A mesh-free convex approximation scheme for Kohn–Sham density functional theory

Phanish Suryanarayana, Kaushik Bhattacharya *, Michael Ortiz

Division of Engineering and Applied Science, California Institute of Technology, CA 91125, USA

A R T I C L E   I N F O

Article history:
Received 9 August 2010
Received in revised form 27 February 2011
Accepted 9 March 2011
Available online 16 March 2011

Keywords:
Convex approximation scheme
Mesh-free methods
Kohn–Sham
Density functional theory
Maximum-entropy

A B S T R A C T

Density functional theory developed by Hohenberg, Kohn and Sham is a widely accepted, reliable ab initio method. We present a non-periodic, real space, mesh-free-convex approximation scheme for Kohn–Sham density functional theory. We rewrite the original variational problem as a saddle point problem and discretize it using basis functions which form the Pareto optimum between competing objectives of maximizing entropy and minimizing the total width of the approximation scheme. We show the utility of the approximation scheme in performing both all-electron and pseudopotential calculations, the results of which are in good agreement with literature.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Density functional theory (DFT) [1] is a viable and accurate method for calculating the ground state properties of materials where the many-body electronic wavefunction is replaced by the electron density as the fundamental unknown field, thereby significantly reducing the complexity of the problem. A common implementation of DFT is through the Kohn–Sham method [2] (KS-DFT) where the intractable problem of interacting electrons is reduced to a tractable problem of non-interacting electrons moving in an effective potential. In this method, models are used to quantify the so called exchange-correlation energy, which consists of all the unknown components of the energy. Commonly used models include the local density approximation (LDA) [2] and the generalized gradient approximation (GGA) [3,4].

The plane-wave basis is frequently chosen for solving the Kohn–Sham problem [5–8] because it is a complete, orthonormal basis and the convolutions can be efficiently evaluated using fast Fourier transform (FFT). However, the plane-wave basis is ideally suited to periodic systems and the study of non-periodic systems requires the introduction of artificial supercell periodicity, which is unnatural and could potentially lead to spurious results. This limitation has lead to the development of a number of real space approaches [9–18], including the finite-element method [19–21]. The finite-element method is particularly well suited for the study of non-periodic systems because its unstructured nature allows for efficient adaptive resolution of space [19,22]. Additionally, it employs a complete basis and allows for efficient parallelization of the calculations owing to their local nature. However, the finite-element method requires the generation of a mesh with particular care to nodal connectivity. Further, a large number of basis functions are required [19,20], rendering the method to be an expensive one.

In this paper, we develop a non-periodic, real space, mesh-free convex approximation scheme for KS-DFT. We use max-ent basis functions which are the Pareto optimum between maximizing the entropy and the locality of the approximation

* Corresponding author.
E-mail address: bhatta@caltech.edu (K. Bhattacharya).

0021-9991/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.03.018
scheme [23]. The finite-element method can be recovered as a special case of the approximation scheme presented here [23].

It provides advantages over conventional mesh-based methods in terms of the flexibility and adaptivity of the spatial discretization. Additionally, it is known to be significantly more accurate than the corresponding simplicial finite-elements [23,24], also verified by the results obtained in this work. The limitations of the method in comparison to the finite-element method include the larger support-size of the basis functions and the need for expensive quadrature rules.

The outline of the paper is as follows. In Sections 2 and 3 we provide a brief introduction to KS-DFT and the convex approximation scheme respectively. In Section 4, we discuss in some detail the numerical implementation which we validate through select examples in Section 5. Finally, we conclude in Section 6.

2. Kohn–Sham density functional theory

The energy functional for KS-DFT is given by [25,26]

$$
\mathcal{E}(\Psi, \mathbf{R}) = T_s(\rho_\sigma, \rho_\beta) + E_{xc}(\rho_\sigma, \rho_\beta) + E_{hi}(\rho) + E_{ext}(\rho, \mathbf{R}) + E_{zz}(\mathbf{R}),
$$

(1)

where

$$
\rho_\sigma(\mathbf{x}) = \sum_{i=1}^{N_\sigma} \psi_{\sigma i}^*(\mathbf{x})\psi_{\sigma i}(\mathbf{x}) = \sum_{i=1}^{N_\sigma} |\psi_{\sigma i}(\mathbf{x})|^2,
$$

(2)

$$
\rho_\beta(\mathbf{x}) = \sum_{i=1}^{N_\beta} \psi_{\beta i}^*(\mathbf{x})\psi_{\beta i}(\mathbf{x}) = \sum_{i=1}^{N_\beta} |\psi_{\beta i}(\mathbf{x})|^2,
$$

(3)

$$
\rho = \rho_\sigma + \rho_\beta, \quad N = N_\sigma + N_\beta.
$$

(4)

The vector of Kohn–Sham wavefunctions is denoted by $\Psi = \{\psi_{1z}, \psi_{2z}, \ldots, \psi_{Nz, \sigma}, \psi_{1\beta}, \psi_{2\beta}, \ldots, \psi_{N\beta, \beta}\}$ and $\mathbf{R} \in \mathbb{R}^{3M}$ is the collection of all the nuclear positions $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M)$. The subscript $\sigma \in \{\alpha, \beta\}$ is used to signify the spin of the electron. The number of electrons is represented by $N$ and the number of nuclei by $M \in \mathbb{N}$. The wavefunctions are orthonormal, thus satisfying the relation

$$
\int \psi_{\sigma i}^*(\mathbf{x})\psi_{\beta j}(\mathbf{x})d\mathbf{x} = \delta_{ij}, \quad \sigma \in \{\alpha, \beta\}, \quad i, j = 1, 2, \ldots, N_\sigma.
$$

(5)

