$$

In the above calculation we have used that $\mathrm{Tr}\,\mathbf{N}\,\Upsilon_{n+1}=\mathrm{Tr}\,\mathbf{N}\,\Upsilon_{n-1}=N-N_b$. We have also used that Υ_{n+1} is the negative spectral projector of $\mathbf{F}_{\Upsilon_n}-\mu_{n+1}\mathbf{N}$, such that we can write

$$\big(\mathbf{F}_{\Upsilon_n}-\mu_{n+1}\mathbf{N}\big)=\big|\mathbf{F}_{\Upsilon_n}-\mu_{n+1}\mathbf{N}\big|\big(\Upsilon_{n+1}^{\perp}-\Upsilon_{n+1}\big).$$

Finally, we have used that $0\leqslant\gamma\leqslant 1$ is equivalent to $(\gamma-P)^2\leqslant P^{\perp}(\gamma-P)P^{\perp}-P(\gamma-P)P$, for any orthogonal projector P, thus

$$\Upsilon_{n+1}^{\perp}\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)\Upsilon_{n+1}^{\perp}-\Upsilon_{n+1}\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)\Upsilon_{n+1}\geqslant\big(\Upsilon_{n-1}-\Upsilon_{n+1}\big)^2.$$

Since $\tilde{\mathcal{E}}(\Upsilon_n, \Upsilon_{n+1})$ converges to a limit ℓ, we deduce that

$$\sum_{n\geqslant 1} \|\Upsilon_{n+1} - \Upsilon_{n-1}\|^2 < \infty.$$

In particular, $\|\Upsilon_{n+1} - \Upsilon_{n-1}\| \to 0$ which is called *numerical convergence* in [12, 13].

Step 3: *convergence of* Υ_{2n} *and* Υ_{2n+1}. In order to upgrade the numerical convergence to true convergence, we use a recent method of Levitt [31]. Namely we show that

$$\Big(\tilde{\mathcal{E}}(\Upsilon_n, \Upsilon_{n-1}) - \ell\Big)^\theta - \Big(\tilde{\mathcal{E}}(\Upsilon_n, \Upsilon_{n+1}) - \ell\Big)^\theta \geqslant \frac{\eta'}{2} \|\Upsilon_{n-1} - \Upsilon_{n+1}\| \tag{4.16}$$

for a well chosen $0 < \theta \leqslant 1/2$. Summing over n and using the convergence of $\tilde{\mathcal{E}}(\Upsilon_n, \Upsilon_{n-1})$, hence of $(\tilde{\mathcal{E}}(\Upsilon_n, \Upsilon_{n-1}) - \ell)^\theta$, then gives the convergence of Υ_{2n} and Υ_{2n+1}.

For the proof of (4.16), we argue as follow. Consider a (real, no-spin) pure HFB state Υ. It is possible to parametrize the manifold of pure HFB states around Υ by using Bogoliubov transformations as follows:

$$H \mapsto e^H \Upsilon e^{-H}$$

where H is assumed to be of the form

$$\begin{pmatrix} h & p \\ -p & -h \end{pmatrix}, \qquad h^T = -\overline{h} = -h, \quad p^T = \overline{p} = p.$$

These constraints ensure that iH is a self-adjoint Hamiltonian such that $e^H = e^{-i(iH)}$ is a Bogoliubov rotation. They also ensure that $e^H \Upsilon e^{-H}$ stays real. That $H \mapsto e^H \Upsilon e^{-H}$ is a local chart of the manifold of pure HFB states around Υ follows from the arguments in [3] as well as simple considerations in linear algebra.

Let us now consider the energy $\tilde{\mathcal{E}}$ in a neighborhood of any fixed (Υ, Υ'). The map

$$f : (H, H') \mapsto \tilde{\mathcal{E}}\big(e^H \Upsilon e^{-H}, e^{H'} \Upsilon' e^{-H'}\big) - \frac{\mu'}{2} \operatorname{Tr} \mathbf{N} e^H \Upsilon e^{-H} - \frac{\mu}{2} \operatorname{Tr} \mathbf{N} e^{H'} \Upsilon' e^{-H'}$$

is real analytic in a neighborhood of $(0, 0)$ for any fixed $\mu, \mu' \in \mathbb{R}$ and any fixed pure HFB states (Υ, Υ'). The Łojasiewicz inequality (Theorem 2.1 in [31]) then tells us that there exist $0 < \theta \leqslant 1/2$ and a constant $\kappa > 0$ such that $\|H\| + \|H'\| \leqslant \kappa$ implies

$$|f(H, H') - f(0)|^{1-\theta} \leqslant \kappa^{-1} \Big(|\nabla_H f(H, H')| + |\nabla_{H'} f(H, H')| \Big).$$

A simple computation shows that

$$\nabla_H f(H, H') = \frac{1}{2} \big[\mathbf{F}_{\Upsilon'} - \mu' \mathbf{N}\,,\, e^H \Upsilon e^{-H} \big], \qquad \nabla_{H'} f(H, H') = \frac{1}{2} \big[\mathbf{F}_{\Upsilon} - \mu \mathbf{N}\,,\, e^{H'} \Upsilon' e^{-H'} \big].$$

If we rephrase all this in our setting, this means that for any fixed pure HFB states $(\Upsilon_1, \Upsilon_1')$ and any $\mu, \mu' \in \mathbb{R}$, there is a $\kappa > 0$ such that for any $(\Upsilon_2, \Upsilon_2')$ another pure HFB state which is at most at a distance κ from $(\Upsilon_1, \Upsilon_1')$, we have

$$\begin{aligned} \Big| \tilde{\mathcal{E}}\big(\Upsilon_1, \Upsilon_1'\big) - \tilde{\mathcal{E}}\big(\Upsilon_2, \Upsilon_2'\big) + \mu' \operatorname{Tr} \mathbf{N}(G_2 - G_1) + \mu \operatorname{Tr} \mathbf{N}(G_2' - G_1') \Big|^{1-\theta} \\ \leqslant \kappa^{-1} \Big(\big\| \big[\mathbf{F}_{\Upsilon_2'} - \mu' \mathbf{N}\,,\, \Upsilon_2 \big] \big\| + \big\| \big[\mathbf{F}_{\Upsilon_2} - \mu \mathbf{N}\,,\, \Upsilon_2' \big] \big\| \Big). \end{aligned} \tag{4.17}$$

The constants κ and θ depend on μ, μ' and of the reference point $(\Upsilon_1, \Upsilon_1')$. But they stay positive as soon as μ, μ' and $(\Upsilon_1, \Upsilon_1')$ stay in a compact set. By a simple compactness argument, we therefore deduce that there exists a neighborhood of the compact set

$$\Big\{(\Upsilon, \Upsilon') \ : \ \tilde{\mathcal{E}}(\Upsilon, \Upsilon') = \ell, \ \mathrm{Tr}(G) = \mathrm{Tr}(G') = N/2\Big\}$$

such that for any (Υ, Υ') in this neighborhood and μ, μ' in a compact set in $\mathbb{R}$, we have

$$\begin{aligned}\Big|\tilde{\mathcal{E}}\big(\Upsilon, \Upsilon'\big) - \ell + \mu'\big(N/2 - \mathrm{Tr}\,\mathbf{N}G\big) + \mu\big(N/2 - \mathrm{Tr}\,\mathbf{N}G'\big)\Big|^{1-\theta} \\ \leqslant \kappa^{-1}\Big(\big\|\big[\mathbf{F}_{\Upsilon'} - \mu'\mathbf{N}\,,\,\Upsilon\big]\big\| + \big\|\big[\mathbf{F}_{\Upsilon} - \mu\mathbf{N}\,,\,\Upsilon'\big]\big\| \Big) \quad (4.18)\end{aligned}$$

for some $0 < \theta \leqslant 1/2$ and some $\kappa > 0$. We recall that ℓ is by definition the limit of $\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n+1}\big)$.

Recall our inequality (4.13) which says that μ_n is uniformly bounded. Also, we know that $\tilde{\mathcal{E}}(\Upsilon_n, \Upsilon_{n+1})$ converges to ℓ so, for n large enough, $(\Upsilon_n, \Upsilon_{n+1})$ must be in the neighborhood of the level set ℓ. Choosing $\mu = \mu_{n+1}$ and $\mu' = \mu_n$ and using that G_n and G_{n+1} have the correct trace, we get the estimate

$$\begin{aligned}\Big(\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n-1}\big) - \ell\Big)^{1-\theta} &\leqslant \kappa^{-1}\left(\big\|\big[\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}\,,\,\Upsilon_{n-1}\big]\big\| + \big\|\big[\mathbf{F}_{\Upsilon_{n-1}} - \mu_n\mathbf{N}\,,\,\Upsilon_n\big]\big\|\right) \\ &= \kappa^{-1}\big\|\big[\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}\,,\,\Upsilon_{n-1} - \Upsilon_{n+1}\big]\big\| \\ &\leqslant C\,\|\Upsilon_{n-1} - \Upsilon_{n+1}\|\end{aligned}$$

for n large enough. Here we have used that Υ_n commutes with $\mathbf{F}_{\Upsilon_{n-1}} - \mu_n\mathbf{N}$ and that Υ_{n+1} commutes with $\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}$ by construction, and that $\|\mathbf{F}_{\Upsilon_n}\|$ and μ_{n+1} are both uniformly bounded. In order to conclude, we use the concavity of $x \mapsto x^\theta$ and (4.15) like in [31] to obtain

$$\begin{aligned}\Big(\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n-1}\big) - \ell\Big)^{\theta} - \Big(\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n+1}\big) - \ell\Big)^{\theta} \\ &\geqslant \frac{\theta}{\Big(\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n-1}\big) - \ell\Big)^{1-\theta}}\Big(\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n-1}\big) - \tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n+1}\big)\Big) \\ &\geqslant \frac{\eta\,\theta}{2\Big(\tilde{\mathcal{E}}\big(\Upsilon_n, \Upsilon_{n-1}\big) - \ell\Big)^{1-\theta}}\,\|\Upsilon_{n+1} - \Upsilon_{n-1}\|^2 \\ &\geqslant \eta\theta/(2C)\ \|\Upsilon_{n+1} - \Upsilon_{n-1}\|\end{aligned}$$

by (4.15). This concludes the proof of the inequality (4.16), hence the proof of the convergence of $(\Upsilon_{2n}, \Upsilon_{2n+1})$, towards some pure HFB states (Υ, Υ').

