

遺伝的交叉を用いた並列シミュレーテッドアニーリングの検討

小掠 真貴[†], 廣安 知之^{††}, 三木 光範^{††}, 角 美智子[‡], 岡本 祐幸^{‡‡}

[†] 同志社大学大学院 ^{††} 同志社大学工学部 [‡] 同志社大学工学部(学)

^{‡‡} 岡崎国立共同研究機構 分子科学研究所

本研究では, 遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) を提案する. PSA/GAc は, 複数のプロセス上で並列にシミュレーテッドアニーリングの操作を行い, 一定期間のアニーリングを行った後, 2 プロセス間で遺伝的アルゴリズムのオペレータである交叉により情報交換を行う. これにより, 大局的にはいくつかの準最適解が存在し, 局所的には無数の準最適解を有するような問題に特に有効であると考えられる. 数種のテスト関数に適用した結果, PSA/GAc の優れた性能が明らかとなった. また, 実問題への適用例として最適化問題の一つであるタンパク質の構造解析を行い, 従来用いられていた手法と比較し, PSA/GAc の有効性を検討した.

Examination of Parallel Simulated Annealing using Genetic Crossover

Maki OGURA[†], Tomoyuki HIROYASU^{††}, Mitsunori MIKI^{††}, Michiko SUMI[‡], and Yuko OKAMOTO^{‡‡}

[†] Graduate School of Engineering, Doshisha University

^{††} Knowledge Engineering Dept., Doshisha University

[‡] Undergraduate student, Doshisha University ^{‡‡} Institute for Molecular Science

This paper proposes Parallel Simulated Annealing using Genetic Crossover (PSA/GAc). In this algorithm, there are several processes of simulated annealing (SA) working parallel. To exchange information between the solutions, the operation of genetic crossover is performed. Through the continuous test problems, it is found that PSA/GAc can search the solution effectively. The proposed algorithm is also applied to the protein folding problem. Comparing PSA/GAc to the conventional algorithm, it is also found that PSA/GAc is effective algorithm for real world problems.

1 はじめに

シミュレーテッドアニーリング (Simulated Annealing : SA) は最適化問題の手法の一つであり, 多くの組合せ最適化問題に対して有効な手法である¹⁾. SA の探索は最適解へ収束するという保証を持つが, 解を得るまでの計算量が非常に多いという短所を持つ. 特に連続最適化問題を対象とした場合, 最適解を得るまでに多くの計算量が必要となり実用的ではない. そのため逐次処理である SA を並列化し高速化を図る研究がされている²⁾.

本研究では, 連続最適化問題を対象とする場合においても, 比較的少ない計算量で良好な解を得ることのできる新たな並列 SA アルゴリズムを提案する. 提案する手法は, 並列に複数実行している SA の解の伝達時に, 遺伝的アルゴリズム (Genetic Algorithm : GA) のオペレータである遺伝的交叉を用いたものである. 局所的な探索が得意な SA に, 大局的な探索が得意でかつ部分解の組み合わせで最適解が得られる GA オペレータを取り入れ

ることで, 適用可能な対象問題の範囲を拡張することが可能だと考えられる.

本研究ではそのような手法である遺伝的交叉を用いた並列 SA (Parallel Simulated Annealing using Genetic Crossover : PSA/GAc) を提案し, テスト関数に適用させることにより, その性能を評価する. また, 最適化問題の一つであるタンパク質の構造解析を行い, その有効性を検討する.

2 遺伝的交叉を用いた

並列シミュレーテッドアニーリング

本研究で提案するアルゴリズムは図 1 に示すように, 複数のシミュレーテッドアニーリング (SA) を並列に実行する. さらに, 一定期間ごとの SA の解の伝達時に, 遺伝的アルゴリズム (GA) のオペレータである遺伝的交叉を用いたものである. GA のオペレータを用いた SA であるため, SA の探索点の総数 (SA の並列数) を個体数, アニーリングステップ (計算繰り返し回数) を世代数と呼ぶこと

