Variational projector augmented-wave method: theoretical analysis and preliminary numerical results

X Blanc, Eric Cancès, Mi-Song Dupuy

Variational projector augmented-wave method: theoretical analysis and preliminary numerical results

X. Blanc[1], E. Cancès[2], and M.-S. Dupuy[1]

[1] *Univ. Paris Diderot, Sorbonne Paris Cité, Laboratoire Jacques-Louis Lions, UMR 7598, UPMC, CNRS, F-75205 Paris, France*

[2] *CERMICS - Ecole des Ponts ParisTech, 6 & 8 avenue Blaise Pascal, Cité Descartes, 77455 Marne la Vallée Cedex 2, France*

November 17, 2017

Abstract

In Kohn-Sham electronic structure computations, wave functions have singularities at nuclear positions. Because of these singularities, plane-wave expansions give a poor approximation of the eigenfunctions. In conjunction with the use of pseudo-potentials, the PAW (projector augmented-wave) method circumvents this issue by replacing the original eigenvalue problem by a new one with the same eigenvalues but smoother eigenvectors. Here a slightly different method, called VPAW (variational PAW), is proposed and analyzed. This new method allows for a better convergence with respect to the number of plane-waves. Some numerical tests on an idealized case corroborate this efficiency. This work has been recently announced in [3].

Introduction

Solving the N-body electronic problem is numerically impossible even for small molecular systems. Various nonlinear one-body approximations of this problem have been proposed, among which Hartree-Fock theory, post-Hartree-Fock methods and density functional theory, which describe fairly well the ground-state electronic structure of many molecules. However it can still be expensive to compute the desired properties with these approximations. In solid-state physics, plane-wave methods are often the method of choice to compute the lowest eigenvalues of Kohn-Sham operators. However Coulomb potentials considerably impedes the rate of convergence of plane-wave expansion because of the cusps [13, 11, 9] located at each nucleus. Over the years, several strategies have been developped to tackle this problem.

In most situations, the knowledge of the whole energy spectrum of the Kohn-Sham Hamiltonian is not needed. Only eigenvalues belonging to a small energy range are. Indeed, it is well-known that the chemical properties mostly come from the valence electrons. So it would be satisfactory to replace the Coulomb potential and nonlinear interactions with the core electrons by a smooth potential that reproduces the exact spectrum in the relevant range. This is the main idea behind pseudopotential methods. Specifically, pseudopotentials are designed to match the eigenvalues of the original atomic model in a fixed energy range. So when used in molecular or solid-state simulations, it seems reasonable to hope that they will accurately approximate the sought eigenvalues. Pseudopotentials can also be introduced to take into account relativistic effects at lower cost [8]. Thus a wide range of pseudopotentials with different properties have been developped, among which, Troullier-Martins [23] and Kleinman-Bylander [14] norm conserving pseudopotentials, Vanderbilt [24] ultrasoft pseudopotentials and Goedecker [10] pseudopotentials. A mathematical study of the generation of optimal norm-conserving pseudopotentials for the reduced Hartree-Fock theory and the local density approximation to the density-functional

theory has already been achieved [6]. So far as we know, it is the first mathematical result on pseudopotentials.

Another strategy is to use a better suited basis set. This is the spirit of the augmented plane-wave (APW) method [21, 15]. The APW basis functions are discontinuous: inside balls centered at each nucleus, a basis function reproduces the cusp of atomic electronic wave functions and outside these balls, it is a plane-wave. One thus tries to get the best of both worlds: having the singularity behavior encoded in the basis functions and at the same time, having the plane-wave convergence property outside the singularity region. This method can be viewed as a discontinuous Galerkin method and its mathematical analysis has been carried out in [7].

The projector augmented-wave (PAW) [4] method relies on the same idea as the APW method but here, instead of using another set of basis functions, the eigenvalue problem is modified by an invertible transformation which carries the cusp behavior and/or fast oscillations in the vicinities of the nuclei. Specifically, the operator acts locally in a ball around each nucleus and maps atomic wave functions to smooth functions called pseudo wave functions. The form of the transformation is compatible with slowly varying pseudopotentials in the Hamiltonian. Doing so, one replaces the Coulomb potential by a smooth potential without changing the spectrum of the original operator. The solution of the corresponding *generalized* eigenvalue problem can then be advantageously expanded in plane-waves. Because of the one-to-one correspondance between the pseudo and the actual wave functions and its efficiency to produce accurate results, the PAW method has become a very popular tool and has been implemented in several popular electronic structure simulation codes (AbInit [22], VASP [16, 17]).

A crucial assumption in the PAW method is the completeness of the basis of atomic wave functions used to build the PAW transformation. In practice, infinite expansions appearing in the generalized eigenvalue problem are truncated, introducing an error which is rarely quantified. In the variational PAW method (which will be referred to as VPAW in the following), a finite number of functions is used right from the beginning, avoiding this truncation error. Although pseudopotentials can no longer be incorporated, an acceleration of convergence is obtained. This acceleration can be precisely characterized in the case of the double Dirac potential in a one-dimensional periodic setting.

Before moving to the 1D model we analyzed, we will briefly introduce both PAW and VPAW methods applied to a 3D Hamiltonian and explain the major differences between both approaches. More detailed expositions of the PAW method in various settings can be found in [1, 12, 20].

1 PAW vs VPAW methods

1.1 General setting

The general setting of the VPAW method for finite molecular systems has been presented in [3]. In this paper, we will focus on the derivation of the VPAW equations in the periodic setting. For simplicity, we restrict ourselves to a linear model.

The crystal is modelled as an infinite periodic motif of N_{at} point charges at positions $\mathbf{R}_I$ in the unit cell

$$\Gamma = \left\{ \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3, (\alpha_1, \alpha_2, \alpha_3) \in [-1/2, 1/2)^3 \right\},$$

and repeated over the periodic lattice

$$\mathcal{R} = \mathbb{Z}\mathbf{a}_1 + \mathbb{Z}\mathbf{a}_2 + \mathbb{Z}\mathbf{a}_3,$$

where $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are linearly independent vectors of $\mathbb{R}^3$.

In the linear model under consideration, the electronic properties of the crystal are encoded in the spectral properties of the periodic Hamiltonian $H_{\rm per}$ acting on $L^2(\mathbb{R}^3)$:

$$H_{\rm per} = -\frac{1}{2}\Delta + V_{\rm per} + W_{\rm per},$$

where $V_{\rm per}$ is the $\mathcal{R}$-periodic potential defined (up to an irrelevant addition constant) by

$$\begin{cases} -\Delta V_{\rm per} = 4\pi \left(\displaystyle\sum_{\mathbf{T}\in\mathcal{R}} \sum_{I=1}^{N_{at}} Z_I \left(\delta_{\mathbf{R}_I}(\cdot + \mathbf{T}) - \frac{1}{|\Gamma|} \right) \right) \\ V_{\rm per} \text{ is } \mathcal{R}\text{-periodic.} \end{cases} \tag{1.1}$$

For simplicity, $W_{\rm per}$ is a regular enough $\mathcal{R}$-periodic potential. In practice, $W_{\rm per}$ is a nonlinear potential depending on the model chosen to describe the Hartree and exchange-correlation terms (typically a Kohn-Sham LDA potential).

The standard way to study the spectral properties of $H_{\rm per}$ is through Bloch theory which will be outlined in the next few lines. Let $\mathcal{R}^*$ be the dual lattice

$$\mathcal{R}^* = \mathbb{Z}\mathbf{a}_1^* + \mathbb{Z}\mathbf{a}_2^* + \mathbb{Z}\mathbf{a}_3^*,$$

where $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ satisfies $\mathbf{a}_i \cdot \mathbf{a}_j^* = 2\pi\delta_{ij}$. The reciprocal unit cell is defined by

$$\Gamma^* = \left\{ \alpha_1\mathbf{a}_1^* + \alpha_2\mathbf{a}_2^* + \alpha_3\mathbf{a}_3^*, (\alpha_1,\alpha_2,\alpha_3) \in [-1/2, 1/2)^3 \right\}.$$

As $H_{\rm per}$ commutes with $\mathcal{R}$-translations, $H_{\rm per}$ admits a Bloch decomposition in operators $H_{\mathbf{q}}$ acting on

$$L^2_{\rm per}(\Gamma) = \{ f \in L^2_{\rm loc}(\mathbb{R}^3) \mid f \text{ is } \mathcal{R}\text{-periodic} \},$$

with domain

$$H^2_{\rm per}(\Gamma) = \{ f \in H^2_{\rm loc}(\mathbb{R}^3) \mid f \text{ is } \mathcal{R}\text{-periodic} \}.$$

The operator $H_{\mathbf{q}}$ is given by:

$$H_{\mathbf{q}} = \frac{1}{2}|-i\nabla + \mathbf{q}|^2 + V_{\rm per} + W_{\rm per}, \quad \mathbf{q} \in \Gamma^*.$$

For each $\mathbf{q} \in \Gamma^*$, the operator $H_{\mathbf{q}}$ is self-adjoint, bounded below and with compact resolvent. It thus has a discrete spectrum. Denoting by $E_{1,\mathbf{q}} \leq E_{2,\mathbf{q}} \leq \ldots$, with $E_{n,\mathbf{q}} \underset{n\to+\infty}{\longrightarrow} +\infty$, its eigenvalues counted with multiplicities, there exists an orthonormal basis of $L^2_{\rm per}(\Gamma)$ consisting of eigenfunctions $(\psi_{n,\mathbf{q}})_{n\in\mathbb{N}^*}$

$$H_{\mathbf{q}}\psi_{n,\mathbf{q}} = E_{n,\mathbf{q}}\psi_{n,\mathbf{q}}. \tag{1.2}$$

The spectrum of $H_{\rm per}$ is purely continuous and can be recovered from the discrete spectra of all the operators $H_{\mathbf{q}}$, $\mathbf{q} \in \Gamma^*$

$$\sigma(H_{\rm per}) = \bigcup_{\mathbf{q}\in\Gamma^*} \sigma(H_{\mathbf{q}}).$$

The PAW and VPAW methods aim to ease the numerical approximation of the eigenvalue problem (1.2). For clarity, we will only present the case $\mathbf{q} = 0$ and denote H_0 by H as this special case encloses all the main difficulties encountered in numerically solving (1.2). Transposition to $\mathbf{q} \neq 0$ can be done without problem.

1.2 The VPAW method for solids

Following the idea of the PAW method, an invertible transformation $(\mathrm{Id}+T)$ is applied to the eigenvalue problem (1.2), where T is the sum of operators T_I, each T_I acting locally around nucleus I. For each operator T_I, two parameters N_I and r_c need to be fixed:

1. the number N_I of PAW functions used to build T_I,
2. a cut-off r_c radius which will set the acting domain of T_I, more precisely:
   - for all $f \in L^2_{\mathrm{per}}(\Gamma)$, $\mathrm{supp}(T_I f) \subset \bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{R}_I+\mathbf{T}, r_c)$, where $\overline{B}(\mathbf{R}, r)$ is the closed ball of $\mathbb{R}^3$ with center $\mathbf{R}$ and radius r,
   - if $\mathrm{supp}(f) \bigcap \bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{R}_I+\mathbf{T}, r_c) = \emptyset$, then $T_I f = 0$.

The cut-off radius r_c must be chosen small enough to avoid pairwise overlaps of the balls $(\overline{B}(\mathbf{R}_I+\mathbf{T}, r_c))_{1\le I\le N, \mathbf{T}\in\mathcal{R}}$.

The operator T_I is given by:

$$T_I = \sum_{k=1}^{N_I} (\phi_k^I(\mathbf{r}-\mathbf{R}_I) - \widetilde{\phi}_k^I(\mathbf{r}-\mathbf{R}_I)) \langle \tilde{p}_k^I(\cdot - \mathbf{R}_I)\,, \bullet \rangle, \tag{1.3}$$

where $\langle \bullet\,, \bullet \rangle$ is the usual L^2_{per}-scalar product and the functions ϕ_k^I, $\widetilde{\phi}_k^I$ and $\tilde{p}_k^I$ are functions in $L^2_{\mathrm{per}}(\Gamma)$.

These functions, which will be referred to as the PAW functions in the sequel, are chosen as follows:

1. first, let $(\varphi_k^I)_{1\le k\le N_I} \in (L^2(\mathbb{R}^3))^{N_I}$ be eigenfunctions of an atomic *non-periodic* Hamiltonian
   $$H_I \varphi_k^I = \epsilon_k \varphi_k^I, \quad \epsilon_1^I \le \epsilon_2^I \le \epsilon_3^I \le \ldots, \quad \int_{\mathbb{R}^3} \varphi_k^I \varphi_{k'}^I = \delta_{kk'},$$
   with H_I defined by
   $$H_I = -\frac{1}{2}\Delta - \frac{Z_I}{|\mathbf{r}|} + W(|\mathbf{r}|),$$
   where W is a regular enough bounded potential. The operator H_I is self-adjoint on $L^2(\mathbb{R}^3)$ with domain $H^2(\mathbb{R}^3)$. Again, in practice, W is a radial nonlinear potential belonging to the same family of models as W_{per} in Equation (1.1). The PAW atomic wave functions $(\phi_k^I)_{1\le k\le N_I} \in (L^2_{\mathrm{per}}(\Gamma))^{N_I}$ satisfy:
   - for $1 \le k \le N_I$ and $\mathbf{r} \in \Gamma$, $\phi_k^I(\mathbf{r}) = \varphi_k^I(\mathbf{r})$,
   - ϕ_k^I is $\mathcal{R}$-periodic;
2. the pseudo wave functions $(\widetilde{\phi}_k^I)_{1\le k\le N_I}$, with $N_I \le Z_I$, are determined by the next conditions:
   (a) inside the unit cell Γ, $\widetilde{\phi}_k^I$ is smooth and matches ϕ_k^I and several of its derivatives on the sphere $\{|\mathbf{r}| = r_c\}$,
   (b) for $\mathbf{r} \in \mathbb{R}^3 \setminus \bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{T}, r_c)$, $\widetilde{\phi}_k^I(\mathbf{r}) = \phi_k^I(\mathbf{r})$;
3. the projector functions $(\tilde{p}_k^I)_{1\le k\le N_I}$ are defined such that:
   (a) each projector function $\tilde{p}_k^I$ is supported in $\bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{T}, r_c)$,
   (b) they form a dual family to the pseudo wave functions $(\widetilde{\phi}_k^I)_{1\le k\le N_I}$: $\langle \tilde{p}_k^I\,, \widetilde{\phi}_{k'}^I \rangle = \delta_{kk'}$.

By our choice of the pseudo wave functions $\widetilde{\phi}_k^I$ and the projectors $\tilde{p}_k^I$, T_I acts in $\bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{R}_I + \mathbf{T}, r_c)$.

The VPAW equations to solve are then:

$$\widetilde{H}\tilde{\psi} = E\widetilde{S}\tilde{\psi}, \tag{1.4}$$

where

$$\widetilde{H} = (\mathrm{Id} + T)^* H (\mathrm{Id} + T), \quad \widetilde{S} = (\mathrm{Id} + T)^* (\mathrm{Id} + T), \tag{1.5}$$

and

$$T = \sum_{I=1}^{N_{at}} T_I.$$

Thus if $(\mathrm{Id} + T)$ is invertible, it is easy to recover the eigenfunctions of H by the formula

$$\psi = (\mathrm{Id} + T)\tilde{\psi}, \tag{1.6}$$

and the eigenvalues of H coincide with the generalized eigenvalues of (1.4).

By construction, the operator $(\mathrm{Id} + T_I)$ maps the pseudo wave functions $\tilde{\phi}$ to the atomic eigenfunctions ϕ:

$$(\mathrm{Id} + T_I)\tilde{\phi}_k^I = \phi_k^I,$$

so if locally around each nucleus, the function ψ "behaves" like the atomic wave functions ϕ_k, we can expect that the cusp behavior of ψ is captured by the operator T, thus $\tilde{\psi}$ is smoother than ψ and the plane-wave expansion of $\tilde{\psi}$ converges faster than the expansion of ψ.

1.3 Differences with the PAW method

The PAW equations solved in materials science simulation packages are different from the VPAW equations (1.4). As in [4], the construction of T_I involves "complete" infinite sets of functions ϕ_k^I, $\widetilde{\phi}_k^I$ and $\tilde{p}_k^I$ in the sense that for a function $f \in L^2_{\mathrm{per}}(\Gamma)$ supported in the balls $\bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{R}_I + \mathbf{T}, r_c)$, we have :

$$\sum_{k=1}^{\infty} \langle \tilde{p}_k^I \,, f \rangle \widetilde{\phi}_k^I(x) = f(x), \quad \text{for a.a. } x \in \bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{R}_I + \mathbf{T}, r_c). \tag{1.7}$$

This relation enables one to simplify the expression of $(\mathrm{Id} + T^*)H(\mathrm{Id} + T)$ and $(\mathrm{Id} + T^*)(\mathrm{Id} + T)$ to

$$H^{PAW} = (\mathrm{Id} + T^*)H(\mathrm{Id} + T) = H + \sum_{I=1}^{N_{at}} \sum_{i,j=1}^{\infty} \tilde{p}_i^I(\cdot - \mathbf{R}_I) \Big(\langle \phi_i^I(\cdot - \mathbf{R}_I) \,, H\phi_j^I(\cdot - \mathbf{R}_I) \rangle - \langle \widetilde{\phi}_i^I(\cdot - \mathbf{R}_I) \,, H\widetilde{\phi}_j^I(\cdot - \mathbf{R}_I) \rangle \Big) \langle \tilde{p}_j^I(\cdot - \mathbf{R}_I) \,, \bullet \rangle, \tag{1.8}$$

and

$$S^{PAW} = (\mathrm{Id} + T^*)(\mathrm{Id} + T) = \mathrm{Id} + \sum_{I=1}^{N_{at}} \sum_{i,j=1}^{\infty} \tilde{p}_i^I(\cdot - \mathbf{R}_I) \Big(\langle \phi_i^I(\cdot - \mathbf{R}_I) \,, \phi_j^I(\cdot - \mathbf{R}_I) \rangle - \langle \widetilde{\phi}_i^I(\cdot - \mathbf{R}_I) \,, \widetilde{\phi}_j^I(\cdot - \mathbf{R}_I) \rangle \Big) \langle \tilde{p}_j^I(\cdot - \mathbf{R}_I) \,, \bullet \rangle. \tag{1.9}$$

In practice, the double sums on i, j appearing in the operators H^{PAW} and S^{PAW} are then truncated and the so-obtained generalized eigenvalue problem is solved. Thus the identity $\psi = (\mathrm{Id} + T)\tilde{\psi}$ does not hold anymore and the eigenvalues of the truncated problem are different

from the exact ones. An analysis of the eigenvalue problem $H^{PAW} f = E S^{PAW} f$ in our 1D-toy model will be provided in another paper [2].

In contrast, the VPAW approach makes use of a finite number of wave functions ϕ_i^I right from the beginning, avoiding truncation approximations.