Step 4: *the limit* (Υ, Υ') *of* $(\Upsilon_{2n}, \Upsilon_{2n+1})$ *is a critical point of* $\tilde{\mathcal{E}}$. Since we have $\Upsilon_{2n} \to \Upsilon$ and $\Upsilon_{2n+1} \to \Upsilon'$, we deduce that $\mathbf{F}_{\Upsilon_{2n}} \to \mathbf{F}_\Upsilon$ and $\mathbf{F}_{\Upsilon_{2n+1}} \to \mathbf{F}_{\Upsilon'}$, by continuity of the map $\Upsilon \mapsto \mathbf{F}_\Upsilon$. Extracting a subsequence, we can assume that $\mu_{2n_k} \to \mu'$ and $\mu_{2n_k+1} \to \mu$. We have

$$\Upsilon_{2n_k} = \mathbb{1}_{(-\infty,0)}(\mathbf{F}_{\Upsilon_{2n_k-1}} - \mu_{2n_k}\mathbf{N}), \qquad \Upsilon_{2n_k+1} = \mathbb{1}_{(-\infty,0)}(\mathbf{F}_{\Upsilon_{2n_k}} - \mu_{2n_k+1}\mathbf{N})$$

and, by uniform well-posedness,

$$\big|\mathbf{F}_{\Upsilon_{2n_k-1}} - \mu_{2n_k}\mathbf{N}\big| \geqslant \eta, \qquad \big|\mathbf{F}_{\Upsilon_{2n_k}} - \mu_{2n_k+1}\mathbf{N}\big| \geqslant \eta.$$

Passing to the limit $k \to \infty$ we get

$$\Upsilon = \mathbb{1}_{(-\infty,0)}(\mathbf{F}_{\Upsilon'} - \mu'\mathbf{N}) \quad \text{and} \quad \Upsilon' = \mathbb{1}_{(-\infty,0)}(\mathbf{F}_{\Upsilon} - \mu\mathbf{N}).$$

This exactly means that (Υ, Υ') is a critical point of $\tilde{\mathcal{E}}$ on $\mathcal{P}_h(N/2) \times \mathcal{P}_h(N/2)$. Note that we have also $\left|\mathbf{F}_{\Upsilon'} - \mu'\mathbf{N}\right| \geqslant \eta$ and $\left|\mathbf{F}_{\Upsilon} - \mu\mathbf{N}\right| \geqslant \eta$.

The remaining statements are verified exactly like in the HF case. This concludes the proof of Theorem 4.1. □

4.2. *Optimal Damping Algorithm*

In the previous section we have studied the convergence properties of the Roothaan algorithm, which consists in solving the self-consistent equation by a fixed point method. We have seen that the algorithm can either converge or oscillate between two states, none of them being a solution to the problem.

Examples of such oscillations in quantum chemistry have been exhibited by Cancès and Le Bris [12, 13]. In this case the potential W is repulsive and there is no pairing. In order to cure this problem of oscillations, Cancès and Le Bris proposed in [13] a *relaxed algorithm* called the *Optimal Damping Algorithm* (ODA). This method makes use of the important fact that one can minimize over mixed states and get the same ground state as when minimizing over pure states only (Theorem 2.1).

The same oscillations can *a priori* happen in HFB with an attractive potential W. They are frequently seen with the Roothaan algorithm and we will give several numerical examples later in Section 5. Even when the sequence Υ_n eventually converges towards a single state Υ, these oscillations can slow down the convergence considerably. This phenomenon is well known in nuclear physics. Dechargé and Gogny already advocate in [16] the use of a *damping parameter* between two successive iterations, in order to *"slow down the convergence on the density matrix. In this way the average field varies slowly and we can insure the convergence on the pairing tensor step by step"* (see [16] page 1574). Even in the modern computations, this damping parameter is fixed all along the algorithm (Nathalie Pillet, private communication).

We suggest to transpose the method of Cancès and Le Bris to the HFB setting by using an optimal damping parameter, chosen such as to minimize the energy. This means resorting to mixed states even if the final ground state is always a pure HFB state. This is theoretically justified when the assumptions of the Bach-Fröhlich-Jonsson Theorem 2.2 are fulfilled.

The ODA involves two density matrices Υ_n and $\tilde{\Upsilon}_n$. The HFB state Υ_n is always pure but $\tilde{\Upsilon}_n$ can (and will usually) be a mixed HFB state. The starting point $\Upsilon_0 = \tilde{\Upsilon}_0$ being chosen, the sequence is then constructed by induction as follows:

(1) One finds $(\Upsilon_{n+1}, \mu_{n+1})$ solving

$$\Upsilon_{n+1} = \mathbb{1}_{(-\infty,0)}\big(\mathbf{F}_{\tilde{\Upsilon}_n} - \mu_{n+1}\mathbf{N}\big) \quad \text{and} \quad \mathrm{Tr}(G_{n+1}) = N/2.$$

This is always possible, by Lemma 4.1 and we can take as before

$$\mu_{n+1} := \frac{\partial_- I_{\tilde{\Upsilon}_n}(N/2) + \partial_+ I_{\tilde{\Upsilon}_n}(N/2)}{2},$$

in case 0 is in the spectrum of $\mathbf{F}_{\Upsilon_n} - \mu_{n+1}\mathbf{N}$.

(2) One lets

$$\tilde{\Upsilon}_{n+1} = t_{n+1}\tilde{\Upsilon}_n + (1 - t_{n+1})\Upsilon_{n+1}$$

where the damping parameter $t_{n+1} \in [0,1]$ is chosen such as to minimize the (quadratic) function

$$t \mapsto \mathcal{E}\big(t\tilde{\Upsilon}_n + (1-t)\Upsilon_{n+1}\big).$$

(3) The algorithm is stopped when $\|[\Upsilon_n, \mathbf{F}_{\Upsilon_n}]\|$ and/or $\|\Upsilon_{n+1} - \Upsilon_n\|$ are smaller than a prescribed ε.

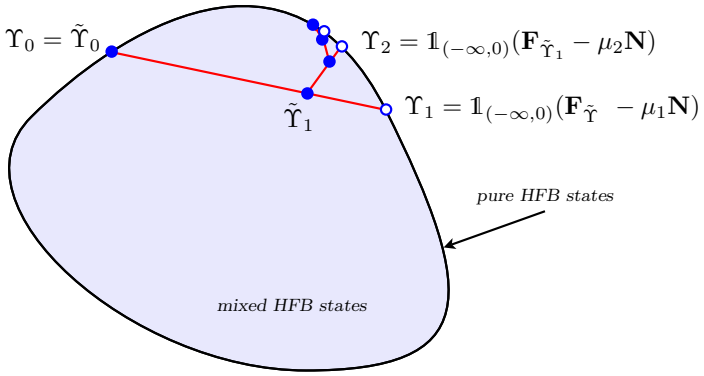

Fig. 1. The Optimal Damping Algorithm of Cancès & Le Bris in the HFB case.

The general strategy of the ODA is displayed in Figure 1. By construction we see that $\mathcal{E}(\tilde{\Upsilon}_n)$ is a non-increasing sequence. This guarantees the convergence of the ODA. The result is the following

Theorem 4.2 (Convergence of the ODA). *Assume that* $0 < N/2 < N_b$. *Let* $\Upsilon_0 = \tilde{\Upsilon}_0$ *be an initial HFB state such that the sequence* (Υ_n) *generated by the ODA is* uniformly well posed, *that is*

$$\forall n, \qquad \big|\mathbf{F}_{\tilde{\Upsilon}_n} - \mu_{n+1}\mathbf{N}\big| \geqslant \eta > 0. \tag{4.19}$$

Then

- *The sequence* $\mathcal{E}(\tilde{\Upsilon}_n)$ *decreases towards a critical value of* $\mathcal{E}$;
- *The sequence* Υ_n numerically *converges towards a critical point* Υ *of* $\mathcal{E}$, *in the sense that* $\Upsilon_{n+1} - \Upsilon_n \to 0$, $\Upsilon_{n+1} - \tilde{\Upsilon}_n \to 0$ *and that all the limit points* Υ *of subsequences of* (Υ_n) *solve* $\Upsilon = \mathbb{1}_{(-\infty,0)}(\mathbf{F}_\Upsilon - \mu\mathbf{N})$.

Proof. The proof is exactly the same as in the Hartree-Fock case [10, 14] and we only sketch it. First we have by definition $\mathcal{E}(\tilde{\Upsilon}_{n+1}) \leqslant \mathcal{E}(\tilde{\Upsilon}_n)$, so $\mathcal{E}(\tilde{\Upsilon}_n)$ must converge to a limit ℓ. Now we have

$$\mathcal{E}(\tilde{\Upsilon}_{n+1}) = \mathcal{E}\big((1-t_{n+1})\tilde{\Upsilon}_n + t_{n+1}\Upsilon_{n+1}\big) = \mathcal{E}(\tilde{\Upsilon}_n) - t_{n+1}a_{n+1} + t_{n+1}^2 b_{n+1}$$

where

$$t_{n+1} = \operatorname{argmin}_{t\in[0,1]}\big(-ta_{n+1} + t^2 b_{n+1}\big)$$

and with

$$a_{n+1} := \operatorname{Tr} \mathbf{F}_{\tilde{\Upsilon}_n}\big(\tilde{\Upsilon}_n - \Upsilon_{n+1}\big) = \operatorname{Tr} |\mathbf{F}_{\tilde{\Upsilon}_n} - \mu\mathbf{N}|\big(\tilde{\Upsilon}_n - \Upsilon_{n+1}\big)^2 \geqslant \eta \left\|\tilde{\Upsilon}_n - \Upsilon_{n+1}\right\|^2,$$

$$b_{n+1} = 2\operatorname{Tr}(\tilde{G}_{n+1} - G_n)J(\tilde{G}_{n+1} - G_n) - \operatorname{Tr}(\tilde{G}_{n+1} - G_n)K(\tilde{G}_{n+1} - G_n) + \operatorname{Tr}(\tilde{A}_{n+1} - A_n)K(\tilde{A}_{n+1} - A_n).$$

In finite dimension we have $|b_{n+1}| \leqslant C \left\|\tilde{\Upsilon}_{n+1} - \Upsilon_n\right\|^2 \leqslant (C/\eta)a_{n+1}$. This can be used to prove that

$$-t_{n+1}a_{n+1} + t_{n+1}^2 b_{n+1} \leqslant -\epsilon\, a_{n+1} \leqslant -\epsilon\,\eta \left\|\tilde{\Upsilon}_n - \Upsilon_{n+1}\right\|^2$$

for some $\epsilon > 0$ independent of n. This now proves that

$$\sum_n \left\|\tilde{\Upsilon}_n - \Upsilon_{n+1}\right\|^2 < \infty,$$

hence that $\tilde{\Upsilon}_n - \Upsilon_{n+1} \to 0$. In order to conclude the proof, we notice that

$$\Upsilon_{n+1} - \Upsilon_n = \Upsilon_{n+1} - \tilde{\Upsilon}_n + (1 - t_n)\big(\tilde{\Upsilon}_{n-1} - \Upsilon_n\big)$$

which finally implies

$$\sum_n \|\Upsilon_n - \Upsilon_{n+1}\|^2 < \infty \quad \text{and} \quad \sum_n \left\|\tilde{\Upsilon}_n - \tilde{\Upsilon}_{n+1}\right\|^2 < \infty.$$

Since $\Upsilon_{n+1} = \mathbb{1}_{(-\infty,0)}(\mathbf{F}_{\tilde{\Upsilon}_n} - \mu_{n+1}\mathbf{N})$ by definition, the proof that any limit Υ of a subsequence of (Υ_n) satisfies the self-consistent equation is elementary. □

