数理モデル化と問題解決 33－12
(2001. 3. 16)

分子ポテンシャル最小化問題に関する並列局所探索法の比較評価

田中秀俊*, Daniel Eriksson+, 白石將*, 佐藤裕幸*, 中島克人*
*新情報処理開発機構　並列応用三菱研究室（三菱電機　情報総研）
+新情報処理開発機構　並列応用三菱研究室（スウェーデン王立工科大学）

Lennard Jones 関数と静電ポテンシャル関数を用いた最小化問題について、Very Fast Simulated Annealing(VFA)、Simulated Quenching(SQ)、Orthogonal Design Local Search(ODLS)、Multi-Directional Search(MDS)、Genetic Algorithm(GA)を比較した。SQ により最も良い結果が得られ、エネルギー関数の評価回数が少ないケースでは VFA が良く、多くなると MDS の方が良いという結果が得られた。GA は評価回数が増しても解の質の向上は鈍かった。

Comparative Study of Parallel Local Search Methods on Lennard Jones Energy and Electrostatic Potential Minimization

Hidetoshi TANAKA, Daniel ERIKSSON, Masashi SHIRAISHI, Hiroyuki SATO, Katsuto NAKAJIMA
Parallel Application Mitsubishi Lab. Real World Computing Partnership
Information Technology R&D Center, Mitsubishi Electric Corp.

Selected global optimization algorithms are tested on a designed minimization problem. The target function consists of Lennard Jones potential function and electrostatic potential for 40 chained atoms. The two Simulated Quenching (SQ) algorithms performs best on the test levels 10000, 25000, 50000 and 100000 maximum function evaluations. The Very Fast Annealing (VFA) algorithm comes in third on the first three test but is outperforms at 100000 function evaluations by the Multi-Directional Search Method (MDS). Local Search with Orthogonal Design (ODLS) and MDS perform very poorly with few function evaluations. Genetic Algorithm (GA) performs poorly when more function evaluations are allowed.

1. はじめに

新情報処理開発機構（RWC）並列応用三菱研究室では、アミノ酸配列からタンパク質の立体構造を予測することを目標として研究を進めている。タンパク質立体構造予測方法の 1 つとして、適当な初期構造から出発してエネルギー最小の原子配置を反復改善により求める方法がある。本稿では炭素原子連鎖のエネルギー最小化問題を対象に、反復改善系の最適化アルゴリズムを適用し、その性能を評価した。エネルギー最小化問題の評価関数は複数の局所解を含むので、局所解に捕らわれずに効率的に真の最適解に近づく方法を工夫しなければならない。反復改善系では、一般に以下の 4 段階の(2)～(4)の反復により実施される。

(1)現在値として、適当な初期値を与える。
(2)現在値に対して、予め定められた操作を加え、得られる値について評価関数を評価する。
(3)現在値を更新する。
(4)予め定められた終了条件が満たされれば終了、そうでない場合は手順(2)へ。

反復改善系最適化の具体的方法として、現在まで様々な方法が提案されている。本稿ではそのうち Orthogonal Design Local Search (ODLS)【1】, Multi-Directional Search (MDS)【2】, Very Fast Annealing (VFA)【3】, Simulated Quenching (SQ)【4,5,6】, Genetic Algorithm (GA)【7】を比較評価の対象とした。

2. 局所探索法

本稿で用いた最適化アルゴリズムについて簡単に説明する。

2.1. 局所探索(Local Search)

局所探索は、探索空間における現在値の近傍点を定義し、近傍点の中から評価関数値が小さくなるものを選んで現在値を更新していく。局所探索では、評価関数値の減少幅が最も大きくなる方向に進んでいくことによって、効率よく探索が進むと考えられる。それには、探索空間を張る変数について 1 つ 1 つ独立に適当な幅だけ増加および減少をさせて、その評価関数値への影響を見るという方法が考えられる。ODLS ではこのとき、実験計画法の考え方を利用することにより、少ない評価回数で確度の高い探索方向を決定する【1】。ODLS 自体に局所解対策はないため、初期値依存性が強い。対策としては、多スタート化や SA、後述の MDS との融合【2】が考えられる。

2.2. パターン探索(Pattern Search)

