

ランダム個体群に基づくグラフにリンケージ同定最小化手法

鐘 睿^{1,a)} 棟朝 雅晴^{2,b)}

概要: 大規模最適化問題を解決するために協調的共進化 (Cooperative co-evolution) を使用する. 理論的に完璧な分割法は最適化精度を損失せずに探索空間を指数的に減少できる. 高精度の分解法は, 変数間の依存関係を検出するために大量の評価回数を使用する. 代表的なアルゴリズムは, リンケージ同定と Differential Grouping (LINC-R / DG). 大量の計算コストと局所的な非線形性チェックはこれらの手法の制約である. この論文の提案手法では, 変数分解を最適化問題と見なし, LINC-R/DG に基づいて目的関数を設計する. 目的関数の実現可能性を数学的に説明する. 数値実験では, 我々の提案手法が大規模最適化問題を解決するため有望な技術を示す.

Random population-based decomposition method by Linkage identification with non-linearity minimization on graph

1. はじめに

Cooperative co-evolution (CC) [1] は分割統治法に基づく大規模最適化問題を解決するために提案された. CC の基本的な考えは, 元の問題を複数の分離不可なサブ問題に分割し, サブ問題毎を最適化する.

変数分割手法の設計は CC の中心である. 多くの研究 [2], [3] は, CC が問題の分割手法に敏感ということを示す, 分割手法の設計はホット研究トピックになる.

本論文では, ランダム個体群に基づくグラフにリンケージ同定最小化手法 (graphLINC-Rmin) という新しい変数分割手法を提案し, 変数分割問題を最適化問題として自動的に行うことが可能になる. 目的関数は LINC-R に基づいて設計される.

大規模最適化問題の解決における提案の優位性を説明するために, CEC2013 LSGO suite[4] で提案手法をテストし, standard Differential Grouping (DECC-DG) [5] と比較する. これで特定の競争力を示す.

残りの論文は次のように構成される. 第 2 節は, 関連研究について, 遺伝子の依存関係を説明する. また, CC を紹介し, 現在ポピュラー分割方法を簡単にレビューする. 第 4 節では, 我々の提案を詳しく説明する. 第 3 節では,

数値実験を示し, 実験結果を分析する. 第 5 節で論文の結論と今後の研究方向を示す.

2. 関連研究

2.1 遺伝子の依存関係

遺伝子間の依存関係は, エピスタシス [6] また Gene interaction[5] として知られている. 生物学では, 表現型レベルの特徴が 2 つ以上の遺伝子によって決定される場合, これらの遺伝子は依存関係を持っていると考えられ, 構成されるゲノムはリンケージセット [7] と呼ばれる. 遺伝子依存関係の概念を最適化問題の定義に拡張する. $f(x_1, x_2, \dots, x_n) = \sum_{i=1}^m f(x_{i_1}, \dots, x_{i_k}), L_i = \{i_1, \dots, i_k\}$ はリンケージセットとして定義され, 同じ変数でも異なるリンケージセットにあることも可能, これはオーバーラップ変数と呼ばれまる. この数式では, $f(x)$ は部分的に分離可能関数と呼ばれる. リンケージセット内の変数ペアに直接的あるいは間接的な依存関係が存在し, 異なるリンケージセット間の変数ペアに依存関係はないと考えられる.

2.2 CC とポピュラー変数分割手法

CC は大規模最適化問題を解決するためのメインフレームワークとして知られている. Cooperative Coevolutionary Genetic Algorithm (CCGA)[1] は CC と同時に提案される変数分割手法である. CCGA は n 次元の問題を $n \times 1$ 次元のサブ問題に分割する. この手法は変数間の依存関係を

¹ 北海道大学大学院情報科学院

² 北海道大学情報基盤センター

a) zhongrui19961230@gmail.com

b) munetomo@iic.hokudai.ac.jp

完全に無視して、CC のパイオニアと考えられる。Random Grouping[8] も簡単に n 次元の問題を $m \times k$ 次元のサブ問題に分割する。 ($n = m \times k$), ハイパーパラメータはユーザーで決められる。数学的説明では、いくつかの試行実験で、2つの依存関係がある変数を少なくとも1回同じサブ問題内に分割する確率はかなり高いことを示した。

関数の fitness 景観に関する知識がほとんどない場合、高精度の分割は不可能。Delta Grouping[9] は、この問題を緩くする。Delta Grouping は x_i と x_j が分割可能な変数である場合、初期のランダム個体群から最適化された個体群への座標の差が大きくなることを気付いた。Delta Grouping はこの現象を利用して、ソートした次元の差に従って、変数を $m \times k$ 次元サブ問題に分割する ($n = m \times k$)。CEC2008 LSGO Suite と CEC2010LSGO Suite での実験結果は、この大まかな推定方法が、先験的な知識がほとんどなくても依存関係をうまく捉えることができることを示す。