A further modification is used in practice. As the pseudo wave functions $\widetilde{\phi}_k^I$ are equal to ϕ_k^I outside the balls $\bigcup_{\mathbf{T}\in\mathcal{R}} \overline{B}(\mathbf{R}_I + \mathbf{T}, r_c)$, the integrals appearing in (1.8) can be truncated to the ball $\overline{B}(\mathbf{R}_I, r_c)$. Doing so, another expression of H^{PAW} can be obtained :

$$H^{PAW} = (\mathrm{Id} + T^*)H(\mathrm{Id} + T) = H + \sum_{I=1}^{N_{at}} \sum_{i,j=1}^{\infty} \tilde{p}_i^I(\cdot - \mathbf{R}_I) \\ \left(\langle \phi_i^I(\cdot - \mathbf{R}_I), H\phi_j^I(\cdot - \mathbf{R}_I)\rangle_{\mathbf{R}_I} - \langle \widetilde{\phi}_i^I(\cdot - \mathbf{R}_I), H\widetilde{\phi}_j^I(\cdot - \mathbf{R}_I)\rangle_{\mathbf{R}_I} \right) \langle \tilde{p}_j^I(\cdot - \mathbf{R}_I), \bullet \rangle,$$

where

$$\langle f, g\rangle_{\mathbf{R}_I} = \int_{B(\mathbf{R}_I, r_c)} f(x)g(x)\,\mathrm{d}x.$$

Using this expression of the operator H^{PAW}, it is possible to introduce an $\mathcal{R}$-periodic potential V^{PP} such that :

1. $V^{\mathrm{PP}} = V_{\mathrm{per}}$ outside $\bigcup_{\mathbf{T}\in\mathcal{R}} \bigcup_{I=1}^{N_{at}} \overline{B}(\mathbf{R}_I + \mathbf{T}, r_c)$,
2. V^{PP} is smooth inside $\bigcup_{\mathbf{T}\in\mathcal{R}} \bigcup_{I=1}^{N_{at}} \overline{B}(\mathbf{R}_I + \mathbf{T}, r_c)$.

The expression of H^{PAW} is

$$H^{PAW} = H_{\mathrm{ps}} + \sum_{I=1}^{N_{at}} \sum_{i,j=1}^{\infty} \tilde{p}_i^I(\cdot - \mathbf{R}_I) \\ \left(\langle \phi_i^I(\cdot - \mathbf{R}_I), H\phi_j^I(\cdot - \mathbf{R}_I)\rangle_{\mathbf{R}_I} - \langle \widetilde{\phi}_i^I(\cdot - \mathbf{R}_I), H_{\mathrm{ps}}\widetilde{\phi}_j^I(\cdot - \mathbf{R}_I)\rangle_{\mathbf{R}_I} \right) \langle \tilde{p}_j^I(\cdot - \mathbf{R}_I), \bullet \rangle, \quad (1.10)$$

with H_{ps}:

$$H_{\mathrm{ps}} = -\frac{1}{2}\Delta + V^{\mathrm{PP}},$$

where V^{PP} is a smooth pseudopotential. Note that, in practice, the sum of i, j in (1.10) is truncated to some level N_I.

A study of the associated eigenvalue problem will also be provided in the aforementioned paper [2]. Some numerical tests comparing the VPAW and PAW methods in our one-dimensional toy model can be found in Section 6.

1.4 Computational complexity

A detailed analysis of the computational cost of the PAW method can be found in [19]: the cost scales like $\mathcal{O}(NM + M\log M)$ where $N = \sum_I N_I$ is the total number of projectors and M the number of plane-waves. Usually, N is chosen relatively small, but M may be large, so it is important to avoid a computational cost of order M^2.

In practice, we are interested in the cost of the computation of $\widetilde{H}\tilde{\psi}$ and $\widetilde{S}\tilde{\psi}$ where $\tilde{\psi}$ is expanded in M plane-waves as the generalized eigenvalue problem is solved by a conjugate gradient algorithm. We will only focus on $\widetilde{H}\tilde{\psi}$ since the analysis $\widetilde{S}\tilde{\psi}$ is similar. Let us split $\widetilde{H}$ into four terms:

$$\widetilde{H}\tilde{\psi} = H\tilde{\psi} + PD_H P^T\tilde{\psi} + H\left(\Phi - \widetilde{\Phi}\right)P^T\tilde{\psi} + PH\left(\Phi - \widetilde{\Phi}\right)^T\tilde{\psi},$$

where P is the $M \times N$ matrix of the projector functions, $H(\Phi - \widetilde{\Phi})$ the $M \times N$ matrix of the Fourier representation of the N functions $H(\phi_i - \widetilde{\phi}_i)$, and D_H is the $N \times N$ matrix $\langle \phi_i - \widetilde{\phi}_i \, , H(\phi_j - \widetilde{\phi}_j) \rangle$.

The computational cost can be estimated as follows (the cost at each step is given in brackets):

1. $H\tilde{\psi}$ is assembled in two steps. First, $-\frac{1}{2}\Delta\tilde{\psi}$ is computed in $\mathcal{O}(M)$ since the operator $\frac{1}{2}\Delta$ is diagonal in Fourier representation. For the potential V, apply an inverse FFT to $\tilde{\psi}$ to have the real space representation of $\tilde{\psi}$, multiply pointwise by V and apply a FFT to the whole result ($\mathcal{O}(M \log M)$);
2. for $PD_HP^T\tilde{\psi}$, compute the N projections $P^T\tilde{\psi}$ ($\mathcal{O}(MN)$), then successively apply the matrices D_H ($\mathcal{O}(N^2)$) and P ($\mathcal{O}(MN)$);
3. for $PH(\Phi - \widetilde{\Phi})^T\tilde{\psi}$, similarly apply successively $H(\Phi - \widetilde{\Phi})^T$ to $\tilde{\psi}$ ($\mathcal{O}(MN)$) and P to $H(\Phi - \widetilde{\Phi})^T\tilde{\psi}$ ($\mathcal{O}(MN)$);
4. for $H(\Phi - \widetilde{\Phi})P^T\tilde{\psi}$, we proceed as in step 3.

Thus, the total numerical cost is of order $\mathcal{O}(MN + M \log M)$ which is the same as for the PAW method.

The matrix $H(\Phi - \widetilde{\Phi})$ is approximated by a plane-wave expansion, which may be a poor approximation because of the singularities of Φ. However, it should be noticed that this is only an intermediary in the computation of $\tilde{\psi}$, which is well approximated by plane-waves. Hence it is not clear that a poor approximation of $H(\Phi - \widetilde{\Phi})$ should imply a poor approximation of $\tilde{\psi}$.

1.5 Generation of the pseudo wave functions

In practice, there are two main ways to generate the pseudo wave functions $\tilde{\phi}_k$ and the projectors $\tilde{p}_k$ introduced by Blöchl [4] and Vanderbilt [18]. For both schemes, the generation of the PAW functions has to be done for *each* angular momentum ℓ.

1.5.1 Vanderbilt scheme

Atomic wave function The functions φ_k are simply the atomic wavefunctions defined earlier *i.e.* solutions to the atomic eigenvalue problem

$$H_I\varphi_k^I = \epsilon_k\varphi_k^I, \quad \epsilon_1^I \leq \epsilon_2^I \leq \epsilon_3^I \leq \ldots, \quad \int_{\mathbb{R}^3} \varphi_j^I\varphi_k^I = \delta_{jk},$$

with

$$H_I = -\frac{1}{2}\Delta - \frac{Z_I}{|\mathbf{r}|} + W(|\mathbf{r}|).$$

The eigenfunctions φ_k can be decomposed into a spherical part $Y_{\ell m}$ -the real Laplace spherical harmonics- and a radial part $R_{n\ell}$

$$\varphi_k(\mathbf{r}) = \frac{R_{n\ell}(r)}{r} Y_{\ell m}(\omega),$$

where k stands for the multiple indices (n, ℓ, m) and $\mathbf{r} = (r, \omega)$ is written in polar coordinates.

Pseudo wave function The pseudo wave functions $\widetilde{\phi}_k$ are given by :

$$\forall \mathbf{r} \in \Gamma, \ \widetilde{\phi}_k(\mathbf{r}) = \frac{\widetilde{R}_{n\ell}(r)}{r} \, Y_{\ell m}(\omega).$$

Various choices of $\widetilde{R}_{n\ell}$ are possible, for example, in [18], $\widetilde{R}_{n\ell}$ is a polynomial inside the augmentation region $\overline{B}(0, r_c)$:

$$\widetilde{R}_{n\ell}(r) = \begin{cases} r^{\ell+1} \sum_{k=0}^{p} c_{2k} r^{2k} & \text{for } 0 \leq r \leq r_c, \\ R_{n\ell}(r) & \text{for } r > r_c, \end{cases}$$

or in [17], a sum of spherical Bessel functions j_ℓ :

$$\widetilde{R}_{n\ell}(r) = \begin{cases} r\sum_{k=1}^{p} \alpha_k j_\ell(q_k r) \ , & \text{for } 0 \le r \le r_c, \\ R_{n\ell}(r) & \text{for } r > r_c, \end{cases}$$

and the coefficients are chosen to match as many derivatives of $R_{n\ell}$ as possible at r_c.

Projector function First, define :

$$\chi_{n\ell}(r) = \frac{1}{2}\widetilde{R}''_{n\ell}(r) + \left(E_n - \frac{\ell(\ell+1)}{2r^2} - V_\ell^{\mathrm{PP}}(r)\right)\widetilde{R}_{n\ell}(r),$$

where $V_\ell^{\mathrm{PP}}(r)$ is usually the Troullier-Martins pseudopotential [23] although other choices are possible.

By construction, $\mathrm{supp}(\chi_{n\ell}) \subset [0, r_c]$. Let B be the matrix

$$B_{n,n'} = \int_0^{r_c} \widetilde{R}_{n\ell}(r)\chi_{n'\ell}(r)\,\mathrm{d}r.$$

The radial parts of the projector functions are given by

$$p_{n\ell}(r) = \sum_{n'=1}^{N_I} \chi_{n'\ell}(r)\left(B^{-1}\right)_{n'n}.$$

The projector functions are defined by

$$\widetilde{p}_{n\ell}(\mathbf{r}) = \frac{p_{n\ell}(r)}{r} Y_{\ell m}(\omega).$$

This ensures that $\langle \widetilde{p}_{n\ell}\,, \widetilde{\phi}_{n'\ell'}\rangle = \delta_{nn'}\delta_{\ell\ell'}$.

1.5.2 Blöchl scheme

The PAW functions are generated in two steps. For each angular momentum ℓ, we define auxiliary functions $\widetilde{R}^0_{n\ell}$ and $p^0_{n\ell}$:

Auxiliary functions Let $\chi(r)$ be the cut-off function

$$\chi(r) = \begin{cases} \frac{\sin(\pi r/r_c)}{(\pi r/r_c)} & \text{for } r \le r_c, \\ 0 & \text{for } r > r_c, \end{cases}$$

and let $(C_{n\ell}, \widetilde{R}^0_{n\ell})_{n\in\mathbb{N}^*}$ be the unique solution to:

$$\begin{cases} -\frac{1}{2}(\widetilde{R}^0_{n\ell})''(r) + \frac{\ell(\ell+1)}{2r^2}\widetilde{R}^0_{n\ell} + (V_\ell^{\mathrm{PP}} - E_n)\widetilde{R}^0_{n\ell} = C_{n\ell}\chi(r)\widetilde{R}^0_{n\ell}, \quad 0 \le r \le r_c \\ \widetilde{R}^0_{n\ell}(0) = 0 \\ \widetilde{R}^0_{n\ell}(r_c) = R^0_{n\ell}(r_c), \quad (\widetilde{R}^0_{n\ell})'(r_c) = (R^0_{n\ell})'(r_c). \end{cases} \tag{1.11}$$

Let $p^0_{n\ell}$ be the auxiliary functions:

$$p^0_{n\ell}(r) = \frac{\chi(r)\widetilde{R}^0_{n\ell}(r)}{\left(\widetilde{R}^0_{n\ell}(r)|\chi(r)\widetilde{R}^0_{n\ell}(r)\right)},$$

where

$$(f|g) = \int_0^{r_c} f(r)g(r)\,\mathrm{d}r.$$

PAW functions Finally the radial part of all the PAW functions are constructed with a Gram-Schmidt process. We describe it here assuming that only two quantum numbers $n_1 < n_2$ are needed for the computation. However, one should bear in mind that, on the one hand, usually, $n_2 = n_1 + 1$, and on the other hand, although limiting the procedure to two quantum numbers is in general sufficient for practical purposes, it is straightforward to generalize the following orthogonalization procedure to an arbitrary number of quantum numbers.

1. *Basis:* the first set of functions $\widetilde{R}_{n_1\ell}$, $p_{n_1\ell}$ and $R_{n_1\ell}$, corresponding to the lowest principal quantum number n used, are defined by

$$\widetilde{R}_{n_1\ell} = \widetilde{R}^0_{n_1\ell}, \quad p_{n_1\ell} = p^0_{n_1\ell}, \quad R_{n_1\ell} = R^0_{n_1\ell}.$$

2. *Inductive step:* if there is a second radial basis function for $n_2 > n_1$,

   - first, the function $\widetilde{R}^0_{n_2\ell}$ is orthogonalized against $p_{n_1\ell}$:

$$\widetilde{R}_{n_2\ell}(r) = \mathcal{F}_{n_2\ell}\left(\widetilde{R}^0_{n_2\ell}(r) - \widetilde{R}_{n_1\ell}(r)\left(p_{n_1\ell}|\widetilde{R}^0_{n_2\ell}\right)\right), \tag{1.12}$$

     where the factor

$$\mathcal{F}_{n_2\ell} = \frac{1}{\left(1 - \left(\widetilde{R}^0_{n_2\ell}|p_{n_1\ell}\right)\left(\widetilde{R}^0_{n_1\ell}|p_{n_2\ell}\right)\right)^{1/2}}$$

     is a normalization constant;

   - similarly, the function $p^0_{n_2\ell}$ is orthogonalized against $\widetilde{R}_{n_1\ell}$ by noticing that $\left(\widetilde{R}^0_{n_2\ell}|p_{n_1\ell}\right) = \left(\widetilde{R}^0_{n_1\ell}|p^0_{n_2\ell}\right)$:

$$p_{n_2\ell}(r) = \mathcal{F}_{n_2\ell}\left(p^0_{n_2\ell} - p_{n_1\ell}\left(\widetilde{R}^0_{n_2\ell}|p_{n_1\ell}\right)\right);$$

   - finally, to ensure the continuity between the radial functions $R_{n_2\ell}$ and $\widetilde{R}_{n_2\ell}$, we apply to $R^0_{n_2\ell}$ the same linear combination in Equation (1.12)

$$R_{n_2\ell}(r) = \mathcal{F}_{n_2\ell}\left(R^0_{n_2\ell}(r) - R_{n_1\ell}(r)\left(p_{n_1\ell}|\widetilde{R}^0_{n_2\ell}\right)\right).$$

The PAW functions are given by

$$\phi_{n\ell m}(x) = \frac{R_{n\ell}(r)}{r}\, Y_{\ell m}(\theta,\varphi),$$

$$\widetilde{\phi}_{n\ell m}(x) = \frac{\widetilde{R}_{n\ell}(r)}{r}\, Y_{\ell m}(\theta,\varphi),$$

$$\tilde{p}_{n\ell m}(x) = \frac{p_{n\ell}(r)}{r}\, Y_{\ell m}(\theta,\varphi).$$

Organization of the paper In this paper, we apply the VPAW formalism to the double Dirac potential with periodic boundary conditions in one dimension. The eigenfunctions of this model have a derivative jump at the positions of the Dirac potentials which is similar to the electronic wave function cusp. Furthermore, the eigenvalues and eigenfunctions being known analytically, it is possible to confront our theoretical results to very accurate numerical tests.

In Section 2, we carefully present the VPAW method in our framework. In Section 3, Fourier decay estimates of the pseudo wave functions are given as well as estimates on the computed eigenvalues. Proofs of these results are gathered in Section 4. In Section 5, we discuss the effect of the addition of a smooth potential to the double Dirac model. Numerical simulations which confirm the obtained theoretical results are provided in Section 6.

Notation

From now on, $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $L^2_{\text{per}}(0,1)$.

Let f be a piecewise continuous function. We denote by:

$$[f]_x := f(x_+) - f(x_-),$$

where $f(x_+)$ and $f(x_-)$ are respectively the right-sided and left-sided limits of f at x.

Let f be a continuous function. We denote by

$$\underbrace{\int \cdots \int}_{2j+2} f(x) = \int_0^x \int_0^{t_1} \cdots \int_0^{t_{2j+1}} f(t_{2j+2})\, \mathrm{d}t_{2j+2} \ldots \mathrm{d}t_1$$

the $(2j+2)$-primitive function of f vanishing at 0 as well as its first $(2j+1)$-st derivatives.

For a and b in $\mathbb{R}^N$, $a \cdot b$ is the Euclidean inner product. e_k is the k-th canonical vector of $\mathbb{R}^d$ or $\mathbb{R}^N$. I_N is the identity matrix of size N.

2 The VPAW method for a one-dimensional model

2.1 The double Dirac potential

We are interested in the lowest eigenvalues of the 1-D periodic Schrödinger operator H on $L^2_{\text{per}}(0,1) := \{\phi \in L^2_{\text{loc}}(\mathbb{R}) \mid \phi \text{ 1-periodic}\}$ with form domain $H^1_{\text{per}}(0,1) := \{\phi \in L^2_{\text{loc}}(\mathbb{R}) \mid \phi' \in L^2_{\text{loc}}(\mathbb{R})\}$:

$$H = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} - Z_0 \sum_{k\in\mathbb{Z}} \delta_k - Z_a \sum_{k\in\mathbb{Z}} \delta_{k+a}, \tag{2.1}$$

where $0 < a < 1$, $Z_0, Z_a > 0$.

A mathematical analysis of this model has been carried out in [5]. There are two negative eigenvalues $E_0 = -\omega_0^2$ and $E_1 = -\omega_1^2$ which are given by the zeros of the function

$$f(\omega) = 2\omega^2(1 - \cosh(\omega)) + (Z_0 + Z_a)\omega \sinh(\omega) - Z_0 Z_a \sinh(a\omega)\sinh((1-a)\omega).$$

The corresponding eigenfunctions are

$$\psi_k(x) = \begin{cases} A_{1,k}\cosh(\omega_k x) + B_{1,k}\sinh(\omega_k x)\ , & 0 \le x \le a, \\ A_{2,k}\cosh(\omega_k x) + B_{2,k}\sinh(\omega_k x)\ , & a \le x \le 1, \end{cases}$$

where the coefficients $A_{1,k}$, $A_{2,k}$, $B_{1,k}$ and $B_{2,k}$ are determined by the continuity conditions and the derivative jumps at 0 and a.

