Adaptive compression for Hartree-Fock-like equations

Michael Lindsey, UC Berkeley
(Joint work with Lin Lin)

May 5, 2018

Outline

- Setup and motivation
- Describe the adaptive compression method
- Numerical results
- Discuss key properties of adaptively compressed operators
- Local convergence
- Global convergence
- See Lin Lin, M.L. *Comm. Pure Appl. Math* (in press)

Our setting

We consider the linear eigenvalue problem (lowest n eigenvalues)

$$(A + B)v_i = \lambda_i v_i, \quad i = 1, \ldots, n,$$

where

- $A, B \in \mathbb{C}^{N \times N}$ Hermitian, $N \gg n \gg 1$
- $\|A\|_2 \gg \|B\|_2$
- $B \prec 0$ (note: can be guaranteed by level shifting)
- A-multiplies are 'easy,' B-multiplies are 'hard' (bottleneck)

Hartree-Fock-like equations

By HF-like equations we mean the nonlinear eigenvalue problem

$$H[P]\psi_i = \left(-\frac{1}{2}\Delta + V_{\text{ion}} + V_{\mathsf{Hxc}}[P] + V_X[P]\right)\psi_i = \varepsilon_i \psi_i,$$

$$\int \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\,\mathrm{d}\mathbf{r} = \delta_{ij}, \quad P(\mathbf{r},\mathbf{r}') = \sum_{i=1}^{N_e} \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}'),$$

where

- N_e is the number of electrons
- P is the density matrix, an orthogonal projector of rank N_e
- $V_{\mathsf{Hxc}}[P]$ depends only on the density $\rho(\mathbf{r}) := P(\mathbf{r},\mathbf{r})$ (includes Hartree and exchange-correlation contributions)
- V_X is an integral operator with kernel

$$V_X[P](\mathbf{r},\mathbf{r}') = -P(\mathbf{r},\mathbf{r}')K(\mathbf{r},\mathbf{r}')$$

- In HF theory, $K(\mathbf{r},\mathbf{r}') = 1/|\mathbf{r}-\mathbf{r}'|$

Hartree-Fock-like equations

- Note: $-V_X[P]$ is a Hadamard product of positive definite operators, hence $V_X[P] \prec 0$
- But $V_X[P]$ is dense and not low-rank
- $V_X[P]$-multiplies often take over 95% of computational time
- Density matrix P needs to be computed self-consistently
- Can fix P so that $H[P]$ is fixed and solve linear eigenvalue problem
- Iterate to solve nonlinear fixed-point problem for self-consistent P
- Numerical discretization of linear problem yields matrix of form $A + B$ as above ($B \prec 0$)

Reminder: our setting

We consider the linear eigenvalue problem (lowest n eigenvalues)

$$(A + B)v_i = \lambda_i v_i, \quad i = 1, \ldots, n,$$

where

- $A, B \in \mathbb{C}^{N\times N}$ Hermitian, $N \gg n \gg 1$
- $\|A\|_2 \gg \|B\|_2$
- $B \prec 0$
- A-multiplies are 'easy,' B-multiplies are 'hard'
- Also, define gap $\lambda_g := \lambda_{n+1} - \lambda_n$

Adaptively compressing operators

- Let $V = [v_1, \ldots, v_n]$, linearly independent columns
- Define $\underline{B}[V] =:= BV(V^*BV)^{-1}V^*B$ (rank n)
- Observe $\underline{B}[V]V = BV$
- Therefore, if $v_1, \ldots, v_n$ are the lowest n eigenvectors of $A + B$...
- then they are eigenvectors of $A + \underline{B}[V]$ as well!
- But are they the lowest eigenvectors?
- **Note:** keep this convention for V

Adaptively compressing operators

- Let P be orthogonal projector onto $\mathrm{span}[V]$
- P encodes subspace as a matrix, useful for analyzing convergence of subspaces
- Compression depends only on subspace, define $\underline{B}[P] = \underline{B}[V]$

Adaptive compression method

- Solve

$$(A + \underline{B}[V^{(k)}])v_i^{(k+1)} = \lambda_i^{(k+1)} v_i^{(k+1)}, \quad i = 1, \ldots, n,$$

