Efficient iterative method for solving the Dirac–Kohn–Sham density functional theory

Lin Lin [a], Sihong Shao [b,*], Weinan E [c,d]

[a] Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[b] LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China
[c] Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544, USA
[d] Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China

ABSTRACT

We present for the first time an efficient iterative method to directly solve the four-component Dirac–Kohn–Sham (DKS) density functional theory. Due to the existence of the negative energy continuum in the DKS operator, the existing iterative techniques for solving the Kohn–Sham systems cannot be efficiently applied to solve the DKS systems. The key component of our method is a novel filtering step (F) which acts as a preconditioner in the framework of the locally optimal block preconditioned conjugate gradient (LOBPCG) method. The resulting method, dubbed the LOBPCG-F method, is able to compute the desired eigenvalues and eigenvectors in the positive energy band without computing any state in the negative energy band. The LOBPCG-F method introduces mild extra cost compared to the standard LOBPCG method and can be easily implemented. We demonstrate our method in the pseudopotential framework with a planewave basis set which naturally satisfies the kinetic balance prescription. Numerical results for Pt_2, Au_2, TlF, and Bi_2Se_3 indicate that the LOBPCG-F method is a robust and efficient method for investigating the relativistic effect in systems containing heavy elements.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The electron, as an elementary particle, has spin and charge, and further acquires an angular momentum quantum number corresponding to a quantized atomic orbital when binded to the atomic nucleus. The importance of the spin–orbit coupling (SOC) effect in semiconductors and other materials have been extensively explored in quantum physics and quantum chemistry. For example, SOC causes shifts in the atomic energy level of an electron [1], and leads to protected metallic surface or edge states as a consequence of the topology of the bulk electronic wave functions [2]. As a typical relativistic effect, SOC has magnitude of the order $Z^4\alpha^{-2}$ for a hydrogen like atom [3], where Z is the nuclear charge, $\alpha = c^{-1}$ in the atomic unit is the fine structure constant and c is the speed of light. For systems containing heavy elements with large Z, the nonrelativistic Schrödinger type equations such as the Kohn–Sham (KS) density functional theory [4,5] leads to large error [6,7]. The extension of the density functional theory is not straightforward as quantum electrodynamics has to been used for charged particles in which complicated renormalization is necessary to get finite expressions for charge, energy, etc. [8]. The relativistic density functional theory, first laid out by Rajagopal and Callaway [9], can be rigorously derived from quantum electrodynamics. However, the resulting equations are too complicated to solve in practice and proper renormalization has to be performed to eliminate the divergent terms. The frequently-used working equations are the four-component

* Corresponding author. Tel.: +86 1062751801.
  E-mail addresses: linlin@lbl.gov (L. Lin), sihong@math.pku.edu.cn (S. Shao), weinan@math.princeton.edu (W. E).

Dirac–Kohn–Sham (DKS) equations derived by Rajagopal [10] and independently by MacDonald and Vosko [11] after making several physically reasonable approximations. The extension to the time-dependent scenario can be found in *e.g.* [12,13].

Mathematically, the DKS operator is fundamentally different from the KS operator. The spectrum of the DKS operator, sketched in Fig. 1, consists of the point spectrum (blue dots), the positive energy continuum (thick blue line from $+mc^2$ to $+\infty$) and the negative energy continuum (thin red line from $-\infty$ to $-mc^2$), and is therefore unbounded from blow [14]. Meanwhile the spectrum of the KS operator does not contain any negative energy and is bounded from below. Here m is the electron mass and we have $m = 1$ and $c \approx 137$ in atomic unit. When solving the DKS equations numerically, the DKS operator will be discretized by a basis set of finite size, and both the positive energy continuum and the negative energy continuum will be truncated before infinity is reached. Hereafter we refer to the set of eigenvalues of the discretized DKS operator in the point spectrum and the positive energy continuum as the *positive energy band*, and the set of eigenvalues of the discretized DKS operator in the negative energy continuum as the *negative energy band*. There are approximately equal number of eigenvalues in the positive and the negative energy bands separated by a large gap, which is also called the forbidden region with its width denoted by Δ. We denote by w the maximum of the widths of the positive and the negative energy band. The desired eigenvalues and eigenvectors in the positive energy band are usually contained in the point spectrum or in very rare cases cover the entire point spectrum and a very few low lying states in the positive energy continuum.

Attempts to solve the DKS equations are plagued by the so-called "variational collapse" [15], which specifically means two difficulties. The first difficulty is the so-called spectral pollution – the appearance of spurious eigenvalues which are the limiting points of the eigenvalues of the discretized DKS operator as the basis set becomes complete, but are not the eigenvalues of the true DKS operator. They often lie deeper than the desired solutions or may even be degenerate with them. The second difficulty is the occurrence of the spurious eigenvalues in the forbidden region when solving the discretized DKS operator using inappropriate numerical schemes. The reason for such collapse is that the spectrum of the DKS operator cannot be defined variationally [14]. Several prescriptions to avoid variational collapse were summarized and analyzed by Kutzelnigg [16] as well as Lewin and Séré [17]. Among all the prescriptions, the kinetic balance prescription [18–20] addresses the first difficulty most successfully from practical point of view, and serves as the foundation for most successful attempts to solve the DKS equations via finite dimensional matrix eigenvalue problems. Many effective numerical implementations for performing four-component DKS calculations with the localized basis sets (*e.g.* Gaussian type orbitals and atomic orbitals) have been presented by quantum chemists in the last four decades, including the DVM scheme [21,22], the BDF package [23], the program of Fricke and coworkers [24], the REL4D module in Utchem [25], the DKS module in DIRAC [26] and the BERTHA code [27]. In all these methods, the finite dimensional DKS matrix is diagonalized directly as a dense matrix, which computes all the eigenvalues and eigenvectors in both positive and the negative energy bands. The direct diagonalization methods avoid the second difficulty in the variational collapse originating from inappropriate numerical schemes. However, these methods are prohibitively expensive when the dimension of the discretized DKS operator becomes large.

While the localized basis sets are widely used in the community of quantum chemistry, the planewave basis set is more popular in the community of condensed matter physics for the reason that systematic convergence can be achieved by only increasing the kinetic energy cutoff. The planewave basis set can be used as the only basis set in the pseudopotential framework [28], and as an augmented basis set such as in the linearized augmented planewave method (LAPW) [29]. In terms of the DKS density functional theory, the planewave basis set automatically satisfies the kinetic balance prescription and is free from the spectral pollution [17]. However, compared to the localized basis set, the planewave basis set leads to matrix eigenvalue problems of much larger size which is impractical to be diagonalized directly as dense matrices. Iterative methods have to be designed to solve the matrix eigenvalue problems for the desired eigenvalues and eigenvectors contained in the positive energy band. There are approximately the same number of eigenvalues in the positive and the negative energy bands, and standard iterative techniques for solving the KS systems cannot be efficiently applied without necessary modification. For example, the conjugate gradient method [30,31], the locally optimal block preconditioned conjugate gradient (LOBPCG) method [32,33], the RMM-DIIS method [34] and the Lanczos method [35] need to evaluate all the states in the negative energy band and therefore is prohibitively expensive; It is difficult to design a set of efficient filtering polynomials as in the Chebyshev filtering method [36], since the separation between the occupied and unoccupied eigenvalues in the positive energy band (usually in the order of 0.1 au or smaller) is much smaller compared to the width of the forbidden region ($\Delta \approx 37000$ au). The spectral radius of the matrix resulting from the spectrum folding technique [37] will be in the order of $\Delta^2 \approx 10^9$ for the DKS equations, and the resulting positive definite eigenvalue problem cannot be solved efficiently with iterative methods.

In this work, an efficient iterative method to directly solve the four-component DKS density functional theory will be presented for the first time. The key component of our method is a filtering step (F) which can act as a preconditioner in

Fig. 1. A sketch of the spectrum of the DKS operator. For explanations see the text.

preconditioned iterative diagonalization methods. In particular, our method is demonstrated based on the LOBPCG method, and is therefore dubbed the LOBPCG-F method. Compared to other types of preconditioned conjugate gradient methods, the LOBPCG method has been shown to be effective for evaluating a relatively large number of eigenvalues and eigenvectors, and its efficiency has been illustrated in large scale electronic structure calculation such as in ABINIT [38,39]. However, for the DKS systems, the standard LOBPCG method amplifies the error of the eigenfunctions projected to the negative energy band in each iteration, and therefore needs to evaluate all the eigenfunctions in the negative energy band together with the desired eigenfunctions in the positive energy band. We illustrate that such error can be efficiently controlled by the filtering step, and the desired eigenvalues and eigenvectors in the positive energy band can be evaluated without computing any negative energy state. The filtering step only introduces 2 extra matrix–vector multiplication per iteration, and can be implemented simply as a preconditioner with little coding effort.