There is an infinity of positive eigenvalues $E_{k+2} = \omega_{k+2}^2$ which are given by the k-th zero of the function :

$$f(\omega) = 2\omega^2(1 - \cos(\omega)) + (Z_0 + Z_a)\omega \sin(\omega) + Z_0 Z_a \sin(a\omega)\sin((1-a)\omega),$$

and the corresponding eigenfunctions $H\psi_k = \omega_k^2 \psi_k$ are

$$\psi_k(x) = \begin{cases} A_{1,k}\cos(\omega_k x) + B_{1,k}\sin(\omega_k x)\ , & 0 \le x \le a, \\ A_{2,k}\cos(\omega_k x) + B_{2,k}\sin(\omega_k x)\ , & a \le x \le 1, \end{cases}$$

where again the coefficients $A_{1,k}$, $A_{2,k}$, $B_{1,k}$ and $B_{2,k}$ are determined by the continuity conditions and the derivative jumps at 0 and a.

2.2 The VPAW method

The principle of the VPAW method consists in replacing the original eigenvalue problem

$$H\psi = E\psi,$$

by the generalized eigenvalue problem:

$$(\mathrm{Id}+T)^* H(\mathrm{Id}+T)\tilde{\psi} = E(\mathrm{Id}+T)^*(\mathrm{Id}+T)\tilde{\psi}, \tag{2.2}$$

where $\mathrm{Id}+T$ is an invertible bounded linear operator on $L^2_{\mathrm{per}}(0,1)$. Thus both problems have the same eigenvalues and it is straightforward to recover the eigenfunctions of the former from the generalized eigenfunctions of the latter:

$$\psi = (\mathrm{Id}+T)\tilde{\psi}. \tag{2.3}$$

T is the sum of two operators acting near the atomic sites

$$T = T_0 + T_a.$$

To define T_0, we fix an integer N and a radius $0 < \eta < \min(\frac{a}{2}, \frac{1-a}{2})$ so that T_0 and T_a act on two disjoint regions $\bigcup\limits_{k\in\mathbb{Z}}[-\eta+k, \eta+k]$ and $\bigcup\limits_{k\in\mathbb{Z}}[a-\eta+k, a+\eta+k]$ respectively.

Atomic wave function ϕ_k Let H_0 be the operator defined by :

$$H_0 = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} - Z_0 \sum_{k\in\mathbb{Z}} \delta_k.$$

By parity, the eigenfunctions of this operator are even or odd. The odd eigenfunctions are in fact $x \mapsto \sin(2\pi k x)$ and the even ones are the 1-periodic functions such that

$$\begin{cases} \phi_0(x) := \cosh(\omega_0(x-\frac{1}{2})), \text{for } x \in [0,1], \\ \phi_k(x) := \cos(\omega_k(x-\frac{1}{2})), \text{for } x \in [0,1], k \in \mathbb{N}^*. \end{cases}$$

To construct T_0, we will only select the non-smooth thus even eigenfunctions $(\phi_k)_{1\le k\le N}$ and denote by $(\epsilon_k)_{1\le k\le N}$ the corresponding eigenvalue:

$$H_0\phi_k = \epsilon_k\phi_k.$$

Pseudo wave function $\tilde{\phi}_i$ The pseudo wave functions $(\tilde{\phi}_i)_{1\le i\le N} \in \left(H^1_{\mathrm{per}}(0,1)\right)^N$ are defined as follows:

1. for $|x| \notin \bigcup\limits_{k\in\mathbb{Z}}[-\eta+k, \eta+k]$, $\tilde{\phi}_i(x) = \phi_i(x)$.
2. for $|x| \in \bigcup\limits_{k\in\mathbb{Z}}[-\eta+k, \eta+k]$, $\tilde{\phi}_i$ is an even polynomial of degree at most $2d-2$, $d \ge N$.
3. $\tilde{\phi}_i$ is C^{d-1} at η *i.e.* $\tilde{\phi}_i^{(k)}(\eta) = \phi_i^{(k)}(\eta)$ for $0 \le k \le d-1$.

Projector functions $\tilde{p}_i$ Let ρ be a positive, continuous function with support $[-1,1]$ and $\rho_\eta(t) = \sum_{k\in\mathbb{Z}} \rho(\frac{t-k}{\eta})$. The projector functions $(\tilde{p}_i)_{1\le i\le N}$ are obtained by an orthonormalization procedure from the functions $p_i(t) = \rho_\eta(t)\tilde{\phi}_i(t)$ in order to satisfy the duality condition :

$$\langle \tilde{p}_i \, , \tilde{\phi}_j \rangle = \delta_{ij}.$$

More precisely, we compute the matrix $B_{ij} := \langle p_i \, , \tilde{\phi}_j \rangle$ and invert it to obtain the projector functions

$$\tilde{p}_k = \sum_{j=1}^{N} (B^{-1})_{kj} p_j.$$

The matrix B is the Gram matrix of the functions $\tilde{\phi}_j$ for the weight ρ_η. The orthogonalization is possible only if the family $(\tilde{\phi}_i)_{1\le i\le N}$ is independent - thus necessarily $d \ge N$.

T_0 and T_a are given by :

$$T_0 = \sum_{i=1}^{N} (\phi_i - \tilde{\phi}_i) \langle \tilde{p}_i \, , \bullet \rangle, \qquad T_a = \sum_{i=1}^{N} (\phi_i^a - \tilde{\phi}_i^a) \langle \tilde{p}_i^a \, , \bullet \rangle \ , \tag{2.4}$$

where ϕ_i^a are singular eigenfunctions of the operator $H_a = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} - Z_a \sum_{k\in\mathbb{Z}} \delta_{a+k}$ and $\tilde{\phi}_i^a$, $\tilde{p}_i^a$ are defined as before.

In the VPAW method, the generalized eigenvalue problem (2.2) is solved by expanding $\tilde{\psi}$ in plane-waves.

Remark 2.1. *Here we have followed the Vanderbilt scheme to generate the pseudo wave functions and the projector functions with the difference that the orthogonalized functions p are taken from the Blöchl construction.*

2.3 Well-posedness of the VPAW method

To be well-posed the VPAW method requires

1. the family of pseudo wave functions $(\tilde{\phi}_i)_{1\le i\le N}$ to be independent on $[-\eta,\eta]$, so that the projector functions $(\tilde{p}_k)_{1\le k\le N}$ are well defined,
2. $(\mathrm{Id} + T)$ to be invertible.

The conditions on the VPAW functions and parameters are given by the following propositions. Proofs can be found in Section 4.

Proposition 2.2 (Linear independence of the pseudo wave functions)**.** *Let $N \in \mathbb{N}^*$ and $d \ge N$. There exists $\eta_0 > 0$ such that for all $0 < \eta \le \eta_0$, the family $(\tilde{\phi}_i|_{[-\eta,\eta]})_{1\le i\le N}$ is linearly independent.*

Proposition 2.3 (Invertibility of $\mathrm{Id} + T$)**.** *The operator $\mathrm{Id} + T$ is invertible in $L^2_{\mathrm{per}}(0,1)$ if and only if the matrix $(\langle \tilde{p}_k \, , \phi_\ell \rangle)_{1\le k,\ell\le N}$ is invertible.*

From now on, we will establish our results under the following

Assumption : the matrix $(\langle \tilde{p}_k \, , \phi_\ell \rangle)_{1\le k,\ell\le N}$ is invertible for all $0 < \eta \le \eta_0$.

3 Main results

We know from (2.3) that

$$\tilde{\psi} = \psi - \sum_{i=1}^{N} (\phi_i - \tilde{\phi}_i) \langle \tilde{p}_i \, , \tilde{\psi} \rangle - \sum_{i=1}^{N} (\phi_i^a - \tilde{\phi}_i^a) \langle \tilde{p}_i^a \, , \tilde{\psi} \rangle.$$

In addition, $\tilde{\psi}$ is a piecewise smooth function with first derivative jumps (due to ψ and the atomic wave function ϕ_k) at points of $\mathbb{Z}$, $\mathbb{Z}+a$ and d-th derivative jumps (due to the pseudo wave functions $\tilde{\phi}_k$) at points of $\mathbb{Z}\pm\eta$ and $\mathbb{Z}+a\pm\eta$. These singularities drive the decays of the Fourier coefficients. Thus to study the Fourier convergence rate, it suffices to study the dependency of the different singularities with respect to N -the number of PAW functions used-, d -the smoothness of the pseudo wave functions $\tilde{\phi}_k$- and η -the cut-off radius.

Proposition 3.1 (Derivative jumps at 0)**.** *Let $N \in \mathbb{N}^*$ and $d \geq N$. Then, there exists a positive constant C independent of η such that for $0 \leq j \leq N-1$*

$$\forall\, 0 < \eta \leq \eta_0,\ \left|\left[\tilde{\psi}^{(2j+1)}\right]_0\right| \leq C\eta^{2N-2j}\ , \tag{3.1}$$

and for $j \geq N$

$$\forall\, 0 < \eta \leq \eta_0,\ \left|\left[\tilde{\psi}^{(2j+1)}\right]_0\right| \leq C. \tag{3.2}$$

The proof of Proposition 3.1 relies on the particular structure induced by the equations satisfied by ψ and ϕ_i. Locally around a Dirac potential, their singularities have the same behavior. More precisely, if we consider the even part ψ_e of ψ, the best approximation of ψ_e by N eigenfunctions ϕ_i is of order $2N$. It is then possible to rewrite the singularity at 0 of $\tilde{\psi}$ to make use of this approximation.

Proposition 3.2 (d-th derivative jump at η)**.** *Let $N \in \mathbb{N}^*$ and $d \geq N$. There exists a constant C independent of η such that for $d \leq k \leq 2d-2$*

$$\forall\, 0 < \eta \leq \eta_0,\ \left|\left[\tilde{\psi}^{(k)}\right]_\eta\right| \leq \frac{C}{\eta^{k-1}}.$$

The derivative jump of $\tilde{\psi}$ at η is due to the lack of regularity of the pseudo wave functions $\tilde{\phi}_j$ at η. The latter can be written as rescaled polynomials $P_{2d-2}(\frac{x}{\eta})$ where P_{2d-2} is of degree at most $2d-2$. If we suppose that the coefficients of P_{2d-2} are uniformly bounded in η and if the dependence on η of the projector functions $\tilde{p}_k$ is neglected, by deriving k times the polynomials $P_{2d-2}(\frac{x}{\eta})$, $k \geq d$, we can see why the derivative jump of $\tilde{\psi}$ at η is expected to grow as η^{-k}. Tracking all the dependencies on η, we can in fact show that a factor η can be gained, which is in full agreement with Figure 2.

Using Proposition 4.1 and classical estimates on eigenvalue approximations [25], we have the following theorems.

Theorem 3.3 (Estimates on the Fourier coefficients)**.** *Let $N \in \mathbb{N}^*$ and $d \geq N$. Let $\widehat{\tilde{\psi}}_m$ be the m-th Fourier coefficient of $\tilde{\psi}$. There exists a constant $C > 0$ independent of η and m such that for all $0 < \eta \leq \eta_0$ and $m \geq \frac{1}{\eta}$*

$$\left|\widehat{\tilde{\psi}}_m\right| \leq C\left(\frac{\eta^{2N}}{m^2} + \frac{1}{\eta^{d-1}m^{d+1}}\right).$$

Theorem 3.4 (Estimates on the eigenvalues)**.** *Let $N \in \mathbb{N}^*$ and $d \geq N$. Let E_M^η be an eigenvalue of the variational approximation of (2.2) in a basis of M plane-waves and for a cut-off radius $0 < \eta \leq \eta_0$, and let E be the corresponding exact eigenvalue. There exists a constant $C > 0$ independent of η and M such that for all $0 < \eta \leq \eta_0$ and $M \geq \frac{1}{\eta}$*

$$0 < E_M^\eta - E \leq C\left(\frac{\eta^{4N}}{M} + \frac{1}{\eta^{2d-2}}\frac{1}{M^{2d-1}}\right). \tag{3.3}$$

The first term has the same asymptotic decay in M as the brute force discretization of the problem with the original Dirac potential. However the prefactor η^{4N} can be made small by using a small cut-off radius η and/or a large N. Doing so, we introduce another error term which decays as M^{1-2d}, with a prefactor of order η^{2-2d}. A natural strategy would thus be to balance these two error terms. This allows one to choose the numerical parameters in a consistent way. The numerical tests in Section 6 suggest that the estimate (3.3) is optimal.

4 Proofs

This section is organized as follows. First, we prove that the VPAW method is well defined. The remainder of the section is then devoted to the proofs of Theorems 3.3 and 3.4. After estimating the decay of the Fourier coefficients of the pseudo wave function $\tilde{\psi}$, we will precisely characterize the singularities of the functions ψ and ϕ_k in order to estimate the derivative jumps of $\tilde{\psi}$.

4.1 Well-posedness of the VPAW method

Proof of Proposition 2.2. To prove the linear independence of the pseudo wave functions is equivalent to show that the matrix $(\phi_{j+1}^{(k)}(\eta))_{0\le j\le N-1, 0\le k\le d-1}$ is full rank. In fact, we will show that the submatrix $(\phi_{j+1}^{(k)}(\eta))_{0\le j,k\le N-1}$ is invertible. Using the expression of ϕ_{j+1}, we have for $0 \le j, k \le N-1$,

$$\phi_{j+1}^{k}(\eta) = \begin{cases} (-1)^{k/2}\omega_{j+1}^{k}\cos(\omega_{j+1}(\eta-\frac{1}{2})) & \text{if } k \text{ is even,} \\ (-1)^{(k+1)/2}\omega_{j+1}^{k}\sin(\omega_{j+1}(\eta-\frac{1}{2})) & \text{if } k \text{ is odd} \end{cases}$$

Let $A(\eta)$ the matrix defined by

$$A(\eta) := \begin{pmatrix} \cos(\omega_{j+1}(\eta-\frac{1}{2})) & -\omega_{j+1}\sin(\omega_{j+1}(\eta-\frac{1}{2})) & -\omega_{j+1}^2\cos(\omega_{j+1}(\eta-\frac{1}{2})) & \cdots \end{pmatrix}.$$

The function $\eta \mapsto \det A(\eta)$ is complex analytic, thus if it is not identically equal to 0, there exists an interval $(0, \eta_0)$ with $\eta_0 > 0$ such that the matrix $A(\eta)$ is invertible. It suffices to show that there exists $\eta \in \mathbb{C}$ such that $A(\eta)$ is invertible. Let $\eta = -ix + \frac{1}{2}$, $x > 0$. Then for x large, we have $\cos(\omega_{j+1}x) \sim \frac{1}{2}e^{\omega_{j+1}x}$ and $\sin(\omega_{j+1}x) \sim \frac{1}{2}e^{\omega_{j+1}x}$, thus

$$A(-ix+\tfrac{1}{2}) = \frac{1}{2}\begin{pmatrix} e^{\omega_{j+1}x} & -\omega_{j+1}e^{\omega_{j+1}x} & -\omega_{j+1}^2e^{\omega_{j+1}x}\cdots \end{pmatrix} + \varepsilon(x),$$

where

$$\|\varepsilon(x)\| \ll \left\| \begin{pmatrix} e^{\omega_{j+1}x} & -\omega_{j+1}e^{\omega_{j+1}x} & -\omega_{j+1}^2e^{\omega_{j+1}x}\cdots \end{pmatrix} \right\|.$$

We have

$$\begin{pmatrix} e^{\omega_{j+1}x} & -\omega_{j+1}e^{\omega_{j+1}x} & -\omega_{j+1}^2e^{\omega_{j+1}x}\cdots \end{pmatrix} = \begin{pmatrix} e^{\omega_1 x} & & 0 \\ & \ddots & \\ 0 & & e^{\omega_N x} \end{pmatrix} \begin{pmatrix} 1 & \omega_1 & & \omega_1^{N-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \omega_N & & \omega_N^{N-1} \end{pmatrix}$$

which is invertible because the phases $(\omega_j)_{1\le j\le N}$ are pairwise distinct. Hence $A(-ix+\frac{1}{2})$ is invertible for x large enough. □

Proof of Proposition 2.3. As T is a finite rank and thus compact operator, proving the statement is equivalent to show that $\mathrm{Ker}(\mathrm{Id}+T) = \{0\}$. First, suppose that the matrix $(\langle \tilde{p}_k\,, \phi_i\rangle)_{k,i}$ is invertible and let $f \in \mathrm{Ker}(\mathrm{Id}+T)$. We have

$$\sum_{i=1}^{N} \langle \tilde{p}_i\,, f\rangle(\phi_i - \tilde{\phi}_i) + f = 0 \tag{4.1}$$

Since $\phi_i - \tilde{\phi}_i$ is supported in $[-\eta, \eta]$ we also have $\mathrm{supp}(f) \subset [-\eta, \eta]$.

By multiplying each side of Equation (4.1) by $\tilde{p}_k$, $1 \leq k \leq N$ and integrating on $[-\frac{1}{2}, \frac{1}{2}]$, we obtain:

$$0 = \sum_{i=1}^{N} \langle \tilde{p}_i \, , f \rangle \langle \tilde{p}_k \, , \phi_i - \tilde{\phi}_i \rangle + \langle \tilde{p}_k \, , f \rangle = \sum_{i=1}^{N} \langle \tilde{p}_i \, , f \rangle (\langle \tilde{p}_k \, , \phi_i \rangle - \underbrace{\langle \tilde{p}_k \, , \tilde{\phi}_i \rangle}_{=\delta_{ki}}) + \langle \tilde{p}_k \, , f \rangle \ ,$$

so that

$$\forall 1 \leq k \leq N, \ 0 = \sum_{i=1}^{N} \langle \tilde{p}_i \, , f \rangle \langle \tilde{p}_k \, , \phi_i \rangle.$$

Since we assumed that the matrix $(\langle \tilde{p}_k \, , \phi_i \rangle)_{k,i}$ is invertible,

$$\forall \, 1 \leq i \leq N, \ \langle \tilde{p}_i \, , f \rangle = 0 \ .$$

Going back to (4.1), this implies $f = 0$ and $\mathrm{Id} + T$ is invertible.

Now we suppose that the matrix $(\langle \tilde{p}_k \, , \phi_i \rangle)_{k,i}$ is not invertible. Thus there is $(\alpha_i)_{1 \leq i \leq N}$ such that

$$\forall 1 \leq i \leq N, \ \sum_{j=1}^{N} \alpha_j \langle \tilde{p}_i \, , \phi_j \rangle = 0.$$

Let $f(x) = \sum_{j=1}^{N} \alpha_j (\phi_j - \tilde{\phi}_j)$. Then

$$\begin{aligned}
(\mathrm{Id} + T) f &= \sum_{j=1}^{N} \alpha_j (\phi_j - \tilde{\phi}_j) + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_j \langle \tilde{p}_i \, , \phi_j - \tilde{\phi}_j \rangle (\phi_j - \tilde{\phi}_j) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_j \langle \tilde{p}_i \, , \phi_j \rangle (\phi_j - \tilde{\phi}_j) \\
&= 0
\end{aligned}$$

Thus $\mathrm{Ker}(I + T) \neq \{0\}$ and $(I + T)$ is not invertible. □

4.2 Structure and approximation lemmas

A key intermediate result in our study is the estimation of the decay of the Fourier coefficients of $\tilde{\psi}$ as a function of its derivative jumps.

