

Numerical Methods for Kohn-Sham Density Functional Theory

Jianfeng Lu (鲁剑锋)

Duke University

jianfeng@math.duke.edu

January 2019, Banff Workshop on “Optimal Transport Methods in Density Functional Theory”

Reference

Lin Lin, L., and **Lexing Ying**, Numerical Methods for Kohn-Sham Density Functional Theory, *Acta Numerica*, 2019 (under review)

Electronic structure theory

We consider the electronic structure, which is given by the many-body electronic Schrödinger equation

$$\left(-\frac{1}{2} \sum_{i=1}^N \Delta_{x_i} + \sum_{i=1}^N V_{\text{ext}}(x_i) + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} \right) \Psi(x_1, \dots, x_N) = E\Psi$$

under non-relativistic and Born-Oppenheimer approximations. Here the atom types and positions enter V_{ext} as parameters.

P.A.M. Dirac, 1929

The fundamental laws necessary to the mathematical treatment of large parts of physics and the whole of chemistry are thus fully known, and the difficult lies **only** in the fact that application of these laws leads to equations that are **too complex to be solved**.

Electronic structure theory

We consider the electronic structure, which is given by the many-body electronic Schrödinger equation

$$\left(-\frac{1}{2} \sum_{i=1}^N \Delta_{x_i} + \sum_{i=1}^N V_{\text{ext}}(x_i) + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} \right) \Psi(x_1, \dots, x_N) = E\Psi$$

under non-relativistic and Born-Oppenheimer approximations. Here the atom types and positions enter V_{ext} as parameters.

Still **too complex to be solved**, even after almost 90 years:

- Curse of dimensionality;
- Anti-symmetry of Ψ due to Pauli's exclusion principle

$$\Psi(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = -\Psi(x_1, \dots, x_j, \dots, x_i, \dots, x_N);$$

- Ψ has complicated singularity structure.

Solving electronic Schrödinger equations:

- Tight-binding approximations (LCAO)
- Density functional theory
 - Orbital-free DFT
 - Kohn-Sham DFT
- Wavefunction methods
 - Hartree-Fock
 - Møller-Plesset perturbation theory
 - Configuration interaction
 - Coupled cluster
 - Multi-configuration self-consistent field
 - Neural network ansatz
- GW approximation; Bethe-Salpeter equation
- Quantum Monte Carlo (VMC, DMC, etc.)
- Density matrix renormalization group (DMRG) / tensor networks
- ...

Solving electronic Schrödinger equations:

- Tight-binding approximations (LCAO)
- Density functional theory

Orbital-free DFT

Kohn-Sham DFT

10 of 18 most cited papers in physics [Perdew 2010]

More than 50,000 citations on Google scholar

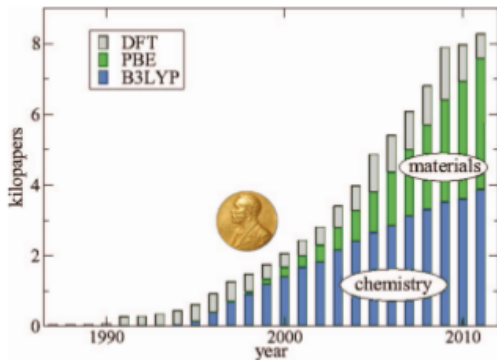


Figure: DFT papers count [Burke 2012].

Density functional theory [Hohenberg-Kohn 1964, Kohn-Sham 1965]: View energy as a functional of the one-body electron density $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$:

$$\rho(x) = N \int |\Psi|^2(x, x_2, \dots, x_N) dx_2 \cdots dx_N.$$

Density functional theory [Hohenberg-Kohn 1964, Kohn-Sham 1965]: View energy as a functional of the one-body electron density $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$:

$$\rho(x) = N \int |\Psi|^2(x, x_2, \dots, x_N) dx_2 \cdots dx_N.$$

Levy-Lieb variational principle [Levy 1979, Lieb 1983]:

$$E_0 = \inf_{\Psi} \langle \Psi, H\Psi \rangle = \inf_{\rho} \inf_{\Psi, \Psi \mapsto \rho} \langle \Psi, H\Psi \rangle$$

Density functional theory [Hohenberg-Kohn 1964, Kohn-Sham 1965]: View energy as a functional of the one-body electron density $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$:

$$\rho(x) = N \int |\Psi|^2(x, x_2, \dots, x_N) dx_2 \cdots dx_N.$$

Levy-Lieb variational principle [Levy 1979, Lieb 1983]:

$$E_0 = \inf_{\Psi} \langle \Psi, H\Psi \rangle = \inf_{\rho} \inf_{\Psi, \Psi \mapsto \rho} \langle \Psi, H\Psi \rangle = \inf_{\rho} E(\rho).$$

