

Pacific Northwest National Laboratory, K8-91, P.O. Box
999, Richland WA 99352, USA

Green's Function Coupled-Cluster Library (GFCCLib)

Bo Peng, Ajay Panyala, Sriram Krishnamoorthy, Karol Kowalski

I. INTRODUCTION

GFCCLib is designed for the Green's function calculation of molecular system at the coupled-cluster level. The library is developed in the C++ programming language with a framework that enables scalability (via scalable design of the algorithm), efficiency via multi-layer parallelism, memory management, and scheduling of tensor operations, and flexibility by providing a long-term sustainable platform for development of methods in this area on machines ranging from workstations to modern supercomputers.

II. GFCC METHODOLOGY

For a review of the GFCC method employed in this work, we refer the readers to Refs. [1]–[6], [8]–[10]. Briefly, the matrix element of the retarded part of the analytical frequency dependent Green's function of an N -electron system can be expressed as

$$G_{pq}^R(\omega) = \langle \Psi | a_q^\dagger(\omega + (H - E_0) - i\eta)^{-1} a_p | \Psi \rangle \quad (1)$$

where H is the electronic Hamiltonian of the N -electron system, $|\Psi\rangle$ is the normalized ground-state wave function of the system, E_0 is the ground state energy, and the a_p (a_p^\dagger) operator is the annihilation (creation) operator for electron in the p -th spin-orbital. Besides, ω is the frequency, η is the broadening factor, and p, q, r, s, \dots refers to general spin-orbital indices (we also use i, j, k, l, \dots to label occupied spin-orbital indices, and a, b, c, d, \dots to label virtual spin-orbital indices). By introducing bi-orthogonal CC formalism, the CC Green's function can then be expressed as

$$G_{pq}^R(\omega) = \langle \Phi | (1 + \Lambda) \bar{a}_q^\dagger(\omega + \bar{H}_N - i\eta)^{-1} \bar{a}_p | \Phi \rangle \quad (2)$$

where $|\Phi\rangle$ is the reference function, and the normal product form of similarity transformed Hamiltonian \bar{H}_N is defined as $\bar{H} - E_0$. The similarity transformed operators \bar{A} ($A = H, a_p, a_q^\dagger$) are defined as $\bar{A} = e^{-T} A e^T$. The cluster operator T and the de-excitation operator Λ are obtained from solving the conventional CC equations. Now we can introduce an ω -dependent IP-EOM-CC type operators $X_p(\omega)$ mapping the N -electron Hilbert space onto an $(N-1)$ -electron Hilbert space

$$\begin{aligned} X_p(\omega) &= X_{p,1}(\omega) + X_{p,2}(\omega) + \dots \\ &= \sum_i x^i(p, \omega) a_i + \sum_{i < j, a} x_a^{ij}(p, \omega) a_a^\dagger a_j a_i + \dots \end{aligned} \quad (3)$$

that satisfies

$$(\omega + \bar{H}_N - i\eta) X_p(\omega) | \Phi \rangle = \bar{a}_p | \Phi \rangle. \quad (4)$$

Substituting this expression into Eq. (2), we end up with a compact expression for the matrix element of the retarded CC Green's function

$$G_{pq}^R(\omega) = \langle \Phi | (1 + \Lambda) \bar{a}_q^\dagger X_p(\omega) | \Phi \rangle, \quad (5)$$

which becomes

$$G_{pq}(\omega) = \langle \Phi | (1 + \Lambda_1 + \Lambda_2) \bar{a}_q^\dagger (X_{p,1}(\omega) + X_{p,2}(\omega)) | \Phi \rangle \quad (6)$$

in the GFCCSD approximation (GFCC with singles and doubles) with $X_{p,1}/\Lambda_1$ and $X_{p,2}/\Lambda_2$ being one- and two-body component of X_p/Λ operators, respectively. The spectral function is then given by the trace of the imaginary part of the retarded GFCCSD matrix,

$$A(\omega) = -\frac{1}{\pi} \text{Tr} [\Im(\mathbf{G}^R(\omega))] = -\frac{1}{\pi} \sum_p \Im(G_{pp}^R(\omega)) \quad (7)$$

As discussed in our previous work on this subject [1], [2], [6], [8]–[10], the practical calculation of GFCCSD matrix employing the above method involves the solution of the conventional CCSD calculations (to get converged T and Λ cluster amplitudes), solving linear systems of the form of Eq. (4) for all the orbitals (p 's) and frequencies of interest (ω 's), and performing Eq. (6). The key step is to solve Eq. (4) for $X_p(\omega)$ for given orbital p and frequency ω , and the overall computational cost approximately scales as $\mathcal{O}(N_\omega N^6)$ with the N_ω being the number of frequencies in the designated frequency regime. Therefore, if a finer or broader frequency range needs to be computed, N_ω would constitute a sizable pre-factor.