4.3. *Handling constraints*

Both the Roothaan algorithm and the ODA are based on Lemma 4.1 which says that for any given $\mathbf{F}_\Upsilon$, there exist μ', δ' and Υ' such that

$$\begin{cases} \Upsilon' = \mathbb{1}_{(-\infty,0)}\big(\mathbf{F}_\Upsilon - \mu'\mathbf{N}\big) + \delta', \\ \operatorname{Tr} \mathbf{N}\Upsilon' = N - N_b. \end{cases} \tag{4.20}$$

The purpose of this section is to explain how to solve this problem numerically. To simplify our notation, we consider in this section a generic matrix

$$\mathbf{F} = \begin{pmatrix} h & p \\ p & -h \end{pmatrix}, \qquad \text{with } p = \overline{p} = p^T \text{ and } h = \overline{h} = h^T \tag{4.21}$$

and we study the problem consisting in finding Υ, μ and δ such that

$$\begin{cases} \Upsilon = \mathbb{1}_{(-\infty,0)}\big(\mathbf{F} - \mu\mathbf{N}\big) + \delta, \\ \operatorname{Tr} \mathbf{N}\Upsilon = N - N_b. \end{cases} \tag{4.22}$$

Assume first that $p \equiv 0$ (Hartree-Fock case). Then we have

$$\mathbf{F} = \begin{pmatrix} h & 0 \\ 0 & -h \end{pmatrix}$$

which commutes with $\mathbf{N}$. The solution of (4.22) is then given by the *aufbau principle*,

$$\Upsilon = \begin{pmatrix} G & 0 \\ 0 & 1-G \end{pmatrix}, \qquad G = \mathbb{1}_{(-\infty,\mu)}(h) + \delta$$

where μ is the $(N/2)$th eigenvalue of h, counted with multiplicity and δ lives in the corresponding eigenspace. Equivalently,

$$G = \sum_{i=1}^{K} v_i v_i^T + \sum_{i=K+1}^{K'} n_i \, v_i v_i^T$$

where the v_i's solve the eigenvalue equation

$$h \, v_i = \epsilon_i \, v_i,$$

$K = \operatorname{Tr} \mathbb{1}_{(-\infty,\epsilon_{N/2})}(h)$ is the dimension of the direct sum of all the eigenspaces corresponding to the eigenvalues $< \epsilon_{N/2}$ and $K' = \operatorname{Tr} \mathbb{1}_{(-\infty,\epsilon_{N/2}]}(h)$ is the dimension of the direct sum of all the eigenspaces corresponding to the eigenvalues $\leqslant \epsilon_{N/2}$. The n_i's are chosen such that

$$0 \leqslant n_i \leqslant 1, \qquad K + \sum_{i=K+1}^{K'} n_i = \frac{N}{2}.$$

Therefore, finding Υ, μ and δ in the Hartree-Fock case only requires to diagonalize h once.

In the Hartree-Fock-Bogoliubov case ($p \neq 0$), the situation is more complicated since $\mathbf{N}$ does *not* commute with $\mathbf{F}$. Let us consider the real function

$$\nu_{\mathbf{F}} : \mu \mapsto \nu(\mu) = \frac{\operatorname{Tr} \mathbf{N} \, \mathbb{1}_{(-\infty,0)}\big(\mathbf{F} - \mu \mathbf{N}\big) + N_b}{2}. \tag{4.23}$$

We are interested in solving the equation

$$\nu_{\mathbf{F}}(\mu) = N/2.$$

In the Hartree-Fock case, $\nu_{\mathbf{F}}$ is a non-decreasing piecewise constant function. There is a solution μ to $\nu_{\mathbf{F}}(\mu') = N/2$ when $N/2$ belong to the range of $\nu_{\mathbf{F}}$. Otherwise, one has to partially fill a shell using the matrix δ.

In the Hartree-Fock-Bogoliubov case, $\nu_{\mathbf{F}}$ is also non-decreasing and in general it is much smoother when $p \neq 0$. The following lemma summarizes some important properties of $\nu_{\mathbf{F}}$ in both the HF and HFB cases.

Lemma 4.3 (Elementary properties of ν). *Let $\mathbf{F}$ be as in (4.21). Then the function $\nu_{\mathbf{F}}$ defined in (4.23) is increasing with respect to μ. It can only have finitely many jumps. It satisfies for some constant C depending only on N_b*

- $\nu(\mu) \leqslant C/\mu$ *for* $\mu \leqslant -C$;
- $\nu_{\mathbf{F}}(\mu) \geqslant N_b - C/\mu$ *for* $\mu \geqslant C$.

If $0 \notin \sigma\big(\mathbf{F} - \mu \mathbf{N}\big)$, then

$$\frac{d\nu_{\mathbf{F}}}{d\mu}(\mu) = 2 \sum_{\substack{\epsilon_i < 0 \\ \epsilon_j > 0}} \frac{|\langle v_j, \mathbf{N} v_i \rangle|^2}{\epsilon_j - \epsilon_i} \geqslant 0 \tag{4.24}$$

where $(\mathbf{F} - \mu \mathbf{N}) v_i = \epsilon_i \, v_i$.

Proof. The behavior of $\nu_{\mathbf{F}}$ for $|\mu| \gg 1$ was already studied in Lemma 4.2.

The matrix $\mathbf{F} - \mu\mathbf{N}$ is a linear function of $\mu \in \mathbb{R}$, hence by [28], we know that its eigenvalues form a set of real analytic functions. They cannot be constant because the matrix $\mathbf{N}$ does not vanish. The eigenvalues of $\mathbf{F} - \mu\mathbf{N}$ all behave like $\pm\mu$ for large μ, by perturbation theory. We conclude that 0 can be an eigenvalue of $\mathbf{F} - \mu\mathbf{N}$ for a finite number of μ's, say $\mu_1 < \cdots < \mu_K$. On the other hand, the map $\mu \mapsto \mathbb{1}_{(-\infty,0)}\big(\mathbf{F} - \mu\mathbf{N}\big)$ is real-analytic outside of the μ_k's (see [28]). So $\nu_{\mathbf{F}}$ is itself real-analytic outside of this set and it can have at most a finite number of jumps.

Outside of the μ_k's, it is possible to compute the derivative of $\nu_{\mathbf{F}}$ by usual perturbation methods [28]. The answer is (4.24) and the fact that $d\nu_{\mathbf{F}}/d\mu \geqslant 0$ proves that $\nu_{\mathbf{F}}$ is increasing with respect to μ, in between these points. That the jumps are all positive can be easily seen by an approximation argument using Lemma 4.4 below. We skip the details. □

The shape of the function $\nu_{\mathbf{F}}$ is very different in the HF and HFB cases. For a Hartree-Fock state, the function $\nu_{\mathbf{F}}$ is piecewise constant and it has jumps at the eigenvalues $\epsilon_1 < \cdots < \epsilon_{N_b}$ of h_G. The size of the jumps is equal to the multiplicity of the associated eigenvalue. An HFB state will most always have a very smooth $\nu_{\mathbf{F}}$. Of course, the smaller p in the Hamiltonian $\mathbf{F}$, the more $\nu_{\mathbf{F}}$ looks like a step function.

In Figure 2 below, we show the function $\nu_{\mathbf{F}}$ for different values of the pairing term. More precisely, we have randomly chosen two symmetric real matrices h and p of size $N_b = 5$, and we display the function $\nu_{\mathbf{F}}$ when the pairing is replaced by tp for $t = 0$ (Hartree-Fock case), $t = 0.1$ and $t = 1$. Figure 3 is a plot of the eigenvalues of $\mathbf{F} - \mu\mathbf{N}$ for $t = 0.1$, as functions of μ. Note that there are some crossings of eigenvalues above and below the real line (recall that the spectrum is symmetric with respect to 0). But, around 0 the crossings are avoided and there is a gap.

If we repeat the numerical experiment with several random matrices h and p, we *never* see any jump for $\nu_{\mathbf{F}}$. The purpose of the next result is to clarify this observation.

Lemma 4.4 (Generic behavior of $\nu_{\mathbf{F}}$). *The Fock matrix $\mathbf{F} - \mu\mathbf{N}$ is invertible if and only if $h \pm ip - \mu$ are invertible. More precisely,*

$$\min \sigma\big(|\mathbf{F} - \mu\mathbf{N}|\big) = \min\left(\left\|(h + ip - \mu)^{-1}\right\|^{-1}, \left\|(h - ip - \mu)^{-1}\right\|^{-1}\right). \tag{4.25}$$

The set of real symmetric matrices h and p such that

$$\sigma\big(\mathbf{F} - \mu\mathbf{N}\big) \cap \{0\} = \emptyset \quad \textit{for all } \mu \in \mathbb{R}$$

is open and dense in $\{(h, p) \ : \ h = h^T = \overline{h},\ p = p^T = \overline{p}\}$. For h and p in this set, $\nu_{\mathbf{F}}$ is real-analytic on $\mathbb{R}$.

It is obvious that there are matrices h and p for which $\mathbf{F} - \mu\mathbf{N}$ has 0 as eigenvalue for some $\mu \in \mathbb{R}$. The simplest examples are HF Hamiltonians for which $p \equiv 0$ and $\mathbf{F} - \mu\mathbf{N}$ is not invertible each time μ equals an eigenvalue of h. If p does not vanish but commutes with h, then we have $|h + ip - \mu|^2 = |h - ip - \mu|^2 = (h - \mu)^2 + p^2$ and we see that 0 is never in the spectrum of $\mathbf{F} - \mu\mathbf{N}$ when the kernel of p does not contain the eigenvectors of h. However there are counterexamples with p invertible not commuting with h. For instance, $\mathbf{F} + \mathbf{N}$ is not invertible for

$$h = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad p = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}.$$

We now turn to the proof of Lemma 4.4.

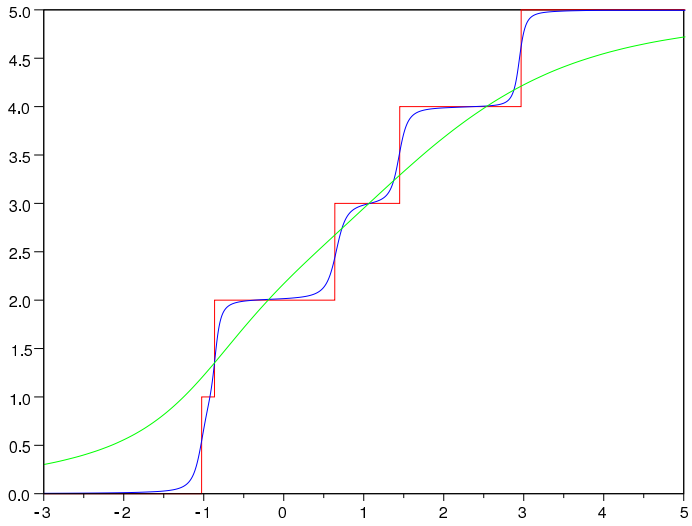

Fig. 2. The function $\nu_{\mathbf{F}}(\mu)$ which gives the average number of particles in the state $\mathbb{1}_{(-\infty,0)}(\mathbf{F}-\mu\mathbf{N})$, in terms of the chemical potential μ. The pairing term in $\mathbf{F}$ is equal to tp with $t=0$ (Hartree-Fock case, red curve), $t=0.1$ (blue curve) and $t=1$ (green curve).