  $V^{(k)} = [v_1^{(k)} \cdots v_n^{(k)}]$
  $P^{(k)}$ orthogonal projector onto $\mathrm{span}[V^{(k)}]$
- In other words, given $P^{(k)}$, define $P^{(k+1)}$ to be the orthogonal projector onto the span of the lowest n eigenvectors of $A + \underline{B}[P^{(k)}]$
- Hopefully $P^{(k)} \to P$
- Solve 'easy' eigenvalue problem at each step, though we have exchanged a linear eigenvalue problem for a nonlinear fixed-point iteration!
- Need to know that P is a fixed point, equivalently that 'lowest-n space' of $A + \underline{B}[P]$ is the same as that of $A + B$
- Would be true if $\underline{B}[P] \succeq B$ (more on this later)

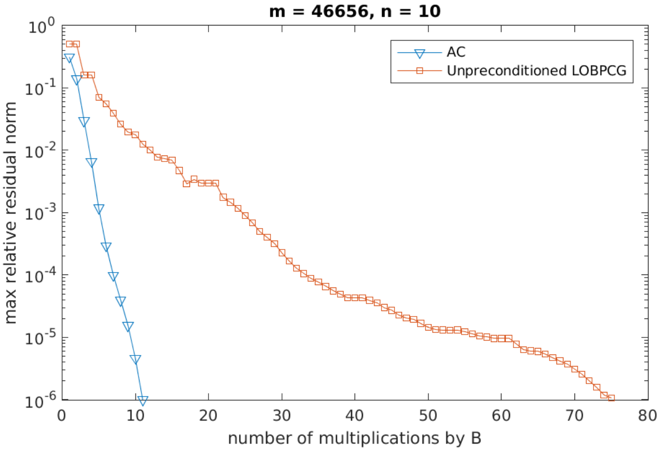
Extensions

- An extension of results allows us to consider ‘metallic’ case of nearly degenerate eigenvalues
- For linear problems, can accelerate adaptive compression fixed point iteration via DIIS ideas
- For (nonlinear) HF-like equations, can delay update of adaptively compressed operator to outer loop for additional speedup
- Framework can be adapted to other structured eigenvalue problems

Numerical results (linear problem)

[also joint with Jeffmin Lin at UC Berkeley]

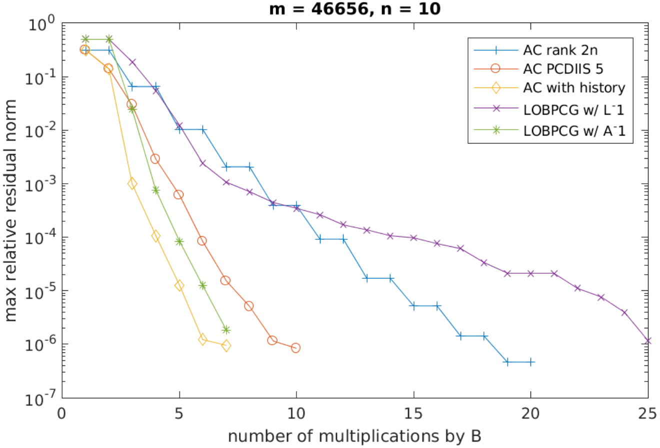
$$A = -\frac{1}{2}\Delta + V, \quad B(x, y) = 1/\sqrt{1 + |x - y|^2}$$



Numerical results (linear problem)

[also joint with Jeffmin Lin at UC Berkeley]

$$A = -\frac{1}{2}\Delta + V, \quad B(x,y) = 1/\sqrt{1+|x-y|^2}$$



Numerical results (nonlinear: DFT with hybrid functional)

- inner iteration to converge density $\rho(\mathbf{r})$
- outer iteration to convergence density matrix $P(\mathbf{r}, \mathbf{r}')$
- only update compressed exchange operator once per outer loop

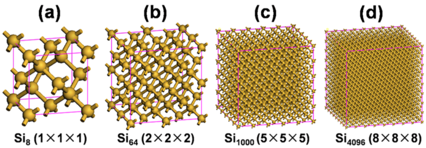


Figure 3. Geometric structures of bulk silicon systems used for the ACE formulation in PWDFT. (a) Si_8 in the unitcell, (b) Si_{64} in the $2 \times 2 \times 2$ supercell, (c) Si_{1000} in the $5 \times 5 \times 5$ supercell, and (d) Si_{4096} in the $8 \times 8 \times 8$ supercell.

methods	ACE HSE06 (LOBPCG)		conventional HSE06 (LOBPCG)	
no. outer SCF	no. inner SCF	time (s)	no. inner SCF	time (s)
1	6	356	6	2518
2	5	320	5	2044
3	5	308	4	1665

Hu, Lin, Banerjee, Vecharynski, & Yang. *JCTC* (2017)

Optimality of adaptive compression

Theorem (Optimality)

For $B \prec 0$ and any $N \times n$ matrix U with linearly independent columns, the adaptive compression $\underline{B}[U]$ is the unique rank-n Hermitian matrix that agrees with B on $\mathrm{span}[U]$. Furthermore, $B \preceq \underline{B}[U] \preceq 0$.

