A posteriori error estimation for the non-self-consistent Kohn-Sham equations

Michael F. Herbst,* Antoine Levitt,† and Eric Cancès‡
CERMICS, Ecole des Ponts and Inria Paris, 6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée, France

We address the problem of bounding rigorously the errors in the numerical solution of the Kohn-Sham equations due to (i) the finiteness of the basis set, (ii) the convergence thresholds in iterative procedures, (iii) the propagation of rounding errors in floating-point arithmetic. In this contribution, we compute fully-guaranteed bounds on the solution of the non-self-consistent equations in the pseudopotential approximation in a plane-wave basis set. We demonstrate our methodology by providing band structure diagrams of silicon annotated with error bars indicating the combined error.

I. INTRODUCTION

Experimental results are provided with error bars in most scientific fields. In an ideal world, this should also be the case for results obtained by numerical simulation. Complementing simulation results with error bars is becoming mandatory in some branches of engineering, such as aeronautics or car industry, in which simulation has partially, or even totally, replaced experiment (e.g. virtual wind tunnels [1] or crash simulators [2]). Obviously, uncontrolled errors in the numerical simulations used to design and test an aircraft are likely to have dramatic consequences for passengers and crews.

What about molecular simulation? *Statistical* error bars are often displayed in molecular dynamics (MD) and quantum Monte Carlo (QMC), where stochastic processes (e.g. Langevin equation for MD, drift-diffusion stochastic differential equations for QMC) are at the core of the simulation. However, these error bars are purely statistical in nature and are not guaranteed: they only reflect the finiteness of the statistical sample and the actual solution of the model has a positive probability to lay outside the confidence interval. In addition, they only take into account one of the components of the error between the exact value of the quantity of interest (q.o.i.) and its numerical approximation. The other components of the error are deterministic in nature and are also present in simulations of purely deterministic models, such as those based on density-functional theory (DFT) which are dealt with in the article.

Consider as an example of q.o.i. the lattice constant of Silicon at 0 K. In order to compute a numerical approximation of this q.o.i., we first have to select a model. We have at our disposal an outstanding model: the many-body Schrödinger equation with relativistic corrections. However, solving this equation directly is completely out of reach. The first approximation is to replace this reference model with a cruder, but tractable, reduced model. We consider as a reduced model the one obtained by successively using the Born-Oppenheimer approximation to decouple as much as possible nuclear and electronic degrees of freedom, and the Local Density Approximation of the Kohn-Sham density-functional theory (KS-LDA), together with a pseudopotential model to avoid representing core electrons and the singularities of the orbitals near atoms. This gives rise to a well-defined mathematical model, hopefully with a unique solution. We refer to the difference between this approximation and the physical reality as the *model error*.

The reduced model, though simpler than the reference model, still has an infinite number of degrees of freedom, and must be discretized to be simulated. For the crystalline phase, a typical (simplified) workflow is as follows [3]:

1. The infinite computational domain is truncated to a finite supercell with periodic boundary conditions, numerically handled through Brillouin zone sampling;
2. In this supercell, crystalline orbitals are discretized using a finite basis set of plane waves;
3. The self-consistent Kohn-Sham equations are solved by a self-consistent field (SCF) algorithm;
4. At each step of the SCF algorithm, a linear eigenvalue equation is solved with an iterative eigensolver;
5. All computations are performed with finite precision.

We refer to the error in steps 1 and 2 as the discretization error, to the error in steps 3 and 4 as the algorithmic error, and to the error in step 5 as the arithmetic error. To these must be added programming errors, not to be neglected in codebases consisting of millions of lines of code, and hardware errors, which are expected to become significant for exascale architectures [4]. These two latter kinds of errors we will not treat.

Note that all steps mentioned above are systematically improvable: by increasing parameters, such as Brillouin zone sampling, basis set cutoff, convergence thresholds, or using higher-precision arithmetic, results get more accurate at the price of longer computation times. The usual procedure for controlling these errors is to perform convergence studies: on the system of interest or a related system with similar characteristics, the parameter is increased until the variation of the quantity of interest is below a specified tolerance. Over time, knowledge of "good" parameter values solidifies into rules of thumb that are automated in codes or suggested to users in manuals. When used appropriately, this results in acceptable errors that are below the target accuracy (for instance, for pseudopotential methods, the error compared to all-electron models) [5, 6].

This empirical process is however still problematic. First, it might result in suboptimal performance by excess of caution. Second, it still requires a degree of hand-tuning and is thus problematic for fully automated computations, which are useful for *in silico* design of novel compounds and for building databases of material or chemical properties. Third, the rules of thumb

* michael.herbst@inria.fr
† antoine.levitt@inria.fr
‡ eric.cances@enpc.fr

can fail for unusual systems with unexpected behavior, i.e. exactly those where accurate simulations are important to aid understanding.

The purpose of *a posteriori* error analysis is to automate and rationalize this process by providing accurate, computable and guaranteed bounds on the error of each step. The purpose of this is twofold. First, bounds of the total error on the q.o.i. can be obtained by simply summing up the different components of the error, allowing one to bound the accuracy of the final result. Second, the computer resources necessary to reach a given accuracy on the final result can be optimized by *error balancing* techniques. For instance, convergence thresholds should be adapted to the discretization: if the discretization is coarse, it is not necessary to use extremely tight convergence criteria, since the final accuracy will be limited by the discretization error anyway. For the same reason, it might suffice to perform most operations in single-precision arithmetic to save computational time without losing much on the accuracy of the final result. Error balancing based on *a posteriori* error bounds allows one to turn these common sense remarks into black-box numerical strategies. The user provides the target accuracy and the software automatically chooses, in an adaptive way along the iterative process, the reduced models, discretization bases, convergence thresholds, and data structures to obtain the desired accuracy at a quasi-optimal computational cost.

Applying this methodology to electronic structure calculations is a challenge due to the complexity of the equations. Despite this, significant progress has been made in the past decade towards rigorous error control, which makes this perspective realistic in the medium term. Let us mention in particular recent works on *a priori* and *a posteriori* discretization error bounds for DFT [7–13], including k-point sampling [14], and on the numerical analysis of SCF algorithms [15–18].

In this contribution, we provide for the first time a combined analysis of the errors in step 2 (basis set truncation), 4 (inexact solution of the eigenvalue problem) and 5 (arithmetic error). The q.o.i. is the band diagram of silicon, or more precisely the energies of the 3 highest occupied and 4 lowest unoccupied crystalline orbitals for specific k-points. The models under consideration are the Cohen-Bergstresser model [19] on the one-hand, and the non-self-consistent (electron-electron interaction completely neglected) periodic KS-LDA model with GTH pseudopotentials [20, 21] on the other hand, for which we will give fully guaranteed error bounds.

Some aspects of our analysis, such as the computation of the residual, are general and can be extended to other quantities of interest, basis sets and models; others, such as the gap estimation, rely more specifically on properties of the setup considered. We refer to the conclusion for more details on this point. Our analysis is, however, limited in its present state due to the neglect of the errors in steps 1 (Brillouin zone sampling) and 3 (self-consistency). We hope to address both these challenging aspects in future publications.

Let us finally point out that the model error arising from the choice of density-functional theory is not easily systematically improvable. For wavefunctions methods (e.g. coupled-cluster methods [22]) guaranteed *a posteriori* error bounds on the model error — with respect to the many-body Schrödinger equation (MBSE) — can in principle be derived, since a residual of the MBSE can be computed from the approximate wavefunction and energy. For DFT, on the other hand, this is not easily the case such that guaranteed model errors are probably out of reach and going beyond machine-learned confidence intervals obtained from big data sets of reference calculations seems very difficult. Regarding pseudopotentials and related approaches, the Projector Augmented Wave [23] (PAW) method should be amenable to the derivation of guaranteed error bounds - with respect to reference full-electron DFT calculations - since approximate all-electron Kohn-Sham orbitals can be reconstructed from PAW pseudo-orbitals allowing the construction of a residual.

II. KOHN-SHAM DENSITY-FUNCTIONAL THEORY IN THE PSEUDOPOTENTIAL APPROXIMATION

As already mentioned in the introduction, the q.o.i. under consideration are the energies of crystalline orbitals at specific $\mathbf{k}$-points, in a non-self-consistent pseudopotential model. In the following we denote by $\mathscr{R}$ the periodic lattice, and assume that the Bloch wavevector $\mathbf{k}$ is fixed in the Brillouin zone. The equation we solve is

$$Hu = \varepsilon u, \quad \int_\Omega |u(\mathbf{r})|^2 d\mathbf{r} = 1, \tag{1}$$

with u an $\mathscr{R}$-periodic function and Ω the unit cell. The periodic Hamiltonian is given by

$$H = \frac{1}{2}(-i\nabla + \mathbf{k})^2 + V,$$

where the (possibly nonlocal) effective potential V is assumed to be known.