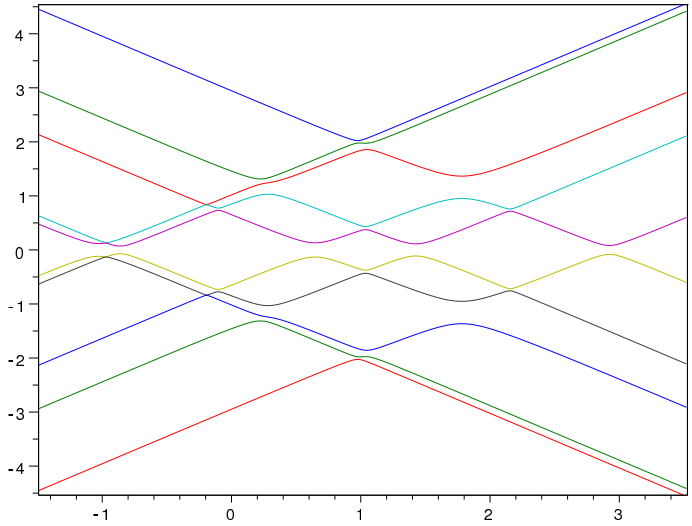

Fig. 3. The eigenvalues of $\mathbf{F}-\mu\mathbf{N}$ in terms of μ for $t=0.1$.

Proof. The operator $\mathbf{F}-\mu\mathbf{N}$ is unitarily equivalent to

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} (\mathbf{F}-\mu\mathbf{N}) \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = -i \begin{pmatrix} 0 & h+ip-\mu \\ -(h-ip-\mu) & 0 \end{pmatrix}.$$

From this we deduce that $\mathbf{F}-\mu\mathbf{N}$ is invertible if and only if $h+ip-\mu$ and $h-ip-\mu$ are

invertible. Then we have

$$\big(\mathbf{F}-\mu\mathbf{N}\big)^{-1} = i\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 0 & -(h-ip-\mu)^{-1} \\ (h+ip-\mu)^{-1} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

and

$$\Big\|\big(\mathbf{F}-\mu\mathbf{N}\big)^{-1}\Big\| = \max\Big(\,\big\|(h+ip-\mu)^{-1}\big\|\,,\,\big\|(h-ip-\mu)^{-1}\big\|\,\Big).$$

The statement now follows from the fact that, on a dense open set, the spectra of $h \pm ip$ do not intersect the real axis. □

Lemma 4.4 is interesting when we apply the Roothaan or the ODA, because it means that, as soon as $p \neq 0$, most often we will have no choice for μ_{n+1} and we will take $\delta_{n+1} = 0$. Saying differently, it is really reasonable to assume that the sequence generated by the Roothaan and the ODA are uniformly well posed (of course when the final state is believed to have a non vanishing pairing), as we did in Theorem 4.1 and 4.2.

Even if it is in general smooth, the function ν can still vary quickly and this will be the case when the pairing term p is small. The appropriate method to find the solution of $\nu_{\mathbf{F}}(\mu) = N/2$ then depends on the properties of $\nu_{\mathbf{F}}$. If the Hamiltonian $\mathbf{F}$ has a large enough pairing matrix p, then $\nu_{\mathbf{F}}$ is smooth and we can use a Newton-like method to solve the equation $\nu_{\mathbf{F}}(\mu) = N/2$. A trial chemical potential μ^0 being given, we compute the derivative $\partial\nu_{\mathbf{F}}/\partial\mu(\mu^0)$ using Formula (4.24) and then let

$$\mu^1 := \mu^0 + \big(N/2 - \nu_{\mathbf{F}}(\mu^0)\big)\left(\frac{\partial\nu_{\mathbf{F}}}{\partial\mu}(\mu^0)\right)^{-1}.$$

The method can be iterated until convergence of μ^n towards the desired μ. The convergence is very fast, as soon as $\nu_{\mathbf{F}}$ is smooth.

If the Hamiltonian $\mathbf{F}$ has a small pairing matrix p, the function $\nu_{\mathbf{F}}$ will be smooth but close to a step function. Its derivative varies very quickly and the previous Newton method is not appropriate. In this case we can use a simple bisection method. The bounds on $\nu_{\mathbf{F}}(\mu)$ for large $|\mu|$ can be used to find a good starting interval $[\mu_l, \mu_r]$ such that $\nu_{\mathbf{F}}(\mu_l) < N/2$ and $\nu_{\mathbf{F}}(\mu_r) > N/2$.

We have to find a new μ_{n+1} at each step of the Roothaan or ODA. It is of course not efficient to find μ_{n+1} with a very high precision all along the algorithm. Dechargé and Gogny advice in Section II.E of [16] to apply the Newton scheme only once at each step. This means that

$$\mu_{n+1} = \mu_n + \big(N/2 - \nu_n(\mu_n)\big)\left(\frac{\partial\nu_n}{\partial\mu}(\mu_n)\right)^{-1}$$

where ν_n is the function ν corresponding to $\mathbf{F}_\Upsilon = \mathbf{F}_{\Upsilon_n}$. This is then the same as doing perturbation theory on first order. We use a slightly different strategy which we explain in the next section.

5. Numerical results

In this section, we present some numerical results for two very simple interactions W. We start by considering in Section 5.2 a purely 3-dimensional gravitational model in which

$$W(x) = -\frac{g}{|x|}, \qquad g > 0.$$

Then we consider in Section 5.3 a (repulsive) Coulomb potential which is perturbed at intermediate distances by an attractive effective potential, as is usually employed in nuclear physics:

$$W(x) = \frac{\kappa}{|x|} - a_1 \exp\left(-b_1|x|^2\right) + a_2 \exp\left(-b_2|x|^2\right), \qquad \kappa > 0.$$

In the next section we quickly explain our numerical technique to treat these two simple systems.

5.1. *Method*

To simulate our physical systems, we have used the open source software Scilab [43]. Our potential W is always real, radial and spin-independent. To reduce the numerical cost we have therefore always imposed the spin, time-reversal and spherical symmetry. This means that we have to cope with $\ell_{\max}+1$ real and symmetric $N_b \times N_b$ matrices G^ℓ and A^ℓ (the one-particle density matrix and the pairing density matrix in the ℓth angular momentum sector). The total energy of the system is given by Equation (3.29) and we have to impose the constraints (3.27) and (3.28) which we recall here for convenience:

$$0 \leqslant \Upsilon^\ell \mathbf{S} \Upsilon^\ell \leqslant \Upsilon^\ell, \quad \text{with} \quad \Upsilon^\ell := \begin{pmatrix} G^\ell & A^\ell \\ A^\ell & S^{-1} - G^\ell \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}, \tag{5.1}$$

$$\sum_{\ell=0}^{\ell_{\max}} (2\ell+1) \operatorname{Tr}(SG^\ell) = N/2. \tag{5.2}$$

We choose a simple basis set $(\chi_1, ..., \chi_{N_b})$ of $L^2([0,\infty), r^2 dr)$, made of "hat functions" associated with a chosen grid

$$0 = r_0 < r_1 < \cdots < r_{N_b} < r_{N_b+1} := r_{\max}.$$

We impose Dirichlet boundary conditions at $r_{\max}$. We have tested different types of grids and there was no important difference between them. The results presented here are all with regular grids. As we will explain later, for a given basis size N_b, the results usually depend a lot on the value of the radius $r_{\max}$ of the ball in which the system is placed.

Our main goal is to investigate the existence of pairing. We therefore always start by doing a precise Hartree-Fock calculation, for which we use the Optimal Damping Algorithm described in Section 4.2. We take as initial state a simple uniform state

$$G_{\text{init}} = \frac{N}{2\operatorname{Tr}(S)} \mathrm{Id}_{N_b} \tag{5.3}$$

and we run HF until convergence. We have observed a global stability of the results with respect to initial states, hence the previous simple choice is appropriate (but more clever choices might decrease the total number of iterations). Then, we use the converged HF state G_{opt} as initial datum for the HFB algorithm. Of course we have to perturb it a little bit since any HF solution is also an HFB solution. We proceed as follows. Assuming that the overlap matrix $S = \mathrm{Id}_{N_b}$ and that $\ell_{\max} = 0$ for simplicity, the optimal HF state G_{opt} can be written in the form

$$G_{\text{opt}} = \sum_{k=1}^{N/2} v_k v_k^T,$$

where v_k are the $N/2$ first eigenvectors of the mean-field matrix h,

$$hv_k = \epsilon_k \, v_k.$$

We then choose a number n_v of valence orbitals and a mixing parameter θ, and we perturb $G_{\rm opt}$ as follows

$$G'_{\rm init} = \sum_{k=1}^{N/2-n_v} v_k v_k^T + \theta \sum_{k=N/2-n_v+1}^{N/2} v_k v_k^T + (1-\theta) \sum_{k=N/2+1}^{N/2+n_v} v_k v_k^T,$$

$$A'_{\rm init} = \sqrt{\theta(1-\theta)} \sum_{k=N/2-n_v+1}^{N/2+n_v} v_k v_k^T.$$

In most cases, we have observed that $n_v = 1$ and $\theta = 0.95$ works perfectly well, that is, the algorithm escapes from the HF solution $G_{\rm opt}$ and converges towards an optimal HFB state. But other values of n_v and θ seem to work fine also.

When the maximum angular momentum $\ell_{\max}$ is larger than 0, we often first run the algorithm with $\ell_{\max} = 0$ for a few iterations before switching to the actual value of $\ell_{\max}$. We stop the algorithm when the commutators $[F_n^\ell, \Upsilon_n^\ell]$ are smaller than a prescribed error. We know from (3.30) and (3.31) that these commutators must all vanish for an exact solution of the discretized HFB minimization problem. In terms of the matrix $\mathbf{S}$, the right quantity to look at is

$$\sum_{\ell=0}^{\ell_{\max}} \left\| \mathbf{S}^{-\frac{1}{2}} \big(\mathbf{F}_n^\ell \Upsilon_n^\ell \mathbf{S} - \mathbf{S} \Upsilon_n^\ell \mathbf{F}_n^\ell \big) \mathbf{S}^{-\frac{1}{2}} \right\|$$

where $\|\cdot\|$ is the usual operator norm for $(2N_b) \times (2N_b)$ matrices. There is a similar formula in the HF case [11].

As we have explained, in the HFB case, ensuring the constraint (5.2) is not as easy as in the HF case. In the beginning of the algorithm, our state Υ is rather close to an HF state by construction. Therefore, the function $\nu_{\mathbf{F}}(\mu)$ defined in Section 4.3 is close to a step function. We choose an error ε and look for the next states Υ_{n+1}^ℓ having a total number of particles $\sum_{\ell=0}^{\ell_{\max}} (2\ell+1) \operatorname{Tr}(S G_{n+1}^\ell)$ close to $N/2$, within the error ε, using a simple bisection method. We use the bisection for a fixed number of global iterations. Then, when the pairing term is large enough, we switch to a Newton method in order to find the state Υ_{n+1}. We have observed that even if in the beginning several Newton iterations can be employed at each step, usually only one Newton iteration is necessary after a while. To guarantee a good value of the average number of particles in the end, we decrease the error ε on $|\sum_{\ell=0}^{\ell_{\max}} (2\ell+1) \operatorname{Tr}(S G_{n+1}^\ell) - N/2|$ along the algorithm.

