Massively Parallel Eigensolvers based on Unconstrained Energy Functionals Methods

Osni Marques, Mauro Del Ben and Andrew Canning

[oamarques,mdelben,acanning]@lbl.gov

Lawrence Berkeley National Laboratory

ABSTRACT

This poster focuses on a preconditioned conjugate gradient based iterative eigensolver using an unconstrained energy functional minimization scheme. This scheme avoids an explicit reorthogonalization of the trial eigenvectors and becomes an attractive alternative for the solution of very large problems. The unconstrained formulation is implemented in the first-principles materials and chemistry CP2K code [1], which performs electronic structure calculations based on a density functional theory approximation to the solution of the many-body Schrödinger equation. The systems we use in our studies have a number of atoms ranging from 2,247 to 12,288. We study the convergence of the unconstrained formulation and its scaling on a Cray XC40 (a partition with 9,688 Intel KNL nodes). We show that there is a trade-off between the preconditioner that leads to fast convergence and lower cost preconditioners that lead to best time to solution.

1 INTRODUCTION

Many scientific applications require the solution of eigenvalue problems. When some small percentage of the eigenpairs is required rather than the full spectrum, iterative eigensolvers are typically used. This is because the computational cost of a direct solver scales as the cube of the matrix dimension, while iterative solvers scale as the square of the number of required eigenpairs times the matrix dimension. The prefactor for the scaling of the iterative solver is much larger than for the direct solver. The crossover point for which method becomes the fastest varies with the class of systems under study. Electronic structure calculations done in codes based on an approximate solution to the Schrödinger equation is an example of this class of problems, where typically a small percentage of the lowest eigenpairs is required. Iterative methods are often used for this class of problem, as they can scale well on large parallel computers for large problems. Most of the operations in the algorithms are large matrix multiplies, where at least one dimension of the matrices has the dimension of the full matrix being diagonalized (solution of and eigenvalue problem). The commonly used iterative solvers for materials and chemistry codes are often referred to as constrained energy functional approaches, as the orthogonality of the eigenvectors requires some form of reorthogonalization of the trial vectors. This can result in a poor parallel scaling of the reorthogonalization step. In contrast, unconstrained energy functionals avoid the explicit reorthogonalization step and have a potential for better parallel scaling than both direct eigensolvers and constrained energy functional iterative solvers.

2 FORMALISM AND ALGORITHMS

First-principles methods based on Density Functional Theory (DFT) in the Kohn-Sham (KS) formalism [2] are widely used for electronic structure calculations in materials science and chemistry, due to their computational efficiency and favorable scaling with system size. The KS single particle eigenfunction equations are usually written in atomic units as

$$\hat{H}\psi_i(\mathbf{r}) = \left[-\frac{1}{2}\nabla^2 + \mathbf{V}\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{1}$$

where $\hat{H}$ is the Hamiltonian, $\psi_i(\mathbf{r})$ are the wavefunctions for each electron, ε_i is the energy of the electron, ∇^2 is the kinetic energy operator, and $\mathbf{V}$ is the potential. The probability of finding the i'th electron at position $\mathbf{r}$ is given by $\psi_i(\mathbf{r})\psi_i^*(\mathbf{r})$, where $*$ denotes the complex conjugate.
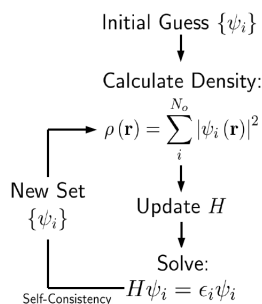

Figure 1: SCF method.

The commonly used approach for solving the nonlinear eigenvalue problem (1) is the self-consistent field (SCF) method. The problem is linearized in an inner loop by fixing the charge density, which is then updated at each step until convergence of the charge density and potential field. One starts from an initial guess for $\psi_i(\mathbf{r})$ in the first step, from which a total charge density and $\hat{H}$ can be calculated. See Fig. 1.

To solve Eq. (1) $\psi_i(\mathbf{r})$ are usually expanded in a basis set or discretized on a real space grid. Commonly used basis are plane waves (Fourier expansion) and atom centered Gaussian functions. Using a set of basis functions $\{\phi_\alpha(\mathbf{r})\}$ of size N_b, one can write $\psi_i(\mathbf{r}) = \sum_\alpha \mathbf{C}_{\alpha i}\phi_\alpha(\mathbf{r})$. If $\{\phi_\alpha(\mathbf{r})\}$ are non-orthogonal, the inner part of the SCF loop (see Fig. 1) requires the solution of the associated generalized eigenvalue problem

$$\mathbf{HC} = \mathbf{SCE} \tag{2}$$

where $\mathbf{C}$ is the matrix of generalized eigenvectors (coefficients of the basis functions) obeying the orthogonality condition $\mathbf{C}^T\mathbf{SC} = \mathbf{I}$, and $\mathbf{E}$ is the diagonal matrix of eigenvalues. For an orthonormal basis $\mathbf{S} = \mathbf{I}$ and (2) becomes a standard eigenvalue problem.

