

Nonlinear Subspace Iteration with applications to DFT

Yousef Saad

Department of Computer Science and Engineering

University of Minnesota

SIAM Applied Linear Algebra

Mini-symposium on DFT

Monterey, CA, Oct. 26-29 th 2009

Density Function Theory - Kohn-Sham Eqns.

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_{tot}[\rho(\mathbf{r}), \mathbf{r}] \right] \Psi(\mathbf{r}) = E \Psi(\mathbf{r})$$

With

$$V_{tot} = V_{ion} + V_H + V_{xc}$$

- V_H = Hartree potential local
- V_{xc} = Exchange & Correlation potential local (LDA)
- V_{ion} = Ionic potential Non-local

▶▶ Electron Density:

$$\rho(\mathbf{r}) = \sum_i^{occup} |\Psi_i(\mathbf{r})|^2$$

Kohn-Sham as a nonlinear eigenvalue problem

$$\left\{ \begin{array}{l} 1. \quad \left[-\frac{\hbar^2}{2m} \nabla^2 + V_{tot}[\rho(r)] \right] \Psi_i(r) = E_i \Psi_i(r) \\ 2. \quad \rho(r) = \sum_i^{occup} |\Psi_i(r)|^2 \\ 3. \quad \nabla^2 V_H = -4\pi\rho(r) \rightarrow V_{tot} = V_H + V_{xc} + V_{ion} \end{array} \right.$$

► Both V_{xc} and V_H , depend on ρ .

The potentials and charge densities must be self-consistent. One can view the KS equations as (1) a nonlinear eigenvalue problem; or (2) a system of nonlinear equations; or (3) a nonlinear optimization problem

► Common approach: Broyden-type quasi-Newton technique. [Typically, a small number of iterations are required]

Self-Consistent Iteration

- ▶▶ Most time-consuming part: **diagonalization**
- ▶▶ Difficulty: large number of wanted eigenvalues/eigenvectors [number of occupied states].
- ▶▶ Consequence: orthogonalize a basis of m vectors of length N , at cost of $O(m^2N)$ – Both m and N are proportional to number of particles. No matter what - cost will scale like $O(N_{part})^3$
- ▶▶ BUT: prefactor can be reduced.

Diagonalization

Specificity of problem:

- 1) Large number of eigenvectors.
- 2) Nonlinear

- ▶▶ Actual problem is to compute a large invariant subspace
- ▶▶ Needed only to compute the diagonal of the projector onto the subspace
- ▶▶ Problem with general purpose software: difficult to take into account nonlinearity.

CHEBYSHEV FILTERING

Chebyshev Subspace iteration

▶▶ Main ingredient: Chebyshev filtering

Given a basis $[v_1, \dots, v_m]$, 'filter' each vector as

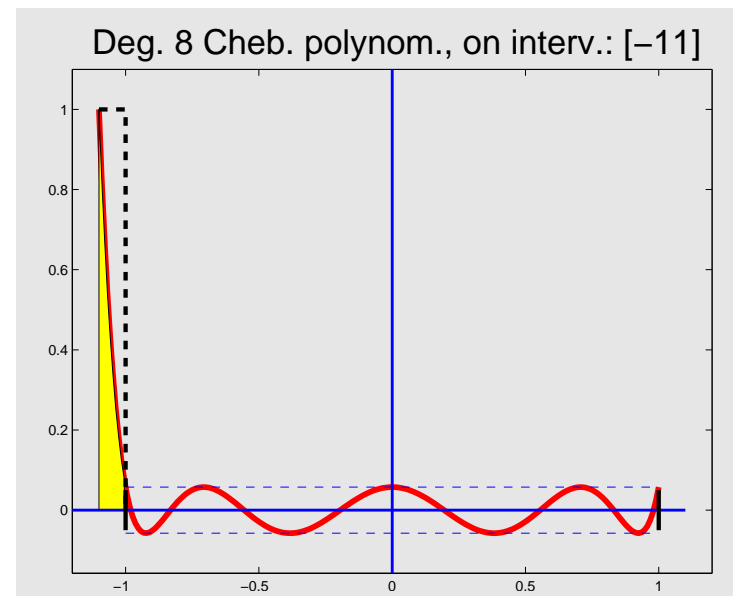
$$\hat{v}_i = P_k(A)v_i$$

▶▶ $p_k =$ Low deg. polynomial. Enhances wanted eigen-components

The **filtering step** is not used to compute eigenvectors accurately ▶▶

SCF & diagonalization loops merged

Important: convergence still good and robust



Main step:

Previous basis $V = [v_1, v_2, \dots, v_m]$

↓

Filter $\hat{V} = [p(A)v_1, p(A)v_2, \dots, p(A)v_m]$

↓

Orthogonalize $[V, R] = qr(\hat{V}, 0)$

▶▶ The basis V is used to do a Ritz step (basis rotation)

$$C = V^T A V \rightarrow [U, D] = \text{eig}(C) \rightarrow V := V * U$$

▶▶ Update charge density using this basis.

