Discretization error cancellation in electronic structure calculation: toward a quantitative study

Eric Cancès, Geneviève Dusson

Discretization error cancellation in electronic structure calculation: toward a quantitative study

Eric Cancès[†] Geneviève Dusson[‡]

June 2, 2017

Abstract

It is often claimed that error cancellation plays an essential role in quantum chemistry and first-principle simulation for condensed matter physics and materials science. Indeed, while the energy of a large, or even medium-size, molecular system cannot be estimated numerically within chemical accuracy (typically 1 kcal/mol or 1 mHa), it is considered that the energy difference between two configurations of the same system can be computed in practice within the desired accuracy.

The purpose of this paper is to initiate the quantitative study of discretization error cancellation. Discretization error is the error component due to the fact that the model used in the calculation (e.g. Kohn-Sham LDA) must be discretized in a finite basis set to be solved by a computer. We first report comprehensive numerical simulations performed with Abinit [14, 15] on two simple chemical systems, the hydrogen molecule on the one hand, and a system consisting of two oxygen atoms and four hydrogen atoms on the other hand. We observe that errors on energy differences are indeed significantly smaller than errors on energies, but that these two quantities asymptotically converge at the same rate when the energy cut-off goes to infinity. We then analyze a simple one-dimensional periodic Schrödinger equation with Dirac potentials, for which analytic solutions are available. This allows us to explain the discretization error cancellation phenomenon on this test case with quantitative mathematical arguments.

AMS subject classifications: 65N25, 35P15, 65G99, 81-08
Key words: Electronic structure calculation; Schrödinger operators; Error analysis

1 Introduction

Error control is a central issue in molecular simulation. The error between the computed value of a given physical observable (e.g. the dissociation energy of a molecule) and the exact one, has several origins. First, there is always a discrepancy between the physical reality and the reference model, here the N-body Schrödinger equation, possibly supplemented with Breit terms to account for relativistic effects. However, at least for the atoms of the first three rows of the periodic table, this reference model is in excellent agreement with experimental data, and can be considered as exact in most situations of interest. The overall error is therefore the sum of the following components:

1. the *model error*, that is the difference between the value of the observable for the reference model, which is too complicated to solve in most cases, and the value obtained with the chosen approximate model (e.g. the Kohn-Sham LDA model), assuming that the latter can be solved exactly;

[†]CERMICS, Ecole des Ponts and INRIA Paris, 6 & 8 Avenue Blaise Pascal, 77455 Marne-la-Vallée, France. email: cances@cermics.enpc.fr

[‡]Sorbonne Universités, UPMC Univ. Paris 06 and CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, and Sorbonne Universités, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France. email: dusson@ljll.math.upmc.fr

*The authors are grateful to Yvon Maday for useful discussions, as well as the anonymous reviewers for interesting suggestions. This work was partially undertaken in the framework of CALSIMLAB, supported by the public grant ANR-11-LABX- 0037-01 overseen by the French National Research Agency (ANR) as part of the Investissements d'avenir program (reference: ANR-11-IDEX-0004-02).

2. the *discretization error*, that is the difference between the value of the observable for the approximate model and the value obtained with the chosen discretization of the approximate model. Indeed, the approximate model is typically an infinite dimensional minimization problem, or a system of partial differential equations, which must be discretized to be solvable by a computer, using e.g. a Gaussian atomic basis set, or a planewave basis;

3. the *algorithmic error*, which is the difference between the value of the observable obtained with the exact solution of the discretized approximate model, and the value computed with the chosen algorithm. The discretized approximate models are indeed never solved exactly; they are solved numerically by iterative algorithms (e.g. SCF algorithms, Newton methods), which, in the best case scenario, only converge in the limit of an infinite number of iterations. In practice, stopping criteria are used to exit the iteration loop when the error at iteration k, measured in terms of differences between two consecutive iterates or, better, by some norm of some residual, is below a prescribed threshold. If the stopping criterion is very tight, the algorithmic error can become very small, ... or not! For instance, if the discretized approximate model is a non convex optimization problem, there is no guarantee that the numerical algorithm will converge to a global minimum. It may converge to a local, non-global minimum, leading to a non-zero algorithmic error even in the limit of an infinitely tight stopping criterion;

4. the *implementation error*, which may, obviously, be due to bugs, but does not vanish in the absence of bugs, because of round-off errors: in molecular simulation packages, most operations are implemented in double precision, and the resulting round-off errors can accumulate, especially for very large systems;

5. the *computer error*, due to random hardware failures (miswritten or misread bits). This component of the error is usually negligible in today's standard computations, but is expected to become critical in future exascale architectures [24].

Quantifying these different sources of errors is an interesting purpose for two reasons. First, guaranteed estimates on these five components of the error would allow one to supplement the computed value of the observable returned by the numerical simulation with guaranteed error bars (certification of the result). Second, they would allow one to choose the parameters of the simulation (approximate model, discretization parameters, algorithm and stopping criteria, data structures, etc.) in an optimal way in order to minimize the computational effort required to reach the target accuracy.

The construction of guaranteed error estimators for electronic structure calculation is a very challenging task. Some progress has however been made in the last few years, regarding notably the discretization and algorithmic errors for Kohn-Sham LDA calculations. *A priori* discretization error estimates have been constructed in [3] for planewave basis sets, and then in [8] for more general variational discretization methods. *A posteriori* error estimators of the discretization error have been proposed in [5, 7, 19]. A combined study of both the discretization and algorithmic errors was published in [4] (see also [11]). We also refer to [26, 9, 10, 23, 25, 17, 22, 30, 31, 33] and references therein for other works on error analysis for electronic structure calculation.

In all the previous works on this topic we are aware of, the purpose was to estimate, *for a given nuclear configuration R of the system*, the difference between the ground state energy E_R (or another observable) obtained with the continuous approximate model under consideration (e.g. Kohn-Sham LDA) and its discretized counterpart denoted by $E_{R,N}$, where N is the discretization parameter. The latter is typically the number of basis functions in the basis set for local combination of atomic orbitals (LCAO) methods [18], the inverse fineness of the grid or the mesh for finite difference (FD) and finite element (FE) methods [16, 34, 29, 28], the cut-off parameter in energy or momentum space for planewave (PW) discretization methods [14, 12, 21], or the inverse grid spacing and the coarse and fine region multipliers for wavelet (WL) methods [27]. In variational approximation methods (LCAO, FE, PW, and WL), the discretization error $E_{R,N} - E_R$ is always nonnegative by construction. In systematically improvable methods (FD, FE, PW, and WL), this quantity goes to zero when N goes to infinity with a well-understood rate of convergence

depending on the smoothness of the pseudopotential (see [3] for the PW case). However, in most applications, the discretization parameters are not tight enough for the discretization error to be lower than the target accuracy, which is typically of the order of 1 kcal/mol or 1 mHa (recall that 1 mHa $\simeq$ 0.6275 kcal/mol $\simeq$ 27.2 meV, which corresponds to an equivalent temperature of about 316 K). It is often advocated that this is not an issue since the real quantity of interest is not the value of the energy E_R for a particular nuclear configuration R, but the energy difference $E_{R_1} - E_{R_2}$ between two different configurations R_1 and R_2. It is indeed expected that

$$|(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})| \ll |E_{R_1,N} - E_{R_1}| + |E_{R_2,N} - E_{R_2}|,$$

that is, the numerical error on the energy difference between the two configurations is much smaller than the sum of the discretization errors on the energies of each configuration. This expected phenomenon goes by the name of (discretization) error cancellation in the Physics and Chemistry literatures.

Obviously, for variational discretization methods, $E_{R_j,N} - E_{R_j} \geq 0$ so that both discretization errors have the same sign, leading to

$$\begin{aligned} |(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})| &= |(E_{R_1,N} - E_{R_1}) - (E_{R_2,N} - E_{R_2})| \\ &\leq \max\left(E_{R_1,N} - E_{R_1}, E_{R_2,N} - E_{R_2}\right), \end{aligned}$$

but this does not explain the magnitude of the error cancellation phenomenon. The commonly admitted *qualitative* argument usually raised to explain this phenomenon is that the errors $E_{R_1,N} - E_{R_1}$ and $E_{R_2,N} - E_{R_2}$ are of the same nature and almost annihilate one another.

The purpose of this article is to provide a *quantitive* analysis of discretization error cancellation for PW discretization methods. First, we report in Section 2 two systematic numerical studies on, respectively, the hydrogen molecule and a simple system consisting of six atoms. For these systems, we are able to perform very accurate calculations with high PW cut-offs and tight convergence criteria, which provide excellent approximations of the ground state energy E_R. We then compute, for two different configurations R_1 and R_2, the error cancellation factor

$$0 \leq Q_N := \frac{|(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})|}{|E_{R_1,N} - E_{R_1}| + |E_{R_2,N} - E_{R_2}|} \leq 1.$$

We observe that this ratio is indeed small (typically between 10^{-3} and 10^{-1} depending on the system and on the configurations R_1 and R_2), and that it does not vary much with N. In Section 3, we introduce a toy model consisting of seeking the ground state of a one-dimensional linear periodic Schrödinger equation with Dirac potentials:

$$\left(-\frac{d^2}{dx^2} - \sum_{m\in\mathbb{Z}} z_1\delta_m - \sum_{m\in\mathbb{Z}} z_2\delta_{m+R}\right) u_R = E_R u_R, \qquad \int_0^1 u_R^2(x)dx = 1,$$

for which we can prove that the error cancellation factor Q_N converges to a fixed number $0 < Q_\infty < 1$ when N goes to infinity. Interestingly, it is possible to obtain a simple explicit expression of Q_∞, which only depends on z_1, z_2 and on $u_{R_1}(0)^2$, $u_{R_2}(0)^2$, $u_{R_1}(R_1)^2$, $u_{R_1}(R_2)^2$, i.e. on the values of the densities $\rho_{R_1} = u_{R_1}^2$ and $\rho_{R_1} = u_{R_2}^2$ at the singularities of the potential.

An alternative way to estimate the error on the energy difference between two configurations R_1 and R_2 is to integrate the error on the atomic forces on a smooth path linking R_1 and R_2. We conclude Section 2 by showing that the latter approach is not efficient in general.