Note that if the domain of integration is not specified, it refers to all of space $\mathbb{R}^3$. The kinetic energy of the non-interacting electrons is denoted by $T_s(\rho_\sigma, \rho_\beta)$, the classical electrostatic interaction energy of the electron density by $E_{hi}(\rho)$, the interaction energy with the external potential ($V_{ext}(\mathbf{x}, \mathbf{R})$) by $E_{ext}(\rho, \mathbf{R})$ and the repulsive energy between the nuclei by $E_{zz}(\mathbf{R})$. They can be expressed as

$$
T_s(\rho_\sigma, \rho_\beta) = -\frac{1}{2} \sum_\sigma \sum_{i=1}^{N_\sigma} \int \psi_{\sigma i}^*(\mathbf{x})\nabla^2 \psi_{\sigma i}(\mathbf{x})d\mathbf{x},
$$

(6)

$$
E_{hi}(\rho) = \frac{1}{2} \int \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}d\mathbf{x}d\mathbf{x}',
$$

(7)

$$
E_{ext}(\rho, \mathbf{R}) = \int \rho(\mathbf{x})V_{ext}(\mathbf{x}, \mathbf{R})d\mathbf{x},
$$

(8)

$$
E_{zz}(\mathbf{R}) = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{Z_iZ_j}{|\mathbf{R}_i - \mathbf{R}_j|}.
$$

(9)

$E_{hi}(\rho)$ is commonly referred to as Hartree energy. The second term $E_{xc}(\rho_\sigma, \rho_\beta)$ in Eq. (1) denotes the exchange-correlation energy for which we use the LDA, in which the exchange-correlation energy is further separated into the exchange and correlation parts:

$$
E_{xc}(\rho_\sigma, \rho_\beta) = E_x(\rho_\sigma, \rho_\beta) + E_c(\rho_\sigma, \rho_\beta),
$$

(10)

where

$$
E_x(\rho_\sigma, \rho_\beta) = -\frac{3}{4} \left(\frac{6}{\pi}\right)^{1/3} \int (\rho_{\sigma}^{5/3}(\mathbf{x}) + \rho_{\beta}^{5/3}(\mathbf{x}))d\mathbf{x},
$$

(11)

$$
E_c(\rho_\sigma, \rho_\beta) = \int \epsilon_c(\rho_\sigma(\mathbf{x}), \rho_\beta(\mathbf{x}))\rho(\mathbf{x})d\mathbf{x}.
$$

(12)

We use the parametrization of Perdew and Wang [27] for the correlation energy $\epsilon_c(\rho_\sigma, \rho_\beta)$.

The electrostatic interaction energy $E_{hi}(\rho)$ and the repulsive energy of the nuclei $E_{zz}(\mathbf{R})$ as put forth in Eqs. (7) and (9) are non-local in nature and thus are not amenable to a local discretization. We overcome this difficulty [8,19] by representing
the nuclear charge of magnitude \( Z_l \) at a site \( R_l \in \mathbb{R}^3 \) by means of a regularized bounded charge distribution \( -Z_l \delta_{R_l}(x) \) with a compact support in a neighborhood of a small ball around \( R_l \) and such that \( \int \delta_{R_l}(x) dx = 1 \) for \( 1 \leq l \leq M \). The repulsive nuclear energy can then be rewritten as

\[
E_{\text{ext}}(R) = \frac{1}{2} \int \int \frac{\mathbf{b}(x, R) \cdot \mathbf{b}(x', R)}{|x - x'|} dx dx',
\]

where

\[
b(x, R) = -\sum_{l=1}^{M} Z_l \delta_{R_l}(x)
\]

denotes the sum of regularized distributions of all the nuclei present in the system. Note that Eq. (13) differs from Eq. (9) by the self-energy of the nuclei, but this is an inconsequential constant depending only on the nuclear charges. The electrostatic potential due to the nuclei and electron charge distribution can now be computed as a solution to the Poisson equation

\[
\frac{1}{4\pi} \nabla^2 \phi(x, R) = \rho(x) + b(x, R).
\]

This equation has the unique solution

\[
\phi(x, R) = \int \frac{\rho(x')}{|x - x'|} dx' + \int \frac{b(x', R)}{|x - x'|} dx' = V_{\text{H}}(x) + V_{\text{ext}}(x, R).
\]

Consequently, the problem of determining the ground state electron density and the equilibrium positions of the nuclei can now be expressed as the saddle point problem [19]

\[
\inf_{\psi \in (H_0^1(\Omega))^N} \sup_{R \in [0, \text{LIM}]} \mathcal{E}(\psi', R) = \inf_{R \in [0, \text{LIM}]} \sup_{\psi \in (H_0^1(\Omega))^N} \mathcal{L}(\psi', R, \phi)
\]

subject to the constraints

\[
\int \psi_{\alpha}^{(i)}(x) \psi_{\beta}^{(j)}(x) dx = \delta_{ij}, \quad \alpha \in \{\alpha, \beta\}, \quad i, j = 1, 2, \ldots, N_e,
\]

where

\[
\mathcal{L}(\psi', R, \phi) = -\frac{1}{2} \sum_{\sigma} \sum_{i=1}^{N_e} \int \psi_{\sigma}^{(i)}(x) \nabla^2 \psi_{\sigma}^{(i)}(x) dx + E_{\text{ext}}(\rho_\sigma, \rho_\beta) - \frac{1}{8\pi} \int |\nabla \phi(x, R)|^2 dx + \int (\rho(x) + b(x, R)) \phi(x, R) dx.
\]

This variational problem is indeed equivalent to the commonly solved eigenvalue problem [19,25]

\[
\mathcal{H}_\sigma \psi_{\sigma}^{(i)} = E_{\sigma} \psi_{\sigma}^{(i)}, \quad \sigma \in \{\alpha, \beta\}, \quad i = 1, 2, \ldots, N_e,
\]

where

\[
\mathcal{H}_\sigma = -\frac{1}{2} \nabla^2 + V_{\text{ext}}(x, R) + V_{\text{H}}(x) + \frac{\delta E_{\text{ext}}(\rho_\sigma, \rho_\beta)}{\delta \rho_\sigma(x)}
\]

and the choice of one over the other is strictly a matter of convenience.

