Fully guaranteed and computable error bounds on the energy for periodic Kohn-Sham equations with convex density functionals

Andrea Bordignon, Éric Cancès, Geneviève Dusson, Gaspard Kemlin, Rafael Antonio Lainez Reyes, Benjamin Stamm

FULLY GUARANTEED AND COMPUTABLE ERROR BOUNDS ON THE ENERGY FOR PERIODIC KOHN–SHAM EQUATIONS WITH CONVEX DENSITY FUNCTIONALS *

ANDREA BORDIGNON *, ERIC CANCÈS *, GENEVIÈVE DUSSON †, GASPARD KEMLIN ‡, RAFAEL ANTONIO LAINEZ REYES §, AND BENJAMIN STAMM §

Abstract. In this article, we derive fully guaranteed error bounds for the energy of convex nonlinear mean-field models. These results apply in particular to Kohn–Sham equations with convex density functionals, which includes the reduced Hartree–Fock (rHF) model, as well as the Kohn-Sham model with exact exchange-density functional (which is unfortunately not explicit and therefore not usable in practice). We then decompose the obtained bounds into two parts, one depending on the chosen discretization and one depending on the number of iterations performed in the self-consistent algorithm used to solve the nonlinear eigenvalue problem, paving the way for adaptive refinement strategies. The accuracy of the bounds is demonstrated on a series of test cases, including a Silicon crystal and an Hydrogen Fluoride molecule simulated with the rHF model and discretized with planewaves. We also show that, although not anymore guaranteed, the error bounds remain very accurate for a Silicon crystal simulated with the Kohn–Sham model using nonconvex exchange-correlation functionals of practical interest.

1. Introduction. Ab initio simulations within the Born–Oppenheimer approximation [2] are performed routinely for simulating molecular and materials systems in several fields including condensed matter physics, chemistry and materials science. The fact that the equations at stake do not depend on empirical parameters except a few fundamental constants of physics make them very appealing for systematic and accurate simulations. A typical problem in this field, on which we will focus in this article, is the problem of finding the electronic ground state of the considered system, that is the state of lowest energy for the electrons, the nuclei (considered as point-like particles) being fixed at some given positions. Among many models used in ab initio simulations, Density Functional theory (DFT) and especially Kohn–Sham models [36] are the most popular ones, as they offer a good compromise between accuracy and computational cost. From a mathematical perspective, the Kohn–Sham models consist of a partial differential equation in the form of a nonlinear eigenvalue problem, as the corresponding differential operator, that needs to be diagonalized, depends on the eigenfunctions themselves.

[1]Université Paris Est, CERMICS, Ecole des Ponts and INRIA, 6 & 8 Av. Pascal, 77455 Marne-la-Vallée, France, bordignon@cermics.enpc.fr, cances@cermics.enpc.fr.

[2]Université de Franche-Comté, CNRS, LmB, 25000 Besançon, France, genevieve.dusson@math.cnrs.fr.

[3]LAMFA, UMR CNRS 7352, Université de Picardie Jules Verne, 80039 Amiens, France, gaspard.kemlin@u-picardie.fr.

[4]IANS-NMH, University of Stuttgart, 70569 Stuttgart, Germany, rafael-antonio.lainez-reyes@ians.uni-stuttgart.de, benjamin.stamm@ians.uni-stuttgart.de.

*This publication is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme – Grant Agreement EMC2 No. 810367. GD was supported by the French 'Investissements d'Avenir' program, project Agence Nationale de la Recherche (ISITE-BFC) (contract ANR-15-IDEX-0003). GD was also supported by the Ecole des Ponts-ParisTech and region Bourgogne Franche-Comté. RL and BS acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project 516782692. We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for supporting this work by funding - EXC2075 – 390740016 under Germany's Excellence Strategy. We acknowledge the support by the Stuttgart Center for Simulation Science (SimTech).

The numerical computations of solutions to the Kohn–Sham equations requires first to choose a discretization space, that is a finite-dimensional space on which the eigenvalue equations are projected and then solved. Standard discretization methods for DFT equations include planewaves, linear combination of atomic orbitals (LCAO) or finite elements. Once the discretization space is chosen, the nonlinear equations have to be solved, usually with a fixed-point algorithm called self-consistent field (SCF) algorithm, which is stopped when a given tolerance criterium is met. Finally, in the SCF algorithm, a linear matrix eigenvalue problem is solved at each iteration, often resorting to an iterative linear algebra solver also stopped after a finite number of iterations. Each of these steps leads to approximations, and it is important to estimate them for two reasons. First, obtaining guaranteed bounds for the error between the exact and computed solutions allows to certify the accuracy of the computed solutions. Second, quantifying the size of the error coming from each approximation allows to optimize the parameters of the simulation in order to reduce the global computational cost necessary to reach a final total accuracy.

The purpose of this article is to derive fully guaranteed and computable *a posteriori* energy bounds for DFT models with convex density functionals using a planewave discretization, as well as to propose a decomposition of the total error between two contributions: one standing for the nonlinear solver error and the other for the discretization error. In this first contribution, we do not (yet) account for the error coming from the iterative eigenvalue solver of the successive (linear) matrix eigenvalue problems. Note that our results could be applied in theory to the Kohn-Sham model with exact exchange-density functional, since the exact density functional is convex (in the Vallone-Lieb version of DFT [57, 40]). However, as the exact exchange-correlation functional has no known explicit expression, our bounds are useless in practice for the exact density functional. On the other hand, they can be applied to the reduced Hartree–Fock (rHF) model, i.e. to the Kohn-Sham model with exchange-correlation terms set to zero. In addition, they can be used in practice for usual density-functional approximations, namely LDA, GGA, etc. (see e.g. [55] for a review on currently available approximate exchange-correlation functionals). Although they are no longer guaranteed, numerical simulations indicate that our bounds give a fairly good approximation of the actual error.

Let us now put our work into perspective. The *a posteriori* error estimation of elliptic boundary value problems is already well developed, see e.g. [51, 3, 20, 25, 37, 59]. Regarding eigenvalue problems, the computation of guaranteed error bounds were first proposed by Kato [34], Forsythe [26], Weinberger [58] or Bazley and Fox [1], and more recently *e.g.* in [9, 10, 18, 21, 32, 33, 38, 42, 43], see also [47] for a recent monograph on the subject, and [24] for a recent review. These error bounds are however only valid for single eigenvalues, while in practice, degenerate or near-degenerate eigenvalues often appear in electronic structure calculations. *A posteriori* error estimates for conforming approximations of eigenvalue clusters of second-order self-adjoint elliptic operators with compact resolvent have more recently been proposed in [11], as well as in the context of electronic structure calculations for linear eigenvalues equations in [30]. When it comes to nonlinear eigenvalue equations as the ones of interest in this article, no guaranteed error bounds have been published to our knowledge, apart from the plane-wave discretization of a toy Gross–Pitaevskii equation in one-dimension [23]. However, asymptotic error bounds can be found in *e.g.* [44] for the Hartree–Fock equations. For Kohn–Sham DFT, practical, not guaranteed, error bounds on the

energy and the density matrix have been proposed for various discretization methods, notably planewave [8, 22], and finite elements [19, 41, 46, 60, 61]. Let us also mention the recent work [7], which includes practical bounds on quantities of interest other than the energy and the density matrix, such as the interatomic forces.

The rest of the article is organized as follows. In Section 2, we present the model of interest together with the discretization of the equations and the self-consistent field (SCF) algorithm used to solve the discretized equations in practice. We then derive the proposed guaranteed bounds on the considered energy in Section 3. Finally, we present numerical results on a set of test systems in Section 4, demonstrating the accuracy of the error bound on the energy for convex density functionals and showing that the bounds also provide very accurate error estimates in the case of nonconvex approximate density functionals used in practice.

2. Problem setting, discretization, and practical resolution. In this section we introduce the model of interest using an abstract framework, similar to [7, 14]. Our description, formulated using a *density matrix formalism*, allows us to cover different models such as the (spinless) Kohn–Sham models, the Hartree–Fock model, or the stationary Gross–Pitaevskii equation.

2.1. Functional setting. Let $\Omega \subset \mathbb{R}^3$ denote the unit cell of an arbitrary periodic lattice $\mathcal{R}$, with dual lattice $\mathcal{R}^*$. For $p > 0$, we denote by $L^p_{\mathrm{per}}(\Omega)$ the space of complex-valued, p-integrable $\mathcal{R}$-periodic functions

$$L^p_{\mathrm{per}}(\Omega) = \{u \in L^p_{\mathrm{loc}}(\mathbb{R}^3;\mathbb{C}) : u \text{ is } \mathcal{R}\text{-periodic}\}. \tag{2.1}$$

In particular, the space $\mathcal{H} \coloneqq L^2_{\mathrm{per}}(\Omega)$ with inner product $\langle\cdot,\cdot\rangle$ admits an orthonormal basis consisting of planewaves:

$$e_{\boldsymbol{G}} : \boldsymbol{x} \in \mathbb{R}^3 \mapsto |\Omega|^{-\frac{1}{2}} e^{\imath \boldsymbol{G}\cdot\boldsymbol{x}},\ \boldsymbol{G} \in \mathcal{R}^*. \tag{2.2}$$

For $s \in \mathbb{R}$, the $\mathcal{R}$-periodic Sobolev space of order s is then defined as

$$H^s_{\mathrm{per}}(\Omega) \coloneqq \left\{ u(\boldsymbol{x}) = \sum_{\boldsymbol{G}\in\mathcal{R}^*} \hat{u}_{\boldsymbol{G}} e_{\boldsymbol{G}}(\boldsymbol{x}) : \sum_{\boldsymbol{G}\in\mathcal{R}^*} \left(1 + \frac{|\boldsymbol{G}|^2}{2}\right)^s |\hat{u}_{\boldsymbol{G}}|^2 < \infty \right\}, \tag{2.3}$$

where $\hat{u}_{\boldsymbol{G}}$ represents the Fourier coefficient

$$\hat{u}_{\boldsymbol{G}} = \langle e_{\boldsymbol{G}}, u\rangle = |\Omega|^{-\frac{1}{2}} \int_\Omega u(\boldsymbol{x}) e^{-\imath \boldsymbol{G}\cdot\boldsymbol{x}} \mathrm{d}\boldsymbol{x}. \tag{2.4}$$

Endowed with the inner product

$$\langle u, v\rangle_{H^s_{\mathrm{per}}(\Omega)} \coloneqq \sum_{\boldsymbol{G}\in\mathcal{R}^*} \left(1 + \frac{|\boldsymbol{G}|^2}{2}\right)^s \hat{u}^*_{\boldsymbol{G}} \hat{v}_{\boldsymbol{G}}, \quad \forall u, v \in H^s_{\mathrm{per}}(\Omega), \tag{2.5}$$

the space $H^s_{\mathrm{per}}(\Omega)$ is a Hilbert space. We denote by $L^p_{\mathrm{per}}(\Omega;\mathbb{R})$ and $H^s_{\mathrm{per}}(\Omega;\mathbb{R})$ the spaces of real-valued functions in $L^p_{\mathrm{per}}(\Omega)$ and $H^s_{\mathrm{per}}(\Omega)$ respectively.

2.2. Traces of operators and density matrices. Let us now recall the definition of the trace of an operator. Suppose A is a bounded, positive linear operator on $\mathcal{H}$, then the *trace* of A is equal to

$$\mathrm{Tr}(A) \coloneqq \sum_{i\in\mathbb{N}} \langle \psi_i, A\psi_i\rangle \ \in \mathbb{R}_+ \cup \{+\infty\}, \tag{2.6}$$

where $(\psi_i)_{i\in\mathbb{N}}$ is any orthonormal basis of $\mathcal{H}$. A bounded linear operator is *trace class* if $\mathrm{Tr}(|A|) < +\infty$ where $|A| := (A^*A)^{1/2}$. For more details on trace-class operators and the underlying functional analysis setting, we refer to [53], [11, Section 2.2], and references therein.

More generally, let us explain how to define $\mathrm{Tr}(AB)$, in the case where (i) A is a linear, self-adjoint, bounded from below, operator with domain $\mathfrak{D}(A)$ and form domain $\mathfrak{Q}(A)$, and (ii) B is a finite-rank operator such that $\mathrm{Ker}(B)^\perp \subset \mathfrak{Q}(A), \mathrm{Ran}(B) \subset \mathfrak{Q}(A)$. First we write B in canonical form, that is

$$B = \sum_{i=1}^{r} \sigma_i |\phi_i\rangle\langle\psi_i|, \tag{2.7}$$

where $r \in \mathbb{N}$ is the rank of B, $0 < \sigma_r \leq \cdots \leq \sigma_1$ the singular values of B, $\phi_i \in \mathfrak{Q}(A)$, $\psi_i \in \mathfrak{Q}(A)$, and $(\phi_i)_{1\leq i\leq r}$ and $(\psi_i)_{1\leq i\leq r}$ are orthonormal with respect to the $L^2_{\mathrm{per}}(\Omega)$-inner product. Then we define

$$\mathrm{Tr}(AB) \coloneqq \sum_{i=1}^{r} \sigma_i \langle\psi_i, A\phi_i\rangle. \tag{2.8}$$

It is easy to check that this definition does not depend on the chosen canonical decomposition of B and obviously coincides with the usual trace when A is a bounded operator.