In the context of high performance computing, one can divide the full computational task posed by the GFCC method into several smaller tasks according to the number of orbitals and frequencies desired. In so doing, one can distribute these smaller tasks over the available processors to execute them concurrently. In this way, the overall computational cost remains the same, but the time-to-solution can be significantly reduced. In order to reduce the formal computational cost of the GFCC method, we further introduce model-order-reduction (MOR) technique in the context of the GFCC method. [11]

We can first represent Eqs. (4) and (6) as a linear multiple-input multiple-output (MIMO) system Θ ,

$$\Theta(\omega) = \begin{cases} (\omega - i\eta + \bar{\mathbf{H}}_N)\mathbf{X}(\omega) & = \mathbf{b}, \\ \mathbf{G}^R(\omega) & = \mathbf{c}^T\mathbf{X}(\omega). \end{cases} \quad (8)$$

Here, the dimension of $\bar{\mathbf{H}}_N$ is D (For the GFCCSD method, D scales as $\mathcal{O}(N_o^2 N_v)$ with N_o being the total number of occupied spin-orbitals and N_v being the total number of virtual spin-orbitals). The columns of \mathbf{b} corresponding to free terms, and the columns of \mathbf{c} corresponding to $\langle \Phi | (1 + \Lambda) a_q^\dagger$. The transfer function of the linear system Θ ,

$$\gamma(\omega) = \mathbf{c}^T (\omega - i\eta + \bar{\mathbf{H}}_N)^{-1} \mathbf{b} \quad (9)$$

describes the relation between the input and output of Θ , and is equal to its output $\mathbf{G}^R(\omega)$ for (8).

To apply the interpolation based MOR technique for Θ , we need to construct an orthonormal subspace $\mathbf{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ with $m \ll D$ and $\langle \mathbf{v}_i | \mathbf{v}_j \rangle = \delta_{ij}$, such that the original linear system (8) can be projected into to form a model system $\hat{\Theta}$,

$$\hat{\Theta}(\omega) = \begin{cases} (\omega - i\eta + \hat{\mathbf{H}}_N)\hat{\mathbf{X}}(\omega) & = \hat{\mathbf{b}}, \\ \hat{\mathbf{G}}^R(\omega) & = \hat{\mathbf{c}}^T\hat{\mathbf{X}}(\omega), \end{cases} \quad (10)$$

where $\hat{\mathbf{H}}_N = \mathbf{S}^T \bar{\mathbf{H}}_N \mathbf{S}$, $\hat{\mathbf{X}}(\omega) = \mathbf{S}^T \mathbf{X}(\omega)$, $\hat{\mathbf{b}} = \mathbf{S}^T \mathbf{b}$, and $\hat{\mathbf{c}}^T = \mathbf{c}^T \mathbf{S}$. With the proper construction of the subspace \mathbf{S} , we expect $\hat{\Theta}(\omega) \approx \Theta(\omega)$ for designated frequency regime.

In practice, the subspace \mathbf{S} is composed of the orthonormalized auxiliary vectors, \mathbf{X}_p , converged at selected frequencies ω_k in a given frequency regime, $[\omega_{\min}, \omega_{\max}]$. Hence, the transfer function of the reduced model $\hat{\Theta}$ interpolates the original model Θ at these selected frequencies, i.e.,

$$\hat{\mathbf{G}}^R(\omega_k) = \mathbf{G}^R(\omega_k)$$

for $k = 1, \dots, m$. The sampling of the selected frequencies in the regime follows the adaptive refinement strategy described in Ref. [12]. Basically, one can start with a uniformly sampled frequencies in the regime to construct a preliminary level reduced order model. Then, based on the error estimates of the computed spectral function between adjacent frequencies over the entire regime, one can decide whether the corresponding midpoints between these adjacent frequencies need to be added to refine the sampling. This refinement process continues until the maximal error estimate of the computed spectral function at the entire frequency window is below the threshold or when the refined model order exceeds a prescribed upper bound.