The problem (1) is discretized in a Fourier basis: any $\mathscr{R}$-periodic function can be expanded on the orthonormal plane-wave basis

$$e_{\mathbf{G}}(\mathbf{r}) = \frac{1}{\sqrt{|\Omega|}} e^{i\mathbf{G}\cdot\mathbf{r}}$$

where $|\Omega|$ is the volume of the unit cell, and $\mathbf{G}$ belongs to the reciprocal lattice $\mathscr{R}^*$. The complete set $(e_{\mathbf{G}})_{\mathbf{G}\in\mathscr{R}^*}$ is truncated to obtain a finite approximation space

$$X = \mathrm{Span}\left\{ e_{\mathbf{G}}, \frac{1}{2}|\mathbf{G}+\mathbf{k}|^2 \le E_{\mathrm{cut}} \right\}$$

with dimension $N_b := \dim(X)$, for some finite cutoff energy $E_{\mathrm{cut}} > 0$. We will use the convenient abuse of notation consisting in writing $\mathbf{G} \in X$ to denote a $\mathbf{G} \in \mathscr{R}^*$ such that $e_{\mathbf{G}} \in X$. The linear eigenvalue equation (1) is then discretized by invoking the variational principle: the $N_b \times N_b$ complex Hermitian matrix

$$\langle e_{\mathbf{G}}|\, H\, |e_{\mathbf{G}'}\rangle = \frac{1}{2}|\mathbf{G}+\mathbf{k}|^2 \delta_{\mathbf{G}\mathbf{G}'} + \langle e_{\mathbf{G}}|V|e_{\mathbf{G}'}\rangle, \quad \mathbf{G},\mathbf{G}' \in X,$$

is diagonalized using an iterative eigensolver [24, 25], employing fast Fourier transform to perform matrix-vector products efficiently.

The potential V can be a purely local potential, or have a nonlocal component in the case of norm-conserving pseudopotentials. We will consider two cases here, chosen for their simple analytic forms.

First we consider the Cohen-Bergstresser pseudopotentials for semiconductors in the diamond and zinc-blende structure [19]. These are extremely purely local $\mathscr{R}$-periodic potentials with a small number of non-zero Fourier coefficients. More precisely

$$\langle e_{\mathbf{G}}|V|e_{\mathbf{G}'}\rangle = \frac{1}{|\Omega|}\widehat{v}_{\mathrm{CB}}(\mathbf{G}-\mathbf{G}') \qquad \text{(Cohen-Bergstresser)}, \quad (2)$$

where $\widehat{v}_{\mathrm{CB}}(\Delta\mathbf{G})$ is only nonzero for $\Delta\mathbf{G}$ in a small finite set (the first five shells of reciprocal vectors). The coefficients $\widehat{v}_{\mathrm{CB}}(\Delta\mathbf{G})$ were adjusted to reproduce spectroscopic data.

We next consider more realistic Goedecker–Teter–Hutter (GTH) norm-conserving potentials [20, 21]. These potentials are composed of a local and a non-local part:

$$V = V_{\mathrm{loc}} + V_{\mathrm{nl}} \qquad \text{(Goedecker–Teter–Hutter, GTH)},$$

where

$$\begin{aligned}\langle e_{\mathbf{G}}|V_{\mathrm{loc}}|e_{\mathbf{G}'}\rangle &= \frac{1}{|\Omega|}\widehat{v}_{\mathrm{loc}}(\mathbf{G}-\mathbf{G}'),\\ \langle e_{\mathbf{G}}|V_{\mathrm{nl}}|e_{\mathbf{G}'}\rangle &= \sum_a \sum_{lm} \sum_{ij} d_{almij}\, p_{almi}(\mathbf{k}+\mathbf{G})\overline{p_{almj}(\mathbf{k}+\mathbf{G}')}. \quad (3)\end{aligned}$$

Here a runs over all the atoms in the unit cell, the range of angular momentum l depends on the chemical element considered, $m = -l,\dots,l$ and the projection indices i, j are summed over a small number of integers (two in the case of silicon). The coefficients $\widehat{v}_{\mathrm{loc}}(\Delta\mathbf{G})$ are finite sums of products of three terms: a structure phase factor depending on the position of the atoms in the unit cell, a radial Gaussian envelope and a radial rational function. The $p_{almi}(\mathbf{q})$ are the product of a structure phase factor, a spherical harmonics Y_{lm}, a radial Gaussian envelope and a radial polynomial of degree l. Among the many types of pseudopotentials available, we chose this type because their analytic form (rather than tabulated data) lends itself well to rigorous error analysis.

All computations presented in this paper have been performed using the density-functional toolkit (DFTK) [26], a recent Julia [27] implementation for Kohn-Sham DFT and related methods using plane-wave basis sets. Our implementation of the discussed estimates and the code producing the figures of this paper is available on Github [28].

III. FROM RESIDUALS TO ERRORS

Assume for simplicity that ε is an isolated eigenvalue of H. Given a finite E_{cut}, the result of the iterative procedure is a normalized vector $\widetilde{u} \in \mathscr{H}$ and a Riesz approximation $\widetilde{\varepsilon} = \langle \widetilde{u}|H|\widetilde{u}\rangle$ of the eigenvalue $\widetilde{\varepsilon}$, such that the algebraic residual $P_X(H\widetilde{u} - \widetilde{\varepsilon}\widetilde{u})$, where P_X is the orthogonal projector on X, is small. Note that the algebraic residual vanishes for an exact matrix eigensolver, in contrast with the global residual $\widetilde{r} = H\widetilde{u} - \widetilde{\varepsilon}\widetilde{u}$, which is always nonzero due to the finite basis discretization. The question we are interested in is: can we bound rigorously the error between $\widetilde{\varepsilon}$ and ε?

Several *a posteriori* error bounds for elliptic eigenvalue problems are available in the literature [29–37], the most accurate of them requiring lower bounds on the distance between the eigenvalue or cluster of eigenvalues[13, 38–42] of interest and the rest of the spectrum. We focus here on two basic bounds for the sake of simplicity: the implementation of more accurate ones is work in progress.

Theorem 1 (Bauer-Fike)**.** There exists an eigenvalue ε of H such that

$$|\widetilde{\varepsilon} - \varepsilon| \leq \|\widetilde{r}\|.$$

This bound is very easy to handle: it only requires an upper bound on the $\mathscr{H}$-norm of the residual. It is however not sharp. In particular, it is well-known that the error on isolated eigenvalues of Hermitian matrices behaves as the square of the residual. An *a posteriori* error bound having this property is the following.

Theorem 2 (Kato-Temple)**.** Let ε be the eigenvalue of H closest to $\widetilde{\varepsilon}$. Assume that ε is an isolated point of the spectrum of H with a distance $\delta > 0$ to the rest of the spectrum. Then

$$|\widetilde{\varepsilon} - \varepsilon| \leq \frac{\|\widetilde{r}\|^2}{\delta}. \quad (4)$$

The proof of both these theorems can be found in standard textbooks [24].

To apply these bounds, we need three ingredients: (i) construct an upper bound on the residual $\|\widetilde{r}\|$, (ii) construct a lower bound on the gap δ, (iii) implement these bounds in the presence of roundoff errors. We address these problems in sequence.

IV. COMPUTING THE RESIDUAL

Assuming that the variational numerical method provides an approximate normalized eigenfunction

$$\widetilde{u} = \sum_{\mathbf{G}\in X} \widetilde{u}(\mathbf{G}) e_{\mathbf{G}} \in X,$$

with

$$\widetilde{u}(\mathbf{G}) := \langle e_{\mathbf{G}}|\widetilde{u}\rangle \quad \text{and} \quad \|\widetilde{u}\|^2 = \sum_{\mathbf{G}\in X} |\widetilde{u}(\mathbf{G})|^2 = 1,$$

and a Riesz approximation $\widetilde{\varepsilon} = \langle \widetilde{u}|H|\widetilde{u}\rangle$ of the associated eigenvalue, then the square norm of the residual

$$\widetilde{r} = H\widetilde{u} - \widetilde{\varepsilon}\widetilde{u}$$

can be decomposed as

$$\begin{aligned}\|\widetilde{r}\|^2 &= \|P_X\widetilde{r}\|^2 + \|P_{X^\perp}\widetilde{r}\|^2\\ &= \|P_X(H\widetilde{u} - \widetilde{\varepsilon}\widetilde{u})\|^2 + \|P_{X^\perp}V\widetilde{u}\|^2\end{aligned} \quad (5)$$

where P_X and $P_{X^\perp}$ are the orthogonal projectors on X and $X^\perp$ respectively. Here we have used that the kinetic energy operator is diagonal in reciprocal space, so that $P_{X^\perp}H\widetilde{u} = P_{X^\perp}V\widetilde{u}$. The first term, which is easily computed in the Fourier basis of X, is the square of the norm of the in-space, algebraic residual $P_X(H\widetilde{u} - \widetilde{\varepsilon}\widetilde{u})$. It is driven to zero by the iterative eigensolver but does not vanish because the iterations stop when the convergence thresholds are reached. This is the origin of the algorithmic error, and is easily computed explicitly. The second term is the out-of-space residual and is the source of the discretization error. This term is more difficult to compute, and most often in practice only an upper bound can be obtained at a reasonable computational cost.