5.2. *Pure Newtonian interaction*

5.2.1. *Model*

Here we consider a system of N spin-1/2 neutral particles, only interacting through the Newtonian interaction

$$W(x) = -\frac{g}{|x|}, \qquad g > 0. \tag{5.4}$$

This potential is strongly attractive at short distances. Since $1/|x|$ does not decay too fast at infinity, it is also quite attractive at large distances. The kinetic energy does not scale the same as the potential energy. By a simple scaling argument, we can therefore always assume that

$$g \equiv 1.$$

This model can be used to describe neutron stars and white dwarfs when $N \gg 1$. It has been particularly studied from a theoretical point of view in the pseudo-relativistic case where the kinetic energy is given by $T = \sqrt{c^4 m^2 - c^2 \Delta} - mc^2$, see [36, 37, 30]. In our simulations we restrict ourselves to the non-relativistic case of the Laplacian $T = -\Delta/(2m)$ (in units such that $m = 1/2$). It would be interesting to take N large but this is of course much too difficult from a numerical point of view.

As mentioned before, we always impose the spin and time-reversal symmetries, which is perfectly justified for the ground state since the interaction (5.4) satisfies the assumption of the Bach-Fröhlich-Jonsson Theorem 2.2. We also impose spherical symmetry which, on the contrary, is not known to hold for the true ground state.

One advantage of the Newtonian interaction (5.4) is that the operators J and $K^{\ell\ell'}$ can be explicitely computed in the basis of hat functions. We have shown in Section 3.3.2 that the energy can be expressed in terms of

$$(ij|mn)_{\ell,\ell'} = \int_0^\infty r^2\,dr \int_0^\infty s^2\,ds\;\chi_i(r)\,\chi_j(r)\,\chi_m(s)\,\chi_n(s)\,w_{\ell,\ell'}(r,s)$$

where

$$w_{\ell,\ell'}(r,s) = \frac{1}{2}\int_{-1}^1 W\left(\sqrt{r^2+s^2-2rst}\right)\,P_\ell(t)\,P_{\ell'}(t)\,dt = -\frac{1}{2}\int_{-1}^1 \frac{P_\ell(t)\,P_{\ell'}(t)}{\sqrt{r^2+s^2-2rst}}\,dt.$$

Using the well-known formula

$$\frac{1}{\sqrt{r^2+s^2-2rst}} = \sum_{n=0}^\infty \frac{\min(r,s)^n}{\max(r,s)^{n+1}} P_n(t)$$

we deduce that

$$w_{\ell,\ell'} = -\frac{1}{2}\sum_{n=0}^\infty \left(\int_{-1}^1 P_n P_\ell P_{\ell'}\right)\frac{\min(r,s)^n}{\max(r,s)^{n+1}}.$$

The integral over the Legendre polynomials is related to the usual Clebsch-Gordan coefficients as follows

$$\frac{1}{2}\int_{-1}^1 P_n(t)\,P_\ell(t)\,P_{\ell'}(t)\,dt = \begin{pmatrix} \ell & \ell' & n \\ 0 & 0 & 0 \end{pmatrix}^2$$

and only a finite number of terms are non zero in the sum over n. The final result can be expressed as

$$\begin{aligned}&(ij|mn)_{\ell,\ell'}\\ &\quad= -\sum_{n=0}^\infty \begin{pmatrix} \ell & \ell' & n \\ 0 & 0 & 0 \end{pmatrix}^2 \int_0^\infty r^2\,dr \int_0^\infty s^2\,ds \frac{\min(r,s)^n}{\max(r,s)^{n+1}}\;\chi_i(r)\,\chi_j(r)\,\chi_m(s)\,\chi_n(s). \quad (5.5)\end{aligned}$$

These integrals can be explicitely computed for hat functions and $0 \leqslant \ell, \ell' \leqslant \ell_{\max}$ with $\ell_{\max}$ not too large. In our numerical experiments we have put the explicit formulas in Scilab for $\ell_{\max} = 1$. The integrals were stored in memory during the whole calculation.

5.2.2. *Roothaan vs ODA*

In the HF case, we have observed that the Roothaan algorithm very often oscillates between two states, none of them being the solution of the problem (as described in Theorem 4.1 and in [13]). The Roothaan algorithm seems more well behaved in the HFB case. With the model presented in this section, we never got real oscillations for HFB. Sometimes the convergence is improved by using the ODA, but in most cases the Roothaan algorithm always converges towards the same state as the ODA in the end. As we will see later, the situation is very different for the model studied in Section 5.3, which is inspired of nuclear physics.

We start by comparing Roothaan and ODA in the HF case. There, oscillations seem to be related to the size of the gap between the largest filled eigenvalue and the smallest unfilled one. Indeed, oscillations in HF seem to only occur when there is pairing in HFB, an effect which is also well-known to be related to the size of the gap (see, e.g., Theorem 5 in [2]). When there is no pairing, the HF Roothaan algorithm always behaves like the ODA. However, the situation is complex and there is no exact rule. Sometimes the Roothaan algorithm does *not* oscillate even when the gap is rather small and there is pairing.

In Figure 4 we display the value of the energy obtained along the algorithm for the Roothaan and the ODA, for the following choice of parameters: $N = 6$, $N_b = 200$, $\ell_{\max} = 0$ and $r_{\max} = 30$. The ODA converges in about 17 iterations, whereas the Roothaan algorithm oscillates. We also show the value of the norms $\|G_n - G_{n-1}\|$ and $\|G_n - G_{n-2}\|$ along the Roothaan algorithm. The oscillation between two points is clearly demonstrated.

When we decrease the parameter $r_{\max}$ but keep $N_b = 200$ constant, the gap is seen to increase slightly and the Roothaan algorithm behaves better. In Table 1, we give the numerical value of the last filled eigenvalue and the corresponding gap. The Roothaan algorithm slowly converges for $r_{\max} = 25$ and it coincides with the ODA when $r_{\max} = 20$. The gap for $r_{\max} = 20$ is 2.5 times the one for $r_{\max} = 30$. We will discuss the occurence of pairing in terms of the parameter $r_{\max}$ in the next section.

$r_{\max}$	$\epsilon_{N/2}$	$\epsilon_{N/2+1} - \epsilon_{N/2}$	behavior of HF Roothaan
20	-0.532430	0.159430	fast convergence
25	-0.536706	0.081016	slow convergence
30	-0.529200	0.061928	oscillations
	-0.548554	0.067422	

Table 1. Value of the last filled eigenvalue $\epsilon_{N/2}$ and the corresponding gap $\epsilon_{N/2+1} - \epsilon_{N/2}$ in HF, for $N = 6$, $N_b = 200$ and $\ell_{\max} = 0$. For $r_{\max} = 30$ the Roothaan algorithm oscillates and we display the last filled eigenvalue and the gap for the two states.

As we have mentioned the Roothaan algorithm is usually much more well behaved in the HFB case. However, sometimes the convergence can be improved dramatically by using the ODA. In Figure 5 we display the energy along the iterations of the algorithm in both the Roothaan and ODA cases, for $N_b = 500$, $N = 16$, $\ell_{\max} = 1$ and $r_{\max} = 10$. In this case the Roothaan algorithm is very badly behaved. It passes very close to the HF ground state and it takes it a very long time to escape from it. On the other hand, the ODA does not suffer from this problem and it converges much more rapidly.

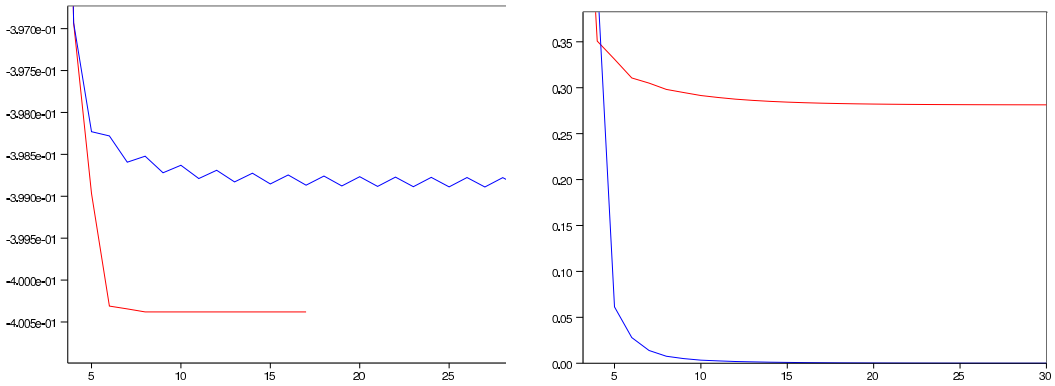

Fig. 4. *Left:* HF energy along the iterations for the Roothaan Algorithm (blue) and the ODA (red). *Right:* Values of $\|G_n - G_{n-1}\|$ (red) and $\|G_n - G_{n-2}\|$ (blue) along the Roothaan algorithm, showing the oscillations between two states. Here $N = 6$, $N_b = 200$, $\ell_{\max} = 0$ and $r_{\max} = 30$.

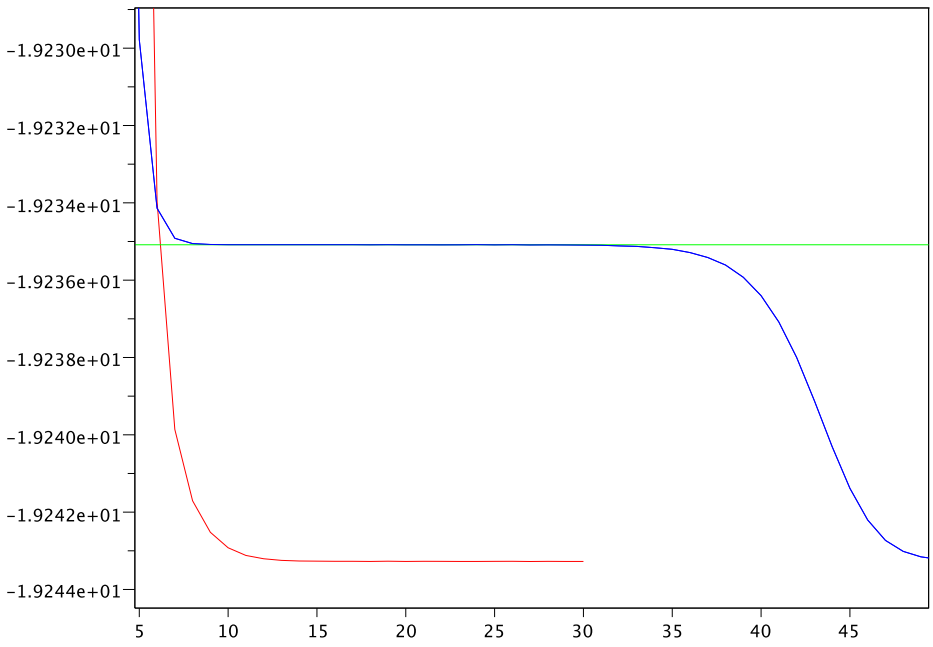

Fig. 5. HFB energy along the iterations of the Roothaan (blue) and the ODA (red) for $N_b = 500$, $N = 16$, $\ell_{\max} = 1$ and $r_{\max} = 10$. The optimal HF energy is also displayed (green).

