

On DFT Molecular Simulation for Non-Adaptive Kernel Approximation

Maharavo Randrianarivony

Virtual Material Design, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss
Birlinghoven, Sankt Augustin, Germany
Email: randrian@ins.uni-bonn.de

Received 30 April 2014; revised 4 June 2014; accepted 11 June 2014

Copyright © 2014 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Using accurate quantum energy computations in nanotechnologic applications is usually very computationally intensive. That makes it difficult to apply in subsequent quantum simulation. In this paper, we present some preliminary results pertaining to stochastic methods for alleviating the numerical expense of quantum estimations. The initial information about the quantum energy originates from the Density Functional Theory. The determination of the parameters is performed by using methods stemming from machine learning. We survey the covariance method using marginal likelihood for the statistical simulation. More emphasis is put at the position of equilibrium where the total atomic energy attains its minimum. The originally intensive data can be reproduced efficiently without losing accuracy. A significant acceleration gain is perceived by using the proposed method.

Keywords

DFT, Energy, Stochastic, Covariance, Hyperparameter

1. Introduction

Nanotechnology tends to be the technology of focus in this century [1]. Many of the macroscopically visible continuum applications seem to have attained nearly optimal state in computer implementation as well as in theoretical analysis. Not only the analytical and numerical treatments of nanotechnology are interesting but it has real-world applications in various disciplines including aircraft, automobile, electronic and medical engineering. This work is partially a continuation of our former studies in [2]-[4] but we consider here electronic structures instead of solute-solvent interactions. In this paper, we would like to concentrate on the efficient treatment of quantum information. We present a method which is efficient for noisy data that contain measure-

ment imperfection because quantum data measurements are not completely ab-initio. Among other methods, DFT (Density Functional Theory) contains some parameters which are obtained from fitting procedures. Our motivation is to generate a system which is both accurate and fairly inexpensive to evaluate. Our method applies to all sorts of quantum information but we focus on the atomic energy in this paper. The availability of efficient and accurate evaluations can be used for subsequent applications. The region where the system admits its equilibrium is examined more carefully than the remaining region during the approximation. That is, a geometry optimization must be performed initially in order to determine the position where the energy attains its minimum. The most general setup enables an estimation of an unknown function in arbitrary dimension such that noisy sample functions together with their higher derivatives are provided. In this document, we treat the simplified version where only the sample values are provided. Depending on the desired accuracy and the size of the atomic system, the determination of the energy can last several minutes till several hours by using direct DFT computations. The molecular configurations consist of germanium, silicon or composites of them using several space group symmetries. The function to be reconstructed is the energy surface of a configuration when some tensor transformations are applied to DFT. The possible transformations include isotropic stretching, anisotropic stretchings and Voigt tensor. Our numerical experiments reveal that the proposed statistical pre-fitting process seems to be a reasonable approach to obtain a good accuracy in an affordable computing time. The preparation overhead lasts very long but it needs only to be performed once for all and its output can be stored. The acceleration factor between the direct DFT and the kernel based approximation is approximately of order 10^5 . Due to this significant acceleration gain, the preparation process shows to be worth calculating.

2. Theoretical Methodology

2.1. Ground State Energy

For a configuration of nuclei $\{\mathbf{a}_j\}$ of number N_u , the main problem in electronic structure computation involves the general Hamiltonian operator. We denote the coordinates of the i -th electron by $\mathbf{x}_i = (x_{i,1}, x_{i,2}, x_{i,3})$ and $\Delta_{\mathbf{x}_i} := \sum_{j=1}^3 \partial^2 / \partial x_{i,j}^2$. For the *Born-Oppenheimer* or adiabatic approximation, one assumes that the mass and the volume of the atoms are very large in comparison to those of the electrons. Thus, the atoms move comparatively slower than the electrons. As a consequence, one treats the time-independent Hamiltonian operator with respect to a set of nuclei \mathbf{a}_i which are supposed to be stationary. That is, the electronic structure is governed by the expression

$$\mathcal{H}_{\text{BO}} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{N_e} \Delta_{\mathbf{x}_i} + \sum_{i=1}^{N_e} \sum_{j=1}^{N_u} \frac{q_e Z_j}{\|\mathbf{x}_i - \mathbf{a}_j\|} + \frac{1}{2} \sum_{j=1}^{N_e} \sum_{i=1}^{j-1} \frac{q_e^2}{\|\mathbf{x}_i - \mathbf{x}_j\|}. \quad (1)$$

The above three terms are related to the kinetic energy, the atom-electron interaction and the inter-electron interaction while m_e , q_e , Z_j , \hbar are quantum parameters. The domain of computation is $\Omega \subset \mathbb{R}^3$ which is supposed to be sufficiently larger than the nuclei cloud $\{\mathbf{a}_j\}_{j=1}^{N_u}$ such that the above operator has neglecting influence beyond Ω . By considering the electron spins which take values $+1/2$ or $-1/2$, we are searching for the electronic wave function

$$\Psi : (\Omega \times \{-1/2, +1/2\})^{N_e} \rightarrow \mathbb{R} \quad \text{verifying the eigenproblem}$$

$$\mathcal{H}_{\text{BO}} \Psi(\gamma_1, \dots, \gamma_{N_e}) = E \Psi(\gamma_1, \dots, \gamma_{N_e}) \quad \text{where } \gamma_i = (\mathbf{x}_i, \sigma_i) \in \Omega \times \{-1/2, +1/2\} \quad (2)$$

such that Ψ is antisymmetric $\Psi(\dots, \gamma_i, \dots, \gamma_j, \dots) = -\Psi(\dots, \gamma_j, \dots, \gamma_i, \dots)$ for all distinct $i, j = 1, \dots, N_e$. The ground state energy corresponds to the smallest eigenvalue E_{\min} of (2). The anti-symmetrization operator \mathcal{A} applied to any N_e -variate function f is defined by