2 Discretization error cancellation in planewave calculations

We present here some numerical simulations on two systems: the H_2 molecule and a system consisting of two oxygen atoms and four hydrogen atoms. The simulations are done in a cubic supercell of size $10\times10\times10$

bohrs with the Abinit simulation package [14, 15]. The chosen approximate model is the periodic Kohn-Sham LDA model [20] with the parametrization and the pseudopotential proposed in [13]. Note that, in this work, we consider the approximation consisting of replacing the original problem set on the whole space $\mathbb{R}^3$ with a problem set on a cubic supercell with periodic boundary conditions as a *model error*. Alternatively, this error could be regarded as a discretization error: the supercell problem can indeed be seen as a non-consistent, non-conforming approximation of the original problem set on the whole space (see [6], in which this point of view was adopted to study the case of a local defect embedded in a perfect crystal).

For each configuration R, we compute a reference ground state energy E_R taking a high energy cutoff $E_{\text{cut}} = 400$ Ha. We then compute approximate energies for $N = E_{\text{cut}}$ varying from 5 to 105 Ha by steps of 5 Ha. The so-obtained energies are denoted by $E_{R,N}$.

For two given configurations R_1 and R_2 of the same system, we compute S_N, the sum of the discretization errors on the energies of the two configurations (note that $E_{R,N} - E_R \geq 0$ since PW is a variational approximation method), and D_N, the discretization error on the energy difference:

$$S_N = (E_{R_1,N} - E_{R_1}) + (E_{R_2,N} - E_{R_2}) \quad \text{and} \quad D_N = |(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})|,$$

as well as the error cancellation factor

$$Q_N = \frac{D_N}{S_N} = \frac{|(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})|}{(E_{R_1,N} - E_{R_1}) + (E_{R_2,N} - E_{R_2})}.$$

The two chemical systems considered in this section are very simple. We can therefore safely assume that for each configuration, our numerical simulations provide good approximations of the Kohn–Sham ground state. Besides, very tight convergence criteria are used, so that algorithmic errors are negligible. Implementation and computer errors are not expected to be significant in this context.

2.1 Ground state potential energy surface of the H_2 molecule

In all our calculations, the H_2 molecule lies on the x axis and is centered at the origin. The parameter R is here the interatomic distance in bohrs.

We numerically observe that D_N is smaller than S_N by a factor of 10 to 100, and that the error cancellation factor Q_N is smaller when the two interatomic distances are close to each other ($R_1 \simeq R_2$). Morevoer, Q_N is almost constant with respect to the cut-off energy N.

In Figure 1, we present detailed results for two different pairs of configurations. On the top, the configurations are rather close since the interatomic distances are $R_1 = 1.464$ and $R_2 = 1.524$ bohr. For this approximate model, the equilibrium distance is about $R_{\text{eq}} \simeq 1.464$ bohrs (the experimental value is $R_{\text{eq}}^{\text{exp}} \simeq 1.401$ bohrs). The energy difference is better approximated by a factor of about 50 compared to the energies ($Q_N \simeq 0.02$). Moreover the log-log plots of S_N and D_N are almost parallel, which suggests that there is no improvement in the order of convergence when considering energy differences instead of energies; only the prefactor is improved. This is confirmed by the plots of the error cancellation factor Q_N, showing that this ratio does not vary much with N. On the bottom, the configurations are further apart. The interatomic distances are $R_1 = 1.344$ and $R_2 = 1.704$ bohrs. We observe a similar behavior except that the error cancellation phenomenon is less pronounced ($Q_N \simeq 0.1$).

We then compare in Table 1 the values of S_N and D_N for different pairs of configurations and for two values of $N = E_{\text{cut}}$: a rather coarse energy cut-off $N = 30$ Ha, and a quite fine one $N = 100$ Ha. One configuration is kept fixed ($R_1 = 1.284$ bohrs), while the second one varies from $R_2 = 1.344$ bohrs (close configurations) to $R_2 = 1.764$ bohrs (distant configurations). We also report, for each pair of configurations, the minimum, maximum, and mean values of Q_N over the different tested energy cutoffs $5 \leq N \leq 105$ Ha. We also observe that Q_N increases with $R_2 - R_1$ on the range $R_2 = [1.344, 1.764]$.

2.2 Energy of a simple chemical reaction

In this section, we consider the energy difference between two very different configurations of a system consisting of two oxygen atoms and four hydrogen atoms. The first configuration, denoted by R_1, corresponds

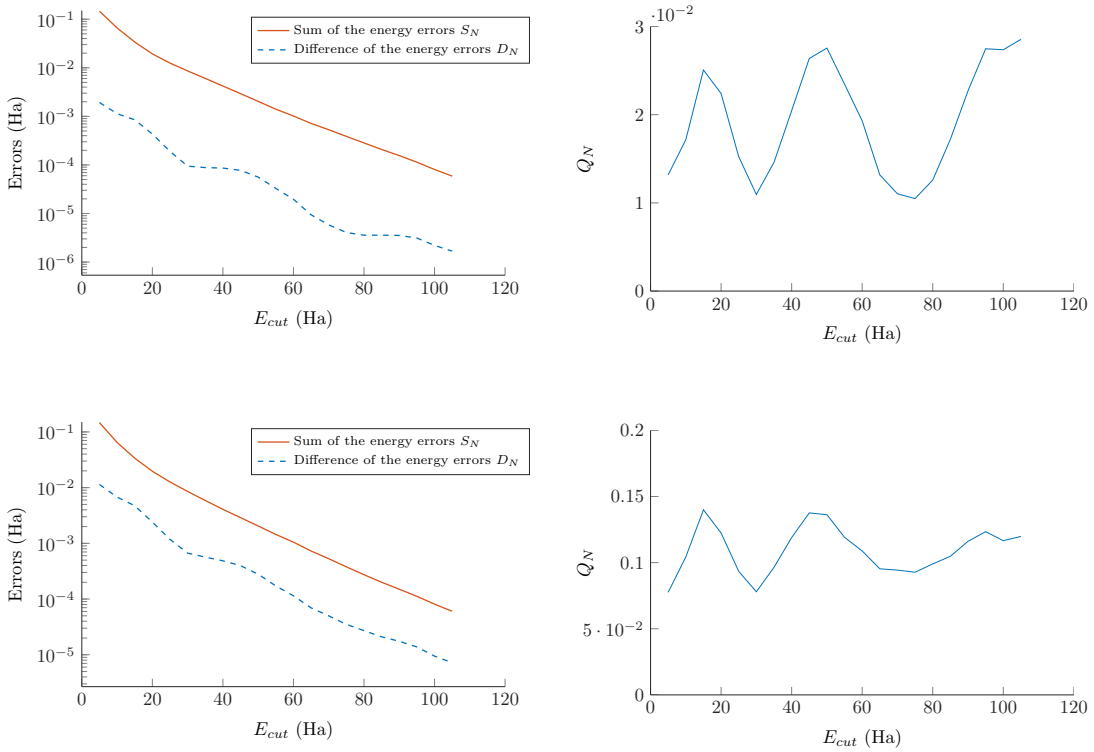

Figure 1: Convergence plots of the quantities S_N and D_N (left) and of the error cancellation factor $Q_N = D_N/S_N$ (right) for two different pairs of interatomic distances for the H_2 molecule. Top: $R_1 = 1.464$ and $R_2 = 1.524$ bohrs. Bottom: $R_1 = 1.344$ and $R_2 = 1.704$ bohrs.

to the chemical system 2 H_2O (two water molecules) and the second one, denoted by R_2, to the chemical system 2 H_2 + O_2, all these molecules being in their equilibrium geometry (see Figure 2). The energy difference between the two configurations thus provides a rough estimate of the energy of the chemical reaction

$$2\ H_2 + O_2 \longrightarrow 2\ H_2O.$$

We can observe on Figure 3 and Table 2 a similar behavior as for H_2, but with a better error cancellation factor ($Q_N \simeq 0.005$).

3 Mathematical analysis of a toy model

We now present a simple one-dimensional periodic linear Schrödinger model for which the discretization error cancellation phenomenon observed in the previous section can be explained with full mathematical rigor.

We denote by

$$L^2_{\mathrm{per}} := \left\{ u \in L^2_{\mathrm{loc}}(\mathbb{R}) \;\middle|\; u \text{ is } 1-\text{periodic} \right\}$$

the vector space of the 1-periodic locally square integrable real-valued functions on $\mathbb{R}$, and by

$$H^1_{\mathrm{per}} := \left\{ u \in L^2_{\mathrm{per}} \;\middle|\; u' \in L^2_{\mathrm{per}} \right\}$$

the associated order-1 Sobolev space. For two given parameters $z_1, z_2 > 0$, we consider the family of problems,

R_1	R_2	$S_{N=30}$	$D_{N=30}$	$S_{N=100}$	$D_{M=100}$	$\min(Q_N)$	$\max(Q_N)$	$\text{mean}(Q_N)$
1.284	1.344	9.410	0.1985	0.09157	0.00112	0.0103	0.0340	0.0212
1.284	1.404	9.268	0.3408	0.08990	0.00279	0.0216	0.0633	0.0413
1.284	1.464	9.160	0.4491	0.08772	0.00497	0.0375	0.0895	0.0610
1.284	1.524	9.065	0.5436	0.08552	0.00717	0.0544	0.1107	0.0802
1.284	1.584	8.969	0.6394	0.08380	0.00889	0.0713	0.1285	0.0985
1.284	1.644	8.863	0.7456	0.08274	0.00995	0.0841	0.1455	0.1151
1.284	1.704	8.744	0.8646	0.08213	0.01056	0.0983	0.1642	0.1302
1.284	1.764	8.615	0.9937	0.08154	0.01115	0.1072	0.1802	0.1440

Table 1: Comparison of S_N, D_N and Q_N for different atomic configurations of the H_2 molecule. Distances are in bohrs, energies in mHa.