We implement the LOBPCG-F method to solve the DKS equations in the pseudopotential framework with a planewave basis set (module DKS). To benchmark our numerical results, we also implement the standard LOBPCG method for solving the KS density functional theory (module KS), and for solving the two-component KS density functional theory with SOC in the pseudopotential framework (module KSSO). We directly compare the total energies obtained from the modules KS and KSSO with those obtained from the corresponding modules in a popular electronic structure software ABINIT [39]. We apply the modules KS, KSSO and DKS to solve systems including Pt_2, Au_2, TlF, and Bi_2Se_3. The differences in the total free energy are less than 1 meV per atom compared to ABINIT for all systems under study. Despite that the relativistic correction is relatively small for valence-valence interaction, our numerical results for solving the DKS equations indicate that the LOBPCG-F method is robust and efficient in studying the relativistic effect in systems containing heavy elements.

All the discussion in the rest of the manuscript will be presented in the pseudopotential framework in the context of orthogonal basis set (the planewave basis set is orthogonal). We remark that the LOBPCG-F method can be used for all-electron calculation and for non-orthogonal basis functions as well. This is the case for the LAPW method such as implemented in the WIEN2k software [40], and also for the localized basis set given that the overlap matrix is not very ill-conditioned.

The paper is organized as follows. We briefly review the four-component DKS density functional theory in Section 2 and discuss in detail the variational collapse and the choice of basis set. We develop the LOBPCG-F algorithm in Section 3. The numerical results are presented with discussion in Section 4. The paper is concluded in Section 5. Throughout the paper, the atomic units ($e = \hbar = 1$) will be used unless otherwise noted.

2. DKS density functional theory

We will now describe the main working equations – the DKS equations derived by Rajagopal [10] and independently by MacDonald and Vosko [11], and refer to Engel [8] and van Wüllen [41] for a elaborate and general discussion on relativistic density functional theory which can be rigorously derived from quantum electrodynamics. Several related numerical issues are then discussed, such as the variational collapse and the choice of basis in solving the DKS matrix eigenvalue problems.

2.1. DKS equations

Under the no-pair and electrostatic approximations [8,41], the total energy of an interacting n-electron system corresponding to the Dirac-Coulomb Hamiltonian takes the form

$$E[\rho, \mathbf{m}] = T[\rho, \mathbf{m}] + \int V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 + E_{\text{xc}}[\rho, \mathbf{m}], \tag{1}$$

where $\rho(\mathbf{r})$ is the electron density, $\mathbf{m}(\mathbf{r})$ is the spin density, and $V_{\text{ext}}(\mathbf{r})$ is the nuclear attractive potential. The first term T is the kinetic energy of the noninteracting reference system suggested by Kohn and Sham [5], the third term is the electrostatic energy, and E_{xc} is the exchange-correlation functional containing both the difference between the true many body electron–electron repulsive energy and the electrostatic energy, and the difference between the true many body kinetic energy and the kinetic energy of the noninteracting reference system. This factious system, which could be represented by a single Slater determinant in terms of one-electron spinors $\{\psi_i\}$ corresponding to the energy levels $\{\epsilon_i\}$, has the same electron density as the interacting many body system. In consequence, we could use $\{\psi_i\}$ to evaluate the kinetic energy and the densities

$$T = \sum_{i}^{occ} \psi_i(\mathbf{r})^\dagger h_{\text{D}} \psi_i(\mathbf{r}), \quad h_{\text{D}} = \begin{pmatrix} mc^2\mathbf{I}_2 & c\boldsymbol{\sigma} \cdot \mathbf{p} \\ c\boldsymbol{\sigma} \cdot \mathbf{p} & -mc^2\mathbf{I}_2 \end{pmatrix}, \tag{2}$$

$$\rho(\mathbf{r}) = \sum_{i}^{occ} \psi_i(\mathbf{r})^\dagger \psi_i(\mathbf{r}), \quad \mathbf{m}(\mathbf{r}) = \sum_{i}^{occ} \psi_i(\mathbf{r})^\dagger \begin{pmatrix} \boldsymbol{\sigma} & \mathbf{0}_2 \\ \mathbf{0}_2 & -\boldsymbol{\sigma} \end{pmatrix} \psi_i(\mathbf{r}), \tag{3}$$

where h_{D} is the Dirac operator corresponding to free particles, $\mathbf{p} = -i\nabla$ is the momentum operator, $\mathbf{I}_n$ and $\mathbf{0}_n$ are the $n \times n$ unit and null matrices, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of the Pauli spin matrices, and the each spinor ψ_i is a complex vector-valued function: $\mathbb{R}^3 \to \mathbb{C}^4$, which are often rewritten as $\psi_i = (\phi_i, \chi_i)^T$ with ϕ_i, χ_i being functions: $\mathbb{R}^3 \to \mathbb{C}^2$. In what follows

we will often refer to the two-spinor ϕ_i (resp. χ_i) as the large (resp. small) component of the four-spinor ψ_i. Here the summations are only restricted to the occupied states in the positive energy band.

The Euler–Lagrange equation with respect to $E[\rho, \mathbf{m}]$ gives rise to the DKS equation

$$\left[\begin{pmatrix} \boldsymbol{\sigma} \cdot \mathbf{B}_{xc} + mc^2 \mathbf{I}_2 & c\boldsymbol{\sigma} \cdot \mathbf{p} \\ c\boldsymbol{\sigma} \cdot \mathbf{p} & -\boldsymbol{\sigma} \cdot \mathbf{B}_{xc} - mc^2 \mathbf{I}_2 \end{pmatrix} + V_{hxc}\mathbf{I}_4 + V_{ext}\mathbf{I}_4 \right] \begin{pmatrix} \phi_i \\ \chi_i \end{pmatrix} = \begin{pmatrix} \phi_i \\ \chi_i \end{pmatrix} \epsilon_i, \tag{4}$$

where

$$V_{hxc}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}_1)}{|\mathbf{r} - \mathbf{r}_1|} d\mathbf{r}_1 + \frac{\delta E_{xc}}{\delta \rho}(\mathbf{r}), \tag{5}$$

$$\mathbf{B}_{xc}(\mathbf{r}) = \frac{\delta E_{xc}}{\delta \mathbf{m}}(\mathbf{r}). \tag{6}$$

Here for simplicity the exchange–correlation functional under local density approximation (LDA) [42,43] is used. Our discussion below is not restricted to the LDA approximation.

Ideally all-electron calculations should be performed to obtain accurate results, but with large computational cost in order to resolve the sharp gradient of core orbitals and the oscillation of valance orbitals in the neighborhood of nuclei. A more practical way striking the balance between accuracy and efficiency is to employ effective core potentials [44], in which the core electrons are frozen and the valance-only problem is solved. The pseudopotential technique is one branch of effective core potential approaches. One should be careful in choosing pseudopotential for a relativistic Hamiltonian such as DKS, since most of the existing pseudopotentials are generated for the nonrelativistic Hamiltonian. The mismatch between the pseudopotential and the relativistic Hamiltonian introduce possibly double counting of the relativistic effect [45]. We adopt here the widely known HGH pseudopotential [46] which includes the scalar relativistic effect and the spin–orbit coupling effect of the core electrons by construction, and can be specified by a very small number of parameters due to its dual-space Gaussian form. Rigorously speaking, the HGH pseudopotential may still suffer from double counting of the relativistic effect, since the parameters are optimized using the non-relativisitic equations. On the other hand, the numerical method developed in this paper does not depend on the specific choice of parameters of the HGH pseudopotential, and the numerical results can be readily improved when the HGH pseudopotential is reparametrized for the DKS calculation. The HGH pseudopotential consists of three parts