$$\mathcal{A}(f) := \frac{1}{N_e!} \sum_{P \in \Pi_{N_e}} \text{sgn}(P) f \circ P \quad (3)$$

in which Π_{N_e} is the set of permutations over $\{1, \dots, N_e\}$ and $\text{sgn}(P)$ designates the signature of a permu-

tation P . If f is a tensor product function as $f = \otimes_{i=1}^{N_e} f_i$, then the anti-symmetrizer coincides with the Slater determinant or wedge product

$$\mathcal{A} \left[\bigotimes_{i=1}^{N_e} f_i \right] (\gamma_1, \dots, \gamma_{N_e}) = \frac{1}{N_e!} \begin{vmatrix} f_1(\gamma_1) & f_1(\gamma_2) & \cdots & f_1(\gamma_{N_e}) \\ f_2(\gamma_1) & f_2(\gamma_2) & \cdots & f_2(\gamma_{N_e}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{N_e}(\gamma_1) & f_{N_e}(\gamma_2) & \cdots & f_{N_e}(\gamma_{N_e}) \end{vmatrix}. \quad (4)$$

The anti-symmetrization operator \mathcal{A} has the properties that it commutes with the Hamiltonian operator and that for an $N_e \times N_e$ matrix \mathbf{M} one has $\mathcal{A}(\mathbf{M}\Phi) = \det(\mathbf{M})\mathcal{A}(\Phi)$. The Hartree-Fock approach is the variational formulation on the Hamiltonian operator (1) where the trial functions are antisymmetric functions. The main difficulty is that the problem is of $3N_e$ -dimension without taking the electronic spins into account. In addition, on account of the antisymmetric property, a direct use of the Slater determinant often produces computations having orders of $\mathcal{O}(N_e!)$ which are very expensive. Counting the electronic spins leads usually to a factor of 2^{N_e} which makes the computation even more intractable. Some simplifications of the stationary Hamilton operators have already been proposed. For the DFT, one solves a set of equations for each electron. The similarity of the solutions is then derived from the theory of Kohn-Sham [5] which consists in replacing the complicated initial problem into several ones. For each $i = 1, \dots, N_e$

$$\left(-\frac{1}{2}\Delta + V_{\text{eff}}(\mathbf{x}) \right) \psi_i(\mathbf{x}) = E_i \psi_i(\mathbf{x}) \quad (5)$$

where V_{eff} is the effective potential energy which depends implicitly on the total electron density $\rho(\mathbf{x}) = \sum_{i=1}^{N_e} |\psi_i(\mathbf{x})|^2$ such that $V_{\text{eff}}(\mathbf{x}) = V_{\text{eff}}[\rho(\mathbf{x})](\mathbf{x})$. The problem (2) is then reduced from dimensions $3N_e$ to N_e sets of $3D$ smaller problems (5). The influence of one electron with respect to the other electrons is measured by the total electron density. These approaches enable the treatment of Hamiltonian problem even for an electronic structure having a large number of particles on a single desktop. The effective potential is constituted of the Hartree potential V_H , the exchange correlation potential V_{XC} and the external electrostatic field such as

$$V_{\text{eff}}[\rho(\mathbf{x})](\mathbf{x}) = V_H[\rho(\mathbf{x})](\mathbf{x}) + V_{\text{XC}}[\rho(\mathbf{x})](\mathbf{x}) + V_{\text{ext}}[\rho(\mathbf{x})](\mathbf{x}) \quad (6)$$

in which the Hartree potential is the inverse of the Poisson operator such as $\Delta V_H(\mathbf{x}) = -4\pi\rho(\mathbf{x})$. For its evaluation, either a Poisson problem is solved or one convolves with the Green fundamental solution such as $V_H(\mathbf{x}) = \int \rho(\tilde{\mathbf{x}}) / \|\mathbf{x} - \tilde{\mathbf{x}}\| d\tilde{\mathbf{x}}$. The exchange-correlation potential is some correction term [6] which is usually done by LDA (Local Density Approximation) or GGA (Generalized Gradient Approximation). Analytic expressions of the correlation energy are only known in a few special cases which mainly consist of the high and low density limits. The exchange-correlation potential is related to the exchange-correlation energy by $V_{\text{XC}} = \delta E_{\text{XC}} / \delta \rho$ where one expresses $E_{\text{XC}} = E_X + E_C$ as the exchange and the correlation parts. In term of the exchange-correlation energy density ε_{XC} one has

$$E_{\text{XC}}[\rho] = \int \varepsilon_{\text{XC}}[\rho](\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x} \quad \text{where} \quad \varepsilon_{\text{XC}} = \varepsilon_X + \varepsilon_C. \quad (7)$$

For the local density approximation (LDA), the exchange energy density is expressed as

$$\varepsilon_X^{\text{LDA}}(\rho) = -0.75(3\rho(\mathbf{x})/\pi)^{1/3} \quad \text{so that} \quad E_X[\rho] = -0.75 \left(\frac{3}{\pi} \right)^{1/3} \int \rho(\mathbf{x})^{4/3} d\mathbf{x}. \quad \text{Analytic values of the correlation}$$

energy density are only known for some extreme cases. The external electrostatic field potential V_{ext} is provided by the kernel $\sum_{i=1}^{N_u} z_i / \|\mathbf{x} - \mathbf{x}_i\|$. Once the solution E_i to (5) for all $i = 1, \dots, N_e$ becomes known, the Kohn-Sham approach uses the approximation to E of (2) by

$$E_{\text{KS}} = \sum_{i=1}^{N_e} E_i - \frac{1}{2} \int \rho(\mathbf{x}) V_H(\mathbf{x}) + \int \rho(\mathbf{x}) (E_{\text{XC}}[\rho(\mathbf{x})] - V_{\text{XC}}[\rho(\mathbf{x})]) d\mathbf{x}. \quad (8)$$