With a CG formulation, problem (2) can be cast in terms of a minimization problem subject to the orthogonality constraint,

$$min \ \text{Tr}\left[\mathbf{C}^T\mathbf{HC}\right], \quad \mathbf{C}^T\mathbf{SC} = \mathbf{I}. \tag{3}$$

$\mathbf{C}$ is now the matrix of trial generalized eigenvectors with dimension $N_b \times N_o$, where N_b is the size of the basis set and N_o is the number of wavefunctions. At the solution, we obtain the total electronic energy of the system, $\sum_i \varepsilon_i$.

The unconstrained energy functional approach offers the possibility of eliminating explicit diagonalization and operations on the small overlap matrix. We define the transformation $\mathbf{C} = \mathbf{X}\mathcal{S}^{-\frac{1}{2}}$, $\mathcal{S} = \mathbf{X}^{\mathbf{T}}\mathbf{X}$, where $\mathbf{X}$ spans the same subspace as $\mathbf{C}$ but is not required to be orthogonal. Note that we use $\mathbf{S}$ for the $N_b \times N_b$ overlap matrix of the basis functions and $\mathcal{S}$ for the $N_o \times N_o$ overlap matrix of the trial vectors $\mathbf{X}$, where $N_o \ll N_b$. The minimization problem of Eq. (3) can now be rewritten as

$$E[\mathbf{X}] = \min \mathrm{Tr}\left[\mathcal{S}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{HX}\right], \qquad (4)$$

relaxing the orthogonality constraint and introducing the so-called unconstrained energy functional $E[\mathbf{X}]$. Near the solution, $\mathcal{S}^{-1}$ is close to the identity matrix $\mathbf{I}$, and most implementations use the first order expansion $(2\mathbf{I} - \mathcal{S})$ to approximate $\mathcal{S}^{-1}$, therefore avoiding a matrix inversion. (See [3, 4] for related work.) Notably, approaches based on expansions of $\mathcal{S}^{-1}$ never involve operations solely on the small $\mathcal{S}$ matrix and this is a key ingredient for improved parallel scaling. A caveat is that the removal of the orthogonality constraint, and a different functional to minimize, affects convergence. In (4), $E[\mathbf{X}]$ can be expressed as $E[\mathbf{X}] = \mathrm{Tr}\left[\left(2\mathbf{I} - \mathbf{X}^T\mathbf{S}\mathbf{X}\right)\mathbf{X}^T\mathbf{HX}\right]$, whose gradient is $\mathbf{G} = 4\mathbf{HX} - 2\mathbf{SX}\mathcal{H} - 2\mathbf{HX}\mathcal{S}$, where $\mathcal{H} = \mathbf{X}^{\mathbf{T}}\mathbf{HX}$ (i.e. $N_o \times N_o$). It has been shown that the minimization of this functional with respect to $\mathbf{X}$ converges to the same energy as that of the correct functional, that at convergence the subspace spanned by $\mathbf{X}$ is the same as that spanned by the true eigenvectors $\mathbf{C}$, and at the minimum $\mathbf{X}^T\mathbf{SX} = \mathbf{I}$ [5–8].

We use a preconditioned conjugate gradient (PCG) procedure to minimize the energy functional, and the Polak-Ribière formula for computing the parameters in the line searches. We have explored different preconditioners for $\mathbf{G}$; in particular, by using $\mathbf{A}$, the Hessian of the unconstrained functional, and applying $\mathbf{A}^{-1}$ to $\mathbf{G}$ (in a Newton-like step). In practice, computing $\mathbf{A}^{-1}$ is not feasible and we obtain its action by solving an iterative system with $\mathbf{A}$, up to K steps. For preconditioning the inner PCG we can use either $\mathbf{S}^{-1}$, which is robust and cheap to compute, or $\tilde{\mathbf{A}}^{-1}$, where $\tilde{\mathbf{A}}$ is an approximation for the Hessian. We name these strategies `FullHess-KIter-InvOvl` and `FullHess-KIter-AprxHess`, respectively. Alternatively, we can turn off the preconditioned with $\mathbf{A}$ and only use either $\mathbf{S}^{-1}$ or $\tilde{\mathbf{A}}^{-1}$ as preconditioner. We name these strategies `InvOvl` and `Iter-AprxHess`, respectively. Then, we obtain a set of strategies as shown in Fig. 2, where the left branch indicates inner preconditioning only, and the right branch indicates preconditioning with $\mathbf{A}$. A detailed discussion about these preconditioners is given in [9].
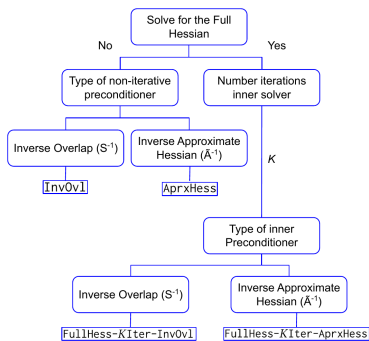