▶▶ Update Hamiltonian — repeat

▶▶ In effect: **Nonlinear subspace iteration**

▶▶ Main advantages: (1) very inexpensive, (2) uses minimal storage (m is a little \geq # states).

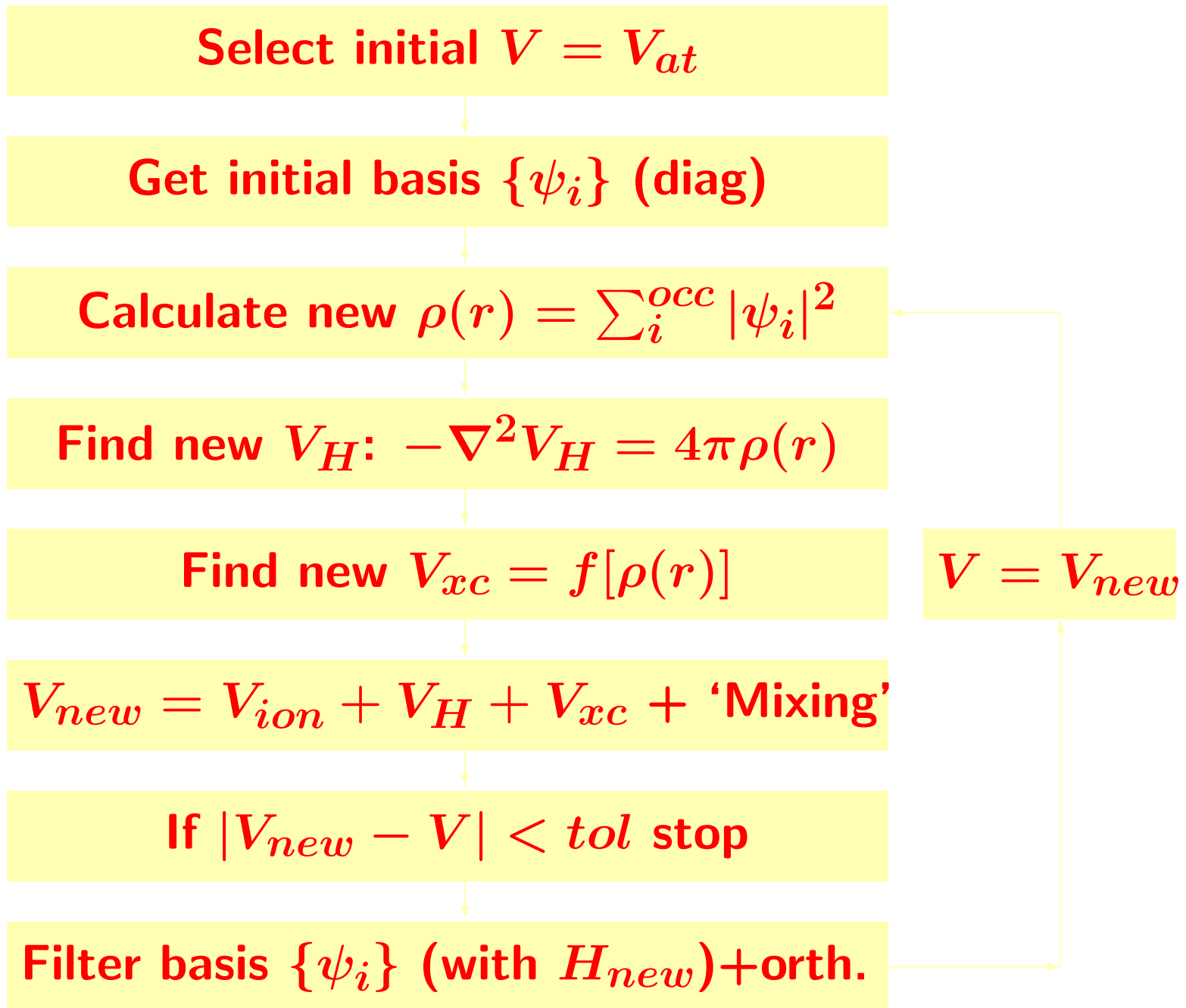
▶▶ Filter polynomials: if $[a, b]$ is interval to dampen, then