Density functional theory [Hohenberg-Kohn 1964, Kohn-Sham 1965]: View energy as a functional of the one-body electron density $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}_{\geq 0}$:

$$\rho(x) = N \int |\Psi|^2(x, x_2, \dots, x_N) dx_2 \cdots dx_N.$$

Levy-Lieb variational principle [Levy 1979, Lieb 1983]:

$$E_0 = \inf_{\Psi} \langle \Psi, H\Psi \rangle = \inf_{\rho} \inf_{\Psi, \Psi \mapsto \rho} \langle \Psi, H\Psi \rangle = \inf_{\rho} E(\rho).$$

The energy functional has the form

$$E(\rho) = T_s(\rho) + \int \rho V_{\text{ext}} + \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} + E_{\text{xc}}(\rho)$$

$T_s(\rho)$: Kinetic energy of non-interacting electrons;

$E_{\text{xc}}(\rho)$: **Exchange-correlation energy**, which encodes the many-body interaction between electrons (chemistry).

Kohn-Sham density functional theory

Kohn-Sham density functional theory introduces one-particle orbitals to better approximate the kinetic and exchange-correlation energies.

It is today the most widely used electronic structure theory, which achieves the best compromise between accuracy and cost.

The energy functional is minimized for N orbitals $\{\psi_i\} \subset H^1(\mathbb{R}^3)$.

$$E_{\text{KS}}(\{\psi_i\}) = \sum_{i=1}^N \frac{1}{2} \int |\nabla \psi_i|^2 + \int \rho V_{\text{ext}} + \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} + E_{\text{xc}}(\rho)$$

where (we consider “spin-less” electrons through the talk)

$$\rho(x) = \sum_{i=1}^N |\psi_i|^2(x).$$

Note that E_{KS} can be still viewed as a functional of ρ , implicitly.

Kohn-Sham density functional theory

$$E_{\text{KS}}(\{\psi_i\}) = \sum_{i=1}^N \frac{1}{2} \int |\nabla \psi_i|^2 + \int \rho V_{\text{ext}} + \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} + E_{\text{xc}}$$

The **exchange-correlation functionals** counts for the corrections from the many-body interactions between electrons:

- Semi-local exchange-correlation functionals

- Local density approximation: $E_{\text{xc}} = \int e_{\text{xc}}(\rho(x))$
- Generalized gradient approximation:

$$E_{\text{xc}} = \int e_{\text{xc}}(\rho(x), \frac{1}{2} |\nabla \sqrt{\rho(x)}|^2)$$

- Nonlocal exchange-correlation functionals

Kohn-Sham density functional theory

$$E_{\text{KS}}(\{\psi_i\}) = \sum_{i=1}^N \frac{1}{2} \int |\nabla \psi_i|^2 + \int \rho V_{\text{ext}} + \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} + E_{\text{xc}}$$

- Nonlocal exchange-correlation functionals

e.g., exact exchange + RPA correlation, $E_{\text{xc}} = E_x + E_c$:

$$E_x = -\frac{1}{2} \sum_{i,j=1}^N \iint \frac{\psi_i^*(x)\psi_j(x)\psi_i(y)\psi_j^*(y)}{|x-y|} dx dy;$$

$$E_c = \frac{1}{2\pi} \int_0^\infty \text{tr}[\ln(1 - \chi_0(i\omega)v) + \chi_0(i\omega)v] d\omega,$$

with v the Coulomb operator and χ_0 Kohn-Sham polarizability operator:

$$\chi_0(x, y, i\omega) = 2 \sum_j^{\text{occ}} \sum_k^{\text{unocc}} \frac{\psi_j^*(x)\psi_k(x)\psi_k^*(y)\psi_j(y)}{\epsilon_j - \epsilon_k - i\omega}.$$

Kohn-Sham density functional theory has a similar structure as a mean field type theory (at least for semilocal xc): The electrons interact through an effective potential.

The electron density is given by an effective Hamiltonian:

$$H_{\text{eff}}(\rho)\psi_i = \varepsilon_i\psi_i, \quad \rho(x) = \sum_i^{\text{occ}} |\psi_i|^2(x);$$

where the first N orbitals are occupied (aufbau principle). The effective Hamiltonian captures the interactions of electrons:

$$H_{\text{eff}}(\rho) = -\frac{1}{2}\Delta + V_{\text{eff}}(\rho);$$
$$V_{\text{eff}}(\rho) = V_c(\rho) + V_{\text{xc}}(\rho).$$

Note that this is a nonlinear eigenvalue problem. We can view it as a fixed-point equation for the density ρ :

$$\rho = F_{\text{KS}}(\rho).$$

Kohn-Sham map

Kohn-Sham fixed-point equation

$$\rho = F_{\text{KS}}(\rho),$$

where F_{KS} is known as the **Kohn-Sham map**, defined through the eigenvalue problem associated with $H_{\text{eff}}(\rho)$.

Given an effective Hamiltonian H_{eff} , we ask for its low-lying eigenspace: the range of the spectral projection

$$P = \chi_{(-\infty, \epsilon_F]}(H_{\text{eff}}),$$

thus ρ is given as the diagonal of the kernel of the operator P .