2.1. Pseudopotential approximation

The motivation for the pseudopotential approximation is as follows. The core states are localized in the vicinity of the nucleus leading to oscillations of the valence wavefunctions in this region. Irrespective of the basis set used, a large number of basis functions are required to capture these oscillations. Additionally, the tightly bound core electrons are chemically inactive and hence have a negligible contribution towards determining physical properties. In view of this, the core electrons are eliminated and an effective nuclear potential is introduced to describe the effect of the core electrons, resulting in nodeless pseudo-wavefunctions. This amounts to replacing the all-electron potential \( V_{\text{ext}}(x, R) \) in Eq. (20) or equivalently the all-electron energy \( E_{\text{ext}}(\rho, R) \) with an effective potential \( V_{\text{PS}}^{\text{ext}}(x, R) \). The pseudopotentials can be broadly classified as either local or non-local based on their spatial dependence. The local pseudopotentials are explicit functions "V_{\text{PS}}^{\text{ext}}(x, R)" with ||V_{\text{PS}}^{ext}||_{L^\infty} \leq C. They can be incorporated into our formulation by letting \( b(x, R) = \frac{1}{4\pi} \nabla^2 V_{\text{PS}}^{\text{ext}} \). In contrast, non-local pseudopotentials are operators on the wavefunction with angular momentum dependence and are designed to accurately reproduce the scattering properties of the all-electron potential. These include norm conserving [28–30] and ultrasoft pseudopotentials [31], which are usually employed in the Kleinman–Bylander form [32].
\[ V^I_{\text{ion}}(\mathbf{x}, \mathbf{R}_j) \psi(\mathbf{x}) = V^I_{\text{loc}}(\mathbf{x}, \mathbf{R}_j) \psi(\mathbf{x}) + \sum_{lm} C^I_{lm} u^I_{lm}(\mathbf{x}, \mathbf{R}_j) \Delta V^I_j(\mathbf{x}, \mathbf{R}_j), \quad J \in \{1, 2, \ldots, M\}, \]  

(22)

where

\[ \Delta V^I_j(\mathbf{x}, \mathbf{R}_j) = V^I(\mathbf{x}, \mathbf{R}_j) - V^I_{\text{loc}}(\mathbf{x}, \mathbf{R}_j), \]  

(23)

\[ C^I_{lm} = \frac{\int_\Omega u^I_{lm}(\mathbf{x}, \mathbf{R}_j) \psi(\mathbf{x}) d\mathbf{x} \Delta V^I_j(\mathbf{x}, \mathbf{R}_j) u^I_{lm}(\mathbf{x}, \mathbf{R}_j) d\mathbf{x}}{\int_\Omega u^I_{lm}(\mathbf{x}, \mathbf{R}_j) \Delta V^I_j(\mathbf{x}, \mathbf{R}_j) u^I_{lm}(\mathbf{x}, \mathbf{R}_j) d\mathbf{x}}. \]  

(24)

\[ V^I(\mathbf{x}, \mathbf{R}_j) \] is the ionic pseudopotential component corresponding to the azimuthal quantum number \( l \). \( V^I_{\text{loc}}(\mathbf{x}, \mathbf{R}_j) \) is the local ionic potential and \( u^I_{lm}(\mathbf{x}, \mathbf{R}_j) \) represents the pseudo-wavefunction for the valence states of interest, all for a single atom. The superscript \( J \) is for the atom number and the subscript \( m \) denotes the magnetic quantum number. Therefore

\[ V^P_{\text{ext}}(\mathbf{x}, \mathbf{R}) \psi(\mathbf{x}) = \sum_{j=1}^M V^I_{\text{loc}}(\mathbf{x}, \mathbf{R}_j) \psi(\mathbf{x}) + \sum_{j=1}^M \sum_{lm} C^I_{lm} u^I_{lm}(\mathbf{x}, \mathbf{R}_j) \Delta V^I_j(\mathbf{x}, \mathbf{R}_j). \]  

(25)

Let us redefine \( V_{\text{ext}}(\mathbf{x}, \mathbf{R}) = \sum_{j=1}^M V^I_{\text{loc}}(\mathbf{x}, \mathbf{R}_j) \), for which we can obtain the corresponding \( b(\mathbf{x}, \mathbf{R}) \). The energy functional for the pseudopotential approximation can then be written as

\[ \mathcal{E}^P(\Psi, \mathbf{R}) = \mathcal{E}(\Psi, \mathbf{R}) + \mathcal{K}(\Psi, \mathbf{R}), \]  

(26)

where

\[ \mathcal{K}(\Psi, \mathbf{R}) = \sum_{\sigma} \sum_{i=1}^{N_{\sigma}} \sum_{j=1}^{M} \sum_{lm} \frac{1}{C^I_{lm}} \left| \int_\Omega f(x, \mathbf{R}_j) \psi_{\sigma}(x) d\mathbf{x} \right|^2, \]  

(27)

\[ f(x, \mathbf{R}_j) = u^I_{lm}(x, \mathbf{R}_j) \Delta V^I_j(x, \mathbf{R}_j), \]  

(28)

\[ C^I_{lm} = \int_\Omega u^I_{lm}(\mathbf{x}, \mathbf{R}_j) \Delta V^I_j(\mathbf{x}, \mathbf{R}_j) u^I_{lm}(\mathbf{x}, \mathbf{R}_j) d\mathbf{x}. \]  

(29)

\( N_{\sigma} \) is the number of electrons after the pseudopotential approximation, i.e., valence electrons.

