Solving Nonlinear Eigenvalue Problems in Electronic Structure Calculations

Chao Yang Juan Meza Lin-wang Wang

Computational Research Division

Lawrence Berkeley National Lab

Quantum Many-Body Problems

$$H\psi_i(r_1, r_2, \ldots, r_N) = E_i\psi_i(r_1, r_2, \ldots, r_N)$$

$$H = -\frac{\hbar}{2m}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}v(r_i) + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|r_i - r_j|}$$

- ψ contains all the information of the system;
- $|\psi_i|^2$ probability density of finding electrons at a certain state *i*.
- E_i quantized energy of the system.

- 3N-coordinate space
- Manageable N
  - < 3 for analytic calculations (e.g. H, He, etc.)
  - O(10) for configuration interaction simulations
  - O(100) for quantum Monte Carlo simulations
- N ≤ 100 for an atom or crystal; for nano-structures ~ (10^3 - 10^6)

Reducing many-body problem to a (nonlinear) single-particle problem (Density Functional Theory)

The ground state density of an atomistic system can be found by minimizing a total energy functional with respect to single particle wave functions (Hohenberg, Kohn, Sham)

$$E_{total}(X) = E_{kinetic} + E_{ionic} + E_{Hartree} + E_{xc}$$

$$E_{kinetic} = \frac{1}{2}\mathrm{trace}(X^T L X)$$

$$E_{ionic} = \frac{1}{2}\left[\mathrm{trace}\left(X D_{ion} X^T\right) + \sum_i \sum_\ell \left(x_i^T w_\ell\right)^2\right]$$

$$E_{Hartree} = \frac{1}{4}\rho(X)^T S \rho(X)$$

$$E_{xc} = e^T\left(f_{xc}[\rho(X)]\right)$$

- single particle wave functions
  $$X = (x_1, x_2, ..., x_k),\ x_i \in \mathbf{R}^n$$
- n – real space grid size, e.g. 32^3~32000
- k – number of occupied states, 1-10% of n
- Charge density
  $$\rho(X) = \mathrm{diag}\left(X X^T\right)$$

Constrained Optimization

- **Solve**

$$\min_X E_{total}(X)$$

$$\text{s.t. } X^T X = I$$

- **First order necessary (KKT) condition**

$$\nabla_X \mathbf{L}(X) = 0,$$

$$X^T X = I,$$

$$\text{where } \mathbf{L}(X) = E_{total}(X) - \text{trace}(\Lambda^T (X^T X - I_k)).$$

Nonlinear Eigenvalue Problem

- **KKT condition (Kohn-Sham equation)**

$$\left[\underbrace{L + D_{ion} + \sum_{\ell} w_{\ell} w_{\ell}^{T} + \mathrm{Diag}\left(S\rho\left(X\right)\right) + \mathrm{Diag}\left(g_{xc}\left(\rho(X)\right)\right)}_{H(X)}\right] X = X\Lambda$$

$$X^{T} X = I$$

- **The solution is not unique (look for an invariant subspace)**

$$\rho(X) = \rho(XQ), \text{ for any } Q \text{ such that } Q^{T} Q = I_{k}$$

The Self Consistent Field Iteration

- **Input: initial guess** X_0 **and** L, D_{ion}, S, w_ℓ
- **Output:** $X = \arg\min\left(E_{total}(X)\right)$

  $\text{s.t. } X^T X = I_k$
- **Major steps**
  - **For** i=1,2,…,**until converged** $H\left(X^{(i)}\right)X^{(i)} \approx X^{(i)}\Lambda^{(i)}$

    1) Form $H^{(i)} = H(X^{(i)})$

    2) Compute k **smallest eigpairs of** $H^{(i)}$

$$X^{(i+1)} = \arg\min\left[\text{trace}\left(X^T H^{(i)} X\right)\right]$$
$$\text{s.t. } X^T X = I_k$$

SCF Iteration

- Most widely used
- Does not always converge
  - —The objective $E_{total}(X^{(i)})$ does not always decrease
- Little convergence theory
  - — Fixed point iteration?