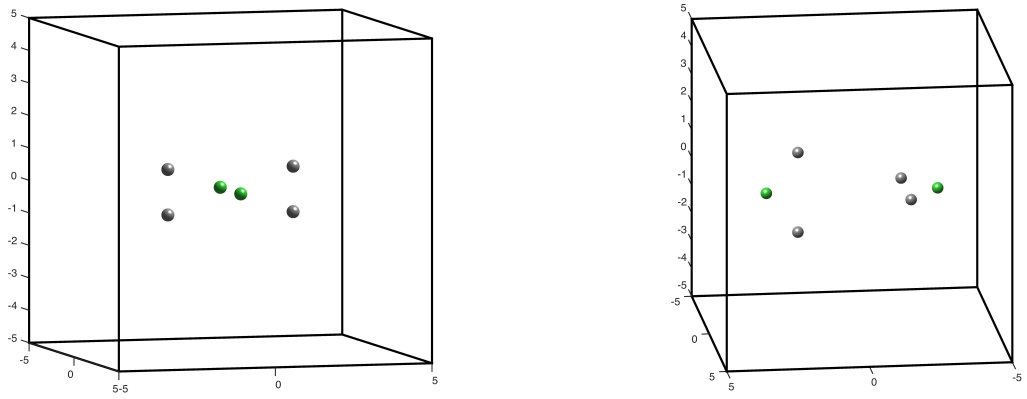

Figure 2: Graphical representation of the two atomic configurations whose energies are compared. Oxygen atoms are in green, hydrogen atoms in black.

indexed by $R \in (0,1)$, consisting in finding the ground state $(u_R, E_R) \in H^1_{\text{per}} \times \mathbb{R}$ of

$$\begin{cases} \left(-\dfrac{d^2}{dx^2} - \displaystyle\sum_{m\in\mathbb{Z}} z_1\delta_m - \sum_{m\in\mathbb{Z}} z_2\delta_{m+R}\right) u_R = E_R u_R, \\ \displaystyle\int_0^1 u_R^2(x)dx = 1, \quad u_R \geq 0, \end{cases} \tag{1}$$

where δ_a denotes the Dirac mass at point $a \in \mathbb{R}$. A variational formulation of the problem is: find the ground state $(u_R, E_R) \in H^1_{\text{per}} \times \mathbb{R}$ of

$$\begin{cases} \forall v \in H^1_{\text{per}}, \ \displaystyle\int_0^1 u_R'(x)v'(x)dx - z_1 u_R(0)v(0) - z_2 u_R(R)v(R) = E_R \int_0^1 u_R(x)v(x)dx, \\ \displaystyle\int_0^1 u_R^2(x)dx = 1, \quad u_R \geq 0. \end{cases} \tag{2}$$

Remark 1. *The ground state eigenvalue E_R is negative. Indeed, using the variational characterization of the ground state energy, we get*

$$E_R = \min_{v \in H^1_{\text{per}} \setminus \{0\}} \frac{\displaystyle\int_0^1 v'(x)^2dx - z_1 v(0)^2 - z_2 v(R)^2}{\displaystyle\int_0^1 v^2(x)dx} < 0,$$

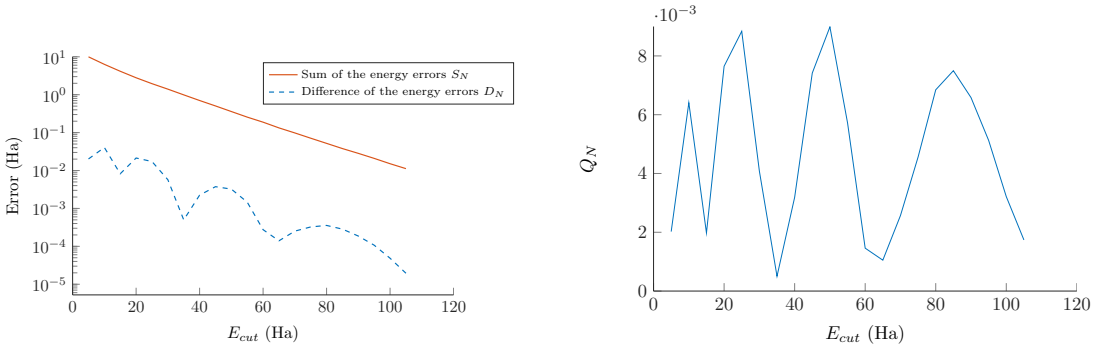

Figure 3: Convergence plots of the quantities S_N and D_N (left) and of the error cancellation factor $Q_N = D_N/S_N$ (right) for the two different configurations displayed on Figure 2.

$S_{N=30}$	$D_{N=30}$	$S_{N=100}$	$D_{N=100}$	$\min(Q_N)$	$\max(Q_N)$	$\mathrm{mean}(Q_N)$
1403	5.726	15.12	0.0485	0.0005036	0.008986	0.004640

Table 2: Comparison of S_N, D_N (in mHa) and Q_N for the two different configurations displayed on Figure 2.

since the Rayleigh quotient is equal to $-z_1 - z_2 < 0$ *for the constant test function* $v = 1$.

Denoting by $k_R = \sqrt{-E_R}$, we have

$$\begin{cases} u_R(x) = Ae^{k_R x} + Be^{-k_R x}, & \forall x \in [0, R], \\ u_R(x) = Ce^{k_R x} + De^{-k_R x}, & \forall x \in [R-1, 0), \end{cases} \tag{3}$$

where A, B, C, and D are real-valued constants. Since the function u_R is 1-periodic and continuous on $\mathbb{R}$ and its derivative satisfies the jump conditions $u_R'(m+0) - u_R'(m-0) = -z_1 u_R(m)$ and $u_R'(m+R+0) - u_R'(m+R-0) = -z_2 u_R(m+R)$ for all $m \in \mathbb{Z}$, the coefficients A, B, C, D solve the linear system

$$\underbrace{\begin{pmatrix} 1 & 1 & -1 & -1 \\ e^{k_R R} & e^{-k_R R} & -e^{k_R(R-1)} & -e^{-k_R(R-1)} \\ k_R + z_1 & -k_R + z_1 & -k_R & k_R \\ (k_R - z_2)e^{k_R R} & -(k_R + z_2)e^{-k_R R} & -k_R e^{k_R(R-1)} & k_R e^{-k_R(R-1)} \end{pmatrix}}_{M(k_R)} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The wave vector k_R is the lowest positive root of the function $k \mapsto \det(M(k))$. The coefficients (A, B, C, D) are then uniquely determined by the normalization condition $\|u_R\|_{L^2_{\mathrm{per}}} = 1$ and the positivity of u_R. Exact solutions for two different values of the triplet of parameters (z_1, z_2, R) are plotted in Figure 4.

An approximate solution of the problem is obtained using the PW discretization method. Denoting by

$$X_N := \mathrm{Span}\left\{ v_N(x) = \sum_{k \in \mathbb{Z},\ |k| \le N} \widehat{v}_k e^{2\pi i k x} \;\middle|\; \widehat{v}_k \in \mathbb{C},\ \widehat{v}_{-k} = \overline{\widehat{v}_k} \right\} \subset H^1_{\mathrm{per}},$$

the variational approximation of problem (2) in X_N consists in computing the ground state $(u_{R,N}, E_{R,N}) \in X_N \times \mathbb{R}$ of

$$\begin{cases} \forall v_N \in X_N, & \displaystyle\int_0^1 u_{R,N}' v_N' - z_1 u_{R,N}(0) v_N(0) - z_2 u_{R,N}(R) v_N(R) \;= E_{R,N} \int_0^1 u_{R,N} v_N, \\ \displaystyle\int_0^1 u_{R,N}^2 = 1, & \displaystyle\int_0^1 u_{R,N} \ge 0. \end{cases} \tag{4}$$

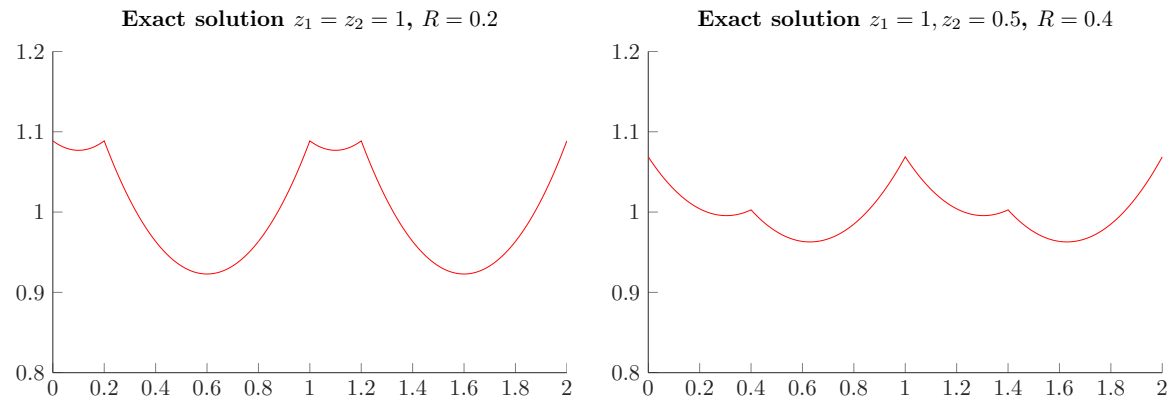

Figure 4: Plot of the exact solutions of (1) for two sets of parameters.

The conditions $\widehat{v}_{-k} = \overline{\widehat{v}_k}$ in the definition of X_N is equivalent to imposing that the elements of X_N are real-valued functions. For convenience, the discretization parameter N here corresponds to the cut-off in momentum space. As above, we consider the error cancellation factor

$$Q_N = \frac{|(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})|}{(E_{R_1,N} - E_{R_1}) + (E_{R_2,N} - E_{R_2})} \tag{5}$$

associated with the pair of configurations (R_1, R_2).

Note that imposing the condition $\int_0^1 u_{R,N} \geq 0$, we ensure that the discrete eigenfunction $u_{R,N}$ will approximate the positive eigenfunction u_R to the continuous problem (1) and not $-u_R$.

Theorem 1 (Asymptotic expressions of the energy error and of the error cancellation factor)**.** *For all $z_1, z_2 > 0$ and $R \in (0,1)$, we have for all $\epsilon > 0$,*

$$E_{R,N} - E_R = \frac{\alpha_R}{N} - \frac{\alpha_R}{2N^2} + \frac{\beta^{(1)}_{R,N}}{N} + \frac{\gamma_R}{N}\eta_{R,N} + o\left(\frac{1}{N^{3-\epsilon}}\right), \tag{6}$$

where

$$\alpha_R := \frac{z_1^2 u_R(0)^2 + z_2^2 u_R(R)^2}{2\pi^2}, \quad \gamma_R := \frac{z_1 z_2 u_R(0) u_R(R)}{\pi^2}, \quad \eta_{R,N} := N \sum_{k=N+1}^{+\infty} \frac{\cos(2\pi k R)}{k^2},$$

$$\beta^{(1)}_{R,N} := \frac{z_1^2 u_R(0)(u_{R,N}(0) - u_R(0)) + z_2^2 u_R(R)(u_{R,N}(R) - u_R(R))}{2\pi^2}.$$

In addition

$$|\eta_{R,N}| \leq \min\left(1, \frac{2 + \frac{\pi^3}{8}}{|\sin(\pi R)| N}\right),$$

and for all $\epsilon > 0$, there exists $C_\epsilon \in \mathbb{R}_+$ such that

$$|\beta^{(1)}_{R,N}| \leq \frac{C_\epsilon}{N^{1-\epsilon}}.$$

As a consequence, we have for all $z_1, z_2 > 0$ and all $R_1, R_2 \in (0,1)$,

$$\lim_{N \to +\infty} Q_N = \frac{|\alpha_{R_1} - \alpha_{R_2}|}{\alpha_{R_1} + \alpha_{R_2}} = \frac{\left|z_1^2\left(u_{R_1}(0)^2 - u_{R_2}(0)^2\right) + z_2^2\left(u_{R_1}(R_1)^2 - u_{R_2}(R_2)^2\right)\right|}{z_1^2(u_{R_1}(0)^2 + u_{R_2}(0)^2) + z_2^2(u_{R_1}(R_1)^2 + u_{R_2}(R_2)^2)}. \tag{7}$$

The proof of the above theorem is given in Appendix. We deduce from (6) that the discretization error $E_{R,N} - E_R$ on the energy of the configuration R is the sum of

1. a leading term $\alpha_R N^{-1}$ of order 1 (in N^{-1});
2. three terms $-1/2\alpha_R N^{-2}$, $\beta^{(1)}_{R,N} N^{-1}$, and $\gamma_R N^{-1} \eta_{R,N}$ which are roughly of order 2;
3. higher order terms which are roughly of order 3 and above.