### 3. Convex approximation scheme – max-ent basis functions

In this section, we introduce the convex approximation scheme based on max-ent basis functions [23]. Consider a set of distinct nodes \( Y = \{ \mathbf{x}_a, a = 1, \ldots, P \} \subset \mathbb{R}^3 \) whose convex hull is denoted by \( \Omega \). We are looking to create approximations to any function \( u : \Omega \rightarrow \mathbb{R} \) of the form

\[ u^p(\mathbf{x}) = \sum_{a=1}^{P} u_a p_a(\mathbf{x}), \]  

(30)

where \( p_a : \Omega \rightarrow \mathbb{R} \) are the basis/interpolation functions. We require that the basis functions satisfy the following constraints \( \forall \mathbf{x} \in \Omega \)

\[ \sum_{a=1}^{P} p_a(\mathbf{x}) = 1, \]  

(31)

\[ \sum_{a=1}^{P} p_a(\mathbf{x}) \mathbf{x}_a = \mathbf{x}, \]  

(32)

\[ p_a(\mathbf{x}) \geq 0, \quad a = 1, \ldots, P. \]  

(33)

Eqs. (31) and (32) ensure that we are able to exactly reproduce affine functions, akin to simplicial linear finite-element basis functions. The approximation scheme is called convex since it follows from Eqs. (31)–(33), that the basis functions at \( \mathbf{x} \in \Omega \) define a convex combination of nodes which evaluates to \( \mathbf{x} \).

Let us define by

\[ H(\mathbf{x}, \mathbf{p}) = -\sum_{a=1}^{P} p_a(\mathbf{x}) \log(p_a(\mathbf{x})), \]  

(34)

\[ U(\mathbf{x}, \mathbf{p}) = \int_\Omega \sum_{a=1}^{P} p_a(\mathbf{x}) h_a(\mathbf{x} - \mathbf{x}_a) d\mathbf{x}, \]  

(35)

the information entropy and the total width of the convex approximation scheme respectively. The vector of basis functions is given by \( \mathbf{p}(\mathbf{x}) = \{ p_1(\mathbf{x}), \ldots, p_P(\mathbf{x}) \} \) and \( h_a(\mathbf{x} - \mathbf{x}_a) \) denotes some generalized notion of distance between points \( \mathbf{x}, \mathbf{x}_a \). Since Eq. (35) does not involve any basis function derivatives, we can work with the pointwise width, using the functional...
\[
W(\mathbf{x}, \mathbf{p}) = \sum_{a=1}^{p} p_a \phi_a(\mathbf{x}) h_a(|\mathbf{x} - \mathbf{x}_a|).
\]

Ideally, we would want an approximation scheme that maximizes the entropy and minimizes the width. However, this is not always possible and therefore we look to find the Pareto optimum: a convex approximation scheme such that there is none better, by minimizing for fixed \( \mathbf{x}, \beta \in (0, \infty) \)
\[
F(\mathbf{x}, \mathbf{p}) = -H(\mathbf{x}, \mathbf{p}) + \beta W(\mathbf{x}, \mathbf{p})
\]
subject to the constraints (31)–(33). Thus, we end up with the Lagrangian
\[
\mathcal{L}(\mathbf{x}, \mathbf{p}, \lambda, \lambda_0) = -H(\mathbf{x}, \mathbf{p}) + \beta W(\mathbf{x}, \mathbf{p}) + \lambda_0 \left( \sum_{a=1}^{p} p_a - 1 \right) + \lambda \sum_{a=1}^{p} p_a (\mathbf{x}_a - \mathbf{x}),
\]
where \( \lambda_0 \in \mathbb{R} \) and \( \lambda \in \mathbb{R}^3 \) are Lagrange multipliers used to enforce the constraints given by Eqs. (31) and (32) respectively. As in [23], we differentiate between the interior points \( \mathbf{x} \in \text{int}(\Omega) \) and points on the boundary \( \mathbf{x} \in \text{bd}(\Omega) \).

3.1. Interior points

**Definition 1.** The partition function \( Z : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R} \) associated with the node set \( \text{Y} \) is
\[
Z(\mathbf{x}, \lambda) \equiv \sum_{a=1}^{p} \exp(-\beta h_a(|\mathbf{x} - \mathbf{x}_a|)) + \lambda . (\mathbf{x} - \mathbf{x}_a).
\]

**Proposition 2.** Suppose affine \( \text{Y} = \mathbb{R}^3 \) and \( \mathbf{x} \in \text{int}(\Omega) \). Then the solution of minimizing (37) subject to the constraints (31)–(33) is
\[
p_a(\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \lambda^*(\mathbf{x}))} \exp(-\beta h_a(|\mathbf{x} - \mathbf{x}_a|)) + \lambda^*(\mathbf{x}).(\mathbf{x} - \mathbf{x}_a), \quad a = 1, \ldots, p,
\]
where
\[
\lambda^*(\mathbf{x}) = \arg\min_{\lambda \in \mathbb{R}^3} \log Z(\mathbf{x}, \lambda).
\]
Additionally, the minimizer \( \lambda^*(\mathbf{x}) \) is unique.

**Proof.** It follows from the Karush–Kuhn–Tucker conditions
\[
\beta h_a + \log(p_a) + 1 + \lambda^* \beta h_a + \lambda^*. (\mathbf{x} - \mathbf{x}_a) = 0, \quad a = 1, \ldots, p.
\]

Therefore
\[
p_a = \exp(-\beta h_a + \lambda^*. (\mathbf{x} - \mathbf{x}_a) - \lambda^* - 1) \quad \forall \mathbf{x} \in \text{int}(\Omega).
\]
The optimal Lagrange multipliers \( \lambda^*_0 \) and \( \lambda^* \) are the maximizers of the Lagrange dual function
\[
g(\lambda_0, \lambda) = \inf_{\mathbf{p} \in \mathcal{P}} \mathcal{L}(\mathbf{p}, \lambda_0, \lambda) = -\lambda_0 - \sum_{a=1}^{p} \exp(-\beta h_a + \lambda . (\mathbf{x} - \mathbf{x}_a) - \lambda_0 - 1).
\]

On maximizing this dual function with respect to \( \lambda_0 \), we obtain
\[
Z(\mathbf{x}, \lambda) = \exp(\lambda^*_0 + 1).
\]
Substituting Eq. (45) into (44), we obtain the reduced Lagrange dual function
\[
\tilde{g}(\lambda) = -\log Z(\mathbf{x}, \lambda),
\]
from which we can conclude that if \( \mathbf{x} \in \text{int}(\Omega) \), then the shape functions are given by Eqs. (40) and (41). The existence of minimizer of \(-\tilde{g}(\lambda)\) is guaranteed by the Kuhn–Tucker theorem. The uniqueness follows from the strict convexity of \(\log Z(\mathbf{x}, \lambda)\), which can easily be verified. \( \square \)