$$X^{(i+1)} = F(X^{(i)}), \quad \text{but what is } F(\cdot)?$$

- Must solve a large-scale eigenvalue problem at each iteration
  - — Dense eigensolver (very expensive)
  - — Iterative solver: preconditioned conjugate gradient (accuracy, stopping criterion)

Optimization Point of View

SCF minimizes a sequence of surrogate models

- **True objective**

- $E_{kinetic} = \frac{1}{2}\mathrm{trace}(X^T L X)$
- $E_{ionic} = \frac{1}{2}\left[\mathrm{trace}\left(X D_{ion} X^T\right) + \sum_i \sum_\ell \left(x_i^T w_\ell\right)^2\right]$
- $E_{Hartree} = \frac{1}{4}\rho(X)^T S \rho(X)$
- $E_{xc} = e^T\left(f_{xc}[\rho(X)]\right)$

- **Surrogate model**

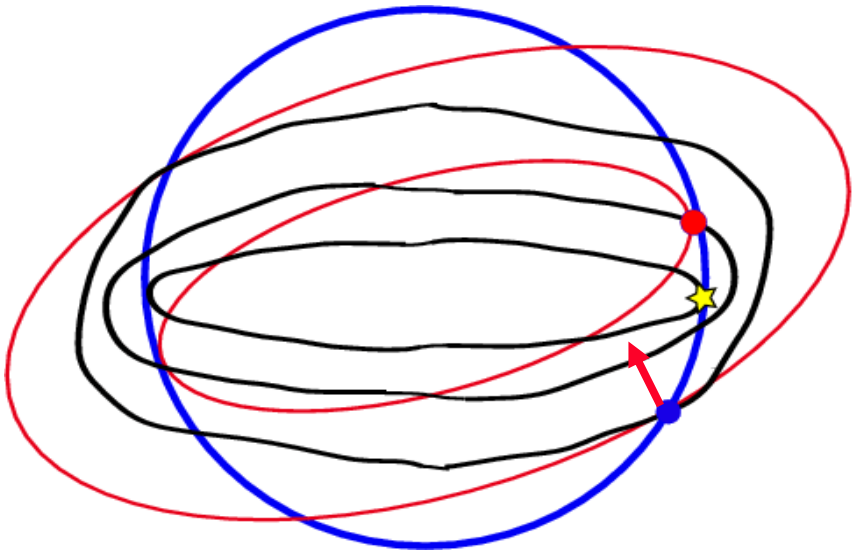
$$E_{surrogate} = \frac{1}{2}\mathrm{trace}\left(X^T H(X^{(i)}) X\right)$$

$$\begin{aligned} H(X) &= L + D_{ion} + \sum_\ell w_\ell w_\ell^T \\ &+ \mathrm{Diag}\left(S\rho(X)\right) \\ &+ \mathrm{Diag}\left(g_{xc}(\rho(X))\right) \end{aligned}$$

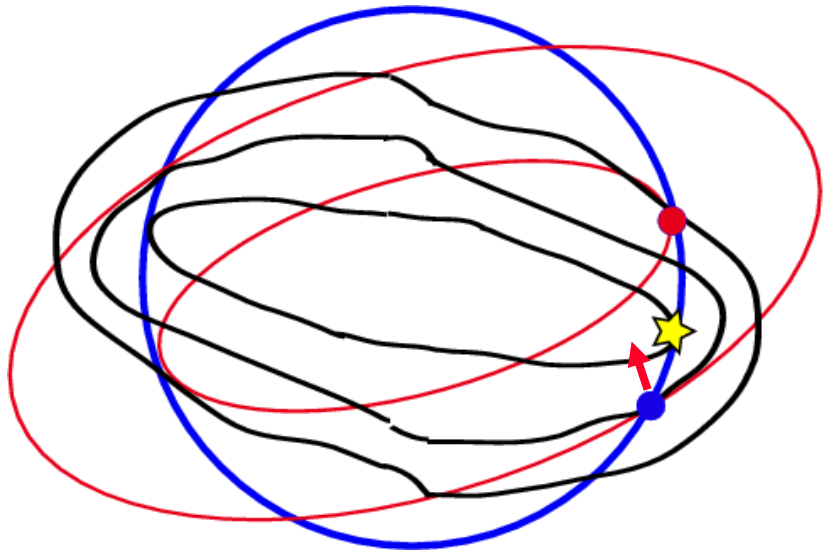
- **Gradient:** $\nabla_X E_{total} = H(X)X$
- **Gradient:** $\nabla_X E_{surrogate} = H(X^{(i)})X$