$$p_k(t) = \frac{C_k(l(t))}{C_k(l(c))}; \quad \text{with} \quad l(t) = \frac{2t - b - a}{b - a}$$

● $c \approx$ eigenvalue farthest from $(a + b)/2$ – used for scaling

▶▶ 3-term recurrence of Chebyshev polynomial exploited to compute $p_k(A)v$. If $B = l(A)$, then $C_{k+1}(t) = 2tC_k(t) - C_{k-1}(t) \rightarrow$

$$w_{k+1} = 2Bw_k - w_{k-1}$$



Reference:

Yunkai Zhou, Y.S., Murilo L. Tiago, and James R. Chelikowsky, **Parallel Self-Consistent-Field Calculations with Chebyshev Filtered Subspace Iteration**, *Phy. Rev. E*, vol. 74, p. 066704 (2006).

[See <http://www.cs.umn.edu/~saad>]

Chebyshev Subspace iteration - example

*Si*₉₀₄₁*H*₁₈₆₀

n_{state}	# $A * x$	# SCF	$\frac{total_eV}{atom}$	1st CPU	total CPU
19015	4804488	18	-92.00412	102.12 h.	294.36 h.

PEs = 48; $n_H = 2,992,832$. $m = 17$ for Chebyshev-Davidson; $m = 8$ for CheFSI.

Done in 2006 –

Iron clusters [symmetry of 12]

Fe_{388}

n_{state}	# $A * x$	# SCF	$\frac{total_eV}{atom}$	1st CPU	total CPU
2328×2	18232215	187	-795.247	16.22	247.05 h.

Fe_{388}

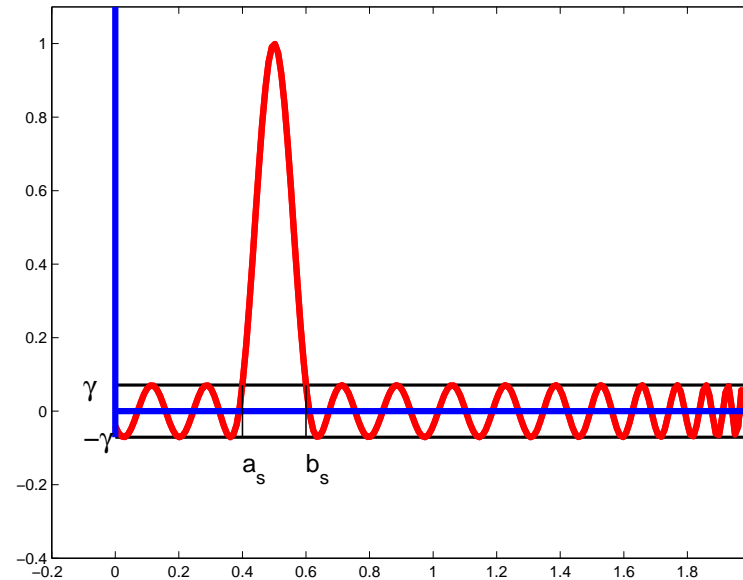
#PE= 24. $n_H = 3332856$. $m = 20$ for Chebyshev-Davidson; $m = 18$ for CheFSI.

Reference:

M. L. Tiago, Y. Zhou, M. M. G. Alemany, YS, and J.R. Chelikowsky, The evolution of magnetism in iron from the atom to the bulk, Physical Review Letters, vol. 97, pp. 147201-4, (2006).