Proposition 4.1. *Let $N \in \mathbb{N}^*$ and $d \geq N$. Let $\widehat{\tilde{\psi}}_m$ be the m-th Fourier coefficient of $\tilde{\psi}$. Then*

$$\begin{aligned}
\widehat{\tilde{\psi}}_m = & \sum_{j=0}^{\lfloor \frac{d}{2} \rfloor - 1} \frac{1}{(2i\pi m)^{2+2j}} \left[\tilde{\psi}^{(2j+1)} \right]_0 + \sum_{k=d}^{2d-2} \frac{e^{\mp 2i\pi m \eta}}{(2i\pi m)^{k+1}} \left[\tilde{\psi}^{(k)} \right]_{\pm \eta} \\
& + e^{-2i\pi m a} \left(\sum_{j=0}^{\lfloor \frac{d}{2} \rfloor - 1} \frac{1}{(2i\pi m)^{2+2j}} \left[\tilde{\psi}^{(2j+1)} \right]_a + \sum_{k=d}^{2d-2} \frac{e^{\mp 2i\pi m \eta}}{(2i\pi m)^{k+1}} \left[\tilde{\psi}^{(k)} \right]_{a \pm \eta} \right) \\
& + \frac{1}{(2i\pi m)^{2d-1}} \int_0^1 \tilde{\psi}^{(2d-1)}(x) e^{-2i\pi m x} \, \mathrm{d}x.
\end{aligned}$$

Proof. This result follows from the definition of the Fourier coefficients and integration by parts. □

In view of the Proposition 4.1, the decay of the Fourier coefficients can be inferred from the derivative jumps of $\tilde{\psi}$ according to the VPAW parameters. The singularities of $\tilde{\psi}$ at integer values are caused by the singularity of the functions ψ and ϕ_k. Thus, to get an accurate characterization of the singularities of $\tilde{\psi}$, we need to precisely know how the functions ψ and ϕ_k behave in a neighborhood of their singularities.

Lemma 4.2 (Structure lemma)**.** *Let ψ be an eigenfunction (2.1) associated to the eigenvalue E. Then in a neighborhood of* 0*, we have the following expansion :*

$$\psi(x) = \psi(0)\left(\sum_{j=0}^{k}\frac{(-E)^j}{(2j)!}x^{2j} - \frac{Z_0}{2}\sum_{j=0}^{k}\frac{(-E)^j}{(2j+1)!}|x|^{2j+1}\right) + \frac{\psi'(0_+)+\psi'(0_-)}{2}\sum_{j=0}^{k}\frac{(-E)^j}{(2j+1)!}x^{2j+1} + \psi_{2k+2}(x), \quad (4.2)$$

where ψ_{2k+2} is a C^{2k+2} function satisfying in a neighbourhood of 0*,*

$$\begin{cases}\psi_{2k+2}^{(2k+2)} = (-E)^{k+1}\psi,\\ |\psi_{2k+2}(x)| \le C\frac{|-E|^{k+1}}{(2k+2)!}|x|^{2k+2}.\end{cases}$$

Proof. This lemma is proved by induction.

Initialization For $k = 0$, let

$$\theta_2(x) = \psi(x) + Z_0\frac{|x|}{2}\psi(0).$$

We differentiate θ_2 twice:

$$\theta_2''(x) = \psi''(x) - [\psi']_0\delta_0 = -E\psi(x), \quad \text{on } (-\tfrac{1}{2},\tfrac{1}{2}). \quad (4.3)$$

The function ψ being continuous, θ_2 is C^2 in a neighborhood of 0. Moreover,

$$\theta_2(0) = \psi(0),$$

and

$$\theta_2'(x) = \psi'(x) - \frac{\operatorname{sign}(x)}{2}[\psi']_0.$$

When x tends to 0_+ or 0_-, we obtain the same expression:

$$\theta_2'(0) = \frac{\psi'(0_+)+\psi'(0_-)}{2}.$$

Setting

$$\psi_2(x) = \theta_2(x) - \psi(0) - \frac{\psi'(0_+)+\psi'(0_-)}{2}x,$$

the statement is true for $k = 0$.

Inductive step Suppose the statement is true for $k-1$. Then, we have in a neighbourhood of 0,

$$\psi_{2k}^{(2k)}(x) = (-E)^k\psi(x).$$

Let

$$\theta_{2k+2}(x) = \psi_{2k}(x) - (-E)^k\frac{|x|}{2}[\psi']_0\frac{x^{2k}}{(2k+1)!}.$$

Then

$$\theta_{2k+2}^{(2k)}(x) = (-E)^k\left(\psi(x) - \frac{|x|}{2}[\psi']_0\right),$$

so that in view of (4.3),

$$\theta_{2k+2}^{(2k+2)}(x) = (-E)^k\left(\psi''(x) - [\psi']_0\delta_0\right) = (-E)^{k+1}\psi(x),$$

in the neighbourhood of 0.

So θ_{2k+2} is a C^{2k+2} function in a neighbourhood of 0 and we have:

$$\psi(x) = \psi(0)\left(\sum_{j=0}^{k-1}\frac{(-E)^j}{(2j)!}x^{2j} - Z_0\sum_{j=0}^{k}\frac{(-E)^j}{(2j+1)!}\frac{|x|^{2j+1}}{2}\right) + \sum_{j=0}^{k-1}\frac{(-E)^j}{(2j+1)!}\frac{\psi'(0_+)+\psi'(0_-)}{2}x^{2j+1} + \theta_{2k+2}(x).$$

Il suffices to evaluate $\theta_{2k+2}^{(2k)}(0)$ and $\theta_{2k+2}^{(2k+1)}(0)$ to conclude the proof. We have

$$\theta_{2k+2}^{(2k)}(0) = (-E)^k\psi(0)\ ,$$

and

$$\theta_{2k+2}^{(2k+1)}(x) = (-E)^k\left(\psi'(x) - \frac{\mathrm{sign}(x)}{2}[\psi']_0\right)\ ,$$

so if x tends to 0, we have :

$$\theta_{2k+2}^{(2k+1)}(0) = \frac{\psi'(0_+)+\psi'(0_-)}{2}(-E)^k.$$

Define

$$\psi_{2k+2}(x) = \theta_{2k+2}(x) - \psi(0)\frac{x^{2k}}{(2k)!} - \frac{\psi'(0_+)+\psi'(0_-)}{2}\frac{x^{2k+1}}{(2k+1)!}\ ,$$

and the induction is proved. □

Let ψ_e be the even part of ψ. We have in a neighbourhood of 0

$$\psi_e(x) = \psi(0)\sum_{k=0}^{N-1}\left(\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\frac{|x|^{2k+1}}{(2k+1)!}\right)(-E)^k + \frac{1}{2}(\psi_{2N}(x)+\psi_{2N}(-x)). \tag{4.4}$$

Lemma 4.3 (Approximation)**.** *There exist constants $(c_j)_{1\le j\le N}\in\mathbb{R}^N$ satisfying*

$$\psi_e(x) = \sum_{j=1}^{N} c_j\phi_j(x) + \mathcal{O}\left(x^{2N}\right), \qquad \text{as } x\to 0.$$

Proof. By Lemma 4.2 applied to ϕ_j, we have

$$\phi_j(x) = \phi_j(0)\left(\sum_{k=0}^{N-1}\frac{(-\epsilon_j)^k}{(2k)!}x^{2k} - \frac{Z_0}{2}\sum_{k=0}^{N-1}\frac{(-\epsilon_j)^k}{(2k+1)!}|x|^{2k+1}\right) + \phi_{j,2N}(x)\ ,$$

where

$$\phi_{j,2N}(x) = \mathcal{O}(x^{2N}).$$

So

$$\psi_e(x) - \sum_{j=1}^{N} c_j\phi_j(x) = \sum_{k=0}^{N-1}\left(\psi(0)(-E)^k - \sum_{j=1}^{N}(-\epsilon_j)^k c_j\phi_j(0)\right)\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\sum_{k=0}^{N-1}\left(\psi(0)(-E)^k - \sum_{j=1}^{N}(-\epsilon_j)^k c_j\phi_j(0)\right)\frac{|x|^{2k+1}}{(2k+1)!} + \psi_{2N}(x) - \sum_{j=1}^{N} c_j\phi_{j,2N}(x).$$

To prove the lemma, it remains to show that there exist coefficients (c_j) such that :

$$\forall\, 0\le k\le N-1,\ \sum_{j=1}^{N}(-\epsilon_j)^k c_j\phi_j(0) = \psi(0)(-E)^k\ .$$

We have chosen the functions ϕ_j so that $\phi_j(0)\neq 0$. By defining $\alpha_j = c_j\phi_j(0)$, we recognize a Vandermonde linear system. The eigenvalues ϵ_j are all different so the system is invertible and the lemma is proved. □

4.3 Derivative jumps at 0

Recall $p_i(t) = \rho_\eta(t)\tilde{\phi}_i(t)$. We will introduce some notation used in the next proofs:

$$\begin{aligned}
p(t) &:= (p_1(t), \dots, p_N(t))^T \in \mathbb{R}^N, \\
\langle \tilde{p}\,, \tilde{\psi}\rangle &:= (\langle \tilde{p}_1\,, \tilde{\psi}\rangle, \dots, \langle \tilde{p}_N\,, \tilde{\psi}\rangle)^T \in \mathbb{R}^N, \\
\langle \tilde{p}\,, \psi\rangle &:= (\langle \tilde{p}_1\,, \psi\rangle, \dots, \langle \tilde{p}_N\,, \psi\rangle)^T \in \mathbb{R}^N, \\
\Phi(t) &:= (\phi_1(t), \dots, \phi_N(t))^T \in \mathbb{R}^N, \\
\widetilde{\Phi}(t) &:= (\tilde{\phi}_1(t), \dots, \tilde{\phi}_N(t))^T \in \mathbb{R}^N, \\
B &:= (\langle p_i, \tilde{\phi}_j\rangle)_{1\le i,j\le N} \in \mathbb{R}^{N\times N}, \\
\widetilde{A} &:= (\langle \tilde{p}_i, \phi_j\rangle)_{1\le i,j\le N} \in \mathbb{R}^{N\times N}, \\
A &:= (\langle p_i, \phi_j\rangle)_{1\le i,j\le N} = B\widetilde{A} \in \mathbb{R}^{N\times N}.
\end{aligned}$$

Lemma 4.4. *Let $(c_k)_{1\le k\le N}$ be any vector of $\mathbb{R}^N$ and $j \in \mathbb{N}$. Then,*

$$[\tilde{\psi}^{(2j+1)}]_0 = -Z_0\left((-E)^j\psi(0) - \sum_{k=1}^N c_k(-\epsilon_k)^j\phi_k(0) - \langle A^{-1}p\,, \psi - \sum_{k=1}^N c_k\phi_k\rangle \cdot \mathcal{E}^j\Phi(0)\right).$$

where $\mathcal{E}$ is the diagonal matrix

$$\mathcal{E} = \begin{pmatrix} -\epsilon_1 & 0 & \dots & 0 \\ 0 & -\epsilon_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & -\epsilon_N \end{pmatrix}.$$

Proof. We first prove the statement for $j = 0$. In a neighbourhood of 0, we have:

$$\tilde{\psi} = \psi - \sum_{i=1}^N \langle \tilde{p}_i\,, \tilde{\psi}\rangle(\phi_i - \tilde{\phi}_i). \tag{4.5}$$

Using the equations satisfied by ψ and ϕ_i gives for the first derivative jump at 0 of $\tilde{\psi}$:

$$\begin{aligned}
[\tilde{\psi}']_0 &= [\psi']_0 - \sum_{i=1}^N \langle \tilde{p}_i\,, \tilde{\psi}\rangle[\phi_i']_0 \\
&= -Z_0\left(\psi(0) - \sum_{i=1}^N \langle \tilde{p}_i\,, \tilde{\psi}\rangle\phi_i(0)\right).
\end{aligned} \tag{4.6}$$

Multiplying equation (4.5) by $\tilde{p}_k$ and integrating on $[-1/2, 1/2]$,

$$\langle \tilde{p}_k\,, \psi\rangle - \sum_{i=1}^N \langle \tilde{p}_i\,, \tilde{\psi}\rangle\langle \tilde{p}_k\,, \phi_i\rangle = \langle \tilde{p}_k\,, \tilde{\psi}\rangle - \sum_{i=1}^N \langle \tilde{p}_i\,, \tilde{\psi}\rangle \underbrace{\langle \tilde{p}_k\,, \tilde{\phi}_i\rangle}_{=\delta_{ki}} = 0\ ,$$

$$\langle \tilde{p}_k\,, \psi\rangle = \sum_{i=1}^N \langle \tilde{p}_i\,, \tilde{\psi}\rangle\langle \tilde{p}_k\,, \phi_i\rangle\ .$$

Therefore

$$\widetilde{A}\langle \tilde{p}\,, \tilde{\psi}\rangle = \langle \tilde{p}\,, \psi\rangle \tag{4.7}$$

Likewise

$$B\langle \tilde{p}\,, \psi\rangle = \langle p\,, \psi\rangle,$$

since for $1 \leq i \leq N$:

$$\begin{aligned}\sum_{j=1}^{N} B_{ij}\langle \tilde{p}_j \, , \psi\rangle &= \sum_{j=1}^{N}\langle p_i \, , \tilde{\phi}_j\rangle\langle \tilde{p}_j \, , \psi\rangle \\ &= \langle p_i \, , \psi\rangle\end{aligned}$$

According to Proposition 2.3, the matrix $\widetilde{A}$ is invertible and so is B since $B = A\widetilde{A}$, with A invertible by assumption. We therefore have

$$\begin{aligned}\sum_{i=1}^{N}\langle \tilde{p}_i \, , \tilde{\psi}\rangle\phi_i(0) &= \langle \tilde{p} \, , \tilde{\psi}\rangle \cdot \Phi(0) \\ &= \widetilde{A}^{-1}\langle \tilde{p} \, , \psi\rangle \cdot \Phi(0) \\ &= (B\widetilde{A})^{-1}\langle p \, , \psi\rangle \cdot \Phi(0) \\ &= \langle A^{-1}p \, , \psi\rangle \cdot \Phi(0).\end{aligned}$$

We thus obtain the more compact form:

$$[\tilde{\psi}']_0 = -Z_0\left(\psi(0) - \langle A^{-1}p \, , \psi\rangle \cdot \Phi(0)\right).$$

To complete the proof the lemma, it suffices to show that

$$\langle A^{-1}p \, , \phi_{i+1}\rangle = e_i,$$

where e_i is the i-th vector of the canonical basis of $\mathbb{R}^N$. This is straightforward since $\langle p \, , \phi_{i+1}\rangle$ is simply the i-th column of the matrix A.

For $j \geq 1$, we proceed in the same way using

$$\begin{cases}[\psi^{(2j+1)}]_0 = -Z_0(-E)^j\psi(0), \\ [\phi_k^{(2j+1)}]_0 = -Z_0(-\epsilon_k)^j\phi_k(0).\end{cases}$$

□

Remark 4.5. *Notice that we showed*

$$\langle \tilde{p} \, , \tilde{\psi}\rangle = \langle A^{-1}p \, , \psi\rangle. \tag{4.8}$$

This equality will be used later in the estimation of the d-th derivative jump.

To prove Lemma 3.1, it remains to study the behavior of $A^{-T}\mathcal{E}^j\Phi(0)$ as η goes to 0. By assumption, A is invertible for all $\eta > 0$ but when $\eta = 0$, A is a rank 1 matrix. Actually, in the special case of the 1D Schrödinger operator with Dirac potentials, we have a precise characterization of the behavior of $A^{-T}\mathcal{E}^j\Phi(0)$ as η goes to 0.

Lemma 4.6. *Let f be a function in $L^2_{\mathrm{per}}(0,1)$ and $Q(t) = (Q_0(t), \ldots, Q_{d-1}(t))^T$ be a vector of even polynomials which forms a basis of the space of even polynomials of degree at most $2d-2$. Let G_η be the $d \times N$ matrix and C_η the $N \times d$ matrix defined by:*

$$\begin{aligned}G_\eta &= \int_{-1}^{1} \rho(t)Q(t)\Phi(\eta t)^T \, \mathrm{d}t, \\ \widetilde{\Phi}(t) &= C_\eta Q(t/\eta), \ \forall t \in (-\eta, \eta).\end{aligned}$$

Then we have

$$\langle A^{-1}p \, , f\rangle = \int_{-1}^{1} \rho(t)(C_\eta G_\eta)^{-1}C_\eta Q(t)f(\eta t) \, \mathrm{d}t. \tag{4.9}$$

Proof. We have

$$\begin{aligned} A_{ij} &= \langle p_{i+1}\,, \phi_{j+1}\rangle \\ &= \int_{-\eta}^{\eta} \rho(t/\eta)\tilde{\phi}_{i+1}(t)\phi_{j+1}(t)\,\mathrm{d}t \\ &= \eta\int_{-1}^{1} \rho(t)\tilde{\phi}_{i+1}(\eta t)\phi_{j+1}(\eta t)\,\mathrm{d}t. \end{aligned}$$

Therefore,

$$\begin{aligned} A &= \eta\int_{-1}^{1} \rho(t)\widetilde{\Phi}(\eta t)\Phi(\eta t)^T\,\mathrm{d}t \\ &= \eta\, C_\eta \int_{-1}^{1} \rho(t)Q(t)\Phi(\eta t)^T\,\mathrm{d}t \\ &= \eta\, C_\eta G_\eta. \end{aligned}$$

Since $(\tilde{\phi}_i)_{1\le i\le N}$ is free, the matrix C_η is invertible :

$$A^{-1}p = \frac{1}{\eta}\rho(t/\eta)(C_\eta G_\eta)^{-1}C_\eta Q(t/\eta).$$

Thus

$$\begin{aligned} \langle A^{-1}p\,, f\rangle &= \frac{1}{\eta}\int_{-\eta}^{\eta} \rho(t/\eta)(C_\eta G_\eta)^{-1}C_\eta Q(t/\eta)f(t)\,\mathrm{d}t \\ &= \int_{-1}^{1} \rho(t)(C_\eta G_\eta)^{-1}C_\eta Q(t)f(\eta t)\,\mathrm{d}t. \end{aligned}$$

□

Before moving to the next lemma, we introduce the following notation. Let Q_k be the even polynomials of degrees at most $2d-2$ defined by

$$\int_{-1}^{1} \rho(t)Q_k(t)t^{2j}\,\mathrm{d}t = \delta_{kj},\ 0\le k,j\le d-1\ ,$$

and let P_k, $0\le k\le d-1$ and P be defined by

$$\begin{aligned} P_k(t) &= \frac{1}{2^k k!}(t^2-1)^k\ , \\ P(t) &= (P_0(t),\dots,P_{d-1}(t))^T. \end{aligned}$$

It is easy to see that P_k satisfies

$$\begin{cases} P_k^{(j)}(1) = 0,\ 0\le j\le k-1\ , \\ P_k^{(k)}(1) = 1. \end{cases}$$

Let Π be the transition matrix from Q to P :

$$\Pi Q = P.$$

Finally we denote by C_η the matrix of the expansion of $\widetilde{\Phi}$ in the basis Q and $C_\eta^{(P)}$ the matrix of the expansion of $\widetilde{\Phi}$ in the basis P :

$$\begin{cases} \widetilde{\Phi}(t) = C_\eta Q(t/\eta)\ , \\ \widetilde{\Phi}(t) = C_\eta^{(P)}P(t/\eta). \end{cases}$$

It is easy to see that

$$C_\eta = C_\eta^{(P)}\Pi.$$

Lemma 4.7. *For $0 \leq j \leq N-1$, we have*

$$\frac{\eta^{2j}}{(2j)!}(\mathcal{E}^j\Phi(0))^T(C_\eta G_\eta)^{-1}C_\eta \underset{\eta\to 0}{\longrightarrow} e_j^T\begin{pmatrix} I_N & M_\pi\end{pmatrix}, \tag{4.10}$$

where M_π is a $(d-N)\times N$ matrix.