In this work, we will focus on the set $\mathcal{M}$ of finite-energy *density matrices*, defined as the set of all rank-N_{el} orthogonal projectors on $\mathcal{H} = L^2_{\mathrm{per}}(\Omega)$ with range in $H^1_{\mathrm{per}}(\Omega)$, *i.e.*

$$\mathcal{M} \coloneqq \Big\{\gamma \in \mathcal{S}(\mathcal{H}),\ \gamma^2 = \gamma,\ \mathrm{Tr}(\gamma) = N_{\mathrm{el}},\ \mathrm{Tr}(-\Delta\gamma) < \infty\Big\}, \tag{2.9}$$

where $\mathcal{S}(\mathcal{H})$ is the set of all bounded self-adjoint operators over $\mathcal{H}$, and N_{el} is a fixed integer, whose physical meaning will be clarified in the following section.
For the sake of clarity, let us rewrite the term $\mathrm{Tr}(-\Delta\gamma)$ in (2.9) in a more explicit form. Since $\gamma \in \mathcal{M}$, there exists an $L^2_{\mathrm{per}}(\Omega)$-orthonormal basis $\Phi = (\phi_i)_{1\leq i\leq N_{\mathrm{el}}}$ of $\mathrm{Ran}(\gamma) \subset \mathfrak{Q}(-\Delta) = H^1_{\mathrm{per}}(\Omega)$ such that, using Dirac bra-ket notation, the projector γ can be written as

$$\gamma = \gamma_\Phi = \sum_{i=1}^{N_{\mathrm{el}}} |\phi_i\rangle\langle\phi_i|, \tag{2.10}$$

and using (2.8), we can write

$$\mathrm{Tr}(-\Delta\gamma) = \sum_{i=1}^{N_{\mathrm{el}}} \int_\Omega |\nabla\phi_i|^2 < +\infty. \tag{2.11}$$

2.3. Problem formulation. From now on the constant N_{el} appearing in definition (2.9) will represent the number of electron in the system. Now consider $\gamma \in \mathcal{M}$, written as in (2.10) for some $\Phi = (\phi_i)_{1\leq i\leq N_{\mathrm{el}}}$ of $\mathrm{Ran}(\gamma)$. We can associate to γ its *kernel*, still denoted by γ

$$\gamma(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{N_{\mathrm{el}}} \phi_i(\boldsymbol{x})\phi_i^*(\boldsymbol{x}'), \tag{2.12}$$

and define the electronic density $\rho_\gamma \in L^1_{\mathrm{per}}(\Omega;\mathbb{R})$ as

$$\rho_\gamma(\boldsymbol{x}) \coloneqq \gamma(\boldsymbol{x},\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{el}}} |\phi_i(\boldsymbol{x})|^2 =: \rho_\Phi(\boldsymbol{x}). \tag{2.13}$$

We would like to remark that as a consequence of $\phi_i \in H^1_{\mathrm{per}}(\Omega)$ for $i = 1,\dots,N_{\mathrm{el}}$, ρ_γ actually belongs to $L^3_{\mathrm{per}}(\Omega;\mathbb{R})$ hence in $L^2_{\mathrm{per}}(\Omega;\mathbb{R})$ since Ω is bounded. Finally we will consider a real-valued functional E taking values over $\mathcal{M}$, which we will refer to as the *energy functional*. When looking for the ground state of the system, we seek the density matrix $\gamma \in \mathcal{M}$ that minimizes E:

$$\min\{E(\gamma),\ \gamma \in \mathcal{M}\}. \tag{2.14}$$

In the models studied in this work (see Section 2.4), the energy E can be expressed as the sum of a linear term and a nonlinear term:

$$E(\gamma) \coloneqq \mathrm{Tr}(h\gamma) + F(\rho_\gamma). \tag{2.15}$$

Here, h is some bounded-from-below self-adjoint operator on $\mathcal{H}$ with domain $H^2_{\mathrm{per}}(\Omega)$ and form domain $H^1_{\mathrm{per}}(\Omega)$, and F is a function that explicitly depends on the electronic density ρ_γ of the density matrix $\gamma \in \mathcal{M}$. We assume in the following that $F : L^2_{\mathrm{per}}(\Omega;\mathbb{R}) \to \mathbb{R}$ is continuously differentiable. Under these conditions, problem (2.14) admits a set of Euler–Lagrange equations, which we proceed to derive.

First, note that the assumptions on F imply that for each $\gamma \in \mathcal{M}$ there exists $V_{\rho_\gamma} \in L^2_{\mathrm{per}}(\Omega;\mathbb{R})$ such that, for any $\widetilde{\gamma} \in \mathcal{M}$:

$$\langle F'(\rho_\gamma), \rho_{\widetilde{\gamma}}\rangle = \int_\Omega V_{\rho_\gamma}\rho_{\widetilde{\gamma}}. \tag{2.16}$$

With a slight abuse of notation, as in (2.8), we denote

$$\mathrm{Tr}(V_{\rho_\gamma}\widetilde{\gamma}) \coloneqq \int_\Omega V_{\rho_\gamma}\rho_{\widetilde{\gamma}}. \tag{2.17}$$

Now, if $\gamma \in \mathcal{M}$ is a solution to problem (2.14) the first-order variation of E at γ reads:

$$\forall\ \zeta \in T_\gamma\mathcal{M}, \quad \langle E'(\gamma), \zeta\rangle = 0,$$

where $T_\gamma\mathcal{M}$, the tangent space at γ, is defined (see [16]) by

$$T_\gamma\mathcal{M} \coloneqq \{\zeta \in \mathcal{S}(\mathcal{H}),\ \gamma\zeta + \zeta\gamma = \zeta,\ \mathrm{Tr}(\zeta) = 0,\ \mathrm{Tr}(-\Delta\zeta) < \infty\}. \tag{2.18}$$

As before $\rho_\zeta \in L^2_{\mathrm{per}}(\Omega;\mathbb{R})$, and as a consequence

$$\forall\ \zeta \in T_\gamma\mathcal{M}, \quad 0 = \langle E'(\gamma), \zeta\rangle = \mathrm{Tr}((h + V_{\rho_\gamma})\zeta) = \mathrm{Tr}(H_{\rho_\gamma}\zeta), \tag{2.19}$$

where, as a consequence of Kato-Rellich theorem [52, Theorem X.12] and the assumptions made on h, the operator $H_{\rho_\gamma} \coloneqq h + V_{\rho_\gamma}$ is self-adjoint on $\mathcal{H}$ with domain $H^2_{\mathrm{per}}(\Omega)$ and form domain $H^1_{\mathrm{per}}(\Omega)$, and bounded from below.

Choosing $\zeta = |\phi\rangle\langle\psi| + |\psi\rangle\langle\phi| \in T_\gamma\mathcal{M}$ with $\phi \in \mathrm{Ran}(\gamma)$ and $\psi \in \mathrm{Ker}(\gamma)$ (see [16, Lemma 3]) we finally obtain the Euler–Lagrange equations corresponding to problem

(2.14): find eigenvectors $\Phi = (\phi_i)_{1\le i\le N_{\rm el}} \in (H^1_{\rm per}(\Omega))^{N_{\rm el}}$ of both H_{ρ_γ} and γ, and eigenvalues $\Lambda = (\lambda_i)_{1\le i\le N_{\rm el}} \in \mathbb{R}^{N_{\rm el}}$ of H_{ρ_γ} such that

$$\begin{cases} H_{\rho_\gamma}\phi_i = \lambda_i\phi_i, \quad i = 1,\dots,N_{\rm el}, \\ \langle\phi_i,\phi_j\rangle = \delta_{ij}, \quad i,j = 1,\dots,N_{\rm el}, \\ \gamma = \displaystyle\sum_{i=1}^{N_{\rm el}} |\phi_i\rangle\langle\phi_i|. \end{cases} \tag{2.20}$$

This is a *nonlinear eigenvector problem*: the Hamiltonian H_{ρ_γ} we seek to diagonalize depends on its own eigenvectors through the electronic density ρ_γ.

Remark 2.1 (*Aufbau* principe). We will assume in the following that for any minimizer $\gamma_\star$ of (2.14),

1. there is a positive gap between the $N_{\rm el}$-th and $(N_{\rm el}+1)$-st eigenvalues of the mean-field operator $H_{\rho_{\gamma_\star}}$ (counting multiplicities);
2. that $\gamma_\star$ is obtained from the eigenvectors associated to the lowest $N_{\rm el}$ eigenvalues of (2.20) (*Aufbau* principle).

Note that for the reduced Hartree–Fock (rHF) model in Ω mentioned above, the function F is in fact strictly convex, which implies that the minimizers of E over the convex hull $\mathrm{CH}(\mathcal{M})$ of $\mathcal{M}$ all share the same density $\rho_\star$. It can then be shown (see e.g. [5, 54]) that if $H_{\rho_\star}$ is gapped, then the minimizer of E on $\mathrm{CH}(\mathcal{M})$ is unique, belongs to $\mathcal{M}$, and is the spectral projector on the lowest $N_{\rm el}$ eigenvalues of $H_{\rho_\star}$.

As a consequence, under reasonable assumptions (see *e.g.* [6] for the reduced Hartree–Fock model), the solution γ to the minimization problem (2.14) is unique, from where it follows that the eigenvectors, up to unitary transformations, are unique. This assumption is also crucial to use the results for eigenvalues clusters of self-adjoint operators presented in [11] in order to derive *a posteriori* error estimates.

2.4. Kohn–Sham DFT equations. The periodic Kohn–Sham equations are commonly used in numerical simulations of condensed matter systems. We now proceed to give a description of this model. Let $\Phi \in (H^1_{\rm per}(\Omega))^{N_{\rm el}}$ be a set of L^2-orthonormal orbitals on Ω. Denoting the electronic density $\rho_\Phi = \sum_{i=1}^{N_{\rm el}} |\phi_i|^2$, the Kohn–Sham energy is defined as

$$\mathcal{E}(\Phi) \coloneqq \frac{1}{2}\sum_{i=1}^{N_{\rm el}} \int_\Omega |\nabla\phi_i|^2 + \int_\Omega V\rho_\Phi + \frac{1}{2}\mathcal{D}(\rho_\Phi,\rho_\Phi) + E_{\rm xc}(\rho_\Phi), \tag{2.21}$$

where the first term represents the part of the kinetic energy corresponding to non-interacting electrons, the external potential $V \in L^2_{\rm per}(\Omega;\mathbb{R})$ represents the interaction between the nuclei and the electrons, while the last two terms describe the interaction between the electrons. In particular, $E_{\rm xc}(\rho_\Phi)$ denotes the exchange-correlation energy while $\mathcal{D}(\rho_1,\rho_2)$ stands for the Coulomb interaction-energy per unit cell

$$\mathcal{D}(\rho_1,\rho_2) = \int_\Omega V_{\rm H}[\rho_1](\boldsymbol{x})\rho_2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{2.22}$$

where the Hartree potential $V_{\rm H}[\rho]$ is the unique zero-mean solution in $L^2_{\rm per}(\Omega;\mathbb{R})$ to the periodic Poisson equation

$$-\Delta V_{\rm H}[\rho] = 4\pi\left(\rho - \frac{1}{|\Omega|}\int_\Omega \rho\right), \tag{2.23}$$

on Ω.

In order to frame this minimization problem in the context of (2.15), recall that $\gamma_\Phi = \sum_{i=1}^{N_{\rm el}} |\phi_i\rangle\langle\phi_i|$ denotes the density matrix corresponding to the orbitals Φ, so we rewrite (2.21) in terms of the generalized trace (2.8) as

$$\mathcal{E}(\Phi) = \mathrm{Tr}\left(\Big(-\frac{1}{2}\Delta + V\Big)\gamma_\Phi\right) + \frac{1}{2}\mathcal{D}(\rho_\Phi, \rho_\Phi) + E_{\rm xc}(\rho_\Phi) =: E(\gamma_\Phi). \tag{2.24}$$

Thus, in the density matrix formalism, the Kohn–Sham minimization problem reads as in (2.14)-(2.15) with the core Hamiltonian

$$h = -\frac{1}{2}\Delta + V \tag{2.25}$$

and density functional

$$F(\rho) = \frac{1}{2}\mathcal{D}(\rho, \rho) + E_{\rm xc}(\rho). \tag{2.26}$$

Moreover, the corresponding Kohn–Sham Hamiltonian appearing in the Euler–Lagrange equations (2.20) takes the form

$$H_{\rho_\gamma} = -\frac{1}{2}\Delta + V + V_{\rho_\gamma}, \tag{2.27}$$

where $V_\rho = V_{\rm H}[\rho] + V_{\rm xc}[\rho]$ and $V_{\rm xc}[\rho] = \frac{\mathrm{d}E_{\rm xc}(\rho)}{\mathrm{d}\rho}$ is the exchange-correlation potential.

Remark 2.2 (Reduced Hartree–Fock model). In Section 4, we will provide numerical simulations on the restricted Hartree–Fock model, which amounts to choosing $E_{\rm xc} \equiv 0$ in (2.21).

Remark 2.3 (Spins). To better fit the geometrical framework allowed by the density matrices, the Kohn–Sham model is presented for systems of "spinless" electrons. Real systems, as well as the numerical simulations performed at the end of this paper, include the spins. In case of systems with a positive band gap, which we consider here, everything works the same except that $N_{\rm el}$ represents the number electron *pairs*, the energy reads $E(\gamma) = 2\mathrm{Tr}(h\gamma) + F(\rho_\gamma)$ with $\rho_\gamma(\boldsymbol{x}) = 2\gamma(\boldsymbol{x}, \boldsymbol{x})$ and the Kohn–Sham Hamiltonian is defined as $H_{\rho_\gamma} = 2h + 2V_{\rho_\gamma}$.

Remark 2.4 (Brillouin zone discretization). Let us mention that the eigenproblems we presented are naturally equipped with periodic boundary conditions. However this introduces artificial interactions between the sample of the material in the unit cell Ω and its periodic images. In the case of a perfect crystal with Bravais lattice $\mathbb{L}$ and unit cell Ω, it is recommended to choose a periodic simulation (super)cell $\omega = L\Omega$ consisting of L^3 unit cells, so that $\mathcal{R} = L\mathbb{L}$. Using Bloch transform [53, Section XIII.16], the problem becomes in the thermodynamic limit $L \to +\infty$

$$\begin{cases} H_{\boldsymbol{k},\rho}\phi_{i,\boldsymbol{k}} = \lambda_{i,\boldsymbol{k}}\phi_{i,\boldsymbol{k}}, \quad i = 1, \dots, N_{\rm el}, \quad \boldsymbol{k} \in \mathcal{B}, \\ \langle \phi_{i,\boldsymbol{k}}, \phi_{j,\boldsymbol{k}} \rangle = \delta_{ij}, \quad i, j = 1, \dots, N_{\rm el}, \quad \boldsymbol{k} \in \mathcal{B}, \\ \rho(\boldsymbol{x}) = \fint_{\mathcal{B}} \sum_{i=1}^{N_{\rm el}} |\phi_{i,\boldsymbol{k}}(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{k}, \end{cases} \tag{2.28}$$

where $\mathcal{B}$ is the first Brillouin zone of the crystal and $H_{\boldsymbol{k},\rho}$ are the Bloch fibers of the Kohn–Sham Hamiltonian acting on $L^2_{\rm per}(\Omega)$, with domain $H^2_{\rm per}(\Omega)$. Finally, we

consider in practice a finite subset of $\boldsymbol{k}$-points in $\mathcal{B}$. The geometrical framework and the bounds derived in this paper then easily extend to the case of several $\boldsymbol{k}$-points in the Brillouin zone discretization, see Appendix A for details. We also refer for instance to [28] for the numerical analysis of the limit $L \to +\infty$ in the case of the reduced Hartree–Fock model.