Kohn-Sham map

Kohn-Sham fixed-point equation

$$\rho = F_{\text{KS}}(\rho),$$

where F_{KS} is known as the **Kohn-Sham map**, defined through the eigenvalue problem associated with $H_{\text{eff}}(\rho)$.

Given an effective Hamiltonian H_{eff} , we ask for its low-lying eigenspace: the range of the spectral projection

$$P = \chi_{(-\infty, \epsilon_F]}(H_{\text{eff}}),$$

thus ρ is given as the diagonal of the kernel of the operator P .

It is often more advantageous to represent P in terms of Green's functions $(\lambda - H_{\text{eff}})^{-1}$, so to turn the eigenvalue problem into solving equations.

Let \mathcal{C} be a contour around the occupied spectrum, we have

$$P = \chi_{(-\infty, \epsilon_F]}(H_{\text{eff}}) = \frac{1}{2\pi i} \oint_{\mathcal{C}} (\lambda - H_{\text{eff}})^{-1} d\lambda.$$

Numerical methods

Kohn-Sham fixed-point equation

$$\rho = F_{\text{KS}}(\rho).$$

Usually solved numerically based on self-consistent field (SCF) iteration (alternative methods, such as direct minimization, are not as widely used)

→ see the talk of [Ziad Musslimani](#)

Numerical methods

Kohn-Sham fixed-point equation

$$\rho = F_{\text{KS}}(\rho).$$

Usually solved numerically based on self-consistent field (SCF) iteration (alternative methods, such as direct minimization, are not as widely used)

- Self-consistent iteration
Linearize the nonlinear problem to a linear one at each step
- Discretization
Making the problem finite dimensional
- Evaluation of the Kohn-Sham map
 - Formation of effective Hamiltonian: $\rho \mapsto H_{\text{eff}}(\rho)$
 - Evaluation of density: $H_{\text{eff}} \mapsto \rho$
Most of the actual computational work.

Self-consistent iteration

Solving for the fixed point equation $\rho = F_{\text{KS}}(\rho)$.

(Note: often more convenient to view the iteration in terms of V_{eff})

- First idea: Fixed point iteration

$$\rho_{n+1} = F_{\text{KS}}(\rho_n)$$

Usually does not converge as F_{KS} is not necessarily a contraction.

Self-consistent iteration

Solving for the fixed point equation $\rho = F_{\text{KS}}(\rho)$.

(Note: often more convenient to view the iteration in terms of V_{eff})

- First idea: Fixed point iteration

$$\rho_{n+1} = F_{\text{KS}}(\rho_n)$$

Usually does not converge as F_{KS} is not necessarily a contraction.

- Simple mixing

$$\begin{aligned}\rho_{n+1} &= (1 - \alpha)\rho_n + \alpha F_{\text{KS}}(\rho_n) \\ &= \rho_n - \alpha(\rho_n - F_{\text{KS}}(\rho_n))\end{aligned}$$

where $\rho_n - F_{\text{KS}}(\rho_n)$ is the residual.

$\alpha = 1$ corresponds to fixed-point iteration.

Self-consistent iteration

Solving for the fixed point equation $\rho = F_{\text{KS}}(\rho)$.

(Note: often more convenient to view the iteration in terms of V_{eff})

- First idea: Fixed point iteration

$$\rho_{n+1} = F_{\text{KS}}(\rho_n)$$

Usually does not converge as F_{KS} is not necessarily a contraction.

- Simple mixing

$$\begin{aligned}\rho_{n+1} &= (1 - \alpha)\rho_n + \alpha F_{\text{KS}}(\rho_n) \\ &= \rho_n - \alpha(\rho_n - F_{\text{KS}}(\rho_n))\end{aligned}$$

where $\rho_n - F_{\text{KS}}(\rho_n)$ is the residual.

$\alpha = 1$ corresponds to fixed-point iteration.

- Preconditioning

$$\rho_{n+1} = \rho_n - \mathcal{P}(\rho_n - F_{\text{KS}}(\rho_n))$$

e.g., $\mathcal{P} = \alpha I$ corresponds to simple mixing.

Kerker mixing for simple metals (jellium like)

→ see the talk of [Antoine Levitt](#)

- Newton method

$$\rho_{n+1} = \rho_n - \mathcal{J}_n^{-1}(\rho_n - F_{\text{KS}}(\rho_n))$$

where \mathcal{J}_n is the Jacobian matrix.

Locally quadratic convergence, but each iteration is quite expensive (involves iterative methods e.g., GMRES and finite difference approximation to Jacobian matrix)

- Newton method

$$\rho_{n+1} = \rho_n - \mathcal{J}_n^{-1}(\rho_n - F_{\text{KS}}(\rho_n))$$

where \mathcal{J}_n is the Jacobian matrix.

