

電子相関理論のための数式処理システムに向けて Toward a Computer Algebra System for Electron Correlation Theory

小副川 健

TAKESHI OSOEKAWA

立教大学

RIKKYO UNIVERSITY *

望月 祐志

YUJI MOCHIZUKI

立教大学

RIKKYO UNIVERSITY †

横山 和弘

KAZUHIRO YOKOYAMA

立教大学

RIKKYO UNIVERSITY ‡

Abstract

本研究の目的は量子化学計算のための数式処理システムの開発である。このシステムは特に、電子相関理論に現れる代数表式の操作を目的としている。本稿はシステムの設計について概説し、開発の状況について報告したものである。また、実装のできたモジュールのベンチマークも紹介する。

Abstract

The aim of this research is to develop a computer algebra system for quantum chemistry. The system is designed to manipulate algebraic formulas that appeared in electron correlation theory. We present the architecture of the system and the progress of the development in this paper. This paper also reports some benchmarks of the implementation of a module of the system.

1 はじめに

電子相関理論は量子化学において、高精度の数値計算をするうえで必要な理論である [10, 9]。電子相関理論を用いた計算にはテンソル縮約式と呼ばれる代数表式が頻出する。従来はこの代数表式を、記号計算で人手によって簡約し、数値計算のためのプログラムに翻訳していた。しかし、より高精度の結果を求めるためにはより高次の代数表式を扱うことが要求されるが、代数表式の次数が上がるにつれて式の複雑さは飛躍的に増すため、人手による代数表式の操作ではすぐに限界がくる。そのため、代数表式の操作の自動化は量子化学の発展にとって重要な意味を持つ。

以下の式は代表的な電子相関理論である Møller-Plesset 摂動論における二次の代数表式である。

$$E_{MP2} = \frac{1}{4} \left(\sum_{ijab} \frac{[ia|jb][ai|bj] - [ia|jb][aj|bi] - [ib|ja][ai|bj] + [ib|ja][aj|bi]}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \right). \quad (1)$$

式における i, j, a, b は分子軌道のインデックス、 ε_i など軌道 i の軌道エネルギーを意味する実数で、また $[ij|ab]$ は二電子積分であり

$$[rs|tu] = \int \chi_r^*(\mathbf{x}_1) \chi_s(\mathbf{x}_1) r_{12}^{-1} \chi_t^*(\mathbf{x}_2) \chi_u(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2, \quad (2)$$

*osoken@rikkyo.ac.jp

†fullmoon@rikkyo.ac.jp

‡yokoyama@rkmath.rikkyo.ac.jp

である。これらの物理的な意味については [10] を参照されたい。問題は、それぞれ $B_{rstbu} := [rs|tu]$, $\epsilon_{ijab} := (\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b)^{-1}$ とおいて、代数表式をこれらの多項式と見做した時の式操作の自動化である。これらの変数は定義から

$$B_{rstu} = B_{srtu} = B_{rsut} = B_{srut} = B_{turs} = B_{utrs} = B_{tusr} = B_{utsr}, \quad (3)$$

$$\epsilon_{ijab} = \epsilon_{jiab} = \epsilon_{ijba} = \epsilon_{jiba}, \quad (4)$$

を満たし、また、分子軌道のインデックス i と j , a と b はそれぞれ等価であり、入れ替えられることに注意すると、式 (1) は最終的に

$$E_{MP2} = \frac{1}{2} \sum_{ijab} \frac{[ia|jb]^2 - [ia|jb][ib|ja]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (5)$$

という表現と同じであることがわかる。

実際の量子化学計算ではこの式を使った数値計算へと移るが、ここまでの計算は記号計算である。

このように量子化学には計算機代数に対する潜在的な需要があり、それに対して計算機代数的手法を供給するのが本研究の意義である。計算機代数の発展にとっても応用分野の開拓は重要であり、本研究は学際的な研究体制による応用分野開拓のモデルケースとなるべく、研究を進めている。

本稿に先立って著者らは [7] でこの簡約操作のアルゴリズムを、変数の等値性を表現した多項式集合の Gröbner 基底を用いて実現したが、この方法では代数表式の次数が上がるにつれて、多項式集合のサイズが指数関数的に増大するため、実用的とは言い難い。実用レベルの問題は、Møller-Plesset 摂動論の式でいえば四次以上であるが、多項式集合のサイズは四次の時点で 10^{12} 程になり、既存の汎用の数式処理システムを利用した実装では、到底扱いきれないものとなる。