パターン探索は、局所探索の類似手法である。探索空間において現在値の近傍を探索する際、その変更方向および変更幅が動的に決定される。本稿では MDS を評価対象とした【3】。MDS では、探索空間における単体(simplex)の各頂点での評価関数値に基づいて単体を反転・拡大・縮小させることにより、探索を進める。

2.3. SA(Simulated Annealing)/SQ(Simulated Quenching)

SA、SQ もまた局所探索の類似手法である。探索空間において現在値の近傍を探索する際、評価関数値が高くなる方向への現在値更新を許す点に特徴があり、これによって局所解からの脱出が可能となる。評価関数値が高くなる方向への現在値更新の頻度を制御するために、SA・SQ では「温度」と呼ばれるパラメータを用いる。温度が高い場合には、現在値近傍の範囲が大きくなり、評価関数値が高くなる現在値への更新確率が増える【4】。

SA・SQ は melting と cooling の 2 段階の処理からなる。melting 過程では、低温から出発し近傍全点が選択可能になるまで温度を高めていく。melting により十分高温になると次に cooling 過程では所与の冷却スケジュールによって温度を徐々に下げていく。現在値近傍にはガウシアン分布を用いる方法やコーシー分布を用いる方法などが提案されており、それぞれについて最適解に到達する確率が 1 となるような冷却スケジュールが存在する。本稿ではこれを SA と呼び、その中で特に効率の良いとされる VFA(Very Fast Annealing)を評価対象とした【5】。また、SA よりも急速な冷却は、確率 1 での最適解到達が保証されないが、実行時間の都合上用いられることが多く、経験的には良い解が得られる。これを区別して SQ と呼ぶ【5,6】。本稿では、近傍に 1 次元コーシー分布を用いる SQ のうち 2 種類を評価対象とし、多次元コーシー分布を冷却スケジュールに用いたものを SQ1【7】、VFA と同じ冷却スケジュールを用いたものを SQ2 とした。

2.4. 遺伝的アルゴリズム(GA：Genetic Algorithm)

現在値の集団を世代(generation)と呼び、これを更新していく手法である。各現在値は一定長のコードとして表現される。世代更新の際には、世代から 2 つの現在値を親として選択し、親同士を交叉(crossover)させて新たな現在値を子として生成、さらにその一部を突然変異(mutation)させる。親の選択の際には、評価関数値の低い現在値の選択確率が高くなるようにする。これらの操作を施すことにより、世代が進むに従って評価関数値が十分低い現在値が得られるようになると期待される。本稿では、おおよそ Galib【8】のデフォルト設定の周辺で最適な運用パラメータを探索して設定した上で、突然変異の処理のみが異なる 3 種類の GA を評価対象とした。固定長更新幅の突然変異を GA_FIXED、一様乱数による更新幅の突然変異を GA_UNIFORM、1 次元コーシー分布による乱数の更新幅の突然変異を GA_CAUCHY とした。

3. 最小化問題

最適化アルゴリズムの適用対象とした、原子鎖の折り畳み問題について説明する。原子 i と原子 i+1 が共有結合によって距離 1 で結合しているものとし、結合角と二面角が自由になるものとする。原子鎖のエネルギーを算出するモデルとしては様々なものがあり、それらは正確さと計算効率との妥協により決められている【9,10】。ここでは静電相互作用に基づくエネルギーとファンデルワールス相互作用に基づくエネルギーよって原子鎖のエネルギーが構成されるものとし、ファンデルワールス相互作用に基づくエネルギーは Lennard Jones ポテンシャルでモデル化したものを使用した。評価関数を簡単に式(1)に示す。Lij が Lennard Jones ポテンシャル、Eij が静電ポテンシャル、r_{ij}が原子間距離である。Aij, Bij, Cij は原子 i と原子 j に関して計算した係数である。

$$F = \sum_{i,j \in \Omega(i,j)} F_{ij} \qquad F_{ij} = L_{ij} + E_{ij} \qquad L_{ij} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \qquad E_{ij} = \frac{C_{ij}}{r_{ij}} \tag{1}$$

$$\Omega(i,j) = \{i = 0 \ldots n-1; i+1 < j < n\}$$

原子の数 n は 40 個とした。各原子の自由度は 2 なので変数の数は 80 個である。評価関数は約 800 個のローカルミニマムを含む。

4. 実験結果

上記手順(4)の終了条件には、エネルギー関数評価の回数を 10000 回、25000 回、50000 回、100000 回の 4 種類を設定した。各最適化アルゴリズム・各終了条件に関し、それぞれ初期値を乱数で変更して 100 回の試行を行い、結果として得られたエネルギー最小値の平均および最小値を比較した。結果を表 1 および表 2 に示す。