The main improvement from LDA to GGA is that the exchange-correlation energy does not depend only on the total electron density but also on its gradient such as $E_{\text{XC}}^{\text{GGA}} = E_{\text{XC}}^{\text{GGA}}[\rho, \nabla\rho](\mathbf{x})$. Because of the imperfections resulting from the estimation of parameters from experimental measurements, the DFT approach is not ab-initio as in the initial equations. But the statistical fitting approach proposed in this paper can deal with noisy data. The eigenvalue problem in (5) is nonlinear because its variational operator

$$\left\langle \Psi \left| -\frac{1}{2}\nabla^2 + V_{\text{eff}}[\rho^{(k)}](\mathbf{x}) \right| \Psi \right\rangle, \quad (9)$$

depends on ρ which in turn depends on ψ_i . It is solved by using a sequence of the linear eigenvalue problems SCF (Self Consistent Field). After assembling the operator (9), the linear eigenvalue problem is solved for the smallest eigenvalue $\lambda_{\text{min}}^{(k)}$. Then, the new iterate $\psi_i^{(k)}(\mathbf{x})$ is determined as the eigenfunction corresponding to $\lambda_{\text{min}}^{(k)}$. In practice, one represents Ψ as a set of basis [7] which are usually plane waves, Finite Element Method, or wavelets.

2.2. Kernel-Based Approximation

In this section, we will survey the main points about kernel-based approximation which are relevant in the quantum approximation. The most general setup of that approximation in any dimension D consists in accepting some inputs which are $\mathbf{x}_i \in \mathbb{R}^D$, $y_i \in \mathbb{R}$, $\mathbf{g}_i \in \mathbb{R}^D$ and $\mathbf{h}_i \in \mathbb{R}^{D \times D}$ such that

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, \quad \mathbf{x}_i \in \mathbb{R}^D, \quad i \in \mathcal{J} \quad (10)$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i) + \boldsymbol{\kappa}_i, \quad i \in \mathcal{N} \quad (11)$$

$$\mathbf{h}_i = \left[\frac{\partial f}{\partial x_p \partial x_q}(\mathbf{x}_i) \right]_{p,q=1}^D + \boldsymbol{\lambda}_i, \quad i \in \mathcal{M} \quad (12)$$

in which \mathcal{N} and \mathcal{M} are some subsets of \mathcal{J} while ε_i , $\boldsymbol{\kappa}_i$, $\boldsymbol{\lambda}_i$ are measurement imperfections. In general, using the gradients is computationally very expensive so that the set \mathcal{N} is much smaller than the set \mathcal{J} while the Hessians are even more restrictively used. In this paper, we use only data measurements such that $\mathcal{N} = \emptyset$ and $\mathcal{M} = \emptyset$. The correlation [8] function to be used is the Matern function

$$k_{\text{MATERN}}(r) := \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu r}}{\ell} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu r}}{\ell} \right) \quad (13)$$

where (ν, ℓ) are positive hyper-parameters and K_ν is the modified Bessel function. In dimension D , the spectral density of the Matern function is

$$S(s) = \frac{2^D \pi^{D/2} \Gamma(\nu + D/2) (2\nu)^\nu}{\Gamma(\nu) \ell^{2\nu}} \left(\frac{2\nu}{\ell^2 + 4\pi^2 s^2} \right)^{-(\nu + D/2)}. \quad (14)$$

If ν tends to infinity, one obtains the squared exponential case. For the special case of half-integers such as $\nu = p + 1/2$, one obtains a product of an exponential and a polynomial of order p such that (13) becomes

$$k_\nu(r) = \exp\left(-\frac{\sqrt{2\nu r}}{\ell}\right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^p \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8\nu r}}{\ell}\right)^{p-i}. \quad (15)$$

For the most applied cases where $\nu = 3/2$ and $\nu = 5/2$, one has

$$k_{3/2}(r) = \left(1 + \frac{\sqrt{3r}}{\ell}\right) \exp\left(-\frac{\sqrt{3r}}{\ell}\right), \quad (16)$$

$$k_{5/2}(r) = \left(1 + \frac{\sqrt{5r}}{\ell} + \frac{5r^2}{3\ell^2}\right) \exp\left(-\frac{\sqrt{5r}}{\ell}\right). \quad (17)$$

Suppose N is the number of observations such that $\text{Card}(\mathcal{J}) = N$ in (10). For noise-free observations, one has the covariance

$$\text{Cov}(f(\mathbf{x}_p), f(\mathbf{x}_q)) = K(\mathbf{x}_p, \mathbf{x}_q). \quad (18)$$

For a noisy data $y_i = f(\mathbf{x}_i) + \varepsilon$ where the additional noise follows the Gaussian distribution $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, the covariance becomes

$$\text{Cov}(\mathbf{y}_p, \mathbf{y}_q) = K(\mathbf{x}_p, \mathbf{x}_q) + \sigma_n^2 \delta_{pq}. \quad (19)$$

Let us denote the sampling points by $X = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$ which is a matrix having the size $N \times D$ where D is the dimension of the trial variables \mathbf{x}_i . The latent function and the set of observations are defined in a similar fashion such that $\mathbf{f} = [f_1, \dots, f_N]^T$ and $\mathbf{y} = [y_1, \dots, y_N]^T$. The marginal likelihood is the integral of the likelihood and the prior such as

$$p(\mathbf{y}|X) = \int p(\mathbf{y}|\mathbf{f}) p(\mathbf{f}|X) d\mathbf{f}. \quad (20)$$

We denote by $K = K(X, X)$ the covariance matrix whose entries are $K(\mathbf{x}_p, \mathbf{x}_q) = k_{\text{MATERN}}(r)$ where r will be defined below as a generalized distance between \mathbf{x}_p and \mathbf{x}_q . We will use also $|K|$ to denote the determinant of K . If $\mathbf{f}|X$ follows the Gaussian distribution $\mathcal{N}(\mathbf{0}, K)$ and one has no noise, then the marginal likelihood is

$$p(\mathbf{f}|X) = (2\pi)^{-N/2} |K|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{f}^T K^{-1} \mathbf{f}\right]. \quad (21)$$