3.2. Boundary points

The treatment of boundary points \( \mathbf{x} \in \text{bd}(\Omega) \) can be reduced to the problem analyzed in the preceding section by exploiting the reduced face property of convex approximation schemes [23].
The adaptability of the basis functions stems from the flexibility provided by the choice of $h_a(x - x_d)$ and $\beta(x)$. Let us first discuss the effect of the parameter $\beta(x)$. The support size of the basis functions, which can also be optimized in a variational approach [33], is determined by $\beta(x)$. Reducing $\beta$ increases the support of the basis functions and in general the accuracy of the calculations [23]. However, it comes at the cost of higher order quadrature rules and increased band-width of the equations. Alternatively, increasing $\beta$ increases the locality of the basis functions at the price of reduced accuracy. Detailed examples and illustrations can be found elsewhere [23].

Further adaptivity is provided by the choice of $h_a(x - x_d)$. By choosing $h_a(x - x_d) = |x - x_d|$, we obtain long range exponential decay and Kato like cusp condition at short range, similar to Slater orbitals. As a consequence, these basis functions could be particularly useful for all-electron calculations, where such a behavior is expected. This is indeed confirmed by results obtained in Section 5. However, in pseudopotential calculations, where the solution is expected to be smooth, $h_a(x - x_d) = |x - x_d|^2$ is a more appropriate choice. In this case, the resulting basis functions will have Gaussian type behavior, similar to Gaussian orbitals used in electronic structure calculations. It is also possible to vary $h_a(x - x_d)$ within the node set, which allows for the seamless transition between basis functions with different properties and behavior. For example, it is possible to easily transition between Gaussian-type basis functions and linear finite-elements within the domain [23].

In our previous work with finite-elements for KS-DFT [19], we have proved the convergence of the finite-element approximation using the technique of $\Gamma$-convergence. For this to be extended to the approximation scheme presented here, the space generated by the basis functions needs to be dense in $H_0^1(\Omega)$. However, this result has so far only been proven for the choice $h_a(x - x_d) = |x - x_d|^2$ [34] and is still open for other choices of $h_a(x - x_d)$. Further, because of the nature of the basis functions, the convergence with numerical quadratures does not follow from previous work, and is still an open problem.

4. Numerical implementation

We now turn to the numerical implementation of the variational formulation described in Section 2. The variational problem (Eq. (17)) is discretized using the max-ent basis functions introduced in Section 3

$$
\psi_{ie}(x) = \sum_{j=1}^{p_{ie}} \psi_{ie}^j p_j(x),
$$

$$
\phi(x) = \sum_{j=1}^{p_{ie}} \phi_j p_j(x)
$$

to obtain

$$
\sum_{j=1}^{p_{ie}} \left[ \int_\Omega \left( \frac{1}{2} \nabla p_j(x) \nabla p_k(x) + V_{\text{eff}}^\sigma(x,R)p_j(x)p_k(x) \right) \right] \psi_{ie}^j = \sum_{q=1}^{p_{ie}} \sum_{s=1}^{p_{ie}} \int_\Omega \psi_{qs}^j \psi_{qs}^k \int_\Omega p_j(x)p_k(x)dx,
$$

$$
\sum_{j=1}^{p_{ie}} \left[ \int_\Omega \nabla p_j(x) \nabla p_j(x) \right] \phi_j = \int_\Omega (\rho_j(x) + b(x,R))p_j(x)dx,
$$

$$
\sum_{s=1}^{p_{ie}} \sum_{r=1}^{p_{ie}} \left[ \int_\Omega p_s(x)p_r(x)dx \right] \psi_{rs}^\sigma \psi_{rs}^\sigma = \delta_{mn},
$$

where the Lagrange multipliers $\lambda_{qs}^\sigma$ are used to enforce the constraints (Eq. (18)). $V_{\text{eff}}^\sigma , \rho_j$ denote the discretized $V_{\text{eff}}^\sigma$, $\rho$, respectively and $p_{ie}$ represents the number of nodes in the node set. The equivalent generalized nonlinear eigenvalue problem corresponding to Eqs. (49) and (51) is given by

$$
A_\sigma \bar{\psi}_\sigma = eM \bar{\psi}_\sigma, \quad \sigma \in \{\alpha, \beta\}.
$$

Here,

$$
A_\alpha = \frac{1}{2} \int_\Omega \nabla p_j(x) \nabla p_j(x)dx + \int_\Omega V_{\text{eff}}^\sigma(x,R)p_j(x)p_j(x)dx,
$$

$$
M = \int_\Omega p_j(x)p_j(x)dx.
$$

and $\bar{\psi}_\sigma$ is a vector of the nodal contribution to the wavefunction as given by Eq. (30). The self-consistent field (SCF) method is commonly used for solving Eqs. (50) and (52).

The Newton–Raphson method is an attractive option for solving the nonlinear system of Eqs. (49)–(51) because of the quadratic rate of convergence achieved in the vicinity of the solution. However, the convergence rate is asymptotic and convergence is assured only if the initial guess lies within the radius of convergence of the solution. On the other hand, the SCF method is an expensive proposition on account of Eq. (52) being a generalized eigenvalue problem.
Keeping the above discussion in mind, we employ the following methodology to solve the Kohn–Sham problem. We position the nodes at a geometric coarsening rate away from the nuclei to enhance the efficiency of the calculations. We enforce zero Dirichlet boundary conditions for both the wavefunctions and electrostatic potential by utilizing the Kronecker-delta property satisfied by the basis functions on the boundary of $X$. We incorporate the best attributes of the SCF and Newton–Raphson methods into our solution procedure along the lines of our previous work [19]. We rewrite Eq. (52) as

