



Learning models for electron densities with Bayesian regression

Eric Schmidt, Andrew T. Fowler, James A. Elliott, Paul D. Bristowe*

Department of Materials Science & Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, United Kingdom



ARTICLE INFO

Keywords:

Bayesian linear regression
Relevance vector machine
Density functional theory
Embedded atom method
Genetic algorithm

ABSTRACT

The Hohenberg-Kohn theorems posit the ground state electron density as a property of fundamental importance in condensed matter physics, finding widespread application in much of solid state physics in the form of density functional theory (DFT) and, at least in principle, in semi-empirical potentials such as the Embedded Atom Method (EAM). Using machine learning algorithms based on parametric linear models, we propose a systematic approach to developing such potentials for binary alloys based on DFT electron densities, as well as energies and forces. The approach is demonstrated on the technologically important Al-Ni alloy system. We further demonstrate how ground state electron densities, obtained with DFT, can be predicted such that total energies have an accuracy of order meV atom^{-1} for crystalline structures. The set of crystalline structures includes a range of materials representing different phases and bonding types, from Al structures to single-wall carbon nanotubes.

1. Introduction

Density functional theory (DFT) exploits the Hohenberg-Kohn theorems by casting the ground state energy of an interacting system in an external field as a unique functional of its electron density [1]. This approach has been successfully used to obtain ground state properties of materials, discover unknown phases *via* methods such as AIRSS [2] and USPEX [3] and to perform high throughput calculations which systematically scan entire phase diagrams using methods such as AFLOW [4], AiiDA [5] and the materials project [6] among others. In addition to its ubiquitous use in DFT, the ground state electron density, $\eta_{\text{GS}}(\mathbf{r})$, also finds application in methods like Bader's atoms-in-molecules analysis and Laplacian-based critical points bond analysis [7]. Pure, or orbital-free (OF), DFT retains functionals of only the electron density, while the Kohn-Sham (KS) approach expresses the kinetic energy functional in terms of non-interacting single particle wave functions [8]. Both OF DFT and KS DFT methods minimize constrained expressions of the total energy until $\eta_{\text{GS}}(\mathbf{r})$ is found. The improved accuracy of the KS wavefunction-based kinetic energy term, however, comes at a significant computational cost, ultimately limiting the size of any system that can tractably be studied. The method described in this paper aims to circumvent this expensive self-consistent field (SCF) minimization and to approximate $\eta_{\text{GS}}(\mathbf{r})$ directly from a given structure. Throughout this work, we use CASTEP [9] and PROFESS [8] for the training and cross validation of linear models reproducing $\eta_{\text{GS}}(\mathbf{r})$ for KS and OF DFT calculations respectively.

To our knowledge very few attempts to rigorously regress $\eta_{\text{GS}}(\mathbf{r})$

from *ab initio* calculations have been made. Recently, Brockherde et al. [10] adopted a kernel-based approach to learn weights for a basis function representation of $\eta_{\text{GS}}(\mathbf{r})$ for H_2 and H_2O as well as small hydrocarbons. They cast the atomic structures into a scalar field, $v(\mathbf{r})$, consisting of squared exponentials and learned the basis function coefficients of their representation for $\eta_{\text{GS}}(\mathbf{r})$ given $v(\mathbf{r})$. Another attempt was made by Mitev et al. [11] who approximated $\eta_{\text{GS}}(\mathbf{r})$ using a non-linear parametric model based on the 3 s atomic orbitals for fcc Al. Although little quantitative measure of the error in $\eta_{\text{GS}}(\mathbf{r})$ was given, a surprising transferability was demonstrated in projections of $\eta_{\text{GS}}(\mathbf{r})$ along high symmetry crystal directions, with the electron density predictions, $\eta_{\text{pred}}(\mathbf{r})$, closely resembling $\eta_{\text{GS}}(\mathbf{r})$ of vacancy, divacancy and adatom defects, without these structures being present in the training set. While the Embedded Atom Method (EAM) potential given by Mitev et al. made reasonable predictions, the contents of its training set were not explicitly given and thus it is unclear if the transferability of the electron density prediction translates to a similar transferability of their potential.

Empirical methods to calculate the total energy of a system rely on forming approximations of the energy per atom $E_i(\mathbf{R}, \mathbf{r}_i, r_c)$ as some multivariate expression of the positions of the atoms under consideration \mathbf{R} and those surrounding the atom at \mathbf{r}_i to within some finite distance r_c . As might be expected, the more general the multivariate approximation to $E_i(\mathbf{R}, \mathbf{r}_i, r_c)$, the more accurate it can be, at the risk of overfitting. Rigorous attempts to approximate high-dimensional multivariate expressions for $E_i(\mathbf{R}, \mathbf{r}_i, r_c)$ have been seen with notable success particularly in the application of kernels which employ Gaussian approximation potentials [12,13] and deep neural networks [14]. Both

* Corresponding author.