By taking the logarithm, the computation of the marginal likelihood is as follows

$$p(\mathbf{f}|X) |K|^{1/2} (2\pi)^{N/2} = \exp\left[-\frac{1}{2} \mathbf{f}^T K^{-1} \mathbf{f}\right] \quad (22)$$

$$\log[p(\mathbf{f}|X)] + \frac{1}{2} \log|K| + \frac{N}{2} \log(2\pi) = -\frac{1}{2} \mathbf{f}^T K^{-1} \mathbf{f} \quad (23)$$

from which one obtains the log marginal likelihood

$$\log[p(\mathbf{f}|X)] = -\frac{1}{2} \mathbf{f}^T K^{-1} \mathbf{f} - \frac{1}{2} \log|K| - \frac{N}{2} \log(2\pi). \quad (24)$$

In the presence of noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ by using (19), the negative log marginal likelihood becomes

$$-\log[p(\mathbf{y}|X)] = \frac{1}{2} \mathbf{y}^T (K + \sigma_n^2 I_{N,N})^{-1} \mathbf{y} + \frac{1}{2} \log|K + \sigma_n^2 I_{N,N}| + \frac{N}{2} \log(2\pi). \quad (25)$$

The first term $\frac{1}{2} \mathbf{y}^T (K + \sigma_n^2 I_{N,N})^{-1} \mathbf{y}$ is the *data-fit term* while the second term $\frac{1}{2} \log|K + \sigma_n^2 I_{N,N}|$ and the last term $\frac{N}{2} \log(2\pi)$ are respectively the *complexity term* and the *normalization term*. For a set of test points $\mathbf{x}_i^* \in \mathbb{R}^D$, the expressions X^* , $K(X, X^*)$ and $K(X^*, X^*)$ are defined in a similar way as X and $K(X, X)$. The predicted estimation follows the Gaussian distribution

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I_{N,N} & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right). \quad (26)$$

That is to say, for given X and \mathbf{y} , the predictive distribution for a covariate vector X^* is a Gaussian admitting the following mean and variance

$$\mathbb{E}[\mathbf{f}^* | X^*, X] = K(X^*, X) [K(X, X) + \sigma_n^2 I_{N,N}]^{-1} \mathbf{y} \quad (27)$$

$$\mathbb{V}[\mathbf{f}^* | X^*, \mathbf{y}, X] = K(X^*, X^*) - K(X^*, X) [K(X, X) + \sigma_n^2 I_{N,N}]^{-1} K(X, X^*) \quad (28)$$

where those expressions are dependent on some set of hyper-parameters which we discuss below. The value of the generalized distance r for given $\mathbf{x}_p \in \mathbb{R}^D$ and $\mathbf{x}_q \in \mathbb{R}^D$ is

$$r^2 = (\mathbf{x}_p - \mathbf{x}_q)^T \begin{bmatrix} 1/\lambda_1^2 & & \\ & \ddots & \\ & & 1/\lambda_D^2 \end{bmatrix} (\mathbf{x}_p - \mathbf{x}_q) \quad (29)$$

in which $\lambda_1, \dots, \lambda_D$ are related to hyper-parameters. In general, the set of hyperparameters include $\theta = \log(\ell)$ where ℓ is from (16) and (17) as well as $\theta = \log(\lambda_i)$ where $i = 1, \dots, D$. The main objective is to determine the hyperparameters by optimizing the marginal likelihood. In practice, using the log marginal likelihood is more efficient to implement. In order to accelerate the speed of the nonlinear optimization, we need the gradients of the functional (25) with respect to the hyperparameters. For a hyper-parameter θ , one has the partial derivative

$$\frac{\partial}{\partial \theta} K^{-1} = -K^{-1} \frac{\partial K}{\partial \theta} K^{-1} \quad \text{and} \quad \frac{\partial}{\partial \theta} \log |K| = \text{trace} \left[K^{-1} \frac{\partial K}{\partial \theta} \right]. \quad (30)$$

From the chain rule, one obtains

$$\partial_\theta [k(r(\mathbf{x}_p, \mathbf{x}_q))] = k'(r) \partial_\theta [r(\mathbf{x}_p, \mathbf{x}_q)], \quad \text{where} \quad r(\mathbf{x}_p, \mathbf{x}_q) = \sqrt{\sum_{i=1}^D \frac{1}{\lambda_i} (x_{p,i} - x_{q,i})^2}. \quad (31)$$

This is singular when $\mathbf{x}_p = \mathbf{x}_q$ which occurs on the diagonal entries of the covariance matrix. In fact, one has

$$\partial_{\lambda_\mu} r(\mathbf{x}_p, \mathbf{x}_q) = -\frac{1}{r(\mathbf{x}_p, \mathbf{x}_q)} \left(\frac{x_{p,\mu} - x_{q,\mu}}{\lambda_\mu} \right)^2 \quad \text{for } \mu = 1, \dots, D. \quad (32)$$

Working with $R := r^2$ provides a regularization everywhere as $\partial_{\lambda_\mu} R(\mathbf{x}_p, \mathbf{x}_q)$ is regular. In addition, $\tilde{k}(R) := k_{\text{MATERN}}(\sqrt{r})$ is smooth at the origin because one has indeed $k(R) = p_{\text{MATERN}}(\sqrt{2\nu}r/\ell)$ such that

$$p_{\text{MATERN}}(t) := f_{\text{MATERN}}(t) \exp(-t) \quad (33)$$

$$p'_{\text{MATERN}}(t) = [f'_{\text{MATERN}}(t) - f_{\text{MATERN}}(t)] \exp(-t) \quad (34)$$

where f_{MATERN} is a polynomial. In particular, for the cases where $\nu = 3/2$ and $\nu = 5/2$ in (16) and (17), one has $f_{3/2} = 1+t$ and $f_{5/2} = 1+t+(1/3)t^2$. In fact, one obtains the relation