2.5. Discretization. In order to approximate the solution to any problem in the framework of (2.14) by solving the eigenproblem (2.20), it must be first discretized in a finite-dimensional space. To this end, let N be a positive integer, $\mathcal{V}_N$ be a finite-dimensional subspace of $H^1_{\rm per}(\Omega)$ with $\dim(\mathcal{V}_N)$ depending on N such that the larger N, the better the Galerkin approximation of (2.14):

$$\min\left\{E(\gamma),\ \gamma \in \mathcal{M},\ \mathrm{Ran}(\gamma) \subset \mathcal{V}_N\right\}. \tag{2.29}$$

The corresponding first-order optimality conditions read: find eigenvectors $\Phi_N := (\phi_{i,N})_{1\le i\le N_{\rm el}} \in \mathcal{V}_N^{N_{\rm el}}$ and eigenvalues $\Lambda_N := (\lambda_{i,N})_{1\le i\le N_{\rm el}} \in \mathbb{R}^{N_{\rm el}}$ such that

$$\begin{cases} \left(\Pi_N H_{\rho_{\gamma_N}} \Pi_N\right) \phi_{i,N} = \lambda_{i,N}\phi_{i,N}, \quad i = 1, \dots N, \\ \langle \phi_{i,N}, \phi_{j,N}\rangle = \delta_{ij}, \quad i = 1, \dots, N_{\rm el}, \\ \gamma_N = \displaystyle\sum_{i=1}^{N_{\rm el}} |\phi_{i,N}\rangle\langle\phi_{i,N}|, \end{cases} \tag{2.30}$$

where Π_N denotes the orthogonal projector onto $\mathcal{V}_N$ for the inner product $\langle\cdot,\cdot\rangle$.

More precisely, given $\mathrm{E_{cut}} \in \mathbb{N}$, we define, for $N = \sqrt{2\mathrm{E_{cut}}}$,

$$\mathcal{V}_N = \mathrm{Span}\left(e_{\boldsymbol{G}} :\ |\boldsymbol{G}| \le N\right) = \mathrm{Span}\left(e_{\boldsymbol{G}} :\ \frac{1}{2}|\boldsymbol{G}|^2 \le \mathrm{E_{cut}}\right). \tag{2.31}$$

The parameter $\mathrm{E_{cut}}$ appearing in the previous definition is known in the materials science community as the *energy cutoff*.

2.6. Self-consistent field iterations. Once a discretization space has been properly defined, it is possible to solve (2.14) via fixed-point-like algorithms, commonly called self-consistent field (SCF) algorithms, which we briefly recall here (see [17] and references therein for a more detailed presentation of such algorithms). We start from an initial guess $\Phi_{N,0} \in \mathcal{V}_N^{N_{\rm el}}$ and, at each iteration $m \in \mathbb{N}$, we solve the following linear eigenvalue problem

$$\begin{cases} \left(\Pi_N H_{\rho_{\gamma_{N,m}}} \Pi_N\right) \phi_{i,N,m+1} = \lambda_{i,N,m+1}\phi_{i,N,m+1}, \quad i = 1, \dots, N_{\rm el}, \\ \langle \phi_{i,N,m+1}, \phi_{j,N,m+1}\rangle = \delta_{ij}, \quad i,j = 1, \dots, N, \\ \gamma_{N,m+1} = \displaystyle\sum_{i=1}^{N_{\rm el}} |\phi_{i,N,m+1}\rangle\langle\phi_{i,N,m+1}|, \end{cases} \tag{2.32}$$

where the Hamiltonian $H_{\rho_{\gamma_{N,m}}}$ at the current iteration is diagonalized in order to build the density matrix $\gamma_{N,m+1}$ using the lowest $N_{\rm el}$ eigenvalues $\lambda_{1,N,m+1}, \dots, \lambda_{N_{\rm el},N,m+1}$ and so on and so forth. By continuity arguments, the existence of a spectral gap for the infinite-dimensional problem (2.20) ensures that, for N and m large enough, the variational eigenvalue problem at iteration m and its infinite dimensional counterpart also admit such a gap. They thus fit the assumptions required in [11] and error bounds can be derived for the linear eigenvalue problem (2.32).

3. *A posteriori* **analysis.** In this section we provide an *a posteriori* error analysis, from which we obtain guaranteed error bounds for the ground-state energy of the continuous problem (2.14).

3.1. Abstract analysis. We start with a short abstract analysis, which indicates how one can obtain guaranteed error bounds on the ground-state energy as soon as the nonlinear term F is a convex functional of the density and lower bounds of the linearized eigenvalue problems are available.

LEMMA 3.1. *For any $\mu \in \mathbb{R}$, for any $\gamma_1, \gamma_2 \in \mathcal{M}$, there holds*

$$\begin{aligned} E(\gamma_2) - E(\gamma_1) = \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_2\big) - \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) \\ - \left(F(\rho_{\gamma_1}) - F(\rho_{\gamma_2}) - \langle F'(\rho_{\gamma_2}), \rho_{\gamma_1} - \rho_{\gamma_2}\rangle\right) . \end{aligned} \tag{3.1}$$

Proof. Define the following intermediate quantity for any $\gamma_1, \gamma_2 \in \mathcal{M}$:

$$J(\gamma_1, \gamma_2) \coloneqq \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_2\big) - \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) . \tag{3.2}$$

Then, for any $\mu \in \mathbb{R}$ and any $\gamma_1, \gamma_2 \in \mathcal{M}$, there holds, using that $\mathrm{Tr}(\gamma_2) = \mathrm{Tr}(\gamma_1) = N_{\mathrm{el}}$, (2.16) and (2.15)

$$\begin{aligned} J(\gamma_1, \gamma_2) &= \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}})\gamma_2\big) - \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}})\gamma_1\big) \\ &= E(\gamma_2) + \mathrm{Tr}(V_{\rho_{\gamma_2}}\gamma_2) - F(\rho_{\gamma_2}) - E(\gamma_1) - \mathrm{Tr}(V_{\rho_{\gamma_2}}\gamma_1) + F(\rho_{\gamma_1}) \\ &= E(\gamma_2) - E(\gamma_1) + F(\rho_{\gamma_1}) - F(\rho_{\gamma_2}) - \langle F'(\rho_{\gamma_2}), \rho_{\gamma_1} - \rho_{\gamma_2}\rangle , \end{aligned}$$

hence the result. ∎

In order to obtain guaranteed error bounds, we now make the following, crucial, assumption.

Assumption 3.2 (Convexity of the functional F). The functional F is convex on $\mathcal{M}$, so that

$$\forall \gamma_1, \gamma_2 \in \mathcal{M}, \quad F(\rho_{\gamma_1}) - F(\rho_{\gamma_2}) - \langle F'(\rho_{\gamma_2}), \rho_{\gamma_1} - \rho_{\gamma_2}\rangle \geq 0. \tag{3.3}$$

COROLLARY 3.3. *Let Assumption 3.2 be satisfied and $\gamma_1, \gamma_2 \in \mathcal{M}$ be fixed. Assume that $\mu \in \mathbb{R}$ is such that*

$$\mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) \geq 0. \tag{3.4}$$

Then

$$E(\gamma_2) - E(\gamma_1) \leq \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_2\big) . \tag{3.5}$$

Proof. Using Assumption 3.2 and $\mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) \geq 0$ in (3.1), we immediately obtain the result. ∎

Remark 3.4. Note that condition (3.4) is equivalent to $\mathrm{Tr}((h + V_{\rho_{\gamma_2}})\gamma_1) \geq \mu N_{\mathrm{el}}$.

Therefore, provided that we are able to choose μ such that $\mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) \geq 0$, we obtain a guaranteed bound on the energy error which reads

$$E(\gamma_2) - E(\gamma_1) \leq \mathrm{Tr}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_2\big) \tag{3.6}$$

and does not involve γ_1. We explain in the next section how to choose this parameter μ.

3.2. Error bounds for linear eigenvalue problems. As a preliminary step, we focus on a linear eigenvalue problem in an abstract setting. Let A be a generic self-adjoint positive operator on $H^1_{\mathrm{per}}(\Omega)$ with compact resolvent, and let $(\varphi_i)_{1\le i\le N_{\mathrm{el}}} \in (H^1_{\mathrm{per}}(\Omega))^{N_{\mathrm{el}}}$, $(\varepsilon_i)_{1\le i\le N_{\mathrm{el}}} \in \mathbb{R}^{N_{\mathrm{el}}}$, $\epsilon_1 \le \epsilon_2 \le \ldots \le \epsilon_{N_{\mathrm{el}}}$, be the solutions to the linear (infinite-dimensional) eigenvalue problem

$$\begin{cases} A\varphi_i = \varepsilon_i\varphi_i, \quad i = 1,\ldots,N_{\mathrm{el}}, \\ \langle\varphi_i,\varphi_j\rangle = \delta_{ij}, \quad i,j = 1,\ldots,N_{\mathrm{el}}, \\ \gamma^0 = \displaystyle\sum_{i=1}^{N_{\mathrm{el}}} |\varphi_i\rangle\langle\varphi_i|. \end{cases} \tag{3.7}$$

It is well known [12, Section 2.2] that γ^0 minimizes $\gamma \mapsto \mathrm{Tr}(A\gamma)$ over $\mathcal{M}$. Let $\mathcal{V}_N \subset H^1_{\mathrm{per}}(\Omega)$ be a finite dimensional subspace and let $(\varphi_{i,N})_{1\le i\le N_{\mathrm{el}}} \in (\mathcal{V}_N)^{N_{\mathrm{el}}}$, $(\varepsilon_{i,N})_{1\le i\le N_{\mathrm{el}}} \in \mathbb{R}^{N_{\mathrm{el}}}$, be the solutions to the linear (finite-dimensional) eigenvalue problem

$$\begin{cases} \big(\Pi_N A \Pi_N\big)\varphi_{i,N} = \varepsilon_{i,N}\varphi_{i,N}, \quad i = 1,\ldots,N_{\mathrm{el}}, \\ \langle\varphi_{i,N},\varphi_{i,N}\rangle = \delta_{ij}, \quad i,j = 1,\ldots,N_{\mathrm{el}}, \\ \gamma_N^0 = \displaystyle\sum_{i=1}^{N_{\mathrm{el}}} |\varphi_{i,N}\rangle\langle\varphi_{i,N}|. \end{cases} \tag{3.8}$$

In [11, Theorem 5.9], the authors introduce a fully computable error bound η, which depends only on the residual, the discretization parameters and a lower bound of $\varepsilon_{N_{\mathrm{el}}+1}$, such that

$$0 \le \sum_{i=1}^{N_{\mathrm{el}}} (\varepsilon_{i,N} - \varepsilon_i) \le \eta^2. \tag{3.9}$$

If such a bound is available, defining

$$\mu_N^{\mathrm{lb}} \coloneqq \frac{1}{N_{\mathrm{el}}}\left(\sum_{i=1}^{N_{\mathrm{el}}} \varepsilon_{i,N} - \eta^2\right) \le \frac{1}{N_{\mathrm{el}}}\sum_{i=1}^{N_{\mathrm{el}}} \varepsilon_i, \tag{3.10}$$

we obtain a computable constant μ_N^{lb} such that for any $\gamma \in \mathcal{M}$, $\mathrm{Tr}\big((A - \mu_N^{\mathrm{lb}})\gamma\big) \ge 0$ holds, meaning (3.4) holds.

In order to obtain a computable η^2 satisfying (3.9) above, we follow [11] adapted to the current settings. This bound depends on

(i) the single eigenpair residuals operators $\mathrm{Res}(\varphi_{i,N},\varepsilon_{i,N})$ [11, Definition 3.4]; for each i this operator is defined by its action over $\varphi \in H^1_{\mathrm{per}}(\Omega)$:

$$\begin{aligned} \langle \mathrm{Res}(\varphi_{i,N},\varepsilon_{i,N}),\varphi\rangle_{H^{-1}_{\mathrm{per}}(\Omega),H^1_{\mathrm{per}}(\Omega)} \coloneqq{}& \varepsilon_{i,N}\langle\varphi_{i,N},\varphi\rangle \\ & - \langle A\varphi_{i,N},\varphi\rangle_{H^{-1}_{\mathrm{per}}(\Omega),H^1_{\mathrm{per}}(\Omega)}, \end{aligned} \tag{3.11}$$

so that it can be identified with the standard residual $\varepsilon_{i,N}\varphi_{i,N} - A\varphi_{i,N}$, as a vector in $H^{-1}_{\mathrm{per}}(\Omega)$. With reasonable assumptions on the potential V (see Remark 3.6), we can assume that this residual actually belongs to $L^2_{\mathrm{per}}(\Omega)$;

(ii) the 2-Schatten norm, denoted by $\|\cdot\|_{\mathfrak{S}_2(\mathcal{H})}$, of the cluster residual [11, Definition 3.5] $\mathrm{Res}(\gamma_N)$ defined by

(3.12)
$$\|\mathrm{Res}(\gamma_N)\|^2_{\mathfrak{S}_2(\mathcal{H})} = \sum_{i=1}^{N_{\mathrm{el}}} \langle \mathrm{Res}(\varphi_{i,N}, \varepsilon_{i,N}), A^{-1}\mathrm{Res}(\varphi_{i,N}, \varepsilon_{i,N})\rangle_{H^{-1}_{\mathrm{per}}(\Omega), H^1_{\mathrm{per}}(\Omega)},$$

(3.13)
$$\|A^{-1/2}\mathrm{Res}(\gamma_N)\|^2_{\mathfrak{S}_2(\mathcal{H})} = \sum_{i=1}^{N_{\mathrm{el}}} \langle A^{-1}\mathrm{Res}(\varphi_{i,N}, \varepsilon_{i,N}), A^{-1}\mathrm{Res}(\varphi_{i,N}, \varepsilon_{i,N})\rangle;$$

(iii) $\underline{\varepsilon}_{N_{\mathrm{el}}+1}$, a (computable) lower bound of the exact eigenvalue $\varepsilon_{N_{\mathrm{el}}+1}$.

Then, by employing equations (4.1), (4.4) and (4.16) from [11] we obtain

$$\eta^2 \coloneqq \|\mathrm{Res}(\gamma_N)\|^2_{\mathfrak{S}_2(\mathcal{H})} + 4\varepsilon_{N_{\mathrm{el}},N} c_N^2 \|A^{-1/2}\mathrm{Res}(\gamma_N)\|^2_{\mathfrak{S}_2(\mathcal{H})}, \tag{3.14}$$

where

$$c_N = \left(1 - \frac{\varepsilon_{N_{\mathrm{el}},N}}{\underline{\varepsilon}_{N_{\mathrm{el}}+1}}\right)^{-1}. \tag{3.15}$$

There are two difficulties in the evaluation of this bound: it involves the inversion of the operator A and the computation of a lower bound on $\varepsilon_{N_{\mathrm{el}}+1}$. The inversion can be circumvented by recasting it as a linear systems of equations, which is still computationally expensive. This is the focus of the next section, after a short remark on the use of negative Sobolev norms. The computation of the lower bound will be commented in Section 4.

Remark 3.5 (Negative Sobolev norms). The computation of the error bounds using negative Sobolev norms of the residuals in [11] relies on additional assumptions on the linear operator A, for instance $A \geq -\frac{1}{2}\Delta + 1$, a much stronger condition than $A \geq 0$. This is fine from a mathematical point of view since shifting can ensure that this condition holds. However, in our numerical experiments it appeared that such shifts were so significant that, even for systems with large gaps, the *relative* gap then becomes very close to 1, making in turn the constant c_N from (3.15) of order 10^6, thus rendering the bound unusable. Second, [11, Theorem 5.9] actually relies not only on $A \geq -\frac{1}{2}\Delta + 1$, but also on $A^2 \geq (\frac{1}{2}\Delta + 1)^2$, which is *not* implied by the first condition if A and Δ do not commute.