E-mail address: pdb1000@cam.ac.uk (P.D. Bristowe).

methods use the atom positions and numbers as features in their map. Consider if instead, one were to use $\eta_{GS}(\mathbf{r})$, which we assume is already known. An empirical DFT-like map could be imagined with complete or partial integrals over the crystal cell. This idea in fact, is already embodied in existing classical potentials, such as the EAM, the Modified Embedded Atom Method (MEAM) [15] and Angular-Dependent Potentials [16], where usually physical electron densities of complete structures are not approximated explicitly. Mitev et al. noted this too and employ their approximation of $\eta_{GS}(\mathbf{r})$ to a novel EAM-like potential for fcc Al.

In this work, we demonstrate how parametric linear models with Bayesian regression can be used to approximate $\eta_{GS}(\mathbf{r})$ for a range of crystalline and high symmetry structures with different bonding types including: Al (fcc, bcc and hcp), NaCl, BaTiO₃, α -W & W₃O, H₂ crystal, graphite and single-wall carbon nanotubes (SWCNTs). The resulting approximation of $\eta_{GS}(\mathbf{r})$ for a range of systems leads to analytical functions that can be used to estimate $\eta_{GS}(\mathbf{r})$. Unlike conventional KS DFT, our model complexity scales linearly with the number of atoms and $\eta_{GS}(\mathbf{r})$ can straightforwardly be approximated at any arbitrary position \mathbf{r} . Finally, we demonstrate explicitly for the Al-Ni system that EAM potentials can be developed to accurately reproduce DFT energies and forces based on our regression of $\eta_{GS}(\mathbf{r})$ near atom cores.

2. Theory and algorithms

2.1. Linear models and $\eta(\mathbf{r})$

In the following we show how the “design matrix” Φ can be set up to regress electron densities using parametric linear models. Since the Bayesian treatment of regression provides a theoretical background to regularization and allows for model comparison [17], it will be discussed as well. Relevance vector machines (RVMs) [18] are one of a range of *linear* Bayesian models used for regression [19]. They are particularly useful since they allow for sparse solutions of large basis sets, scaling linearly in the number of observations, and can themselves be seen as a special case of the expectation-maximization algorithm [20].

At the heart of assuming linear models for the approximation of $\eta_{GS}(\mathbf{r})$ is that a linear superposition of functions can be used to describe the electron density in the first place. Specifically, and in contrast to Brockherde et al. [10], we consider multivariate functions of the atomic positions and atom species explicitly. Importantly, we construct multivariate functions that remain invariant to arbitrary rotations and translations of all atomic positions considered as contributing to $\eta_{GS}(\mathbf{r})$.

To remain in keeping with the common style of notation used in machine learning, we consider for now the ground state electron density at \mathbf{r} as the n th observation ρ_n in a larger set of density values. The aim of the regression with linear models is then to obtain a vector of weights \mathbf{w} , of size M , which allows for the calculation of the n th electron density value ρ_n given numerical features, \mathbf{x}_n , describing its neighborhood. \mathbf{x}_n may contain the spatial distances to the respective L neighboring atoms paired with their species or, generally, any other features. Using some basis set of non-linear functions, one can create a map $\phi(\mathbf{x}): \mathbb{R}^{3L} \mapsto \mathbb{R}^M = [\phi_1(\mathbf{x}), \dots, \phi_M(\mathbf{x})]$ to the M -dimensional *feature space*. The linear model can then be defined as $\rho_n = \phi(\mathbf{x}_n) \cdot \mathbf{w}$. If one includes uncertainty in the form of a noise parameter ε which is distributed normally as $\mathcal{N}(0, \beta)$ with noise level β , in the following denoted as $\varepsilon \sim \mathcal{N}(0, \beta)$, we find the *Bayesian linear model* as:

$$\rho_n = \phi(\mathbf{x}_n) \cdot \mathbf{w} + \varepsilon \quad (1)$$

Considering N observations, we can form the vector ρ . The neighboring information together with a chosen basis forms the *design matrix* Φ , where $\Phi_{nk} = \phi_k(\mathbf{x}_n)$. Distributing the k th weight normally as $w_k \sim \mathcal{N}(0, \alpha_k)$, we can constrain each basis function individually. This is at the core of the RVM as originally proposed by Tipping and Faul [18,21,22]. The first step in the iterative process is the calculation of

the weights, which can be obtained analytically because both the noise and the weight distributions are Gaussians. Given the new weights, the hyperparameters α and β are calculated using the type-II maximum likelihood. It can be shown that unique maxima regarding α can be utilized to update, remove or include basis functions [21,22].

To ensure translational and rotational invariance of $\phi(\mathbf{x})$, with respect to the reference frame of \mathbf{x} , we must choose appropriate values of ϕ_k . As in classical dynamics, we ensure translational invariance by adopting displacement vectors $\Delta \mathbf{r}_i = \mathbf{r}_i - \mathbf{r}$ between the current density point and the i th neighboring atom out of the total L neighboring atoms. $\Delta \mathbf{r}_i$ can be used to describe a linear combination of neighboring contributions to calculate the predicted density as:

$$\eta_{\text{pred}}(\mathbf{r}) = \underbrace{\sum_i^L f_i(\Delta \mathbf{r}_i)}_{2\text{-body}} + \underbrace{\sum_i^L \sum_{j>i}^L g_{ij}(\Delta \mathbf{r}_i, \Delta \mathbf{r}_j)}_{3\text{-body}} + \dots \quad (2)$$

where g_{ij} is symmetric with respect to permutation of i and j . Note that “ n -body” here refers to $n-1$ atoms and a single density point.