Locally quadratic convergence, but each iteration is quite expensive (involves iterative methods e.g., GMRES and finite difference approximation to Jacobian matrix)

- Quasi-Newton type algorithms (the most widely used approach)
 - Broyden's second method [Fang, Saad 2009; Marks, Luke 2008]
Low-rank update to $\hat{\mathcal{J}}_n^{-1}$ (proxy for inverse Jacobian)
 - Anderson mixing [Anderson 1965] (a variant of Broyden)
 - Direct Inversion of Iterative Subspace (DIIS) [Pulay 1980]
Minimize a linear combination of residual to get linear combination coefficients; other variants include E-DIIS [Cances, Le Bris 2000].
 - commutator-DIIS [Pulay 1983] (for nonlocal functionals)
Choice of residual as the commutator of $H_{\text{eff}}[P]$ and P

Discretization

- Large basis sets ($\mathcal{O}(100) \sim \mathcal{O}(10,000)$ basis functions per atom)
 - Planewave method (Fourier basis set / pseudospectral method) [Payne et al 1992, Kresse and Furthmüller 1996]
 - Finite element method [Tsuchida and Tsukada 1995, Suryanarayana et al 2010, Bao et al 2012]
 - Wavelet method [Genovese et al 2008]
 - Finite different method [Chelikowski et al 1994]
- Small basis sets ($\mathcal{O}(10) \sim \mathcal{O}(100)$ basis functions per atom)
 - Gaussian-type orbitals (GTO) (see review [Jensen 2013])
 - Numerical atomic-orbitals (NAO) [Blum et al 2009]
- Adaptive / hybrid basis sets ($\mathcal{O}(10) \sim \mathcal{O}(100)$ basis fctns per atom)
 - Augmented planewave method [Slater 1937, Andersen 1975]
 - Nonorthogonal generalized Wannier function [Skylaris et al 2005]
 - Adaptive local basis set [Lin et al 2012]
 - Adaptive minimal basis in BigDFT [Mohr et al 2014]

Fourier basis set

Computational domain $\Omega = [0, L_1] \times [0, L_2] \times [0, L_3]$ with periodic boundary condition (*i.e.*, Γ -point for simplicity).

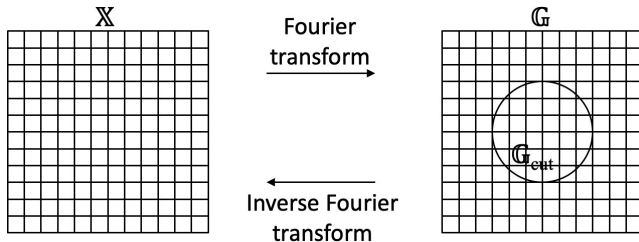
Reciprocal lattice in the frequency space

$$\mathbb{L}^* = \left\{ \mathbf{g} = \left(\frac{2\pi}{L_1}i_1, \frac{2\pi}{L_2}i_2, \frac{2\pi}{L_3}i_3 \right), \quad i \in \mathbb{Z}^3 \right\}$$

Basis functions

$$\phi_{\mathbf{g}}(\mathbf{r}) = \frac{1}{|\Omega|^{1/2}} \exp(i\mathbf{g} \cdot \mathbf{r})$$

with $\mathbf{g} \in \mathbb{G}_{\text{cut}} = \left\{ \mathbf{g} \in \mathbb{L}^* : \frac{1}{2}|\mathbf{g}|^2 \leq E_{\text{cut}} \right\}$.



Real space representation

Fourier basis set (without energy cutoff) can be equivalently viewed in the real space using psinc function as basis set [Skylaris et al 2005]

$$\varphi_{\mathbf{r}'}(\mathbf{r}) = \frac{1}{\sqrt{N_g|\Omega|}} \sum_{\mathbf{g} \in G} \exp(i\mathbf{g} \cdot (\mathbf{r} - \mathbf{r}')).$$

This is a numerical δ -function on the discrete set \mathbb{X} , satisfying

$$\varphi_{\mathbf{r}'}(\mathbf{r}) = \sqrt{\frac{N_g}{|\Omega|}} \delta_{\mathbf{r}, \mathbf{r}'}, \quad \mathbf{r}, \mathbf{r}' \in \mathbb{X}.$$

Thus a function in the finite dimensional approximation space can be represented by its values on the grid \mathbb{X} .

We can keep in mind discretization based on real space representation for the discussion of algorithms below (pseudopotential assumed).

Numerical atomic-orbitals

Basis functions are given by

$$\left\{ \varphi_{ilm}(\mathbf{r} - \mathbf{R}_I) = \frac{u_i(|\mathbf{r} - \mathbf{R}_I|)}{|\mathbf{r} - \mathbf{R}_I|} Y_{lm} \left(\frac{\mathbf{r} - \mathbf{R}_I}{|\mathbf{r} - \mathbf{R}_I|} \right) \mid i = 1, \dots, n_I, I = 1, \dots, M \right\},$$

where Y_{lm} are spherical harmonics and the radial part is obtained by solving (numerically) Schrödinger-like radial equation

$$\left(-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r) \right) u_i(r) = \epsilon_i u_i(r)$$

- $v_i(r)$: a radial potential chosen to control the main behavior of u_i ;
- $v_{\text{cut}}(r)$: a confining potential to ensure the rapid decay of u_i beyond a certain radius and can be treated as compactly supported.