実数コード GA にリンケージ同定 (LINC-R) [7] は高分割精度を目指して提案される。LINC-R は、 x_i と x_j の依存関係を判断する場合に、 x_i と x_j に摂動を加えて、fitness 差によって判断する。Diffenretial Grouping (DG) [5] は、LINC-R のメカニズムを 1000 次元問題に適用して、CEC2010 LSGO Suite に評価する。現在、LINC-R/DG に基づく変数分割法は高精度の分割法と考えられて、graphDG [10] は CEC2010LSGO Suite のほとんどのベンチマーク関数で 100% の精度を達成した。ただし、LINC-R/DG-based アルゴリズムの欠点は、変数間の依存関係を検出するために大量の計算コストが必要。最近いくつかの研究では、計算コストが $O(n^2)$ [5] から $O(n \log(n))$ [11] に削減されたが、もっと高次元の問題（数千数万変数を含む）では、単純に LINC-R/DG-based アルゴリズムを使う計算コストは許容できない。また、計算コストが高いため、変数間の依存関係は局所的な線形／非線形特徴を大局的な線形／非線形特徴に表わすことが多い。これも改善する必要がある。

3. graphLINC-Rmin

第3節では、提案を詳しく紹介する。我々の提案のフローチャートを図1に示す。

次に、フローチャートに従って提案を説明する。

我々の提案では、変数間の依存関係をグラフに変換し、隣接行列として保存する。5つの変数を含む例を図2に示す。

グラフを隣接行列として保存することで、自適応ランダムローカルサーチを行列に直接使用できる。さらに、初期化で n サイズの母集団をランダムに生成して、大規模最適化問題での LINC-R の局所的な線形／非線形を調べるという不足を克服する。

分割ソリューションをどのように評価するかが私たちの研究の中心。過去の研究では、Fitness Difference Minimization (FDM) [2] という分割法は、最初に n 次元の

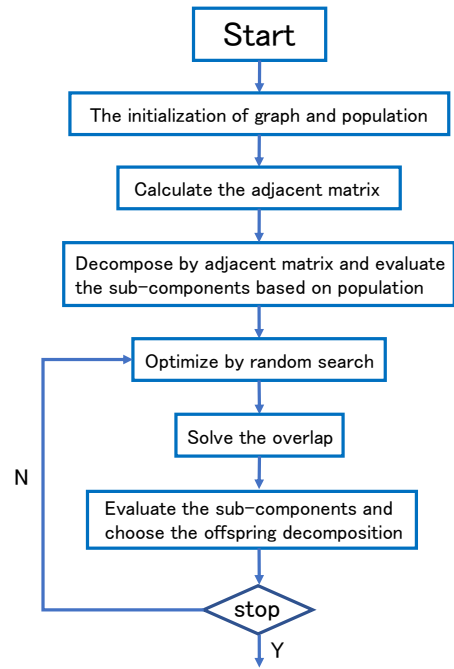


図1 我々の提案のフローチャート

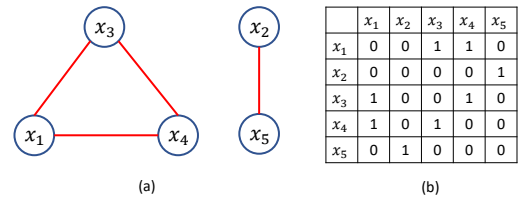


図2 グラフから隣接行列への変換。(a) グラフは変数間の依存関係を示す。(b) (a) から隣接行列に変換する。

問題を $m \times k$ 次元のサブ問題 ($n = m \times k$) に分割し、組み合わせ最適化問題と見なされる。式 (1) は FDM の目的関数である。

$$\min(f(x) - \sum_{i=1}^m (f_i(x))^2) \quad (1)$$

$f_i(x)$ は k 個の変数を含むサブ問題である。FDM は CEC2010LSGO Suite に良い結果を得た。私たちの実験では、variant LINC-R を目的関数とする。詳細な説明は以下になる。

LINC-R のオリジナルバージョンを式 (2) に示す。

$$\begin{aligned} & \text{if } |(f(s_{ij}) - f(s_i)) - (f(s_j) - f(s))| < \epsilon \\ & \text{then } x_i \text{ and } x_j \text{ are separable} \end{aligned} \quad (2)$$

s は LINC-R の基準点、 s_i と s_j は s に i^{th} 次元と j^{th} 次元を δ 摂動を加える個体であり、 s_{ij} は i^{th} 次元と j^{th} 次元両方に摂動を加える個体である。我々の提案に LINC-R の条件 (if の中身) を式 (3) に変換する。

$$|(f(s_{ij}) - f(s)) - ((f(s_i) - f(s)) + (f(s_j) - f(s)))| < \epsilon \quad (3)$$

variant LINC-R はベクトルの足し算形式に理解できる．この発見に基づいて，次に LINC-R を 3 次元空間に応用する．3 次元空間に，LINC-R を応用する形は図 3 に示す．