III. KEY FEATURES

Version 1.0

- Concurrent computation of frequency-dependent Green's function matrix elements and spectral function in the CCSD/GFCCSD level (enabled via MPI process groups),

- Supporting multidimensional real/complex hybrid tensor contractions and slicing on both CPU and GPU,
- On-the-fly Cholesky decomposition for atomic-orbital based two-electron integrals,
- Direct inversion of the iterative subspace (DIIS) is customized and implemented as the default complex linear solver,
- Gram-Schmidt orthogonalization for multidimensional complex tensors,
- Model-order-reduction (MOR) procedure for complex linear systems,
- Automatic resource throttling for various inexpensive operations,
- Checkpointing (or restarting) calculation employing parallel IO operations for reading (writing) tensors from (to) disk.

Version 1.1

- The default linear solver has been changed from DIIS to Generalized minimal residual method (GMRES),
- Load balancing across computation of different orbitals.

IV. REMARKS

Version 1.0

- In this version, the Cholesky vectors are used in all the post Hartree-Fock calculations supported by the library (to save the memory cost from using the two-electron integrals). As we have shown earlier the accuracy of using Cholesky vectors in the post Hartree-Fock calculation can be well-controlled by a pre-defined Cholesky threshold. [7]
- In this version, in the GFCCSD calculation, the Λ in Eq. (6) is approximated by T^\dagger amplitude. Our recent tests show that this approximation gives very close results to the conventional GFCCSD results (only the positions of satellites slightly shift) but saves almost one third of the entire computational cost of CCSD/GFCCSD. [11]
- In this version, the MOR technique is automatically enabled in the GFCCSD calculation, from which for a given frequency range one can either interpolate or extrapolate the GFCCSD results to obtain good approximations for the same or extended frequency range. [11]
- In this version, the GFCCSD calculations are only performed for the retarded part in occupied molecular orbital space. For molecular systems, this is a good approximation, since energy gap between the occupied and virtual molecular orbital is relatively large, the contribution of the virtual molecular space to the retarded part of the Green's function is negligible.

V. EXAMPLE

In the following, we will use carbon monoxide as an example to demonstrate how to perform the CCSD/GFCCSD

computation employing GFCCLib. Here, we want to compute the spectral function of CO molecule over $[-0.50, -0.30]$ a.u. (i.e. $[-13.60, -8.16]$ eV) in the CCSD/cc-pVDZ level, and extrapolate the GFCCSD results to $[-2.00, 0.00]$ a.u. (i.e. $[-54.42, 0.00]$ eV).

A. Input File

The following shows the input file for the CCSD/GFCCSD calculation.

```
L1  2
L2
L3  C    0.000    0.000    0.000
L4  O    0.000    0.000    1.128
L5
L6  COMMON {
L7      basis cc-pvdz
L8      geometry units angstrom
L9  }
L10
L11 SCF {
L12     tol_int 1e-10
L13     tol_lindep 1e-6
L14     conve 1e-10
L15     convd 1e-9
L16 }
L17
L18 CD {
L19     diagtol 1e-8
L20 }
L21
L22 CCSD {
L23     threshold 1e-7
L24     writet true
L25     gf_ngmres 10
L26     gf_maxiter 500
L27     gf_eta -0.01
L28     gf_threshold 1e-2
L29     gf_omega_min_ip -0.4
L30     gf_omega_max_ip -0.2
L31     gf_omega_min_ip_e -2.0
L32     gf_omega_max_ip_e 0
L33     gf_omega_delta 0.01
L34     gf_omega_delta_e 0.01
L35     gf_nprocs_poi 2
L36 }
```

Line #1: number of atoms in the molecular system.

Line #3-#4: atom coordinates.

Line #6-#9: COMMON section, listing some common parameters such as basis set and the coordinate units

(angstrom or bohr).

Line #11-#16: SCF section, listing parameters used in the self-consistent field calculation.