Furthermore

$$\|(C_\eta G_\eta)^{-1}C_\eta\| = \mathcal{O}\left(\frac{1}{\eta^{2N-2}}\right).$$

Remark 4.8. *The main idea of the proof is to use the particular structures of the matrices C_η and G_η. We denote by C_1, C_2, G_1 and G_2 the matrices defined by*

$$\begin{cases} C_\eta = \Big(C_1 \,\Big|\, C_2\Big), \\ G_\eta = \begin{pmatrix} G_1 \\ \hline G_2 \end{pmatrix}. \end{cases}$$

Suppose that C_1 is invertible and such that $\|C_1^{-1}C_2\| = \mathcal{O}(1)$ as $\eta \to 0$, and there exists an invertible matrix H_1 such that

$$\begin{cases} G_1H_1 = I_N + \mathcal{O}(\eta)\ , \\ G_2H_1 = \mathcal{O}(\eta)\ , \\ e_0^TH_1^{-1} = \Phi(0)^T. \end{cases}$$

Then it is easy to see that

$$\Phi(0)^T(C_\eta G_\eta)^{-1}C_\eta = e_0^T\Big(I_N \ \Big|\ C_1^{-1}C_2\Big) + \mathcal{O}(\eta).$$

Using Lemma 4.2 applied to Φ, it is easy to unveil the dependence in η of the matrix G_η but we have no hint on the structure of C_η. Likewise, $C_\eta^{(P)}$ is easy to study but the matrix $\int_{-1}^1 \rho(t)P(t)\Phi(\eta t)^T\,\mathrm{d}t$ is not. So we have to work with both bases P and Q, exhibit the structures of the matrices $C_\eta^{(P)}$ and G_η and recombine everything with the transition matrix Π.

Before proving Lemma 4.7, we state some properties of the matrix $C_\eta^{(P)}$ and its submatrices C_1, C_2.

Lemma 4.9. *Let $N \in \mathbb{N}^*$ and $d \geq N$. Let $C_1 \in \mathbb{R}^{N\times N}$ and $C_2 \in \mathbb{R}^{N\times(d-N)}$ be the matrices such that:*

$$C_\eta^{(P)} = \Big(C_1 \,\Big|\, C_2\Big). \tag{4.11}$$

Let $(g_k)_{0\leq k\leq N-1}$ be the dual family of the vectors $(\eta^k\Phi^{(k)}(\eta))_{0\leq k\leq N-1}$ and K_1 be the matrix

$$K_1 = \begin{pmatrix} g_0^T \\ \vdots \\ g_{N-1}^T \end{pmatrix} \in \mathbb{R}^{N\times N}.$$

Then, there exists an upper triangular matrix $\mathcal{P}$ independent of η of the form

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & * & * \\ \vdots & 0 & \ddots & * \\ 0 & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N\times N}$$

such that

$$\begin{cases} C_1^{-1} = \mathcal{P}K_1, \\ C_1^{-1}C_2 = M + \mathcal{O}(\eta), \end{cases}$$

where $M \in \mathbb{R}^{N\times(d-N)}$ is a matrix independent of η.

Remark 4.10. *The particular form of the matrix $\mathcal{P}$ will be used in the estimation of the d-th derivative jump of $\tilde{\psi}$ at η (Lemma 3.2) and of $Tf(x)$ (Lemmas 4.13 and 4.14).*

Proof. Let $(c_k)_{0\leq k\leq d-1}$ be the columns of $C_\eta^{(P)}$. By the continuity conditions at η and our choice of the polynomials P_k, we have

$$\forall 0 \leq j \leq d-1,\ c_j = \eta^j \Phi^{(j)}(\eta) - \sum_{k=0}^{j-1} P_k^{(j)}(1) c_k. \tag{4.12}$$

Thus c_j is a linear combination of the vectors $\eta^k \Phi^{(k)}(\eta)$ for $k \leq j$ whose coefficients are *independent* of η. Moreover as $P_0 = 1$, we have $P_0^{(j)} = 0$ for $j \geq 1$. So in fact, for $j \geq 1$, c_j is spanned by the vectors $\eta^k \Phi^{(k)}(\eta)$ for $1 \leq k \leq j$. Then, by definition of K_1 and the vectors g_k,

$$K_1 C_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & * & * \\ \vdots & 0 & \ddots & * \\ 0 & \dots & 0 & 1 \end{pmatrix}. \tag{4.13}$$

Note that this matrix is independent of η. Let us denote it by $\mathcal{P}^{-1}$. Then the inverse of C_1 is $\mathcal{P}K_1$ and $\mathcal{P}$ has the same structure as $\mathcal{P}^{-1}$. Recall that for $k \geq N$, c_k is a linear combination of the vectors $\eta^j \Phi^{(j)}(\eta)$ for $j \leq k$. As g_j is the dual family of the vectors $(\eta^l \Phi^{(l)}(\eta))_{0\leq l\leq N-1}$ and $\|g_j\| = \mathcal{O}(\eta^{1-N})$, we have

$$C_1^{-1} C_2 = M + \mathcal{O}(\eta), \tag{4.14}$$

where M is a $N \times (d-N)$ matrix independent of η. □

Proof of Lemma 4.7. Let $G_1 \in \mathbb{R}^{N\times N}$ and $G_2 \in \mathbb{R}^{(d-N)\times N}$ be the unique matrices such that

$$G_\eta = \begin{pmatrix} G_1 \\ \hline G_2 \end{pmatrix} \in \mathbb{R}^{d\times N}.$$

By definition

$$G_\eta = \int_{-1}^{1} \rho(t) Q(t) \Phi(\eta t)^T \,\mathrm{d}t.$$

By Lemma 4.2 applied to each ϕ_j,

$$\Phi(\eta t) = \sum_{k=0}^{N-1} \left(t^{2j} - \eta \frac{Z_0}{2} \frac{|t|^{2j+1}}{2j+1} \right) \frac{\eta^{2j}}{(2j)!} \mathcal{E}^j \Phi(0) + \mathcal{O}(\eta^{2N}).$$

Let $a_j \in \mathbb{R}^d, a_j^N \in \mathbb{R}^N, a_j^{d-N} \in \mathbb{R}^{d-N}$ be defined by

$$a_j := -\int_{-1}^{1} \rho(t) \frac{Z_0}{2} \frac{|t|^{2j+1}}{2j+1} Q(t) \,\mathrm{d}t =: \begin{pmatrix} a_j^N \\ \hline a_j^{d-N} \end{pmatrix}.$$

Then by definition of the polynomials Q we have

$$\begin{cases} G_1 = \sum_{j=0}^{N-1} (e_j + \eta a_j^N) \frac{\eta^{2j}}{(2j)!} (\mathcal{E}^j \Phi(0))^T + \mathcal{O}(\eta^{2N}), \\ G_2 = \sum_{j=0}^{N-1} \eta a_j^{d-N} \frac{\eta^{2j}}{(2j)!} (\mathcal{E}^j \Phi(0))^T + \mathcal{O}(\eta^{2N}). \end{cases}$$

Let $(f_0, \dots, f_{N-1})$ be the dual basis of $\left(\mathcal{E}^j \Phi(0) \frac{\eta^{2j}}{(2j)!}\right)_{0\leq j\leq N-1}$ in $\mathbb{R}^N$ and H_1 be the matrix

$$H_1 := \begin{pmatrix} f_0 & \cdots & f_{N-1} \end{pmatrix} \in \mathbb{R}^{N\times N}.$$

It is straightforward to see that $\|H_1\| = \mathcal{O}(\eta^{2-2N})$ and

$$\begin{cases} G_1 H_1 = I_N + \mathcal{O}(\eta) \ , \\ G_2 H_1 = \mathcal{O}(\eta). \end{cases}$$

The matrix H_1 is invertible and its inverse is

$$H_1^{-1} = \sum_{j=0}^{N-1} e_j \frac{\eta^{2j}}{(2j)!} (\mathcal{E}^j \Phi(0))^T.$$

Let us now prove (4.10) for $j = 0$. We have

$$\begin{aligned} \Phi(0)^T (C_\eta G_\eta)^{-1} C_\eta &= e_0^T H_1^{-1} \left(C_\eta^{(P)} \Pi G_\eta \right)^{-1} C_\eta^{(P)} \Pi \\ &= e_0^T \left(C_1^{-1} C_\eta^{(P)} \Pi G_\eta H_1 \right)^{-1} \Pi \Big(I_N \ \Big| \ C_1^{-1} C_2 \Big) \\ &= e_0^T \left(\Big(I_N \ \Big| \ M + \mathcal{O}(\eta) \Big) \Pi \Big(I_N \ \Big| \ \mathcal{O}(\eta) \Big)^T \right)^{-1} \Big(I_N \ \Big| \ M + \mathcal{O}(\eta) \Big) \Pi. \end{aligned}$$

Decomposing Π into four blocks

$$\Pi = \begin{pmatrix} \Pi_1 & \Pi_2 \\ \Pi_3 & \Pi_4 \end{pmatrix}, \text{ with } \Pi_1 \in \mathbb{R}^{N \times N},$$

we obtain

$$\begin{aligned} \Phi(0)^T (C_\eta G_\eta)^{-1} C_\eta &= e_0^T \left(\Pi_1 + M\Pi_3 + \mathcal{O}(\eta) \right)^{-1} \Big(\Pi_1 + M\Pi_3 + \mathcal{O}(\eta) \ \Big| \ \Pi_2 + M\Pi_4 + \mathcal{O}(\eta) \Big) \\ &= e_0^T \Big(I_N \ \Big| \ M_\pi \Big) + \mathcal{O}(\eta). \end{aligned}$$

For $1 \leq j \leq N-1$ we proceed in the same way, using $e_j^T H_1^{-1} = \frac{\eta^{2j}}{(2j)!} (\mathcal{E}^j \Phi(0))^T$. □

Proof of Proposition 3.1. Let $0 \leq j \leq N-1$ and let $(c_k)_{1 \leq k \leq N}$ be as in Lemma 4.3. Then by Lemma 4.4 we have :

$$\begin{aligned} [\tilde{\psi}^{(2j+1)}]_0 &= \underbrace{(-E)^j \psi(0) - \sum_{k=1}^N c_k (-\epsilon_k)^j \phi_k(0)}_{=0} + \langle A^{-1} p \, , \psi - \sum_{k=1}^N c_k \phi_k \rangle \cdot \mathcal{E}^j \Phi(0) \\ &= \langle A^{-1} p \, , \psi - \sum_{k=1}^N c_k \phi_k \rangle \cdot \mathcal{E}^j \Phi(0) \\ &= \langle A^{-1} p \, , \psi_e - \sum_{k=1}^N c_k \phi_k \rangle \cdot \mathcal{E}^j \Phi(0) \ , \end{aligned}$$

as p is even.

Combining Lemmas 4.6 and 4.7, we get

$$(\mathcal{E}^j \Phi(0))^T A^{-1} p = \rho(t/\eta) (\mathcal{E}^j \Phi(0))^T (C_\eta G_\eta)^{-1} C_\eta Q(t/\eta) = \rho(t/\eta) x_\eta^T Q(t/\eta) \ ,$$

where $\|x_\eta\| = \mathcal{O}(\eta^{-2j})$.

Using again Lemma 4.3, we obtain

$$\begin{aligned} \left| \langle A^{-1} p \, , \psi - \sum_{k=1}^N c_k \phi_k \rangle \cdot \mathcal{E}^j \Phi(0) \right| &\leq C \|x_\eta\| \left\| \rho(t) \sum_{\ell=0}^{N-1} |Q_\ell(t)| \right\|_{L^1[-1,1]} \left\| \psi - \sum_{k=1}^N c_k \phi_k \right\|_{L^\infty[-\eta,\eta]} \\ &\leq C \eta^{2N-2j}. \end{aligned}$$

We therefore obtain

$$|[\tilde{\psi}^{(2j+1)}]_0| \leq C\eta^{2N-2j}.$$

For $j \geq N$, we then have

$$(-E)^j\psi(0) - \sum_{k=1}^{N} c_k(-\epsilon_k)^j\phi_k(0) \neq 0,$$

and by Lemmas 4.3 and 4.7,

$$\left|\langle A^{-1}p\,, \psi - \sum_{j=1}^{N} c_j\phi_j\rangle \cdot \mathcal{E}^j\Phi(0)\right| \leq C \underbrace{\|(C_\eta G_\eta)^{-1}C_\eta\|}_{=\mathcal{O}(\eta^{2-2N})} \underbrace{\left\|\psi - \sum_{k=1}^{N} c_k\phi_k\right\|_{L^\infty[-\eta,\eta]}}_{=\mathcal{O}(\eta^{2N})} \leq C\eta^2.$$

We therefore have

$$[\tilde{\psi}^{(2j+1)}]_0 = \underbrace{(-E)^j\psi(0) - \sum_{k=1}^{N} c_k(-\epsilon_k)^j\phi_k(0)}_{\neq 0} + \underbrace{\langle A^{-1}p\,, \psi - \sum_{k=1}^{N} c_k\phi_k\rangle \cdot \mathcal{E}^j\Phi(0)}_{=\mathcal{O}(\eta^2)}.$$

Thus,

$$|[\tilde{\psi}^{(2j+1)}]_0| \leq C,$$

which completes the proof. □

4.4 d-th derivative jump

We use the notation introduced in the previous section.

Proof of Proposition 3.2. We give the proof only for $k = d$ as the proof for $d+1 \leq k \leq 2d-2$ is very similar. By definition of $\tilde{\psi}$ and $\widetilde{\Phi}$,

$$\begin{aligned}
[\tilde{\psi}^{(d)}]_\eta &= \langle \tilde{p}\,, \tilde{\psi}\rangle \cdot [\widetilde{\Phi}^{(d)}]_\eta \\
&= \frac{1}{\eta^d}\langle \tilde{p}\,, \tilde{\psi}\rangle \cdot (C_\eta^{(P)}P^{(d)}(1) - \eta^d\Phi^{(d)}(\eta)) \\
&= \frac{1}{\eta^d}\langle A^{-1}p\,, \psi\rangle \cdot (C_\eta^{(P)}P^{(d)}(1) - \eta^d\Phi^{(d)}(\eta)) \qquad \text{(by Equation (4.8))}, \\
&= \frac{1}{\eta^d}\int_{-1}^{1} \rho(t)\psi(\eta t)Q(t)\,\mathrm{d}t \cdot C_\eta^T(G_\eta^T C_\eta^T)^{-1}(C_\eta^{(P)}P^{(d)}(1) - \eta^d\Phi^{(d)}(\eta)) \text{ (by Lemma 4.6).}
\end{aligned}$$

We know from (4.12) that the columns of $C_\eta^{(P)}$ are linear combinations of $\eta^k\Phi^{(k)}(\eta)$. Let us apply Lemma 4.2 to Φ. As the remainder is C^{2d-2}, for $k \leq d-1$, we can differentiate k times and we have for k even :

$$\eta^k\Phi^{(k)}(\eta) = \sum_{j=\frac{k}{2}}^{N-1}\left(\frac{\eta^{2j}}{(2j-k)!} - \frac{Z_0}{2}\frac{\eta^{2j+1}}{(2j+1-k)!}\right)\mathcal{E}^j\Phi(0) + \mathcal{O}(\eta^{2N}),$$

and for k odd, we have

$$\eta^k\Phi^{(k)}(\eta) = -\frac{Z_0}{2}\eta^k D^{\frac{k-1}{2}}\Phi(0) + \sum_{j=\frac{k+1}{2}}^{N-1}\left(\frac{\eta^{2j}}{(2j-k)!} - \frac{Z_0}{2}\frac{\eta^{2j+1}}{(2j+1-k)!}\right)\mathcal{E}^j\Phi(0) + \mathcal{O}(\eta^{2N}).$$

But by Lemma 4.7, for $0 \leq k \leq 2N-2$, k even, we have :

$$\begin{aligned} C_\eta^T (G_\eta^T C_\eta^T)^{-1} \eta^k \Phi^{(k)}(\eta) &= \sum_{j=\frac{k}{2}}^{N-1} \left(\frac{\eta^{2j}}{(2j-k)!} - \frac{Z_0}{2} \frac{\eta^{2j+1}}{(2j+1-k)!} \right) C_\eta^T (G_\eta^T C_\eta^T)^{-1} \mathcal{E}^j \Phi(0) + \mathcal{O}(\eta^2) \\ &= \sum_{j=\frac{k}{2}}^{N-1} \frac{(2j)!}{(2j-k)!} \begin{pmatrix} I_N \\ M_\pi^T \end{pmatrix} e_j + \mathcal{O}(\eta). \end{aligned} \tag{4.15}$$

Similarly for $0 \leq k \leq 2N-1$, k odd :

$$\begin{aligned} C_\eta^T (G_\eta^T C_\eta^T)^{-1} \eta^k \Phi^{(k)}(\eta) &= -\frac{Z_0}{2} C_\eta^T (G_\eta^T C_\eta^T)^{-1} \eta^k D^{\frac{k-1}{2}} \Phi(0) \\ &\quad + \sum_{j=\frac{k+1}{2}}^{N-1} \left(\frac{\eta^{2j}}{(2j-k)!} - \frac{Z_0}{2} \frac{\eta^{2j+1}}{(2j+1-k)!} \right) C_\eta^T (G_\eta^T C_\eta^T)^{-1} \mathcal{E}^j \Phi(0) + \mathcal{O}(\eta^2) \\ &= \sum_{j=\frac{k+1}{2}}^{N-1} \frac{(2j)!}{(2j-k)!} \begin{pmatrix} I_N \\ M_\pi^T \end{pmatrix} e_j + \mathcal{O}(\eta) \ , \end{aligned} \tag{4.16}$$

and for $k \geq 2N$, using $\|C_\eta^T (G_\eta^T C_\eta^T)^{-1}\| = \mathcal{O}(\eta^{2-2N})$ we have

$$C_\eta^T (G_\eta^T C_\eta^T)^{-1} \eta^k \Phi^{(k)}(\eta) = \mathcal{O}(\eta).$$

We have proved that $[\tilde{\psi}^{(d)}]_\eta = \mathcal{O}(\eta^{-d})$ but it is possible to have a slightly better estimate.