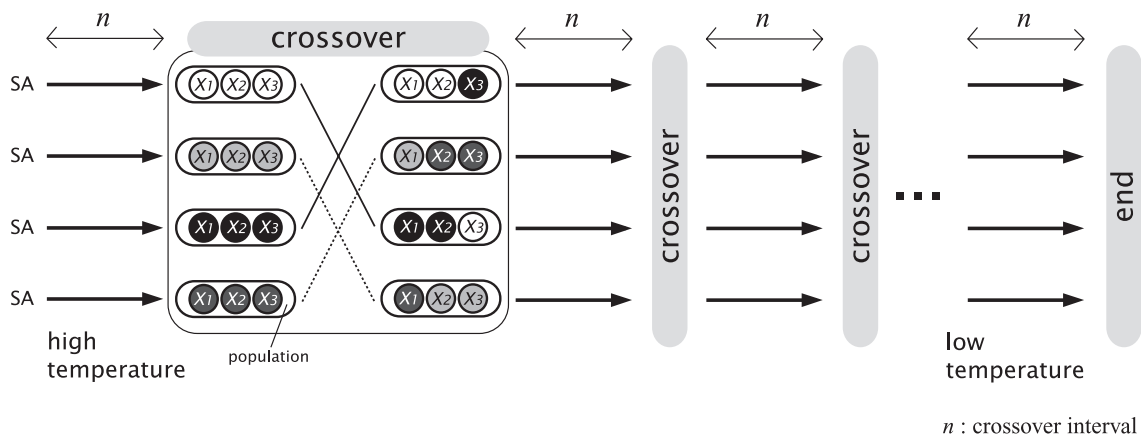


図 1: 遺伝的交叉を用いた並列 SA(PSA/GAc) の概念図

とする。

提案する遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) では、一定期間 n に達するとその時点での解を個体とし、個体をランダムに 2 つ選択して遺伝的交叉を行う。ここで用いる遺伝的交叉は一点の設計変数間交叉である。設計変数間交叉とは各設計変数の間でのみ交叉を行うことをいう。図 2 に示すように、もとの親と生成した子との 4 個体間のうち評価値の高い 2 個体を選択して、選択された 2 個体から次の探索を続ける。ある設計変数の最適値が求まっている場合、遺伝的交叉によってその設計変数の最適値を他の SA 探索に伝達することができるため、アニーリングの収束を早めることができる。これらの処理を繰り返し、終了条件を満たすまで探索を続ける。

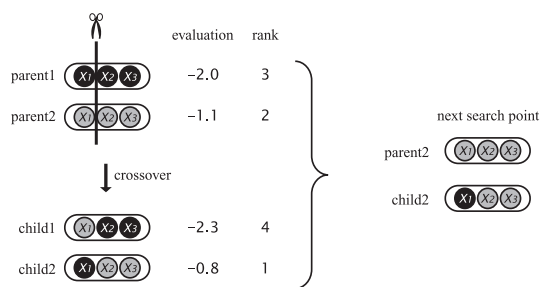


図 2: 設計変数間交叉と選択

3 数値実験

3.1 GA オペレータを用いた 3 種の並列 SA との比較

並列に実行している各シミュレーテッドアニーリング (SA) の探索途中の解の伝達に遺伝的アルゴ

リズム (GA) のオペレータである遺伝的交叉を用いることの有効性を検証するため、遺伝的交叉以外の GA オペレータを解の伝達方法として用いた他の 3 種の並列 SA と解探索能力を比較した。これらは一定間隔ごとに以下のような方法で解の伝達を行うものである。

- エリート選択を用いた並列 SA (elitePSA) : 1 つのエリート個体の解を他のすべての個体の新たな探索点とする
- ルーレット選択を用いた並列 SA (roulettePSA) : ルーレット選択によりすべての個体の新たな探索点を決定する
- エリート保存を含むルーレット選択を用いた並列 SA (e-roulettePSA) : エリート保存を用いたルーレット選択によりすべての個体の新たな探索点を決定する

対象問題として、大域的に局所解の多い Rastrigin 関数と、大域的にはなだらかだが局所的に多くの極小値を持つ Griewank 関数の 2 つを用いた。設計変数の数は 2 とした。これらの関数は、各設計変数の値が 0 のときに最適値 0 をとる。探索に用いる個体数は 20, 200 とした。なお、本数値実験では、終了条件を「解の $1.0e-4$ 以上の値が 100 世代変化しないこと」とし、 $1.0e-4$ 以下の解が得られればそれを最適解とした。

図 3 は、それぞれの並列 SA の個体数を 20 として Rastrigin 関数を解いた結果である。横軸は初期温度、縦軸は 1 個体の持つエネルギーつまり解の値であり、10 試行の平均値を示している。図 3 からは各並列 SA の結果に大きな違いがあることがわかる。遺伝的交叉を用いた並列 SA(PSA/GAc)

は初期温度によらずに常に最適解を求められているが、他の3つの手法ではどのような初期温度でも常に最適解を求めることができなかった。

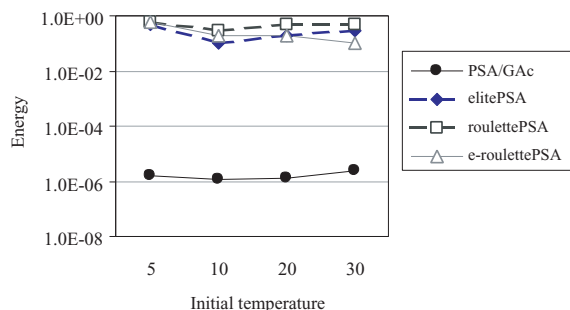


図 3: Rastrigin 関数の結果 (個体数 20)