5.2.3. *Numerical evidence of pairing*

Pairing effects in a finite discretization basis might depend on the properties of the basis. As we have explained, the occurence of pairing is related to the size of the gap in HF theory and this gap varies with the radius $r_{\max}$ in which the system is confined. If $r_{\max}$ is decreased the system is more condensed and the HF gap increases.

In Figure 6 we display the HF and HFB ground state energies computed for $N = 16$ in a basis set of size $N_b = 200$, in terms of $r_{\max}$. The HF and HFB curves are distinct for $r_{\max}$ large enough and they merge at $r_{\max} = 6$ approximately. This observation is confirmed by the value of the norm of A plotted on the right of the same figure. The minima of the HF and HFB ground state energies are attained at about $r_{\max} \simeq 10$ in the HF case and $r_{\max} \simeq 10.5$ in the HFB case, which is sufficiently far from the merging point. The minima of these curves correspond to the best possible approximation for a given basis size N_b (here $N_b = 200$) and a given type of grid (here regular). The difference between the corresponding HF and HFB energies is significant. The HF ground state energy at $r_{\max} = 10$ is -19.232176 (in our units in which $m = 1/2$ and $e = 1$), whereas the HFB ground state energy at $r_{\max} = 10.5$ is -19.240176. The norm of the pairing matrix A is rather large at this point:

$$\|A\| = \sqrt{\mathrm{Tr}(SA_0SA_0) + 3\,\mathrm{Tr}(SA_1SA_1)} \simeq 0.462129.$$

This goes in favour of the conclusion that pairing really occurs for $N = 16$ in this model. This intuition is confirmed by a more precise calculations with $N_b = 500$ which we discuss below.

The observation of pairing requires to have an appropriate $r_{\max}$ but it does *not* require to have a very large basis set. Even for $N_b = 30$ and $r_{\max} = 10.5$, we already find that the HFB energy is approximately -19.078416 whereas the HF energy is about -19.072954. The corresponding norm of the pairing matrix A is $\|A\| \simeq 0.424124$.

Pairing is a subtle effect which decreases the energy by a small amount (much less than one percent here). Catching this effect requires to be very careful when choosing the radius $r_{\max}$. Taking $r_{\max}$ too small might lead to the conclusion that there is no pairing. In our simulations we have always observed the occurence of pairing, but provided we choose $r_{\max}$ appropriately. The values of $r_{\max}$ at which the HF and HFB energies attain their minimum were always found on the right of the merging point of the two curves. In Table 2 below we give our results for $N_b = 200$ and $N = 6, 10, 16$ and 20. The HFB ground state energy is always smaller than the HF energy.

In the paper [30], Lenzmann and Lewin have rigorously studied the gravitational model of this section. They showed the existence of a ground state in both the HF and HFB cases. But, so far, no proof that pairing occurs has been provided. The numerical results of this section tend to show that there is actually always pairing, at least for N not too large.

5.2.4. *Properties of the HFB ground state*

In Table 2 below we give our results for $N_b = 200$ and $N = 6, 10, 16$ and 20, for the optimal values of $r_{\max}$. With $\ell_{\max} = 1$ we have observed that the shells are filled alternatively. In HF theory, the cases $N = 10$ and $N = 16$ correspond to closed shells, whereas for $N = 6$ and $N = 20$ the last shell is only partially filled. This is a simple explanation for the fact that the pairing matrix is much bigger in these cases.

In Table 3 we display the occupation numbers for the optimal HFB ground state in the

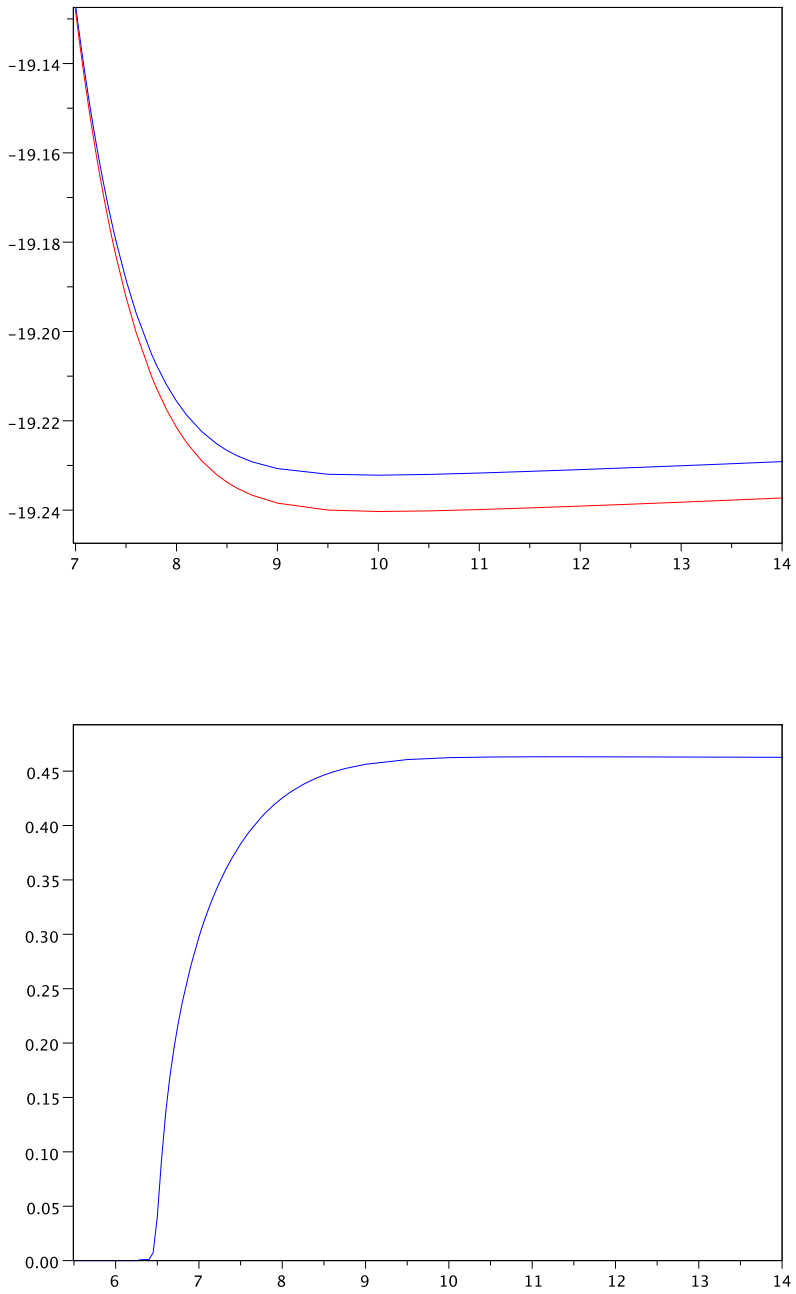

Fig. 6. Value of the ground state HF and HFB energies (top) and of the norm of the pairing matrix A (bottom), as functions of the radius $r_{\max}$ in which the system is confined, for $N = 16$, $N_b = 200$ and $\ell_{\max} = 1$.

closed shell case $N = 16$ and in the open shell case $N = 20$. Because of the spin, these are the eigenvalues of G_0 multiplied by 2 and that of G_1 multiplied by 6. Even in the closed shell case $N = 16$, a rather important pairing effect is observed between the last filled orbital (the second $\ell = 1$ eigenvalue) and the first unfilled one (the third $\ell = 0$ eigenvalue). This results in a decrease of the last occupation number of the HF one-particle density matrix by approximately 0.228.

N	$r_{\max}$	HF gap	HF energy	HFB energy	$\|A\|$
6	15	0	-1.7327688	-1.9934252	1.0242134
10	11	1.023642	-6.7911634	-6.8148576	0.5871951
16	10	1.404396	-19.232177	-19.2403096	0.4623593
20	9	0	-30.010574	-30.174576	0.8235512

Table 2. Results for $N_b = 200$ and $\ell_{\max} = 1$.

$N = 16$	
$\ell = 0$	$\ell = 1$
1.9999318	5.9997504
1.9980654	5.7710970
0.2281366	0.0026448
0.0002694	0.0002720
0.0000120	0.0000084
0.0000014	0.0000012
0.0000002	3.316D-07
6.896D-08	1.005D-07
⋮	⋮

$N = 20$	
$\ell = 0$	$\ell = 1$
1.9999832	5.9999490
1.9997458	5.9983572
1.9875134	2.0084946
0.0045222	0.0012960
0.0000690	0.0000552
0.0000058	0.0000066
0.0000010	0.0000002
0.0000002	3.072D-07
⋮	⋮

Table 3. Occupation numbers of the HFB minimizer, for $N_b = 200$ and $\ell_{\max} = 1$.

5.2.5. *Quality of the approximation in terms of the number N_b of points*

In Table 4 we display the HF and HFB ground state energies for $N = 16$, $\ell_{\max} = 1$ for the the optimal value of $r_{\max}$, in terms of the number of discretization points N_b of the regular grid. The convergence is not very fast, but we see that the difference between the HF and the HFB energy, as well as the norm of the pairing matrix are of the same order for small N_b as they are for larger N_b's. From this observation we can conclude that it is probably not necessary to take N_b very large in order to decide whether pairing occurs or not.

5.3. *A simplified model for protons and neutrons*

In this section we report on our numerical results concerning a simple model inspired of nuclear physics. The interaction between protons and neutrons is not a fundamental law of nature because these are composite particles made of quarks, which interact through weak, strong and electrostatic forces. A common procedure used in nuclear physics is to use *empirical forces* [41] which involve a small number of parameters which are fitted to

N_b	$r_{\max}$	HF energy	HFB energy	difference	$\|A\|$
30	9	-19.112314	-19.117948	0.005634	0.425604
50	9	-19.189066	-19.196012	0.006946	0.445728
100	9	-19.222300	-19.229872	0.007572	0.454173
150	10	-19.229494	-19.237574	0.008080	0.461725
200	10	-19.232176	-19.240308	0.008132	0.462363
250	10	-19.233420	-19.241576	0.008156	0.462659
300	10	-19.234094	-19.242264	0.008170	0.462821
400	11	-19.234826	-19.243068	0.008242	0.463905
500	11	-19.235206	-19.243456	0.008250	0.463985

Table 4. Value of the HF and HFB energies for $N = 16$ and $\ell_{\max} = 1$ and the (approximate) optimal $r_{\max}$.

experiment or to the known behavior of the model in some limits. The most common forces used in practice are the so-called Skyrme [45] and Gogny [20, 21, 16] forces and they depend nonlinearly on the state itself. Here we consider an effective force which is fixed and does not depend on the quantum state. We also take it spin-independent and isospin-independent. Our goal is to test some simple ideas and not to do a real nuclear physics calculation.