$$[f'_{\text{MATERN}}(t) - f_{\text{MATERN}}(t)] = tq_{\text{MATERN}}(t) \quad (35)$$

where q_{MATERN} is a polynomial. Therefore, one deduces from (35)

$$\frac{d\tilde{k}}{dR} = \frac{\sqrt{2\nu}}{2\ell\sqrt{R}} \left[f'_{\text{MATERN}} \left(\frac{\sqrt{2\nu}\sqrt{R}}{\ell} \right) - f_{\text{MATERN}} \left(\frac{\sqrt{2\nu}\sqrt{R}}{\ell} \right) \right] \exp \left(-\frac{\sqrt{2\nu}\sqrt{R}}{\ell} \right) \quad (36)$$

$$= \frac{\sqrt{2\nu}}{2\ell} q_{\text{MATERN}} \left(\frac{\sqrt{2\nu}\sqrt{R}}{\ell} \right) \exp \left(-\frac{\sqrt{2\nu}\sqrt{R}}{\ell} \right) \quad (37)$$

which is regular for all values of $r(\mathbf{x}_p, \mathbf{x}_q)$.

3. Simulation Results

In this section, we would like to report some results from computer simulation of the formerly described approach. First, we will present some results pertaining to general real valued multi-variate functions. That will be

followed by some application in quantum simulation.

The former theoretical approach was implemented by using C/C++, BLAS/LAPACK and NLOPT. The BLAS packet is used for the fast vector operations. We use LAPACK for the linear operations such as Cholesky factorization and dense matrix solvers. We use NLOPT for the nonlinear operations for both the geometry optimization and the optimal hyper-parameters in the log marginal likelihood (25). NLOPT supports diverse nonlinear optimization operations [9] in which local optimizers are involved. A local one searches only inside a neighborhood of a certain provided starting initial guess. The optimizers are performed by using derivative free or gradient based algorithm which are available in NLOPT. Derivative free algorithms include BOBYQA (Bond Optimization BY Quadratic Approximation), COBYLA (Constrained Optimization BY Linear Approximation), NEWUOA (NEW Unconstrained Optimization Algorithm). Gradient-based methods include MMA (Method of Moving Asymptotes) and LBFGB (Limited memory Broyden-Fletcher-Goldfarb-Shanno). The code is a very developed version of the matlab implementation provided in [8]. The new additional enhancements from the matlab version consist of the following features. First, using NLOPT provides a lot of improvements as compared to the original matlab nonlinear conjugate gradients. That can be observed when both the number of points and the dimension become large in the nonlinear optimization of the hyper-parameters. In addition, we can also accept higher derivatives in the input apart from the functional values. One can use the entire gradient or only some components of it. Furthermore, the gradient of the kernel-based approximation can be evaluated. That can be done analytically instead of using finite difference. In addition, our C/C++ code admits some python interface enabling direct application to ATK which is the quantum package we use.

As a first test, we consider the reconstruction of the function $f(x_1, \dots, x_D) = \sum_{i=1}^D \sin(x_i)$ by using the kernel-based approximation. Since that function does not present any special feature such as cusp or boundary layer or any special interesting region, we use only randomly generated points. The initial guess of the hyper-parameters is provided by the users. One considers the determination of the final hyper-parameters as an unconstrained nonlinear optimization. The results of the computations are collected in **Table 1** where the errors are computed for the function values as well as for the gradient values. The results exhibit the performance for different numbers of data for each dimension. In addition, we examine the effect of increasing the dimension D . In **Figure 1(c)**, we display the error plots in function of the number of data. We consider dimensions $D = 3, 4, 5, 6$ by analyzing the errors in the function evaluation. We observe that the increase of the dimension does not really deteriorate the accuracy too much. One needs certainly more points for higher dimensions but the increase of the number of additional points is not too significant.

For the application to nano-simulation, we consider the unit cell generated by three vectors $\mathbf{u}_A, \mathbf{u}_B, \mathbf{u}_C \in \mathbb{R}^3$. After a transformation $T_{(\lambda_1, \dots, \lambda_D)}$, they become $\tilde{\mathbf{u}}_A := T(\mathbf{u}_A)$, $\tilde{\mathbf{u}}_B := T(\mathbf{u}_B)$, $\tilde{\mathbf{u}}_C := T(\mathbf{u}_C) \in \mathbb{R}^3$. For a bulk configuration \mathcal{K} inside the unit cell, we denote by $T(\mathcal{K})$ the transformed configuration where the fractional coordinates belonging to the transformed unit cell $T(\mathcal{K})$ remain the same as in the original configuration \mathcal{K} . Their values are within $[0, 1]$ before and after the transformations. That is, the fractional coordinates of the reference configuration \mathcal{K} remain unchanged from beginning till the end of the computation. We are interested in the impact of the DFT energy by applying the transformation to the unit cell vectors. The transformation $T_{(\lambda_1, \dots, \lambda_D)}$ depends on D parameters so that the D -variate function to be approximated maps $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_D) \in \mathbb{R}^D$ to the energy such as

$$(\lambda_1, \dots, \lambda_D) \mapsto E = \text{DFT} \left[T_{(\lambda_1, \dots, \lambda_D)}(\mathcal{K}) \right].$$