The leading term $\alpha_R N^{-1}$ has a very simple expression and the prefactor α_R does not vary much with respect to R (see Figure 5). This explains the phenomenon of discretization error cancellation. Regarding the second order corrections on $E_{R,N} - E_R$, we have observed numerically (see Figure 6) that

- the terms $-\frac{1}{2}\alpha_R N^{-2}$ and $\gamma_R N^{-1} \eta_{R,N}$ are of about the same order of magnitude in absolute values, that the former is always negative (since $\alpha_R > 0$), but that the latter can be either positive or negative, so that the sum of these two contributions can be either significant or negligible;
- the term $\beta^{(1)}_{R,N} N^{-1}$ is smaller in absolute value than the other two terms, and seems to be always negative. Our numerical calculations indeed show that $u_{R,N}(0) < u_R(0)$ and $u_{R,N}(R) < u_R(R)$, which is not very surprising since the function u_R has cusps at points $x = 0$ and $x = R$ (see Figure 4). These inequalities have not been rigorously established though.

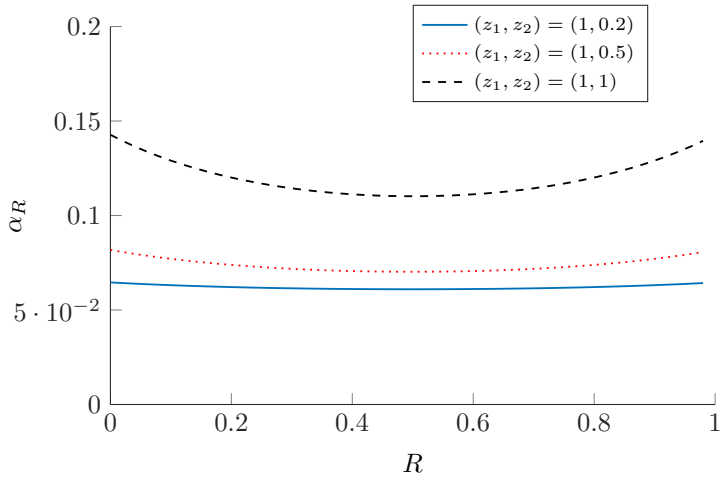

Figure 5: Plots of the function $R \mapsto \alpha_R$ for three sets of parameters (z_1, z_2).

Finally, we observe on Figure 7 that Q_N converges to the asymptotic value Q_∞ when N goes to infinity very smoothly for large values of R, and with oscillations when R becomes close to zero. Moreover, $Q_N - Q_\infty$ is of order N^{-2}.

Remark 2. *The 1D model studied in this section involves Dirac potentials, for which the exact solutions* (3)*, as well as the lowest-order terms of the discretization error* (6)*, can be computed explicitly. It would have been possible to use more regular potentials with explicit solutions, such as piecewise constant potentials for instance. However, the calculations would have been more tedious than for the Dirac case, and we anticipate that, qualitatively, the results would have been similar. Loosely speaking, the faster convergence of the energy difference originates from the fact that the leading term of the error depends on the nuclear configuration, but not that much. This explains why the convergence rate is not improved, while the prefactor is improved.*

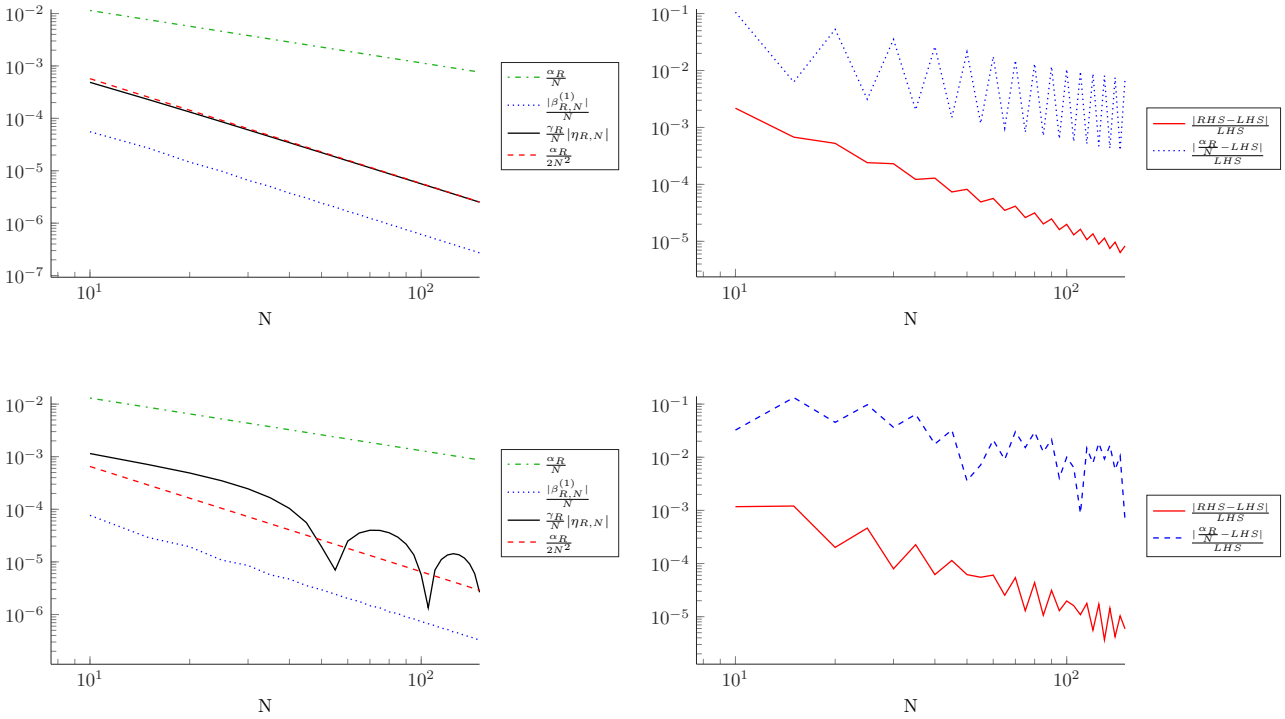

Figure 6: Convergence plots of the four quantities $\frac{\alpha_R}{N}$, $\frac{\alpha_R}{2N^2}$, $\frac{|\beta^{(1)}_{R,N}|}{N}$, and $\frac{\gamma_R}{N}|\eta_{R,N}|$ (left) and plots of $\frac{|(\frac{\alpha_R}{N}-\frac{\alpha_R}{2N^2}+\frac{\beta^{(1)}_{R,N}}{N}+\frac{\gamma_R}{N}\eta_{R,N})-(E_{R,N}-E_R)|}{E_{R,N}-E_R}$ and $\frac{|\frac{\alpha_R}{N}-(E_{R,N}-E_R)|}{E_{R,N}-E_R}$ (right). Top: $z_1 = z_2 = 1, R = 0.3$. Bottom: $z_1 = z_2 = 1, R = 0.09$.

For smoother potentials, as well as for pseudopotentials, it is expected that most of the error on the energy remains concentrated in the vicinities of the core regions, where, for different nuclear configurations, the electronic orbitals change, but not much.

Remark 3. *Note that a variant of the projected augmented wave (PAW) method [2] was recently studied for the 1D model considered here [1]: it is shown that the error on the energy has two contributions, the first one scaling as $r_c^{4N_0} N^{-1}$, and the second one as $r_c^{-p} N^{-(p+1)}$, where r_c is the core radius, N_0 the number of pseudo-orbitals, p the degree of the (polynomial) pseudo-orbitals in the core region, and N the number of planewaves. However, it is not clear how to use the estimates in [1] to obtain estimates on energy differences. We intend to investigate this point in the future.*

To conclude, let us comment on the alternative approach to estimate the error on the energy difference between two configurations consisting in integrating the error on the atomic forces along a path in the nuclear configuration space liking the two configurations. In this simple 1D setting, we have, for $R_1 < R_2$,

$$|(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})| = \left| \int_{R_1}^{R_2} (F_{R,N} - F_R)\, dR \right|, \text{ where } F_{N,R} := -\frac{dE_{R,N}}{dR} \text{ and } F_R := -\frac{dE_R}{dR}.$$

The use of a variational method guaranties that the energy error $E_{R,N} - E_R$ is nonnegative for all N and all R. On the other hand, the error on the force $F_{R,N} - F_R$ does not have a constant sign (it integrates to zero on the interval $[0, 1]$), so that, in general,

$$|(E_{R_1,N} - E_{R_2,N}) - (E_{R_1} - E_{R_2})| = \left| \int_{R_1}^{R_2} (F_{R,N} - F_R)\, dR \right| \leq \int_{R_1}^{R_2} |F_{R,N} - F_R|\, dR.$$

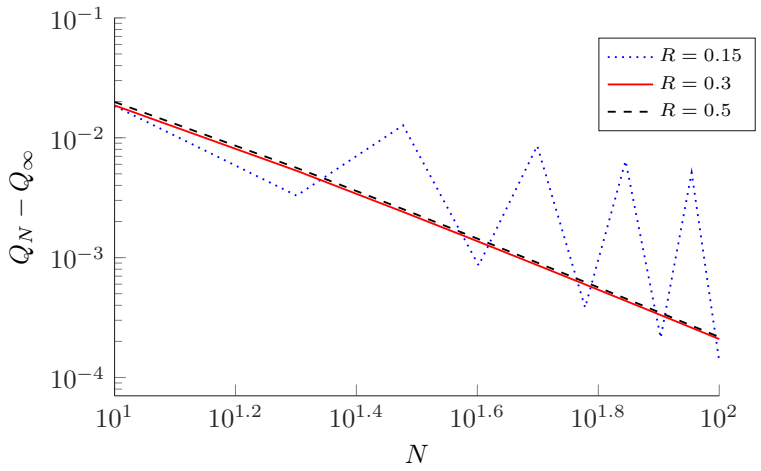

Figure 7: Plot of $Q_N - Q_\infty$ for three values of R.