表 1 比較結果（100 試行平均）

評価回数	10000	25000	50000	100000
ODLS	190.330	−251.005	−398.515	−708.044
MDS	71.095	−347.891	−498.232	−704.670
GA_FIXED	−258.213	−385.035	−481.326	−579.278
GA_UNIFORM	−192.494	−200.455	−294.122	−377.649
GA_CAUCHY	−208.015	−159.602	−486.624	−561.776
SQ1	−481.319	−670.677	−832.162	−970.054
SQ2	−439.559	−670.463	−790.848	−903.995
VFA	−300.539	−430.241	−504.248	−594.664

表 2 比較結果（100 試行最小）

評価回数	10000	25000	50000	100000
ODLS	−46.767	−564.495	−773.137	−969.417
MDS	−186.206	−646.438	−810.039	−989.782
GA_FIXED	−481.082	−661.170	−800.671	−941.817
GA_UNIFORM	−530.644	−620.232	−721.034	−778.859
GA_CAUCHY	−541.003	−645.671	−838.964	−831.206
SQ1	−754.007	−978.759	−1092.800	−1177.280
SQ2	−764.784	−1019.880	−1039.840	−1178.400
VFA	−623.960	−689.285	−743.911	−911.115

5. 考察とまとめ

全最適化アルゴリズムで、評価回数を増すと得られる結果の質も改善されるという傾向は共通する。どのアルゴリズムにおいても、100 回の試行による結果の質には大きなばらつきが見られる。アルゴリズム間の比較評価では、2 種の SQ で最も良い結果が得られた。SQ に次ぐものは、評価回数が 1 万回、2 万 5 千回、5 万回では VFA、10 万回では MDS および ODLS だった。

現状の ODLS は更新幅が一定であるため、今回の結果は妥当である。MDS は、更新幅を動的に変更する方式だが、最適値近辺の挙動は ODLS と類似しているものと思われる。SA(VFA)の最適値近辺の挙動が悪い点は、冷却スケジュールの設定の通りと言える。GA は他アルゴリズムに比べ、評価回数増加に対する解の質の向上が鈍かった。

最適化アルゴリズム適用の際にエネルギー関数の性質は特に利用しなかったので、今回の評価結果は一般の非線形最適化問題に対しても当てはまると考えられる。実際のタンパク質立体構造予測では、変数の数がさらに多くなるため、変数の数を実際に合わせた場合の比較を今後実施する予定である。

【参考文献】

【1】 Tanaka, H. Local Search Using Orthogonal Design of Experiment. PDPTA '00 (2000)

【2】 田中,Eriksson,白石,佐藤,中島. 局所探索法の比較評価. 情処学会全国大会(2001).

【3】 Torczon, V. J. Multi-Directional Search: A Direct Search Algorithm for Parallel Machines. Ph.D Thesis, Rice University, Houston, Texas (1989).

【4】 Frost, R. and Hieneman, P. Simulated Annealing: A Heuristic for Parallel Stochastic Optimization. PDPTA '97 (1997)

【5】 Ingber, L. Very-Fast Simulated Annealing. (1989)

【6】 Ingber, L. Simulated Annealing; Practice versus theory. Lester Ingber Research (1993).

【7】 SA C++ package. http://www.taygeta.com/

【8】 Wall, M. Galib: A C++ Library of Genetic Algorithm Components, Version 2.4, Document Revision B. http://lancet.mit.edu/ga/ (1996).

【9】 Comparison of Force Fields: http://chemcca51.ucsd.edu/~chem215/lectures/special/node1.html

【10】 Potential Energy Function: http://wwwch.embnet.org/MD_tutorial/pages/MD.Part2.html