A Unified Formulation of Tensor HyperContraction Coupled-Cluster Theory

Yao Zhao and Edward Hohenstein

Department of Chemistry and Biochemistry, City College of New York

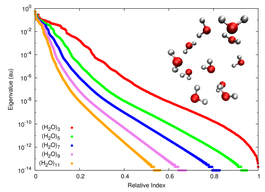
Abstract

One approach to removing the scaling bottleneck from coupled-cluster theory is the use of low-rank factorization of the electron repulsion integrals and wavefunction coefficients within the framework of Tensor HyperContraction (THC). This technique allows most common correlated methods including single and double excitations to be evaluated with effort scaling as O(N^4). We posit that, in the context of coupled-cluster theory, approximate triples corrections (both perturbative and iterative) can be formulated in a manner that will scale O(N^5). Our previous efforts to develop reduced-scaling implementations of coupled-cluster methods have been limited to CC2 (where an exact factorization of doubles amplitudes is possible) and a proof-of-concept implementation of THC-CCSD. A more general approach to the factorization of the doubles amplitudes is CCSD is required if we are to extend these methods to EOM methods and response properties. We are in the process of developing a unified formulation of rank-reduced coupled-cluster singles and doubles methods from the EOM perspective that will allow us to target these molecular properties.

To expedite the development of these coupled-cluster methods, we are working to remove a bottleneck from the code development process: namely, the automatic factorization and code generation of THC contractions. By working with heavily factorized intermediates, a new bottleneck to code development was introduced. Consider the Coulomb term in the MP2 energy expression. When THC is applied, this term is expressed as a contraction of 14 matrices; there are now 13! possible ways to implement this term. In general, this search space is far too large to identify the optimal contraction path by brute force. We have developed a highly efficient graphical approach to identifying the least computationally demanding path and, subsequently, automatically generating a GPU-accelerated implementation of these contractions. This technique is enabling rapid development of new THC coupled-cluster algorithms.

Tensor HyperContraction

The ubiquitous electron repulsion integral tensor emits a low-rank factorization that can be exploited in many contexts in electronic structure theory.



Left: Eigenvalues of the (ia|jb) electron repulsion integral matrix. Eigenvalues are sorted from smallest to largest and plotted against a relative index where 0 is the largest eigenvalue and 1 is the smallest.

The density fitting approximation,

$$(pq|rs) \approx \sum_{AB} (pq|A)[J^{-1}]_{AB}(B|rs)$$

and the use of an incomplete Cholesky decomposition,

$$(pq|rs) \approx \sum_{A} L^A_{pq} L^A_{rs}$$

are two common strategies for building such a low-rank representation. However, these approaches leave the indices in the bra and ket "pinned," which limits the available algebraic flexibility. The Tensor HyperContraction factorization, however, exposes all available flexibility and represents the ideal factorization of such a fourth-order tensor.

$$(pq|rs) \approx \sum_{PQ} X^P_p X^P_q Z^{PQ} X^Q_r X^Q_s$$

Rank-Reduced Coupled-Cluster Theory

The challenge of developing a rank-reduced treatment of the coupled-cluster singles and doubles (CCSD) method is in the handling of the double amplitudes. In CCSD, these amplitudes are solutions to the nonlinear system of equations:

$$\langle \Phi^{ab}_{ij} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_0 \rangle = 0$$

Solving these equations is rate-limiting in CCSD and scales O(N^6). Unfortunately, application of the THC factorized electron repulsion integrals alone is insufficient to reduce the computational complexity. A scaling reduction to O(N4) is possible for THC-CCSD, but it requires that a low-rank factorization of the doubles amplitudes be available and that one solves for the low-rank factorization, directly. In contrast to the electron repulsion integrals, it is less obvious that the doubles amplitudes will emit a low-rank factorization. Here, it is useful to consider a simpler case of the MP2 amplitudes.

$$t^{ab}_{ij} = \frac{(ia|jb)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}$$

If a Laplace transformation of the orbital energy denominator is applied,

$$\frac{1}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} = \int_0^\infty e^{-(\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)t} \mathrm{d}t \approx \sum_\nu \tau^\nu_i \tau^\nu_j \tau^\nu_a \tau^\nu_b$$

in conjunction with the THC representation of the ERIs, then a low-rank factorization of the doubles amplitudes can be easily obtained.