$$\mathbf{V}_{HGH}(\mathbf{r}, \mathbf{r}') = V_{loc}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')\mathbf{I}_4 + \sum_l \left(V_l(\mathbf{r}, \mathbf{r}')\mathbf{I}_4 + \Delta V_l^{SO}(\mathbf{r}, \mathbf{r}')\mathbf{L}' \cdot \mathbf{S} \right), \tag{7}$$

where the subscript l is the angular momentum number, V_{loc}, V_l, and ΔV_l^{SO} represent in the order the local contribution, the nonlocal contribution and the SOC effect of the HGH pseudopotential, of which the formulas can be found in [46] and thus are skipped here to save space. $\delta(\mathbf{r})$ is the Dirac delta function, $\mathbf{L}'$ is the angular momentum at position $\mathbf{r}'$, and $\mathbf{S} = \frac{1}{2} \begin{pmatrix} \boldsymbol{\sigma} & \mathbf{0}_2 \\ \mathbf{0}_2 & \boldsymbol{\sigma} \end{pmatrix}$ is the spin operator. The HGH pseudopotential V_{HGH} is an integral operator on the spinors. After denoting the first matrix operator of the LHS term of Eq. (4) by $\mathbf{H}_0$, we finally arrive the DKS equation for valence electrons

$$[\mathbf{H}_0 + V_{hxc}\mathbf{I}_4]\psi_i(\mathbf{r}) + \int \mathbf{V}_{HGH}(\mathbf{r}, \mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}' = \psi_i(\mathbf{r})\epsilon_i. \tag{8}$$

The nuclear attraction term V_{ext} in Eq. (4) is replaced here by the $\mathbf{V}_{HGH}$ term that describes the electrostatic interaction between the effective nuclei and valance electrons. The DKS Eqs. (4) or (8) are usually solved with the self-consistent field (SCF) iteration as in the case of the nonrelativistic KS equations [47].

2.2. Variational collapse and kinetic balance

As we have mentioned in Section 1, the most salient feature of the DKS operator is that its spectrum contains the negative energy continuum and is therefore unbounded from below. The desired eigenvalues and eigenvectors corresponding to the occupied bound states lie in the positive energy band (see Fig. 1), *i.e.* we are seeking for highly excited states above all the negative energy states. When solving the resulting matrix eigenvalue problems by expanding ψ_i in Eq. (8) with a finite basis set, such feature of the DKS operator imposes a large obstacle, *i.e.* the variational collapse called by Schwarz and Wallmeier [15], in the sense that the eigenvalues of the discretized DKS operator may not systematically converge to the eigenvalues of the true DKS operator as the basis size increases. This variational collapse, is imputed by Kutzelnigg [16] to the wrong nonrelativistic limit of the matrix representation of the Dirac operator and several prescriptions have been proposed to avoid it. The first strategy is to replace minimization procedures by min–max procedures [48–50]. The min–max methods find the minimum over the large component of the spinors and the maximum over the small component of the spinors for the DKS energy functional, and these methods are not used often practically in electronic structure calculation. The second strategy is the two-component relativistic theory which seeks for operators which are bounded from below and agrees with

the DKS operator in the nonrelaticistic limit [51]. A very good discussion of approaches made in this direction can be found in [52]. The third strategy is to carefully choose the basis set with which the discretized DKS operator is free of the spectral pollution. Among all the strategies, the kinetic balance method [18–20], falling into the third category, is widely used by theoretical chemists [52], and serves as the basis of most successful attempts to solve matrix eigenvalue equations based on finite-dimensional representations of the DKS operator. The feature of the kinetic balance is the one-to-one correspondence of the basis sets that are used to expand the large and small components, achieved via a linear operator $\boldsymbol{\sigma} \cdot \mathbf{p}$. Specifically, the basis $\{f_k, , k = 1, \ldots, N\}$ for expanding the small components are obtained directly from the basis $\{g_k, k = 1, \ldots, N\}$ for expanding the large components, according to the so-called kinetic balance prescription

$$f_k = \frac{1}{2mc} \boldsymbol{\sigma} \cdot \mathbf{p} g_k, \quad k = 1, \ldots, N. \tag{9}$$

The use of the kinetic balance avoids the worst aspects of variational collapse and could generate the desired eigenvalues and eigenvectors in the positive energy band with good accuracy if a sufficiently flexible basis set is employed [19,20,17,53]. The localized basis sets satisfying the kinetic balance prescription (9) are often used in quantum chemistry and all the eigenvalues and eigenvectors of the discretized DKS operator are obtained by the direct diagonalization method including the negative energy band that are not used in practice.

2.3. Choice of basis set: planewaves

Compared to the standard localized basis sets used in the relativistic quantum chemistry community, the planewave basis set has the advantage that systematic convergence can be achieved by tuning one parameter – the kinetic energy cutoff [47]. Furthermore, the planewave basis set automatically satisfies the kinetic balance prescription (9), *i.e.* the space spanned by planewaves within a certain cutoff is invariant when applied by the $\boldsymbol{\sigma} \cdot \mathbf{p}$ operator, and therefore the planewave basis set is a good basis set for both the large component and the small component. Namely, the planewave basis set can avoid the spectral pollution in practice. However, the planewave basis set usually results in a much larger degrees of freedom per atom (in the order of $500 \sim 5000$ or more for standard norm conserving pseudopotential [28] with only valence electrons taken into consideration), which is much larger than the degrees of freedom per atom used by localized basis set. We remark that the degrees of freedom used by the finite difference method and the finite element method will also be much larger than the degrees of freedom used by localized basis set, but it is less obvious how the kinetic balance prescription will be enforced in these methods.

Due to the large matrix size after the planewave discretization, the direct diagonalization method will not be applicable. The discretized Hamiltonian operator in (8) generated by the planewave basis set, hereafter also referred to as the Hamiltonian matrix, is dense in both the Fourier space and the real space, which discourages the usage of the shift–invert Lanczos method [54], the contour integral based spectrum slicing methods [55], or the Fermi operator expansion based low order scaling algorithms [56–58]. These methods involve matrix inversion and become less competitive when the matrix is dense. Furthermore, techniques that directly eliminate the negative energy band, such as exact two-component theory [52] involves direct manipulation of the Hamiltonian matrix, which is again not possible to handle in the case of the planewave basis set as a result of the large matrix size. Iterative methods have to be used to diagonalize the discretized Hamiltonian operator, which will be discussed in detail in Section 3.

3. An efficient iterative method for solving the DKS equations

As mentioned in Section 1, there are approximately the same number of eigenvalues and eigenvectors in the positive and the negative energy bands, separated by a large spectral gap with width Δ as shown in Fig. 1. Existing iterative techniques for solving the KS equations cannot be efficiently applied to solve the DKS equations without necessary modification. In this section we demonstrate that the negative energy band of the DKS operator can be efficiently eliminated using only 2 additional matrix–vector multiplications per iteration, based on the locally optimal block preconditioned conjugate gradient (LOBPCG) method [32]. The key component of the proposed new method is an additional filtering (F) step, and we therefore refer to our new method as the LOBPCG-F method. We remark that a similar filtering procedure can also be applied to other preconditioned iterative solvers to eliminate the negative energy band, such as the preconditioned Davidson type methods [59].

3.1. LOBPCG-F algorithm

The LOBPCG-F algorithm is described in Algorithm 1. It can be easily found there that removing the new filtering steps underlined in Algorithm 1 yields directly the standard LOBPCG algorithm as in [32]. The matrices H, T and $f(H)$ used in Algorithm 1 can be defined in the "matrix-free" form, in the sense that they are defined according to subroutines which only computes $Hx, Tx, f(H)x$ for any vector x without forming the matrix elements explicitly. C is a constant used as a shift in the filtering function, which is chosen from the negative energy continuum as shown in Fig. 1.

Algorithm 1. The LOBPCG-F algorithm for solving the DKS density functional theory. The underlined steps describe the new filtering step compared to the standard LOBPCG method. The filter is given in Eq. (16).

Input:
Subroutine Hx (Tx) to multiply the Hamiltonian (preconditioning) operator to a vector, the shift constant C, the normalization constant Ω, the number of eigenvalues n, the initial guess of eigenvectors $X_0 \in \mathbb{R}^{N \times n}$, the number of initial filtering steps N_{init}, the maximum number of iterations N_{iter}, the tolerance for convergence ϵ.