A first approximation with $\eta_{\text{pred}}(\mathbf{r})$ that is also invariant to rotation of the reference frame is to consider only isotropic or 2-body contributions from neighboring atoms using the magnitude of the displacement $\Delta \mathbf{r} = r$. To introduce rotationally invariant 3-body contributions, a map $(\Delta \mathbf{r}_i, \Delta \mathbf{r}_j): \mathbb{R}^6 \mapsto \mathbb{R}^3$ is adopted, where $(\Delta \mathbf{r}_i, \Delta \mathbf{r}_j) = (r_i, r_j, \theta_{ij})$ and $\theta_{ij} = \cos^{-1}\left(\frac{r_i r_j}{r_i r_j}\right)$. With this we formulate $\eta_{\text{pred}}(\mathbf{r})$ with $\eta_i^{2\text{-body}}(r_i)$ and $\eta_{ij}^{3\text{-body}}(r_i, r_j, \theta_{ij})$ as:

$$\eta_{\text{pred}}(\mathbf{r}) \approx \underbrace{\sum_i^L \eta_i^{2\text{-body}}(r_i)}_{\rho^{2\text{-body}}} + \underbrace{\sum_{i,j>i}^{L,L} \eta_{ij}^{3\text{-body}}(r_i, r_j, \theta_{ij})}_{\rho^{3\text{-body}}} \quad (3)$$

With a small modification for small distances where $r_i \approx 0$, the $\rho^{2\text{-body}}$ part is exactly the approximation made with the embedding density in the EAM. Adopting a cosine basis for $\eta_i^{2\text{-body}}(r_i)$ we have:

$$\rho^{2\text{-body}} = \sum_i^L \sum_{k=0}^{M_{\text{el}_i}-1} w_{\text{el}_i, k} \cos\left(\frac{\pi k}{r_c} r_i\right), \quad (4)$$

where M_{el_i} is the number of basis functions for element el corresponding to the i th neighbor, $w_{\text{el}_i, k}$ is the weight for that k th basis function and r_c is the 2-body cutoff distance. $\rho^{2\text{-body}}$ can be re-written grouping contributions by atom elements, denoting the neighboring atoms for element “ el ” as “ $\text{neigh}(\text{el})$ ”. Re-arranging Eq. (4) we find:

$$\rho^{2\text{-body}} = \sum_{\text{el} \in \text{element types}} \sum_{k=0}^{M_{\text{el}}-1} w_{\text{el}, k} \underbrace{\sum_{i \in \text{neigh}(\text{el})} \cos\left(\frac{\pi k}{r_c} r_i\right)}_{\equiv \phi_k(\mathbf{x}_{\text{el}})} \quad (5)$$

which, with $\Phi_{nk} = \phi_k(\mathbf{x}_n)$ for $\phi_k(\mathbf{x}_{\text{el}})$ leads to recovering the form of the linear models with $\mathbf{w}_{\text{el}} = [w_{\text{el}, 0}, \dots, w_{\text{el}, M_{\text{el}}}]$

$$\rho^{2\text{-body}} = \sum_{\text{el} \in \text{element types}} \phi(\mathbf{x}_{\text{el}}) \cdot \mathbf{w}_{\text{el}} \quad (6)$$

Stacking all elements and basis function types together we recover the original linear model. The 3-body part $\rho^{3\text{-body}}$ can be re-written in an analogous way grouping by atom element pairs instead resulting in:

$$\phi_{k_0, k_1, k_2}(\mathbf{x}_{\text{el}_a, \text{el}_b}) = \sum_{i, j \in \text{neigh}(\text{el})} \cos\left(\frac{\pi k_0}{r_{\text{cut}}} r_i\right) \cos\left(\frac{\pi k_1}{r_{\text{cut}}} r_j\right) \cos(k_2 \theta_{ij}) \quad (7)$$

with $\theta_{ij} = \cos^{-1}\left(\frac{r_i r_j}{r_i r_j}\right)$. To denote the smoothness of functions in their zeroth and first derivative towards cutoff points c , $y^{(0)}(x \geq c) = 0$ and $y^{(1)}(x \geq c) = 0$, we will use the notation $y^{(k)}(x \geq c) = 0$ if the function and its derivatives up to k are zero for $x \geq 0$. To ensure smoothness of functions in their zeroth and first derivative towards the cutoff point we multiply with $\psi(\tilde{x}) = \frac{\tilde{x}^4}{1 + \tilde{x}^4}$, where $\tilde{x} = \frac{x-c}{0.1}$, with $c = r_c$ (the distance cutoff) and $x = r$.

Download English Version:

<https://daneshyari.com/en/article/7957576>

Download Persian Version:

<https://daneshyari.com/article/7957576>

[Daneshyari.com](https://daneshyari.com)