A. Cohen-Bergstresser model

In this model, V is given by (2), and only a small number of terms $\widetilde{v}_{\mathrm{CB}}(\Delta\mathbf{G})$ are non-zero. In this case

$$\|P_{X^\perp} V\widetilde{u}\|^2 = \frac{1}{|\Omega|^2} \sum_{\mathbf{G}\in X^\perp} \left| \sum_{\mathbf{G}'\in X} \widehat{v}_{\mathrm{CB}}(\mathbf{G}-\mathbf{G}')\widetilde{u}(\mathbf{G}') \right|^2 .$$

This computation extends over a finite range of $\mathbf{G}$. Denoting by $G_{\max}$ the norm of the largest non-zero Fourier mode of $\widehat{v}_{\mathrm{CB}}$, these can be identified to be of the form $\mathbf{G}+\Delta\mathbf{G}$ for $\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \leq E_{\mathrm{cut}}$ and $|\Delta\mathbf{G}| \leq G_{\max}$. It follows that $P_{X^\perp} V\widetilde{u}$ belongs to the finite-dimensional space

$$Y = \mathrm{Span}\left\{ e_{\mathbf{G}}, \mathbf{G}\in\mathscr{R}^*, \frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \leq E_{\mathrm{cut}}^{(2)} \right\} \tag{6}$$

with $E_{\mathrm{cut}}^{(2)} = \frac{1}{2}\left(\sqrt{2E_{\mathrm{cut}}}+G_{\max}\right)^2$. We therefore extend $\widetilde{u}$ to this new basis Y by zero-padding, and compute $V\widetilde{u}$ in this new basis, resulting in an exact computation of the residual. The very quick decay of the residual with increasing E_{cut} is shown in Figure 1 for the first eigenvalue at the Γ point.

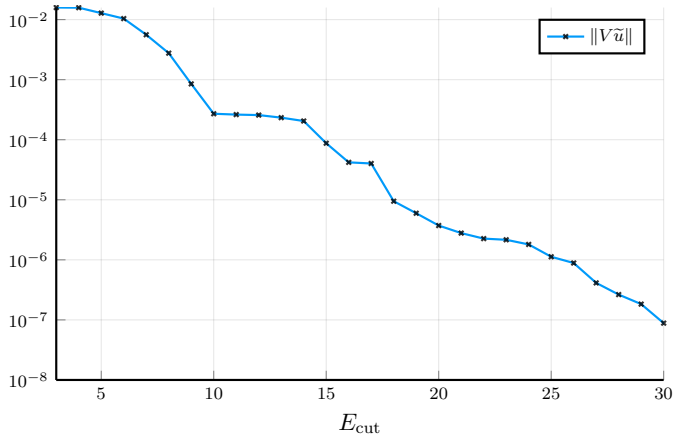

FIG. 1. Exact residual norm for the first eigenvalue at the Γ-point of the Cohen-Bergstresser model of silicon.

B. Goedecker–Teter–Hutter (GTH) pseudopotentials

With the GTH pseudopotentials, $H\widetilde{u}$ extends on all wave vectors in $\mathscr{R}^*$, and therefore we cannot compute $\|P_{X^\perp} V\widetilde{u}\|^2$ explicitly. Rather, as before, we compute $\|P_Y V\widetilde{u}\|^2$, where the larger approximation subspace Y, still defined by (6) is determined by a chosen $E_{\mathrm{cut}}^{(2)} > E_{\mathrm{cut}}$. We then bound the remaining term

$$\|P_{Y^\perp} V\widetilde{u}\|^2 = \sum_{\mathbf{G}\in Y^\perp} \left| \sum_{\mathbf{G}'\in X} \langle e_{\mathbf{G}}|\, V\, |e_{\mathbf{G}'}\rangle \widetilde{u}(\mathbf{G}') \right|^2 .$$

We do this by exploiting the fact that the matrix elements $\langle e_{\mathbf{G}}|V|e_{\mathbf{G}'}\rangle$ connecting small wave vectors in $\mathbf{G}'\in X$ to large wave vectors in $\mathbf{G}\in Y^\perp$ are small. We obtain an explicit bound using our knowledge of V and the decay properties of Gaussians, see Appendix A for details.

Our rigorous bound of the residual is plotted Figure 2 along with its different components as a function of $E_{\mathrm{cut}}^{(2)}$, for an initial $\widetilde{u}$ obtained with $E_{\mathrm{cut}} = 20$. As can be seen clearly, the residual

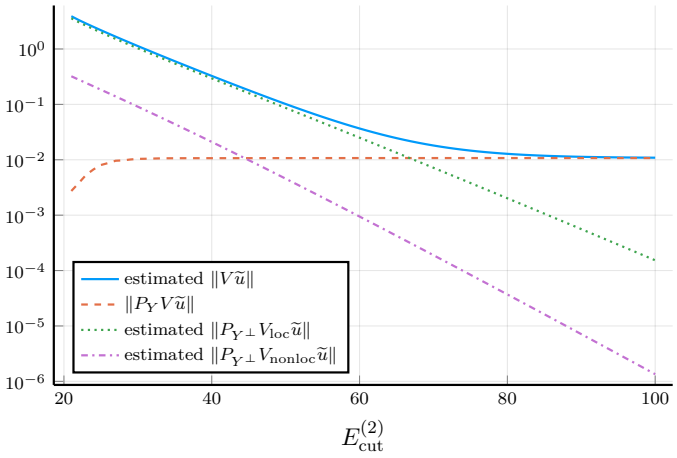

FIG. 2. Bound on the residual $\|P_{X^\perp} V\widetilde{u}\|$ and its three components for the first eigenvalue at the Γ point of the GTH model of Silicon as a function of $E_{\mathrm{cut}}^{(2)}$. The initial $\widetilde{u}$ is obtained at $E_{\mathrm{cut}} = 20$.

is essentially in Y as soon as $E_{\mathrm{cut}}^{(2)} \geq 30$, our estimate, however, is only accurate for larger values of $E_{\mathrm{cut}}^{(2)} \approx 80$. This is since our bounds for $\|P_{Y^\perp} V_{\mathrm{loc}}\widetilde{u}\|$ are still rather crude (see Appendix A). Improving these is work in progress.

In theory, one could choose $E_{\mathrm{cut}}^{(2)}$ dynamically based on the size of the estimated residual. In the following, we use the simple heuristic $E_{\mathrm{cut}}^{(2)} = 4E_{\mathrm{cut}}$, deduced from $E_{\mathrm{cut}} = 20$ and Figure 2. As can be seen in Figure 3, $E_{\mathrm{cut}} = 20$ represented the worst case for our heuristic. With this choice of $E_{\mathrm{cut}}^{(2)}$ the component on $Y^\perp$ of the residual is negligible for all values of E_{cut} and our bound is nearly optimal.

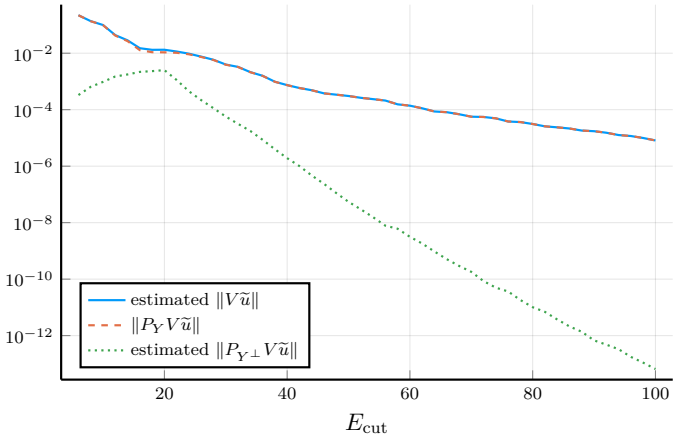

FIG. 3. Computed and bounded terms of the residual norm for the first band at the Γ-point of the GTH model of silicon, using $E_{\mathrm{cut}}^{(2)} = 4E_{\mathrm{cut}}$. The orange dashed and the solid blue curve are almost superimposed, indicating that the bounded $\|P_{Y^\perp} V\widetilde{u}\|$ term is negligible.

V. ESTIMATING THE GAP

The Kato-Temple bound (4) requires a lower bound on the gap δ between the (unknown) exact eigenvalue ε and the rest of the (unknown) spectrum of H. Let N be the index of the target eigenvalue. Given an operator H with eigenvalues $\varepsilon_1 \leq \varepsilon_2 \leq \ldots$, how can we obtain a lower bound on both $\varepsilon_{N+1} - \varepsilon_N$ in the one direction and on $\varepsilon_N - \varepsilon_{N-1}$ in the other? In the rest of this section we focus on a lower bound for $\varepsilon_{N+1} - \varepsilon_N$, the other one being obtained similarly.