$$\frac{(ia|jb)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \approx -\sum_\nu \tau^\nu_i \tau^\nu_j \tau^\nu_a \tau^\nu_b \sum_{PQ} X^P_i X^P_a Z^{PQ} X^Q_j X^Q_b$$

In the context of the CCSD doubles amplitudes, it is not possible to obtain an analytic factorization of this form. However, empirically, we have found that the opposite-spin doubles amplitudes emit a low-rank factorization similar to that of the MP2 amplitudes.

$$t^{a\bar{b}}_{i\bar{j}} \approx \sum_{PQ} X^P_i X^P_a T^{PQ} X^Q_{\bar{j}} X^Q_{\bar{b}}$$

To exploit this property of the doubles amplitudes, the usual CCSD amplitude equations must be reformulated to solve for the core T^{PQ} matrix. We approach this from the perspective of the coupled-cluster Lagrangian and introduce rank-reduced T and lambda operators.

$$\hat{T}_2 = \frac{1}{4} \sum_{ijab} a^\dagger_a a^\dagger_b a_j a_i \sum_{PQ} X^P_i X^P_a T^{PQ} X^Q_j X^Q_b$$

$$\hat{\Lambda}_2 = \frac{1}{4} \sum_{ijab} a^\dagger_i a^\dagger_j a_b a_a \sum_{PQ} X^P_i X^P_a \Lambda^{PQ} X^Q_j X^Q_b$$

Now, we can take derivatives with respect to the rank-reduced lambda operator in order to obtain equations defining the T operator.

$$\frac{\partial \mathcal{L}(\hat{T}, \hat{\Lambda})}{\partial \lambda^{ij}_{ab}} \frac{\partial \lambda^{ij}_{ab}}{\partial \Lambda^{PQ}} = 0$$

This approach accomplishes several desirable things: it defines a unique set of equations for T^{PQ}, it restores size extensivity and size consistency, and it provides a path to obtain response properties.

Automatic Code Generation

Goal: Automate the derivation, factorization and high performance implementation of Tensor HyperContraction coupled-cluster methods.

Second-quantized expression

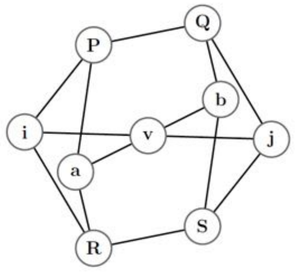
$$\langle \Phi_0 | (\hat{V}\hat{T}_2)_c | \Phi_0 \rangle = \frac{1}{16} \sum_{pqrs} \sum_{ijab} \langle pq||rs\rangle t^{ab}_{ij} \langle \Phi_0 | \{a^\dagger_p a^\dagger_q a_s a_r\} \{a^\dagger_a a^\dagger_b a_j a_i\} | \Phi_0 \rangle$$

Spin-adapted equation

$$E_{\mathrm{MP2}} = \sum_{ijab} t^{ab}_{ij} \left[2(ia|jb) - (ib|ja) \right]$$

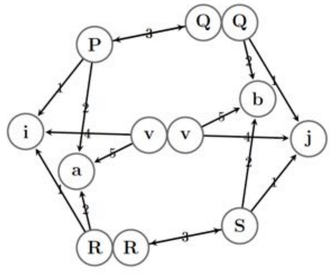
$$E_{\mathrm{THC-MP2}} = \sum_\nu \sum_{ijab} \sum_{PQRS} \tau^\nu_i \tau^\nu_j \tau^\nu_a \tau^\nu_b \cdot X^P_i X^P_a Z^{PQ} X^Q_j X^Q_b \times \left[2 \cdot X^R_i X^R_a Z^{RS} X^S_j X^S_b - X^R_i X^R_b Z^{RS} X^S_j X^S_a \right]$$

Graphical representation



$$E \leftarrow 2 \sum_\nu \sum_{ijab} \sum_{PQRS} \tau^\nu_i \tau^\nu_j \tau^\nu_a \tau^\nu_b \cdot X^P_i X^P_a Z^{PQ} X^Q_j X^Q_b \cdot X^R_i X^R_a Z^{RS} X^S_j X^S_b$$

