

CD 法に基づくタンパク質全電子カノニカル分子軌道計算プログラムの開発

(東大生研)平野 敏行, 佐藤 文俊

Development of canonical molecular orbital calculation program for proteins using Cholesky decomposition

(Institute of Industrial Science, the University of Tokyo)

Toshiyuki HIRANO, Fumitoshi SATO

【概要】

分散メモリ型並列計算機において、大規模タンパク質の全電子カノニカル分子軌道計算を行うためには、計算に必要な大規模行列を各計算ノードに分散保持する必要がある[1]。Kohn-Sham 行列におけるクーロン項ならびに Fock の交換項の行列要素を求める際に、計算律速ルーチンの一つである 4 中心 2 電子積分が必要であるとともに、疎とはみなせない密度行列要素を参照しなければならない。したがって、大規模分子の全電子カノニカル分子軌道計算では、高速かつ高度に並列化された 4 中心 2 電子積分計算に加え、分散保持した大規模な密度行列に対する並列性能の良いアクセスが求められる。

本研究では、次世代スーパーコンピュータ「京」をはじめとする、分散メモリ型並列計算機に適したカノニカル分子軌道計算プログラムの開発を目的として、コレスキー分解[2]を用いた並列計算法を開発・実装した。巨大メモリを必要とするコレスキー分解法であるが、その行列を縮約し、また大規模行列を分散保持することにより、その欠点を克服することができた。

【コレスキー分解による大規模電子状態計算】

原子軌道のペア $N_{AO-pairs}$ の次元を持つ、4 中心 2 電子積分を行列要素とする行列 \mathbf{V} は正値対称行列であり、コレスキー分解によりランク N_{CD} の行列 \mathbf{L} に分解できる。

$$V_{pq,rs} = (pq|rs) \approx \sum_I^{N_{CD}} L_{pq}^I L_{rs}^I$$

ここで p, q, r, s は原子軌道のインデックスをあらわす。このときコレスキー分解による計算誤差は以下のように求められ、任意の閾値 δ に対して計算精度を設定することができる。

$$\left| (pq|rs) - \sum_I^{N_{CD}} L_{pq}^I L_{rs}^I \right| \leq \delta$$

ここで L のサイズは $N_{AO-pairs} \times N_{CD}$ となり、巨大な行列になる。そこで本研究では、CDAM 法[3]とピボット CD 法[4]を組み合わせることにより、行列サイズを縮小した。すなわち CDAM 法では、以下の条件式

$$V_{pq,pq} = (pq|pq) \geq \tau$$

で生き残った原子軌道ペア (pq) から構成される行列 \mathbf{V} に対してコレスキー分解を行う。これは Schwartz の不等式でのスクリーニング[5]に相当する。コレスキー分解は、ピボット CD 法によ

り、なるべくランク N_{CD} が少なくなるようにコレスキーベクトルを求める。計算コストの高い 2 電子 4 中心積分は、コレスキー分解処理に必要なタイミング (on-the-fly) で求めることができるため、余計な積分計算が発生しない。なお、計算途中で必要なコレスキーベクトルは、行ベクトルとして各プロセスに分散して保持している。これにより、計算可能なサイズが、各計算ノードではなく、計算機全体に搭載されたメモリ容量に依存することになるとともに、ロードバランスの均等化を図っている。

コレスキー分解法において、クーロン項 \mathbf{J} ならびに Fock の交換項 \mathbf{K} は以下のように求められる [3]。

$$J_{pq} = \sum_{rs} (pq, rs) P_{rs} \approx \sum_{rs} \sum_I L_{I,pq} L_{I,rs} P_{rs}$$

$$K_{pq} = \frac{1}{2} \sum_{rs} (pr, qs) P_{rs} \approx \sum_i \sum_I X_{I,pi} X_{I,qi}$$

$$X_{I,pi} = \sum_r L_{I,pr} Q_{ri}$$

$$P_{pq} = \sum_J Q_{p,J} Q_{q,J}$$

SCF 繰り返し計算におけるクーロン項 \mathbf{J} ならびに Fock の交換項 \mathbf{K} の計算において、シンプルな行列演算のみで求めることができ、並列計算機に適した、ScaLAPACK[6]などの線形演算ライブラリを利用することができる。均等なロードバランスを取ることが難しい、大域密度行列のアクセスと分子積分計算が必要ないことも長所の一つである。

本手法によるコレスキー分解法では、計算精度を左右する閾値は、 δ と τ の 2 つ存在する。適切な閾値を決定するために、いくつかの分子において、これまでの update 法を用いた SCF 計算と本手法を比較した。その結果、分子積分のカットオフ値を 1.0×10^{-10} に設定した計算結果と同程度の計算精度を保つためには、 $\delta = 1.0 \times 10^{-10}$ 、 $\tau = 1.0 \times 10^{-4}$ が妥当であることがわかった。

当日は、上記コレスキー分解を用いた電子状態計算法の各アルゴリズム、計算精度、必要な計算機資源について紹介するとともに、具体的なタンパク質をモデルとした本手法の計算例を示す。

【参考文献】

- [1] 平野, 千葉, 佐藤, 日本シミュレーション学会論文誌, **1**, 50, (2009).
- [2] F. Aquilante, L. Boman, J. Boström, H. Koch, R. Lindh, M. A. Sánchez, T. B. Pederson, *LinearScaling Techniques in Computational Chemistry and Physics*, 301, 2011.
- [3] Y. Okiyama, T. Nakano, K. Yamashita, Y. Mochizuki, N. Taguchi, S. Tanaka, *Chem. Phys. Lett.*, **490**, 84 (2010).
- [4] H. Harbrecht, M. Peters, R. Schneider, *Appl. Num. Math.*, **62**, 428 (2012).
- [5] M. Häser, R. Ahlrichs, *J. Comput. Chem.*, **10**, 104, (1989)
- [6] <http://www.netlib.org/scalapack/>