NAO is used in FHI-aims [Blum et al 2009], which is one of the most accurate all-electron DFT package.

Adaptive / hybrid basis set

Recall ...

- Large basis set
 - pro: accuracy is systematically improvable;
 - con: higher number of DOF per atom;
- Small basis set
 - pro: much smaller number of DOF per atom;
 - con: more difficult to improve its quality in a systematic fashion (rely heavily on tuning and experience).

Adaptive basis aims at combining the best of two worlds. Examples:

- Augmented planewave method [Slater 1937, Andersen 1975]
- Nonorthogonal generalized Wannier function [Skylaris et al 2005]
- Adaptive local basis set [Lin et al 2012]
- Adaptive minimal basis in BigDFT [Mohr et al 2014]

Adaptive local basis set [Lin, Lu, Ying, E, 2012]

Key idea: Use local basis sets numerically obtained on local patches of the domain so that it captures the local information of the Hamiltonian.

- Based on the discontinuous Galerkin framework for flexibility of basis set choices (alternatives, such as partition of unit FEM, can be also used);
- Partition the computational domain into non-overlapping patches $\{\kappa_1, \kappa_2, \dots, \kappa_m\}$;
- For each patch κ , solve on an extended element $\tilde{\kappa}$ eigenvalue problem locally and take the first few eigenfunctions

$$\left(-\frac{1}{2}\Delta + V_{\text{eff}}^{\tilde{\kappa}} + V_{\text{nl}}^{\tilde{\kappa}}\right)\tilde{\phi}_{\kappa,j} = \lambda_{\kappa,j}\tilde{\phi}_{\kappa,j}$$

- Restrict $\tilde{\phi}_{\kappa,j}$ to κ and use SVD to obtain an orthogonal set of local basis functions.

Remark: Similar in spirit to GFEM, GMsFEM, reduced basis functions, etc.

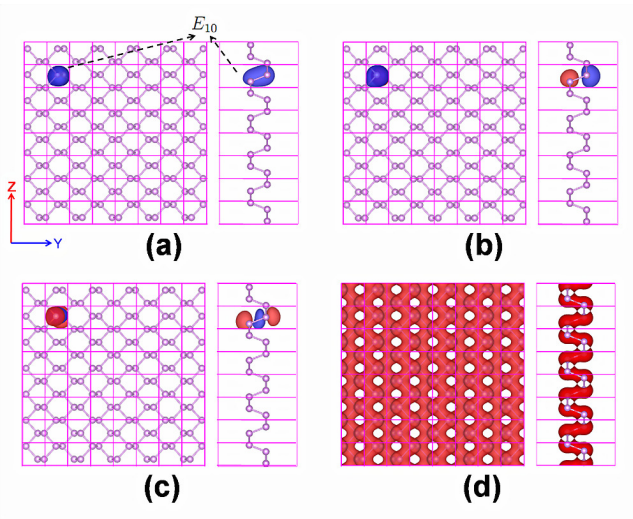


Figure: The isosurfaces ($0.04 \text{ Hartree/Bohr}^3$) of the first three ALB functions belonging to the tenth element E_{10} : (a) ϕ_1 , (b) ϕ_2 , (c) ϕ_3 , and (d) the electron density ρ across in top and side views in the global domain in the example of P_{140} . There are 64 elements and 80 ALB functions in each element, which corresponds to 37 basis functions per atom [Hu et al, 2015].

DGDFT package

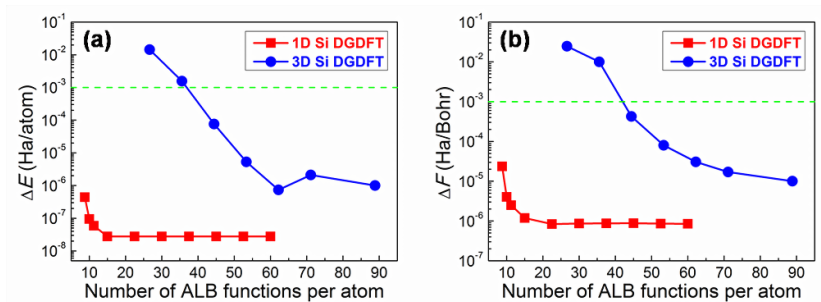


Figure: Convergence of DGDFE total energy and forces for quasi-1D and 3D Si systems to reference planewave results with increasing number of ALB functions per atom. (a) Total energy error per atom ΔE (Ha/atom). (b) Maximum atomic force error ΔF (Ha/Bohr). The dashed green line corresponds to chemical accuracy. Energy cutoff $E_{\text{cut}} = 60$ Ha and penalty parameter $\alpha = 40$. When the number of basis functions per atom is sufficiently large, the error is well below the target accuracy (green dashed line) [Zhang et al 2017].