Output:
$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ where $\{\lambda_i\}_{i=1}^n$ are the lowest n eigenvalues in the positive energy band, and $X \in \mathbb{R}^{N \times n}$ are the associated eigenvectors.

1: **for** $\underline{k = 1, \ldots, N_{\mathrm{init}}}$ **do**
2:    Set $\underline{X_0 \leftarrow f(H)X_0}$.
3: **end for**
4: Solve $V, \Lambda_1 \in \mathbb{R}^{n \times n}$ from the generalized eigenvalue problem $(X_0^T H X_0)V = (X_0^T X_0)V\Lambda_1$, with the diagonal elements of Λ_1 sorted in a non-decreasing order.
5: Set $X_1 \leftarrow X_0 V$.
6: Compute the residual $R \leftarrow HX_1 - X_1\Lambda_1$.
7: Set $P \leftarrow []$ to be an empty matrix.
8: **for** $k = 1, \ldots, N_{\mathrm{iter}}$ **do**
9:    Solve the preconditioned residual Y from $TY = R$.
10:   Set $\underline{Y \leftarrow f(H)Y}$.
11:   Construct a subspace $Z = [X_k, Y, P], Z \in \mathbb{R}^{N \times n_Z}$ and $n_Z > n$.
12:   Solve $V, \Lambda_{k+1} \in \mathbb{R}^{n_Z \times n_Z}$ from the generalized eigenvalue problem $(Z^T H Z)V = (Z^T Z)V\Lambda_{k+1}$, with the diagonal elements of Λ_{k+1} sorted in a non-decreasing order.
13:   Set $V \leftarrow$ the first n columns of V.
14:   Set $\Lambda_{k+1} \leftarrow$ the upper-left $n \times n$ block of Λ_{k+1}.
15:   Set $X_{k+1} \leftarrow ZV$.
16:   Compute the residual $R \leftarrow HX_{k+1} - X_{k+1}\Lambda_{k+1}$.
17:   **if** the norm of each column of R is less than ϵ **then**
18:    Exit the loop.
19:   **end if**
20:   Set $R \leftarrow$ the columns of R with norm larger than ϵ.
21:   Set $V \leftarrow$ the columns of V corresponding to remaining columns of R.
22:   Set $P \leftarrow [0, Y, P]V$.
23: **end for**
24: **return** $X \leftarrow X_{k+1}, \Lambda \leftarrow \Lambda_{k+1}$.

We first demonstrate how the error of the eigenfunctions projected to the negative energy band is propagated in the standard LOBPCG method. To simplify the analysis, we assume here that the preconditioner T to be an identity operator. The qualitative result remains to be valid if standard preconditioners in the electronic structure calculation are applied, such as the preconditioner proposed by Teter et al. [31,60]. Furthermore, we need a key assumption that

$$\Delta \gg w, \tag{10}$$

which is valid in the pseudopotential framework with the planewave basis set for $\Delta \approx 37000$ and $w \approx 100$ hold there. The foregoing assumption is also valid in the all-electron calculation using the LAPW basis set, or using the localized basis set given that the basis set remains well conditioned. However, we remark that w can be artificially large when the basis set is overcomplete and then the assumption (10) will not be valid anymore.

The LOBPCG method computes the residual $R = HX - X\Lambda$ once per iteration (line 16 in Algorithm 1). Denote by x a given column of X, λ the corresponding eigenvalue in Λ, and r and y the corresponding column in the residual R and the preconditioned residual Y, respectively. We express r and x using the eigen-decomposition as

$$
\begin{aligned}
r &= \sum_i \psi_i^+ \tilde{r}_i^+ + \sum_j \psi_j^- \tilde{r}_j^-, \\
x &= \sum_i \psi_i^+ \tilde{x}_i^+ + \sum_j \psi_j^- \tilde{x}_j^-,
\end{aligned}
\tag{11}
$$

Here ψ_i^+ is the spinor corresponding to ϵ_i^+ in the positive energy band, and ψ_j^- the spinor corresponding to ϵ_j^- in the negative energy band. We further assume that the error of the eigenfunctions projected to the negative energy band, characerized by

$\max_j |\tilde{x}_j^-|$, is initially very small but not yet vanishes. Since $r = Hx - x\lambda$, the coefficients $\{\tilde{r}_i^+\}, \{\tilde{r}_j^-\}$ are related to the coefficients $\{\tilde{x}_i^+\}, \{\tilde{x}_j^-\}$ according to

$$\tilde{r}_i^+ = (\epsilon_i^+ - \lambda)\tilde{x}_i^+, \quad \tilde{r}_j^- = (\epsilon_j^- - \lambda)\tilde{x}_j^-, \tag{12}$$

implying that the error in the residual $\tilde{r}_i^+$ (resp. $\tilde{r}_j^-$) is amplified by $\epsilon_i^+ - \lambda$ (resp. $\epsilon_j^- - \lambda$) from $\tilde{x}_i^+$ (resp. $\tilde{x}_j^-$). In order to quantify the relative amplification of the error projected to the negative energy band in the residual r compared to the amplification of the error projected to the positve energy band, we define the *amplification factor* as follows

$$\gamma_{\text{LOBPCG}} = \max_\lambda \frac{\max_j |\epsilon_j^- - \lambda|}{\max_i |\epsilon_i^+ - \lambda|}. \tag{13}$$

Since λ is a desired eigenvalue contained in the positive energy band, the assumption (10) implies further that

$$\left|\epsilon_j^- - \lambda\right| \sim \Delta, \quad \left|\epsilon_i^+ - \lambda\right| \sim w. \tag{14}$$

In consequence, we have

$$\gamma_{\text{LOBPCG}} \sim \frac{\Delta}{w} \tag{15}$$

and the relative error projected to the negative energy band in r is amplified by a factor Δ/w compared to that in x. The error remains in the preconditioned residual y and propagates into the subspace used in the next LOBPCG iteration (line 11 in Algorithm 1). In case of $\frac{\Delta}{w} = 300$, even if the initial error is small, say $\max_j |\tilde{x}_j^-| \sim 10^{-15}$, only after 6 iterations, the error of the eigenfunctions projected to the negative energy band will accumulate to $\mathcal{O}(1)$ in the worst case scenario.

In order to compensate for the amplified error projected to the negative energy band, the LOBPCG-F method applies a filtering step to the preconditioned residual Y before the error propagates to the next LOBPCG iteration (line 10 in Algorithm 1). The filtering function takes the form

$$f(E) = \frac{1}{\Omega^2}(E - C)^2, \tag{16}$$

where the shift constant C is chosen to be in the negative energy continuum (see Fig. 1), and Ω is a normalization constant so that $f(E) = 1$ when E is the largest eigenvalue in the positive energy band. We remark that although filtering functions of other forms can also be employed, the quadartic filtering function (16) requires only 2 extra matrix–vector multiplications, and achieves nearly the minimum cost. Under the assumption (10), we have $\Omega \sim \Delta$. In the following we show how the filter (16) works.

Again using the eigen-decomposition for y after the filtering step (line 10 in Algorithm 1) and assuming the identity preconditioner T lead to

$$y = \sum_i \psi_i^+ \tilde{y}_i^+ + \sum_j \psi_j^- \tilde{y}_j^-, \tag{17}$$

where

$$\tilde{y}_i^+ = f(\epsilon_i^+)(\epsilon_i^+ - \lambda)\tilde{x}_i^+, \quad \tilde{y}_j^- = f(\epsilon_j^-)(\epsilon_j^- - \lambda)\tilde{x}_j^-. \tag{18}$$

The amplification factor of the LOBPCG-F method becomes

$$\gamma_{\text{LOBPCG–F}} = \max_\lambda \frac{\max_j |f(\epsilon_j^-)(\epsilon_j^- - \lambda)|}{\max_i |f(\epsilon_i^+)(\epsilon_i^+ - \lambda)|}. \tag{19}$$

Again under the assumption (10), we have

$$\left|f(\epsilon_j^-)(\epsilon_j^- - \lambda)\right| \sim \frac{w^2\Delta}{\Omega^2}, \quad \left|f(\epsilon_i^+)(\epsilon_i^+ - \lambda)\right| \sim \frac{\Delta^2 w}{\Omega^2} \tag{20}$$

and then

$$\gamma_{\text{LOBPCG–F}} \sim \frac{w}{\Delta}. \tag{21}$$

Therefore after each filtering step, the relative error projected to the negative energy band in y is reduced by a factor w/Δ compared to the relative error in x. As mentioned in Section 1, the spectral radius of the matrix originated from the spectrum folding type methods [37] deteriorates as Δ^2 when Δ increases to infinity, and an efficient iterative method is difficult to be designed for the resulting matrix eigenvalue problem. On the contrary, the efficiency of the LOBPCG-F method improves as Δ increases, characterized by the factor w/Δ.