一方で、我々の問題の特殊性を利用して計算の効率化を行うことで、実用レベルの問題も、実用時間で解けるというベンチマークを得ており、4 節で紹介している。特殊性を最大限利用して効率化を行うには、汎用システム上の実装は、汎用性が弊害となるため困難であり、低レベルのプログラム言語による実装がより現実的である。

また、本研究によって提案された手法は、背景にある理論を必ずしも把握していない、量子化学の研究者によって使える状態で提供して初めて意味を持つ。このためには、ユーザが使えるようになるまでの学習コストは低い方がよく、限定的な機能を持ったシステムという形で提供するのがよい。

そのような理由から、我々は本研究の目標の一つとして、電子相関理論に現れる代数表式処理に特化した独自の数式処理システム開発への着想を得た。

本稿は、電子相関理論に関する記号計算に特化した数式処理システム開発に関する、最初のレポートである。

2 節では量子化学を取り巻くソフトウェア環境を概観する。3 節では本システムに要求される機能を列挙し、設計段階におけるシステムと各モジュールの概要を説明する。4 節では開発状況について報告を行い、代数表式の操作において最も重要な簡約モジュールのベンチマークを紹介する。

2 量子化学と計算システム

本節では量子化学の研究で使われているソフトウェアの一部を紹介し、量子化学計算を取り巻くソフトウェア環境を概観する。

2.1 数値計算用ソフトウェア

量子化学計算の主な目的は分子の構造のデータを入力として、電子のエネルギーなどを表す数値を計算することである。数値計算用のソフトウェアは GAUSSIAN[2], GAMESS (General Atomic and Molecular Electronic Structure System) [3], ABINIT-MP[6] をはじめとした数々のソフトウェアが開発・改良されており、充実しているといえる。

2.2 分子の可視化や構造作成のためのソフトウェア

数値計算を行うには分子の構造のデータが必要であるが、昨今の研究対象になるような分子の構造は非常に複雑で、適切なデータを作成するのは困難である。そのため、計算に使うデータを用意するための支援が充実している。

ソフトウェアではないが、タンパク質の構造はインターネット上のデータバンクがあり、そのデータが研究には広く用いられている。有名なものに Protein Data Bank[8] などがある。

構造を作る、あるいはある分子の一部を別のものに置き換える作業はモデリングと呼ばれているが、そのためのソフトウェアに GAUSS VIEW や MOE (Molecular Operating Environment) などがある。これらのソフトウェアは分子の構造の可視化の役割も持っている。

2.3 代数表式のためのソフトウェア

代数表式の操作により、式の変形や導出を行うソフトウェアも開発されてきている。そのようなソフトウェアとして最も有名なものは TCE (Tensor Contraction Engine) [1, 4, 5] である。

TCE は定式化からくる制約によって、開殻系の分子をきちんと扱うことができないなどの問題がある。そのような問題が起こらないシステム構築が、本研究の目的の一つである。

3 システムの概要

本節では、システムに必要な機能を挙げ、現時点におけるシステム設計の概要について述べる。

既存の多くの数式処理システムがそうであるように、本システムもユーザインターフェースと計算エンジンは分離する。これは、入力規則に対するユーザの細かい要望を反映しやすくする狙いもある。

3.1 代数表式入力 → 簡約 → 評価用プログラムソースコード出力

この処理の流れが本システムにおける想定された用途である。数値計算に使いたい代数表式を、予め決められた構文規則にしたがってシステムに入力するとそれを解釈し、簡約した式を出力するのが簡約モジュールの役割である。MP2 の例でいけば、

$$\frac{1}{4} \left(\sum_{ijab} \frac{[ia|jb][ai|bj] - [ia|jb][aj|bi] - [ib|ja][ai|bj] + [ib|ja][aj|bi]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \right).$$

の入力に対し、

$$\frac{1}{2} \sum_{ijab} \frac{[ia|jb]^2 - [ia|jb][ib|ja]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}$$

を出力する。数値計算のためにはさらに**スピン積分**という処理を行い、最終的に、数値計算に使う式として

$$\sum_{ijab} \frac{2(ia|jb)^2 - (ia|jb)(ib|ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}$$

を出力する。スピン積分も代数表式の記号操作の一種と見做せるので、簡約モジュールで行うこととする。

数値計算は、

$$\begin{aligned} (11|11) &= 4.744 \\ (21|11) &= -0.4177 \\ &\vdots \\ (71|11) &= -0.2379 \\ (12|11) &= -0.4177 \\ &\vdots \\ (77|77) &= 0.5932 \\ \epsilon_1 &= -20.23 \\ &\vdots \\ \epsilon_7 &= 0.5899 \end{aligned}$$