$$\left( A^\sigma - \eta M \right)^{-1} M \tilde{\Psi}^\sigma = \tilde{\epsilon} \tilde{\Psi}^\sigma,$$

where $\eta$ is the shift parameter and $\tilde{\epsilon} = \frac{1}{\lambda^2}$, thereby magnifying the region of interest in the eigenvalue spectrum resulting in significant increase in the rate of convergence without incurring any additional cost. We use a low resolution node set $Y_0$ to solve Eqs. (50) and (55) using the SCF method. Specifically, we employ the implicitly restarted Lanczos method [35] for solving the linear eigenvalue problem and the conjugate gradient algorithm as the linear solver. Since $Y_0$ has low resolution, the computational expense involved for this step is minimal. Next, we generate $Y_1$ which is a refinement of the node set $Y_0$. The solutions obtained on $Y_0$ are transferred to $Y_1$ using the basis functions of $Y_0$. This serves as a starting guess for the solution of Eqs. (49)–(51) simultaneously using Newton’s method with Goldstein-Armijo line-searches [36]. For the associated linear solver we use the generalized minimal residual method (GMRES, [37]) while retaining the option of switching to either the bi-conjugate gradient stable method (Bi-CGSTAB, [38]) or transpose-free quasi-minimal residual method (TFQMR, [39]). Since the Hessian information is required only through its product with a vector, we evaluate it by using directional derivative quotients [40,41] which is significantly faster than evaluating the exact Hessian. The quality of the initial guess generated by the procedure just described allows us to obtain convergence within exceedingly tight tolerances in no more than 6–7 iterations at a quadratic rate.

The equilibrium position of the nuclei are computed using Broyden–Fletcher–Goldfarb–Shanno (BFGS) method. Each force update following a displacement of the nuclei requires the re-evaluation of the electrostatic potential and wavefunctions within an internal loop. In order to expedite this re-evaluation, we use the solution from the previous configuration as an initial guess for the Newton–Raphson method. In order to verify convergence with respect to number of basis functions, we repeat the calculations on increasingly finer node sets $Y_n$ and the solution on the parent node set is used as starting guess for the Newton–Raphson method on the finer node set, which ensures rapid convergence. This process is repeated until convergence is achieved to within a prespecified tolerance. The recursive nature of the calculations is advantageous with respect to solution schemes that require a complete restart every time the discretization is modified, e. g., in plane wave basis calculations, the convergence has to be verified with respect to the energy cut-off and number of $k$-points sampled.

We evaluate all the integrals via Gaussian quadrature [42] on the Delaunay triangulation of $\Omega$ generated using the corresponding max-ent node set $Y_n$. The coarsening nature of the triangulation allows for efficient and accurate evaluation of the integrals. See Fig. 1 for examples of triangulations used for a single atom and dimer. At each quadrature point, we solve the minimization problem (Eq. (41)) using the Newton–Raphson method to obtain values of the shape functions and calculate their derivatives using the expressions derived in Appendix A. It should be noted that since an exact quadrature rule for the max-ent basis functions does not exist, we verify the convergence of the solution with respect to the order of the quadrature used. In order to reduce the error due to the inexact quadrature rule, we use the same node sets for evaluating the energy of both the cluster of atoms as well as the individual component atoms. This results in significant quadrature error cancellation.

Finally, the implementation is parallelized as follows. We decompose the cubical domain $\Omega$ into cubical subdomains $\Omega^i$ and correspondingly the node set $Y$ into subsets $Y_i$. In order to satisfy the conforming patches property [23], the node sets $Y, Y^i$

![Fig. 1](image-url). Examples of sliced background triangulations used for numerical integration.
should be chosen such that \( \Omega' \) is the convex hull of \( Y_i \). Each node set \( Y_i \) and corresponding subdomain \( X_i \) is now handled independently by a different processor. Since the basis functions for \( Y_i \) satisfy the Kronecker-delta property on the boundary of \( \Omega' \) [23], inter-processor communication is required only for the nodes on the boundary. This is exactly the same situation when performing parallel computations using finite-elements via domain decomposition. Therefore, efficient parallel computations can be performed, akin to simplicial finite-elements.

5. Examples and results

In this section, we validate our implementation through a number of examples ranging from single atoms to small clusters of atoms. We do so for both the all-electron case as well as local and nonlocal pseudopotential approximations. For all-electron calculations, we choose \( h_d(\|x - x_i\|) = \|x - x_i\| \), which results in basis functions with the expected behavior of the solution inherent to them. For the pseudopotential calculations, we verify the obtained results with both \( h_d(\|x - x_i\|) = \|x - x_i\| \) and \( \frac{x}{C_0} x a_j \), respectively, where \( \gamma \) is a dimensionless constant and \( h \) is the representative spacing between the nodes.

5.1. All-electron calculations

First we perform all-electron single atom calculations for hydrogen, helium, lithium, beryllium and carbon. The results obtained have been compared with highly accurate calculations of Kotochigova et al. [43] in Table 1. We obtain very good agreement validating the accuracy of our implementation.

Fig. 2a and b provide a comparison of the convergence obtained using max-ent basis functions with simplicial linear finite-elements [19]. Note that the position of the nodes correspond to the optimal triangulations developed for the simplicial linear finite-elements [19]. \( E_0 \) denotes the converged energy and \( P_0 \) denotes the number of nodes required for the converged finite-element solution. For the helium, carbon atoms we have used 1250, 10,000 max-ent basis functions to achieve a converged solution. It is clear that there is a significant reduction in the number of basis functions required to achieve convergence, resulting in computational times differing by an order of magnitude, in spite of their non-locality and necessity for higher order quadrature rules. The relatively small number of basis functions make them a viable choice for performing medium sized all-electron calculations.

The next set of examples considered are the nitrogen (N\(_2\)) and carbon monoxide (CO) molecules. A comparison of the obtained binding energy and bond length with previous studies is provided in Tables 2 and 3. Again, the agreement is good.