The left hand-side of the above inequality can *a priori* be much smaller than the right hand-side. In this case, using bounds on the error on the forces would lead to a dramatic overestimation of the error on the energy difference. This is confirmed by our numerical simulations. The functions

$$(R_1, R_2) \mapsto \left| \int_{R_1}^{R_2} (F_{R,N} - F_R)\, dR \right| \quad \text{and} \quad (R_1, R_2) \mapsto \int_{R_1}^{R_2} |F_{R,N} - F_R|\, dR, \tag{8}$$

plotted in Figure 8, are very different and the latter one is not a good approximation of the former one. Another interesting observation is the following. Numerical simulations show that the forces converge at the

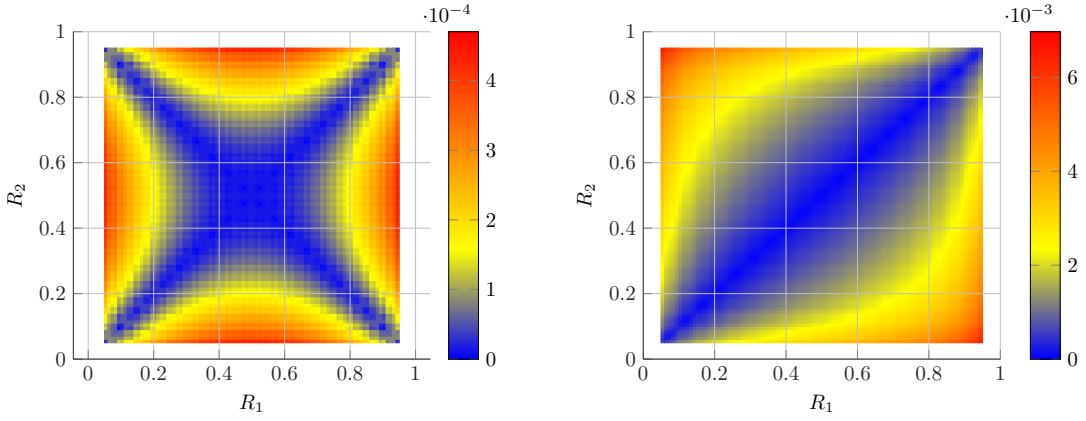

Figure 8: Colorplots of the functions defined in (8). The forces were computed with centered finite difference with step size 10^{-6} and the integrals with Simpson's rule with step length 10^{-2}, chosen equal to the resolution of the figure.

same rate as the energy, i.e. in $1/N$ (see Figure 9), and that, for each value of N in the range $[10, 100]$, the

derivatives of the functions

$$R \mapsto E_{R,N} - E_R \quad \text{and} \quad R \mapsto \chi_{R,N} := \frac{\alpha_R}{N} - \frac{\alpha_R}{2N^2} + \frac{\beta^{(1)}_{R,N}}{N} + \frac{\gamma_R}{N}\eta_{R,N}$$

agree up to very small correction terms. Nevertheless, the derivative of the fourth term in $\chi_{R,N}$ (i.e. of $\gamma_R \eta_{R,N} N^{-1}$) can be much larger than the derivative of the first term (i.e. of $\alpha_R N^{-1}$). The leading term of the error on the force is therefore not in general (minus) the derivative of the leading term of the energy error. In Figure 10, the above functions are plotted for $N = 10$ (top) and $N = 100$ (bottom).

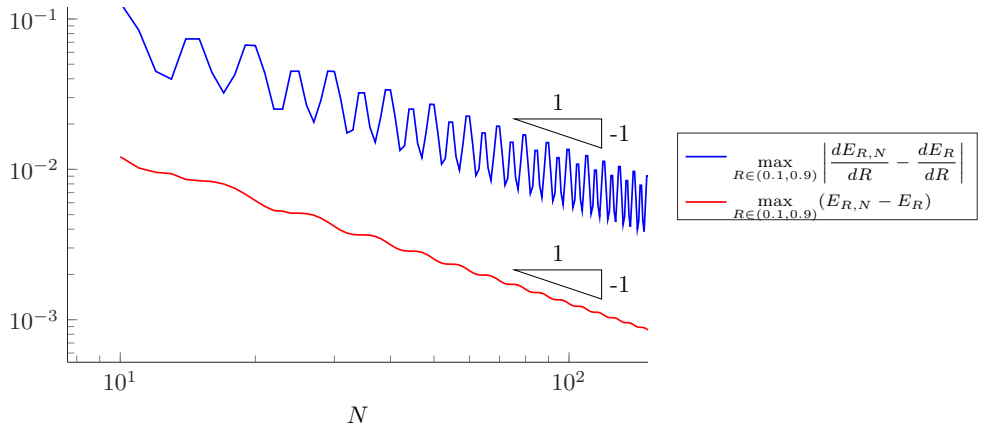

Figure 9: Convergence of the errors on the energy (in red) and on the forces (in blue).

4 Appendix: proof of Theorem 1

In the sequel, z_1 and z_2 are fixed positive real numbers. We endow the functional spaces $L^2_{\rm per}$ and $H^1_{\rm per}$ with their usual scalar products

$$\langle u|v\rangle_{L^2_{\rm per}} := \int_0^1 u(x)v(x)\,dx \quad \text{and} \quad \langle u|v\rangle_{H^1_{\rm per}} := \langle u|v\rangle_{L^2_{\rm per}} + \langle u'|v'\rangle_{L^2_{\rm per}}.$$

More generally, we endow the Sobolev space

$$H^s_{\rm per} := \left\{ v(x) = \sum_{k\in\mathbb{Z}} \widehat{v}_k e^{2i\pi kx} \;\middle|\; \widehat{v}_k \in \mathbb{C},\ \widehat{v}_{-k} = \overline{\widehat{v}_k},\ \sum_{k\in\mathbb{Z}} (1+(2\pi k)^2)^s |\widehat{v}_k|^2 < \infty \right\},$$

$s \in \mathbb{R}$, with the scalar product defined by

$$\langle u|v\rangle_{H^s_{\rm per}} := \sum_{k\in\mathbb{Z}} (1+(2\pi k)^2)^s\, \overline{\widehat{u}_k}\, \widehat{v}_k.$$

Note that the above two definitions of $\langle u|v\rangle_{H^1_{\rm per}}$ coincide and that $H^0_{\rm per} = L^2_{\rm per}$. We also denote by Π_N the orthogonal projection on X_N for the $L^2_{\rm per}$ (and also $H^s_{\rm per}$) scalar product and by $\Pi_N^\perp = 1 - \Pi_N$.

We first recall some useful results on the convergence of $(u_{R,N}, E_{R,N})$ to (u_R, E_R).

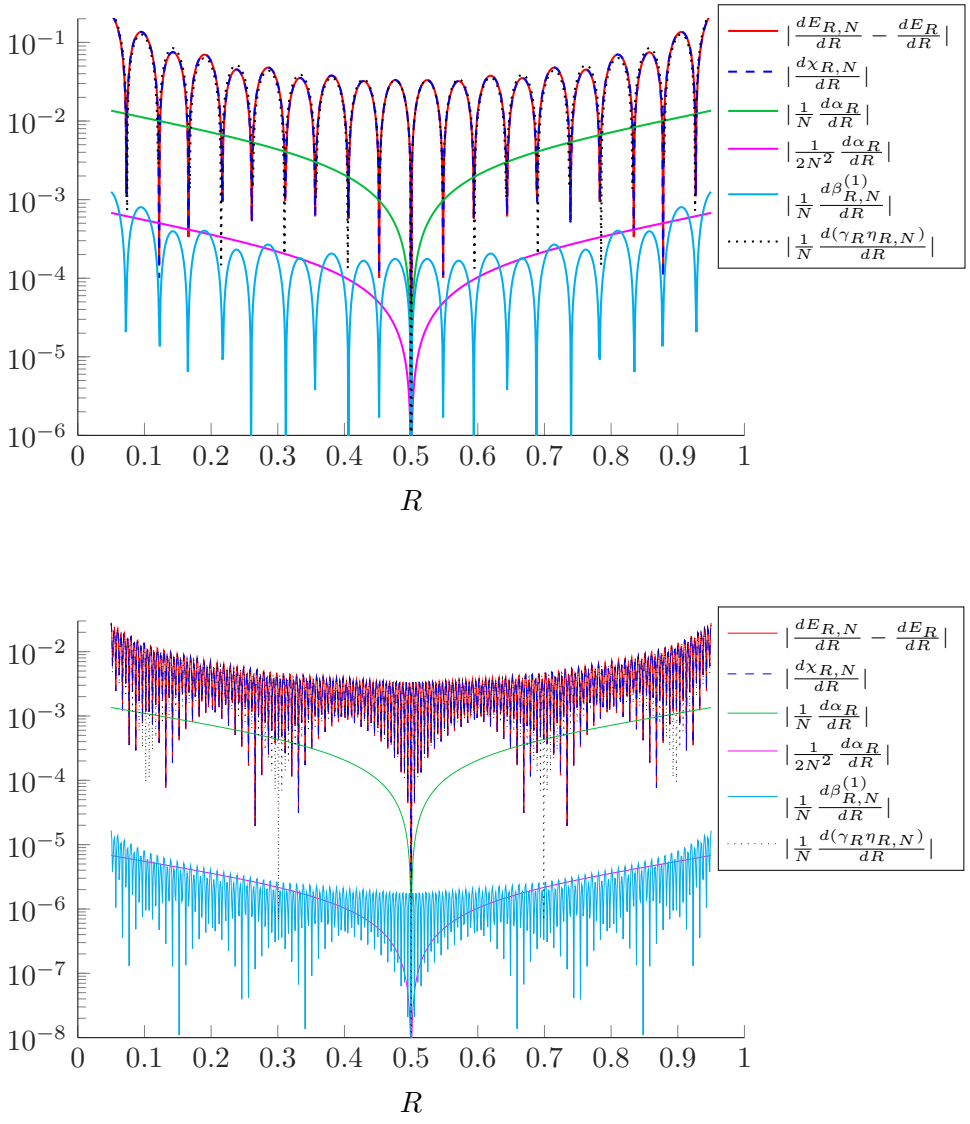

Figure 10: Plots of the functions $R \mapsto \frac{dE_{R,N}}{dR} - \frac{dE_{R,N}}{dR}$ and $R \mapsto \frac{d\chi_{R,N}}{dR}$, and of the derivative of each of the four components of $\chi_{R,N}$, for $N = 10$ (top) and $N = 100$ (bottom). The derivatives were computed numerically by centered finite differences with step size 10^{-6}.