Assume that we have computed the discretization H_{XX} of the Hamiltonian $H = \frac{1}{2}(-\nabla + \mathbf{k})^2 + V$ on a plane-wave basis set X with finite cutoff energy E_{cut}, and let $\widetilde{\varepsilon}_{n,X}$ be its eigenvalues with corresponding orthonormal eigenvectors $\widetilde{u}_{n,X}$. Note that for the purposes of this section we assume that all computations inside the basis set X are exact: we form explicitly the matrix representation of H_{XX} and diagonalize it using a dense eigensolver, and do not consider any arithmetic error.

From the variational principle, $\widetilde{\varepsilon}_{N,X} \geq \varepsilon_N$. We now need to obtain a rigorous lower bound on ε_{N+1}, which is is much more complex. Simply taking the difference $\widetilde{\varepsilon}_{N+1,X} - \widetilde{\varepsilon}_{N,X}$ may lead to an *overestimation* of the gap, see for example $E_{\text{cut}} = 3$ in Figure 4. To compute a proper lower bound on ε_{N+1} we express the operator H on X and its orthogonal complement $X^\perp$ as

$$H = \begin{pmatrix} H_{XX} & V_{XX^\perp} \\ V_{X^\perp X} & H_{X^\perp X^\perp} \end{pmatrix}.$$

In this we have used that the kinetic energy term is diagonal in a plane-wave basis and does not appear in the off-diagonal blocks. We then use the Haynsworth inertia additivity formula [43]: for any real threshold μ not in the spectrum of H_{XX}, the number of eigenvalues of H below μ is equal to the number of negative eigenvalues of the block $H_{XX} - \mu$ plus the number of negative eigenvalues of the Schur complement

$$S_\mu = (H_{X^\perp X^\perp} - \mu) - V_{X^\perp X}\,(H_{XX} - \mu)^{-1}\,V_{XX^\perp}.$$

For pedagogical purposes we briefly sketch the proof of this statement in the simplified case where $H_{XX} - \mu$ is positive. Consider the quadratic form

$$\begin{aligned}\langle u, (H-\mu)u\rangle = \langle u_X, (H_{XX}-\mu)u_X\rangle + \langle u_X, V_{XX^\perp}u_{X^\perp}\rangle \\ + \langle u_{X^\perp}, V_{X^\perp X}u_X\rangle + \langle u_{X^\perp}, (H_{X^\perp X^\perp}-\mu)u_{X^\perp}\rangle\end{aligned}$$

where $u = (u_X, u_{X^\perp})$. For a fixed $u_{X^\perp}$, this is again a quadratic form in u_X, which can be explicitly minimized by solving a linear system:

$$\operatorname*{argmin}_{u_X \in X} \langle u, (H-\mu)u\rangle = -(H_{XX}-\mu)^{-1} V_{XX^\perp} u_{X^\perp}$$

and therefore

$$\min_{u_X \in X} \langle u, (H-\mu)u\rangle = \langle u_{X^\perp}, S_\mu u_{X^\perp}\rangle$$

It follows that S_μ is positive if and only if $H - \mu$ is. The extension to arbitrary number of negative eigenvalues proceeds similarly.

From this, it follows that if we manage to prove that S_μ is positive for some $\mu \in (\widetilde{\varepsilon}_{N,X}, \widetilde{\varepsilon}_{N+1,X})$, we obtain that $\varepsilon_{N+1} \geq \mu$. Assume that we have computed eigenvectors $\widetilde{u}_{1,X}, \ldots, \widetilde{u}_{M,X}$ up until index $M \geq N+1$. Then the Schur complement S_μ can be bounded from below by expanding H_{XX} on its eigenvector basis $\widetilde{u}_{n,X}$

$$\begin{aligned}S_\mu \geq E_{\text{cut}} - \|V_{X^\perp X^\perp}\|_{\text{op}} - \mu \\ \underbrace{- \left\| \sum_{n=N+1}^{M-1} \frac{(V_{X^\perp X}\widetilde{u}_{n,X})(V_{X^\perp X}\widetilde{u}_{n,X})^\dagger}{\widetilde{\varepsilon}_{n,X} - \mu} \right\|_{\text{op}}}_{=B_\mu} - \frac{\|V_{XX^\perp}\|_{\text{op}}^2}{\widetilde{\varepsilon}_{M,X} - \mu},\end{aligned} \tag{7}$$

where $\|\cdot\|_{\text{op}}$ is the operator norm on the space of bounded linear operators on $\mathscr{H}$. Computing the term B_μ requires knowing the potential V on all of the complement $X^\perp$, which is not feasible for GTH pseudopotentials. Similarly as with the residuals, we assume that we are able to compute V on a superset $Y \supset X$ and additionally able to bound it on $Y^\perp$. With this in place we split B_μ into contributions inside and outside of Y:

$$\begin{aligned}B_\mu \geq - \left\| (V_{X^\perp \cap Y, X}\,\widetilde{U})\,(\widetilde{\Lambda} - \mu)^{-1}\,(V_{X^\perp \cap Y, X}\,\widetilde{U})^\dagger \right\|_{\text{op}} \\ - 2 \left\| (V_{X^\perp \cap Y, X}\,\widetilde{U})\,(\widetilde{\Lambda} - \mu)^{-1} \right\|_{\text{op}} \left\| V_{X, Y^\perp} \right\|_{\text{op}} - \frac{\left\| V_{X,Y^\perp} \right\|_{\text{op}}^2}{\widetilde{\varepsilon}_{N,X} - \mu},\end{aligned} \tag{8}$$

where $\widetilde{\Lambda}$ is the diagonal matrix of eigenvalues $\widetilde{\varepsilon}_{N+1,X}, \ldots, \widetilde{\varepsilon}_{M-1,X}$ and $\widetilde{U}$ the orthogonal matrix of corresponding eigenvectors (column-wise).

To compute the first term we use the fact that $V_{X^\perp \cap Y, X}\widetilde{U}$ is a computable, long-and-thin matrix (more rows than columns). We can employ the QR decomposition to factorize it into the product of an orthogonal matrix Q and a (small) triangular matrix R. The first term in the right-hand side of (8) can then be explicitly computed as the largest eigenvalue of the (small) Hermitian matrix $R\,(\widetilde{\Lambda} - \mu)^{-1} R^\dagger$. The second term can be treated similarly. Details on our bounds on the operator norms $\|V_{X^\perp X^\perp}\|_{\text{op}}$, $\|V_{XX^\perp}\|_{\text{op}}$ and $\left\|V_{X,Y^\perp}\right\|_{\text{op}}$ for the Cohen-Bergstresser and the GTH pseudopotential models are given in Appendix A.

For a fixed $\mu \in (\widetilde{\varepsilon}_{N,X}, \widetilde{\varepsilon}_{N+1,X})$, S_μ is positive for all E_{cut} large enough. For a given E_{cut} and M, we find the best lower bound to $\widetilde{\varepsilon}_{N+1,X}$, denoted by μ^*_{N+1}, as the maximum μ, for which we can ensure that S_μ is positive. This is done using a bisection algorithm on our bound of S_μ.

Figures 4 and 5 show gaps obtained using this lower bound on ε_{N+1} for different values of E_{cut} and M for the first eigenvalue at the Γ point of the two pseudopotential models we consider. For the simple Cohen-Bergstresser model and using $M = 8$ our lower bound for the gap is accurate already at $E_{\text{cut}} = 10$. For the GTH pseudopotentials one needs to use larger values of M and E_{cut}.

VI. ESTIMATING THE ARITHMETIC ERROR

To rigorously bound the arithmetic error, we use interval arithmetic, as specified in the IEEE standard 1788-2015 [44]. The main idea of interval arithmetic is to use not one but two floating-point numbers to represent a given quantity, forming an interval

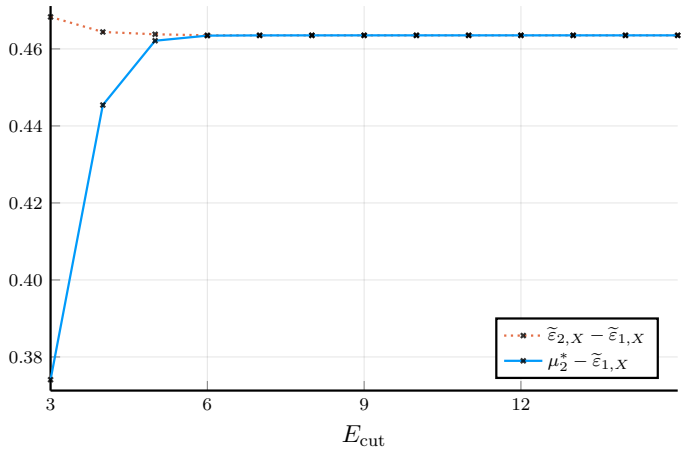

FIG. 4. Comparison of gap lower bounds for the Cohen-Bergstresser model of silicon. The orange dotted is the naïve estimate $\widetilde{\varepsilon}_{2,X} - \widetilde{\varepsilon}_{1,X}$ between the computed approximate eigenvalues. Shown in solid blue is $\mu_2^* - \widetilde{\varepsilon}_{1,X}$ with μ a lower bound to $\widetilde{\varepsilon}_{2,X}$ obtained as described in the main text with $M = 8$ eigenpairs.