Observing that ψ is in fact a Lipschitz function and not only a continuous function, we have for $|t| \leq 1$:

$$\psi(\eta t) = \psi(0) + \mathcal{O}(\eta).$$

By definition of the polynomials Q_k, we have

$$\begin{aligned} \int_{-1}^{1} \rho(t) Q(t) \psi(\eta t) \, \mathrm{d}t &= \psi(0) \int_{-1}^{1} \rho(t) Q(t) \, \mathrm{d}t + \mathcal{O}(\eta) \\ &= \psi(0) e_0 + \mathcal{O}(\eta). \end{aligned}$$

To complete the proof of the proposition, it remains to show

$$e_0 \cdot C_\eta^T (G_\eta^T C_\eta^T)^{-1} (C_\eta^{(P)} P^{(d)}(1) - \eta^d \Phi^{(d)}(\eta)) = \mathcal{O}(\eta).$$

As we have for $j \geq 1$

$$e_0^T \begin{pmatrix} I_N \\ M_\pi^T \end{pmatrix} e_j = 0,$$

then for $d \geq 2$, equations (4.15) and (4.16) lead to

$$e_0 \cdot C_\eta^T (G_\eta^T C_\eta^T)^{-1} \eta^d \Phi^{(d)}(\eta) = \mathcal{O}(\eta).$$

Recall that the columns of $C_\eta^{(P)}$ satisfy the relation

$$\forall 0 \leq j \leq d-1, \ c_j = \eta^j \Phi^{(j)}(\eta) - \sum_{k=0}^{j-1} P_k^{(j)}(1) c_k.$$

But $P_0^{(j)}(1) = 0$ for $j \geq 1$ so in fact, for all $k \geq 1$, c_k is a linear combination of the vectors $\eta^j \Phi^{(j)}(\eta)$ for $1 \leq j \leq k$. Moreover by definition, we have

$$P^{(d)}(1) = (\underbrace{0, \ldots, 0}_{\lfloor \frac{d}{2} \rfloor}, *, \ldots, *)^T,$$

so $C_\eta^{(P)} P^{(d)}(1)$ is a linear combination of the last $\lceil \frac{d}{2} \rceil$ columns of $C_\eta^{(P)}$. Thus $C_\eta^{(P)} P^{(d)}(1)$ is a linear combination of the vectors $\eta^j \Phi^{(j)}(\eta)$ for $1 \leq j \leq d-1$ and therefore in view of (4.15) and (4.16), we have

$$e_0^T C_\eta^T (G_\eta^T C_\eta^T)^{-1} C_\eta^{(P)} P^{(d)}(1) = \mathcal{O}(\eta).$$

□

Proof of Theorem 3.3. First, we need to bound the remainder $\int_0^1 \tilde{\psi}^{(2d-1)}(x) e^{-2i\pi mx} \, \mathrm{d}x$ with respect to η. $\widetilde{\Phi}$ is a vector of polynomials of degree at most $2d-2$, thus $\widetilde{\Phi}^{(2d-1)} = 0$. Thus

$$\begin{aligned}\left| \int_0^1 \tilde{\psi}^{(2d-1)}(x) e^{-2i\pi mx} \, \mathrm{d}x \right| \leq \int_0^1 |\psi^{(2d-1)}(x)| \, \mathrm{d}x + \int_{-\eta}^{\eta} |\langle \tilde{p}\,, \tilde{\psi} \rangle \cdot \Phi^{(2d-1)}(x)| \, \mathrm{d}x \\ + \int_{a-\eta}^{a+\eta} |\langle \tilde{p}^a\,, \tilde{\psi} \rangle \cdot \Phi^{(2d-1)}(x-a)| \, \mathrm{d}x.\end{aligned}$$

We have $\langle \tilde{p}\,, \tilde{\psi} \rangle = \langle A^{-1} p\,, \psi \rangle$ by (4.8) and by Lemmas 4.6 and 4.7,

$$|\langle A^{-1} p\,, \psi \rangle| \leq \frac{C}{\eta^{2N-2}},$$

where C is a positive constant independent of η. Thus

$$\int_{-\eta}^{\eta} |\langle \tilde{p}\,, \tilde{\psi} \rangle \cdot \Phi^{(2d-1)}(x)| \, \mathrm{d}x \leq \frac{C}{\eta^{2N-3}}.$$

Then by Proposition 4.1, using the estimates (3.1) and (3.2) on the derivative jumps

$$\begin{aligned}|\widehat{\tilde{\psi}}_m| &\leq C \left(\sum_{j=0}^{\lfloor \frac{d}{2} \rfloor - 1} \frac{1}{(2\pi m)^{2+2j}} \left| \left[\tilde{\psi}^{(2j+1)} \right]_0 \right| + \sum_{k=d}^{2d-2} \frac{1}{(2\pi m)^{k+1}} \left| \left[\tilde{\psi}^{(k)} \right]_{\pm\eta} \right| \right. \\ &\quad + \sum_{j=0}^{\lfloor \frac{d}{2} \rfloor - 1} \frac{1}{(2\pi m)^{2+2j}} \left| \left[\tilde{\psi}^{(2j+1)} \right]_0 \right| + \sum_{k=d}^{2d-2} \frac{1}{(2\pi m)^{k+1}} \left| \left[\tilde{\psi}^{(k)} \right]_{a\pm\eta} \right| \\ &\quad \left. + \frac{1}{(2\pi m)^{2d-1}} \left| \int_0^1 \tilde{\psi}^{(2d-1)}(x) e^{-2i\pi mx} \, \mathrm{d}x \right| \right) \\ &\leq C \left(\sum_{j=0}^{\lfloor \frac{d}{2} \rfloor - 1} \frac{\eta^{2N-2j}}{m^{2+2j}} + \sum_{k=d}^{2d-2} \frac{1}{\eta^{k-1} m^{k+1}} + \frac{1}{\eta^{2N-3} m^{2d-1}} \right).\end{aligned}$$

Since $N \leq d$ and $m \geq \frac{1}{\eta}$, we have the result. □

4.5 Error bound on the eigenvalues

To derive the estimate on the eigenvalues, we would like to use the following classical result ([25], p. 68).

Proposition 4.11. *Let H be a self-adjoint coercive H^1-bounded operator, $E_1 \leq \cdots \leq E_n$ be the lowest eigenvalues of H and $\psi_1, \ldots, \psi_n$ be L^2-normalized associated eigenfunctions. Let $E_1^{(M)} \leq \cdots \leq E_n^{(M)}$ be the lowest eigenvalues of the Rayleigh quotient of H restricted to the subspace V_M of dimension M.*

Let $w_k \in V_M$ for $1 \leq k \leq n$ be such that

$$\sum_{k=1}^{n} \|w_k - \psi_k\|_{H^1}^2 < 1.$$

Then there exists a positive constant C which depends on the H^1 norm of H and the coercivity constant such that for all $1 \leq k \leq n$

$$\left|E_k^{(M)} - E_k\right| \leq C \sum_{k=1}^{n} \|w_k - \psi_k\|_{H^1}^2.$$

We would like to apply this result to $\psi_M = (\mathrm{Id} + T)\tilde{\psi}_M$ where $\tilde{\psi}_M$ is the truncation of $\tilde{\psi}$ to the first M plane-waves but we need to bound the H^1 norm of T with respect to η. Coercivity for our one-dimensional model has been proved in [5]. To find this bound, we will need to rewrite Tf in a convenient way.

Lemma 4.12. *For $f \in H^1_{\mathrm{per}}(0,1)$, we have for $|x| \leq \eta$:*

$$\begin{aligned} Tf(x) &= \langle \tilde{p}\,, f\rangle \cdot \left(\Phi(x) - \widetilde{\Phi}(x)\right) \\ &= \left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \int_{-1}^{1} \rho(t) f(\eta t) P(t)\,\mathrm{d}t \cdot \left(\begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x) - P(x/\eta)\right), \end{aligned}$$

where $G(P)$ is the following Gram matrix :

$$G(P) = \int_{-1}^{1} \rho(t) P(t) P(t)^T \,\mathrm{d}t,$$

and $C_1 \in \mathbb{R}^{N\times N}$ is the square matrix defined in Lemma 4.9.

Proof. For $|x| \leq \eta$, we have :

$$\begin{aligned} (Tf)(x) &= \langle \tilde{p}\,, f\rangle \cdot (\Phi(x) - \widetilde{\Phi}(x)) \\ &= \langle B^{-1}p\,, f\rangle \cdot (\Phi(x) - \widetilde{\Phi}(x)). \end{aligned}$$

Recall that

$$\begin{aligned} B &= \langle p\,, \widetilde{\Phi}^T\rangle \\ &= \langle \rho(t/\eta)\widetilde{\Phi}\,, \widetilde{\Phi}^T\rangle \\ &= \eta \int_{-1}^{1} \rho(t) C_\eta^{(P)} P(t) P(t)^T \left(C_\eta^{(P)}\right)^T \,\mathrm{d}t \\ &= \eta C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T. \end{aligned}$$

Thus,

$$\begin{aligned} \langle B^{-1}p\,, f\rangle \cdot \widetilde{\Phi}(x) &= \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \int_{-1}^{1} \rho(t) f(\eta t) P(t)\,\mathrm{d}t \cdot C_\eta^{(P)} P(x/\eta) \\ &= \left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \int_{-1}^{1} \rho(t) f(\eta t) P(t)\,\mathrm{d}t \cdot P(x/\eta). \end{aligned}$$

Using the identity

$$C_\eta^{(P)} \begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} = \begin{pmatrix} C_1 \mid C_2 \end{pmatrix} \begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} = I_N,$$

we can formally rewrite $\langle B^{-1}p\,, f\rangle$ as

$$\langle B^{-1}p\,, f\rangle \cdot \Phi(x) = \left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \int_{-1}^{1} \rho(t) f(\eta t) P(t)\,\mathrm{d}t \cdot \begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x), \tag{4.17}$$

and the result follows. □

Lemma 4.13. *There exists a positive constant C independent of η such that for all $f \in H^1_{\mathrm{per}}(0,1)$ and for all $x \in \mathbb{R}$, we have*

$$\forall\, 0 < \eta \leq \eta_0,\ \left|\langle \tilde{p}\,, f\rangle \cdot \left(\Phi(x) - \widetilde{\Phi}(x)\right)\right| \leq C\eta \|f\|_{H^1_{\mathrm{per}}}.$$

Proof. In this proof, C denotes a generic constant that does not depend on η or f. Let $f \in H^1_{\mathrm{per}}(0,1)$. On $|x| \geq \eta$, we have by definition of $\widetilde{\Phi}$

$$\langle \tilde{p}\,, f\rangle \cdot \left(\Phi(x) - \widetilde{\Phi}(x)\right) = 0.$$

We deduce from Lemma 4.12 that for $|x| \leq \eta$ we have

$$\langle \tilde{p}\,, f\rangle \cdot \left(\Phi(x) - \widetilde{\Phi}(x)\right) = \left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \int_{-1}^{1} \rho(t) f(\eta t) P(t)\,\mathrm{d}t \cdot \left(\begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x) - P(x/\eta)\right).$$

The proof of the lemma consists of four steps. We will successively show that

1. $\begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x) - P(x/\eta) = \begin{pmatrix} 0 \\ * \end{pmatrix} + \mathcal{O}(\eta)$, where $\begin{pmatrix} 0 \\ * \end{pmatrix}$ is uniformly bounded in η and x;
2. $\int_{-1}^{1} \rho(t) f(\eta t) P(t)\,\mathrm{d}t = f(0) G(P) e_0 + \mathcal{O}(\eta)\|f\|_{H^1_{\mathrm{per}}}$;
3. the norm of $\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)}$ is uniformly bounded in η;
4. for $j \geq 1$, $e_j^T \left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} G(P) e_0$ is of order $\mathcal{O}(\eta)$.

Indeed, assuming these statements hold, we can infer from statement 2 that

$$\begin{aligned}
&\left|\langle \tilde{p}\,, f\rangle \cdot \left(\Phi(x) - \widetilde{\Phi}(x)\right)\right| \\
&\quad \leq |f(0)| \left|\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} G(P) e_0 \cdot \left(\begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x) - P(x/\eta)\right)\right| \\
&\qquad + \mathcal{O}(\eta)\|f\|_{H^1_{\mathrm{per}}} \left|\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \left(\begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x) - P(x/\eta)\right)\right|.
\end{aligned}$$

We treat both terms separately. For the second term, by statements 1 and 3, we have

$$\left|\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} \left(\begin{pmatrix} C_1^{-1} \\ 0 \end{pmatrix} \Phi(x) - P(x/\eta)\right)\right| \leq C.$$

For the first one, by statement 1, we only have to check that for $j \geq 1$, we have

$$\left|\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} G(P) e_0 \cdot e_j\right| \leq C\eta,$$

which is exactly statement 4. The lemma is then proved using the Sobolev embedding $\|f\|_{L^\infty_{\mathrm{per}}} \leq C\|f\|_{H^1_{\mathrm{per}}}$.

Step 1 Writing down the Taylor expansion of Φ at η, we obtain

$$\begin{aligned}\Phi(x) &= \sum_{k=0}^{N-1} \frac{(x-\eta)^k}{k!}\Phi^{(k)}(\eta) + \mathcal{O}((x-\eta)^N) \\ &= \sum_{k=0}^{N-1} \frac{1}{k!}\left(\frac{x}{\eta}-1\right)^k \eta^k \Phi^{(k)}(\eta) + \mathcal{O}((x-\eta)^N).\end{aligned}$$

By Lemma 4.9, we have

$$C_1^{-1}\eta^k\Phi^{(k)}(\eta) = \mathcal{P}K_1\eta^k\Phi^{(k)}(\eta) = \mathcal{P}e_k,$$

and for $k \neq 0$, $\mathcal{P}e_k \cdot e_0 = 0$. We also know that $\|C_1^{-1}\| = \mathcal{O}(\eta^{1-N})$, so that

$$C_1^{-1}\Phi(x) = \begin{pmatrix}1\\ *\end{pmatrix} + \mathcal{O}(\eta).$$

By definition $P_0 = 1$, and therefore

$$\left(\frac{C_1^{-1}}{0}\right)\Phi(x) - P(x/\eta) = \begin{pmatrix}0\\ *\end{pmatrix} + \mathcal{O}(\eta).$$

Step 2 Since $f \in H^1_{\mathrm{per}}(0,1)$, by the Sobolev embedding theorem, f is continuous and $f(0)$ exists. Thus we can write

$$\begin{aligned}\int_{-1}^1 \rho(t)f(\eta t)P(t)\,\mathrm{d}t &= f(0)\int_{-1}^1 \rho(t)P(t)\,\mathrm{d}t + \int_{-1}^1 \rho(t)(f(\eta t)-f(0))P(t)\,\mathrm{d}t \\ &= f(0)G(P)e_0 + \int_{-1}^1 \rho(t)(f(\eta t)-f(0))P(t)\,\mathrm{d}t,\end{aligned}$$

since $P_0 = 1$. Using

$$f(\eta t) = f(0) + \int_0^{\eta t} f'(x)\,\mathrm{d}x,$$

and Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\left|\int_{-1}^1 \rho(t)(f(\eta t)-f(0))P(t)\,\mathrm{d}t\right| &\leq \left(\int_{-1}^1 \rho(t)^2P^2(t)\,\mathrm{d}t\int_{-1}^1\left(\int_0^{\eta t} f'(x)\,\mathrm{d}x\right)^2\,\mathrm{d}t\right)^{1/2} \\ &\leq C\left(\int_{-1}^1\left(\int_0^{\eta t} f'(x)^2\,\mathrm{d}x\right)\eta^2t^2\,\mathrm{d}t\right)^{1/2} \\ &\leq C\eta\|f\|_{H^1_{\mathrm{per}}}.\end{aligned}$$

Step 3 We want to bound the norm of the matrix

$$\left(C_\eta^{(P)}\right)^T\left(C_\eta^{(P)}G(P)\left(C_\eta^{(P)}\right)^T\right)^{-1}C_\eta^{(P)}.$$

Since $G(P)$ is the Gram matrix of the polynomials P_k for the weight ρ, $G(P)$ is a symmetric positive definite matrix and thus admits a square root. It is easy to check that

$$G(P)^{1/2}\left(C_\eta^{(P)}\right)^T\left(C_\eta^{(P)}G(P)\left(C_\eta^{(P)}\right)^T\right)^{-1}C_\eta^{(P)}G(P)^{1/2}$$

is an orthogonal projector. Its norm is therefore uniformly bounded in η.

Step 4 Let $G_1 \in \mathbb{R}^{N\times N}$, $G_2 \in \mathbb{R}^{N\times(d-N)}$ and $G_3 \in \mathbb{R}^{(d-N)\times(d-N)}$ be the matrices respectively defined by

$$G(P) = \begin{pmatrix} G_1 & G_2 \\ G_2^T & G_3 \end{pmatrix}.$$

Let M_η be the matrix

$$C_1^{-1}C_2 = M_\eta.$$

Recall that by Lemma 4.9, $\|M_\eta\| \leq C$. With this notation, we have

$$C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T = C_1 \left(G_1 + M_\eta G_2^T + G_2 M_\eta^T + M_\eta G_3 M_\eta^T\right),$$

and therefore,

$$\begin{aligned}\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} e_0 \\ = \begin{pmatrix} I_N \\ M_\eta \end{pmatrix} \left(G_1 + M_\eta G_2^T + G_2 M_\eta^T + M_\eta G_3 M_\eta^T\right)^{-1} \left(G_1 + M_\eta G_2^T\right) e_0. \quad (4.18)\end{aligned}$$

We will now show that $M_\eta^T e_0 = \mathcal{O}(\eta)$. By definition,

$$e_0^T M_\eta = e_0^T C_1^{-1} C_2 = e_0^T \mathcal{P} K_1 C_2.$$

By Lemma 4.9, $e_0^T \mathcal{P} = e_0$ and by definition of K_1, $e_0^T K_1 = g_0^T$ where g_0 is the vector satisfying $g_0^T \eta^k \Phi^{(k)}(\eta) = \delta_{0k}$ for $k \leq N-1$. Again by Lemma 4.9, the columns of $C_\eta^{(P)}$ satisfy :

$$\forall 0 \leq j \leq d-1,\ c_j = \eta^j \Phi^{(j)}(\eta) - \sum_{k=0}^{j-1} P_k^{(j)}(1) c_k,$$

with $P_0^{(j)}(1) = 0$ for $j \geq 1$. Consequently c_j is a linear combination of the vectors $\eta^k \Phi^{(k)}(\eta)$ for $1 \leq k \leq j$. Since $\|g_0\| = \mathcal{O}(\eta^{1-N})$, we get $g_0^T C_2 = \mathcal{O}(\eta)$. Coming back to Equation (4.18), we have

$$\left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} e_0 = \begin{pmatrix} I_N \\ M_\eta^T \end{pmatrix} e_0 + \mathcal{O}(\eta).$$

Consequently,

$$\begin{aligned} e_j^T \left(C_\eta^{(P)}\right)^T \left(C_\eta^{(P)} G(P) \left(C_\eta^{(P)}\right)^T\right)^{-1} C_\eta^{(P)} e_0 &= e_j^T \begin{pmatrix} I_N \\ M_\eta^T \end{pmatrix} e_0 + \mathcal{O}(\eta) \\ &= \begin{cases} \mathcal{O}(\eta), & 0 \leq j \leq N-1, \\ e_{j-N}^T M_\eta^T e_0 = \mathcal{O}(\eta), & N \leq j \leq d-1. \end{cases}\end{aligned}$$

□

We can establish a similar result for $\langle \tilde{p}\,, f\rangle \cdot \left(\Phi'(x) - \widetilde{\Phi}'(x)\right)$.