We plot the shape of the filtering function (16) for the case $w = 100$ au with a very small gap $\Delta = 300$ au (correspondingly $c \approx 12$ au) in Fig. 2 (a) and for $w = 100$ au with a larger gap $\Delta = 3000$ au (correspondingly $c \approx 39$ au) in Fig. 2 (b). We observe that even for such relatively small gaps, the filtering function already quickly reduces the error in the negative energy band.
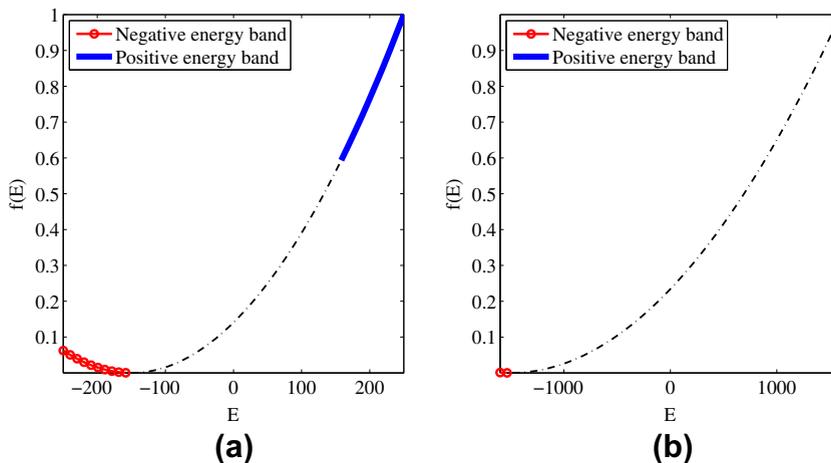
Fig. 2. The filtering function $f(E)$ for $w = 100$ au with a small gap $\Delta = 300$ au (a) and with a large gap $\Delta = 3000$ au (b). The function values correspond to the positive energy band (blue solid line) and the negative energy band (red circles) are separated by the function values in the spectral gap (black dotted line). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

The filtering function (16) is increasingly more effective as the gap Δ increases. In the practical calculation, the gap $\Delta \approx 37000$ au ($c \approx 137$ au), and thus the filtering function can effectively control the error of the eigenfunctions projected to the negative energy band in each LOBPCG-F iteration.

Finally we need to address the initial condition such that $\max_j |\tilde{x}_j^-|$ is small. This can be achieved by the lines 1–3 in Algorithm 1 using N_{init} steps of the same filtering function. A high quality input of the initial vectors X_0 can help reduce the number N_{init}. In the DKS calculation, a good set of initial vectors can be obtained by using the wavefunctions obtained from the KS or the KSSO calculations. The converged KS or KSSO wavefunctions can be used as the initial guess for the large component of the DKS spinors, and the initial guess for the small component of the DKS spinors can be obtained by applying the kinetic balance prescription (9) to them. The good input of the initial vectors is also readily available in consecutive SCF steps, in which X_0 simply takes the output X in the previous SCF step, and the error projected to the negative energy band in X is negligible as controlled by the LOBPCG-F method. Using such strategy, we find that the LOBPCG-F algorithm is robust by setting $N_{\mathrm{init}} = 3$ for the first SCF iteration (for safety), and then setting $N_{\mathrm{init}} = 1$ in the following SCF steps.

3.2. Implementation

Compared to the standard LOBPCG algorithm which applies the Hamiltonian matrix once per iteration step, the LOBPCG-F algorithm applies the Hamiltonian matrix for 2 extra times per iteration due to the filtering step (16). The computational cost of the numerical linear algebra operations, such as for the orthogonalization step and for the projected Rayleigh–Ritz eigenvalue problem, remains to be the same as that in the standard LOBPCG algorithm. Furthermore, the implementation of LOBPCG-F algorithm is very simple. Since the implementation of the LOBPCG algorithm always provides an interface for the preconditioner, one only needs to apply the filtering function after the standard preconditioner to eliminate the negative energy states components as shown in Algorithm 1, and the overall framework for LOBPCG, such as that provided by BLOPEX [33] does not need to be changed.

4. Numerical results

We implement the LOBPCG-F method for solving the DKS system (8) using the HGH pseudopotential [46], with the local and nonlocal pseudopotential implemented fully in the real space [61]. The exchange–correlation functional under local density approximation (LDA) is used. Together with the module DKS for solving the four-component DKS system, we also implement the standard LOBPCG method for solving the KS density functional theory (module KS), and for solving the two-component KS density functional theory with SOC in the pseudopotential framework (module KSSO). Although we implement the modules from scratch, the LOBPCG-F method can also be implemented without too much efforts by developers of other packages such as ABINIT [39], Quantum ESPRESSO [62] etc. The major difference between the module KS and KSSO is that KSSO includes the spin–orbit coupling term between the core electrons and the valence electrons in the form of $\sum_l V_l^{\mathrm{SO}}(\mathbf{r}, \mathbf{r}')\mathbf{L}' \cdot \mathbf{S}$ as in Eq. (7). The real and complex arithmetics are equally handled in our implementation. The description of the modules KS, KSSO and DKS is summarized in Table 1. Since our main focus is the relativistic effect such as SOC, the module KS directly uses a two-component spinor rather than a one-component spinor. However, this does not lead to changes in the computed physical quantities such as the total energy. The availability of modules KS and KSSO allows us to directly benchmark our implementation with existing software such as ABINIT for electronic structure calculation. To

Table 1
The number of components for describing a spinor (N_{spinor}), the arithmetic, and the diagonalization solver for modules KS, KSSO and DKS in our implementation.

Module	N_{spinor}	Arithmetic	Solver	Description
KS	2	Real	LOBPCG	Kohn–Sham
KSSO	2	Complex	LOBPCG	Kohn–Sham with spin–orbit coupling
DKS	4	Complex	LOBPCG-F	Dirac–Kohn–Sham

facilitate the presentation of the numerical results, we adopt the standard convention that all energies are shifted by $-mc^2$, so that the positive energy continuum starts from 0 rather than $+mc^2$.

More computational details of our implementation and the numerical results are as follows. The preconditioner proposed by Teter et al. [31,60] is employed by both LOBPCG and LOBPCG-F methods. Anderson mixing [63] with Kerker preconditioner [64] is used for the SCF iteration. Gamma point Brillouin sampling is used for simplicity for all calculations, even for crystal systems such as Bi_2Se_3. This is because the purpose of this manuscript is to demonstrate the capability of the LOBPCG-F method for solving DKS systems. The support of k-point sampling will be added in the future work. The density and the wavefunction are resolved using the same grid in both real space and Fourier space, and the grid size is measured by the corresponding kinetic energy cutoff in the Fourier space in atomic unit, denoted by E_{cut}. All computational experiments are performed on the Hopper system at the National Energy Research Scientific Computing (NERSC) center. Each Hopper node consists of two twelve-core AMD "MagnyCours" 2.1 GHz processors and has 32 gigabytes (GB) DDR3 1333 MHz memory. Each core processor has 64 kilobytes (KB) L1 cache and 512 KB L2 cache. All modules KS, KSSO and DKS are implemented sequentially, and one core processor is used in each computation.