“tol_int”: tolerance for screening atomic-orbital based two-electron integrals,

“tol_lindep”: tolerance for detecting the linear dependence of basis set,

“conve”: convergence threshold for SCF energy during the iteration,

“convd”: convergence threshold for density matrix during the iteration.

Line #18-#20: CD section, listing the parameters used in the Cholesky decomposition of atomic-orbital based two-electron integral tensor.

“diagonal”: a diagonal threshold used to terminate the decomposition procedure and truncate the Cholesky vectors.

Line #22-#35: CCSD section, listing the parameters used in the Cholesky-based CCSD/GFCCSD calculations (the parameters solely used in the GFCCSD calculation are labelled by ‘gf_’ in their names).

“threshold”: energy threshold used in the ground state CCSD iteration,

“writet”: whether to save the converged T amplitude to disk.

“gf_ngmres”: the micro steps in the GMRES procedure in the Green’s function calculations.

“gf_maxiter”: maximum number of iterations used to solve the GFCCSD linear equation.

“gf_eta”: the broadening factor η .

“gf_threshold”: threshold used to determine whether the GFCCSD linear equation is converged or not.

“gf_omega_min_ip ” and

“gf_omega_max_ip ”: the lower and upper bounds of a user-defined frequency range to perform interpolation.

“gf_omega_min_ip_e ” and

“gf_omega_max_ip_e ”: the lower and upper bounds of a user-defined frequency range to perform extrapolation based on the interpolated model.

“gf_omega_delta ” and

“gf_omega_delta_e ”: frequency intervals used in the above two frequency ranges.

“gf_nprocs_poi ”: number of mpi processes to use to process a single orbital.

B. Output File

The following is the abstract format of the output file, where detailed information has been skipped due to the space.

```

...

Hartree-Fock iterations
...

** Hartree-Fock energy = -112.7493113372152
writing orbitals to file ... done.

...

Begin Cholesky Decomposition ...

#AOs, #electrons = 28 , 7

Number of cholesky vectors = 272

...

CCSD iterations

...

Iterations converged
CCSD correlation energy / hartree =
-0.298038798762180
CCSD total energy / hartree
= -113.047350135977382

...

#occupied , #virtual = 14, 42

-----
GF-CCSD (omega = -0.5)
-----

GF-CCSD (omega = -0.3)
-----

spectral function (omega_npts = 21):
...

omegas processed in level 1 = [-0.5, -0.3,]

-----
GF-CCSD (omega = -0.4)
-----

spectral function (omega_npts = 21):

```

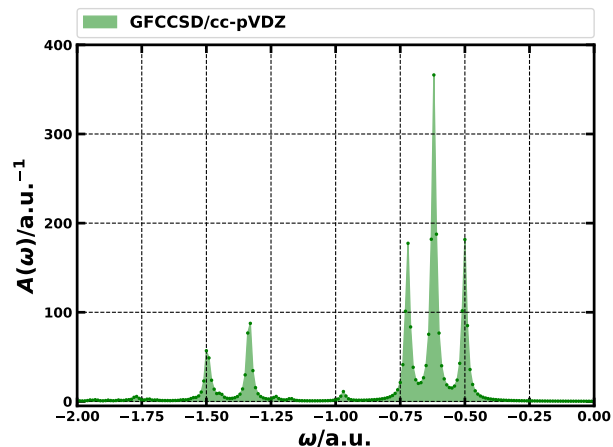


Fig. 1: Spectral functions of the carbon monoxide molecule in the frequency regime of $[-2.00, 0.00]$ a.u. computed by GFCCSD method employing MOR technique. $\eta=0.01$ a.u. and $\Delta\omega \leq 0.002$ a.u.

```

...

omegas processed in level 2 = [-0.4,]

-----extrapolate & converge-----

...

```

As we can see, the output file includes three major parts, namely Hartree-Fock (SCF), CCSD, and GFCCSD. The major information that can be obtained directly includes SCF energy, CCSD ground state correlation energy, GFCCSD spectral function. Since MOR technique is used by default in the GFCCSD calculation, the spectral function computed is printed out at each level. At the end of the output, the intrapolated and extrapolated spectral function over extended frequency range is also printed out.