Remark 3.6 (Regularity of the residuals). The regularity of the potential V has a direct impact on the regularity of the orbitals φ_i and density ρ. For instance, it was shown in [4] for the LDA exchange-correlation functional, that if $V \in H^s_{\mathrm{per}}(\Omega)$ for $s > 3/2$, then the orbitals and the density are in $H^{s+2}_{\mathrm{per}}(\Omega)$, a valid framework for Troullier–Martins pseudopotentials [56]. For GTH pseudopotentials [27, 29], used in the numerical simulations performed with DFTK in Section 4, the pseudopotentials are actually real-analytic, a property that translates to the orbitals, see [15]. For all these reasons, it is reasonable to assume that the residuals have actually at least $L^2_{\mathrm{per}}(\Omega)$ regularity.

3.3. Practical strategies for efficient computation of error bounds based on operator splitting. We now discuss some alternative strategies which do not

require to solve a full linear system for computing (3.14). We assume $A = -\frac{1}{2}\Delta + V$ for some linear potential $V \in L^2_{\mathrm{per}}(\Omega;\mathbb{R})$ such that $A > 0$, as the Hamiltonian at a given SCF iteration will be of this form. Recall that Π_N is the orthogonal projection on the final dimensional subspace $\mathcal{V}_N \subset H^1_{\mathrm{per}}(\Omega)$, with $\Pi_N^\perp$ being the projection on the corresponding orthogonal complement. In the decomposition $\mathcal{H} = \mathcal{V}_N \oplus \mathcal{V}_N^\perp$, the operator A can be written in block representation as

$$A = \begin{bmatrix} \Pi_N A \Pi_N & \Pi_N A \Pi_N^\perp \\ \Pi_N^\perp A \Pi_N & \Pi_N^\perp A \Pi_N^\perp \end{bmatrix}. \tag{3.16}$$

Defining $\langle V \rangle \in \mathbb{R}$ as the average value of the potential V over the unit cell and setting

$$H_0 = \begin{bmatrix} \Pi_N \left(-\frac{1}{2}\Delta + V\right) \Pi_N & 0 \\ 0 & \Pi_N^\perp \left(-\frac{1}{2}\Delta + \langle V \rangle\right) \Pi_N^\perp \end{bmatrix}, \tag{3.17}$$

together with

$$W = \begin{bmatrix} 0 & \Pi_N V \Pi_N^\perp \\ \Pi_N^\perp V \Pi_N & \Pi_N^\perp (V - \langle V \rangle) \Pi_N^\perp \end{bmatrix}, \tag{3.18}$$

we have $A = H_0 + W$. Under the assumption $\|H_0^{-1} W\| < 1$, A^{-1} admits the Neumann expansion

$$A^{-1} = \sum_{n=0}^{+\infty} (-H_0^{-1} W)^n H_0^{-1}. \tag{3.19}$$

Therefore A^{-1} can be approximated, among others, by the zeroth-order term in the series:

$$A^{-1} = H_0^{-1} + \mathcal{O}(\|H_0^{-1} W\|), \tag{3.20}$$

or the first-order term of the series:

$$A^{-1} = H_0^{-1} - H_0^{-1} W H_0^{-1} + \mathcal{O}(\|H_0^{-1} W\|^2). \tag{3.21}$$

This leads to two corresponding approximations for the estimator η defined by (3.14). Using the notation $r_{i,N} = \mathrm{Res}(\varphi_{i,N}, \varepsilon_{i,N})$ for the sake of clarity, this bound reads:

$$\eta_0^2 \coloneqq \sum_{i=1}^{N_{\mathrm{el}}} \langle r_{i,N}, H_0^{-1} r_{i,N} \rangle + 4\varepsilon_{N_{\mathrm{el}},N} c_N^2 \sum_{i=1}^{N_{\mathrm{el}}} \langle H_0^{-1} r_{i,N}, H_0^{-1} r_{i,N} \rangle, \tag{3.22}$$

and

$$\begin{aligned} \eta_1^2 \coloneqq & \sum_{i=1}^{N_{\mathrm{el}}} \langle r_{i,N}, \left(H_0^{-1} - H_0^{-1} W H_0^{-1}\right) r_{i,N} \rangle \\ & + 4\varepsilon_{N_{\mathrm{el}},N} c_N^2 \sum_{i=1}^{N_{\mathrm{el}}} \langle \left(H_0^{-1} - H_0^{-1} W H_0^{-1}\right) r_{i,N}, \left(H_0^{-1} - H_0^{-1} W H_0^{-1}\right) r_{i,N} \rangle. \end{aligned} \tag{3.23}$$

Using either one of these two bounds on the eigenvalue differences, one can compute a bound on the energy, following (3.10), replacing η respectively by η_0 or η_1. By

introducing this approximation, we have replaced the full operator inversion by the inversion of the block diagonal operator H_0 and a multiplication by the operator W. Note that the block $\Pi_N^\perp \left(-\frac{1}{2}\Delta + \langle V\rangle\right)\Pi_N^\perp$ being diagonal in Fourier representation greatly simplifies its inversion. Also, the block $\Pi_N \left(-\frac{1}{2}\Delta + V\right)\Pi_N$ is dense, but its inversion is performed on the finite dimensional space $\mathcal{V}_N$. In fact, since $\Pi_N r_{i,N} = 0$ (if we assume the finite-dimensional linear eigenvalue problem (3.8) is solved exactly), the computation of the zeroth-order approximation is obtained by the inversion of a diagonal system in $\mathcal{V}_N^\perp$ and thus is very efficient. The price of this approach is the introduction of an additional, *a priori* uncontrollable, source of error, originating from the truncation of the Neumann series.

We now obtain an upper bound for the remainder in (3.20) or (3.21), controlling the additional error introduced above. This will result in two more additional bounds that are mathematically guaranteed. We begin by writing the preconditioned residual $A^{-1}r_{i,N}$ as

$$A^{-1}r_{i,N} = \sum_{n=0}^{\infty}(-H_0^{-1}W)^n H_0^{-1} r_{i,N}, \tag{3.24}$$

and split it into two parts, depending on the number of terms in the approximation, which we denote by L. Notice $L = 0$ corresponds to (3.20) and $L = 1$ corresponds to (3.21). We have $A^{-1}r_{i,N} = \chi_{i,N} + e_{i,N}$ with

$$\chi_{i,N} = \sum_{n=0}^{L}(-H_0^{-1}W)^n H_0^{-1} r_{i,N} \tag{3.25}$$

and

$$e_{i,N} = \sum_{n=L+1}^{+\infty}(-H_0^{-1}W)^n H_0^{-1} r_{i,N}. \tag{3.26}$$

The norm of the remainders can then be estimated as

$$\begin{aligned}
\|e_{i,N}\| &= \left\| \sum_{n=L+1}^{\infty}(-H_0^{-1}W)^n H_0^{-1} r_{i,N} \right\| \\
&= \left\| \sum_{n=0}^{\infty}(-H_0^{-1}W)^n (-H_0^{-1}W)^{L+1} H_0^{-1} r_{i,N} \right\| \\
&\leq \frac{\|(-H_0^{-1}W)^{L+1} H_0^{-1} r_{i,N}\|}{1 - \|H_0^{-1}W\|} \leq \frac{\|(-H_0^{-1}W)^{L+1}\| \|H_0^{-1} r_{i,N}\|}{1 - \|H_0^{-1}W\|} =: \tilde{e}_{i,N,L}.
\end{aligned} \tag{3.27}$$

For any $i = 1, \ldots, N_{\mathrm{el}}$, we can now bound the preconditioned residuals appearing in (3.14) as follows:

$$\begin{aligned}
\langle r_{i,N}, A^{-1}r_{i,N}\rangle &= \langle r_{i,N}, \chi_{i,N} + e_{i,N}\rangle \\
&\leq \langle r_{i,N}, \chi_{i,N}\rangle + \|r_{i,N}\|\|e_{i,N}\|,
\end{aligned} \tag{3.28}$$

and

$$\begin{aligned}
\langle A^{-1}r_{i,N}, A^{-1}r_{i,N}\rangle &= \langle \chi_{i,N} + e_{i,N}, \chi_{i,N} + e_{i,N}\rangle \\
&\leq \langle \chi_{i,N}, \chi_{i,N}\rangle + 2\|e_{i,N}\|\|\chi_{i,N}\| + \|e_{i,N}\|^2.
\end{aligned} \tag{3.29}$$

Next, we sum (3.28) and (3.29) over $i = 1, \dots, N_{\mathrm{el}}$, substitute in (3.14), (3.27) and identify the terms corresponding to (3.22) and (3.23) to obtain fully guaranteed upper bounds on η^2. These bounds read, for the zeroth-order approximation,

$$\eta^2 \le \eta_0^2 + \sum_{i=1}^{N_{\mathrm{el}}} \|r_{i,N}\| \tilde{e}_{i,N,0} + 4\varepsilon_{N_{\mathrm{el}},N} c_N^2 \big(2\tilde{e}_{i,N,0} \|H_0^{-1} r_{i,N}\| + \tilde{e}_{i,N,0}^2\big) =: \eta_{0,\mathrm{g}}^2, \tag{3.30}$$

and, for the first-order approximation,

$$\begin{aligned} \eta^2 \le \eta_1^2 &+ \sum_{i=1}^{N_{\mathrm{el}}} \|r_{i,N}\| \tilde{e}_{i,N,1} \\ &+ 4\varepsilon_{N_{\mathrm{el}},N} c_N^2 \big(2\tilde{e}_{i,N,1} \|(H_0^{-1} - H_0^{-1} W H_0^{-1}) r_{i,N}\| + \tilde{e}_{i,N,1}^2\big) =: \eta_{1,\mathrm{g}}^2. \end{aligned} \tag{3.31}$$

Regarding a way to estimate the operator norm $\|H_0^{-1}W\|$, we refer to Appendix B. We now show how to compile everything into *a posteriori* error bounds for the nonlinear problem.

3.4. Error bounds for the nonlinear problem. We now apply the error bounds obtained for a linear operator A to the nonlinear problem of interest in this article, namely (2.20). In practice, for a fixed N, γ_N is not directly computable but is rather obtained as the limit of a sequence $(\gamma_{N,m})_{m\in\mathbb{N}}$ typically generated by SCF iterations (2.32). To this end, we follow Section 3.2 by applying Corollary 3.3 with a global minimizer $\gamma_\star$ of (2.14) (resp. $\gamma_{N,m}$ from (2.32)) in place of γ_1 (resp. γ_2), $A = H_{\rho_{\gamma_{N,m}}}$ (with A shifted by a positive constant if necessary to guarantee $A > 0$) and $\varepsilon_{i,N} = \lambda_{i,N,m+1}$. In other words, we compute $\mu_{N,m+1}^{\mathrm{lb}}$ as a lower bound of the mean of the eigenvalues from the (infinite-dimensional) eigenvalue problem (3.7) with $A = H_{\rho_{\gamma_{N,m}}}$, using (3.10) with one of the proposed bounds on η, *e.g.* η_0 or η_1. This bound then reads

$$0 \le E(\gamma_{N,m}) - E(\gamma_\star) \le \mathrm{Tr}\big((H_{\rho_{\gamma_{N,m}}} - \mu_{N,m+1}^{\mathrm{lb}})\gamma_{N,m}\big), \tag{3.32}$$

and yields the following certification of the energy at iteration m of the SCF:

$$E(\gamma_\star) \in \Big[E(\gamma_{N,m}) - \mathrm{Tr}\big((H_{\rho_{\gamma_{N,m}}} - \mu_{N,m+1}^{\mathrm{lb}})\gamma_{N,m}\big), E(\gamma_{N,m})\Big]. \tag{3.33}$$

We can then also separate the error bound $\mathrm{Tr}\big((H_{\rho_{\gamma_{N,m}}} - \mu_{N,m+1}^{\mathrm{lb}})\gamma_{N,m}\big)$ into two parts: one depending mainly on the discretization and the other depending mainly on the number of performed SCF iterations. More precisely, define

$$\mathtt{err}_{N,m}^{\mathrm{disc}} \coloneqq \mathrm{Tr}\big((H_{\rho_{\gamma_{N,m}}} - \mu_{N,m+1}^{\mathrm{lb}})\gamma_{N,m+1}\big) \ge 0 \tag{3.34}$$

and

$$\mathtt{err}_{N,m}^{\mathrm{SCF}} \coloneqq \mathrm{Tr}(H_{\rho_{\gamma_{N,m}}}\gamma_{N,m}) - \mathrm{Tr}(H_{\rho_{\gamma_{N,m}}}\gamma_{N,m+1}) \ge 0. \tag{3.35}$$

Then, as $\mathrm{Tr}(\gamma_{N,m+1}) = \mathrm{Tr}(\gamma_{N,m})$, we naturally have that $\mathrm{Tr}\big((H_{\rho_{\gamma_{N,m}}} - \mu_{N,m+1}^{\mathrm{lb}})\gamma_{N,m}\big) = \mathtt{err}_{N,m}^{\mathrm{disc}} + \mathtt{err}_{N,m}^{\mathrm{SCF}}$. Putting things together, we proved the following theorem.

THEOREM 3.7 (Fully guaranteed error bound on the energy). *Under Assumption 3.2 and assuming that, at iteration m of the SCF algorithm in $\mathcal{V}_N$, (3.4) holds with $\mu = \mu^{\mathrm{lb}}_{N,m+1}$, then we have*

$$\boxed{E(\gamma_{N,m}) - E(\gamma_\star) \leq \mathtt{err}^{\mathrm{disc}}_{N,m} + \mathtt{err}^{\mathrm{SCF}}_{N,m},} \tag{3.36}$$

where $\mathtt{err}^{\mathrm{disc}}_{N,m}$ and $\mathtt{err}^{\mathrm{SCF}}_{N,m}$ are respectively defined by (3.34) and (3.35).

We remark that (i) $\mathtt{err}^{\mathrm{SCF}}_{N,m}$ goes to zero as m goes to infinity, provided that the SCF algorithm does converge, and (ii) $\mathtt{err}^{\mathrm{disc}}_{N,m}$ tends to zero as the discretization space is enlarged, provided that the limit of $\mathcal{V}_N$ when $N \to +\infty$ is $H^1_{\mathrm{per}}(\Omega)$ in the sense that $\forall \phi \in H^1_{\mathrm{per}}(\Omega)$, $\|\phi - \Pi_N \phi_{H^1_{\mathrm{per}}(\Omega)}\| \to 0$ as $N \to \infty$, and that $\mu^{\mathrm{lb}}_{N,m+1}$ is well chosen, for instance as explained above. We end this paper with a series of test cases, in 1D and 3D, where these bounds and their approximations are applied and compared in term of computational cost and accuracy.

