Density functional theory: algorithms

Antoine Levitt

January 30, 2024

Kohn-Sham density functional theory for a molecule

For the ground state of $2N_{\rm el}$ spin-paired electrons in a potential $V_{\rm nucl}$:

$$\begin{cases} -\frac{1}{2}\Delta\phi_n + V_{\rm nucl}\phi_n + V_{\rm HXC}[\rho]\phi_n = \lambda_n\phi_n, \quad \langle\phi_n,\phi_m\rangle_{L^2(\mathbb{R}^3)} = \delta_{mn} \\ \rho(x) = 2\sum_{n=1}^{N_{\rm el}} |\phi_n(x)|^2 \\ V_{\rm HXC}[\rho] = \frac{1}{|x|} * \rho + V_{\rm XC}[\rho] \end{cases}$$

- $V_{\rm XC}$ is an approximation of the true exchange-correlation potential; e.g. LDA: $(V_{\rm XC}[\rho])(x) = v_{\rm xc}(\rho(x))$
- $N_{\rm el}$ nonlinear coupled PDEs in dimension 3
- Aufbau principle: $\{\lambda_n\}_{n=1,\dots,N_{\rm el}}$ are the lowest eigenvalues of

$$H_\rho = -\frac{1}{2}\Delta + V_{\rm nucl} + V_{\rm HXC}[\rho]$$

- Alternatively: minimize the energy $E(\phi)$ under $\langle\phi_n,\phi_m\rangle_{L^2(\mathbb{R}^3)} = \delta_{mn}$

Mean field methods for electrons

Roadmap:

- Set up the system (boundary conditions)
    - Molecule vs crystal
- Regularize the problem
    - Pseudopotentials
    - Finite temperature
- Discretize (basis sets)
    - Gaussian basis sets, finite differences, finite elements, Fourier expansions...
- Solve the discretized equations
    - Self-consistent field
    - Constrained optimization
- Get energy and derived properties

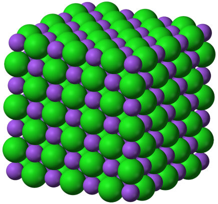
Mature scientific domain ('70s), but new developments (more accurate discretizations, better algorithms, HPC implementations...)

Post-processing

Once you're done solving the ground state:

- Optimize total energy with respect to atomic positions
  - Nested optimization problem
- Move the atoms and solve it again (ab initio molecular dynamics)
  - Very numerically intensive
  - Parametric nonlinear eigenvalue problem
- Perturb the problem and compute the response: $V_{\mathrm{ext}} + \delta V_{\mathrm{ext}} \to \rho + \delta \rho$
  - Extremely useful physically (forces, stresses, phonons, speed of sound...)
  - Mathematically: eigenvalue perturbation theory, derivatives of functions of matrices
- Compute response to non-periodic variations
  - Numerical integration
- Build reduced model
  - Involves unexpected mathematics (geometry, topology)
- Do a sensitivity analysis
- Dynamics (time-dependent DFT): $i\partial_t \phi_n(t) = H_{\rho(t)} \phi_n(t)$
  - Geometric integration
- Make the nuclei quantum
- ...

Discretization



$$-\frac{1}{2}\Delta\phi_n + V_{\mathrm{nucl}}\phi_n + V_{\mathrm{HXC}}[\rho]\phi_n = \lambda_n\phi_n$$

- For a solid, with Brillouin zone sampling, u_n is periodic
- Fourier discretization natural

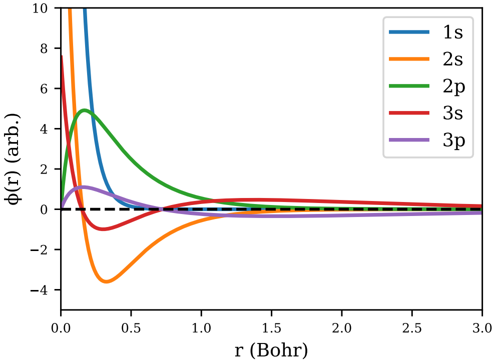
$$\phi_n(x) = \sum_{G\in\mathcal{R}^*} c_G e^{iG\cdot x}$$

- The smoother u is, the faster c_G decays

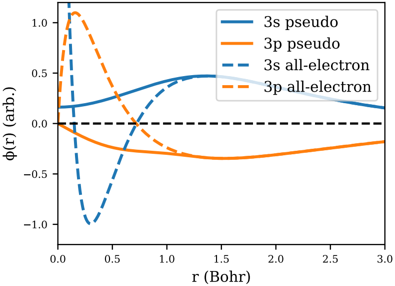
Pseudopotentials

Several problems:

1. Coulomb $1/|x|$ singularity ⇒ cusp at nuclei positions
2. Core states strongly peaked
3. Valence states strongly oscillatory



Pseudopotentials



For an atom, replace $-Z/|x|$ by

$$V_{\rm loc}(x) + \sum_{ij} D_{ij} |p_i\rangle\langle p_j|$$

- Choose $V_{\rm loc}, D_{ij}, p_i$ so that the valence eigenvalues and orbitals after some cutoff $r_{\rm cut}$ match the exact ones.
- Low-rank nonlocal potential

Plane-wave discretization

$$-\tfrac{1}{2}\Delta\phi_n + V_{\mathrm{pseudo}}\phi_n + V_{\mathrm{HXC}}[\rho]\phi_n = \lambda_n\phi_n$$

$$\phi_n(x) = \sum_{G\in\mathcal{R}^*} c_G e^{iG\cdot x}$$

How to truncate G vectors? Kinetic energy cutoff

$$\frac{1}{2}|G|^2 \leq E_{\mathrm{cut}}$$

Galerkin method: require that

$$\forall G \text{ s.t. } \frac{1}{2}|G|^2 \leq E_{\mathrm{cut}},$$

$$\langle e^{iG\cdot r}, H\phi_n\rangle = \lambda_n\langle e^{iG\cdot r}, \phi_n\rangle$$

Plane waves are orthogonal: yields an $N_{\mathrm{pw}} \times N_{\mathrm{pw}}$ algebraic system

$$\sum_{G'} H_{GG'}c_{G'} = \lambda_n c_G$$

The Hamiltonian

$$H = -\frac{1}{2}\Delta + V_{\mathrm{loc,HXC}} + \sum_{mn} D_{mn} |p_m\rangle\langle p_n|$$

$$\begin{aligned} H_{GG'} &= \frac{1}{\Omega}\langle e^{iGr}, He^{iG'r}\rangle \\ &= \frac{1}{2}|G|^2\delta_{GG'} + \widehat{V}_{\mathrm{loc,HXC}}(G' - G) + \sum_{mn} D_{mn}\widehat{p}_m(G)\overline{\widehat{p}_n(G')} \end{aligned}$$

Forming the Hamiltonian

- V_{loc} and V_{nonloc} are built from radial pseudopotential information
- The density

  $$\rho(x) = 2\sum_{nk} |\phi_n(x)|^2$$

  is computed on a real-space grid
- The exchange-correlation potential

  $$V_{\mathrm{XC}}(x) = V_{\mathrm{LDA}}(\rho(x))$$

  also
- The Hartree potential

  $$-\Delta V_{\mathrm{H}} = 4\pi\rho$$

  is computed in Fourier space
- All of these steps are not very costly

Applying the Hamiltonian

$$H_{GG'} = \frac{1}{2}|G|^2\delta_{GG'} + \widehat{V}_{\mathrm{loc,HXC}}(G' - G) + \sum_{mn} D_{mn}\widehat{p}_m(G)\overline{\widehat{p}_n(G')}$$

- Discrete convolutions can be performed with an FFT and zero-padding
- Matrix-vector products can be performed by simple scaling, FFT and low-rank multiplication
- Operation cost for one matvec:

$$O(N_{\mathrm{pw}} + N_{\mathrm{FFT}} \log(N_{\mathrm{FFT}}) + N_{\mathrm{proj}} N_{\mathrm{pw}})$$

- Orders of magnitudes:
  1. $N_{\mathrm{el}} \approx 5N_{\mathrm{at}}$
  2. $N_{\mathrm{proj}} \approx 5N_{\mathrm{at}}$
  3. $N_{\mathrm{pw}} \approx 200N_{\mathrm{at}}$
  4. $N_{\mathrm{FFT}} \approx 5000N_{\mathrm{at}}$
- For small to moderate systems, FFT dominates

Summary

Solve

$$H[\rho]\phi_n = \lambda_n\phi_n$$

$$\rho(x) = 2\sum_{n=1}^{N_{\mathrm{el}}} |\phi_n(x)|^2$$

where H can be applied on a vector efficiently

In linear algebra notation:

$$A[X]x_n = \lambda_n x_n$$

Nonlinear eigenvector problem (see corresponding MS)

Self-consistent equations

$$H[\rho]\phi_n = \lambda_n \phi_n, \quad \rho(x) = 2\sum_n |\phi_n(x)|^2$$

- Nonlinear problem: get iterative solution
- Simple SCF:

$$H[\rho^i]\phi_n^{i+1} = \lambda_n^{i+1}\phi_n^{i+1}, \quad \rho^{i+1}(r) = 2\sum_{n=1}^{N_{\rm el}} |\phi_n^{i+1}(r)|^2$$

- Usually does not converge!
- Simple mixing:

$$\rho^{i+1} = \rho^i + \alpha\left(2\sum_{nk} |\phi_n^{i+1}(r)|^2 - \rho^i\right)$$

  converges for $\alpha > 0$ small enough, but slowly
- State of the art: advanced mixing (Anderson) and preconditioning (Kerker...)