Evaluation of the Kohn-Sham map

Problem: Given H , solve for the electron density ρ .

- Eigensolver
 - Direct diagonalization: ScaLAPACK, ELPA
 - Iterative methods:
 - ★ Traditional iterative diagonalization [Davidson 1975, Liu 1978]
 - ★ Modern Krylov subspace method (with reduced orthogonalization) [Knyazev 2001, Vecharynski et al 2015]
 - ★ Chebyshev filtering method [Zhou et al 2006]
 - ★ Orbital minimization method [Mauri et al 1993, Ordejón et al 1993]
- Linear scaling methods
 - Divide-and-conquer [Yang 1991]
 - Density matrix minimization [Li, Nunes, Vanderbilt 1993; Lai et al 2015]
 - Density matrix purification [McWeeny 1960]
- Pole expansion and selected inversion (PEXSI) method [Lin et al 2009]

Common library approach for Kohn-Sham map:

ELSI (ELectronic Structure Infrastructure) [Yu et al 2018]

<https://wordpress.elsi-interchange.org/>

Chebyshev filtering [Zhou et al 2006]

Key observation: For Kohn-Sham DFT, we do not need individual eigenvalues and eigenvectors of H , but only a representation for the occupied space.

Accelerated subspace iteration

$$X_{n+1} = P_k(H)(X_n).$$

P_k is a k -th degree Chebyshev polynomial to filter out the higher spectrum.

- Orthogonalization of X_{n+1} is needed to avoid collapsing (Rayleigh-Ritz rotation is used);
- Might be bypassed using 2-level Chebyshev filtering (inner Chebyshev for spectrum near Fermi level) [Banerjee et al 2018]
- For insulating systems, can be replaced by localization procedure to achieve linear scaling [E, Li, Lu 2010]

Algorithm 1: Chebyshev filtering method for solving the Kohn-Sham DFT eigenvalue problems $H\psi_i = \varepsilon_i\psi_i$.

Input: Hamiltonian matrix H and an orthonormal matrix $X \in \mathbb{C}^{N_b \times N_s}$

Output: Eigenvalues $\{\varepsilon_i\}_{i=1}^N$ and wave functions $\{\psi_i\}_{i=1}^N$

- 1: Estimate ε_{N+1} and ε_{N_b} using a few steps of the Lanczos algorithm.
 - 2: **while** convergence not reached **do**
 - 3: Apply the Chebyshev polynomial $P_k(H)$ to X : $Y = P_k(H)X$.
 - 4: Orthonormalize columns of Y .
 - 5: Compute the projected Hamiltonian matrix $\tilde{H} = Y^*HY$ and solve the eigenproblem $\tilde{H}\tilde{\Psi} = \tilde{\Psi}\tilde{D}$.
 - 6: Subspace rotation $X = Y\tilde{\Psi}$.
 - 7: **end while**
 - 8: Update $\{\psi_i\}_{i=1}^N$ from the first N columns of X .
-

Linear scaling methods

The conventional diagonalization algorithm scales as $\mathcal{O}(N^3)$, known as the “cubic scaling wall”, which limits the applicability of Kohn-Sham DFT.

For insulating systems (or metallic system at high temperature), $\mathcal{O}(N)$ -scaling algorithms are available, based on the locality of the electronic structure problem:

- Exponential decay of the density matrix;
- Exponentially localized basis of the occupied space (*i.e.*, localized molecular orbitals, Wannier functions)

Linear scaling methods

The conventional diagonalization algorithm scales as $\mathcal{O}(N^3)$, known as the “cubic scaling wall”, which limits the applicability of Kohn-Sham DFT.

For insulating systems (or metallic system at high temperature), $\mathcal{O}(N)$ -scaling algorithms are available, based on the locality of the electronic structure problem:

- Exponential decay of the density matrix;
- Exponentially localized basis of the occupied space (*i.e.*, localized molecular orbitals, Wannier functions)

Side: Localization methods for Wannier functions

- Optimization based approaches [Foster, Boys 1960; Marzari, Vanderbilt 1998; E, Li, Lu 2010; Mustafa et al 2015];
- Gauge smoothing approach [Cances et al 2017];
- SCDM (selected column of density matrix) [Damle et al 2015]

Density matrix purification

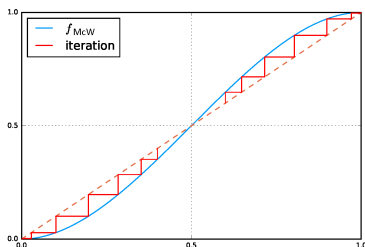
McWeeny's purification

$$P_{n+1} = f_{\text{McW}}(P_n) = 3P_n^2 - 2P_n^3$$

with initial iterate ($\alpha = \min\{(\epsilon_{\max} - \mu)^{-1}, (\mu - \epsilon_{\min})^{-1}\}$)

$$P_0 = \frac{\alpha}{2}(\mu - H) + \frac{1}{2}I.$$

Fixed point of $f_{\text{McW}} = 0, \frac{1}{2}, 1$:



Density matrix purification

Linear scaling can be achieved by truncating the density matrix at each iteration; see e.g., NTPoly [Dawson, Nakajima 2018]

The initialization requires estimate of the extreme eigenvalues of H and also the chemical potential.

$$P_0 = \frac{\alpha}{2}(\mu - H) + \frac{1}{2}I.$$

where $\alpha = \min\{(\epsilon_{\max} - \mu)^{-1}, (\mu - \epsilon_{\min})^{-1}\}$.