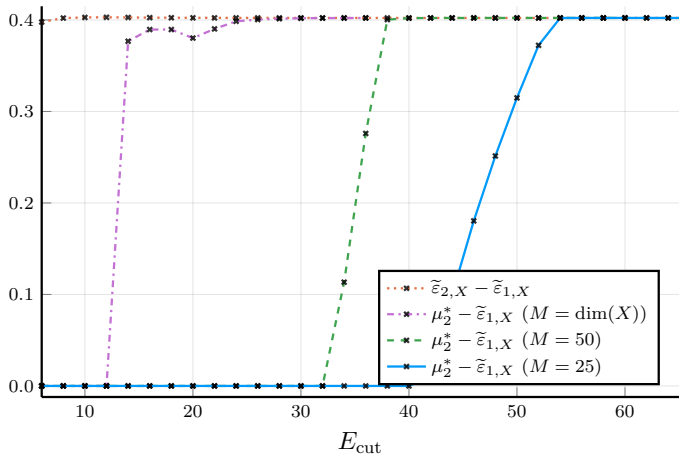

FIG. 5. Gap lower bounds of the GTH pseudopotential model of silicon. Similar to Figure 4, dotted orange indicates the naïve eigenvalue difference, dashed green the lower bound $\mu_2^* - \widetilde{\varepsilon}_{1,X}$ obtained using $M = 50$ eigenpairs and solid blue using $M = 25$ eigenpairs. The purple dashed line refers to the case where H_{XX} is fully diagonalized, i.e. $M = \dim(X)$.

which contains the exact number. Every operation is then performed on the limits of the interval utilizing rounding modes of CPUs to ensure that the upper limit is only rounded upwards and the lower limit only rounded downwards if rounding is needed to represent the outcome of the operation. This ensures that the *exact* answer always lies between the limits of the final interval. This simplistic approach generally overestimates floating-point error, since it considers all floating-point operations independent from each other. The bound obtained in this way is thus too large making an application of interval arithmetic to, e.g. a complete iterative diagonalization algorithm, impractical. Fortunately if we obtain an approximate eigenpair $(\widetilde{\varepsilon}, \widetilde{u})$ inside a discretization basis X using ordinary floating-point arithmetic, we only need to re-evaluate the residual norm $\|P_X(H\widetilde{u} - \widetilde{\varepsilon}\widetilde{u})\|$ in interval arithmetic to obtain a rigorous bound. The upper bound of the resulting interval gives access to the sum of *both* floating-point error and algorithmic error due to the eigensolver.

In Julia IEEE interval arithmetic is available in the `IntervalArithmetic.jl` [45] package as a custom floating-point type. This type can be directly used with any native Julia code, including `GenericLinearAlgebra.jl` [46] and `FourierTransforms.jl` [47], which provide interval-arithmetic equivalents to classical linear algebra and FFT algorithms. These packages allowed us to use the routines of DFTK to also bound the arithmetic error for all our double-precision calculations presented in this paper. Example values are shown in Figure 6 indicating that the obtained arithmetic error is several orders of magnitude smaller than the discretization error until around $E_{\text{cut}} = 65$. Increasing the accuracy of the obtained solution to the Cohen-Bergstresser model beyond this point would not only require to increase the basis, but also to switch to a more accurate floating-point arithmetic beyond IEEE double precision. In a type-generic Julia code like DFTK this is, however, completely seamless and has been used to obtain Figure 7, demonstrating errors below double precision.

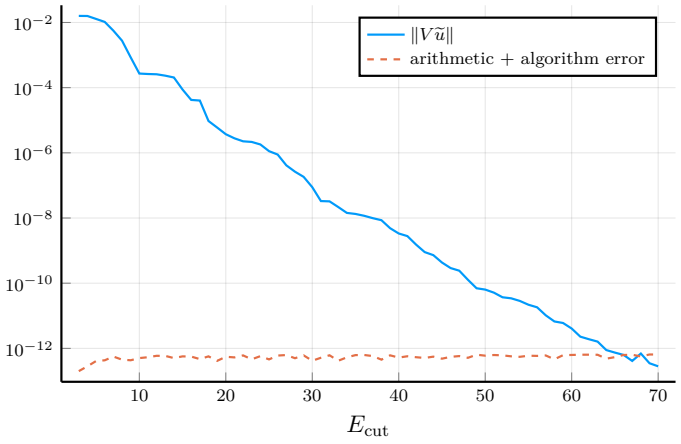

FIG. 6. Total residual and sum of algorithm and floating-point errors at double-precision arithmetic for the first band at the Γ point of the Cohen-Bergstresser model of silicon.

VII. TRACKING ERRORS IN BAND COMPUTATIONS

With the discussed methodologies we are able to

- compute the norm of the in-space residual norm $\|P_X\widetilde{r}\|$ (leading to the algorithmic error);
- compute a guaranteed upper bound of the out-of-space residual norm $\|P_{X^\perp}\widetilde{r}\|$ (leading to the discretization error);
- compute a guaranteed lower bound of the spectral gap δ involved in the Kato-Temple inequality (4);
- compute *a posteriori* errors bounds on the quantities of interest (crystalline orbital energies) using either Bauer-Fike or Kato-Temple inequalities;
- estimate the impact of floating-point errors via interval arithmetic (arithmetic error).

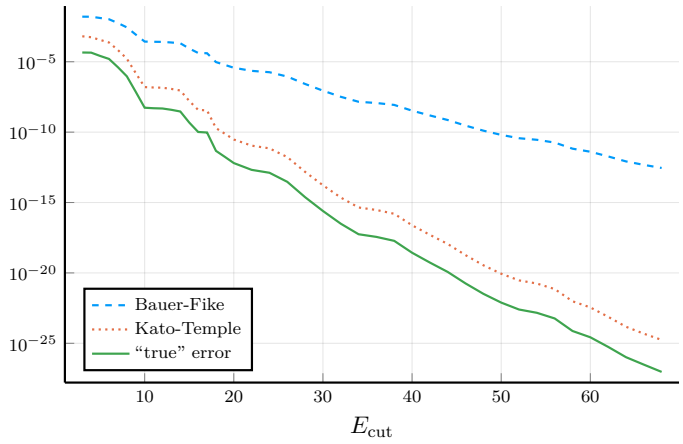

FIG. 7. Overview of discretization error and error bounds at the Γ point for the first band in the Cohen-Bergstresser model of silicon. The `DoubleFloats.jl` [48] package was used to go beyond the accuracy limitations of IEEE double precision. Gap lower bounds for the Kato-Temple bound have been obtained with $M = 8$ eigenpairs. The "true" error was computed by taking the absolute difference to the eigenvalue obtained at $E_{\mathrm{cut}} = 70$.

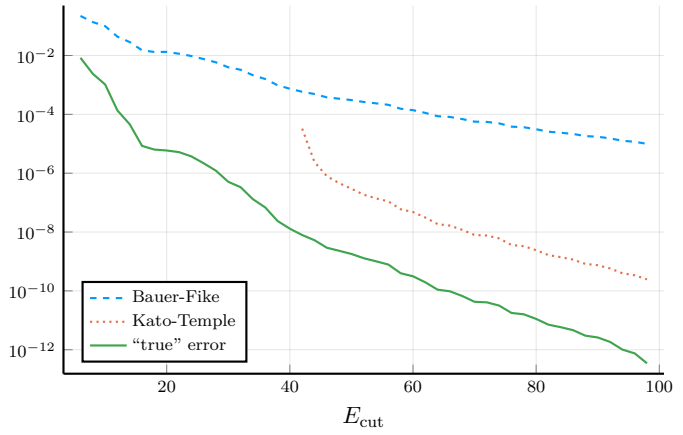

FIG. 8. Overview of discretization error and error bounds at the Γ point for the first band of the GTH model of silicon. Gap lower bounds for the Kato-Temple bound have been obtained with $M = 25$ eigenpairs. The "true" error was computed by taking the absolute difference to the eigenvalue obtained at $E_{\mathrm{cut}} = 100$.