4. Numerical results.

4.1. Numerical framework. All the simulations presented in this section are realized with the DFTK software, a recent `Julia` package to perform planewave DFT calculations [31]. DFTK uses the planewave basis introduced in Section 2.4, through a discretization parameter $\mathrm{E}_{\mathrm{cut}} \coloneqq N^2/2$. The nonlinear eigenproblems (2.30) for the Kohn–Sham Hamiltonian are solved iteratively using the SCF algorithm described previously, with proper tuning to ensure convergence (Anderson/DIIS acceleration, density mixing, damping, etc.), with a tolerance on the L^2-norm between two successive densities set to 10^{-10} to stop the iterations. Notice that the linear eigenproblems are solved at each iteration of the SCF procedure with a LOBPCG solver (see *e.g.* [48]). We assume in this paper that these linear eigenproblems are exactly solved within the variational approximation space $\mathcal{V}_N$, the treatment of the numerical linear algebra error being left for future work.

The simulations are performed within the periodic reduced Hartree–Fock(rHF) model, for which $E_{\mathrm{xc}}(\rho) = 0$ (see *e.g.* [6, 54] for the mathematical properties of the rHF model). Assumption 3.2 is thus satisfied, as the functional $\rho \mapsto \mathcal{D}(\rho, \rho)$ is (strictly) convex [28] on *e.g.* the affine space $\{\rho \in L^2_{\mathrm{per}}(\Omega) : \int_\Omega \rho = N_{\mathrm{el}}\}$.

Each simulation is performed in a reference space, defined by a parameter $\mathrm{E}_{\mathrm{cut,ref}}$, and a computation space, defined by a parameter $\mathrm{E}_{\mathrm{cut}} < \mathrm{E}_{\mathrm{cut,ref}}$. The reference space is assumed to account for the full space and is used to compute the reference solutions, as well as the residuals in various norms. The computation space is used to perform the actual calculation and we seek to bound the error on the energy due to the variational approximation of the Kohn–Sham equations (2.21) in this space.

In all the simulations below, we track the error bound on the energy difference $E(\gamma_{N,m}) - E(\gamma_\star)$ by computing $\mathtt{err}^{\mathrm{SCF}}_{N,m}$ and $\mathtt{err}^{\mathrm{disc}}_{N,m}$ as in (3.36), the latter being evaluated with the different η's highlighted in Sections 3.2 and 3.3. For convenience, we recall in Table 1 the different possibilities, together with their properties and computational cost. Note also that, for the sake of simplicity, the constant $\underline{\varepsilon_{N_{\mathrm{el}}+1}}$ appearing in c_N (3.15) is taken as the variational approximation $\varepsilon_{N_{\mathrm{el}}+1,N}$. Other methods to estimate this lower bound are proposed in [11].

TABLE 1

Names of the bounds and corresponding expressions for the different ways of computing η in (3.10). Each of these expression requires $A = H_{\rho_{\gamma_{N,m}}} > 0$ and has a different computational cost.

Name	Notation	Fully guaranteed	Equation	Condition	Computational cost
full-inversion	η	yes	(3.14)	$A > 0$	full-inversion of A in $\mathcal{H}$
zeroth-order	η_0	no	(3.22)	$A > 0$	diagonal inversion in $\mathcal{V}_N^\perp$
zeroth-order guaranteed	$\eta_{0,\mathrm{g}}$	yes	(3.30)	$A > 0$ $\|H_0^{-1}W\| < 1$	diagonal inversion in $\mathcal{V}_N^\perp$ and remainder estimation
first-order	η_1	no	(3.23)	$A > 0$	full-inversion in $\mathcal{V}_N$
first-order guaranteed	$\eta_{1,\mathrm{g}}$	yes	(3.31)	$A > 0$ $\|H_0^{-1}W\| < 1$	full-inversion in $\mathcal{V}_N$ and remainder estimation

4.2. 1D toy model. We first present simulations obtained with a 1D toy model ($\Omega = (0, 10)$), for which reference solutions with very high accuracy can be obtained with a moderate computational cost. The reference basis is built with $\mathrm{E}_{\mathrm{cut,ref}} = 1000$ Ha and the calculation basis with $\mathrm{E}_{\mathrm{cut}} = 400$ Ha. The potential V we use is defined by its Fourier coefficients as:

$$\forall\, G \in 2\pi\mathbb{Z}, \quad \hat{V}_G = \begin{cases} 1 & \text{if } G = 0, \\ \frac{\omega_G}{|G|^{1.1}} & \text{if } 0 < |G| \leq 100, \\ \frac{1}{|G|^{1.1}} & \text{else,} \end{cases} \tag{4.1}$$

where $(\omega_G)_{0<|G|\leq 100}$ are independent random variables, uniformly distributed between -10 and 10. We then solve iteratively the Kohn–Sham equations for $N_{\mathrm{el}} = 3$ on the unit cell Ω with periodic boundary conditions. Note that we obtained qualitatively similar results with other potentials with comparable Sobolev regularity.

In Figures 1 and 2, we display the bounds from Table 1: the full operator inversion, the zeroth-order and first-order approximations, both with and without the estimation of the remainder terms in the truncation of the Neumann series. Note that, when computing the error bound with $\eta_{0,\mathrm{g}}$ and $\eta_{1,\mathrm{g}}$ at every step of the SCF cycle, not only has the Hamiltonian to be shifted to ensure positivity, but also to satisfy $\|H_0^{-1}W\| < 1$. One then realizes that, while both the zeroth- and first-order approximations already give very satisfactory results in Figure 1 (left), adding the remainder terms in the Neumann series to make the bound fully guaranteed worsen the error estimation by about an order of magnitude (Figure 2). We also optimized the shift s in order to make the estimation of the remainder terms (3.27) as small as possible. The optimal shift is computed by a dichotomy to find a zero of the derivative of $\mathtt{err}_{N,m}^{\mathrm{disc}}$ with respect to s, while keeping the constraint $\|H_0^{-1}W\| < 1$. This clearly improves the resulting bound, which is guaranteed, but the final accuracy is still far from being satisfying (Figure 2). As a conclusion, one sees that the best error estimation, even though it is not guaranteed, is given by the zeroth-order bound, without estimating the remainders. Moreover, in Figure 1 (right), the transition between the two contributions ($\mathtt{err}_{N,m}^{\mathrm{SCF}}$ and $\mathtt{err}_{N,m}^{\mathrm{disc}}$) to the error clearly appears.

In Table 2, we compute the effective ratio between $E(\gamma_{N,m}) - E(\gamma_\star)$ and the upper bound $\mathtt{err}_{N,m}^{\mathrm{SCF}} + \mathtt{err}_{N,m}^{\mathrm{disc}}$ for all the η's mentioned above. We expect this ratio to be close to 1 when the bound is accurate, which confirms the observations made before.

Note that the guaranteed bound obtained with the full-inversion is well approximated by the first-order bound, as there is an additional term taken into account in the Neumann series. The zeroth-order bound also seems to approximate better the true error than the full-inversion bound: this is due to the truncation of the Neumann series used to design this bound, and one can not expect this observation to hold in general.

4.3. 3D insulating systems. We now present some numerical illustrations of the error estimation developed in this paper for real 3D systems. In addition to the previous settings, we use the Goedecker–Teter–Hutter (GTH) pseudopotentials [27, 29], already implemented in DFTK, see Remark 4.1 below. We then consider two physical systems:

- a Silicon crystal, first with a single $\boldsymbol{k}$-point only (the Γ point), then with eight $\boldsymbol{k}$-points.
- a Hydrogen-Fluoride molecule with eight $\boldsymbol{k}$-points.

Remark 4.1 (pseudopotentials norms). The calculations from Appendix B can easily be extended to pseudopotentials in the Kleinmann–Bylander form [35]: in addition to the L^∞ norm of the local contribution, one just has to add the largest of the projection coefficients from the nonlocal contribution.

4.3.1. Silicon (Si) crystal. For the Si crystal, a semiconductor with a positive band gap, we set $\mathrm{E}_{\mathrm{cut,ref}} = 400$ Ha, and $\mathrm{E}_{\mathrm{cut}} = 150$ Ha. For the $\boldsymbol{k}$-point sampling (see Appendix A), we use a single $\boldsymbol{k}$-point for the first test and eight $\boldsymbol{k}$-points for the second one.

In Figure 3 (left), we display, for the first case with a single $\boldsymbol{k}$-point, the error with respect to the reference energy as well as the different error bounds computed from: (i) the energy norm (3.14) using the full inverse, (ii) the zeroth-order approximation (3.20) and (iii) the first-order approximation (3.21). Even though the last two bounds are not mathematically guaranteed because of the truncation of the Neumann series, the accuracy of the estimation is very satisfying. In Figure 3 (right), the transition from a SCF dominating error to a discretization dominating error also clearly appears. As for the 1D toy model, we added in Figure 3 the estimation of the remainders together with their optimization with respect to the shift. Again, the resulting bound, now mathematically guaranteed, is far from being effective.

In Figure 5, we present the results for simulations for the same system, now with eight $\boldsymbol{k}$-points. First, note that this case is not suitable for the rHF model as Silicon behaves as a metallic system in this approximation: one usually needs to introduce a numerical smearing and fractional occupation numbers (see for instance [13] and references therein) to ensure the convergence of the SCF procedure. To avoid this, we use the following workaround. As Silicon is expected to be a semi-conducting system when using the LDA approximation for the exchange-correlation energy (which is not convex), we first run a simulation for this model. Then, we extract the effective (*i.e.* converged) exchange-correlation potential $V_{\mathrm{xc}}[\rho]$, and we use it in the external, linear, potential V before running the SCF algorithm for the rHF model. Again, our main observation here is that there is not a significant difference between using the zeroth-order approximation, the first-order approximation or the full-inversion of the operator to compute the energy norm, with a clear transition between the SCF error regime to the discretization error regime. We also observed in our simulations that

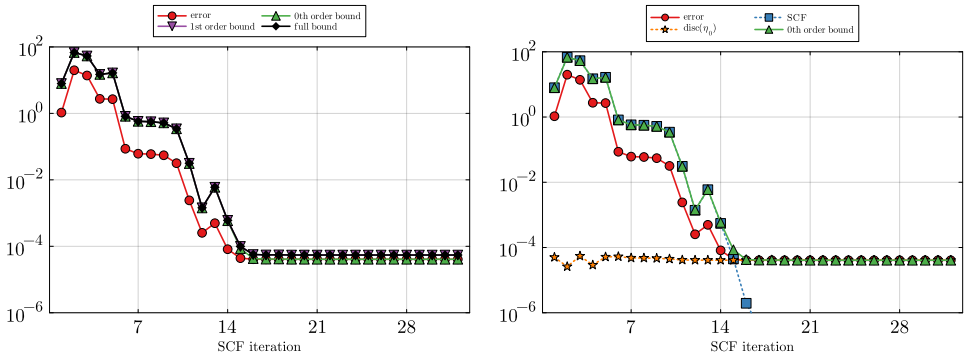

FIGURE 1. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a 1D toy model. (Left) full-inversion bound and its zeroth- and first-order approximations. (Right) Zeroth-order bound and its splitting between SCF and discretization contributions, as in* (3.36).

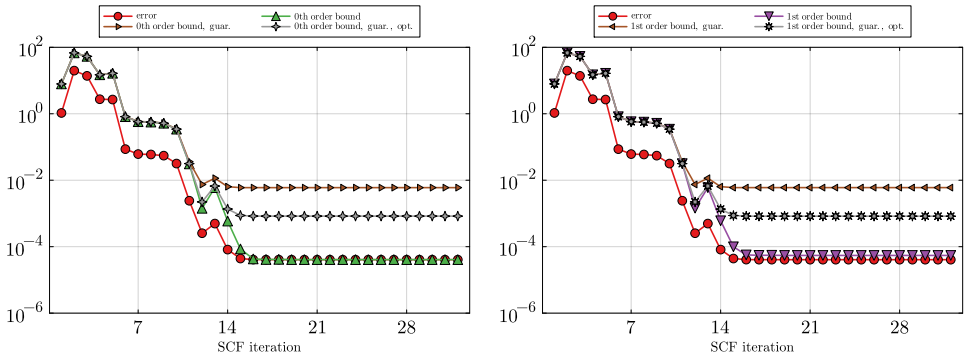

FIGURE 2. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a 1D toy model, with the zeroth-order bound (left) and first-order bound (right). We added the associated guaranteed bounds by estimating the remainders and their optimization with respect to the shift.*

TABLE 2

Ratio between the error $E(\gamma_{N,m}) - E(\gamma_\star)$ and the upper bound $\mathtt{err}_{N,m}^{\mathrm{SCF}} + \mathtt{err}_{N,m}^{\mathrm{disc}}$ for different η's, for a 1D toy model. The closer to 1 the ratio, the better the bound.

SCF it.	η_0	η_1	η	$\eta_{0,\mathrm{g}}$	$\eta_{1,\mathrm{g}}$	$\eta_{0,\mathrm{g,opt}}$	$\eta_{1,\mathrm{g,opt}}$
1	7.47740	7.47740	7.47740	7.48102	7.48106	7.47805	7.47800
5	6.16377	6.16377	6.16377	6.16581	6.16585	6.16424	6.16424
10	10.8366	10.8368	10.8368	1.04030	11.0411	10.8586	10.8585
15	1.94103	2.25435	2.25719	135.935	136.214	19.7679	19.6773
20	1.00401	1.34280	1.34587	147.209	147.513	20.4755	20.3783

estimating the remainders of the Neumann series yields bounds that are not accurate enough to be of any practical interest.

Finally, we tabulate in Table 3 the ratio between the zeroth-order bound or the full-inversion bound and the targeted error: note the difference between the accuracy of the bounds when using one or eight $\boldsymbol{k}$-points. This is due to the chosen linear potential and the rHF model: when using a single $\boldsymbol{k}$-point, the system has a positive gap, but smaller than when using eight $\boldsymbol{k}$-points and a linear potential obtained as above. The constant c_N (3.15) is thus larger when using a single $\boldsymbol{k}$-points, yielding a worse upper bound. This also appears when comparing Figures 3 and 5: the error bound is better in the second case, for which the gap is smaller.

TABLE 3

Ratio between the error $E(\gamma_{N,m}) - E(\gamma_\star)$ and the upper bound $\mathtt{err}_{N,m}^{\mathrm{SCF}} + \mathtt{err}_{N,m}^{\mathrm{disc}}$ using η_0 and η. Data displayed for both the Si crystal (with one and eight $\boldsymbol{k}$-points) and the HF molecule, every five steps of the SCF iterations.