4.1. Setup

We present numerical results for computing the total free energy of three dimers systems Pt_2, Au_2, and TlF, as well as a condensed matter system Bi_2Se_3. For simplicity of demonstration, the electronic structure calculation of the dimers are not computed with their optimal bond lengths. Instead, all the dimers are placed in a supercell of dimension 15.000 au, 10.000 au, 10.000 au along the x, y, z directions respectively. The positions of the dimers are placed at $(5.000, 0.000, 0.000)$ au and $(10.000, 0.000, 0.000)$ au, respectively. In the HGH pseudopotential, the elements Pt, Au and Tl have the option of including semicore electrons [46] or not, and our implementation can handle both cases. Here we treat Pt atoms without semicore electrons, and treat Au atoms and Tl atoms with semicore electrons. Our implementation assumes an orthorhombic supercell. Non-orthorhombic cells can be converted into orthorhombic cells for the total energy computation. For example, the topological insulator phase of Bi_2Se_3 has rhombohedral crystal structure [65]. The rhombohedral structure can be converted into an orthorhombic cell as shown in Fig. 3 with 12 Bi atoms and 18 Se atoms in the supercell. The lengths of the converted orthorhombic supercell is 7.820 au, 13.544 au, and 54.123 au along the x, y, z directions, respectively. The kinetic energy cutoff E_{cut} is set to be the same in ABINIT and in our implementation, and E_{cut} is high enough so that the difference in the total free energy per atom is less than 1 meV. Neither Bi nor Se allows semicore treatment in HGH pseudopotential. More details of the setup of the computational systems are illustrated in Table 2. Note that TlF requires particularly large kinetic energy cutoff. This is mainly due to localized pseudopotential of the F atom, which requires a very fine numerical grid to resolve.

4.2. Calibration with ABINIT

We compare our result with ABINIT [39] which also supports the usage of the HGH pseudopotential and the LDA approximation. ABINIT has the capability of performing KS calculation (by setting nspinor = 1, nspden = 1) and KSSO calculation with vector magnetization of the spin density (by setting nspinor = 2, nspden = 4). At present, ABINIT does not provide the DKS module. As shown by other researchers, the SCF iteration can become oscillatory for spin-unrestricted calculations, which is known as the "spin-sloshing" and exists even for a simple spin-unrestricted O_2 dimer system [66]. In order to accelerate the SCF iteration, the temperature is set to be 5000 K in all simulations reported below. Although the increase of the temperature in the outer SCF loop directly affects the convergence of density, magnetization and potential, it does not directly affect the convergence behavior of individual eigenfunctions, which is controlled by eigensolvers in the inner loop such as
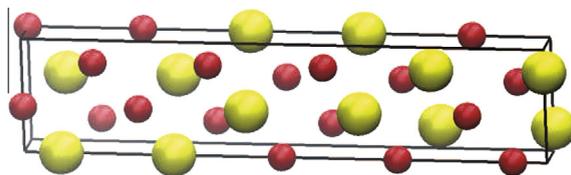
Fig. 3. The converted orthorhombic supercell for describing a topological insulating system Bi_2Se_3 with 12 Bi atoms (large yellow balls) and 18 Se atoms (small red balls). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

Table 2
The kinetic energy cutoff E_{cut}, the number of valance electrons and the treatment of semicore electrons for the systems under consideration. The same kinetic energy cutoff is used in both ABINIT and our implementation.

System	E_{cut}	# Electron	Semicore
Pt_2	50	20	No
Au_2	50	22	Yes
TlF	400	20	Yes
Bi_2Se_3	45	168	No

LOBPCG or LOBPCG-F. Finite temperature formulation of the KS density functional theory [67] is used, and the total free energy is the quantity to be measured [68]. In particular, the total free energy computed from modules KS, KSSO and DKS in our implementation are denoted by $E_{KS}, E_{KSSO}, E_{DKS}$, respectively, while those computed from the corresponding modules KS and KSSO in ABINIT are denoted by $E_{KS}^{ABINIT}, E_{KSSO}^{ABINIT}$, respectively. To facilitate the illustration, all the energies will be measured in the unit of au, and all the differences of energies will be measured in the unit of meV (1 au $\approx$ 27211 meV).

In Table 3, we compare the total free energies for the KS calculation. The differences of energies per atom are less than 1 meV in all cases. In Table 4, we compare the total free energies for the KSSO calculation. Similarly, the differences of energies per atom are also less than 1 meV in all cases. Tables 3 and 4 indicate that the KS and KSSO modules in our implementation are accurate. Comparing the data shown in Tables 3 and 4, we find that the total free energies are consistently lower when SOC effect is considered.

4.3. DKS results

Table 5 demonstrates the total free energies computed using the module DKS in our implementation, for which ABINIT does not provide the same functionality, and shows the differences of the energies per atom between calculations using modules KS, KSSO and DKS. The energies obtained from DKS is systematically lower than those in KS and KSSO. The full relativistic correction described by DKS is large (in the order of hundreds of meV per atom) compared to the result obtained by KS which only takes into account the scalar relativistic effect in the level of pseudopotential. Most of the correction originates from SOC effect. The remaining correction due to the difference between the Dirac description (DKS) and the Pauli description (KSSO, with the relativistic effect taken account only in the HGH pseudopotential) is generally two orders of magnitude smaller than the SOC effect. Furthermore, since the relativistic effect of the core electrons is more significant than that of the valence electrons, the difference between DKS and KSSO is particularly small when the semicore treatment is not present, such as in the case of Pt_2 and Bi_2Se_3. With the same HGH pseudopotential (7), the correction from the Dirac description compared to the Pauli description is found to be smaller than the transferability error of the pseudopotentials. Our study also agrees with the recent study using the planewave implementation of the ZORA equations – one kind of approximate two-component relativistic theory by NWChem [69]. We also note that in TlF, the difference between DKS and KSSO is around 10% of the effect due to SOC. To the extent of our knowledge, our result for the first time reveals the quantitative difference between the Pauli description and the full Dirac description of relativistic effects for the valence electrons in the pseudopotential framework for the systems under study.

Table 3
The total free energies E_{KS}^{ABINIT} calculated from ABINIT, and E_{KS} from module KS in our implementation for systems under consideration.

System	E_{KS}^{ABINIT} au	E_{KS} au	$(E_{KS} - E_{KS}^{ABINIT})/N_{atom}$ meV
Pt_2	−52.364236	−52.364274	−0.51
Au_2	−66.438610	−66.438601	0.12
TlF	−74.156822	−74.156828	−0.08
Bi_2Se_3	−237.357221	−237.357063	0.14

Table 4
The total free energies E_{KSSO}^{ABINIT} calculated from ABINIT, and E_{KSSO} from module KSSO in our implementation for systems under consideration.

System	E_{KSSO}^{ABINIT} au	E_{KSSO} au	$(E_{KSSO} - E_{KSSO}^{ABINIT})/N_{atom}$ meV
Pt_2	−52.411523	−52.411564	−0.56
Au_2	−66.473300	−66.473302	−0.03
TlF	−74.178538	−74.178602	−0.87
Bi_2Se_3	−237.728372	−237.728987	−0.56

Table 5
The total free energies E_{KSSO} from module DKS in our implementation for systems under consideration.

System	E_{DKS} au	$(E_{DKS} - E_{KS})/N_{atom}$ meV	$(E_{DKS} - E_{KSSO})/N_{atom}$ meV
Pt_2	−52.411595	−643.82	−0.41
Au_2	−66.473702	−477.56	−5.44
TlF	−74.181219	−331.85	−35.60
Bi_2Se_3	−237.730530	−338.75	−1.40

Table 6
The wall clock time (in the unit of sec) for performing one step of SCF iteration for systems under consideration.

System	KS	KSSO	DKS
Pt_2	4	26	113
Au_2	6	37	155
TlF	148	709	2787

4.4. Computational efficiency

We demonstrate the computational efficiency of the LOBPCG-F method in terms of the wall clock time per SCF iteration for Pt_2, Au_2 and TlF in Table 6. Each SCF iteration consists of 3 LOBPCG iterations for the KS and the KSSO calculation, and 3 LOBPCG-F iterations for the DKS calculation. The computation of Bi_2Se_3 uses significant amount of virtual memory due to the large number of valance electrons in the system, and the corresponding computational time is therefore not meaningful and not reported here. This will not be a problem when our parallel implementation of our method is introduced in the future. We also remark that compared to the standard LOBPCG method, the LOBPCG-F method does not introduce any additional memory cost.

From KS to KSSO the computational time increased by a factor of $5 \sim 6$. As discussed before, the module KS uses a two-component formulation, and the increase of the computational time is mostly due to the change from real to complex arithmetic. For computing the same quantity, the complex arithmetic is 4 times more expensive than the real arithmetic. The remaining difference comes from that KSSO uses complex to complex (c2c) Fourier transform, and KS uses real to complex (r2c) and complex to real (c2r) Fourier transform.