$$\nabla_X E_{total}(X^{(i)}) = \nabla_X E_{surrogate}(X^{(i)})$$

Geometry



SCF works



SCF fails

Improving SCF

- **Better Surrogate Model?**
  - **—Can't afford to use local quadratic approximation (Hessian too expensive)**
  - **—Charge mixing or DIIS (H is a function of the charge density)** $\rho^{(i+1)} = \rho^{(i)} + \gamma\left(\rho(X^{(i+1)}) - \rho^{(i)}\right)$
- **Use Trust Region?**
  - **—TRSCF by Thogersen, Olsen, Yeager & Jorgensen (2004)**
  - **—Globally convergent trust-region by Francisco, J. Martinez & L. Martinez (2004)**

Direct Minimization

- Minimize the total energy directly
- Maintain the orthonormality constraint
- Search direction
  - Gradient, conjugate gradient based approach
    - Brendt & Zunger (1982)
    - Gillan (1989)
    - Payne, Teter, Allen, Arias & Jaoannopoulos (1992)
    - Kresse & Furthmuller (1996)
  - Quasi-Newton
    - Edelman, Arias & Smith (1998)
    - Voorhis & Head-Gordon (2002)
- Line search
  - Re-parametrize $x^{(i+1)} = x^{(i)} \cos\theta + p^{(i)} \sin\theta$
  - Move along geodesic

New Approach

- **Conjugate Gradient-like scheme (no Hessian approximation)**
- **Block ("all band") method**
- **Determine search direction and "step length" simultaneously**

$$X^{(i+1)} = X^{(i)} G_x^{(i)} + \underbrace{P^{(i-1)} G_p^{(i)} + R^{(i)} G_r^{(i)}}_{\text{search direction } P^{(i)}}$$

$$R^{(i)} \equiv \nabla_X \mathbf{L} = H^{(i)} X^{(i)} - X^{(i)} \Theta^{(i)},$$

$$\Theta^{(i)} = \text{diag}\left(X^{(i)^T} H^{(i)} X^{(i)} \right)$$

Choose $G_x^{(i)}$, $G_P^{(i)}$, and $G_R^{(i)}$ ("step length") to:

1. $\min E_{total}\left(X^{(i+1)}\right)$
2. maintain $X^{(i+1)^T} X^{(i+1)} = I_k$

Yang, Meza &Wang, LBNL Tech Report-57434, 2005

Minimization within a Subspace

- **Let** $Y = \left(X^{(i)}, R^{(i)}, P^{(i-1)} \right)$
- **Solve**

$$\min_{G} E_{total}(YG)$$

$$\text{s.t. } G^T Y^T Y G = I_k$$

- **Equivalent to solving**

$$\hat{H}(G)G = BG\Omega_k,$$

$$G^T B G = I_k,$$

where

$$\hat{H}(G) = Y^T \left(L + D_{ion} + \sum_{\ell} w_\ell w_\ell^T + \text{Diag}\left(S\rho\left(YG\right)\right) + \text{Diag}\left(g_{xc}\left(\rho(YG)\right)\right) \right) Y,$$