SCF it.	Si (one $\boldsymbol{k}$-point)		Si (eight $\boldsymbol{k}$-points)		HF (eight $\boldsymbol{k}$-points)	
	η_0	η	η_0	η	η_0	η
1	1.53019	1.53019	1.48986	1.48986	3.13099	3.13099
5	4.04515	4.04515	1.22005	1.22010	2.73819	2.73871
10	6.80721	6.80721	1.02225	1.06797	0.99125	1.02102
15	4.41846	4.42815	–	–	0.98945	1.01923
20	3.77631	3.93929	–	–	–	–

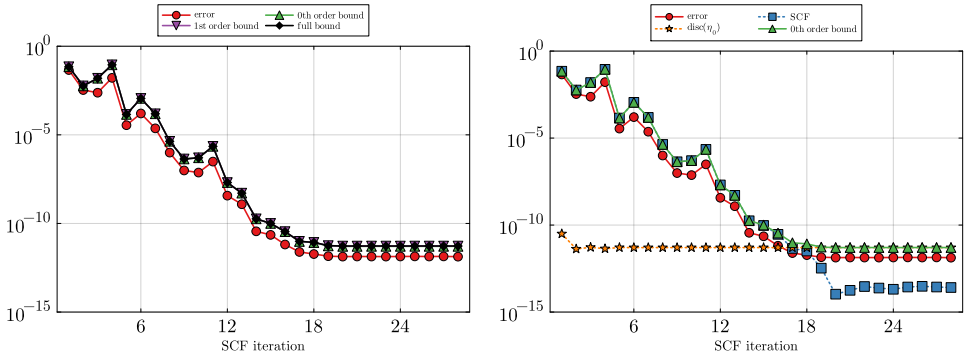

FIGURE 3. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a Si crystal (one $\mathbf{k}$-point). (Left) Full-inversion bound with zeroth- and first-order approximations. (Right) Zeroth-order bound and its splitting between SCF and discretization contributions, as in* (3.36).

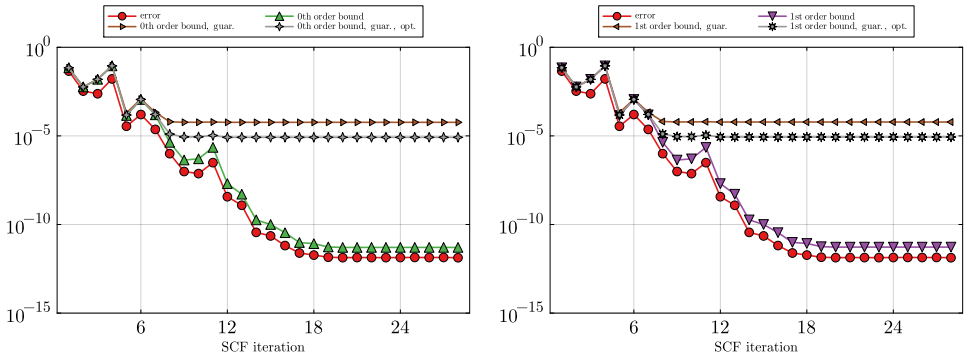

FIGURE 4. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a Si crystal (one $\mathbf{k}$-point), with the zeroth-order bound (left) and first-order bound (right). We added the associated guaranteed bounds by estimating the remainders and their optimization with respect to the shift.*

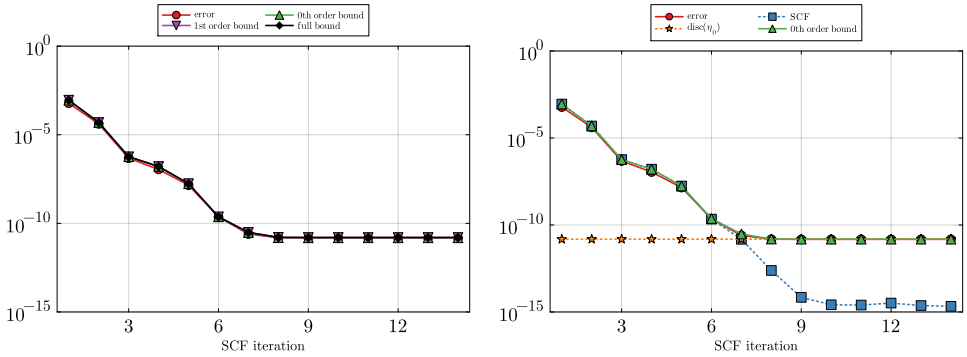

FIGURE 5. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a Si crystal (eight $\mathbf{k}$-points). (Left) Full-inversion bound with zeroth- and first-order approximations. (Right) Zeroth-order bound and its splitting between SCF and discretization contributions, as in* (3.36).

4.3.2. Hydrogen Fluoride (HF) molecule. We run here a simulation for eight $\boldsymbol{k}$-points, with $\mathrm{E_{cut,ref}} = 1500$ Ha, $\mathrm{E_{cut}} = 750$ Ha (we use the same trick to ensure convergence of the SCF as the one we used for the Silicon system). In Figure 6, we plot the results from the said simulation and in Table 3 we display the ratio between the energy difference and the zeroth-order bound for the simulation with eight $\boldsymbol{k}$-points. We notice again that the zeroth-order approximation provides a very satisfying estimation of the true error, with a similar transition between the regime where the SCF error dominates and the one where the discretization error dominates. Note however that, while the full inversion is guaranteed, this time the zeroth-order bound is not, due to the truncation of the Neumann series.

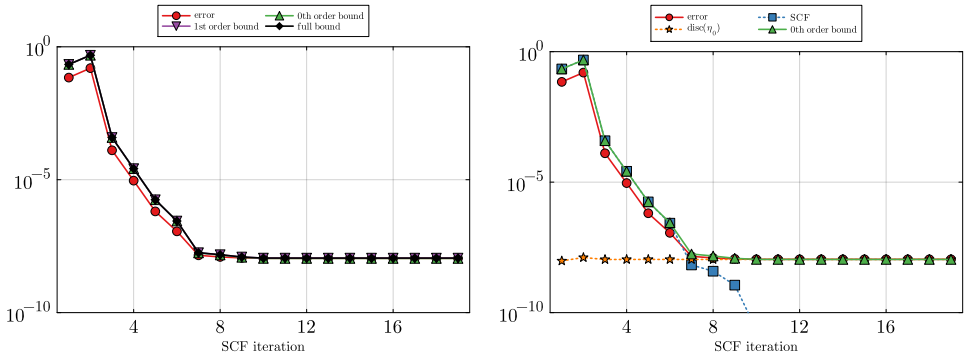

FIGURE 6. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a HF molecule (eight $\boldsymbol{k}$-points). (Left) Full-inversion bound with zeroth- and first-order approximations. (Right) Zeroth-order bound and its splitting between SCF and discretization contributions, as in* (3.36).

4.4. Usage of the bounds in practice and conclusion. We would like to end this section with a brief discussion on the practical aspects of the bounds derived in this paper. Table 4 summarizes the execution time for one of the simulations, which will be useful to decide which bound to employ. First, the bound based on the full-inversion of the Hamiltonian is clearly the closest to the exact error. Unfortunately, performing the full-inversion at each step is too costly to be used in practical simulations. Next, the approximate inversions by means of a zeroth or first-order truncation seem to lead to very similar results. Hence, the zeroth-order approximation being the cheapest to compute, it seems the most relevant one. Indeed, even if it is in the finite dimensional space $\mathcal{V}_N$, it is still necessary to perform a full operator inversion in this space to compute the first-order approximation: solving the underlying linear system is as costly as performing a full SCF cycle in the space $\mathcal{V}_N$, which greatly reduces its applicability. Finally, it appears that adding the estimation of the remainders from the truncation of the Neumann series worsen the bounds by more than one order of magnitude: mathematically guaranteed bounds seem to be only possible at the expanse of accuracy, a conclusion some of us also reached in a previous paper [7] about discretization errors in the calculation of interatomic forces. These considerations therefore support, for practical applications, the choice of the zeroth-order bound, without the estimation of the remainders from the Neumann series.

4.5. Application to nonconvex exchange-correlation functionals (LDA and PBE) for Silicon. Finally, we apply the results from this paper along the SCF iterations for a Si crystal modelled with the LDA [50] and PBE [49] exchange-correlation functionals, where we set $\mathrm{E_{cut,ref}} = 400$ Ha, $\mathrm{E_{cut}} = 50$ Ha, together with

TABLE 4

Time spent in each bound computation with respect to a single simulation for an HF molecule (eight $\boldsymbol{k}$-points). Results for the bound based on the full-inversion of the Hamiltonian (η), the zeroth-order truncation (η_0) and the first-order truncation (η_1).

	Time(hr)	(%)
η	3.45	32.5
η_0	0.11	1.0
η_1	1.47	13.8
Simulation	10.6	

eight $\boldsymbol{k}$-points. In these cases, E_{xc} is not a convex functional of the density anymore, so the bounds cannot be expected to be guaranteed. However, as the additional term is expected to be of higher order in the discretization error (see for instance [4, 22] for the simple $X\alpha$ LDA functional), we can expect the approximation to hold asymptotically. This is confirmed by the numerical experiment in Figure 7, where we notice a very good approximation of the error on the energy along the SCF iterations using the simple zeroth-order bound. In particular, the transition from a SCF dominating error to a discretization dominating error clearly appears, paving the way for efficient adaptive strategies, even for nonconvex density functionals.

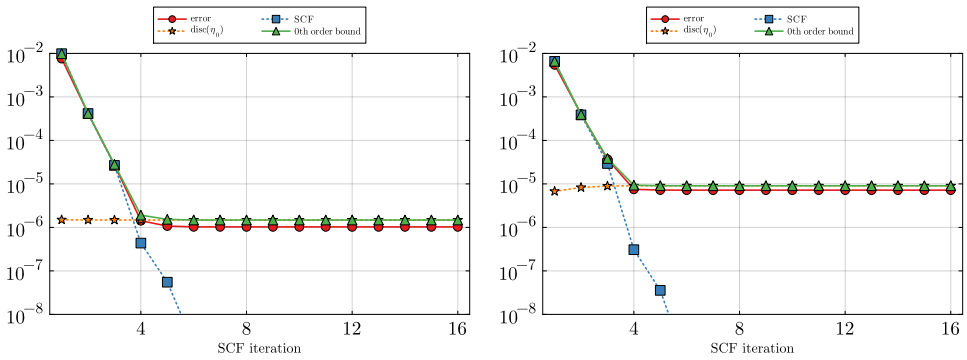

FIGURE 7. *Tracking of the error $E(\gamma_{N,m}) - E(\gamma_\star)$ for a Si crystal with the zeroth-order bound. Data generated with eight $\boldsymbol{k}$-points and the LDA (left) or PBE (right) exchange-correlation functional. "SCF" stands for the SCF error $\mathtt{err}_{N,m}^{\mathrm{SCF}}$ and "disc(η_0)" stands for the discretization error $\mathtt{err}_{N,m}^{\mathrm{disc}}$, computed with η_0 (cf. Table 1). The transition between a regime where the SCF error dominates to a regime where the discretization error dominates clearly appears, even though the bound is not mathematically guaranteed (nonconvex model and truncation of the Neumann series).*

Data availability. All the codes used to generate the plots from this paper are available at https://doi.org/10.18419/darus-4469.

REFERENCES

[1] N. W. BAZLEY AND D. W. FOX, *Lower Bounds for Eigenvalues of Schrödinger's Equation*, Physical Review, 124 (1961), pp. 483–492.

[2] M. BORN AND R. OPPENHEIMER, *Zur quantentheorie der molekeln*, Ann. Phys., 389 (1927), pp. 457–484.

[3] D. Braess, V. Pillwein, and J. Schöberl, *Equilibrated residual error estimates are p-robust*, Computer Methods in Applied Mechanics and Engineering, 198 (2009), pp. 1189–1197.
[4] E. Cancès, R. Chakir, and Y. Maday, *Numerical analysis of the planewave discretization of some orbital-free and Kohn-Sham models*, ESAIM: Mathematical Modelling and Numerical Analysis, 46 (2012), pp. 341–388.
[5] É. Cancès, A. Deleurence, and M. Lewin, *A New Approach to the Modeling of Local Defects in Crystals: The Reduced Hartree-Fock Case*, Communications in Mathematical Physics, 281 (2008), pp. 129–177.
[6] E. Cancés, A. Deleurence, and M. Lewin, *A New Approach to the Modeling of Local Defects in Crystals: The Reduced Hartree-Fock Case*, Communications in Mathematical Physics, 281 (2008), pp. 129–177.
[7] E. Cancès, G. Dusson, G. Kemlin, and A. Levitt, *Practical error bounds for properties in plane-wave electronic structure calculations*, SIAM Journal on Scientific Computing, 44 (2022), pp. B1312–B1340.
[8] E. Cancès, G. Dusson, Y. Maday, B. Stamm, and M. Vohralík, *A perturbation-method-based post-processing for the planewave discretization of Kohn–Sham models*, J. Comput. Phys., 307 (2016), pp. 446–459.
[9] E. Cancès, G. Dusson, Y. Maday, B. Stamm, and M. Vohralík, *Guaranteed and Robust a Posteriori Bounds for Laplace Eigenvalues and Eigenvectors: Conforming Approximations*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 2228–2254.
[10] ———, *Guaranteed and robust a posteriori bounds for Laplace eigenvalues and eigenvectors: A unified framework*, Numerische Mathematik, 140 (2018), pp. 1033–1079.
[11] ———, *Guaranteed a posteriori bounds for eigenvalues and eigenvectors: Multiplicities and clusters*, Mathematics of Computation, (2020).
[12] ———, *Post-processing of the planewave approximation of Schrödinger equations. Part I: Linear operators*, IMA Journal of Numerical Analysis, 41 (2021), pp. 2423–2455.
[13] E. Cancès, V. Ehrlacher, D. Gontier, A. Levitt, and D. Lombardi, *Numerical quadrature in the Brillouin zone for periodic Schrödinger operators*, Numerische Mathematik, 144 (2020), pp. 479–526.
[14] E. Cancès, G. Kemlin, and A. Levitt, *Convergence Analysis of Direct Minimization and Self-Consistent Iterations*, SIAM Journal on Matrix Analysis and Applications, 42 (2021), pp. 243–274.
[15] ———, *A Priori Error Analysis of Linear and Nonlinear Periodic Schrödinger Equations with Analytic Potentials*, Journal of Scientific Computing, 98 (2024), p. 25.
[16] E. Cancès and C. Le Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 34 (2000), pp. 749–774.
[17] E. Cancès, A. Levitt, Y. Maday, and C. Yang, *Numerical Methods for Kohn–Sham Models: Discretization, Algorithms, and Error Analysis*, in Density Functional Theory: Modeling, Mathematical Analysis, Computational Methods, and Applications, E. Cancès and G. Friesecke, eds., Springer International Publishing, Cham, 2023, pp. 333–400.
[18] C. Carstensen and J. Gedicke, *Guaranteed lower bounds for eigenvalues*, Mathematics of Computation, 83 (2014), pp. 2605–2629.
[19] H. Chen, X. Dai, X. Gong, L. He, and A. Zhou, *Adaptive finite element approximations for kohn–sham models*, Multiscale Model. Simul., 12 (2014), pp. 1828–1869.
[20] P. Destuynder and B. Métivet, *Explicit error bounds in a conforming finite element method*, Mathematics of Computation, 68 (1999), pp. 1379–1396.
[21] R. G. Durán, C. Padra, and R. Rodríguez, *A Posteriori Error Estimates for the Finite Element Approximation of Eigenvalue Problems*, Mathematical Models and Methods in Applied Sciences, 13 (2003), pp. 1219–1229.
[22] G. Dusson, *Post-processing of the plane-wave approximation of Schrödinger equations. Part II: Kohn–Sham models*, IMA Journal of Numerical Analysis, 41 (2021), pp. 2456–2487.
[23] G. Dusson and Y. Maday, *A posteriori analysis of a nonlinear Gross–Pitaevskii-type eigenvalue problem*, IMA Journal of Numerical Analysis, 37 (2017), pp. 94–137.
[24] G. Dusson and Y. Maday, *An overview of a posteriori error estimation and post-processing methods for nonlinear eigenvalue problems*, Journal of Computational Physics, 491 (2023), p. 112352.
[25] A. Ern and M. Vohralík, *Polynomial-Degree-Robust A Posteriori Estimates in a Unified Setting for Conforming, Nonconforming, Discontinuous Galerkin, and Mixed Discretizations*, SIAM Journal on Numerical Analysis, 53 (2015), pp. 1058–1081.
[26] G. E. Forsythe, *Asymptotic lower bounds for the frequencies of certain polygonal membranes*,