Lemma 4.14. *There exists a positive constant C independent of η such that for all $f \in H^1_{\mathrm{per}}(0,1)$ and for all $x \in \mathbb{R}$, we have*

$$\forall\, 0 < \eta \leq \eta_0,\ \left|\langle \tilde{p}\,, f\rangle \cdot \left(\Phi'(x) - \widetilde{\Phi}'(x)\right)\right| \leq C\|f\|_{H^1_{\mathrm{per}}},$$

Proof. It is a transposition of the proof of the previous lemma. The first step is simply replaced by

1. $\begin{pmatrix} C_1^{-1} \\ \hline 0 \end{pmatrix} \Phi'(x) - \frac{1}{\eta} P'(x/\eta) = \frac{1}{\eta} \begin{pmatrix} 0 \\ * \end{pmatrix} + \mathcal{O}(1)$, where $\begin{pmatrix} 0 \\ * \end{pmatrix}$ is uniformly bounded in η and x.

To prove the latter statement, we observe that $P_0' = 0$ and by a Taylor expansion of Φ', we obtain

$$\Phi'(x) = \frac{1}{\eta} \sum_{k=1}^{N-1} \frac{1}{(k-1)!} \left(\frac{x}{\eta} - 1\right)^{k-1} \eta^k \Phi^{(k)}(\eta) + \mathcal{O}(\eta^{N-1}),$$

hence

$$C_1^{-1}\Phi'(x) = \frac{1}{\eta} \begin{pmatrix} 0 \\ * \end{pmatrix} + \mathcal{O}(1).$$

□

Lemma 4.15. *There exists a positive constant C independent of η such that for all $f \in H^1_{\mathrm{per}}(0,1)$, we have*

$$\forall\, 0 < \eta \leq \eta_0,\ \|Tf\|_{H^1_{\mathrm{per}}} \leq C\eta^{\frac{1}{2}} \|f\|_{H^1_{\mathrm{per}}}$$

Proof. This is a straightforward consequence of Lemmas 4.13 and 4.14. □

Proof of Theorem 3.4. Applying Proposition 4.11 to $\psi_M = (\mathrm{Id} + T)\Pi_M\tilde{\psi}$, where Π_M is the truncation to the first M plane-waves, we have

$$\begin{aligned} |E_M^\eta - E| &\leq C\|(\mathrm{Id} + T)(\Pi_M\tilde{\psi} - \tilde{\psi})\|^2_{H^1_{\mathrm{per}}} \\ &\leq C(\|\Pi_M\tilde{\psi} - \tilde{\psi}\|^2_{H^1_{\mathrm{per}}} + \|T(\Pi_M\tilde{\psi} - \tilde{\psi})\|^2_{H^1_{\mathrm{per}}}). \end{aligned}$$

By Lemma 4.15,

$$\|T(\Pi_M\tilde{\psi} - \tilde{\psi})\|^2_{H^1_{\mathrm{per}}} \leq C\|\Pi_M\tilde{\psi} - \tilde{\psi}\|^2_{H^1_{\mathrm{per}}}\ ,$$

and we deduce from Theorem 3.3 that

$$\begin{aligned} |E_M^\eta - E| &\leq C\|\Pi_M\tilde{\psi} - \tilde{\psi}\|^2_{H^1_{\mathrm{per}}} \\ &\leq C \sum_{j=M+1}^{\infty} (1 + j^2)|\widehat{\tilde{\psi}}_j|^2 \\ &\leq C \sum_{j=M+1}^{\infty} \left(\frac{\eta^{4N}}{j^2} + \frac{1}{\eta^{2d-2} j^{2d}}\right) \\ &\leq C \left(\frac{\eta^{4N}}{M} + \frac{1}{\eta^{2d-2} M^{2d-1}}\right). \end{aligned}$$

□

5 Perturbation by a continuous potential

In standard electronic structure calculations, the Hartree and exchange-correlation terms are modelled by a potential that is smoother than the Coulomb potential. To reproduce this setting in our one-dimensional toy model, a smoother potential W is added to the Hamiltonian (2.1). In the following, we examine how the VPAW method accelerates the computation of eigenvalues.

Consider the Hamiltonian

$$H = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} - Z_0 \sum_{k\in\mathbb{Z}} \delta_k - Z_a \sum_{k\in\mathbb{Z}} \delta_{a+k} + W\ , \tag{5.1}$$

where W is 1-periodic, continuous, $0 < a < 1$, $Z_0, Z_a > 0$.

With the VPAW method, the generalized eigenvalue problem

$$(\mathrm{Id} + T^*)H(\mathrm{Id} + T)\tilde{\psi} = E(\mathrm{Id} + T^*)(\mathrm{Id} + T)\tilde{\psi}, \tag{5.2}$$

is solved by expanding $\tilde{\psi}$ in plane-waves. Like in Section 2.2, $T = T_0 + T_a$, where T_0 and T_a act on two disjoint regions $\bigcup_{k\in\mathbb{Z}} [-\eta + k, \eta + k]$ and $\bigcup_{k\in\mathbb{Z}} [a - \eta + k, a + \eta + k]$ respectively. The atomic wave functions $(\phi_k)_{0\leq k\leq N-1}$ are the non-smooth solutions of the atomic Hamiltonian

$$H_0 = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} - Z_0 \sum_{k\in\mathbb{Z}} \delta_k + V \ ,$$

where V can be different from W. The eigenvalues associated to $(\phi_k)_{0\leq k\leq N-1}$ are denoted by ϵ_k. To define the pseudo wave functions $(\tilde{\phi}_k)_{0\leq k\leq N-1}$ and the projectors $(\tilde{p}_k)_{0\leq k\leq N-1}$, we proceed as in Section 2.2.

It follows from the study of the double Dirac delta potential Hamiltonian that the key lemma of the analysis is the structure Lemma 4.2, which describes the behavior of eigenfunctions near the singularities. It is possible to establish a similar result for the eigenfunctions of the Hamiltonian (5.1).

Lemma 5.1. *Let ψ be an eigenfunction of the Hamiltonian H given by* (5.1) *for the eigenvalue E. Then in a neighborhood of* 0*, we have the following expansion :*

$$\psi(x) = \psi(0)\left(\sum_{j=0}^{k} \frac{(-E)^j}{(2j)!}x^{2j} - \frac{Z_0}{2}\frac{(-E)^j}{(2j+1)!}|x|^{2j+1}\right) + \frac{\psi'(0_+) + \psi'(0_-)}{2}\sum_{j=0}^{k}\frac{(-E)^j}{(2j+1)!}x^{2j+1} + \sum_{j=0}^{k}\underbrace{\int\cdots\int}_{2j+2} W\psi(x) + \psi_{2k+2}(x),$$

where ψ_{2k+2} is a C^{2k+2} function satisfying in a neighbourhood of 0

$$\begin{cases} \psi_{2k+2}^{(2k+2)} = (-E)^{k+1}\psi, \\ \psi_{2k+2}(x) = \mathcal{O}(x^{2k+2}). \end{cases}$$

Proof. This lemma can be proved by induction. For $k = 0$, we set

$$\theta_2(x) = \psi(x) + \frac{Z_0}{2}|x|\psi(0) - \int_0^x\int_0^t W(s)\psi(s)\mathrm{d}s\mathrm{d}t$$

and then proceed as in the proof of Lemma 4.2. □

We will now make some assumptions on the potentials V and W:

1. V and W are smooth and 1-periodic;
2. V is even. This property would indeed be satisfied by potentials that does not break the crystal symmetry;

Lemma 5.2. *For $N \geq 2$:*

$$\iint V\Phi(x) = \sum_{k=1}^{N-1}(V\Phi)_k\left(\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\frac{|x|^{2k+1}}{(2k+1)!}\right) + \mathcal{O}\left(x^{2N}\right),$$

where $(V\Phi)_k$ is in $\mathrm{span}(\mathcal{E}^j\Phi(0),\ j \leq N-2)$.

Proof. This lemma is proved by induction.

Initialization Applying Lemma 5.1 with $Z_a = 0$ and $W = V$ to each function ϕ_k, we obtain expansions of the atomic PAW functions ϕ_k in the vicinity of 0:

$$\Phi(x) = \Phi(0)\left(1 - \frac{Z_0}{2}|x|\right) + \mathcal{O}(x^2).$$

Deriving twice $\iint V\Phi$ gives

$$\begin{aligned}\left(\iint V\Phi(x)\right)'' &= V(x)\Phi(x)\\ &= V(0)\Phi(0)\left(1 - \frac{Z_0}{2}|x|\right) + \mathcal{O}(x^2).\end{aligned}$$

Therefore

$$\iint V\Phi(x) = V(0)\Phi(0)\left(\frac{x^2}{2} - \frac{Z_0}{2}\frac{|x|^3}{6}\right) + \mathcal{O}(x^4).$$

Inductive step Let us derive twice $\iint V\Phi$:

$$\begin{aligned}\left(\iint V\Phi(x)\right)'' &= V(x)\Phi(x)\\ &= \left(\sum_{k=0}^{2N-2} V^{(2k)}(0)\frac{x^{2k}}{(2k)!}\right)\left(\sum_{k=0}^{N-1}\left(\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\frac{|x|^{2k+1}}{(2k+1)!}\right)D^k\Phi(0)\right.\\ &\quad\left. + \sum_{j=0}^{N-2}\underbrace{\int\cdots\int}_{(2j+2)} V\Phi(x)\right) + \mathcal{O}(x^{2N}).\end{aligned} \tag{5.3}$$

By the induction hypothesis,

$$\underbrace{\int\cdots\int}_{(2j+2)} V\Phi(x) = \sum_{k=j+1}^{N-1}(V\Phi)_{k-j}\left(\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\frac{|x|^{2k+1}}{(2k+1)!}\right) + \mathcal{O}(x^{2N}).$$

Thus,

$$\begin{aligned}\sum_{j=0}^{N-1}\underbrace{\int\cdots\int}_{(2j+2)} V\Phi(x) &= \sum_{j=0}^{N-2}\sum_{k=j+1}^{N-1}(V\Phi)_{k-j}\left(\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\frac{|x|^{2k+1}}{(2k+1)!}\right)\\ &= \sum_{k=1}^{N-1}\sum_{j=0}^{k-1}(V\Phi)_{k-j}\left(\frac{x^{2k}}{(2k)!} - \frac{Z_0}{2}\frac{|x|^{2k+1}}{(2k+1)!}\right).\end{aligned}$$

Going back to (5.3), expanding the equation and using the last equation, we obtain the result. □

Lemma 5.3. *In a neighbourhood of* 0*, the vector* Φ *has the following expansion :*

$$\Phi(x) = \sum_{j=0}^{k}\left(\frac{x^{2j}}{(2j)!} - \frac{Z_0}{2}\frac{|x|^{2j+1}}{(2j+1)!}\right)X_j + \Phi_{2k+2}(x)\ ,$$

where the function Φ_{2k+2} *is* C^{2k+2} *at 0 and* X_j *are vectors satisfying*

$$\begin{cases} X_j \in \operatorname{span}(\mathcal{E}^\ell\Phi(0),\ \ell \le j),\\ X_j - \mathcal{E}^j\Phi(0) \in \operatorname{span}(\mathcal{E}^\ell\Phi(0),\ \ell \le j-1),\end{cases} \tag{5.4}$$

where $\mathcal{E}$ *is the diagonal matrix with entries* $-\epsilon_0, \dots, -\epsilon_{N-1}$.

Proof. We apply Lemmas 5.1 and 5.2 and notice that the vectors $(V\Phi)_k$ are spanned by $(\mathcal{E}^j\Phi(0),\ j \le k-1)$. □

Lemma 5.4. *The even part of ψ satisfies*

$$\psi_e(x) = \sum_{j=0}^{1}\left(\frac{(-E)^j}{(2j)!}x^{2j} - \frac{Z_0}{2}\frac{(-E)^j}{(2j+1)!}|x|^{2j+1}\right) + W(0)\psi(0)\left(\frac{x^2}{2} - \frac{Z_0}{2}\frac{|x|^3}{3!}\right)$$
$$+\left(W'(0)\psi_s'(0) - \frac{E}{2}W(0)\psi(0)\right)\frac{x^4}{4!} - \frac{Z_0}{2}\left(\frac{W''(0)}{2}\psi(0) - EW(0)\psi(0)\right)\frac{|x|^5}{5!}$$
$$+W(0)\psi(0)\left(\frac{x^4}{4!} - \frac{Z_0}{2}\frac{x^5}{5!}\right) + \mathcal{O}\left(x^6\right),$$

where

$$\psi_s'(0) = \frac{\psi'(0_+) + \psi'(0_-)}{2}.$$

Proof. The proof follows from Lemma 5.1 and a careful estimation of the terms $\iint W\psi$ and $\iiint W\psi$. □

Since W is not even, ψ does not have the same structure as for the double delta potential. More precisely, we can show that because of the term $\iint W\psi$ the singularity of the fifth order term cannot be removed by the VPAW approach.

Lemma 5.5. *For $N = 2$ there exist coefficients c_0 and c_1 such that:*

$$\psi_e(x) - c_0\phi_0(x) - c_1\phi_1(x) = \mathcal{O}\left(x^4\right).$$

For $N \ge 3$, there exists a family of coefficients $(c_k)_{0\le k\le N-1}$ such that:

$$\psi_e(x) - \sum_{k=0}^{N-1} c_k\phi_k(x) = \mathcal{O}\left(x^5\right).$$

Following the same steps as in Section 4, we can establish the following theorems.

Theorem 5.6 (Estimates on the Fourier coefficients)**.** *Let $N \in \mathbb{N}^*$ and $d \ge N$. Let $\widehat{\tilde{\psi}}_m$ be the m-th Fourier coefficient of $\tilde{\psi}$. Then, there exists a positive constant C such that for all $0 < \eta \le \eta_0$ and $m \ge \frac{1}{\eta}$*

$$\left|\widehat{\tilde{\psi}}_m\right| \le C\left(\frac{\eta^{2N\wedge 5}}{m^2} + \frac{1}{\eta^{d-1}m^{d+1}}\right),$$

where $a \wedge b = \min(a, b)$.

Theorem 5.7 (Estimates on the eigenvalues)**.** *Let $N \in \mathbb{N}^*$ and $d \ge N$. Let E_M^η be an eigenvalue of the variational approximation of (5.2) in a basis of M plane-waves and for a cut-off radius $0 < \eta \le \eta_0$, and let E be the corresponding exact eigenvalue. There exists a constant $C > 0$ independent of η and M such that for all $0 < \eta \le \eta_0$ and $M \ge \frac{1}{\eta}$*

$$|E_M^\eta - E| \le C\left(\frac{\eta^{4N\wedge 10}}{M} + \frac{1}{\eta^{2d-2}}\frac{1}{M^{2d-1}}\right). \tag{5.5}$$

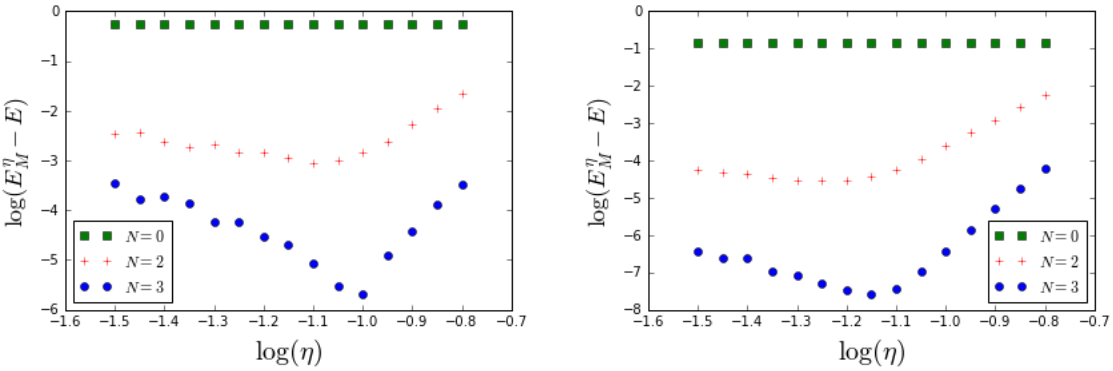
Remark 5.8. *The estimate (5.5) does not seem optimal as shown in Figure 10. It seems that singularities of any order can be removed by the VPAW method.*

6 Numerical tests

The goal of this section is to compare the theoretical estimates determined in Sections 3 and 4 to numerical simulations and show that they are optimal.

All numerical simulations are carried out with $Z_0 = Z_a = 10$ and $a = 0.4$. The Fourier coefficients are evaluated by a very accurate numerical integration.

It is interesting to compare the results (Figure 1) obtained by a direct expansion of the wave function ψ (here displayed by the points $N = 0$) and the VPAW method. Recall that N is the number of pseudo wave functions used to build the operator T. The smoothness of the pseudo wave functions is set to $d = N$.