$$B = Y^T Y$$

DCM Algorithm

For i=1,2,… until convergence

1. Form $H^{(i)} = H(X^{(i)})$
2. Compute $R^{(i)} = K^{-1}\left(H^{(i)}X^{(i)} - X^{(i)}\Theta^{(i)}\right)$,
   where $\Theta^{(i)} = \mathrm{Diag}\left(X^{(i)^T} H^{(i)} X^{(i)}\right)$, K is a preconditioner
3. If (i>1) then
   - set $Y \leftarrow \left(X^{(i)}, R^{(i)}, P^{(i-1)}\right)$
4. else
   - set $Y \leftarrow \left(X^{(i)}, R^{(i)}\right)$
5. Solve $\min_{G^T Y^T Y G = I_k} E_{total}\left(Y^{(i)} G\right)$
6. Set $X^{(i+1)} = YG$
7. If (i>1) then
   - set $P^{(i)} \leftarrow Y^{(i)}(:, k+1:3k)G(k+1:3k,:)$
8. else
   - set $P^{(i)} \leftarrow Y^{(i)}(:, k+1:2k)G(k+1:2k,:)$

Solving the projected problem

- **Use SCF**
  - —**No guarantee of convergence**
  - —**But full convergence may not be necessary, all we need is sufficient decrease in** $E_{total}(YG)$
- **Constrained Quasi-Newton**
  - —**May be feasible due to the much smaller size of the problem**
  - —**Optimization on Grassman manifold (Edelman et al 1998, Van Vooris & Head-Gordon 2002)**

Estimated Computational Complexity

(in terms of the number of MATVECs $y \leftarrow H^{(i)}x$)

DCM

- **Outer iteration:**
  - **Gradient calculation**

  k MATVECs/iter
- **Inner iteration**
  - **Update the Hamiltonian**

$$Y^T\left[\mathrm{Diag}\left(R\rho\left(YG\right)\right)+\mathrm{Diag}\left(g\left(\rho\left(YG\right)\right)\right)\right]Y$$

- **Overall:**

  $k+p$ MATVECs/outer iter

SCF

- **Outer iteration:**
  - **Update the Hamiltonian**

$$\mathrm{Diag}\left(\mathrm{R}\rho\left(X^{(i)}\right)\right)+\mathrm{Diag}\left(g_{xc}\left(\rho\left(X^{(i)}\right)\right)\right)$$