From KSSO to DKS the computational time increases by a factor around 4. In this process, the change of the number of components from 2 to 4 inevitably increases the computational time by a factor of 2. The remaining factor of 2 in the increased computational time originates from the difference between the LOBPCG and the LOBPCG-F method. Each LOBPCG iteration applies the Hamiltonian to all spinors once, and each LOBPCG-F iteration applies the Hamiltonian operator to all spinors for 3 times to filter the components in the negative energy band. However, since the linear algebra operations (such
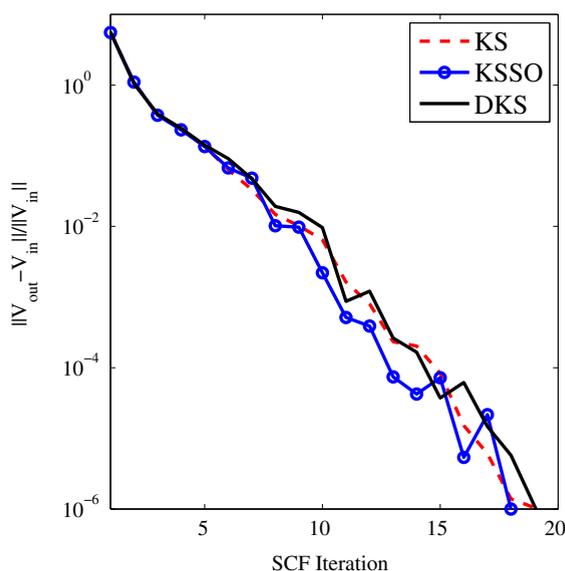
Fig. 4. Convergence of the SCF iteration for KS (red dashed line), KSSO (blue line with circles) and DKS calculations (black solid line) for Au_2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as the orthogonalization procedure of the spinors) are the same in the LOBPCG and the LOBPCG-F method, the usage of LOBPCG-F only increases the computational time by a factor around 2 in practice.

Fig. 4 compares the convergence of the SCF iteration for KS, KSSO and DKS using Au_2 as an example. The convergence of the SCF iteration is measured by the quantity $\|V_{out} - V_{in}\|/\|V_{in}\|$, where V_{in} is the local part of the effective potential before each SCF iteration, and V_{out} is the effective potential after each SCF iteration. Fig. 4 indicates that the usage of the LOBPCG-F method does not deteriorate the convergence rate of the SCF iteration. Combining Table 6 and Fig. 4, we conclude that the LOBPCG-F is a robust and efficient method for solving the DKS system in practice.

5. Conclusion and outlook

In this manuscript we develop for the first time a simple and efficient iterative algorithm, named the LOBPCG-F method, for directly solving the Dirac–Kohn–Sham (DKS) equations in the relativistic density functional theory. By adding an additional filtering step in the preconditioning stage of the LOBPCG method, the LOBPCG-F method is able to compute the desired eigenvalues and eigenvectors in the positive energy band without computing any state in the negative energy band. The LOBPCG-F method requires only 2 extra computational costs and little extra coding effort, and thus remarkably facilitates the transition from nonrelativistic Kohn–Sham (KS) calculations using LOBPCG methods to relativistic DKS calculations in studying the relativistic effect such as spin–orbit coupling, without the need of approximating the DKS equation by two-component relativistic theory. We also remark that even though the LOBPCG-F method is efficient for solving the DKS systems, it is still significantly more expensive to solve the more complicated four-component DKS systems compared to the KS systems or the KSSO systems. In practice one should decide whether two-component theory or the four-component theory should be pursued by balancing the desired efficiency and accuracy.

The efficiency of the filtering step is determined by the factor w/Δ, where Δ is the spectral gap between the positive and the negative energy band, and w is the maximum of the widths of the positive and the negative energy band (see Fig. 1). The smaller the factor is, the more efficiently the filter performs. It is worth highlighting that the proposed filtering technique is more general and can be embedded into other preconditioned iterative solvers, while the LOBPCG method is employed in this work for its high efficiency in the electronic structure calculation. We demonstrate the applicability of the LOBPCG-F method in the pseudopotential framework in the planewave basis set, in which the condition $\Delta \gg w$ is satisfied. The plane-wave basis set automatically satisfies the kinetic balance prescription and is free from the variational collapse. Our results compared with ABINIT indicate that our implementation is accurate, and the LOBPCG-F method does not lead to deterioration in the convergence rate of the SCF iteration. We directly observe that the difference between the total energies between the two-component KS density functional theory with SOC description (module KSSO) and the DKS description (module DKS) is no more than a few tens of meV per atom for systems under study, and thus it is not cost-effective to solve the four-component DKS problem in the pseudopotential framework for most systems in practice. The LOBPCG-F method will be more effective for studying the relativistic effects in solving the all-electron DKS system directly when the condition $\Delta \gg w$ is satisfied, such as using the linearized augmented plane-wave basis set (LAPW) as in the WIEN2k software, or the localized basis set given that the basis set remains well conditioned. This will be our future work.

References

[1] D.J. Griffiths, Introduction to Quantum Mechanics, Prentice Hall, London, 1995.
[2] C.L. Kane, E.J. Mele, Z_2 Topological order and the quantum spin Hall effect, Phys. Rev. Lett. 95 (2005) 146802.
[3] P. Atkins, R. Friedman, Molecular Quantum Mechanics, 4th ed., Oxford University Press, New York, 2005.
[4] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871.
[5] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133.
[6] P. Pyykkö, Relativistic effects in structural chemistry, Chem. Rev. (1988) 563–594.
[7] P. Pyykkö, The physics behind chemistry and the periodic table, Chem. Rev. (2012) 371–384.
[8] E. Engel, Relativistic density functional theory: foundations and basic formalism, in: P. Schwerdtfeger (Ed.), Relativistic Electronic Structure Theory, Part I: Fundamentals, Elsevier Science B.V, Amsterdam, 2002, pp. 523–621.
[9] A.K. Rajagopal, J. Callaway, Inhomogeneous electron gas, Phys. Rev. B 7 (1973) 1912.
[10] A.K. Rajagopal, Inhomogeneous relativistic electron gas, J. Phys. C: Solid State Phys. 11 (1978) L943.
[11] A.H. MacDonald, S.H. Vosko, A relativistic density functional formalism, J. Phys. C: Solid State Phys. 12 (1979) 2977.
[12] A.K. Rajagopal, Time-dependent functional theory of coupled electron and electromagnetic fields in condensed-matter systems, Phys. Rev. A 50 (1994) 3759–3765.