Pacific Journal of Mathematics, 4 (1954), pp. 467–480.
[27] S. Goedecker, M. Teter, and J. Hutter, *Separable dual-space Gaussian pseudopotentials*, Physical Review B, 54 (1996), p. 1703.
[28] D. Gontier and S. Lahbabi, *Convergence rates of supercell calculations in the reduced Hartree-Fock model*, ESAIM: Mathematical Modelling and Numerical Analysis, 50 (2016), pp. 1403–1424.
[29] C. Hartwigsen, S. Gœdecker, and J. Hutter, *Relativistic separable dual-space gaussian pseudopotentials from h to rn*, Physical Review B, 58 (1998), p. 3641.
[30] M. F. Herbst, A. Levitt, and E. Cancès, *A posteriori error estimation for the non-self-consistent Kohn–Sham equations*, Faraday Discussions, 224 (2020), pp. 227–246.
[31] ———, *DFTK: A Julian approach for simulating electrons in solids*, Proceedings of the JuliaCon Conferences, 3 (2021), p. 69.
[32] J. Hu, Y. Huang, and Q. Lin, *Lower Bounds for Eigenvalues of Elliptic Operators: By Nonconforming Finite Element Methods*, Journal of Scientific Computing, 61 (2014), pp. 196–221.
[33] J. Hu, Y. Huang, and Q. Shen, *The Lower/Upper Bound Property of Approximate Eigenvalues by Nonconforming Finite Element Methods for Elliptic Operators*, Journal of Scientific Computing, 58 (2014), pp. 574–591.
[34] T. Kato, *On the Upper and Lower Bounds of Eigenvalues*, Journal of the Physical Society of Japan, 4 (1949), pp. 334–339.
[35] L. Kleinman and D. M. Bylander, *Efficacious Form for Model Pseudopotentials*, Physical Review Letters, 48 (1982), pp. 1425–1428.
[36] W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*, Physical Review, 140 (1965), pp. A1133–A1138.
[37] P. Ladeveze and D. Leguillon, *Error Estimate Procedure in the Finite Element Method and Applications*, SIAM Journal on Numerical Analysis, 20 (1983), pp. 485–509.
[38] M. G. Larson, *A Posteriori and a Priori Error Analysis for Finite Element Approximations of Self-Adjoint Elliptic Eigenvalue Problems*, SIAM Journal on Numerical Analysis, 38 (2000), pp. 608–625.
[39] A. Levitt, *Screening in the Finite-Temperature Reduced Hartree–Fock Model*, Arch Rational Mech Anal, 238 (2020), pp. 901–927.
[40] E. H. Lieb, *Density functionals for coulomb systems*, International Journal of Quantum Chemistry, 24 (1983), pp. 243–277.
[41] L. Lin and B. Stamm, *A posteriori error estimates for discontinuous galerkin methods using non-polynomial basis functions. part II: Eigenvalue problems*, Esaim Math. Model. Numer. Anal., (2016).
[42] X. Liu, *A framework of verified eigenvalue bounds for self-adjoint differential operators*, Applied Mathematics and Computation, 267 (2015), pp. 341–355.
[43] F. Luo, Q. Lin, and H. Xie, *Computing the lower and upper bounds of Laplace eigenvalue problem: By combining conforming and nonconforming finite element methods*, Science China Mathematics, 55 (2012), pp. 1069–1082.
[44] Y. Maday and G. Turinici, *Error bars and quadratically convergent methods for the numerical simulation of the Hartree-Fock equations*, Numer. Math., 94 (2003), pp. 739–770.
[45] H. J. Monkhorst and J. D. Pack, *Special points for Brillouin-zone integrations*, Physical Review B, 13 (1976), pp. 5188–5192.
[46] P. Motamarri, M. R. Nowak, K. Leiter, J. Knap, and V. Gavini, *Higher-order adaptive finite-element methods for Kohn–Sham density functional theory*, Journal of Computational Physics, 253 (2013), pp. 308–343.
[47] M. T. Nakao, M. Plum, and Y. Watanabe, *Numerical Verification Methods and Computer-Assisted Proofs for Partial Differential Equations*, no. 53 in Springer Series in Computational Mathematics, Springer, Singapore, 2019.
[48] T. Nottoli, I. Giannì, A. Levitt, and F. Lipparini, *A robust, open-source implementation of the locally optimal block preconditioned conjugate gradient for large eigenvalue problems in quantum chemistry*, Theoretical Chemistry Accounts, 142 (2023), p. 69.
[49] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Physical Review Letters, 77 (1996), pp. 3865–3868.
[50] J. P. Perdew and Y. Wang, *Accurate and simple analytic representation of the electron-gas correlation energy*, Physical Review B, 45 (1992), pp. 13244–13249.
[51] W. Prager and J. L. Synge, *Approximations in elasticity based on the concept of function space*, Quarterly of Applied Mathematics, 5 (1947), pp. 241–269.
[52] M. Reed and B. Simon, *Fourier Analysis, Self-Adjointness*, no. 2 in Methods of Modern Mathematical Physics, Academic Press, 1975.

[53] ———, *Analysis of Operators*, no. 4 in Methods of Modern Mathematical Physics, Academic Press, 1978.
[54] J. Solovej, *Proof of the ionization conjecture in a reduced Hartree-Fock model*, Invent. Math., 104 (1991), pp. 291–311.
[55] J. Toulouse, *Review of Approximations for the Exchange-Correlation Energy in Density-Functional Theory*, Springer International Publishing, Cham, 2023, pp. 1–90.
[56] N. Troullier and J. L. Martins, *Efficient pseudopotentials for plane-wave calculations*, Physical Review B, 43 (1991), pp. 1993–2006.
[57] S. Valone, *A one-to-one mapping between one-particle densities and some N-particle ensembles*, The Journal of Chemical Physics, 73 (1980), pp. 4653–4655.
[58] H. F. Weinberger, *Upper and lower bounds for eigenvalues by finite difference methods*, Communications on Pure and Applied Mathematics, 9 (1956), pp. 613–623.
[59] N. Yan and A. Zhou, *Gradient recovery type a posteriori error estimates for finite element approximations on irregular meshes*, Computer Methods in Applied Mechanics and Engineering, 190 (2001), pp. 4289–4299.
[60] B. Yang and A. Zhou, *Eigenfunction behavior and adaptive finite element approximations of nonlinear eigenvalue problems in quantum physics*, ESAIM: Mathematical Modelling and Numerical Analysis, 55 (2021), pp. 209–227.
[61] G. Zhang, L. Lin, W. Hu, C. Yang, and J. E. Pask, *Adaptive local basis set for Kohn–Sham density functional theory in a discontinuous galerkin framework II: Force, vibration, and molecular dynamics calculations*, J. Comput. Phys., 335 (2017), pp. 426–443.

Appendix A. Periodic potentials and Brillouin zone discretization.

In this appendix, we explain how to extend the bounds developed in the rest of the paper to the case of Kohn–Sham equations with periodic potentials and Brillouin zone discretization, as explained in Remark 2.4. Note that, in practice, a finite set of $\boldsymbol{k}$-points is used (typically, a uniform Monkhorst–Pack grid [45]), introducing a numerical quadrature error when integrating over the Brillouin zone. Therefore, all the integrals on the Brillouin zone appearing in this appendix are replaced by quadrature rules: this discretization error is not taken into account here. The interested reader is referred for instance to [13] for the *a priori* numerical analysis of such quadrature methods.

First, let us recall some basic properties of the Bloch–Floquet transform (see for instance [53, Section XIII.16] and [39] for more details). For a generic operator A acting on $L^2(\mathbb{R}^3;\mathbb{C})$ which commutes with $\mathcal{R}$-translations, the Bloch–Floquet transform $\mathcal{Z}$ gives the decomposition

$$\mathcal{Z}^* A \mathcal{Z} = \fint_{\mathcal{B}} A_{\boldsymbol{k}} \mathrm{d}\boldsymbol{k}, \tag{A.1}$$

where $\fint_{\mathcal{B}} = \frac{1}{|\mathcal{B}|}\int_{\mathcal{B}}$ is the averaged integral over the Brillouin zone $\mathcal{B}$ (defined as the first Voronoï cell of the reciprocal lattice $\mathcal{R}^*$), $(A_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{B}}$ is the set of the Bloch fibers of A and each $A_{\boldsymbol{k}}$ is an operator acting on $L^2_{\mathrm{per}}(\Omega)$. If the operator A is locally trace-class, then its *trace per unit cell* is given by

$$\underline{\mathrm{Tr}}(A) = \fint_{\mathcal{B}} \mathrm{Tr}(A_{\boldsymbol{k}}) \mathrm{d}\boldsymbol{k}, \tag{A.2}$$

where here and below Tr denotes the trace of operators on $\mathcal{H} = L^2_{\mathrm{per}}(\Omega)$. We now change the density matrix manifold $\mathcal{M}$ defined in (2.9) to

$$\underline{\mathcal{M}} \coloneqq \Big\{\gamma \in \mathcal{S}(L^2(\mathbb{R}^3)),\ \gamma^2 = \gamma,\ \gamma\tau_{\boldsymbol{R}} = \tau_{\boldsymbol{R}}\gamma\ \forall\ \boldsymbol{R}\in\mathcal{R}, \\ \underline{\mathrm{Tr}}(\gamma) = N_{\mathrm{el}},\ \underline{\mathrm{Tr}}(-\Delta\gamma) < +\infty\Big\}, \tag{A.3}$$

where, for a given vector $\boldsymbol{R} \in \mathcal{R}$, $\tau_{\boldsymbol{R}}$ is the translation operator on $L^2(\mathbb{R}^3;\mathbb{C})$ defined by

$$\forall\ \phi \in L^2(\mathbb{R}^3;\mathbb{C}),\ (\tau_{\boldsymbol{R}}\varphi)(\boldsymbol{x}) = \varphi(\boldsymbol{x}-\boldsymbol{R}) \text{ for a.e. } \boldsymbol{x} \in \mathbb{R}^3,$$

and where $N_{\rm el}$ is the number of electrons *per unit cell*. The conditions $\gamma^2 = \gamma$, $\gamma\tau_{\boldsymbol{R}} = \tau_{\boldsymbol{R}}\gamma$ for all $\boldsymbol{R} \in \mathcal{R}$, and $\underline{\rm Tr}(\gamma) = N_{\rm el}$, imply that the Bloch decomposition of γ is given by

$$\mathcal{Z}^*\gamma\mathcal{Z} = \fint_{\mathcal{B}} \gamma_{\boldsymbol{k}} {\rm d}\boldsymbol{k} \quad \text{with} \quad \gamma_{\boldsymbol{k}} = \sum_{i=1}^{N_{\boldsymbol{k}}} |\phi_{i,\boldsymbol{k}}\rangle\langle\phi_{i,\boldsymbol{k}}|, \tag{A.4}$$

where the function $\mathcal{B} \ni \boldsymbol{k} \mapsto N_{\rm k} \in \mathbb{N}$ is integrable and such that

$$\fint_{\mathcal{B}} N_{\boldsymbol{k}} {\rm d}\boldsymbol{k} = N_{\rm el},$$

and where $(\phi_{i,\boldsymbol{k}})_{1\le i\le N_{\boldsymbol{k}}}$ forms an orthonormal family of $L^2_{\rm per}(\Omega)$. The term $\underline{\rm Tr}(-\Delta\gamma)$ is defined in the same spirit of (2.8). As

$$\mathcal{Z}^*\left(-\Delta\right)\mathcal{Z} = \fint_{\mathcal{B}} (-\imath\nabla + \boldsymbol{k})^2 {\rm d}\boldsymbol{k},$$

we can generalize the notion of trace as in (2.8) and obtain

$$\underline{\rm Tr}(-\Delta\gamma) := \fint_{\mathcal{B}} {\rm Tr}\left((-\imath\nabla+\boldsymbol{k})^2\gamma_{\boldsymbol{k}}\right) {\rm d}\boldsymbol{k} = \fint_{\mathcal{B}} \sum_{i=1}^{N_{\boldsymbol{k}}} \|(-\imath\nabla+\boldsymbol{k})\phi_{i,\boldsymbol{k}}\|^2_{L^2_{\rm per}(\Omega)} {\rm d}\boldsymbol{k}.$$

Next, assuming that the potential V is in $L^2_{\rm per}(\Omega;\mathbb{R})$, we can consider $h = -\frac{1}{2}\Delta + V$ as a self-adjoint operator acting on $L^2(\mathbb{R}^3;\mathbb{C})$, with domain $H^2(\mathbb{R}^3;\mathbb{C})$ and form domain $H^1(\mathbb{R}^3;\mathbb{C})$, that commute with every $\mathcal{R}$-translations and is therefore decomposed by the Bloch transform

$$\mathcal{Z}^* h \mathcal{Z} = \fint_{\mathcal{B}} h_{\boldsymbol{k}} {\rm d}\boldsymbol{k}, \tag{A.5}$$

where $h_{\boldsymbol{k}} = \frac{1}{2}(-\imath\nabla+\boldsymbol{k})^2 + V$ is an operator on $L^2_{\rm per}(\Omega)$ with domain $H^2_{\rm per}(\Omega)$ and form domain $H^2_{\rm per}(\Omega)$. Using the generalized trace notation (2.8), we define

$$\underline{\rm Tr}(h\gamma) := \fint_{\mathcal{B}} {\rm Tr}(h_{\boldsymbol{k}}\gamma_{\boldsymbol{k}}) {\rm d}\boldsymbol{k}, \tag{A.6}$$

and we can write the *energy per unit cell* as

$$\underline{E}(\gamma) := \underline{\rm Tr}(h\gamma) + F(\rho_\gamma), \tag{A.7}$$

where the density ρ_γ is computed as

$$\rho_\gamma(\boldsymbol{x}) = \fint_{\mathcal{B}} \gamma_{\boldsymbol{k}}(\boldsymbol{x},\boldsymbol{x}) {\rm d}\boldsymbol{k} = \fint_{\mathcal{B}} \sum_{i=1}^{N_{\boldsymbol{k}}} |\phi_{i,\boldsymbol{k}}(\boldsymbol{x})|^2 {\rm d}\boldsymbol{k}. \tag{A.8}$$

Note that, by arguments similar to those used in Section 2.3, ρ_γ belongs to $L^2_{\rm per}(\Omega;\mathbb{R})$. The nonlinear term $F(\rho_\gamma)$ is then defined for instance by (2.26) in Kohn–Sham DFT.