5.3.1. *Model*

The nucleon-nucleon potential has been observed to be repulsive at short distances and only attractive at medium distances. It decays very fast at infinity. A simple choice to describe this is to take

$$W(x) = \frac{\kappa}{|x|} - a_1\, e^{-b_1|x|^2} + a_2\, e^{-b_2|x|^2}, \tag{5.6}$$

with $a_2, a_1 > 0$, $b_1 < b_2$. The constant κ is 1 for the proton-proton interaction and 0 for the proton-neutron and the neutron-neutron interaction. The other constants usually also depend on the isospin (the quantum variable which determines whether a nucleon is a neutron or a proton). For simplicity we work here with particles having a definite isospin. This means that we assume to have either only protons or only neutrons. In particular we want to ask for which strength of the effective force it becomes possible for the protons to overcome their Coulomb repulsion and form a bound state. In reality a nucleus is made of a certain number of protons and neutrons and one has to use a different HFB state for each species.

In our applications we have chosen for simplicity $b_1 = 1$, $b_2 = 4$, $a_1 = a = 2\,a_2/3$. This means that the effective force takes the form

$$W(x) = \frac{\kappa}{|x|} + a\left(\frac{3}{2}\, e^{-4|x|^2} - e^{-|x|^2}\right). \tag{5.7}$$

When $\kappa = 1$, this force is purely repulsive for $a \leqslant 2.87$ and it becomes attractive at intermediate distances for larger a's. The corresponding force is displayed in Figure 7 for $a = 1$ and $\kappa = 0$.

One can ask several questions concerning the model considered in this section:

(1) For which value of a does a system of N identical nucleons bind in Hartree-Fock theory?

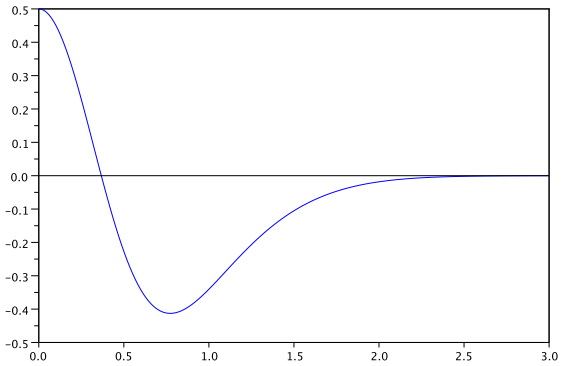

Fig. 7. The effective force $-e^{-|x|^2} + 3\,e^{-4|x|^2}/2$ used in our calculation of Section 5.3. The (repulsive) Coulomb potential must be added for protons.

(2) Is there always pairing when there is binding?
(3) Can pairing effects allow for binding with a smaller a than in HF theory?

These questions are mostly of academic nature for the very simplified model considered in this section. But investigating the same problems with more realistic forces is very important from a physical point of view. From a mathematical point of view, nothing seems to be known for simple models of the same form as in this section. It is not even known that binding always occurs for a large enough with the previous interaction. We hope that our calculations will stimulate some further mathematical studies.

5.3.2. *Some computational details*

We always minimize over states having the spin, time-reversal and rotation symmetries. The Bach-Fröhlich-Jonsson Theorem 2.2 does not apply to the model of this section, hence we are making a further approximation here.

For such symmetric states we have shown in Section 3.3.2 that the energy can be expressed in terms of

$$(ij|mn)_{\ell,\ell'} = \int_0^\infty r^2\,dr \int_0^\infty s^2\,ds\ \chi_i(r)\,\chi_j(r)\,\chi_m(s)\,\chi_n(s)\,w_{\ell,\ell'}(r,s)$$

where, for the model considered in this section,

$$\begin{aligned} w_{\ell,\ell'}(r,s) &= \frac{1}{2}\int_{-1}^{1} W\left(\sqrt{r^2+s^2-2rst}\right)\,P_\ell(t)\,P_{\ell'}(t)\,dt \\ &= \frac{1}{2}\int_{-1}^{1} P_\ell(t)\,P_{\ell'}(t)\Bigg(\frac{\kappa}{\sqrt{r^2+s^2-2rst}} \\ &\qquad - a_1\,e^{-b_1(r^2+s^2-2rst)} + a_2\,e^{-b_2(r^2+s^2-2rst)}\Bigg)\,dt. \end{aligned}$$

For $0 \leqslant \ell, \ell' \leqslant \ell_{\max}$ with $\ell_{\max}$ not too large, the Gaussian integrals can be computed exactly and it is possible to find the exact expression of $w_{\ell,\ell'}(r,s)$.

The computation of the integral $(ij|mn)_{\ell,\ell'}$ against hat functions is much more tedious, however. It is easy to find an exact expression for the Coulomb part, but not so simple for the Gaussian part. So we have performed a numerical calculation of these integrals. Since we have of the order of $(N_b)^4$ integrals, we could not take N_b too large. The results of the previous section indicated that the existence of pairing effects does not depend very much on the size of the basis.

5.3.3. *Slow convergence and oscillations of Roothaan*

We have observed that the Roothaan algorithm *almost always oscillates*, even in the HFB case (see some examples in Figures 8, 9 and 10). This is in stark contrast with the results of the previous section where the Roothaan algorithm was almost always converging. Sometimes it very slowly converges in the HF case (see, e.g., Figure 9). However we have always obtained convergence for the HF Roothaan algorithm when a is small enough, that is, when it is expected that there is actually no binding. For the case displayed in Figure 9 we have $a = 20$ but the critical a is about $\simeq 24$ (see the next section).

We conclude that using the ODA is very important for such attractive potentials. The same might be true with the more involved forces used in nuclear physics.

5.3.4. *The critical strength*

In finite dimension there is always a minimizer. Saying differently, since the particles are trapped in a ball, they always bind. Furthermore we work with rotation-invariant states. So, for the true model in infinite dimension, the particles escaping to infinity cannot form a bound state of the same kind because they are too far from the (fixed) center of symmetry. In this special case they will spread out and have a vanishing energy.

In Hartree-Fock theory, we conclude that we can detect the loss of binding by looking at the last filled HF eigenvalue. When it crosses 0, this corresponds to the last particle becoming a scattering state. We can therefore choose as definition for the critical strength a, the value at which this eigenvalue is 0. In finite dimensional Hartree-Fock-Bogoliubov theory things are less clear and we will not discuss the problem of binding. In our simulations we have observed that the HFB ground state density was always rather close to the HF ground state density, which suggests that there is binding in HFB as well.

We have made some calculations for $N_b = 50$ and $N = 4$. We found that the critical strength is about $a_c \simeq 23.5$ in the proton-proton case and $a_c \simeq 17.5$ in the neutron-neutron case. In Figure 11 we display the HF and HFB energies as functions of the parameter a, for $N = 4$ and $\kappa = 1$ (proton-proton case). Figure 12 is the equivalent result for $\kappa = 0$ (neutron-neutron case). For these calculations we have chosen $r_{\max} = 3$ which is the optimal choice for a in a neighborhood of the critical value. Like in the previous section the results depend on the radius of the ball in which the system is confined. We see that there is always pairing, in the sense that the HFB curve is below the HF curve. This is even more manifest in the neutron-neutron case for which the potential is much more attractive than for protons, which repel with the Coulomb potential. Also, the norm of the pairing matrix A does not vary too much with a, it stays between 0.80 and 0.95 for a in the range $15 \leqslant a \leqslant 30$, for both $\kappa = 0$ and $\kappa = 1$.

From these numerical results we can conclude that pairing seems to happen in this model, for any strength a for which there is binding in Hartree-Fock theory. It is an interesting

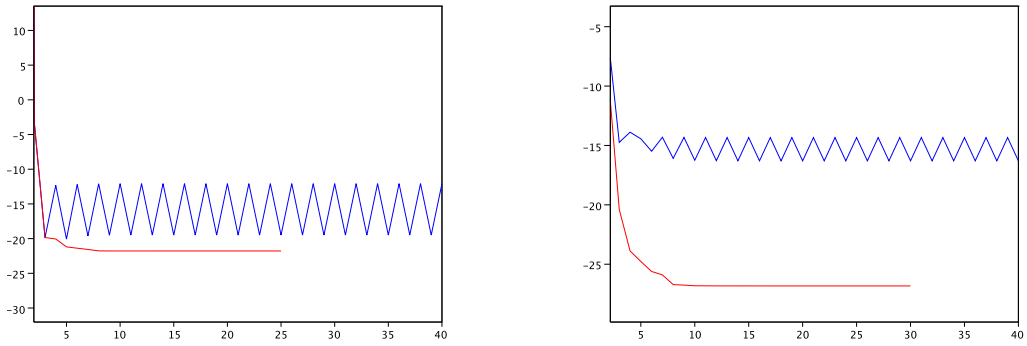

Fig. 8. Energy along the iterations in the HF (left) and HFB (right) cases, for the Roothaan (blue) and the ODA (red), with $N = 4$, $N_b = 20$, $\ell_{\max} = 1$, $r_{\max} = 3$, $a = 35$ and $\kappa = 1$ (proton-proton case).

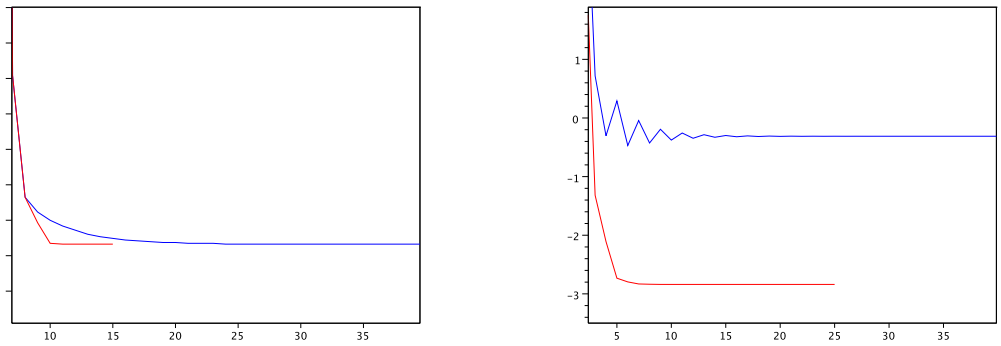

Fig. 9. Same calculation with $a = 20$ and $\kappa = 1$ (proton-proton case). The Roothaan algorithm slowly converges in the HF case and it oscillates in the HFB case but the two values are very close.

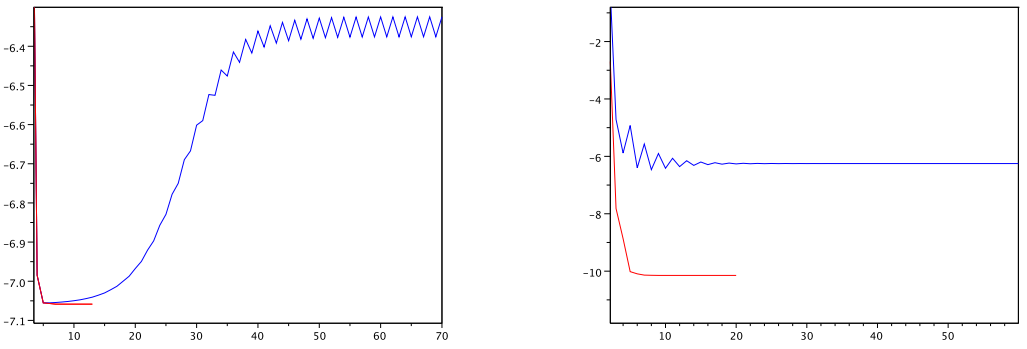

Fig. 10. Same calculation with $a = 20$ and $\kappa = 0$ (neutron-neutron case). The Roothaan algorithm oscillates in the HFB case, but the two values are very close.

problem to actually prove that pairing always occurs, for instance for a large enough. We are not aware of any result of this kind.