<table>
<thead>
<tr>
<th>Element</th>
<th>Max-ent</th>
<th>Kotochigova et al. [43]</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>-0.445</td>
<td>-0.445</td>
</tr>
<tr>
<td>He</td>
<td>-2.830</td>
<td>-2.834</td>
</tr>
<tr>
<td>Li</td>
<td>-7.338</td>
<td>-7.335</td>
</tr>
<tr>
<td>Be</td>
<td>-14.434</td>
<td>-14.447</td>
</tr>
<tr>
<td>C</td>
<td>-37.432</td>
<td>-37.425</td>
</tr>
</tbody>
</table>

Table 1

Ground state energy of selected atoms (a.u.).

Fig. 2. Comparison between the convex approximation scheme and simplicial linear finite-elements for all-electron calculations.
5.2. Local pseudopotential approximation

In this section we present the results obtained using the local ‘Evanescent Core’ (EC) pseudopotential. First, we provide plots of convergence of the energy for the sodium atom using \( h_\alpha(|x - x_a|) = |x - x_a|, |x - x_a|^2 \) in Fig. 3a and b, respectively. For this convergence study, we have used uniformly spaced nodes with \( P_0 = 8000 \). Next, we consider dimers of lithium and sodium, the results of which are presented in Table 4. The \( \ast \) denotes the calculation where we neglect spin polarization for the evaluation of the single atom energy. The agreement is good and the difference from Nogueira et al. arises because of their use of the linear combination of atomic orbitals (LCAO) theory. Electron density contours on the mid-plane of Na\(_2\) has been plotted in Fig. 4a.

Next, we study octahedral clusters of lithium and sodium. Note that we do not try to optimize the geometry with respect to different arrangements of the atoms and we neglect spin polarization of the cluster to facilitate comparison with previous studies. Table 5 lists the binding energy and bond length so obtained. Again, the difference in results obtained by Nogueira et al. arises because of their use of LCAO theory. We plot electron density contours on the mid-plane of Na\(_6\) in Fig. 4b.

Finally, we study a cluster of 3\( \times 3\)\( \times 3 \) body centered cubic (BCC) unit cells of lithium with fixed geometry. The binding energy and lattice constant obtained are listed in Table 6. Contours of electron density on the mid-plane of this cluster of atoms can be seen in Fig. 5.

5.3. Nonlocal pseudopotential approximation

Norm-conserving pseudopotentials are attractive because of their accuracy, transferability and availability for all elements in the periodic table. Here, we look at the results obtained using the Troullier–Martins (TM) pseudopotential.

### Table 2

<table>
<thead>
<tr>
<th>Property</th>
<th>Max-ent</th>
<th>Engel et al. [44]</th>
<th>Suryanarayana et al. [19]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding energy (eV)</td>
<td>11.6</td>
<td>11.593</td>
<td>11.6</td>
</tr>
<tr>
<td>Bond length (a.u.)</td>
<td>2.06</td>
<td>2.068</td>
<td>2.06</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Property</th>
<th>Max-ent</th>
<th>Engel et al. [44]</th>
<th>Suryanarayana et al. [19]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding energy (eV)</td>
<td>13.0</td>
<td>12.967</td>
<td>13.03</td>
</tr>
<tr>
<td>Bond length (a.u.)</td>
<td>2.10</td>
<td>2.128</td>
<td>2.08</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Dimer</th>
<th>Property</th>
<th>Max-ent</th>
<th>Max-ent*</th>
<th>Suryanarayana et al. [19]</th>
<th>Nogueira et al. [46]</th>
<th>Fiolhais et al. [47]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li(_2)</td>
<td>Binding energy (eV/atom)</td>
<td>-0.49</td>
<td>-0.84</td>
<td>-0.49</td>
<td>-0.52</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Bond length (a.u.)</td>
<td>4.82</td>
<td>4.82</td>
<td>4.86</td>
<td>4.92</td>
<td>-</td>
</tr>
<tr>
<td>Na(_2)</td>
<td>Binding energy (eV/atom)</td>
<td>-0.36</td>
<td>-0.60</td>
<td>-0.35</td>
<td>-0.46</td>
<td>-0.59</td>
</tr>
<tr>
<td></td>
<td>Bond length (a.u.)</td>
<td>5.72</td>
<td>5.72</td>
<td>5.72</td>
<td>5.77</td>
<td>5.72</td>
</tr>
</tbody>
</table>

### Fig. 3. Convergence of energy of the sodium atom using the EC pseudopotential for different choices of \( h_\alpha(|x - x_a|) \).
Fig. 4. Contours of electron density on the mid-plane of sodium clusters using the EC pseudopotential.

Table 5
Binding energy and bond length for octahedral clusters of lithium and sodium using the EC pseudopotential.

<table>
<thead>
<tr>
<th>Dimer</th>
<th>Property</th>
<th>Max-ent</th>
<th>Max-ent*</th>
<th>Nogueira et al. [46]</th>
<th>Fiolhais et al. [47]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Binding energy (eV/atom)</td>
<td>−0.5</td>
<td>−0.85</td>
<td>−0.72</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>Bond length (a.u.)</td>
<td>5.66</td>
<td>5.66</td>
<td>5.79</td>
<td>−</td>
</tr>
<tr>
<td>Li₆</td>
<td>Binding energy/atom (eV/atom)</td>
<td>−0.42</td>
<td>−0.66</td>
<td>−0.53</td>
<td>−0.67</td>
</tr>
<tr>
<td></td>
<td>Bond length (a.u.)</td>
<td>6.78</td>
<td>6.78</td>
<td>6.87</td>
<td>6.31</td>
</tr>
<tr>
<td>Na₆</td>
<td>Binding energy/atom (eV/atom)</td>
<td>0.42</td>
<td>0.66</td>
<td>0.53</td>
<td>−0.67</td>
</tr>
</tbody>
</table>

Table 6
Binding energy and lattice constant of 3 × 3 × 3 BCC unit cells of lithium using the EC pseudopotential.