VI. PERFORMANCE

At present, the GFCCLib has been tested for systems consisting of 100~1200 basis functions. For a given frequency point, the computational cost of performing GFCCSD calculation for entire molecular orbital space is $\mathcal{O}(N^6)$, where N represents the number of basis functions. As a perception, Fig. 2 shows the running time of the GFCCSD/cc-pVDZ calculation of C_{60} molecule ($N = 840$) at $\omega = -10.45$ a.u. on OLCF Summit supercomputer.

VII. ACKNOWLEDGMENTS

The development of the GFCCLib is supported by the Center for Scalable, Predictive methods for Excitation and Correlated phenomena (SPEC), which is funded by the U.S.

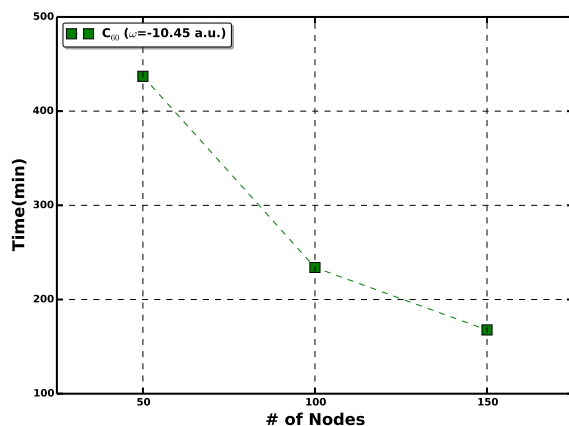


Fig. 2: Running time as a function of number of nodes for carrying out GFCCSD/cc-pVDZ calculation of C_{60} molecule at $\omega = -10.45$ a.u.

Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, the Division of Chemical Sciences, Geosciences, and Biosciences.

REFERENCES

- [1] K. Bhaskaran-Nair, K. Kowalski, and W. A. Shelton. Coupled cluster green function: Model involving single and double excitations. *J. Chem. Phys.*, 144(14):144101, 2016.
- [2] K. Kowalski, K. Bhaskaran-Nair, and W. A. Shelton. Coupled-cluster representation of green function employing modified spectral resolutions of similarity transformed hamiltonians. *J. Chem. Phys.*, 141(9):094102, 2014.
- [3] M. Nooijen and J. G. Snijders. Coupled cluster approach to the single-particle green’s function. *Int. J. Quantum Chem.*, 44(S26):55–83, 1992.
- [4] M. Nooijen and J. G. Snijders. Coupled cluster green’s function method: Working equations and applications. *Int. J. Quantum Chem.*, 48(1):15–48, 1993.
- [5] M. Nooijen and J. G. Snijders. Second order many-body perturbation approximations to the coupled cluster green’s function. *J. Chem. Phys.*, 102(4):1681–1688, 1995.
- [6] B. Peng and K. Kowalski. Coupled-cluster green’s function: Analysis of properties originating in the exponential parametrization of the ground-state wave function. *Phys. Rev. A*, 94:062512, Dec 2016.
- [7] B. Peng and K. Kowalski. Highly efficient and scalable compound decomposition of two-electron integral tensor and its application in coupled cluster calculations. *J. Chem. Theory Comput.*, 13(9):4179–4192, 2017.
- [8] B. Peng and K. Kowalski. Green’s function coupled-cluster approach: Simulating photoelectron spectra for realistic molecular systems. *J. Chem. Theory Comput.*, 14(8):4335–4352, 2018.
- [9] B. Peng and K. Kowalski. Green’s function coupled cluster formulations utilizing extended inner excitations. *J. Chem. Phys.*, 149(21):214102, 2018.
- [10] B. Peng and K. Kowalski. Properties of advanced coupled-cluster green’s function. *Mol. Phys.*, 116(5-6):561–569, 2018.
- [11] B. Peng, R. Van Beeumen, D. B. Williams-Young, K. Kowalski, and C. Yang. Approximate green’s function coupled cluster method employing effective dimension reduction. *Journal of Chemical Theory and Computation*, 15(5):3185–3196, 2019. PMID: 30951302.
- [12] R. Van Beeumen, D. B. Williams-Young, J. M. Kasper, C. Yang, E. G. Ng, and X. Li. Model order reduction algorithm for estimating the absorption spectrum. *J. Chem. Theory and Comput.*, 13(10):4950–4961, 2017.