Several ways have been proposed for auto-tuning the chemical potential (reviewed in [Niklasson 2011]):

- Canonical purification [Paler, Manolopoulos 1998]
- Trace correcting purification [Niklasson 2002]
- Trace resetting purification [Niklasson et al 2003]
- Generalized canonical purification [Truflandier et al 2016]

PEXSI (Pole EXpansion and Selected Inversion)

PEXSI [Lin et al 2009, Lin et al 2011, Jacquelin et al 2016]

- Reduced scaling algorithm based on Fermi operator expansion and sparse direct linear algebra
 - $\mathcal{O}(N)$ for quasi-1D system;
 - $\mathcal{O}(N^{3/2})$ for quasi-2D system;
 - $\mathcal{O}(N^2)$ for bulk 3D system.
- Applicable to general systems (insulating or metallic) with high accuracy;
- Integrated into packages include BigDFT, CP2K, DFTB+, DGDFT, FHI-aims, SIESTA;
- Part of the Electronic Structure Infrastructure (ELSI);
- Available at <http://www.pexsi.org/> under the BSD-3 license.

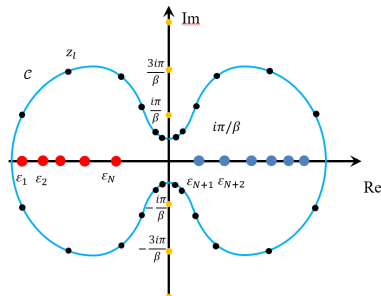
Pole expansion

Contour integral representation for density matrix (at finite temperature); chemical potential can be estimated on-the-fly during SCF [Jia, Lin 2017]

$$f_{\beta}(H - \mu) = \frac{1}{2\pi i} \oint_{\mathcal{C}} f_{\beta}(z) ((z + \mu)I - H)^{-1} dz$$

Discretization with $\mathcal{O}(\log(\beta\Delta E))$ terms [Lin et al 2009; Moussa 2016]

$$P \approx \sum_{l=1}^m \omega_l (H - z_l)^{-1}.$$



Selected inversion

Key observation: We don't need every entry of $(H - z_l)^{-1}$; only those near the diagonal.

Assume A partitioned into a 2×2 block form

$$A = \begin{pmatrix} \alpha & b^\top \\ b & \tilde{A} \end{pmatrix}$$

Pivoting by α (with $S = \tilde{A} - b\alpha^{-1}b^\top$ known as the Schur complement)

$$A = \begin{pmatrix} 1 & \\ \ell & I \end{pmatrix} \begin{pmatrix} \alpha & \\ & \tilde{A} - b\alpha^{-1}b^\top \end{pmatrix} \begin{pmatrix} 1 & \ell^\top \\ & I \end{pmatrix}$$

A^{-1} can be expressed by

$$A^{-1} = \begin{pmatrix} \alpha^{-1} + \ell^\top S^{-1} \ell & -\ell^\top S^{-1} \\ -S^{-1} \ell & S^{-1} \end{pmatrix}$$

Thus the calculation can be organized in a recursive fashion based on the hierarchical Schur complement.

Algorithm 2: Selected inversion based on LDL^\top factorization.

Input: LDL^\top factorization of a symmetric matrix $A \in \mathbb{C}^{N_b \times N_b}$.

Output: Selected elements of A^{-1} , i.e., $\{A_{i,j}^{-1} \mid (L + L^\top)_{i,j} \neq 0\}$.

- 1: Calculate $A_{N_b, N_b}^{-1} \leftarrow (D_{N_b, N_b})^{-1}$.
 - 2: **for** $k = N_b - 1, \dots, 1$ **do**
 - 3: Find the collection of indices $C = \{i \mid i > k, L_{i,k} \neq 0\}$.
 - 4: Calculate $A_{C,k}^{-1} \leftarrow -A_{C,C}^{-1} L_{C,k}$.
 - 5: Calculate $A_{k,C}^{-1} \leftarrow (A_{C,k}^{-1})^\top$.
 - 6: Calculate $A_{k,k}^{-1} \leftarrow (D_{k,k})^{-1} - A_{k,C}^{-1} L_{C,k}$.
 - 7: **end for**
-

In practice, as in LDLT packages, columns of A are partitioned into supernodes (set of contiguous columns) to enable level-3 BLAS for efficiency [Jacquelin et al 2016].