<table>
<thead>
<tr>
<th>Property</th>
<th>Max-ent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding energy (eV/atom)</td>
<td>−0.97</td>
</tr>
<tr>
<td>Lattice constant (a.u.)</td>
<td>6.40</td>
</tr>
</tbody>
</table>

Fig. 5. Contours of electron density on the mid-plane of 3 × 3 × 3 BCC unit cells of lithium obtained using the EC pseudopotential.
implemented in the Kleinman–Bylander (KB) [32] form. The pseudopotential parameters used for the present work are presented in Table 7. Note that we have not incorporated nonlinear core corrections here.

First, we present convergence plots for the lithium atom for $h_a(|x - x_n|) = |x - x_n|$. For this convergence study, we have used uniformly spaced nodes with $P_0 = 8000$.

Next, we present the binding energy and bond length of dimers of lithium and carbon in Tables 8 and 9 respectively. Again, we have excellent agreement with previous such calculations by Engel et al. [44].

6. Conclusions

We have developed a non-periodic, real-space, mesh-free convex approximation scheme for Kohn–Sham density functional theory. By including the electrostatic potential among the unknown fields, we have reformulated the problem as a local saddle-point problem which has been discretized using max-ent basis functions. We have developed a parallel implementation of this formulation capable of performing both all-electron and pseudopotential calculations. The formulation has been tested through a number of examples and the accuracy of the results is in accord with the literature.

We have shown that the use of max-ent basis functions can result in a significant reduction in the required number of basis functions compared to the corresponding simplicial linear finite-elements. However, this comes at the price of increased non-locality of the basis functions as well as the need for significantly higher order quadrature, which in some cases could outweigh the gains from the requirement of fewer basis functions. Also, as in most mesh-free methods, the necessity of a background mesh for performing numerical integration makes the method not completely mesh free. Further work in these aspects is required to make the method competitive with conventional bases used in electronic structure calculations.
Appendix A. Spatial derivatives of the shape functions

In this section we derive expressions for the spatial derivatives of the basis functions. We define the following quantities from Eqs. (39) and (46)

\[
\begin{align*}
\mathbf{f}_a(x, \lambda, \beta) &= -\beta \mathbf{h}_a(\mathbf{x} - \mathbf{x}_a) + \lambda \mathbf{x}, \\
\mathbf{p}_a(x, \lambda, \beta) &= \frac{\exp(\mathbf{f}_a(x, \lambda, \beta))}{\sum_{b=1}^{p} \exp(\mathbf{f}_b(x, \lambda, \beta))}, \\
\mathbf{r}(x, \lambda, \beta) &= -\frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} = \sum_{a=1}^{p} \mathbf{p}_a(x, \lambda, \beta)(\mathbf{x} - \mathbf{x}_a), \\
\mathbf{J}(x, \lambda, \beta) &= -\frac{\partial^2 \mathbf{g}(\lambda)}{\partial \lambda^2} = \sum_{a=1}^{p} \mathbf{p}_a(x, \lambda, \beta)(\mathbf{x} - \mathbf{x}_a) \otimes (\mathbf{x} - \mathbf{x}_a) - \mathbf{r}(x, \lambda, \beta) \otimes \mathbf{r}(x, \lambda, \beta).
\end{align*}
\]

Given a function \( h(x, \lambda, \beta) \), we define \( h^* = h(x, \lambda^*, \beta) \), where \( \lambda^* \) is given by Eq. (41). It follows that

\[
\nabla p_a^* = p_a^* \left( \nabla f_a^* - \sum_{b=1}^{p} p_b^* \nabla f_b^* \right).
\]

Using the chain rule we have

\[
\nabla f_a^* = \left( \frac{\partial f_a}{\partial x} \right)^* \mathbf{D} \lambda^*,
\]

where

\[
\begin{align*}
\left( \frac{\partial f_a}{\partial x} \right)^* &= -\beta \mathbf{h}_a(\mathbf{x} - \mathbf{x}_a) + \lambda^*(\mathbf{x}), \\
\left( \frac{\partial f_a}{\partial \lambda} \right)^* &= \mathbf{x} - \mathbf{x}_a.
\end{align*}
\]

To evaluate \( \mathbf{D} \lambda^* \), we note

\[
\mathbf{Dr} = \left( \frac{\partial \mathbf{r}}{\partial \lambda} \right)^* + \left( \frac{\partial \mathbf{r}}{\partial \lambda} \right)^* \mathbf{D} \lambda = 0,
\]

where

\[
\begin{align*}
\left( \frac{\partial \mathbf{r}}{\partial \lambda} \right)^* &= \mathbf{J}, \\
\left( \frac{\partial \mathbf{r}}{\partial x} \right)^* &= -\beta \sum_{a=1}^{p} \mathbf{p}_a \mathbf{h}_a \frac{(\mathbf{x} - \mathbf{x}_a) \otimes (\mathbf{x} - \mathbf{x}_a)}{\| \mathbf{x} - \mathbf{x}_a \|^2} + \mathbf{I}_d \equiv \mathbf{K}.
\end{align*}
\]

\( \mathbf{I}_d \) denotes the identity matrix. It follows

\[
\nabla f_a^* = -\beta \mathbf{h}_a \frac{(\mathbf{x} - \mathbf{x}_a)}{\| \mathbf{x} - \mathbf{x}_a \|^2} + \lambda^*(\mathbf{x}) - (\mathbf{x} - \mathbf{x}_a) \mathbf{J}^{-1} \mathbf{K}
\]

and

\[
\nabla p_a^* = p_a^* \left[ \beta \sum_{b=1}^{p} p_b \mathbf{h}_b \frac{(\mathbf{x} - \mathbf{x}_a)}{\| \mathbf{x} - \mathbf{x}_a \|^2} - \beta \mathbf{h}_a \frac{(\mathbf{x} - \mathbf{x}_a)}{\| \mathbf{x} - \mathbf{x}_a \|^2} - (\mathbf{x} - \mathbf{x}_a) \mathbf{J}^{-1} \mathbf{K} \right].
\]

References