Helpful observation: matrix of $\underline{B}[U]$ (with respect to $\mathrm{span}[U] \times \mathrm{span}[U]^\perp$ block structure) is

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{12}^* B_{11}^{-1} B_{12} \end{pmatrix}$$

Local convergence

Theorem (Local convergence)

The fixed point iteration $P^{(k)} \mapsto P^{(k+1)}$ converges locally to $P = VV^*$. The number of matrix-vector multiplications Bv needed for k steps of fixed point iteration is nk. Starting from $P^{(0)}$, the asymptotic convergence rate is

$$\|P - P^{(k)}\|_2 \lesssim \gamma^k \|P - P^{(0)}\|_2, \quad \text{where} \quad \gamma \leq \frac{\|B\|_2}{\|B\|_2 + \lambda_g}.$$

Note: rate of convergence does not depend on $\|A\|_2$ (stable under refinement of discretization if A is a differential operator). Can also obtain sharp expression.

Idea of proof

- Let $\mathfrak{F}$ be iteration map of adaptive compression method, so $\mathfrak{F} : P^{(k)} \to P^{(k+1)}$
- Repeated application of $\mathfrak{F}$ induces discrete-time dynamical system on the set of rank-n orthogonal projectors, equivalently the set of n-dimensional subspaces or the Grassmann manifold $\mathbf{Gr}(n, \mathbb{C}^N)$
- Linearize of dynamical system at fixed point P by computing Jacobian of $\mathfrak{F}$ in appropriate coordinates
- Bound eigenvalues of Jacobian to get asymptotic convergence rate

Global convergence

Theorem (Global convergence)

For almost every pair of Hermitian matrices A, B *with* $B \prec 0$*, the fixed point iteration* $\mathfrak{F} : P^{(k)} \mapsto P^{(k+1)}$ *converges globally to* $P = VV^*$ *for almost every initial guess* $P^{(0)}$*.*

Simple counterexamples show that there can exist 'bad' fixed points, so one cannot hope for convergence from arbitrary initial guess.

Idea of proof

- **Eigenvalue monotonicity:** each of the bottom n eigenvalues $\lambda_1^{(k)}, \ldots, \lambda_n^{(k)}$ of $A + \underline{B}[P^{(k-1)}]$ is monotonically non-increasing in k (hence convergent)
- Change of $P^{(k)}$ across one iteration can be controlled by the change of $\sum_{i=1}^n \lambda_i^{(k)}$
- So when k is large, $\mathfrak{F}(P^{(k)}) \approx P^{(k)}$, i.e., the point $P^{(k)}$ is almost fixed by the mapping $\mathfrak{F}$
- Not yet enough to directly imply by general considerations that the sequence $P^{(k)}$ is convergent, but one might hope that a point that is close to being fixed is close to some fixed point!
- This is true

Idea of proof

- Fixed points P_f are in particular orthogonal projectors onto invariant subspaces of $A + B$
- For generic A, B, there are finitely many such P_f and $A + \underline{B}[P_f]$ has a spectral gap
- Already know that $P^{(k)}$ is close to some fixed point for k large. Fixed points are finite in number (hence isolated), so this implies that $P^{(k)}$ converges to some fixed point
- Use spectral gap to perform linearization about all fixed points
- Only the desired fixed point is stable; stable manifold of all others has positive codimension (hence measure zero)
- Convergence to bad fixed point would mean that dynamics live on such a stable manifold for k large enough
- Thus want to show that the preimage under $\mathfrak{F}$ of a measure-zero set has measure zero
- This is the **egg on barn lemma** (technical; $\mathfrak{F}$ not a diffeomorphism, nor even continuous)

Conclusion

- Introduced adaptive compression framework
- Applications to linear and nonlinear eigenvalue problems
- Compression operation has nice linear-algebraic properties
- Convergence theory
- Acknowledgement: NSF GRFP (DGE-1106400)
- Thank you!