The first transformation consists of an isotropic one that corresponds to a unidimensional function which is a scaling. That transformation amounts to stretching and confining the unit cell equally in all x , y , z directions. Thus, the function to be approximated maps $\boldsymbol{\lambda}$ into $\text{DFT} \left[T_{\boldsymbol{\lambda}}(\mathcal{K}) \right]$ where one has $\tilde{\mathbf{u}}_A := \lambda \mathbf{u}_A$, $\tilde{\mathbf{u}}_B := \lambda \mathbf{u}_B$, $\tilde{\mathbf{u}}_C := \lambda \mathbf{u}_C$. The next transformations are anisotropic in the sense that the unit cell is scaled differently in the x , y , z directions. The first anisotropic transformation scales in a way that $\tilde{\mathbf{u}}_A := \lambda_1 \mathbf{u}_A$, $\tilde{\mathbf{u}}_B := \lambda_2 \mathbf{u}_B$, $\tilde{\mathbf{u}}_C := \lambda_3 \mathbf{u}_C$. That transformation corresponds to the case where the dimension is $D = 3$ and $T_{(\lambda_1, \lambda_2, \lambda_3)}$ can be represented as a diagonal matrix having $(\lambda_1, \lambda_2, \lambda_3)$ as entries. For the 2D anisotropic case, the transformation is given by $T_{(\lambda_1, \lambda_2)}$ which has diagonal entries $(\lambda_1 \lambda_2, \lambda_2, \lambda_2)$. For the most general transformation, one utilizes a strain tensor

Table 1. Comparison w.r.t. $f(\mathbf{x}) = \sum_{i=1}^D \sin(x_i)$. Error of the function and gradient values on $[0, \pi]^D$ for dimension $D = 4, 5, 6$.

Dimension	Error in	Number of data				
		10	20	30	40	50
4	Function	1.4727e-01	2.4419e-02	2.3565e-02	1.6221e-02	1.4906e-02
	Gradient	4.7896e-01	1.1478e-01	1.0937e-01	8.6681e-02	7.7206e-02
5	Function	9.0570e-02	3.1160e-02	2.7457e-02	2.6331e-02	1.8296e-02
	Gradient	3.8708e-01	1.3668e-01	1.2931e-01	1.2187e-01	9.8637e-02
6	Function	1.0068e-01	3.5685e-02	2.7100e-02	2.0901e-02	1.8697e-02
	Gradient	3.8986e-01	1.5762e-01	1.2739e-01	1.1043e-01	1.0545e-01

$$T_{(\lambda_1, \dots, \lambda_6)} = T_{(\varepsilon_{xx}, \varepsilon_{xy}, \varepsilon_{xz}, \varepsilon_{yx}, \varepsilon_{yy}, \varepsilon_{yz}, \varepsilon_{zx}, \varepsilon_{zy}, \varepsilon_{zz})} = \begin{bmatrix} 1 + \varepsilon_{xx} & 0.5\varepsilon_{xy} & 0.5\varepsilon_{xz} \\ 0.5\varepsilon_{yx} & 1 + \varepsilon_{yy} & 0.5\varepsilon_{yz} \\ 0.5\varepsilon_{zx} & 0.5\varepsilon_{zy} & 1 + \varepsilon_{zz} \end{bmatrix}. \quad (38)$$

This is the Voigt strain tensor which is usually encountered in anisotropic transformations such as elasticity. This general transformation corresponds to a six-variate function since the Voigt strain matrix is supposed to be symmetric. The isotropic and 2D/3D anisotropic transformations are ensured to be non-singular, provided that the diagonal parameters are strictly positive. The tensor transformation in (38) is non-singular as long as the parameters $[\boldsymbol{\varepsilon}]$ are not too large as the Voigt tensor approaches the identity transformation. Although we are interested in the general approximation of the quantum energy, we put more emphasis on the neighborhood of the position where the energy attains its minimal value because it defines the equilibrium of the system. The main objective is to obtain a good approximation next to the optimal point $(\lambda_1^{\text{opt}}, \dots, \lambda_D^{\text{opt}})$ corresponding to

$$\operatorname{argmin}_{(\lambda_1, \dots, \lambda_D)} \operatorname{Energy} \left[T_{(\lambda_1, \dots, \lambda_D)}(\mathcal{K}) \right] \quad (39)$$

for a molecular or bulk configuration \mathcal{K} . We have built a function that performs the geometry optimization using unpolarized single ζ DFT calculators for any given molecular configuration. This is the first instance in the program where the optimizer NLOPT is applied. The Gaussian approximation are applied on a set of points $\boldsymbol{\lambda}_i = (\lambda_{1,i}, \dots, \lambda_{D,i})$ together with their images $E = \operatorname{DFT} \left[T_{(\lambda_{1,i}, \dots, \lambda_{D,i})}(\mathcal{K}) \right]$. In our applications, we accumulate very dense points in the neighborhood of the optimal value $\boldsymbol{\lambda}^{\text{opt}}$ while only few points are used elsewhere. An illustration of this situation is displayed in **Figure 1(a)** for the unidimensional case where the energy curve is traced together with the samples accumulating at the equilibrium. A typical sampling point distribution for the spatial case is depicted in **Figure 1(b)**. The implementation of the quantum application was realized by using ATK (Atomistix ToolKit) [10] [11] which consisted of python scripts. The ATK has some GUI extension well known as VNL (Virtual NanoLab). For given nuclei coordinates $\{\mathbf{a}_1, \dots, \mathbf{a}_{N_u}\}$, that quantum packet can provide the energy $E_{\min}(\mathbf{a}_1, \dots, \mathbf{a}_{N_u})$ such that $\mathcal{H}\Psi = E_{\min}\Psi$. As an example of using the ATK simulation, we observe the isosurface of the electron density function in **Figure 2** for a composite system of germanium/silicon in which only the atoms inside the unit cell are displayed. The displayed figure corresponds to $\text{Ge}_x\text{Si}_{1-x}$ where $x = 50\%$ in a germanium-silicon combination obtained from a diamond structure. After applying the approach described in the previous sections to several atomic configurations, we collect the major results on **Table 2**. It mainly summarizes the elapsed time for the preparation and the performance of the new method. In the tabulated outcomes, we consider first Body Centered Cubic of germanium. In addition, we use composites which are labeled $\text{Ge}_x\text{Si}_{1-x}$ where $x \in [0, 1]$ controls the amounts of germanium and silicon. More precisely, we consider the composites which are $\text{Ge}_{0.75}\text{Si}_{0.25}$, $\text{Ge}_{0.50}\text{Si}_{0.50}$ and $\text{Ge}_{0.25}\text{Si}_{0.75}$. All those composites are structured by using Face Centered Cubic as space group symmetry. The appropriate parameter values are obtained from the American Mineralogist Crystal Structure Database [12].