Automatic optimal factorization



$$E \leftarrow 2 \sum_{\nu QR} \underbrace{\left(\sum_P \underbrace{\left(\sum_i \tau^\nu_i X^P_i X^R_i \right)}_{O^\nu_{PR}} \underbrace{\left(\sum_a \tau^\nu_a X^P_a X^R_a \right)}_{V^\nu_{PR}} Z^{PQ} \right)}_{A^\nu_{RQ}} \underbrace{\left(\sum_S \underbrace{\left(\sum_j \tau^\nu_j X^Q_j X^S_j \right)}_{O^\nu_{QS}} \underbrace{\left(\sum_b \tau^\nu_b X^Q_b X^S_b \right)}_{V^\nu_{QS}} Z^{RS} \right)}_{A^\nu_{RQ}}$$

High performance implementation

Triples Corrections

A full implementation of CCSDT will scale as O(N^8) with a storage requirement of O(N^6) for the triples amplitudes. Most commonly, approximate triples corrections are applied which scale as O(N^7) and do not increase the storage requirement over CCSD. Tensor HyperContraction methods offer the possibility to reduce the computational complexity down to O(N^5). Consider the linearized CCSDT method, CCSDT-1.

$$t^{abc}_{ijk} \approx \frac{\sum_d (ia|bd) t^{cd}_{kj} + \ldots}{\epsilon_i + \epsilon_j + \epsilon_k - \epsilon_a - \epsilon_b - \epsilon_c}$$

The triples amplitudes may be written entirely in terms of singles and doubles amplitudes. If Tensor HyperContraction approximations and a Laplace transformation of the denominator are applied, the triples amplitudes may be factorized.

$$t^{abc}_{ijk} \approx \sum_{\nu d} \tau^\nu_i \tau^\nu_j \tau^\nu_k \tau^\nu_a \tau^\nu_b \tau^\nu_c \sum_{PQ} X^P_i X^P_a Z^{PQ} X^Q_b X^Q_d \sum_{RS} X^R_k X^R_j T^{RS} X^S_c X^S_d + \ldots$$

This allows any triples correction that can be written as a subset of the terms in CCSDT-3 to be evaluated in O(N^5).

$$\langle \Phi^{abc}_{ijk} | e^{-\hat{T}_1 - \hat{T}_2} \hat{H} e^{\hat{T}_1 + \hat{T}_2} | \Phi_0 \rangle = 0$$

This means that triples corrections could be buried under the cost of a conventional CCSD implementation or that they could be evaluated in O(N^5) in conjunction with O(N^4) THC-CCSD.

SciDAC

SciDAC: Designing Photocatalysts Through Scalable Quantum Mechanics and Dynamics

Todd Martinez[1] (PI), Henry Van Den Bedem[2] (co-PI), T.J. Lane[2] (co-PI), Alex Aiken[1] (co-PI), Lexing Ying[1] (co-PI), Kunle Olukotun[1] (co-PI), Possu Huang[1] (co-PI), Ron Dror[1] (co-PI), Edward Hohenstein[3] (co-PI)

1. Stanford University
2. SLAC National Accelerator Laboratory
3. City College of New York

This project is part of a larger SciDAC program that brings together a diverse group of researchers to the computational design of photocatalysts. Our work on reduced-scaling coupled-cluster methods will provide accurate energetics of the catalytic systems that will inform the computational design process. Over the next three years, we will be leveraging contributions from other members of our SciDAC team in applied mathematics as well as computer science into the development and implementation of these coupled-cluster methods.

References

"Tensor Hypercontraction Density Fitting. I. Quartic Scaling Second- and Third-Order Møller-Plesset Perturbation Theory," E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, *J. Chem. Phys.* **137**, 044103 (2012)

"Tensor Hypercontraction. II. Least-Squares Renormalization," R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, *J. Chem. Phys.* **137**, 224106 (2012)

"Communication: Tensor Hypercontraction. III. Least-Squares Tensor Hypercontraction for the Determination of Correlated Wavefunctions," E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, and T. J. Martínez, *J. Chem. Phys.* **137**, 221101 (2012)