With such a framework, one can easily extend the main results to the Kohn–Sham equations with periodic potentials and Brillouin zone discretization. Indeed, we have

$$\text{(A.9)}\quad \forall\ \gamma_1,\gamma_2 \in \underline{\mathcal{M}},\ \langle F'(\rho_{\gamma_1}),\rho_{\gamma_2}\rangle = \int_\Omega V_{\rho_{\gamma_1}}\rho_{\gamma_2} = \fint_{\mathcal{B}} \mathrm{Tr}(V_{\rho_{\gamma_1}}\gamma_{2,\boldsymbol{k}})\mathrm{d}\boldsymbol{k} = \underline{\mathrm{Tr}}(V_{\rho_{\gamma_1}}\gamma_2).$$

The Bloch fibers of the Kohn–Sham Hamiltonian are then given by $H_{\rho,\boldsymbol{k}} = h_{\boldsymbol{k}} + V_\rho = \frac{1}{2}(-\imath\nabla + \boldsymbol{k})^2 + V + V_\rho$, and have domain $H^2_{\mathrm{per}}(\Omega)$ and form domain $H^1_{\mathrm{per}}(\Omega)$. Note that the Kohn–Sham equations with periodic potentials (2.21) are obtained as the Euler–Lagrange equations of the minimisation problem

$$\text{(A.10)}\qquad \min_{\gamma\in\underline{\mathcal{M}}} \underline{E}(\gamma).$$

Existence and uniqueness of a minimizer to this problem is studied for instance in [6, Theorem 1] for the rHF model. A crucial assumption in our approach is that the band gap ν is positive, where $\nu = \min_{\boldsymbol{k}\in\mathcal{B}} \varepsilon_{N_{\boldsymbol{k}}+1,\boldsymbol{k}} - \max_{\boldsymbol{k}\in\mathcal{B}} \varepsilon_{N_{\boldsymbol{k}},\boldsymbol{k}}$ with $(\varepsilon_{i,\boldsymbol{k}})_{i\in\mathbb{N}}$ the eigenvalues of the Bloch fiber $H_{\rho_\gamma,\boldsymbol{k}}$. We make this assumption in the sequel. From a physical point of view, this means that the system under consideration is an *insulator*. In this case, it holds $N_{\boldsymbol{k}} = N_{\mathrm{el}}$ for a.e. $\boldsymbol{k}\in\mathcal{B}$.

Assuming that F is convex, everything then follows similarly to the computations from Section 3 by (i) replacing the $L^2_{\mathrm{per}}(\Omega)$ trace Tr with the trace per unit cell $\underline{\mathrm{Tr}}$ and (ii) replacing the real number $\mu\in\mathbb{R}$ with a function $\mu : \mathcal{B}\ni\boldsymbol{k}\mapsto\mu_{\boldsymbol{k}}\in\mathbb{R}$. Then, for fixed $\gamma_1,\gamma_2\in\underline{\mathcal{M}}$, if we are able to find a function μ such that,

$$\text{(A.11)}\qquad \underline{\mathrm{Tr}}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) \geq 0,$$

then the extension of Corollary 3.3 to the energy per unit cell reads

$$\text{(A.12)}\qquad \underline{E}(\gamma_2) - \underline{E}(\gamma_1) \leq \underline{\mathrm{Tr}}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_2\big).$$

In order for (A.11) to hold, since

$$\text{(A.13)}\qquad \underline{\mathrm{Tr}}\big((h + V_{\rho_{\gamma_2}} - \mu)\gamma_1\big) = \fint_{\mathcal{B}} \mathrm{Tr}\big((h_{\boldsymbol{k}} + V_{\rho_{\gamma_2}} - \mu_{\boldsymbol{k}})\gamma_{1,\boldsymbol{k}}\big)\mathrm{d}\boldsymbol{k},$$

it is sufficient to compute $(\mu_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathcal{B}}$ such that

$$\text{(A.14)}\qquad \forall\ \boldsymbol{k}\in\mathcal{B},\ \mathrm{Tr}\big((h_{\boldsymbol{k}} + V_{\rho_{\gamma_2}} - \mu_{\boldsymbol{k}})\gamma_{1,\boldsymbol{k}}\big) \geq 0.$$

Therefore, one can apply the strategy presented in Section 3.2 to each Bloch fibers $h_{\boldsymbol{k}} + V_{\rho_{\gamma_2}}$ to compute an admissible function μ. This ultimately amounts to compute (or approximate) an $\eta_{\boldsymbol{k}}$ from (3.14) for every $\boldsymbol{k}$ in $\mathcal{B}$. Everything we presented to compute this quantity can therefore be applied for every $\boldsymbol{k}$ in $\mathcal{B}$, or in practice for every point of the grid used to discretize the integrals over $\mathcal{B}$.

From a practical point of view, the Kohn–Sham equation with Brillouin zone discretization (2.21) are also solved with SCF algorithms, except that each Bloch fiber $H_{\rho,\boldsymbol{k}}$ has to be diagonalized and the density rebuilt by integrating over the Brillouin zone. A similar splitting between discretization error and SCF error, as in (3.36), can thus also be derived: the guaranteed bound on the energy per unit cell then reads, at iteration m of the SCF in the space $\mathcal{V}_N$,

$$\text{(A.15)}\qquad \boxed{\underline{E}(\gamma_{N,m}) - \underline{E}(\gamma_\star) \leq \mathtt{err}^{\mathrm{disc}}_{N,m} + \mathtt{err}^{\mathrm{SCF}}_{N,m},}$$

where each error component has to be computed fiber-wise. In other words,

$$
\begin{aligned}
\mathtt{err}_{N,m}^{\mathrm{disc}} &= \underline{\mathrm{Tr}}\big((H_{\rho_{\gamma_{N,m}}} - \mu_{N,m+1}^{\mathrm{lb}})\gamma_{N,m+1}\big) \\
&= \fint_{\mathcal{B}} \mathrm{Tr}\big((H_{\rho_{\gamma_{N,m}},\boldsymbol{k}} - \mu_{N,m+1,\boldsymbol{k}}^{\mathrm{lb}})\gamma_{N,m+1,\boldsymbol{k}}\big)\mathrm{d}\boldsymbol{k}, \\
\mathtt{err}_{N,m}^{\mathrm{SCF}} &= \underline{\mathrm{Tr}}\big(H_{\rho_{\gamma_{N,m}}}\gamma_{N,m}\big) - \underline{\mathrm{Tr}}\big(H_{\rho_{\gamma_{N,m}}}\gamma_{N,m+1}\big) \\
&= \fint_{\mathcal{B}} \mathrm{Tr}\big(H_{\rho_{\gamma_{N,m}},\boldsymbol{k}}\gamma_{N,m,\boldsymbol{k}}\big) - \mathrm{Tr}\big(H_{\rho_{\gamma_{N,m}},\boldsymbol{k}}\gamma_{N,m+1,\boldsymbol{k}}\big)\mathrm{d}\boldsymbol{k}.
\end{aligned}
\tag{A.16}
$$

Appendix B. Neumann series truncation error.

In this section we provide a way to estimate $\|H_0^{-1}W\|$. As before we make use of the decomposition $\mathcal{H} = \mathcal{V}_N \oplus \mathcal{V}_N^{\perp}$, and write

$$
\|H_0^{-1}W\| \leq \|H_0^{-1}\Pi_N W\| + \|H_0^{-1}\Pi_N^{\perp}W\|. \tag{B.1}
$$

The second term can be easily estimated as follows:

$$
\begin{aligned}
\|H_0^{-1}\Pi_N^{\perp}W\| \leq& \left\| \left(-\frac{1}{2}\Delta + \langle V\rangle\right)^{-1} \Pi_N^{\perp}(V - \langle V\rangle)\Pi_N^{\perp} \right\| \\
&+ \left\| \left(-\frac{1}{2}\Delta + \langle V\rangle\right)^{-1} \Pi_N^{\perp}V\Pi_N \right\| \\
\leq& \left\| \left(-\frac{1}{2}\Delta + \langle V\rangle\right)^{-1} \Pi_N^{\perp} \right\| \Big(\big\|(V - \langle V\rangle)\Pi_N^{\perp}\big\| + \|V\Pi_N\| \Big) \\
\leq& \frac{1}{\mathrm{E}_{\mathrm{cut}} + \langle V\rangle} \Big(\|V - \langle V\rangle\|_{L^\infty} + \|V\|_{L^\infty} \Big).
\end{aligned}
\tag{B.2}
$$

The final inequality is a consequence of the Fourier representation of the operator $-\frac{1}{2}\Delta + \langle V\rangle$ and the definition (2.31) of the subspace $\mathcal{V}_N$.

We now proceed to estimate the first term. To this end let us denote by $\pi_{N_{\mathrm{el}}}$ the orthogonal projection on the vector space generated by the approximate eigenvectors $\mathcal{W}_{N_{\mathrm{el}}} = \mathrm{Span}(\varphi_{i,N})_{i=1,\cdots,N_{\mathrm{el}}} \subset \mathcal{V}_N$ and $\pi_{N_{\mathrm{el}}}^{\perp}$ as the corresponding orthogonal projection in $\mathcal{V}_N$. Introducing $A_N = \Pi_N A \Pi_N = \Pi_N H_0 \Pi_N$, we can estimate $\|H_0^{-1}\Pi_N W\|$ by

$$
\begin{aligned}
\|H_0^{-1}\Pi_N W\| =& \|A_N^{-1}\Pi_N V\Pi_N^{\perp}\| \\
=& \| \left(\pi_{N_{\mathrm{el}}}A_N^{-1}\pi_{N_{\mathrm{el}}} + \pi_{N_{\mathrm{el}}}^{\perp}A_N^{-1}\pi_{N_{\mathrm{el}}}^{\perp}\right)\Pi_N V\Pi_N^{\perp}\| \\
\leq& \|\pi_{N_{\mathrm{el}}}A_N^{-1}\pi_{N_{\mathrm{el}}}\Pi_N V\Pi_N^{\perp}\| + \|\pi_{N_{\mathrm{el}}}^{\perp}A_N^{-1}\pi_{N_{\mathrm{el}}}^{\perp}\Pi_N V\Pi_N^{\perp}\| \\
\leq& \|\pi_{N_{\mathrm{el}}}A_N^{-1}\pi_{N_{\mathrm{el}}}\Pi_N V\Pi_N^{\perp}\| + \|\pi_{N_{\mathrm{el}}}^{\perp}A_N^{-1}\pi_{N_{\mathrm{el}}}^{\perp}\|\|\Pi_N V\Pi_N^{\perp}\| \\
\leq& \|\pi_{N_{\mathrm{el}}}A_N^{-1}\pi_{N_{\mathrm{el}}}\Pi_N V\Pi_N^{\perp}\| + \frac{\|V\|_{L^\infty}}{\varepsilon_{N_{\mathrm{el}},N}}.
\end{aligned}
\tag{B.3}
$$

The first term can be further computed in terms of the residuals from the eigendecomposition (3.8). Indeed, since $\Pi_N^{\perp}\varphi_{i,N} = 0$, we have $\Pi_N^{\perp}r_{i,N} = \Pi_N^{\perp}A\varphi_{i,N}$ from which

we can write, using in addition that $\Pi_N \Delta \Pi_N^\perp = 0$,

$$\begin{aligned}
\|\pi_{N_{\mathrm{el}}} A_N^{-1} \pi_{N_{\mathrm{el}}} \Pi_N V \Pi_N^\perp\| &= \left\| \sum_{i=1}^{N_{\mathrm{el}}} \frac{|\varphi_{i,N}\rangle\langle\varphi_{i,N}|}{\varepsilon_{i,N}} \Pi_N V \Pi_N^\perp \right\| \\
&= \left\| \sum_{i=1}^{N_{\mathrm{el}}} \frac{|\varphi_{i,N}\rangle\langle\varphi_{i,N}|}{\varepsilon_{i,N}} \Pi_N \left(-\frac{1}{2}\Delta + V \right) \Pi_N^\perp \right\| \\
&= \left\| \sum_{i=1}^{N_{\mathrm{el}}} \frac{|\varphi_{i,N}\rangle\langle\varphi_{i,N}|}{\varepsilon_{i,N}} \Pi_N A \Pi_N^\perp \right\| \\
&= \left\| \sum_{i=1}^{N_{\mathrm{el}}} \frac{|\varphi_{i,N}\rangle\langle r_{i,N}|}{\varepsilon_{i,N}} \right\|.
\end{aligned} \tag{B.4}$$

Since the family $(\varphi_{i,N})_{i=1,\dots,N_{\mathrm{el}}}$ is L^2-orthonormal, the matrix

$$X = \sum_{i=1}^{N_{\mathrm{el}}} \frac{|\varphi_{i,N}\rangle\langle r_{i,N}|}{\varepsilon_{i,N}} \in \mathbb{R}^{N \times N}$$

satisfies

$$X^* X = \sum_{i=1}^{N_{\mathrm{el}}} \frac{|r_{i,N}\rangle\langle r_{i,N}|}{(\varepsilon_{i,N})^2} \in \mathbb{R}^{N \times N}.$$

The operator norm $\|X\|$ in (B.4) is thus given by $\sqrt{\rho(X^* X)}$, where ρ denotes the spectral radius. Finally, this norm is actually computable for a negligible computational cost (even for a large basis size N) as the square root of the highest eigenvalue of the smaller matrix $R^* R \in \mathbb{R}^{N_{\mathrm{el}} \times N_{\mathrm{el}}}$ with

$$\mathbb{R}^{N \times N_{\mathrm{el}}} \ni R = \begin{bmatrix} r_{1,N} & r_{2,N} & \cdots & r_{N_{\mathrm{el}},N} \end{bmatrix} \times \begin{bmatrix} 1/\varepsilon_{1,N} & & & \\ & 1/\varepsilon_{2,N} & & \\ & & \ddots & \\ & & & 1/\varepsilon_{N_{\mathrm{el}},N} \end{bmatrix},$$

where the matrix on the right is diagonal.