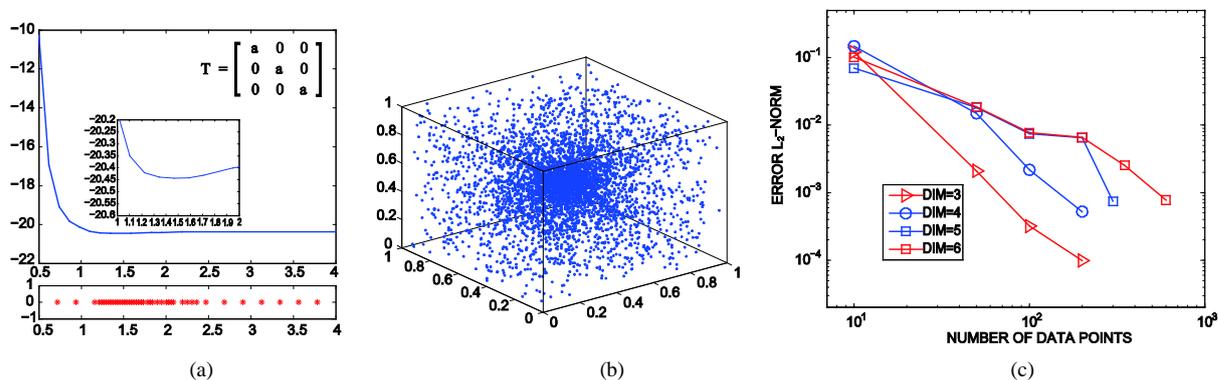


Figure 1. Accumulated sampling points at the geometric optimum: (a) Unidimensional; (b) Tridimensional; (c) Increasing the dimensions.

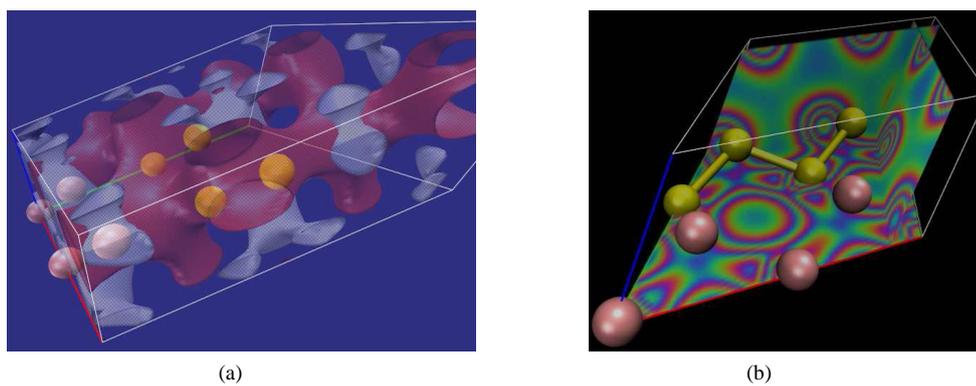


Figure 2. Electronic density for the DFT simulation of composite germanium/silicon.

The preparation step consists of a geometry optimization and the determination of the kernel approximation. The duration of the Gaussian kernel is dominated by the DFT computations related to the point samples. The numbers of point samples are 70, 105, 200 and 250 for isotropic, 2D anisotropic, 3D anisotropic and Voigt tensor respectively. The stochastic computation as described in section 2.2 is very fast. In fact, the application of stochastic simulation is at most 2 percent of the whole preparation. All the computations were performed with the DFT basis unpolarized single ζ which is the least intensive basis available in the implementation. If other bases were used, the time for the DFT computation would last even much longer. In the case of DFT double ζ polarized, the scaling of the computation intensity might be doubled. In all dimensions, the preparation expense has long durations. Depending on the tensor transformation and the number of sample points, the preparation overhead can last a few minutes till several days. But the output of those preparations can be stored so that they need only be computed once for all. The ratio between the direct DFT-evaluation and the evaluation using kernel approximation is displayed in the last column of the **Table 2**. It exhibits that in average, the orders of acceleration are respectively $2.62e-06$, $9.16e-06$, $3.00e-05$ and $4.22e-05$ for the isotropic, 2D anisotropic, 3D anisotropic and 6D tensor cases. Since the acceleration advantage is very good, investing on the preparation process is worth calculating as the results can be stored and subsequently post-processed. In order to observe the improvement of the accuracy, we consider two silicon configurations in **Figure 3**. As a matter of fact, the first configuration consists of silicon admitting a hexagonal lattice using space group $P6_3/mmc$. The second one admits a Face Centered Cubic lattice possessing the space group $Fd3m$. For both configurations, the bond lengths are obtained from the American Mineralogist Crystal Structure Database [12]. They are respectively represented by triangular and circular marks on the diagonal. That figure displays a comparison of the accuracy between the DFT computation and the kernel approximation. Marks which are closer to the diagonal identify good agreements between the two methods. We present only the case of anisotropic 3D transformation in this comparison because the other cases would exhibit similar results. The only difference between the two figures is that more sample points are used for the second one. More precisely, 45 sampling points are used in the first si-

