

Scalable Eigensolvers for Excited States Calculations

EMWG Position Paper
submitted by Zhaojun Bai¹ and Ren-Cang Li²
April 29, 2013

Excited states calculation is an important tool in the study of collective motion of many particle systems, ranging from silicon nanoparticles and nanoscale materials to the analysis of interstellar clouds. In computational quantum physics and chemistry, the excitation states are described by the linear response perturbation analysis in the time-dependent density functional theory (TDDFT). There are immense interests in developing efficient numerical algorithms and simulation techniques for excitation state calculations of molecules and solids for materials design in energy science; see, for example, recently funded DOE SciDAC BES projects (<http://www.ornl.gov/scidac3pi2012/agenda.htm>).

TDDFT is an extension of the density functional theory (DFT), and is used to investigate the properties and dynamics of many-body systems in the presence of time-dependent potentials, such as electric or magnetic fields [1]. The effect of such fields on molecules and solids can be studied with TDDFT to extract features like excitation energies, frequency-dependent response properties, and photoabsorption spectra. Although the conceptual and computational foundations of DFT and TDDFT are analogous, TDDFT calculations are more complex and expensive, most notably because the time-dependent effective potential at any given instant depends on the density value at previous time. To simplify the TDDFT calculation, one assumes that the external perturbation is small in the sense that it does not completely destroy the ground-state structure of the system. In this case, it is sufficient to analyze the linear response of the system. The variation of the system will then depend only on, to the first order, the ground-state wave-functions [1].

The computational kernel of the linear-response TDDFT, such as in electronic structure calculation packages CASTEP, QUANTUM EXPRESSO and OCTOPUS, is to solve the following so-called *Linear Response Eigenvalue Problem* (LREVP):

$$\begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} \Sigma & \Delta \\ \Delta & \Sigma \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix},$$

where A and B are $n \times n$ real symmetric matrices and $A \pm B$ are positive definite, Σ and Δ are also $n \times n$ with Σ being symmetric while Δ skew-symmetric (i.e., $\Delta^T = -\Delta$) and $\Sigma \pm \Delta$ are nonsingular. It can be shown that the eigenvalues of LREVP are real and appear in pairs $\pm\lambda$. The first few smallest positive eigenvalues and the corresponding eigenvectors give rise to the approximates of the excitation energies of the molecular system. Due to the nature of the linear response TDDFT, the order of LREVP is very large. For example, a plane-wave based TDDFT calculation for the excited energies of fullerene C_{60} in QUANTUM EXPRESSO can lead to an LREVP of the order 22 millions [6].

Since the linear response theory was proposed by Bohm and Pines for studying the collective motion of many particles in the early 1950s, the development of numerical methods for solving the LREVP has been an active research subject in computational (quantum) physics and chemistry, and in numerical analysis community for over four decades. A 2009 survey study [2] compared four numerical methods

¹Department of Computer Science and Department of Mathematics, University of California, Davis, CA 95616. Bai is also a Faculty Scientist of Scientific Computing Group at LBL. bai@cs.ucdavis.edu

²Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019, rc.li@uta.edu.

(namely Lanczos, Arnoldi, Davidson, and CG), and discussed the limitations of each of these methods for developing an efficient **scalable** eigensolver. In the study, severe limitations were experienced for the Lanczos-type methods due to the orthogonality constraints, for the CG type methods to compute several eigenpairs simultaneously, and for incorporating preconditioning techniques.

In the past two years, we have conducted a systematical study of the theory, computation, and applications of the LREVP. We uncovered new and surprising minimization principles to characterize the eigenvalues of interest. Although the LREVP is a non-Hermitian eigenvalue problem, these theoretical results mirror the well-known results for the Hermitian eigenvalue problem [4].

With the help of these new theoretical results, we are able to develop the *best* (possible) approximations of few smallest positive eigenvalues via structure-preserving projection, and derive locally optimal steepest descent and conjugate gradient-like algorithms, based on a novel 4D search idea instead of the usual line-search, as well as block algorithms for simultaneously computing these eigenvalues and their associate eigenvectors [3, 5]. A successful excited state calculation of the order 22 millions of the LREVP for fullerene C₆₀ by our methods was recently presented in [6].

The newly developed locally optimal block 4D CG algorithms allow us to use blocking strategies to perform multiple computation steps of the algorithm for each communication step and to incorporate proper preconditioners for fast convergence. Furthermore, unlike the Arnoldi and Lanczos type Krylov subspace methods, the block 4D CG-like algorithms are memory-efficient, and numerical convergence behavior is less stringent on the (orthogonal-like) normalization constraints of projection subspaces. However, towards a fully communication-avoiding approach for exascale computing, we must derive new parallel computation and communication strategies. On modern computer architectures, communication costs in terms of the performance of an algorithm can be much greater than arithmetic computation costs, and the gap is only going to widen in future systems [7]. One of our research issues is how to efficiently computing matrix-vector and matrix-matrix operations by exploiting the matrix (sub)-structures and sparsity of the LREVP. Another issue is how to hide costly global synchronization latency. This is in fact a challenge for the scalability for all preconditioned CG and Krylov subspace type eigensolvers. Furthermore, for the new LREVP solver to be integrated into existing electronic structure calculation packages, we should also study how to provide flexible interfaces which can easily incorporate physical-based preconditioners.

REFERENCES

- [1] M. E. Casida. Time-dependent density-functional response theory for molecules. In D.P. Chong, editor, *Recent advances in Density Functional Methods*, pp.155-189, World Scientific, Singapore, 1995.
- [2] S. Tretiak, C. M. Isborn, A. M. N. Niklasson, and M. Challacombe. Representation independent algorithms for molecular response calculations in time-dependent self-consistent field theories. *J. Chem. Phys.*, 130(5), p.054111, 2009.
- [3] D. Rocca, Z. Bai, R.-C. Li and G. Galli, A block variational procedure for the iterative diagonalization of non-Hermitian random-phase approximation matrices, *J. Chem. Phys.*, 136, p.034111, 2012.
- [4] Z. Bai and R.-C. Li, Minimization principles of the linear response eigenvalue problem I: Theory, *SIAM J. Matrix Anal. Appl.*, 33(4), pp.1075-1100, 2012.
- [5] Z. Bai and R.-C. Li, Minimization principles of the linear response eigenvalue problem II: Computation, *SIAM J. Matrix Anal. App.*, 34(2), pp.392-416, 2013.
- [6] D. Rocca. Iterative diagonalization of non-Hermitian eigenproblems in time-dependent density functional and many-body perturbation theory, APS Annual Meeting, Session B39, Boston, March 2012.
- [7] J. Demmel, M. Hoemmen, M. Mohiyuddin, and K. Yelick. Avoiding communication in sparse matrix computations. in *Proceedings of IEEE International Symposium on Parallel and Distributed Processing*, pp.112, 2008.