(a) Error on the 8-th eigenvalue for $M = 64$ basis functions

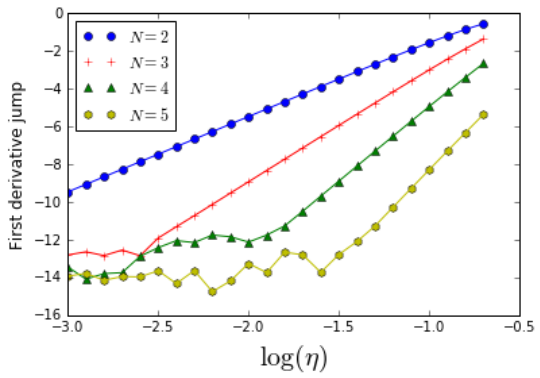
(b) Error on the 8-th eigenvalue for $M = 256$ basis functions

Figure 1: The VPAW method compared to a direct calculation for the 8-th eigenvalue

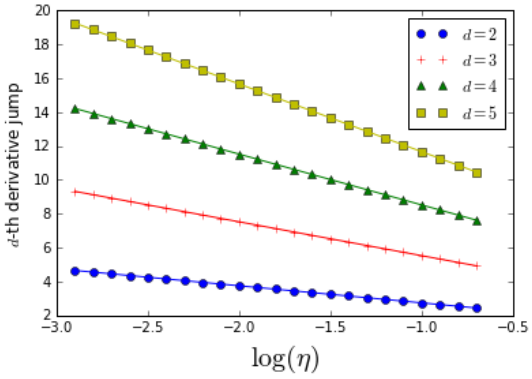
Given a number M of basis functions, the VPAW method is much more accurate than the direct method although it is quite sensitive to the choice of η. More comments on this behavior will be made in Section 6.3. We do not report the computing times for the VPAW method because in this study, each time a simulation is run, we generate all the pseudo wave functions $\tilde{\phi}$, the projector functions $\tilde{p}$ and compute their Fourier coefficients. In practice, these data are precomputed and stored in a file. Thus, the only additional cost compared to the direct method comes from the assembly of the matrices $(\mathrm{Id} + T)^* H (\mathrm{Id} + T)$ and $(\mathrm{Id} + T)^* (\mathrm{Id} + T)$.

6.1 Derivative jumps

Since ψ and the functions ϕ_i are known analytically, it is possible to evaluate the derivative jumps of $\tilde{\psi}$ at 0 and $\pm\eta$ (Figures 2 and 3). The plots are given for the eigenfunction associated to the lowest eigenvalue of H. The behavior is similar for other eigenfunctions.



(a) First derivative jump at 0 as a function of η in log-log scale ($d = N$).



(b) d-th derivative jump at $\pm\eta$ as a function of η in log-log scale ($N = 2$).

Figure 2: Derivative jumps of the pseudo wave function $\tilde{\psi}$

N	Numerics	Theory
2	3.90	4
3	5.94	6
4	7.85	8
5	9.85	10

(a) Numerical and theoretical slopes for the first derivative jump at 0.

d	Numerics	Theory
2	-1.005	-1
3	-2.000	-2
4	-3.000	-3
5	-4.000	-4

(b) Numerical and theoretical slopes for the d-th derivative jump at $\pm\eta$.

Figure 3: Comparison of the theoretical and numerical results for the derivative jumps

These numerical results are in remarkable agreement with Propositions 3.1 and 3.2.

6.2 Comparison of the PAW and VPAW methods in pre-asymptotic regime

The simulations are run for a fixed value of $d = 6$ and two different values of η ($\eta = 0.1$ and $\eta = 0.2$). In Figure 4, E is the lowest eigenvalue of the 1D-Schrödinger operator H given by (2.1).

Recall that our theoretical estimate on the eigenvalue given by the VPAW method is :

$$|E_M^\eta - E| \leq C\left(\frac{\eta^{4N}}{M} + \frac{1}{\eta^{2d-2}}\frac{1}{M^{2d-1}}\right). \tag{6.1}$$

To transpose the PAW method to our one-dimensional setting, we need to account for the use of a pseudo-potential. For this purpose, we replace the Dirac delta potential by some smooth function in Equation (2.1). We choose the 1-periodic function χ_ϵ such that

$$\chi_\epsilon(x) = \begin{cases} \frac{C}{\epsilon}\exp\left(-\frac{1}{1-\left(\frac{x}{\epsilon}\right)^2}\right), & x \in [-\epsilon, \epsilon], \\ 0, & x \in [-1/2, 1/2] \setminus [-\epsilon, \epsilon]. \end{cases}$$

where C ensures that $\int_{-\epsilon}^{\epsilon} \chi_\epsilon = 1$. As ϵ goes to 0, χ_ϵ converges to the 1-periodized Dirac potential in $H^{-1}_{\text{per}}(0,1)$.

As expected, the PAW method quickly converges to a wrong value of E. It is interesting to notice that asymptotically, the VPAW convergence is of order $\mathcal{O}\left(\frac{1}{M}\right)$ but for small enough values of M and η, the second term in the RHS of (6.1) dominates.

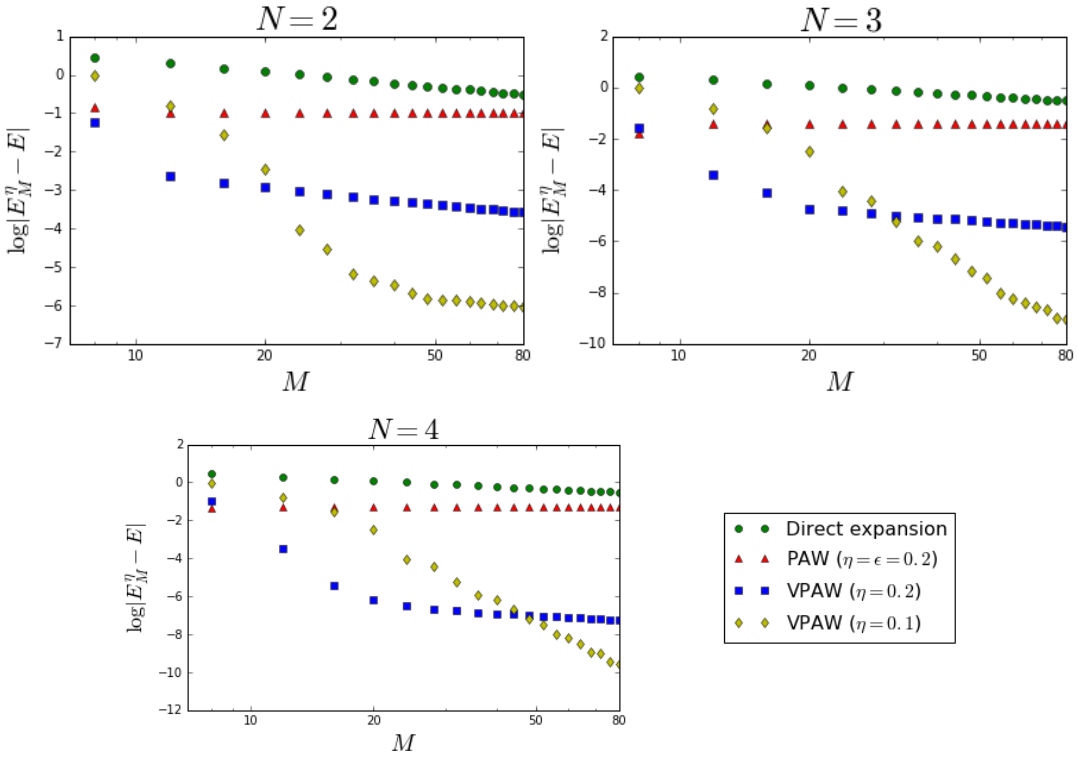
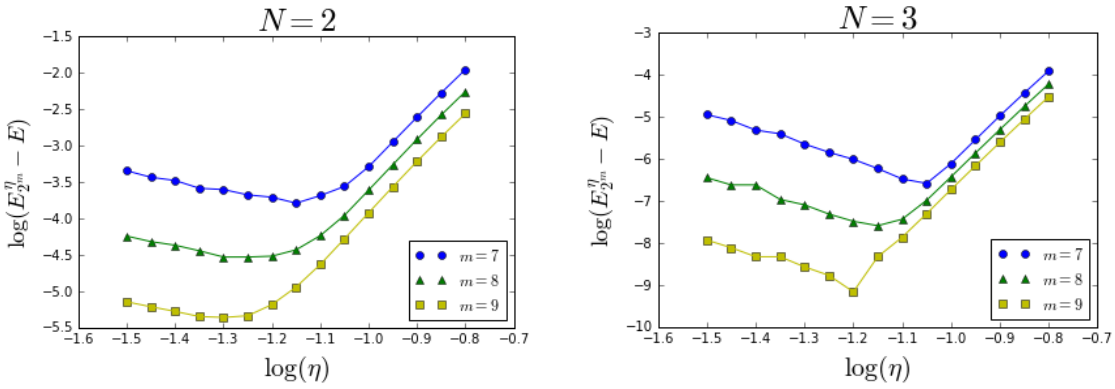


Figure 4: Comparison between the PAW and VPAW methods for the lowest eigenvalue

6.3 Asymptotic regime

6.3.1 Behavior in the plane-wave cut-off M

The next numerical tests (Figures 5 and 6) are run with $d = N$ and $N = 2$, $N = 3$. The pseudo wave function $\tilde{\psi}$ is expanded in $M = 2^m$ plane-waves, $m = 7$ to 9.



(a) Error on the 9-th eigenvalue with $N = 2$ (b) Error on the 9-th eigenvalue with $N = 3$

Figure 5: Error on the eigenvalue for different values of M

Here, we can clearly see two regimes : for η small (resp. η large), the leading term in the error is dominated by the d-th derivative jumps at $k \pm \eta$ and $k + a \pm \eta$, $k \in \mathbb{Z}$ (resp. the first derivative jump at k and $k + a$, $k \in \mathbb{Z}$). In each regime, the gaps between the decreasing and increasing slopes seem constant and their evaluation gives the correct orders of convergence in M (see Figure 6).

	Numerics	Theory
Decreasing lines	0.30	$\log(2) \simeq 0.30$
Increasing lines	0.90	$3\log(2) \simeq 0.90$

(a) Gaps for $N = 2$

	Numerics	Theory
Decreasing lines	0.32	$\log(2) \simeq 0.30$
Increasing lines	1.50	$5\log(2) \simeq 1.50$

(b) Gaps for $N = 3$

Figure 6: Estimation of the order of convergence in M

6.3.2 Dependence of the convergence rate in η on N and d

In each graph of Figure 7, we have kept M constant to track the dependence of the convergence rate in η. By Theorem 3.4, the logarithm of error on the eigenvalue is given by

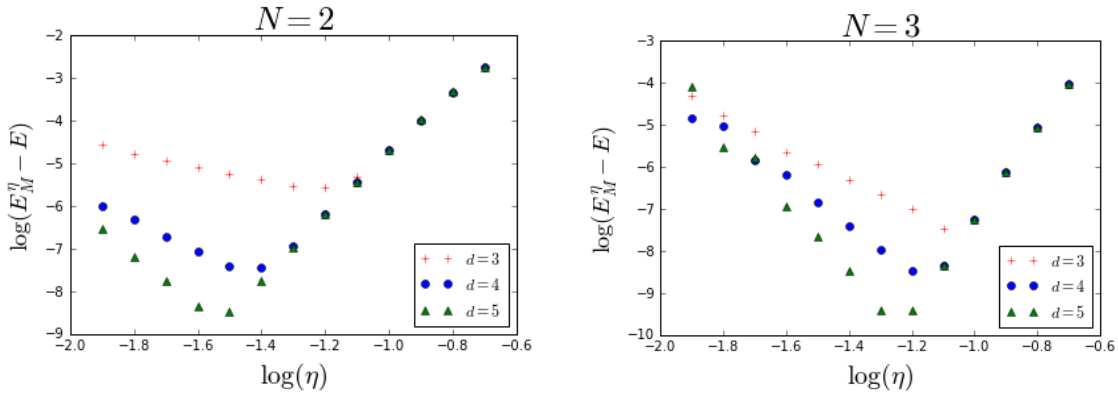
$$\log(E_M^\eta - E) = \log(C) + \log\left(\frac{\eta^{4N}}{M} + \frac{1}{\eta^{2d-2}M^{2d-1}}\right).$$

Hence, when η is large, we have

$$\log(E_M^\eta - E) \simeq \log(C) + 4N\log\eta - \log(M),$$

and when η is small, we have

$$\log(E_M^\eta - E) \simeq \log(C) - (2d-2)\log\eta - (2d-1)\log M.$$



(a) Error on the 8-th eigenvalue with $N = 2$, $M = 256$

(b) Error on the 8-th eigenvalue with $N = 3$, $M = 128$

Figure 7: Error on the eigenvalue for different values of d

Notice that in each graph, for η large, the parameter d has a negligible effect on the error on the eigenvalues, in agreement with our theoretical estimates.

d	Numerics	Theory
2	6.5	8
3	6.9	8
4	7.2	8

(a) $N = 2$

d	Numerics	Theory
3	10.6	12
4	10.7	12
5	10.9	12

(b) $N = 3$

Figure 8: Estimation of the increasing slopes in Figure 7

There is a small discrepancy between the theoretical and numerical values of the increasing slope. A possible explanation could be that the estimates we have given for the first derivative

d	Numerics	Theory
2	-1.6	-2
3	-3.6	-4
4	-6.0	-6

(a) $N = 2$

d	Numerics	Theory
3	-3.8	-4
4	-5.4	-6
5	-7.8	-8

(b) $N = 3$

Figure 9: Estimation of the decreasing slopes in Figure 7

jumps are valid asymptotically as η goes to 0, but the increasing slopes are observed for relatively large values of η.

For the decreasing slopes, our estimate is in very good agreement with the numerical simulations.

6.4 Perturbation by a continuous potential

In this subsection, we study the VPAW method applied to the Hamiltonian (5.1) with $W(x) = 10\sin(2\pi x + 0.2)$. Since this model is not exactly solvable, we use a P2 finite elements method to compute very accurately the eigenvalues (the relative error on the computed eigenvalue is less than 10^{-10}).

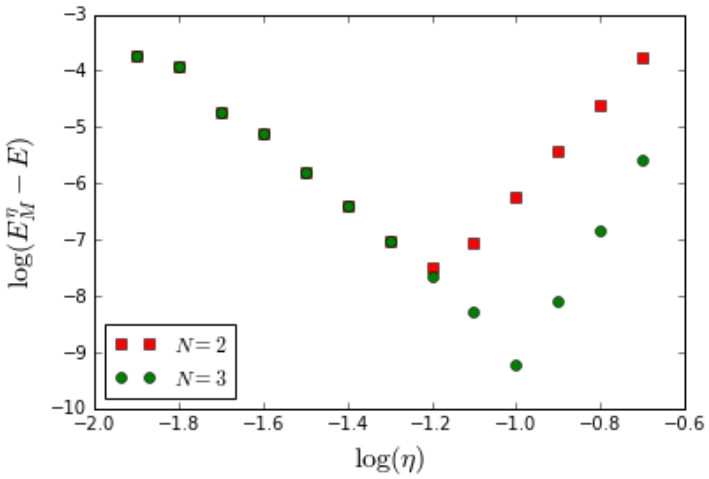


Figure 10: Error on the first eigenvalue as a function of η ($M = 128, d = 4$)

N	Numerics	Theory
2	8.2	8
3	12	10

(a) Increasing slopes

	Numerics	Theory
$N = 2$ and $N = 3$	-5.7	-6

(b) Decreasing slopes

Figure 11: Estimation of the slopes in Figure 10

For $N = 3$, the increasing part of the curve has a slope which is very close to the theoretical estimation of Theorem 3.4 (that is with $W = 0$). This seems to indicate that the VPAW method removes the singularity at the nucleus up to the fifth order, but we are unable to support this observation with rigorous numerical analysis arguments.

References

[1] C. Audouze, F. Jollet, M. Torrent, and X. Gonze, *Projector augmented-wave approach to density-functional perturbation theory*, Physical Review B, 73 (2006), p. 235101.

[2] X. Blanc, E. Cancès, and M.-S. Dupuy, in preparation.

[3] ——, *Variational projector augmented-wave method*, Comptes Rendus Mathematique, 355 (2017), pp. 665 – 670.

[4] P. E. Blochl, *Projector augmented-wave method*, Phys. Rev. B, 50 (1994), pp. 17953–17979.

[5] E. Cancès and G. Dusson, *Discretization error cancellation in electronic structure calculation: a quantitative study*, accepted at ESAIM: Mathematical Modelling and Numerical Analysis, (2017).

[6] E. Cances and N. Mourad, *Existence of a type of optimal norm-conserving pseudopotentials for kohn–sham models*, Communications in Mathematical Sciences, 14 (2016), pp. 1315–1352.

[7] 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 (2015), pp. 755–785.

[8] M. Dolg and X. Cao, *Relativistic pseudopotentials: their development and scope of applications*, Chemical reviews, 112 (2011), pp. 403–480.

[9] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Ø. Sørensen, *Sharp regularity results for coulombic many-electron wave functions*, Communications in mathematical physics, 255 (2005), pp. 183–227.

[10] S. Goedecker, M. Teter, and J. Hutter, *Separable dual-space gaussian pseudopotentials*, Physical Review B, 54 (1996), p. 1703.

[11] M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Ø. Sørensen, *Electron wavefunctions and densities for atoms*, in Annales Henri Poincaré, vol. 2, Springer, 2001, pp. 77–100.

[12] F. Jollet, M. Torrent, and N. Holzwarth, *Generation of projector augmented-wave atomic data: A 71 element validated table in the XML format*, Computer Physics Communications, 185 (2014), pp. 1246–1254.

[13] T. Kato, *On the eigenfunctions of many-particle systems in quantum mechanics*, Communications on Pure and Applied Mathematics, 10 (1957), pp. 151–177.

[14] L. Kleinman and D. Bylander, *Efficacious form for model pseudopotentials*, Physical Review Letters, 48 (1982), p. 1425.

[15] D. Koelling and G. Arbman, *Use of energy derivative of the radial solution in an augmented plane wave method: application to copper*, Journal of Physics F: Metal Physics, 5 (1975), p. 2041.

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

[17] G. Kresse and D. Joubert, *From ultrasoft pseudopotentials to the projector augmented-wave method*, Phys. Rev. B, 59 (1999), pp. 1758–1775.

[18] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt, *Car-Parrinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials*, Phys. Rev. B, 47 (1993), pp. 10142–10153.

[19] A. Levitt and M. Torrent, *Parallel eigensolvers in plane-wave density functional theory*, Computer Physics Communications, 187 (2015), pp. 98–105.

[20] C. Rostgaard, *The projector augmented-wave method*, arXiv preprint arXiv:0910.1921, (2009).

[21] D. J. Singh and L. Nordstrom, *Planewaves, Pseudopotentials, and the LAPW method*, Springer Science & Business Media, 2006.

[22] M. Torrent, F. Jollet, F. Bottin, G. Zérah;, and X. Gonze, *Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure*, Computational Materials Science, 42 (2008), pp. 337 – 351.

[23] N. Troullier and J. L. Martins, *Efficient pseudopotentials for plane-wave calculations*, Physical review B, 43 (1991), p. 1993.

[24] D. Vanderbilt, *Soft self-consistent pseudopotentials in a generalized eigenvalue formalism*, Phys. Rev. B, 41 (1990), pp. 7892–7895.

[25] H. F. Weinberger, *Variational methods for eigenvalue approximation*, vol. 15, Siam, 1974.