各並列 SA の個体数を 10 倍の 200 としたときは解の伝達方法による差はなく、4 つの手法すべてで常に最適解が求められた。個体数を増加させても各個体の繰り返し計算回数が増え、すべての手法で解が求まったと考えられる。

一方、個体数を 20 として Griewank 関数を対象としたときは、PSA/GAc の解が比較的良かったが、どの手法でも 100% の確率では最適解が求められず有意な差はなかった。しかし 200 個体を用いた場合には、図 4 に示すように大きな差が生じた。ルーレット選択を用いた並列 SA とエリート保存を含むルーレット選択を用いた並列 SA ではどの初期温度でも常に最適解は求められなかったが、エリート選択を用いた並列 SA では初期温度によっては 100% で最適解が求められることがあった。PSA/GAc では初期温度によらず常に最適解が求められた。

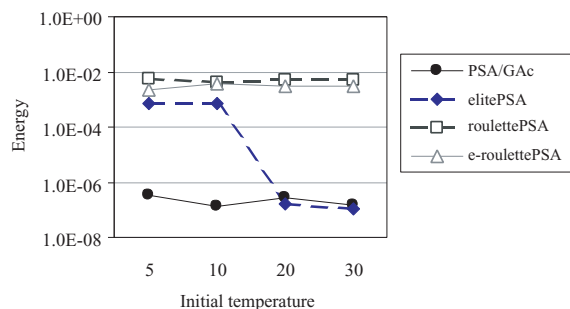


図 4: Griewank 関数の結果 (個体数 200)

Griewank 関数は全体的な景観はなだらかであるが細かく見ると局所解が多数あるため、個体数の少ないときには局所解にはまってしまい、最適解

を求めることができない。個体数を増やすことで最適解にたどり着く可能性が上がったものと考えられる。

これらの結果から提案手法である PSA/GAc の解探索能力が最も優れているといえる。またエリート選択を用いた並列 SA も比較的優れた解探索能力を示しているが、最適解を求められる確率は PSA/GAc より低い。このためエリート選択を用いた並列 SA は、PSA/GAc と比較すると解探索能力は低いといえる。

3.2 分散 GA・逐次 SA との比較

3.1 節の数値実験によって、PSA/GAc の解探索能力が優れていることが明らかとなった。本節では設計変数を増加させた問題を対象として実験を行い、分散 GA との比較を行った。また逐次 SA との探索能力の差も示した。

対象問題は、3.1 節で用いた Rastrigin 関数、Griewank 関数と、設計変数間に強い依存関係のある Rosenbrock 関数とした。設計変数の数は 10、30 とした。Rosenbrock 関数は、各設計変数の値が 1 のときに最適値 0 をとる。

各手法の評価計算回数は同等にし、終了条件を満たしたときに探索を終了した。PSA/GAc では個体数 400、8000 世代の探索を行った。逐次 SA は 3200000 世代 (8000 世代 × 400 個体に相当) の探索を行う SSA-long、8000 世代の探索を独立に 400 回実行する SSA-short の 2 種類とした。また分散 GA は 20 個体 × 20 島の 400 個体とし、8000 世代の探索を行った。PSA/GAc と 2 種の逐次 SA の初期温度は 10 に統一した。各手法について試行はそれぞれ 10 回ずつ行い、最適解が得られた確率を表 1 に示した。Success rate とは、試行回数に対して最適解を得られた回数の割合を示している。なお、最適解は 3.1 節と同様に定義する。

表 1: PSA/GAc と分散 GA、逐次 SA を用いて最適解が得られた確率

		Success rate			
		PSA/GAc	SSA-long	SSA-short	GA
Rastrigin	10dimensions	1.00	0.00	0.00	1.00
	30dimensions	0.00	0.00	0.00	0.10
Griewank	10dimensions	0.90	0.00	0.00	0.00
	30dimensions	1.00	0.00	0.00	0.70
Rosenbrock	10dimensions	1.00	0.00	0.10	0.00
	30dimensions	1.00	0.00	0.00	0.00

まず、PSA/GAc と 2 種の逐次 SA との結果を比較すると、逐次 SA と PSA/GAc の評価計算回数

は等しいことから、単に SA のアニーリング時間や回数を増加しただけでは結果が向上しないことがわかる。また分散 GA との結果を比較すると、GA での探索が困難な設計変数間に依存関係のある問題に関しては、特に PSA/GAc の探索が有効であることがわかる。設計変数間に依存関係のない関数を対象としたときも、GA に劣らない探索が行われている。これらの結果から、提案する PSA/GAc の探索能力は非常に高いといえる。