Lemma 1. *Let $R \in (0,1)$. Let (u_R, E_R) be the ground state of the continuous problem (2), and $(u_{N,R}, E_{R,N})$ be a ground state of the discretized problem (4). Then, for all $\epsilon > 0$ and all $0 \le s < 3/2$, there exists $C_{s,\epsilon} \in \mathbb{R}_+$ such that*

$$\|u_{R,N} - u_R\|_{H^s_{\mathrm{per}}} \le \frac{C_{s,\epsilon}}{N^{3/2-s-\epsilon}}. \tag{9}$$

In addition, there exist $0 < c \le C < \infty$ such that

$$c\|u_{R,N} - u_R\|^2_{H^1_{\mathrm{per}}} \le E_{R,N} - E_R \le C\|u_{R,N} - u_R\|^2_{H^1_{\mathrm{per}}}, \tag{10}$$

and for all $\epsilon > 0$, there exists $C_\epsilon \in \mathbb{R}_+$ such that

$$|u_{R,N}(0) - u_R(0)| + |u_{R,N}(R) - u_R(R)| \leq \frac{C_\epsilon}{N^{1-\epsilon}}. \tag{11}$$

Proof. We denote by C^0_{per} the space of continuous 1-periodic functions from $\mathbb{R}$ to $\mathbb{R}$ endowed with the norm defined by

$$\forall u \in C^0_{\mathrm{per}}, \quad \|u\|_{C^0_{\mathrm{per}}} := \max_{x\in\mathbb{R}} |u(x)|.$$

Recall that H^s_{per} is continuously embedded in C^0_{per} for all $s > 1/2$. In particular, $H^1_{\mathrm{per}} \hookrightarrow C^0_{\mathrm{per}}$ and there exists $K \in \mathbb{R}_+$ such that

$$\forall u \in H^1_{\mathrm{per}}, \quad \|u\|_{C^0_{\mathrm{per}}} \leq K\|u\|_{H^{3/4}_{\mathrm{per}}} \leq K\|u\|^{3/4}_{H^1_{\mathrm{per}}}\|u\|^{1/4}_{L^2_{\mathrm{per}}}. \tag{12}$$

In particular, the bilinear form

$$\forall (u,v) \in H^1_{\mathrm{per}} \times H^1_{\mathrm{per}}, \quad a_R(u,v) = \int_0^1 u'v' - z_1u(0)v(0) - z_2u(R)v(R)$$

is well-defined, symmetric, and continuous on $H^1_{\mathrm{per}} \times H^1_{\mathrm{per}}$, and we have

$$\begin{aligned}\forall u \in H^1_{\mathrm{per}}, \quad a_R(u,u) &\geq \|u\|^2_{H^1_{\mathrm{per}}} - (z_1+z_2)K^2\|u\|^{3/2}_{H^1_{\mathrm{per}}}\|u\|^{1/2}_{L^2_{\mathrm{per}}} - \|u\|^2_{L^2_{\mathrm{per}}} \\ &\geq \frac{1}{2}\|u\|^2_{H^1_{\mathrm{per}}} - \left(1 + \frac{27}{32}(z_1+z_2)^4K^8\right)\|u\|^2_{L^2_{\mathrm{per}}},\end{aligned}$$

using Young's inequality. The quadratic form $H^1_{\mathrm{per}} \ni u \mapsto a_R(u,u) \in \mathbb{R}$ therefore is bounded below and closed. We denote by H_R the unique self-adjoint operator on L^2_{per} associated to $a_R(\cdot,\cdot)$ (see e.g. [32, Theorem VIII.15]). Formally,

$$H_R = -\frac{d^2}{dx^2} - z_1\sum_{m\in\mathbb{Z}}\delta_m - z_2\sum_{m\in\mathbb{Z}}\delta_{m+R}.$$

The domain of H_R being a subspace of H^1_{per}, which is itself compactly embedded in L^2_{per}, the spectrum of H_R is purely discrete: it consists of an increasing sequence of eigenvalues of finite multiplicities going to $+\infty$. It is easily seen that its ground state eigenvalue E_R is simple. Let us denote by $\mu_R > 0$ the gap between the lowest two eigenvalues of H_R. A classical calculation shows that

$$\begin{aligned}E_{R,N} - E_R &= a_R(u_{R,N} - u_R, u_{R,N} - u_R) - E_R\|u_{R,N} - u_R\|^2_{L^2_{\mathrm{per}}} \\ &= \langle u_{R,N}|H_R|u_{R,N}\rangle - E_R.\end{aligned}$$

First, since $E_R < 0$, we have

$$E_{R,N} - E_R \leq a_R(u_{R,N} - u_R, u_{R,N} - u_R) \leq M_R\|u_{R,N} - u_R\|^2_{H^1_{\mathrm{per}}},$$

where M_R is the continuity constant of a_R, which proves the second inequality in (10). Second, since $\|u_R\|_{L^2_{\mathrm{per}}} = \|u_{R,N}\|_{L^2_{\mathrm{per}}} = 1$, we have on the one hand

$$\begin{aligned}E_{R,N} - E_R = \langle u_{R,N}|H_R|u_{R,N}\rangle - E_R &\geq \left(E_R|\langle u_{R,N}|u_R\rangle_{L^2_{\mathrm{per}}}|^2 + (E_R+\mu_R)\left(1 - |\langle u_{R,N}|u_R\rangle_{L^2_{\mathrm{per}}}|^2\right)\right) - E_R \\ &= \mu_R\left(1 - |\langle u_{R,N}|u_R\rangle_{L^2_{\mathrm{per}}}|^2\right) \geq \mu_R\left(1 - \langle u_{R,N}|u_R\rangle_{L^2_{\mathrm{per}}}\right) = \frac{\mu_R}{2}\|u_{R,N} - u_R\|^2_{L^2_{\mathrm{per}}},\end{aligned}$$

and, on the other hand,

$$E_{R,N} - E_R \geq \frac{1}{2}\|u_{R,N} - u_R\|^2_{H^1_{\mathrm{per}}} - \left(1 + \frac{27}{32}(z_1+z_2)^4K^8 + E_R\right)\|u_{R,N} - u_R\|^2_{L^2_{\mathrm{per}}}.$$

Combining the above two inequalities yields the first inequality in (10). Hence, (10) is proved.

We deduce from the min-max principle that for each $v_N \in X_N$ such that $\|v_N\|_{L^2_{\rm per}} = 1$, we have

$$\begin{aligned} E_{R,N} - E_R &\leq a_R(v_N, v_N) - E_R = a_R(v_N - u_R, v_N - u_R) - E_R \|v_N - u_R\|^2_{L^2_{\rm per}} \\ &\leq (M_R - E_R) \|v_N - u_R\|^2_{H^1_{\rm per}}. \end{aligned}$$

Since $z_1 \sum_{m\in\mathbb{Z}} \delta_m + z_2 \sum_{m\in\mathbb{Z}} \delta_{m+R} \in H^{-1/2-\epsilon}_{\rm per}$ for all $\epsilon > 0$, we have that $u_R \in H^{3/2-\epsilon}_{\rm per}$. Applying the above estimate to $v_N = \|\Pi_N u_R\|^{-1}_{L^2_{\rm per}} \Pi_N u_R$, we get $E_{R,N} - E_R \leq \frac{C_\epsilon}{N^{1-\epsilon}}$. Combining with (10), we obtain (9) for $s = 1$. Together with (12), this implies in addition that $(u_{R,N})_{N\in\mathbb{N}}$ converges to u_R in $C^0_{\rm per}$. Since

$$-u''_{R,N} = z_1 u_{R,N}(0)\Pi_N \left(\sum_{k\in\mathbb{Z}} \delta_m\right) + z_2 u_{R,N}(R)\Pi_N \left(\sum_{k\in\mathbb{Z}} \delta_{m+R}\right) + E_{R,N} u_{R,N},$$

and the right hand-side converges to $-u''_R$ in $H^{-1/2-\epsilon}_{\rm per}$ for all $\epsilon > 0$, the sequence $(u_{R,N})_{N\in\mathbb{N}}$ converges to u_R in $H^{3/2-\epsilon}_{\rm per}$ for all $\epsilon > 0$. By interpolation, we then obtain (9) for all $1 \leq s < 3/2$. We finally obtain (9) for $s = 0$ by a classical Aubin-Nitsche argument, and we conclude by interpolation that the result also holds true for all $0 \leq s < 1$.

To prove (11), we infer from the Sobolev embedding $H^{1/2+\epsilon}_{\rm per} \hookrightarrow C^0_{\rm per}$, that

$$|u_{R,N}(0) - u_R(0)| + |u_{R,N}(R) - u_R(R)| \leq 2\|u_{R,N} - u_R\|_{C^0_{\rm per}} \leq 2C'_\epsilon \|u_{R,N} - u_R\|_{H^{1/2+\epsilon}_{\rm per}},$$

and we conclude using (9) with $s = 1/2 + \epsilon$. □

The following lemma provides an expression of the leading term of the energy difference $E_{R,N} - E_R$.

Lemma 2. *Let $z_1, z_2 > 0$. Let $R \in (0,1)$. Let (u_R, E_R) be the ground state of the continuous problem (2), and $(u_{R,N}, E_{R,N})$ be a ground state of the discretized problem (4). Then, for all $\epsilon > 0$,*

$$E_{R,N} - E_R = z_1 u_{R,N}(0)(\Pi_N^\perp u_R)(0) + z_2 u_{R,N}(R)(\Pi_N^\perp u_R)(R) + o\left(\frac{1}{N^{3-\epsilon}}\right), \tag{13}$$

when N goes to $+\infty$.