A summary of the overall numerical error on the quantities of interest bound by Bauer-Fike and Kato-Temple versus the "true" error estimated using the largest basis employed ($E_{\mathrm{cut}} = 50$ and $E_{\mathrm{cut}} = 100$ respectively) is given in Figures 7 and 8 for the Cohen-Bergstresser and GTH models of silicon. Using our methods we were able to compute band structures for both models with fully guaranteed bounds on the numerical error, see Figures 9 and 10, respectively. Notice that neither Kato-Temple nor Bauer-Fike universally provide the best bound on the error, with the Kato-Temple bound being inapplicable for degenerate eigenvalues in particular. In our plots we alternate between both bounds, only showing the one providing the smallest error. The design and use of error bounds robust to degenerate eigenvalues is left to future work.

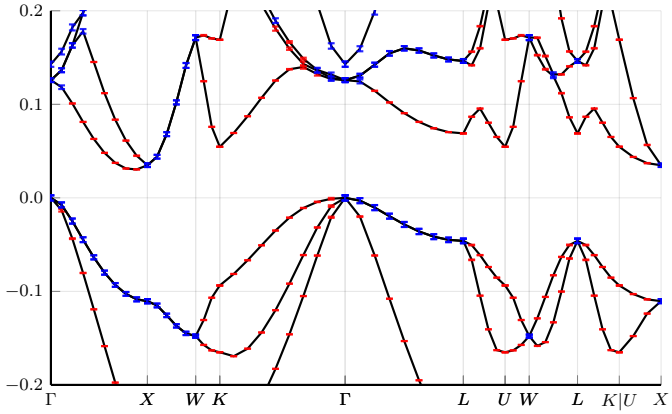

FIG. 9. Band structure of silicon using the Cohen-Bergstresser model at $E_{\mathrm{cut}} = 10$. Error bars indicate the total numerical error, i.e. arithmetic error, algorithm error and discretization error. Red (resp. blue) error bars indicate that the Kato-Temple (resp. Bauer-Fike) bound gave the lowest discretization error and was used. Gap lower bounds have been obtained with $M = 8$ eigenpairs. Band energies are given in Hartrees and relative to the 4th eigenvalue at the Γ point.

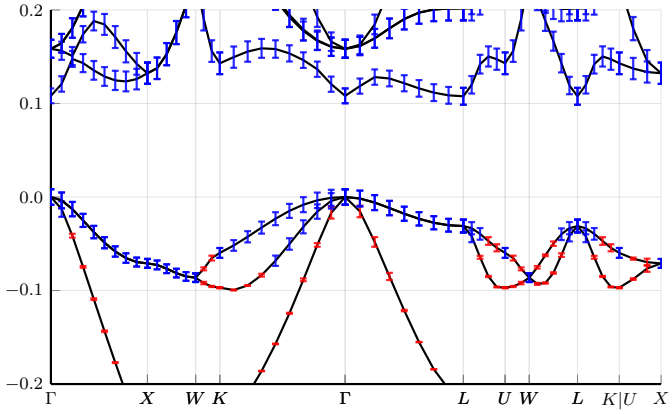

FIG. 10. Band structure with error bars for the GTH model of silicon at $E_{\mathrm{cut}} = 42$. The same conventions as in Figure 9 are used, except that gap lower bounds have been obtained using $M = 35$ eigenpairs.

VIII. CONCLUSIONS AND OUTLOOK

We have discussed the computation of fully guaranteed bounds to the error in the numerical solution of non-self-consistent Kohn-Sham equations. While many other sources of error exist, our treatment focused exclusively on the discretization error, due to finite basis sets, algorithm error, due to non-zero convergence thresholds, and arithmetic error, due to finite-precision floating-point arithmetic. As quantities of interest we considered the band energies around the Fermi level of the Cohen-Bergstresser and GTH pseudopotential models and demonstrated the feasibility of our approach by providing, for each model, band structure diagrams of silicon, in which the total numerical error was annotated in the form of error bars. In this paper we have focused on error bounds on eigenvalues for simplicity, but that is not a major limitation of our approach; error bounds on eigenvectors can be obtained in a similar fashion

to the Kato-Temple bound (see [24], Theorem 3.9). From there it is possible to obtain bounds on other quantities of interest, for instance the density or the forces.

Our presented approach relies on two key ingredients. First, a bound on the residual, including both its in-space component (the one driven to zero by the iterative solvers) and its out-of-space component. For the out-of-space component, we relied in this work on the properties of plane-wave basis and of the analytical pseudopotential; however, this can in principle be applied to any basis set for which exact analytical computations are possible, such as Gaussian basis sets. The second ingredient is to relate the residual to the actual error, which involved for our particular quantities of interest an eigenvalue gap. This is much more complex to perform in full rigor, as it requires a control of the out-of-basis part of the full operator itself. Achieving a similar analysis to ours in Gaussian basis sets (for instance) is a major research challenge.

A limitation of our study is that we only considered non-self-consistent models. Taking into account self-consistency in error bounds one runs into two separate issues. First, the non-convexity of the model due to the exchange-correlation term can introduce multiple energy local minima, and it is impossible to certify that the true ground state has been found. One then has to settle for a less ambitious notion of error control: showing that there exists a solution of the equation close to the numerically obtained approximate solution. Second, rigorously proving error bounds for nonlinear problems involves mathematically more sophisticated techniques to achieve a sufficient control on the nonlinear terms. This has been performed for the Gross-Pitaevskii equation, an equation similar in form but simpler than Kohn-Sham DFT [49]. Extending this to Kohn-Sham DFT is work in progress.

Finally, we hope to use our bounds to enable a fully black-box modeling, where accuracy parameters such as the kinetic energy cutoff, the convergence thresholds of the iterative solver or employed floating-point precision are chosen automatically by the code. The hope is to be able to dynamically adjust such accuracy parameters during a simulation to do as little work as necessary to reach the accuracy desired by the user. For such purposes we expect the presented Kato-Temple-type bounds to be not tight enough, so that better bounds have to be constructed and implemented. Preliminary work in that direction can be found in [13].

CONFLICTS OF INTEREST

There are no conflicts of interest to declare.

ACKNOWLEDGEMENTS

This project has received funding from ISCD (Sorbonne Université) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367).

APPENDIX A: BOUNDS ON POTENTIALS

In this section we fix the two plane-wave spaces X and Y defined by energy cutoffs E_{cut} and $E_{\text{cut}}^{(2)} > E_{\text{cut}}$. Our error bounds for the GTH pseudopotentials require explicit bounds on the quantities $\|P_{Y^\perp} V \widetilde{u}\|$ for a given $\widetilde{u} \in X$ (to bound the residuals), and on the operators $P_{X^\perp} V P_X$, $P_{Y^\perp} V P_X$, and $P_{X^\perp} V P_{X^\perp}$ (to obtain bounds on the gap).

We will bound these quantities using our explicit knowledge of the tails of the functions involved: given a non-negative continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$, we define

$$S_{\mathscr{R}^*}(f,q) = \sum_{\mathbf{G}\in\mathscr{R}^*, |\mathbf{G}|\geq q} f(|\mathbf{G}|).$$

Bounds on $S_{\mathscr{R}^*}(f,q)$ are obtained in Appendix B.

Note that there is a large freedom in designing bounds, and many possible improvements. The ones we present result from a compromise between accuracy and simplicity.

Local potential

In the following we assume a single atom centered at the origin; the case of multiple atoms is obtained simply by adding a contribution for each atom. The function $\widehat{v}_{\text{loc}}(\Delta\mathbf{G})$ is then radial.

To bound $P_Y V_{\text{loc}} \widetilde{u}$ for a given $\widetilde{u}$, we use the fact that $\widetilde{u}$ has small components on high wavevectors:

$$\begin{aligned}
&|\Omega|^2 \|P_{Y^\perp} V_{\text{loc}} \widetilde{u}\|^2 \\
&= \sum_{\mathbf{G}\in Y^\perp} \left| \sum_{\mathbf{G}'\in X} \widehat{v}_{\text{loc}}(\mathbf{G}-\mathbf{G}') \widetilde{u}(\mathbf{G}') \right|^2 \\
&\leq \dim(X) \sum_{\mathbf{G}\in Y^\perp} \sum_{\mathbf{G}'\in X} \left| \widehat{v}_{\text{loc}}(\mathbf{G}-\mathbf{G}') \widetilde{u}(\mathbf{G}') \right|^2 \\
&= \dim(X) \sum_{\mathbf{G}'\in X} |\widetilde{u}(\mathbf{G}')|^2 \sum_{\mathbf{G}\in Y^\perp} \left| \widehat{v}_{\text{loc}}(\mathbf{G}-\mathbf{G}') \right|^2 \\
&\leq \dim(X) \sum_{\mathbf{G}'\in X} |\widetilde{u}(\mathbf{G}')|^2 \sum_{\substack{\Delta\mathbf{G}\in\mathscr{R}^* \\ |\Delta\mathbf{G}|>\sqrt{2E_{\text{cut}}^{(2)}}-|\mathbf{G}'|}} |\widehat{v}_{\text{loc}}(|\Delta\mathbf{G}|)|^2 \\
&= \dim(X) \sum_{\mathbf{G}'\in X} |\widetilde{u}(\mathbf{G}')|^2 S_{\mathscr{R}^*}\left(|\widehat{v}_{\text{loc}}|^2, \sqrt{2E_{\text{cut}}^{(2)}} - |\mathbf{G}'| \right)
\end{aligned}$$