3.3 タンパク質の構造解析

タンパク質は生命現象に直接関わる重要なものであり、構造を解明することは生命現象の仕組みを説明することにもつながる。タンパク質の立体構造はエネルギーの最小状態に対応しており、エネルギーを最小とするような構造を最適化手法を用いて求めることで、構造解析が可能である。これまで、タンパク質の構造解析においては SA が主として使用されてきた³⁾。岡本らは、小規模なタンパク質 (Met-enkephalin) を対象として構造解析における SA の有効性を確かめている⁴⁾。

Met-enkephalin は Tyr-Gly-Gly-Phe-Met という 5 個のアミノ酸からなり、 $E \leq -11\text{kcal/mol}$ の領域で最小エネルギー構造をしている。Met-enkephalin の主鎖における 10 個の二面角と、側鎖における 9 個の二面角をそれぞれ設計変数とする。つまりこのタンパク質は 19 個の設計変数を持っており、1MCsweep によって 19 回のアニーリングの処理が行われるものとする。

本実験では PSA/GAc を用いて Met-enkephalin の解析を行い、岡本らの結果と比較した。岡本らの実験で用いられた逐次 SA では、1MCsweep ごとに 19 個の二面角をそれぞれ生成・受理判定するので、100000MCsweeps の評価計算回数は $100000 \times 19 = 1900000$ 回となる。PSA/GAc では交叉オペレーションを実行するときにも評価計算を行う。そのため、20 個体を用いたときに評価計算回数を約 1900000 回とするために、MCsweep 数は 4992 とした。

それぞれ試行は 10 回ずつ行い、最適構造が得られた確率を表 2 に示した。Success rate は 3.2 節と同様である。

表 2: 最適構造が得られる確率

Number of MCsweeps	Evaluations	Success rate
4992 (PSA/GAc)	100005×19	0.90
100000 (SSA)	100000×19	0.50

表 2 から、最適構造を得る確率は、逐次 SA を用いる場合よりも PSA/GAc を用いた場合の方が高いことが明らかとなった。

4 結論

本研究では、遺伝的アルゴリズム (GA) のオペレータである遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) を提案し、その有効性を検討した。

GA の他のオペレータであるエリート選択やルーレット選択などを用いた並列 SA と、PSA/GAc との性能を比較した数値実験では、2 設計変数の Rastrigin 関数と Griewank 関数を対象問題とした。その結果、PSA/GAc が最も良いふるまいをした。

分散 GA や逐次 SA との比較を行い、PSA/GAc の有効性を検討した数値実験では、探索がより困難な 10 設計変数以上の Rastrigin 関数、Griewank 関数、Rosenbrock 関数を対象問題とした。PSA/GAc と逐次 SA と比較した結果、提案する PSA/GAc は収束が速く解の品質も良いことが示された。また分散 GA と比較した結果、GA での探索が困難な問題に対しては PSA/GAc が有効であることが示された。これらから PSA/GAc は優れた解探索能力を有するといえる。

5 個のアミノ酸をからなる Met-enkephalin を対象に構造解析を行った数値実験では、従来用いられていた逐次 SA よりも PSA/GAc の解探索能力が高いことが明らかとなった。この結果、PSA/GAc はタンパク質の構造解析に有効な手法であることが示された。

今後は、現在用いられている手法では解析が困難である大規模なタンパク質の構造解析に PSA/GAc を適用し、良好な解が得られることを確認する。

謝辞

本研究は文部省からの補助を受けた同志社大学の学術フロンティア研究プロジェクト「知能情報科学とその応用」における研究の一環として行った。

参考文献

- 1) Bruce E. Rosen, 中野良平. シミュレーテッドアニーリング -基礎と最新技術-. 人工知能学会誌, Vol. 9, No. 3, 1994.
- 2) E. H. L. Aarts, J. H. M. Korst. *Simulated annealing and Boltzmann machines*. John Wiley & Sons, 1989.
- 3) 木寺詔紀. コンピュータ解析 -最適化をめぐる-. 蛋白質 核酸 酵素, Vol. 39, No. 7, 1994.
- 4) 岡本祐幸. モンテカルロシミュレーションで探るタンパク質の折り畳み機構. 物性研究, Vol. 70, No. 6, pp. 719-742, 7 1998.

出典：

情報処理学会研究報告 Vol.2001, No.27

ISSN 0919-6072, pp. 57-60

(2001年3月第33回MPS研究会講演論文)

問い合わせ先：

同志社大学工学部/同志社大学大学院工学研究科

知的システムデザイン研究室

(<http://mikilab.doshisha.ac.jp>)