- **Inner iteration:**
  - **Gradient calculation**

  k MATVECs/iter
- **Overall:**

  $k*p$ MATVECs/outer iter

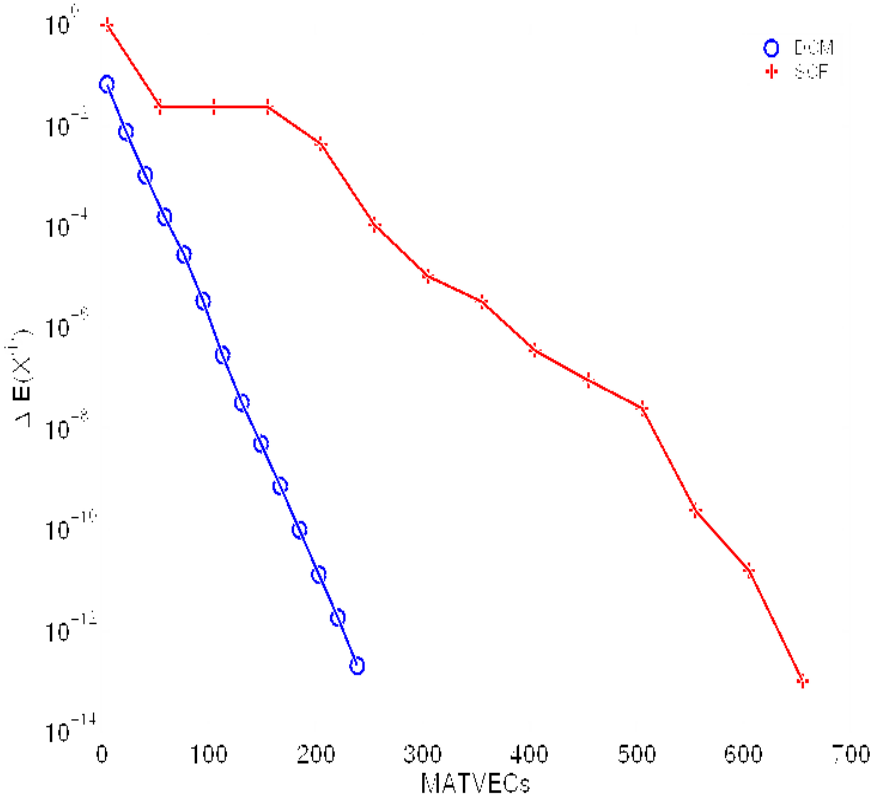
Assumption: k is much smaller than n

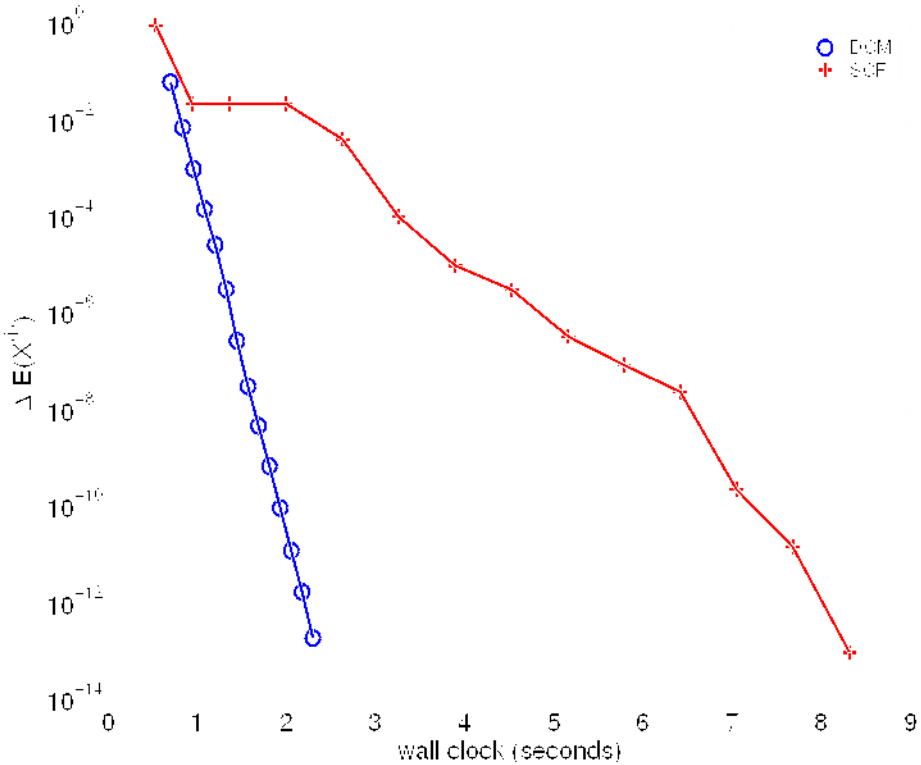
Numerical Example

- **Atomistic system: SiH4**
- **Discretization: spectral method with plane wave basis: $n=32^3$ in real space, $N=2103$ (# of basis functions) in frequency space**
- **Number of occupied states: $k = 4$**
- **PETOT version of SCF uses 10 PCG steps (inner iterations) per outer iteration**
- **DCM: 3 inner iterations**
- **Compare the change of the objective function with respect to the number of MATVECs performed & timing**
- **Parallelized using MPI on 16 IBM Power3 CPUS**

Performance

$$\Delta E(X^{(i)}) = E_{total}(X^{(i)}) - E_{\min}$$





Conclusion

- **Nonlinear eigenvalue problem in electronic structure calculation should be solved as an optimization problem**
- **No need to solve linear eigenvalue problems to full accuracy in SCF**
- **SCF can be further improved (better surrogate model, trust region)**
- **DCM competitive with SCF (Reduced a large-scale nonlinear minimization problem to a sequence of much smaller problems)**
- **Further research required to solve the small nonlinear problem efficiently**