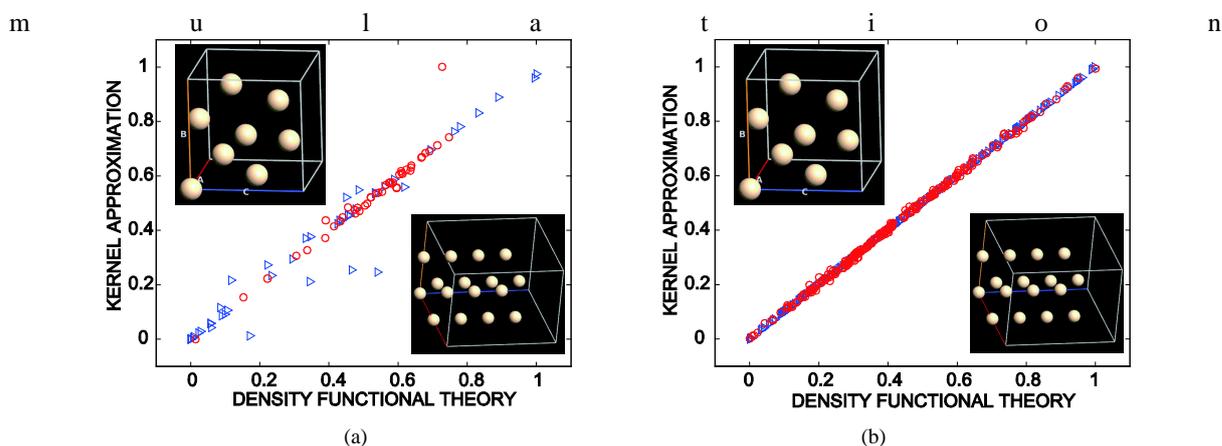


Figure 3. DFT vs. kernel approximation for two silicon configurations: (a) 45 sampling points; (b) 225 sampling points.

Table 2. Preparation overhead and acceleration gain.

CONFIG.	DIM	PREPARATION		EVALUATIONS		
		Geom. opt.	Kernel setup	Direct DFT	Kernel	Ratio
BCC Ge	1D	4.791 mn	58.788 mn	89.116 mn	0.0238 sc	4.45e-06
	2D	8.632 mn	78.076 mn	105.952 mn	0.1257 sc	1.97e-05
	3D	13.353 mn	134.019 mn	182.194 mn	0.5181 sc	4.73e-05
	6D	58.238 mn	6.986 hr	8.346 hr	0.9142 sc	3.04e-05
$Ge_{0.75}Si_{0.25}$	1D	57.627 mn	128.618 mn	172.634 mn	0.0272 sc	2.62e-06
	2D	67.437 mn	176.058 mn	229.228 mn	0.0776 sc	5.64e-06
	3D	139.958 mn	276.281 mn	363.696 mn	0.7345 sc	3.36e-05
$Ge_{0.50}Si_{0.50}$	6D	7.493 hr	107.328 hr	143.634 hr	0.8964 sc	1.73e-06
	1D	33.119 mn	137.369 mn	205.144 mn	0.0225 sc	1.82e-06
	2D	79.887 mn	161.299 mn	212.237 mn	0.0710 sc	5.57e-06
	3D	104.440 mn	279.807 mn	377.709 mn	0.4704 sc	2.07e-05
$Ge_{0.25}Si_{0.75}$	6D	7.896 hr	117.347 hr	133.213 hr	0.8978 sc	1.87e-06
	1D	43.675 mn	144.608 mn	201.094 mn	0.0193 sc	1.59e-06
	2D	71.893 mn	186.089 mn	233.715 mn	0.0807 sc	5.75e-06
	3D	159.021 mn	395.511 mn	498.650 mn	0.5624 sc	1.87e-05
	6D	7.962 hr	89.853 hr	110.410 hr	0.8994 sc	1.35e-04

while 225 are used for the second one. One can observe that the values on the diagonals become more precise as more sampling points are used. Additional sampling points can be used if a better accuracy is desired with the costs of having longer preparation overhead.

References

- [1] Hill, J., Subramanian, L. and Maiti, A. (2005) *Molecular Modeling Techniques in Material Sciences*. Taylor & Francis Group, Boca Raton.
- [2] Harbrecht, H. and Randrianarivony, M. (2011) Wavelet BEM on Molecular Surfaces: Solvent Excluded Surfaces. *Computing*, **92**, 335-364. <http://dx.doi.org/10.1007/s00607-011-0147-y>
- [3] Randrianarivony, M. (2013) On Space Enrichment Estimator for Nonlinear Poisson-Boltzmann. *American Institute of Physics*, 1558, 2365-2369.
- [4] Randrianarivony, M. (2013) Parallel Processing of Analytical Poisson-Boltzmann Using Higher Order FEM. In: Kle-

- ment, E.P., Borutzky, W., Fahringer, T., Hamza, M.H. and Uskov, V., Eds., *Proceeding of Conference "Parallel and Distributed Computing and Networks"*, ACTA Press, 455434, 1-6.
- [5] Kohn, W. and Sham, L. (1965) Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, **140**, 1133-11388. <http://dx.doi.org/10.1103/PhysRev.140.A1133>
- [6] Perdew, J. and Wang, Y. (1992) Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Physical Review B*, **45**, 13244. <http://dx.doi.org/10.1103/PhysRevB.45.13244>
- [7] Suryanarayana, P., Gavini, V., Blesgen, T., Bhattacharya, K. and Ortiz, M. (2010) Non-Periodic Finite-Element Formulation of Kohn-Sham Density Functional Theory. *Journal of the Mechanics and Physics of Solids*, **58**, 256-280. <http://dx.doi.org/10.1016/j.jmps.2009.10.002>
- [8] Rasmussen, C. and Williams, C. (2006) *Gaussian Processes for Machine Learning*. The MIT Press, Massachusetts.
- [9] Johnson, S. (2010) The NLOpt Nonlinear Optimization Package. <http://ab-initio.mit.edu/nlopt>
- [10] Atomistix ToolKit Version 13.8.0, QuantumWise A/S. www.quantumwise.com
- [11] Brandbyge, M., Mozos, J., Ordejon, P., Taylor, J. and Stokbro, K. (2002) Density-Functional Method for Nonequilibrium Electron Transport. *Physical Review B*, **65**, 165401. <http://dx.doi.org/10.1103/PhysRevB.65.165401>
- [12] American Mineralogist Crystal Structure Database. <http://rruff.geo.arizona.edu/AMS/>