Diagonalization of the Hamiltonian

The SCF loop

1. Initialize system, precompute things
2. Diagonalize the Hamiltonian
3. Update the Hamiltonian
4. If not converged, go back to (2)
5. Postprocess

$$H\phi_n = \varepsilon_n \phi_n$$

Characteristics:

- Need to solve for $N_{\mathrm{el}} \ll N_{\mathrm{pw}}$ states ($N_{\mathrm{el}} \approx 1\%$ of N_{pw})
- Matrix dense but efficient matvecs available
- Need smallest eigenvalues
- Scale separation: characteristic gap $\approx 1eV$, largest eigenvalue (spectral width) $\approx 1000eV$
- Usually have good initial guess

A prototype diagonalization algorithm: subspace iteration

Input: set of vectors $X \in \mathbb{C}^{N_{\mathrm{pw}} \times N_{\mathrm{el}}}$

1. Form the block matvec $Y_i = HX_i$
2. Find an orthonormal basis Z_i of the range of (X_i, Y_i)
3. Solve for the N_{pw} lowest eigenvalues of $(Z_i^* H Z_i)c = \lambda c$
4. Obtain the new $X_{i+1} = Z_i c$

A better diagonalization algorithm: LOBPCG

Input: set of vectors $X \in \mathbb{C}^{N_{\mathrm{pw}} \times N_{\mathrm{el}}}$

1. Form the block matvec $Y_i = HX_i$
2. Find an orthonormal basis Z_i of the range of (X_i, Y_i, X_{i-1})
3. Solve for the N_{pw} lowest eigenvalues of $(Z_i^* HZ_i)c = \lambda c$
4. Obtain the new $X_{i+1} = Z_i c$

LOBPCG: advantages and inconvenients

Advantages

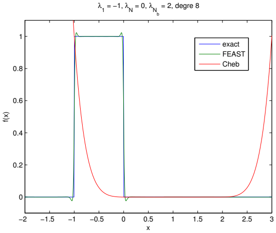
- Krylov-like convergence rate
- Able to use preconditioning (Laplacian diagonal in Fourier basis ⇒ very efficient)

Inconvenients

- Very tricky to code in a stable way
- Diagonalization of $3N_{\mathrm{el}} \times 3N_{\mathrm{el}}$ becomes costly
- Require lots of communications in an HPC setting

Alternative algorithms: Chebyshev filtering and FEAST

Finding the N first eigenvectors can be done by applying $f(H)$ to a set of random vectors, with $f(x) = \mathbb{1}(x \leq \lambda_{N_{\text{el}}})$



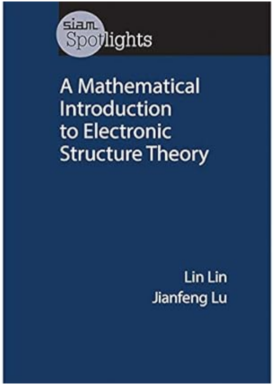
Approximate the step function

- By a polynomial: Chebyshev filtering
  - Same $\sqrt{\kappa}$ convergence rate property as Krylov, but less adaptable
  - Every vector can be filtered in parallel
  - No preconditioning
- By a rational function: FEAST
  - Needs modest number of poles
  - $f(x) \approx \sum_i r_i(x)$, very good parallel properties
  - Need to solve linear systems efficiently...

Even more algorithms

- Optimization-based
- Spectrum slicing
- ...

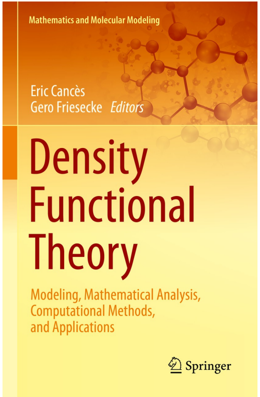
References



siam Spotlights

A Mathematical Introduction to Electronic Structure Theory

Lin Lin
Jianfeng Lu



Mathematics and Molecular Modeling

Eric Cancès
Gero Friesecke *Editors*

Density Functional Theory

Modeling, Mathematical Analysis, Computational Methods, and Applications

Springer