Proof. The variational formulation (2) with $v = u_{R,N}$ gives

$$E_R \int_0^1 u_{R,N} u_R = \int_0^1 u'_{R,N} u'_R - z_1 u_{R,N}(0) u_R(0) - z_2 u_{R,N}(R) u_R(R).$$

The variational formulation (4) with $v_N = \Pi_N u_R$ gives

$$E_{R,N} \int_0^1 u_{R,N}(\Pi_N u_R) = \int_0^1 u'_{R,N}(\Pi_N u_R)' - z_1 u_{R,N}(0)(\Pi_N u_R)(0) - z_2 u_{R,N}(R)(\Pi_N u_R)(R).$$

Subtracting these two equalities, and noting first that $\displaystyle\int_0^1 u_{R,N}(\Pi_N u_R) = \int_0^1 u_{R,N} u_R$, and second that $\displaystyle\int_0^1 u'_{R,N}(\Pi_N u_R)' = \int_0^1 u'_{R,N} u'_R$, since $u_{R,N} \in X_N$ and the orthogonal projection Π_N and the derivation commute, we get

$$(E_{R,N} - E_R) \int_0^1 u_{R,N} u_R = z_1 u_{R,N}(0)(\Pi_N^\perp u_R)(0) + z_2 u_{R,N}(R)(\Pi_N^\perp u_R)(R).$$

Moreover, since $\int_0^1 u_R^2 = \int_0^1 u_{R,N}^2 = 1$, we have

$$\int_0^1 u_{R,N} u_R = 1 - \frac{1}{2}\int u_R^2 - \frac{1}{2}\int_0^1 {u_{R,N}}^2 + \int_0^1 u_{R,N} u_R = 1 - \frac{1}{2}\|u_{R,N} - u_R\|_{L^2_{\mathrm{per}}}^2.$$

Hence,

$$(E_{R,N} - E_R)\left(1 - \frac{1}{2}\|u_{R,N} - u_R\|_{L^2_{\mathrm{per}}}^2\right) = z_1 u_{R,N}(0)(\Pi_N^\perp u_R)(0) + z_2 u_{R,N}(R)(\Pi_N^\perp u_R)(R).$$

Using estimates (9) for $s = 0$ and (10), we obtain that for all $\epsilon > 0$,

$$1 - \frac{1}{2}\|u_{R,N} - u_R\|_{L^2_{\mathrm{per}}}^2 = 1 + o\left(\frac{1}{N^{3-\epsilon}}\right), \quad \text{when } N \to +\infty.$$

This concludes the proof of Lemma 2. □

The following lemma provides an explicit expression of the quantities $(\Pi_N^\perp u_R)(0)$ and $(\Pi_N^\perp u_R)(R)$ appearing in (13).

Lemma 3. *Let $z_1, z_2 > 0$. For all $R \in (0,1)$, all $N \in \mathbb{N}$, and all $x \in \mathbb{R}$,*

$$(\Pi_N^\perp u_R)(x) = \sum_{k=N+1}^{+\infty} \frac{2}{k_R^2 + 4\pi^2 k^2}\left(z_1 u_R(0)\cos(2\pi k x) + z_2 u_R(R)\cos(2\pi k(x-R))\right). \tag{14}$$

Proof. In order to estimate $(\Pi_N^\perp u_R)(x)$, we first need to compute the Fourier coefficients of u_R

$$\forall k \in \mathbb{Z}, \quad \widehat{u_R}(k) := \int_0^1 u_R(x) e^{-2i\pi k x}\, dx. \tag{15}$$

Using the periodicity of u_R, we can rewrite the first equation in (1) as

$$-u_R'' - z_1 u_R(0)\left(\sum_{m\in\mathbb{Z}} \delta_m\right) - z_2 u_R(R)\left(\sum_{m\in\mathbb{Z}} \delta_{m+R}\right) = E_R u_R.$$

Taking the Fourier transform, and using the relation $E_R = -k_R^2$, we obtain

$$4\pi^2 k^2 \widehat{u_R}(k) - z_1 u_R(0) - z_2 u_R(R) e^{-2i\pi k R} = -k_R^2 \widehat{u_R}(k).$$

Hence, for all $k \in \mathbb{Z}$,

$$\widehat{u_R}(k) = \frac{1}{k_R^2 + 4\pi^2 k^2}\left(z_1 u_R(0) + z_2 u_R(R) e^{-2i\pi k R}\right). \tag{16}$$

Consequently,

$$\begin{aligned}(\Pi_N^\perp u_R)(x) &= \sum_{k\in\mathbb{Z},\ |k|>N} \widehat{u_R}(k) e^{2i\pi k x} = \sum_{k\in\mathbb{Z},\ |k|>N} \frac{1}{k_R^2 + 4\pi^2 k^2}\left(z_1 u_R(0) + z_2 u_R(R) e^{-2i\pi k R}\right) e^{2i\pi k x} \\ &= \sum_{k=N+1}^{+\infty} \frac{2}{k_R^2 + 4\pi^2 k^2}\left(z_1 u_R(0)\cos(2\pi k x) + z_2 u_R(R)\cos(2\pi k(x-R))\right),\end{aligned}$$

which completes the proof of Lemma 3. □

The last technical lemma we need provides an estimates of the series in (14) for $x = 0$ and $x = R$.

Lemma 4. *Let $\mathbb{R} \ni R \mapsto k_R \in \mathbb{R}$ be a positive bounded function and $M = \sup_{R\in\mathbb{R}} k_R^2$. We denote by*

$$f_N(R) := \sum_{k=N+1}^{+\infty} \frac{1}{k_R^2 + 4\pi^2 k^2} \quad \text{and} \quad g_N(R) := \sum_{k=N+1}^{+\infty} \frac{\cos(2\pi kR)}{k_R^2 + 4\pi^2 k^2}.$$

For all $R \in \mathbb{R} \setminus \mathbb{Z}$ we have

$$f_N(R) = \frac{1}{4\pi^2 N} a_N + \phi_N(R), \quad \text{with } a_N = N \sum_{k=N+1}^{+\infty} \frac{1}{k^2}, \ |\phi_N(R)| \le \frac{M}{48\pi^4 N^3}, \tag{17}$$

and

$$g_N(R) = \frac{1}{4\pi^2 N} \eta_{N,R} + \psi_N(R), \quad \text{with } \eta_{N,R} = N \sum_{k=N+1}^{+\infty} \frac{\cos(2\pi kR)}{k^2}, \ |\psi_N(R)| \le \frac{M}{48\pi^4 N^3}. \tag{18}$$

Besides,

$$a_N = 1 + \frac{1}{2N} + O\left(\frac{1}{N^2}\right) \qquad \text{and} \qquad |\eta_{N,R}| \le \min\left(1, \frac{2 + \frac{\pi^3}{8}}{|\sin(\pi R)| N}\right). \tag{19}$$

Proof. The function f_N can be decomposed as

$$f_N(R) = \frac{1}{4\pi^2 N} a_N + \phi_N(R),$$

where

$$\phi_N(R) = f_N(R) - \frac{1}{4\pi^2 N} a_N = -\frac{k_R^2}{4\pi^2} \sum_{k=N+1}^{+\infty} \frac{1}{k^2 (k_R^2 + 4\pi^2 k^2)}.$$

We have on the one hand

$$a_N = 1 + N \sum_{k=N+1}^{+\infty} \left(\frac{1}{k^2} - \int_{k-1}^{k} \frac{dt}{t^2} \right) = 1 + N \sum_{k=N+1}^{+\infty} \frac{1}{k^2} \int_0^1 \left(1 - \left(1 - \frac{s}{k}\right)^{-2}\right) ds = 1 + \frac{1}{2N} + O\left(\frac{1}{N^2}\right),$$

and on the other hand, by a sum-integral comparison,

$$|\phi_N(R)| \le \frac{M}{4\pi^2} \sum_{k=N+1}^{+\infty} \frac{1}{4\pi^2 k^4} \le \frac{M}{48\pi^4 N^3}.$$

Thus, (17) and the first statement of (19) are proved. For $N \in \mathbb{N}$ and $R \in \mathbb{R}$, we set

$$h_N(R) := \sum_{k=N+1}^{+\infty} \frac{\cos(2\pi kR)}{4\pi^2 k^2} = \frac{1}{4\pi^2 N} \eta_{R,N}.$$

We have

$$|\psi_N(R)| = |g_N(R) - h_N(R)| = \left| - \sum_{k=N+1}^{+\infty} \frac{k_R^2 \cos(2\pi kR)}{4\pi^2 k^2 (k_R^2 + 4\pi^2 k^2)} \right| \le M \sum_{k=N+1}^{+\infty} \frac{1}{16\pi^4 k^4} \le \frac{M}{48\pi^4 N^3}.$$

Taking the second derivative of h_N in the distribution sense and using Poisson summation formula, we obtain

$$\begin{aligned} h_N''(R) &= \frac{d^2}{dR^2} \left(\sum_{k=N+1}^{+\infty} \frac{e^{2i\pi kR} + e^{-2i\pi kR}}{8\pi^2 k^2} \right) = -\frac{1}{2} \left(\sum_{k\in\mathbb{Z} \,|\, |k|>N} e^{2i\pi kR} \right) \\ &= -\frac{1}{2} \left(\sum_{k\in\mathbb{Z}} e^{2i\pi kR} - \sum_{k=-N}^{N} e^{2i\pi kR} \right) = -\frac{1}{2} \sum_{m\in\mathbb{Z}} \delta_m(R) + \frac{1}{2} \frac{\sin\left((2N+1)\pi R\right)}{\sin(\pi R)}. \end{aligned}$$

Therefore, h_N is smooth on $\mathbb{R} \setminus \mathbb{Z}$. Since it is 1-periodic, it suffices to study it on the open interval $(0,1)$. Since $h_N\left(\frac{1}{2}+t\right) = h_N\left(\frac{1}{2}-t\right)$ for all $|t| < \frac{1}{2}$, we have $h_N'\left(\frac{1}{2}\right) = 0$, so that for all $R \in (0,1)$, and using Taylor formula with integral remainder, we get

$$\begin{aligned} h_N(R) &= h_N\left(\frac{1}{2}\right) + \int_{\frac{1}{2}}^{R} (R-t)\, h_N''(t)\, dt = h_N\left(\frac{1}{2}\right) + \frac{1}{2}\int_{\frac{1}{2}}^{R} (R-t)\, \frac{\sin\left((2N+1)\pi t\right)}{\sin(\pi t)}\, dt \\ &= h_N\left(\frac{1}{2}\right) + \frac{1}{2(2N+1)^2\pi^2}\left((-1)^N - \frac{\sin\left((2N+1)\pi R\right)}{\sin(\pi R)}\right) \\ &\quad - \frac{1}{2(2N+1)^2\pi^2}\int_{\frac{1}{2}}^{R}\left(2\pi\frac{\cos(\pi t)}{\sin(\pi t)} + \frac{(R-t)\pi^2(1+\cos^2(\pi t))}{\sin^2(\pi t)}\right)\frac{\sin\left((2N+1)\pi t\right)}{\sin(\pi t)}\, dt. \end{aligned}$$