といったテーブルが与えられ、分子軌道のインデックスそれぞれに $1, \dots$ を当てはめたときの値を評価して $\sum$ の展開を行う。この例の場合、 $i, j = 1, 2, \dots, 5$ と $a, b = 6, 7$ で展開を行うことになっており、結果 -0.03586 を得る。

この計算自体は単純な繰返し計算とテーブルとの照合なので、代数表式の形が決まれば、数値評価用の処理を行うプログラムソースコードの出力ができる。その役割を担うのが**出力モジュール**である。

FORTRAN などの、量子化学の数値計算用ソフトウェアで利用できる形のプログラムソースコード出力ができれば、既存の数値計算ソフトウェアと連携が取れる。

3.2 その他の機能とモジュールの相関図

高次の代数表式は項数が多く非常に複雑で、システムに正しく入力するのが難しい。一方で、代数表式は帰納的に導出することができる。そこで、特定のユーザの入力によって、帰納的計算によって代数表式を導出する**導出モジュール**の開発も目指したい。

また補助的に、内部形式の代数表式を使って数値計算ができるよう、**数値計算モジュール**も開発する。これは主に式の確認が目的であるが、作りこむことで、システムを代数表式から数値計算まで行える統合環境として昇華できる。

システムに必要なモジュールの相関は、図 1 のようになっている。点線で囲まれている部分がシステムの本体である。

4 開発状況とベンチマーク

前節で述べたモジュール群のうち、構文解析器、数値計算モジュール、出力モジュールと内部形式の実装には理論的困難はない。残りの簡約モジュールと導出モジュールについてはアルゴリズムから構築する必要がある。

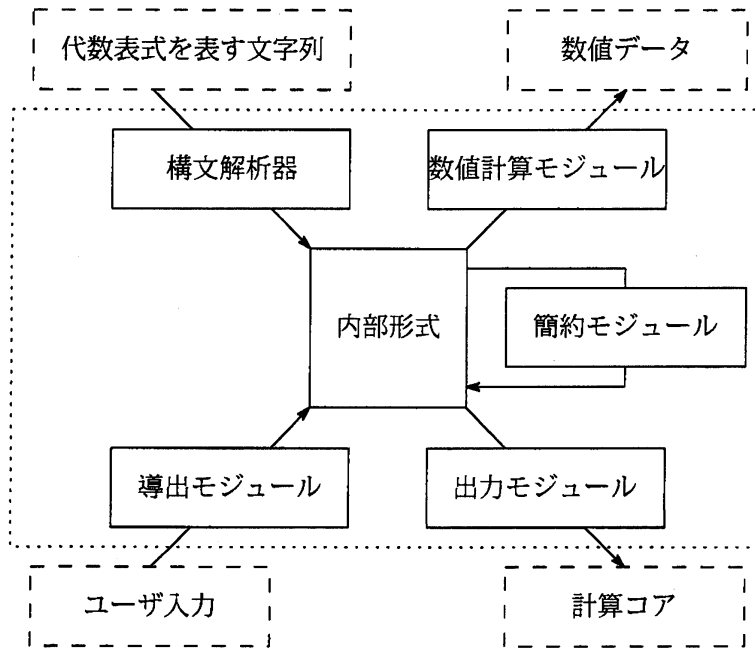


図 1: システムの模式図

ある. 著者らは既に [7] で簡約モジュールの基本となるアルゴリズムを提案し, その改良と実装を行っている.

簡約モジュールのベンチマークを紹介する. 実用レベルの問題である MP4 の代数式に現れる単項に対して, [7] のものを基本として, 効率化したアルゴリズムを適用し, (ある項順序における) その標準形を計算した. 表中の記号はそれぞれ $B_{rstu} = [rs|tu]$, $\epsilon_{ijab} = (\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b)^{-1}$ である.

実行環境は Intel (R) Core (TM) 2 Duo 2.20 GHz, 1GB RAM, Windows XP (32bit) で, 実装は C++ で行った. その結果が表 1 である.