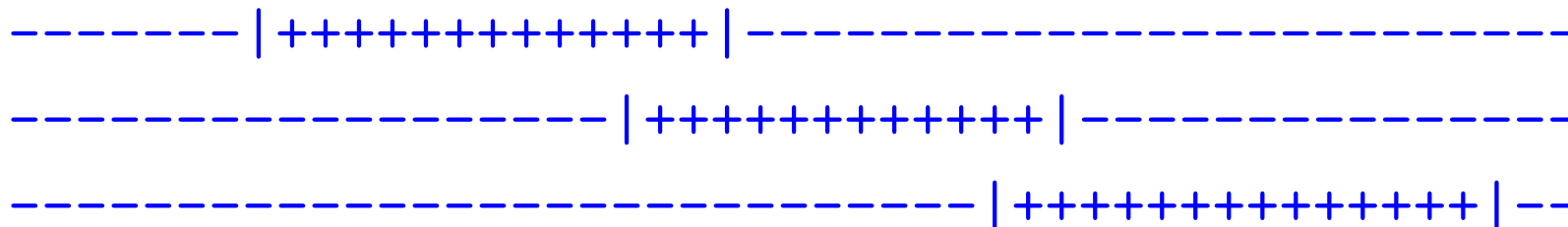
Plans: Spectrum slicing

- ▶ Idea: compute spectrum by pieces.
- ▶ No orthogonalization between pieces which are not nearest neighbors



Main issues:

- 1) Make sure no eigenvalues are missed
- 2) that there are no duplicates
- 3) that process is cost effective



Matlab version: RSDFT

- ▶▶ Goal is to provide (1) prototyping tools (2) simple codes for teaching Real-space DFT with pseudopotentials
- ▶▶ Can do small systems on this laptop – [Demo later?]
- ▶▶ Idea: provide similar input file as PARSEC –
- ▶▶ Can also enter data online
- ▶▶ Many summer interns helped with the project:

Olivier Cots, Yuelian Jia, Sam Handler, Virginie Audin, Long Bui, Nate Born, Amy Coddington, Nick Voshell, Adam Jundt, ...

+ ... others who worked with a related visualization tool (PVOX)

Important step: First iteration of SCF

Issue: good initial guess is needed – else convergence can be delayed – possibly even compromised (?)

- ▶▶ Remedy: do a full diagonalization in first SCF step..
- ▶▶ Far more desirable: **completely bypass** diagonalization
- ▶▶ First alternative: use a memory efficient eigenproblem solver. [example: (linear) subspace iteration]
- ▶▶ Second alternative: use continuation/ homotopy

Use of continuation

- ▶▶ Perform continuation on the charge density:

$$\rho_{new} = \lambda \rho_{out} + (1 - \lambda) \rho_{old}$$

- ▶▶ Initial λ can be set to say 1.0 or 0.5 – should converge to unity with SCF.
- ▶▶ Can set λ manually ... somewhat arbitrarily
- ▶▶ Can also try to set λ automatically
- ▶▶ Criterion: Near convergence $\rho_{old} \approx \rho_{out}$, use $\lambda \approx 1$.

▶▶ We select λ based on degree of variation between ρ_{out} and ρ_{old} :

$$\lambda = \cos^2 (\theta(\rho_{out}, \rho_{old}))$$

▶▶ So with $\theta \equiv \theta(\rho_{out}, \rho_{old})$ we have

$$\rho_{new} = \rho_{out} \cos^2 \theta + \rho_{old} \sin^2 \theta$$

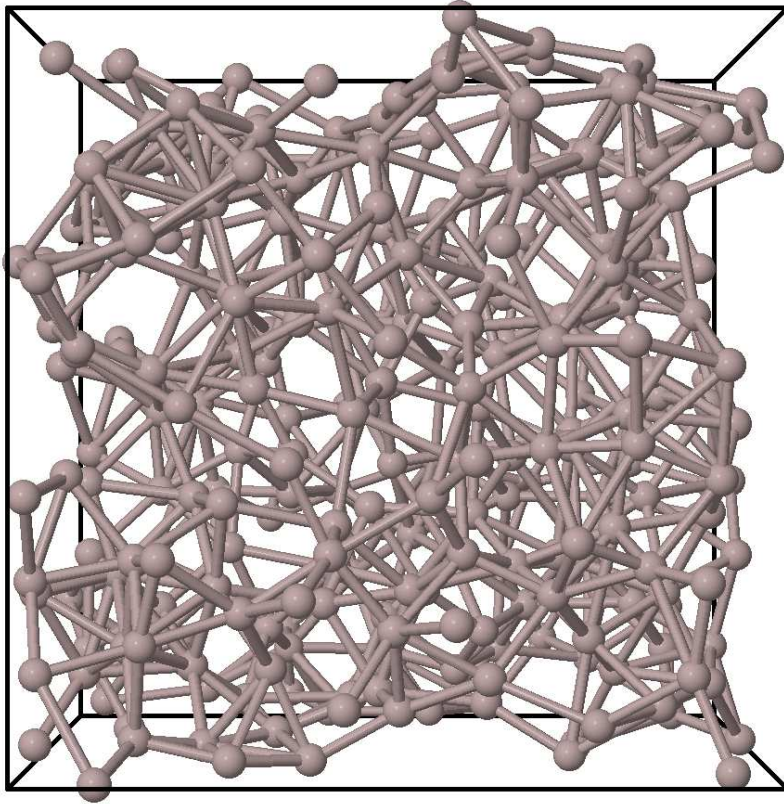
▶▶ Note: this continuation is combined with standard mixing which acts on potential (Broyden, secant, Anderson, etc).