Performance and scalability of PEXSI

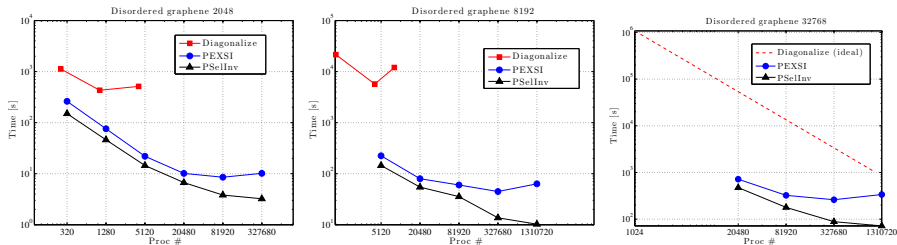


Figure: Wall clock time versus the number of cores for a graphene systems with 2048, 8192, and 32768 atoms [Jacquelin et al 2016]

241 second wall clock time for graphene with 32,768 atoms (DG ALB discretization); infeasible for traditional solvers.

What is happening now and perhaps in the near future?

- Massively scalable software development
e.g., the ELSI project [Yu et al 2018]; the ESL project;
many actively maintained packages available

What is happening now and perhaps in the near future?

- Massively scalable software development
e.g., the ELSI project [Yu et al 2018]; the ESL project;
many actively maintained packages available
- Efficient algorithms for nonlocal functionals (and beyond)

Some recent developments of algorithmic tools
(low-rankness seems to be useful):

- Interpolative separable density fitting (ISDF) for pair products of orbital functions [Lu, Ying 2015];
- Projective eigendecomposition of the dielectric screening (PDEP) for GW calculations [Govoni, Galli 2015]
- Adaptive compressed exchange operator (ACE) method [Lin 2016].
- ...

What is happening now and perhaps in the near future?

- Massively scalable software development
e.g., the ELSI project [Yu et al 2018]; the ESL project;
many actively maintained packages available
- Efficient algorithms for nonlocal functionals (and beyond)

Some recent developments of algorithmic tools
(low-rankness seems to be useful):

- Interpolative separable density fitting (ISDF) for pair products of orbital functions [Lu, Ying 2015];
- Projective eigendecomposition of the dielectric screening (PDEP) for GW calculations [Govoni, Galli 2015]
- Adaptive compressed exchange operator (ACE) method [Lin 2016].
- ...
- Excited states; time-dependent DFT; highly correlated systems;

What is happening now and perhaps in the near future?

- Massively scalable software development
e.g., the ELSI project [Yu et al 2018]; the ESL project;
many actively maintained packages available
- Efficient algorithms for nonlocal functionals (and beyond)

Some recent developments of algorithmic tools
(low-rankness seems to be useful):

- Interpolative separable density fitting (ISDF) for pair products of orbital functions [Lu, Ying 2015];
- Projective eigendecomposition of the dielectric screening (PDEP) for GW calculations [Govoni, Galli 2015]
- Adaptive compressed exchange operator (ACE) method [Lin 2016].
- ...
- Excited states; time-dependent DFT; highly correlated systems;
- Numerical analysis; robust and black-box algorithms for high-throughput calculations;

What is happening now and perhaps in the near future?

- Massively scalable software development
e.g., the ELSI project [Yu et al 2018]; the ESL project;
many actively maintained packages available
- Efficient algorithms for nonlocal functionals (and beyond)

Some recent developments of algorithmic tools
(low-rankness seems to be useful):

- Interpolative separable density fitting (ISDF) for pair products of orbital functions [Lu, Ying 2015];
- Projective eigendecomposition of the dielectric screening (PDEP) for GW calculations [Govoni, Galli 2015]
- Adaptive compressed exchange operator (ACE) method [Lin 2016].
- ...
- Excited states; time-dependent DFT; highly correlated systems;
- Numerical analysis; robust and black-box algorithms for high-throughput calculations;
- Machine learning techniques for electronic structure;

What is happening now and perhaps in the near future?

- Massively scalable software development
e.g., the ELSI project [Yu et al 2018]; the ESL project;
many actively maintained packages available
- Efficient algorithms for nonlocal functionals (and beyond)

Some recent developments of algorithmic tools
(low-rankness seems to be useful):

- Interpolative separable density fitting (ISDF) for pair products of orbital functions [Lu, Ying 2015];
- Projective eigendecomposition of the dielectric screening (PDEP) for GW calculations [Govoni, Galli 2015]
- Adaptive compressed exchange operator (ACE) method [Lin 2016].
- ...
- Excited states; time-dependent DFT; highly correlated systems;
- Numerical analysis; robust and black-box algorithms for high-throughput calculations;
- Machine learning techniques for electronic structure;
- Quantum computing for quantum chemistry

Thank you for your attention

Email: `jianfeng@math.duke.edu`

URL: `http://www.math.duke.edu/~jianfeng/`

Reference:

- **Lin Lin, L.**, and **Lexing Ying**, Numerical Methods for Kohn-Sham Density Functional Theory, *Acta Numerica*, 2019