[13] A.K. Rajagopal, Time-dependent variational principle and the effective action in density-functional theory and Berry's phase, Phys. Rev. A 54 (1996) 3916–3922.
[14] B. Thaller, The Dirac Equation, Springer, Berlin, 1992.
[15] W.H.E. Schwarz, H. Wallmeier, Basis set expansions of relativistic molecular wave equations, Mol. Phys. 46 (1982) 1045.
[16] W. Kutzelnigg, Basis set expansion of the Dirac operator without variational collapse, Int. J. Quantum Chem. 25 (1984) 107.
[17] M. Lewin, E. Séré, Spectral pollution and how to avoid it, Proc. London Math. Soc. 100 (2010) 864.
[18] Y. Ishikawa, R.C. Binning Jr., K.M. Sando, Dirac–Fock discrete-basis calculations on the beryllium atom, Chem. Phys. Lett. 111 (1983) 101.
[19] R.E. Stanton, S. Havriliak, Kinetic balance: a partial solution to the problem of variational safety in Dirac calculations, J. Chem. Phys. 81 (1984) 1910.
[20] K.G. Dyall, K. Fgri Jr., Kinetic balance and variational bounds failure in the solution of the Dirac equation in a finite Gaussian basis set, Chem. Phys. Lett. 174 (1990) 25.
[21] A. Rosen, D.E. Ellis, Relativistic molecular calculations in the Dirac-Slater model, J. Chem. Phys. 62 (1975) 3039.
[22] D.E. Ellis, G.L. Goodman, Self-consistent Dirac–Slater calculations for molecules and embedded clusters, Int. J. Quantum Chem. 25 (1984) 185.
[23] W.J. Liu, G.Y. Hong, D.D. Dai, L.M. Li, M. Dolg, The Beijing four-component density functional program package (BDF) and its application to EuO, EuS, YbO and YbS, Theor. Chem. Acc. 96 (1997) 75.
[24] S. Varga, B. Fricke, H. Nakamatsu, T. Mukoyama, J. Anton, D. Geschke, A. Heitmann, E. Engel, T. Baştuğ, Four-component relativistic density functional calculations of heavy diatomic molecules, J. Chem. Phys. 112 (2000) 3499.
[25] T. Yanai, H. Iikura, T. Nakajima, Y. Ishikawa, K. Hirao, A new implementation of four-component relativistic density functional method for heavy-atom polyatomic systems, J. Chem. Phys. 115 (2001) 8267.
[26] T. Saue, T. Helgaker, Four-component relativistic Kohn–Sham theory, J. Comput. Chem. 23 (2002) 814.
[27] H.M. Quiney, P. Belanzoni, Relativistic density functional theory using Gaussian basis sets, J. Chem. Phys. 117 (2002) 5550.
[28] N. Troullier, J.L. Martins, Efficient pseudopotentials for plane-wave calculations, Phys. Rev. B 43 (1991) 1993–2006.
[29] O.K. Andersen, Linear methods in band theory, Phys. Rev. B 12 (1975) 3060–3083.
[30] I. Štich, R. Car, M. Parrinello, S. Baroni, Conjugate gradient minimization of the energy functional: a new method for electronic structure calculation, Phys. Rev. B 39 (1989) 4997–5004.
[31] M.C. Payne, M.P. Teter, D.C. Allen, T.A. Arias, J.D. Joannopoulos, Iterative minimization techniques for ab initio total energy calculation: molecular dynamics and conjugate gradients, Rev. Mod. Phys. 64 (1992) 1045–1097.
[32] A.V. Knyazev, Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method, SIAM J. Sci. Comput. 23 (2001) 517.
[33] A.V. Knyazev, M.E. Argentati, I. Lashuk, E.E. Ovtchinnikov, Block locally optimal preconditioned eigenvalue xolvers (BLOPEX) in HYPRE and PETSc, SIAM J. Sci. Comput. 29 (2007) 2224–2239.
[34] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169–11186.
[35] K. Wu, A. Canning, H.D. Simon, L.W. Wang, Thick-restart Lanczos method for electronic structure calculations, J. Comput. Phys. 154 (1999) 156.
[36] Y. Zhou, Y. Saad, M.L. Tiago, J.R. Chelikowsky, Self-consistent-field calculations using Chebyshev-filtered subspace iteration, J. Comput. Phys. 219 (2006) 172–184.
[37] L.W. Wang, A. Zunger, Solving schrödinger's equation around a desired energy: application to silicon quantum dots, J. Chem. Phys. 100 (1994) 2394.
[38] F. Bottin, S. Leroux, A. Knyazev, G. Zérah, Large-scale ab initio calculations based on three levels of parallelization, Comput. Mater. Sci. 42 (2008) 329–336.
[39] X. Gonze, B. Amadon, P. Anglade, J.M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Cote, et al, Abinit: first-principles approach to material and nanosystem properties, Comput. Phys. Commun. 180 (2009) 2582–2615.
[40] K. Schwarz, P. Blaha, G.K.H. Madsen, Electronic structure calculations of solids using the WIEN2k package for material sciences, Comput. Phys. Commun. 147 (2002) 71–76.
[41] C. van Wüllen, Relativistic density functional theory, in: M. Barysz, Y. Ishikawa (Eds.), Relativistic Methods for Chemists, Springer, New York, 2010, pp. 191–214.
[42] D.M. Ceperley, B.J. Alder, Ground state of the electron gas by a stochastic method, Phys. Rev. Lett. 45 (1980) 566–569.
[43] J.P. Perdew, A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, Phys. Rev. B 23 (1981) 5048–5079.
[44] M. Dolg, X.Y. Cao, Relativistic pseudopotentials: their development and scope of applications, Chem. Rev. 112 (2012) 403.
[45] Y. Ishikawa, G. Malli, Effective core potentials for fully relativistic Dirac–Fock calculations, J. Chem. Phys. 75 (1981) 5423.
[46] C. Hartwigsen, S. Goedecker, J. Hutter, Relativistic separable dual-space Gaussian pseudopotentials from H to Rn, Phys. Rev. B 58 (1998) 3641.
[47] R. Martin, Electronic Structure – Basic Theory and Practical Methods, Cambridge Univ. Pr, West Nyack, NY, 2004.
[48] J.D. Talman, Minimax principle for the Dirac equation, Phys. Rev. Lett. 57 (1986) 1091.
[49] S.N. Datta, G. Deviah, The minimax technique in relativistic Hartree–Fock calculations, Pramana - J. Phys. 30 (1988) 387.
[50] J.P. Desclaux, J. Dolbeault, M.J. Esteban, P. Indelicato, E. Séré, Computational approaches of relativistic models in quantum chemistry, Handbook Numer. Anal., vol. 10, 2003, pp. 453.
[51] Z.D. Li, S.H. Shao, W.J. Liu, Relativistic explicit correlation: coalescence conditions and practical suggestions, J. Chem. Phys. 136 (2012) 144117.
[52] W.J. Liu, Ideas of relativistic quantum chemistry, Mol. Phys. 108 (2010) 1679.
[53] Q.M. Sun, W.J. Liu, W. Kutzelnigg, Comparison of restricted, unrestricted, inverse, and dual kinetic balances for four-component relativistic calculations, Theor. Chem. Acc. 129 (2011) 423.
[54] B.N. Parlett, Y. Saad, Complex shift and invert strategies for real matrices, Linear Algebra Appl. 88 (1987) 575–595.
[55] E. Polizzi, Density-matrix-based algorithm for solving eigenvalue problems, Phys. Rev. B 79 (2009) 115112.
[56] L. Lin, J. Lu, L. Ying, W. E, Pole-based approximation of the Fermi–Dirac function, Chin. Ann. Math. 30B (2009) 729.
[57] L. Lin, J. Lu, L. Ying, R. Car, W. E, Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems, Commun. Math. Sci. 7 (2009) 755.
[58] T. Ozaki, Efficient low-order scaling method for large-scale electronic structure calculations with localized basis functions, Phys. Rev. B 82 (2010) 075131.
[59] G. Sleijpen, H.A. Van der Vorst, A Jacobi–Davidson iteration method for linear eigenvalue problems, SIAM Rev. 42 (2000) 267–293.
[60] M. Teter, M. Payne, D. Allan, Solution of Schrödinger's equation for large systems, Phys. Rev. B 40 (1989) 12255.
[61] J.E. Pask, P.A. Sterne, Real-space formulation of the electrostatic potential and total energy of solids, Phys. Rev. B 71 (2005) 113101.
[62] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, et al, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, J. Phys.: Condens. Matter 21 (2009) 395502–395520.
[63] D.G. Anderson, Iterative procedures for nonlinear integral equations, J. Assoc. Comput. Mach. 12 (1965) 547–560.
[64] G.P. Kerker, Efficient iteration scheme for self-consistent pseudopotential calculations, Phys. Rev. B 23 (1981) 3082–3084.
[65] H. Zhang, C.X. Liu, X.L. Qi, X. Dai, Z. Fang, S.C. Zhang, Topological insulators in Bi2Se3, Bi2Te3 and Sb2Te3 with a single Dirac cone on the surface, Nat. Phys. 5 (2009) 438–442.
[66] L.D. Marks, D.R. Luke, Robust mixing for ab initio quantum mechanical calculations, Phys. Rev. B 78 (2008) 075114–075125.
[67] N. Mermin, Thermal properties of the inhomogeneous electron gas, Phys. Rev. 137 (1965) A1441.
[68] A. Alavi, J. Kohanoff, M. Parrinello, D. Frenkel, Ab initio molecular dynamics with excited electrons, Phys. Rev. Lett. 73 (1994) 2599.
[69] P. Nichols, N. Govind, E.J. Bylaska, W.A. de Jong, Gaussian basis set and planewave relativistic spin-orbit methods in NWChem, J. Chem. Theory Comput. (2009) 491.