Results

Test with simple algorithm using

$$\lambda = \frac{(\rho_{out}, \rho_{old})^2}{\|\rho_{out}\|^2 \|\rho_{old}\|^2}$$

on models of 300 random atoms of Aluminum [Done within First-principles Molecular Dynamics (MD) simulation for determining melting properties]



Snapshot of model with 300 atoms of Aluminum used in a molecular dynamics simulation (melting of Al).

SCF Iters.	Total Energy [Ry]	λ
1	8371.13518513	0.500000000
2	-809.98294405	0.698983019
3	-1093.14471242	0.965543975
4	-1050.40221019	0.964408629
5	-1151.94407198	0.967302532
6	-1114.00738274	0.929312880
7	-1184.58405142	0.942015013
⋮	⋮	⋮
10	-1232.61311896	0.985936605
⋮	⋮	⋮
20	-1248.00273541	0.999296141
⋮	⋮	⋮
30	-1250.82621163	0.999999860
⋮	⋮	⋮
40	-1250.82629339	0.999999999

Nonlinear eigenvalue problems

▶▶ In numerical linear algebra a nonlinear eigenvalue problem is something like

$$\Phi(\lambda)u = 0 \quad \text{with}$$

$$\Phi(\lambda) = \sum_{i=1}^m A_i \lambda^i$$

[Each A_i is $n \times n$]

▶▶ SCF methods involve nonlinear eigenvalue problems of a different kind:

$$[A + V(U)]U = U\Lambda$$

where U is an orthonormal set of eigenvectors of $A + V(U)$.

▶▶ $V(U)$ depends only on the space spanned by U -

Example: Hartree Fock

$$H\Psi = -\frac{1}{2}\nabla^2 + V_0 + V_H + V_x$$

- As before:
- ▶▶ V_0 (ionic potential)
 - ▶▶ V_H (Hartree potential)
 - ▶▶ V_x is a 2-electron term -

$$V_x \cdot \psi_i(r) = \int \frac{\sum_j \psi_j(r')^* \psi_j(r) \psi_i(r')}{|r - r'|} dr'$$

DFT replaces 2-electron term by a one-electron contribution V_{xc}

Model problems

Question: Can we find problems of this type that are fairly representative of SCF methods and the physics, **without the complications?**

Motivation:

- ▶▶ can study theoretical questions [see J. Meza and C. Yang]
- ▶▶ can develop and test algorithms quickly [Nonlinear eigenvalue problems, 'mixing', $O(N)$ methods,...]
- ▶▶ can help understanding nature of the SCF problem

Example: Can look at a model problem from a material requiring only a local Pseudo-Potential - e.g.: Sodium

- Advantage: close to the physics - simple to implement.
- Disadvantage: Just one example - not amenable to variations

Example: [using matlab notation]

$$\left[-\frac{1}{2}\nabla^2 + V_0 + \text{diag}(|u|^2)\right]u = \lambda u$$

Or: (Notation: $\rho(U) = \text{diag}(UU^H)$)

$$\left[-\frac{1}{2}\nabla^2 + V_0 + L^{-1}\rho(U)\right]u_i = \lambda_i u_i$$

►► Note: Always exists a solution to

$$\min_{U^T U = I} \text{Tr} \left[U^T \left(-\frac{1}{2}\nabla^2 + D(U) \right) U \right]$$

►► Extreme simplifications of typical SCF problem → not clear if these satisfy the requirements.

►► Difficulty: find situations that reflect the issue of the 'gap' in SCF [hard convergence for metallic systems]