We bound the operators $P_{X^\perp} V_{\text{loc}} P_X$, $P_{Y^\perp} V_{\text{loc}} P_X$ and $P_{X^\perp} V_{\text{loc}} P_{X^\perp}$ in operator norm simply by the operator norm of V_{loc}. We use the fact that V_{loc} is a multiplication operator in real space by the periodic function

$$v_{\text{loc}}(\mathbf{r}) = \frac{1}{|\Omega|} \sum_{\mathbf{G}\in\mathscr{R}^*} \widehat{v}_{\text{loc}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$$

and therefore

$$\begin{aligned}
\|V_{\text{loc}}\|_{\text{op}} &\leq \frac{1}{\Omega} \left(\left\| \sum_{\mathbf{G}\in Y} \widehat{v}_{\text{loc}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \right\|_\infty + \sum_{\mathbf{G}\in Y^\perp} |\widehat{v}_{\text{loc}}| \right) \\
&\leq \frac{1}{\Omega} \left(\left\| \sum_{\mathbf{G}\in Y} \widehat{v}_{\text{loc}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \right\|_\infty + S_{\mathscr{R}^*}(|\widehat{v}_{\text{loc}}|, \sqrt{2E_{\text{cut}}}) \right),
\end{aligned}$$

where $\|v\|_\infty = \sup_{\mathbf{r}\in\mathbb{R}^3} |v(\mathbf{r})|$. To bound the first term, we use the fact that for a regular grid $\mathfrak{G}$ of the unit cell, we have

$$\left\| \sum_{\mathbf{G}\in Y} \widehat{v}_{\rm loc}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \right\|_\infty \leq \max_{\mathbf{r}\in\mathfrak{G}} \left| \sum_{\mathbf{G}\in Y} \widehat{v}_{\rm loc}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \right| + \delta \sum_{\mathbf{G}\in Y} |\mathbf{G}||\widehat{v}_{\rm loc}(\mathbf{G})|$$

where δ is the diameter of the grid. We compute the first term using a fast Fourier transform and the second explicitly.

Nonlocal potential

The nonlocal potential is a sum of separable terms [20, 21], over atoms, angular momentum l, magnetic quantum number m and projector channels i, j. We focus on just one of the terms in (3), of the form

$$\langle e_{\mathbf{G}} | v_{\rm nl} | e_{\mathbf{G}'} \rangle = p^{(1)}(\mathbf{k}+\mathbf{G}) \overline{p^{(2)}(\mathbf{k}+\mathbf{G}')},$$

where both $p^{(i)}$ are of the form $p^{(i)}(\mathbf{q}) = Y_{lm}(\mathbf{q}/|\mathbf{q}|)R^{(i)}(|\mathbf{q}|)$ ($i = 1,2,\ldots$). A bound on the total potential can be obtained naively by the triangular inequality. A better bound can be obtained by exploiting orthogonality between the different quantum numbers (l,m) and using Pythagoras theorem as well as using Unsöld's theorem to simplify the sums over m. We do not detail these technicalities here.

For a given $\widetilde{u} \in X$, let $c_{\widetilde{u}} = \sum_{\mathbf{G}\in X} \overline{p^{(2)}(\mathbf{k}+\mathbf{G})}\widetilde{u}(\mathbf{G})$. Then

$$\begin{aligned}
\|P_{Y^\perp} v_{\rm nl} \widetilde{u}\|^2 &= |c_{\widetilde{u}}|^2 \sum_{\mathbf{G}\in Y^\perp} |p^{(1)}(\mathbf{k}+\mathbf{G})|^2 \\
&\leq |c_{\widetilde{u}}|^2 \|Y_{lm}\|_\infty^2 \sum_{\mathbf{G}\in Y^\perp} |R^{(1)}(|\mathbf{G}|-|\mathbf{k}|)|^2 \\
&= |c_{\widetilde{u}}|^2 \|Y_{lm}\|_\infty^2 S_{\mathscr{R}^*}\left(|R^{(1)}(\cdot - |\mathbf{k}|)|^2, \sqrt{2E_{\rm cut}^{(2)}} \right),
\end{aligned}$$

where $\|Y_{lm}\|_\infty = \sup_{\mathbf{q}\in\mathbb{R}^3, |\mathbf{q}|=1} |Y_{lm}(\mathbf{q})|$ and we have assumed that $|R^{(i)}|$ is non-increasing on $[E_{\rm cut}^{(2)} - |\mathbf{k}|, +\infty)$. We compute $c_{\widetilde{u}}$ explicitly and bound the remainder S as before. For the operator norm of $P_{X^\perp} V_{\rm nl} P_X$ and $P_{Y^\perp} V_{\rm nl} P_X$, we use the above computation together with the fact that, for any normalized $\widetilde{u} \in X$,

$$|c_{\widetilde{u}}|^2 \leq \sum_{\mathbf{G}\in X} |p^{(2)}(\mathbf{k}+\mathbf{G})|^2.$$

Finally,

$$\begin{aligned}
\|P_{X^\perp} V_{\rm nl} P_{X^\perp}\|_{\rm op}^2 \leq \|Y_{lm}\|_\infty^4 \quad & S_{\mathscr{R}^*}(|R^{(1)}(\cdot - |\mathbf{k}|)|^2, \sqrt{2E_{\rm cut}}) \\
& S_{\mathscr{R}^*}(|R^{(2)}(\cdot - |\mathbf{k}|)|^2, \sqrt{2E_{\rm cut}}).
\end{aligned}$$

APPENDIX B: BOUNDS ON TAIL SUMS

Our bounds involve quantities of the form

$$S_{\mathscr{R}^*}(f,q) := \sum_{\mathbf{G}\in\mathscr{R}^*, |\mathbf{G}|\geq q} f(|\mathbf{G}|).$$

where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a known smooth non-negative function non-increasing on the internal $[q_{\rm min}, +\infty)$ for some $q_{\rm min} \geq 0$, and more specifically a Gaussian times a polynomial or rational function. We seek to bound $S_{\mathscr{R}^*}(f,q)$ by an explicitly computable quantity for q large enough. For a one dimensional lattice $\mathscr{R}^* = b\mathbb{Z}$, $b > 0$, this can easily be done using a sum-integral comparison. We have for all $q \geq q_{\rm min} + b$, $f(q) \leq b^{-1} \int_{q-b}^q f(q')\,dq'$, so that

$$S_{b\mathbb{Z}}(f,q) \leq 2b^{-1} \int_{q-b}^\infty f(q')\,dq',$$

where the right-hand side can be computed explicitly using integrals of Gaussian functions.

For the multidimensional case we use a similar idea: we bound the value of $f(|\mathbf{G}|)$ by its mean over a unit cell $C_{\mathbf{G}}$ for which $|\mathbf{G}|$ is the vertex of $C_{\mathbf{G}}$ furthest away from the origin (see Figure 11). A technical difficulty lays in the fact that this $C_{\mathbf{G}}$ is not always uniquely defined and that the $C_{\mathbf{G}}$'s do not form a proper tiling. For instance, for the two-dimensional square lattice, we have

$$\begin{aligned}
S_{b\mathbb{Z}^2}(f,q) &= 4 \sum_{\mathbf{G}\in b(\mathbb{N}^*\times\mathbb{N}^*), |\mathbf{G}|\geq q} f(|\mathbf{G}|) + 2 \sum_{\mathbf{G}\in b(\mathbb{Z}\times\{0\}), |\mathbf{G}|\geq q} f(|\mathbf{G}|) \\
&\leq 4 \sum_{\mathbf{G}\in b(\mathbb{N}^*\times\mathbb{N}^*), |\mathbf{G}|\geq q} \int_{C_{\mathbf{G}}} f(\mathbf{G}')d\mathbf{G}' + 2S_{b\mathbb{Z}}(f,q). \\
&\leq 2\pi b^{-2} \int_{q-\sqrt{2}b}^{+\infty} q' f(q')\,dq' + 2S_{b\mathbb{Z}}(f,q).
\end{aligned}$$

with a similar bound for the cubic lattice. For a non-cubic lattice $\mathscr{R}^* = \mathbf{b}_1\mathbb{Z} + \mathbf{b}_2\mathbb{Z} + \mathbf{b}_3\mathbb{Z}$, a partitioning in octants has to be done instead based on the signs of the quantities $\mathbf{G}\cdot\mathbf{b}_i$, in the spirit of what can be seen on Figure 11 for a two-dimensional non-square case. Unit cells then overlap in the vicinity of the three planes

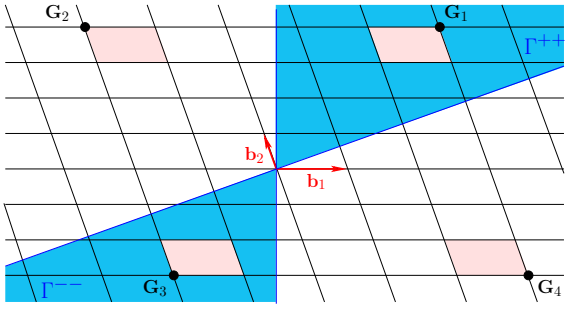

FIG. 11. Partitioning of the space used to decide in which direction to extend the unit cell $C_{\mathbf{G}}$ (red).