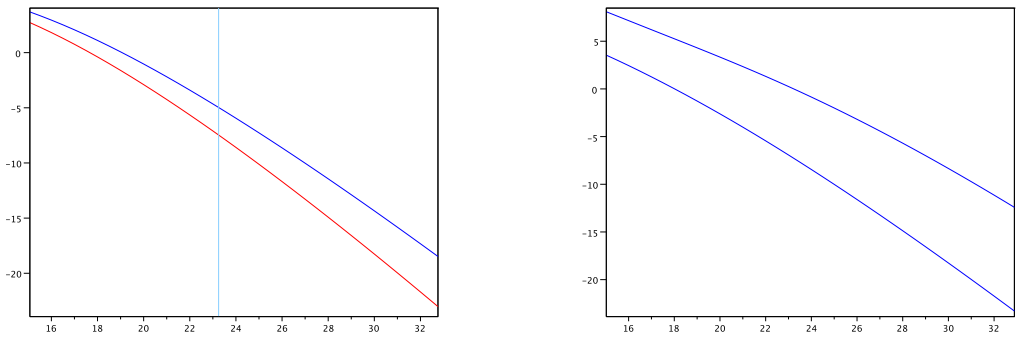

Fig. 11. Left: Values of the HF (blue) and HFB (red) ground state energies as functions of a, with $N = 4$, $N_b = 50$, $\ell_{\max} = 1$, $r_{\max} = 3$ and $\kappa = 1$ (proton-proton case). The vertical line is the value of a for which the last filled eigenvalue vanishes. Right: Values of the two filled HF eigenvalues for the same a.

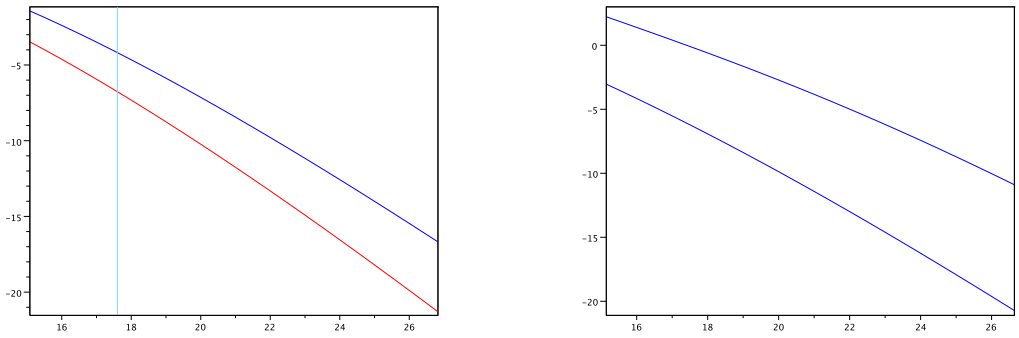

Fig. 12. Same calculations for $\kappa = 0$ (neutron-neutron case).

References

1. V. Bach, *Error bound for the Hartree-Fock energy of atoms and molecules*, Commun. Math. Phys., 147 (1992), pp. 527–548.
2. V. Bach, J. Fröhlich, and L. Jonsson, *Bogolubov-Hartree-Fock mean field theory for neutron stars and other systems with attractive interactions*, J. Math. Phys., 50 (2009), pp. 102102, 22.
3. V. Bach, E. H. Lieb, and J. P. Solovej, *Generalized Hartree-Fock theory and the Hubbard model*, J. Statist. Phys., 76 (1994), pp. 3–89.
4. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Theory of superconductivity*, Phys. Rev., 108 (1957), pp. 1175–1204.
5. P. Billard and G. Fano, *An existence proof for the gap equation in the superconductivity theory*, Commun. Math. Phys., 10 (1968), pp. 274–279.
6. N. N. Bogoliubov, *About the theory of superfluidity*, Izv. Akad. Nauk SSSR, 11 (1947), p. 77.

7. ———, *Energy levels of the imprefect Bose gas*, Bull. Moscow State Univ., 7 (1947), p. 43.
8. ———, *On the theory of superfluidity*, J. Phys. (USSR), 11 (1947), p. 23.
9. N. N. Bogoliubov, *On a New Method in the Theory of Superconductivity*, J. Exp. Theor. Phys., 34 (1958), p. 58.
10. É. Cancès, *SCF algorithms for HF electronic calculations*, in Mathematical models and methods for ab initio quantum chemistry, vol. 74 of Lecture Notes in Chem., Springer, Berlin, 2000, ch. 2, pp. 17–43.
11. É. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday, *Computational quantum chemistry: a primer*, in Handbook of numerical analysis, Vol. X, Handb. Numer. Anal., X, North-Holland, Amsterdam, 2003, pp. 3–270.
12. É. Cancès and C. Le Bris, *Can we outperform the DIIS approach for electronic structure calculations?*, Int. J. Quantum Chem., 79 (2000), pp. 82–90.
13. ———, *On the convergence of SCF algorithms for the Hartree-Fock equations*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 749–774.
14. É. Cancès, C. Le Bris, and Y. Maday, *Méthodes mathématiques en chimie quantique. Une introduction*, vol. 53 of Collection Mathématiques et Applications, Springer, 2006.
15. E. Davies, *Spectral theory and differential operators*, vol. 42 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1995.
16. J. Dechargé and D. Gogny, *Hartree-Fock-Bogolyubov calculations with the D1 effective interaction on spherical nuclei*, Phys. Rev. C, 21 (1980), pp. 1568–1593.
17. C. Fefferman and R. de la Llave, *Relativistic stability of matter. I*, Rev. Mat. Iberoamericana, 2 (1986), pp. 119–213.
18. R. L. Frank, C. Hainzl, S. Naboko, and R. Seiringer, *The critical temperature for the BCS equation at weak coupling*, J. Geom. Anal., 17 (2007), pp. 559–567.
19. G. Friesecke, *The multiconfiguration equations for atoms and molecules: charge quantization and existence of solutions*, Arch. Ration. Mech. Anal., 169 (2003), pp. 35–71.
20. D. Gogny, in Proceedings of the International Conference on Nuclear Physics, J. de Boer and H. Mang, eds., 1973, p. 48.
21. ———, in Proceedings of the International Conference on Nuclear Self-Consistent Fields, M. Porneuf and G. Ripka, eds., Trieste, 1975, p. 333.
22. D. Gogny and P.-L. Lions, *Hartree-Fock theory in nuclear physics*, RAIRO Modél. Math. Anal. Numér., 20 (1986), pp. 571—637.
23. C. Hainzl, E. Hamza, R. Seiringer, and J. P. Solovej, *The BCS functional for general pair interactions*, Comm. Math. Phys., 281 (2008), pp. 349–367.
24. C. Hainzl, E. Lenzmann, M. Lewin, and B. Schlein, *On blowup for time-dependent generalized Hartree-Fock equations*, Ann. Henri Poincaré, 11 (2010), pp. 1023–1052.
25. C. Hainzl and R. Seiringer, *General decomposition of radial functions on $\mathbb{R}^n$ and applications to N-body quantum systems*, Lett. Math. Phys., 61 (2002), pp. 75–84.
26. ———, *The BCS critical temperature for potentials with negative scattering length*, Lett. Math. Phys., 84 (2008), pp. 99–107.
27. M. Hoffmann-Ostenhof and T. Hoffmann-Ostenhof, *Schrödinger inequalities and asymptotic behavior of the electron density of atoms and molecules*, Phys. Rev. A, 16 (1977), pp. 1782–1785.
28. T. Kato, *Perturbation theory for linear operators*, Springer, second ed., 1995.
29. C. Le Bris, *Computational chemistry from the perspective of numerical analysis*, Acta Numerica, 14 (2005), pp. 363–444.
30. E. Lenzmann and M. Lewin, *Minimizers for the Hartree-Fock-Bogoliubov theory of neutron stars and white dwarfs*, Duke Math. J., 152 (2010), pp. 257–315.
31. A. Levitt, *Convergence of gradient-based algorithms for the Hartree-Fock equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 46 (2012), pp. 1321–1336.
32. M. Lewin, *Geometric methods for nonlinear many-body quantum systems*, J. Funct. Anal., 260 (2011), pp. 3535–3595.
33. E. H. Lieb, *Variational principle for many-fermion systems*, Phys. Rev. Lett., 46 (1981), pp. 457–459.

34. E. H. Lieb and R. Seiringer, *The Stability of Matter in Quantum Mechanics*, Cambridge Univ. Press, 2010.
35. E. H. Lieb and B. Simon, *The Hartree-Fock theory for Coulomb systems*, Commun. Math. Phys., 53 (1977), pp. 185–194.
36. E. H. Lieb and W. E. Thirring, *Gravitational collapse in quantum mechanics with relativistic kinetic energy*, Ann. Physics, 155 (1984), pp. 494–512.
37. E. H. Lieb and H.-T. Yau, *The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics*, Commun. Math. Phys., 112 (1987), pp. 147–174.
38. P.-L. Lions, *Solutions of Hartree-Fock equations for Coulomb systems*, Commun. Math. Phys., 109 (1987), pp. 33–97.
39. J. B. McLeod and Y. Yang, *The uniqueness and approximation of a positive solution of the Bardeen-Cooper-Schrieffer gap equation*, J. Math. Phys., 41 (2000), pp. 6007–6025.
40. P. Quentin and H. Flocard, *Self-Consistent Calculations of Nuclear Properties with Phenomenological Effective Forces*, Ann. Rev. Nucl. Part. Sci., 28 (1978), pp. 523–594.
41. P. Ring and P. Schuck, *The nuclear many-body problem*, vol. Texts and Monographs in Physics, Springer Verlag, New York, 1980.
42. C. C. J. Roothaan, *New developments in molecular orbital theory*, Rev. Mod. Phys., 23 (1951), pp. 69–89.
43. Scilab Consortium, *Scilab: The free software for numerical computation*, Scilab Consortium, Digiteo, Paris, France, 2011.
44. B. Simon, *Geometric methods in multiparticle quantum systems*, Commun. Math. Phys., 55 (1977), pp. 259–274.
45. T. Skyrme, *The effective nuclear potential*, Nuclear Physics, 9 (1959), pp. 615–634.
46. J. P. Solovej, *Proof of the ionization conjecture in a reduced Hartree-Fock model.*, Invent. Math., 104 (1991), pp. 291–311.
47. ———, *The ionization conjecture in Hartree-Fock theory*, Ann. of Math. (2), 158 (2003), pp. 509–576.
48. A. Vansevenant, *The gap equation in superconductivity theory*, Phys. D, 17 (1985), pp. 339–344.
49. Y. S. Yang, *On the Bardeen-Cooper-Schrieffer integral equation in the theory of superconductivity*, Lett. Math. Phys., 22 (1991), pp. 27–37.