入力単項式	標準形	時間 (ms)
$\epsilon_{kb}\epsilon_{ijab}\epsilon_{lmbe}B_{ikja}B_{ibja}B_{klmc}B_{lbmc}$	$\epsilon_{ia}\epsilon_{jkab}\epsilon_{lmac}B_{ijkb}B_{ilmc}B_{jakb}B_{lamc}$	172
$\epsilon_{kb}\epsilon_{ijab}\epsilon_{klcd}B_{iajk}B_{iajb}B_{kcld}B_{lcbd}$	$\epsilon_{ia}\epsilon_{ijbc}\epsilon_{klad}B_{ikld}B_{ibjc}B_{jbac}B_{kald}$	359
$\epsilon_{jc}\epsilon_{ijab}\epsilon_{jkde}B_{ibja}B_{ibac}B_{jdke}B_{kecd}$	$\epsilon_{ia}\epsilon_{ijbc}\epsilon_{ikde}B_{ibjc}B_{idke}B_{jcab}B_{kead}$	172
$\epsilon_{ijab}\epsilon_{ikac}\epsilon_{lmac}B_{imkl}B_{ibja}B_{jbkc}B_{lcma}$	$\epsilon_{ijab}\epsilon_{ikac}\epsilon_{lmab}B_{iljm}B_{icka}B_{jbkc}B_{lamb}$	406
$\epsilon_{ijab}\epsilon_{ikac}\epsilon_{ilad}B_{ibja}B_{idla}B_{jbkc}B_{kcld}$	$\epsilon_{ijab}\epsilon_{ikac}\epsilon_{ilad}B_{ibja}B_{icka}B_{jblb}B_{kcld}$	110
$\epsilon_{ijab}\epsilon_{ikac}\epsilon_{ikde}B_{ibja}B_{iekd}B_{jbkc}B_{aecd}$	$\epsilon_{ijab}\epsilon_{ijcd}\epsilon_{ikae}B_{icjd}B_{ieka}B_{jbke}B_{acbd}$	422
$\epsilon_{ijab}\epsilon_{ikbd}\epsilon_{jlac}B_{ibja}B_{idkb}B_{jcla}B_{kcld}$	$\epsilon_{ijab}\epsilon_{ikac}\epsilon_{jlbd}B_{iajb}B_{icka}B_{jdlb}B_{kdlc}$	110

表 1: 簡約モジュールのベンチマーク

この結果は単項式の標準形の計算であり, 本来の目標は多項式の標準形であるが, 多項式に対してかかる時間も, 高々これの定数倍であることから, この実装は十分実用的であると言える.

5 まとめと今後の課題

本稿では量子化学計算, 特に電子相関理論のための数式処理システムに関して, その開発動機と設計の概要, そして現状の報告とベンチマークの紹介を行った.

重複を恐れず強調すると, 本研究において重要なのは, 量子化学計算の実用レベルの問題を実用時間で解ける手段を提供するということである. そのため本研究は, 理論面, 実装面双方において研究を進めることが必要であり, 本稿は実装面における第一歩を報告したものである.

4 節で述べたとおり, 簡約モジュールは十分実用的な実装ができているが, システムの他の面はまだまだ開発が必要であり, 今後の課題である. 実用的なシステムを作るためにはユーザビリティの検証も必要であり, そのためには最低限入出力を整えなくてはならない. その実装を進めつつ, 導出モジュールに関するアルゴリズムを構築することが当面の課題である.

謝辞

本研究は立教大学 SFR の支援のもと行われている.

参 考 文 献

- [1] A. Auer, G. Baumgartner, and D. E. Bernholdt. Automated code generation for many-body electronic structure methods: the tensor contraction engine. *Molecular Physics*, Vol. 104, pp. 211–228, 2006.
- [2] J. B. Foresman and Æ. Frisch. *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc., second edition, 1996.
- [3] Gordon Group/GAMESS Homepage. <http://www.msg.ameslab.gov/gamess/>.
- [4] S. Hirata. Tensor contraction engine: abstraction and automated parallel implementation of configuration-interaction, coupled-cluster, and many-body perturbation theories. *Journal of Physical Chemistry A*, Vol. 107, pp. 9887–9897, 2003.
- [5] S. Hirata. Symbolic algebra in quantum chemistry. *Theoretical Chemistry Accounts*, Vol. 116, pp. 2–17, 2007.
- [6] T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayasi, and K. Kitaura. Fragment molecular orbital method: use of approximate electrostatic potential. *Chemical Physics Letters*, Vol. 351, pp. 475–480, 2002.
- [7] T. Osoekawa, N. Shinohara, Y. Mochizuki, and K. Yokoyama. Gröbner basis technique for algebraic formulas in electron correlation theories. *Proceeding of the 10th International Conference on Computational Science and Its Applications (ICCSA-2010)*, pp. 17–23, 2010.
- [8] RCSB Protein Data Bank. <http://www.pdb.orb/>.
- [9] I. Shavitt and R. J. Bartlett. *Many-Body Methods in Chemistry and Physics*. Cambridge University Press, London, 2009.
- [10] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry*. MacMillan, New York, 1982.