$\mathbf{b}_1^\perp, \mathbf{b}_2^\perp$ and $\mathbf{b}_3^\perp$. A relatively straightforward but tedious bound on these extra overlaps yields the following:

$$\begin{aligned}
S_{\mathscr{R}^*}(f,q) \leq \frac{4\pi}{|C|} \int_{q-\delta}^{+\infty} q'^2 f(q')\,dq' + 2\pi \left(\sum_{j=1}^3 \frac{n_{\mathscr{R}^*,j}}{|C_j|} \right) \int_{q-\delta}^{+\infty} q' f(q')\,dq' \\
+ 2 \left(\sum_{j=1}^3 \frac{n_{\mathscr{R}^*,j}}{|\widetilde{C}_j|} \right) \int_{q-\delta}^{+\infty} f(q')\,dq', \qquad (9)
\end{aligned}$$

with δ the unit cell diameter and

$$
\begin{aligned}
|C| &= |(\mathbf{b}_1 \times \mathbf{b}_2) \cdot \mathbf{b}_3| \\
|C_j| &= \left| \left(\mathbf{b}_{j+1} - \frac{(\mathbf{b}_j \cdot \mathbf{b}_{j+1})}{|\mathbf{b}_j|^2} \mathbf{b}_j \right) \times \left(\mathbf{b}_{j+2} - \frac{(\mathbf{b}_j \cdot \mathbf{b}_{j+2})}{|\mathbf{b}_j|^2} \mathbf{b}_j \right) \right| \\
|\widetilde{C}_j| &= \left| \mathbf{b}_j - \sum_{k \neq j} \frac{(\mathbf{b}_k \cdot \mathbf{b}_j)}{|\mathbf{b}_k|^2} \mathbf{b}_k \right| \\
n_{\mathscr{R}^*,j} &= \prod_{k \neq j} \lceil 2 + \frac{|\mathbf{b}_j \cdot \mathbf{b}_k|}{|\mathbf{b}_j|^2} \rceil .
\end{aligned}
$$

with the convention that $\mathbf{b}_4 = \mathbf{b}_1, \mathbf{b}_5 = \mathbf{b}_2$. Details of this computation will be published in an upcoming paper.

The bound above is numerically observed to be very pessimistic because it uses values of $f(q')$ for $q' < q$, which for a rapidly decaying function are much larger than $f(q)$. In order to obtain a better error, we instead apply this bound to the function $\widetilde{f}(q') = \min(f(q), f(q'))$.

[1] J.-O. Mo, A. Choudhry, M. Arjomandi, and Y.-H. Lee, J. Wind. Eng. Ind. Aerodyn. **112**, 11 (2013).
[2] P. Spethmann, C. Herstatt, and S. H. Thomke, Int. J. Prod. Dev. **8**, 291 (2009).
[3] R. M. Martin and R. M. Martin, *Electronic structure: basic theory and practical methods* (Cambridge University Press, 2004).
[4] M. Snir, R. W. Wisniewski, J. A. Abraham, S. V. Adve, S. Bagchi, P. Balaji, J. Belak, P. Bose, F. Cappello, B. Carlson, *et al.*, Int. J. High Perform. Comput. Appl. **28**, 129 (2014).
[5] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, Crit. Rev. Solid State Mater. Sci. **39**, 1 (2014).
[6] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, *et al.*, Science **351**, aad3000 (2016).
[7] E. Cancès, R. Chakir, and Y. Maday, ESAIM: M2AN **46**, 341 (2012).
[8] H. Chen, X. Gong, L. He, Z. Yang, and A. Zhou, Adv. Comput. Math. **38**, 225 (2013).
[9] H. Chen and R. Schneider, ESAIM: M2AN **49**, 755 (2015).
[10] H. Chen and R. Schneider, Comm. Comput. Phys. **18**, 125 (2015).
[11] J. Kaye, L. Lin, and C. Yang, Comm. Math. Sci. **13**, 1741 (2015).
[12] E. Cancès, G. Dusson, Y. Maday, B. Stamm, and M. Vohralík, J. Comput. Phys. **307**, 446 (2016).
[13] E. Cancès, G. Dusson, Y. Maday, B. Stamm, and M. Vohralik, Math. Comput. (in press).
[14] E. Cancès, V. Ehrlacher, D. Gontier, A. Levitt, and D. Lombardi, Numer. Math. (in press).
[15] M. Chupin, M.-S. Dupuy, G. Legendre, and E. Séré, arXiv:2002.12850 (2020), 2002.12850 [math.NA].
[16] X. Dai, Z. Liu, L. Zhang, and A. Zhou, SIAM J. Sci. Comput. **39**, A2702 (2017).
[17] T. Rohwedder and R. Schneider, J Math Chem **49**, 1889 (2011).
[18] Z. Zhao, Z. Bai, and X. Jin, SIAM J. Matrix Anal. Appl. **36**, 752 (2015).
[19] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).
[20] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[21] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
[22] T. Rohwedder and R. Schneider, ESAIM: Math. Model. and Num. Anal. **47**, 1553 (2013).
[23] P. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[24] Y. Saad, *Numerical Methods for Large Eigenvalue Problems: Revised Edition*, Classics in Applied Mathematics (Society for Industrial and Applied Mathematics, 2011).
[25] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[26] M. F. Herbst and A. Levitt, "Density-functional toolkit," Accessed on 08 April 2020.
[27] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, SIAM Rev. **59**, 65 (2017).
[28] M. F. Herbst, A. Levitt, and E. Cancès, "Implementation of a posteriori error estimates for non-self-consistent Kohn-Sham equations," Accessed on 14 April 2020.
[29] F. Goerisch and Z. Q. He, in *Computer arithmetic and self-validating numerical methods (Basel, 1989)*, Notes Rep. Math. Sci. Engrg., Vol. 7 (Academic Press, Boston, MA, 1990) pp. 137–153.
[30] I. Babuška and J. Osborn, in *Handbook of numerical analysis, Vol. II*, Handb. Numer. Anal., II (North-Holland, Amsterdam, 1991) pp. 641–787.
[31] V. Heuveline and R. Rannacher, Adv. Comput. Math. **15**, 107 (2001).
[32] V. Mehrmann and A. Miedlar, Numer. Linear Algebra Appl. **18**, 387 (2011).
[33] Y. A. Kuznetsov and S. I. Repin, J. Numer. Math. **21**, 135 (2013).
[34] R. E. Bank, L. Grubišić, and J. S. Ovall, Appl. Numer. Math. **66**, 1 (2013).
[35] C. Carstensen and J. Gedicke, Math. Comp. **83**, 2605 (2014).
[36] X. Liu, Appl. Math. Comput. **267**, 341 (2015).
[37] E. Cancès, G. Dusson, Y. Maday, B. Stamm, and M. Vohralík, Numer. Math. **140**, 1033 (2018).
[38] D. Gallistl, Comput. Methods Appl. Math. **14** (2014).
[39] D. Gallistl, Numer. Math. **130**, 467 (2015).
[40] X. Dai, L. He, and A. Zhou, IMA J. Numer. Anal. **35**, 1934 (2015).
[41] A. Bonito and A. Demlow, SIAM J. Numer. Anal. **54**, 2379 (2016).
[42] D. Boffi, D. Gallistl, F. Gardini, and L. Gastaldi, Math. Comp. **86**, 2213 (2017).
[43] E. V. Haynsworth, Linear Algebra Appl. **1**, 73 (1968).
[44] IEEE Std 1788.1-2017 , 1 (2018).
[45] D. P. Sanders, L. Benet, E. Gupta, B. Richard, *et al.*, "JuliaIntervals/IntervalArithmetic.jl: v0.17.0," (2020).
[46] A. Noack *et al.*, "Generic numerical linear algebra in Julia," Accessed on 08 April 2020.
[47] S. G. Johnson, A. Noack, Y. Ma, *et al.*, "Fourier transforms written in Julia," Accessed on 08 April 2020.
[48] J. Sarnoff *et al.*, "DoubleFloats.jl: math with more good bits," Accessed on 14 April 2020.
[49] G. Dusson and Y. Maday, IMA Journal of Numerical Analysis **37**, 94 (2017).