Figure 2: Set of preconditioners studied in this work.
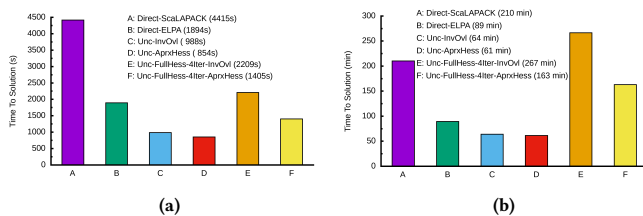

Figure 3: Time to solution for full SCF convergence using unconstrained minimization and setups of Fig. 2, compared to direct solvers (ScaLAPACK and ELPA). (a) Water-1024, (b) BiLayer. The actual times are given in parenthesis.
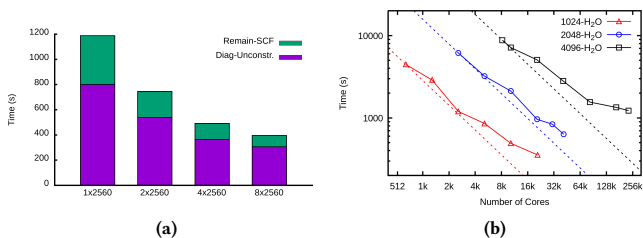

Figure 4: Strong scaling, method D in Fig. 3. (a) OpenMP threads per MPI task for a fixed number of MPI tasks (2560), bulk liquid water with 1024 molecules. (b) Time to solution for bulk liquid water with 1024, 2048 and 4096 molecules. The largest simulation used roughly 38% of the full machine.

3 NUMERICAL EXPERIMENTS

All calculations have been performed on Cori, at (NERSC) [10]. Cori is a Cray XC40 system with two partitions; the one we used consists of over 9600 nodes, each node has one Intel Xeon-Phi 7250 processor (KNL), connected with an Aries interconnect. Our implementation is a hybrid MPI+OpenMP, optimal for many-core architectures such as Cori. The CP2K software package was compiled with the Intel compiler and linked to the latest available versions of MKL, ELPA [11], and LIBXSMM [12] (a library optimized for small matrix-matrix multiplications).

The systems used in the numerical experiments are: (a) 1024 molecules of bulk liquid water (Water-1024), (b) supramolecular catalyst gold(III)-complex (Complex), (c) bilayer of MoS_2-WSe_2 (BiLayer), and (d) divacancy point defect in silicon (SiDivac). These systems present an increasing order of "complexity" for convergence, and their sizes range from 2,247 to 12,288 atoms[1]. Fig. 3 shows the time to solution for full SCF convergence, where one can observe that `Iter-AprxHess` (bar D) leads to best time. Fig. 4 shows strong scaling, in terms of OpenMP threads per MPI task and time to solution, up to 230k cores.

As a final note, the unconstrained approach can be applied to any matrix, and not only in the context of Gaussian basis electronic structure methods (the focus of this poster). In future work we plan to study the unconstrained approach to solve problems in other domains.

---

[1]In this abstract we show results for systems (a) and (c) only; results for all systems are given in the poster.

REFERENCES

[1] CP2K, https://www.cp2k.org, 2019.

[2] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138.

[3] B. G. Pfrommer, J. Demmel, H. Simon, Unconstrained energy functionals for electronic structure calculations, J. Comput. Phys. 150 (1999) 287–298.

[4] F. Corsetti, The orbital minimization method for electronic structure calculations with finite-range atomic basis sets, Comput. Phys. Commun. 185 (2014) 873 – 883.

[5] F. Mauri, G. Galli, R. Car, Orbital formulation for electronic-structure calculations with linear system-size scaling, Phys. Rev. B 47 (1993) 9973–9976.

[6] F. Mauri, G. Galli, Electronic-structure calculations and molecular-dynamics simulations with linear system-size scaling, Phys. Rev. B 50 (1994) 4316–4326.

[7] P. Ordejón, D. A. Drabold, M. P. Grumbach, R. M. Martin, Unconstrained minimization approach for electronic computations that scales linearly with system size, Phys. Rev. B 48 (1993) 14646–14649.

[8] P. Ordejón, D. A. Drabold, R. M. Martin, M. P. Grumbach, Linear system-size scaling methods for electronic-structure calculations, Phys. Rev. B 51 (1995) 1456–1476.

[9] M. Mauro Del Ben, O. Marques, A. Canning, Improved unconstrained energy functional method for eigensolvers in electronic structure calculations, in: 48th International Conference on Parallel Processing, Kyoto, Japan. To appear.

[10] NERSC, http://www.nersc.gov, Last accessed October 14, 2019.

[11] A. Marek, V. Blum, R. Johanni, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H.-J. Bungartz, H. Lederer, The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science, J. Phys. Condens. Matter 26 (2014) 213201.

[12] LIBXSMM, https://github.com/hfp/libxsmm, Last accessed October 14, 2019.