Since

$$\left|h_N\left(\frac{1}{2}\right)\right| = \left|\sum_{k=N+1}^{+\infty} \frac{(-1)^k}{4\pi^2 k^2}\right| \leq \frac{1}{4\pi^2(N+1)^2} \leq \frac{1}{4\pi^2 N^2},$$

and since, for all $R \in (0, 1/2)$,

$$\left|\frac{1}{2(2N+1)^2\pi^2}\left((-1)^N - \frac{\sin\left((2N+1)\pi R\right)}{\sin(\pi R)}\right)\right| \leq \frac{1}{8\pi^2 N^2}\left(1 + \frac{1}{\sin(\pi R)}\right) \leq \frac{1}{4\pi^2 N^2 \sin(\pi R)},$$

$$\left|\int_{\frac{1}{2}}^{R} 2\pi\frac{\cos(\pi t)}{\sin(\pi t)}\frac{\sin\left((2N+1)\pi t\right)}{\sin(\pi t)}\, dt\right| \leq 2\pi\int_{R}^{\frac{1}{2}} \frac{\cos(\pi t)}{\sin^2(\pi t)} dt = 2\left(\frac{1}{\sin(\pi R)} - 1\right),$$

and, using the inequalities $2t < \sin(\pi t) < \pi t$ for all $0 < t < \frac{1}{2}$,

$$\begin{aligned} \left|\int_{\frac{1}{2}}^{R} \frac{(R-t)\pi^2(1+\cos^2(\pi t))}{\sin^2(\pi t)}\frac{\sin\left((2N+1)\pi t\right)}{\sin(\pi t)}\, dt\right| &\leq 2\pi^2\int_{R}^{\frac{1}{2}} \frac{t-R}{\sin^3(\pi t)}\, dt \leq \pi^2\int_{R}^{\frac{1}{2}} \frac{2t}{\sin^3(\pi t)}\, dt \\ &\leq \frac{\pi^2}{4}\int_{R}^{\frac{1}{2}} \frac{1}{t^2} dt \leq \frac{\pi^2}{4R} \leq \frac{\pi^3}{4\sin(\pi R)}, \end{aligned}$$

we finally get

$$\begin{aligned} |\eta_{N,R}| = \left|4\pi^2 N h_N(R)\right| &\leq \frac{1}{N} + \frac{1}{N\sin(\pi R)} + \frac{1}{N}\left(\frac{1}{\sin(\pi R)} - 1\right) + \frac{\pi^3}{8\sin(\pi R)N} \\ &= \left(2 + \frac{\pi^3}{8}\right)\frac{1}{\sin(\pi R)N}, \end{aligned}$$

which concludes the proof. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. Combining Lemmata 1, 2, 3 and 4, we get that for any $R \in (0,1)$,

$$\begin{aligned}
E_{R,N} - E_R &= z_1 u_{R,N}(0)(\Pi_N^\perp u_R)(0) + z_2 u_{R,N}(R)(\Pi_N^\perp u_R)(R) + o\left(\frac{1}{N^{3-\epsilon}}\right) && \text{(Lemma 2)} \\
&= z_1 u_{R,N}(0)\left(2z_1 u_R(0) f_N(R) + 2z_2 u_R(R) g_N(R)\right) \\
&\quad + z_2 u_{R,N}(R)\left(2z_2 u_R(R) f_N(R) + 2z_1 u_R(0) g_N(R)\right) + o\left(\frac{1}{N^{3-\epsilon}}\right) && \text{(Lemma 3)} \\
&= \left(2z_1^2 u_{R,N}(0)u_R(0) + 2z_2^2 u_{R,N}(R)u_R(R)\right) f_N(R) \\
&\quad + 2z_1 z_2 \left(u_{R,N}(0)u_R(R) + u_{R,N}(R)u_R(0)\right) g_N(R) + o\left(\frac{1}{N^{3-\epsilon}}\right) \\
&= \left(2z_1^2 u_{R,N}(0)u_R(0) + 2z_2^2 u_{R,N}(R)u_R(R)\right) \frac{1}{4\pi^2 N} a_N \\
&\quad + 2z_1 z_2 \left(u_{R,N}(0)u_R(R) + u_{R,N}(R)u_R(0)\right) \frac{1}{4\pi^2 N} \eta_{R,N} + o\left(\frac{1}{N^{3-\epsilon}}\right) && \text{(Lemma 4)} \\
&= \frac{\alpha_R}{N} a_N + \frac{\beta_{R,N}^{(1)}}{N} a_N + \frac{\gamma_R}{N^2} \eta_{R,N} + o\left(\frac{1}{N^{3-\epsilon}}\right),
\end{aligned}$$

where we have used the bounds (11) and (19) to obtain the last equality. The proof of (7) easily follows. □

References

[1] X. Blanc, E. Cancès, and M.-S. Dupuy. Variational projector augmented-wave method. *Comptes Rendus Mathématique*, pages –, 2017.

[2] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50:17953–17979, Dec 1994.

[3] 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(2):341–388, 2012.

[4] E. Cancès, G. Dusson, Y. Maday, B. Stamm, and M. Vohralík. A perturbation-method-based a posteriori estimator for the planewave discretization of nonlinear Schrödinger equations. *Comptes Rendus Mathematique*, 352(11):941–946, 2014.

[5] 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. *Journal of Computational Physics*, 307:446–459, 2016.

[6] E. Cancès, V. Ehrlacher, and Y. Maday. Non-consistent approximations of self-adjoint eigenproblems: application to the supercell method. *Numerische Mathematik*, 128(4):663–706, 2014.

[7] H. Chen, X. Dai, X. Gong, L. He, and A. Zhou. Adaptive finite element approximations for Kohn–Sham models. *Multiscale Modeling & Simulation*, 12(4):1828–1869, 2014.

[8] H. Chen, X. Gong, L. He, Z. Yang, and A. Zhou. Numerical analysis of finite dimensional approximations of Kohn–Sham models. *Advances in Computational Mathematics*, 38(2):225–256, 2013.

[9] H. Chen and R. Schneider. Error estimates of some numerical atomic orbitals in molecular simulations. *Communications in Computational Physics*, 18(01):125–146, 2015.

[10] H. Chen and R. Schneider. Numerical analysis of augmented plane wave methods for full-potential electronic structure calculations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(3):755–785, 2015.

[11] G. Dusson and Y. Maday. A posteriori analysis of a nonlinear Gross–Pitaevskii-type eigenvalue problem. *IMA Journal of Numerical Analysis*, page drw001, 2016.

[12] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, 2009.

[13] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space Gaussian pseudopotentials. *Physical Review B*, 54(3):1703–1710, 1996.

[14] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, Ph. Ghosez, M. Giantomassi, S. Goedecker, D.R. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M.J.T. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M.J. Verstraete, G. Zerah, and J.W. Zwanziger. ABINIT: First-principles approach to material and nanosystem properties. *Computer Physics Communications*, 180(12):2582–2615, 2009.

[15] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph Ghosez, J.-Y. Raty, and D.C. Allan. First-principles computation of material properties: the ABINIT software project. *Computational Materials Science*, 25(3):478–492, 2002.

[16] F. Gygi and G. Galli. Real-space adaptive-coordinate electronic-structure calculations. *Physical Review B*, 52(4):R2229–R2232, 1995.

[17] M. Hanrath. Wavefunction quality and error estimation of single- and multi-reference coupled-cluster and CI methods: the H4 model system. *Chemical Physics Letters*, 466(4-6):240–246, 2008.

[18] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, Ltd, Chichester, UK, 2000.

[19] J. Kaye, L. Lin, and C. Yang. A posteriori error estimator for adaptive local basis functions to solve Kohn–Sham density functional theory. *Communications in Mathematical Sciences*, 13(7):1741–1773, 2015.

[20] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965.

[21] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169–11186, 1996.

[22] W. Kutzelnigg. Error analysis and improvements of coupled-cluster theory. *Theoretica Chimica Acta*, 80(4-5):349–386, 1991.

[23] W. Kutzelnigg. Rate of convergence of basis expansions in quantum chemistry. *AIP Conference Proceedings*, 1504(1):15–30, 2012.

[24] S. Li, K. Chen, M.-Y. Hsieh, N. Muralimanohar, C. D. Kersey, J. B. Brockman, A. F. Rodrigues, and N. P. Jouppi. System implications of memory reliability in exascale computing. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis on - SC '11*, page 1, New York, New York, USA, 2011. ACM Press.

[25] L. Lin and B. Stamm. A posteriori error estimates for discontinuous Galerkin methods using non-polynomial basis functions Part I: Second order linear PDE. *ESAIM: Mathematical Modelling and Numerical Analysis*, 50(4):1193–1222, 2016.

[26] Y. Maday and G. Turinici. Error bars and quadratically convergent methods for the numerical simulation of the Hartree-Fock equations. *Numerische Mathematik*, 94(4):739–770, 2003.

[27] S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, and S. Goedecker. Daubechies wavelets for linear scaling density functional theory. *The Journal of Chemical Physics*, 140(20):204110, 2014.

[28] 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*, 253:308–343, 2013.

[29] J. E. Pask and P. A. Sterne. Finite element methods in ab initio electronic structure calculations. *Modelling and Simulation in Materials Science and Engineering*, 13(3):R71–R96, 2005.

[30] P. Pernot, B. Civalleri, D. Presti, and A. Savin. Prediction uncertainty of density functional approximations for properties of crystals with cubic symmetry. *The Journal of Physical Chemistry A*, 119(21):5288–5304, 2015.

[31] S. N. Pieniazek, F. R. Clemente, and K. N. Houk. Sources of error in DFT computations of C–C bond formation thermochemistries: $\pi \rightarrow \sigma$ transformations and error cancellation by DFT methods. *Angewandte Chemie International Edition*, 47(40):7746–7749, 2008.

[32] M. Reed and B. Simon. *Methods of modern mathematical physics. I. Functional analysis*, volume 53. Academic Press Inc., New York, 1972.

[33] T. Rohwedder and R. Schneider. Error estimates for the coupled cluster method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(6):1553–1582, 2013.

[34] Y. Saad, J. R. Chelikowsky, and S. M. Shontz. Numerical methods for electronic structure calculations of materials. *SIAM Review*, 52(1):3–54, 2010.