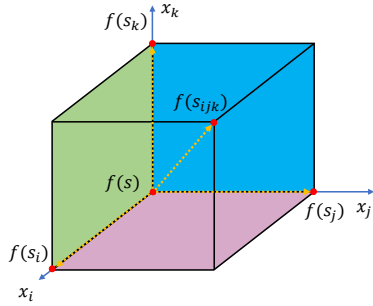


図 3 LINC-R を 3 次元空間に応用

そこで，3 次元空間では，LINC-R の条件を式 4 に変換する．

$$|f(s_{ijk}) + 2f(s) - (f(s_i) + f(s_j) + f(s_k))| < \epsilon \quad (4)$$

したがって，次元数が n になると，LINC-R の条件を式 5 に推測できる．

$$|f(s_{1,2,\dots,n}) + (n-1)f(s) - \sum_{i=1}^n f(s_i)| < \epsilon \quad (5)$$

ただし，式 (5) が満たされない場合に，依存関係が一部の 변수ペア間に存在することが分かるけど，どちらの 변수ペアに存在することが分からない．しかし，アルゴリズムで積極的に変数間の依存関係を検出することができる．この説明によると， n 次元問題を分割するための目的関数は式 (6) で定義される．

$$\min((f(s_{1,2,\dots,n}) + (m-1)f(s) - \sum_{i,j,\dots}^m f(s_{i,j,\dots}))^2) \quad (6)$$

グラフの隣接行列に基づいて，我々は自適応ランダムローカルサーチアルゴリズムを設計した．基本的な考えは，アルゴリズムの初期段階では，グラフ上の多数のコネクション状態を変更して許容可能な分割ソリューションを見つけ，後期段階では，アルゴリズムは少数の接続状態を変更し，どんどん収束に迫り着く．

自適応ランダムローカルサーチによって見つけるソリューションでは，オーバーラップ変数を含む可能性が高い．オーバーラップ変数をどのように処理することがよく研究される問題である．ここで，簡単な戦略を応用して，オーバーラップ変数の依存関係をランダムに無視する．

4. 数値実験と分析

我々の提案手法の優位性を示すために，CEC2013 LSGO Suite に standard DG を比較して，25 回の試行実験を行う．実験条件は次のようにする．表 1 は変数分割最適化のパラメーターであり，表 2 はサブ問題最適化のパラメーターである．

表 1 変数分割最適化のパラメーター

パラメーター	値
最適化方向	最小化
初期コネクション行列	Zero-like
個体群規模	5
ローカルサーチの自適応ストラテジー	Linearity
オーバーラップコネクション無視率	0.5
最大世代数	100

表 2 サブ問題最適化のパラメーター

パラメーター	値
最適化方向	最小化
オブティマイザ	DE/current-to-best/1
評価回数の制約	3,000,000
個体群規模	30 × サブ問題の次元数
スケールファクター	0.7
交差率	0.9

提案の優位性を示すために，サブ問題の最適化段階の最後に，我々の提案と standard DG に Mann-Whitney U 検定を使用する．また，提案が $p < 0.05$ より優れる場合に「*」を付き， $p < 0.01$ より優れる場合に「**」を付く．収束曲線は図 4 に示します．

我々の提案のパフォーマンスは CEC2013 LSGO Suite の多くの関数に standard DG より大幅に優れる．理由はこれから説明する．

我々の提案は，完全分離可能関数の分割段階に大量の評価回数を節約し，サブ問題の最適化段階の計算資源は standard DG より多い．これは私たちの提案が優れる理由．部分的分離不可関数では，2 つの方面で説明する．まず，提案は standard DG が局所的な線形性/非線形性を検出することより，複数のサンプルに基づいて線形性/非線形性を調べる．これは standard DG よりロバストになる．さらに，変数分割最適化段階では，強依存関係は弱依存関係よりも目的関数に大きなインパクトを与えて，そういうコネクションを後代に保留する可能性も高い．簡単な例で説明する． $f(x) = 100(x_1 - x_2)^2 + (x_2 - x_3)^2$, $s = (0, 0, 0)$, $s_{123} = (1, 1, 1)$ ．この例では， x_1 と x_2 の依存関係は強い， x_2 と x_3 の依存関係は比較的弱い．初期化に x_1, x_2, x_3 は全て独立と判定する，目的関数の値は $(f(s_{123}) + 2f(s) - (f(s_1) + f(s_2) + f(s_3)))^2 = 40401$ にな

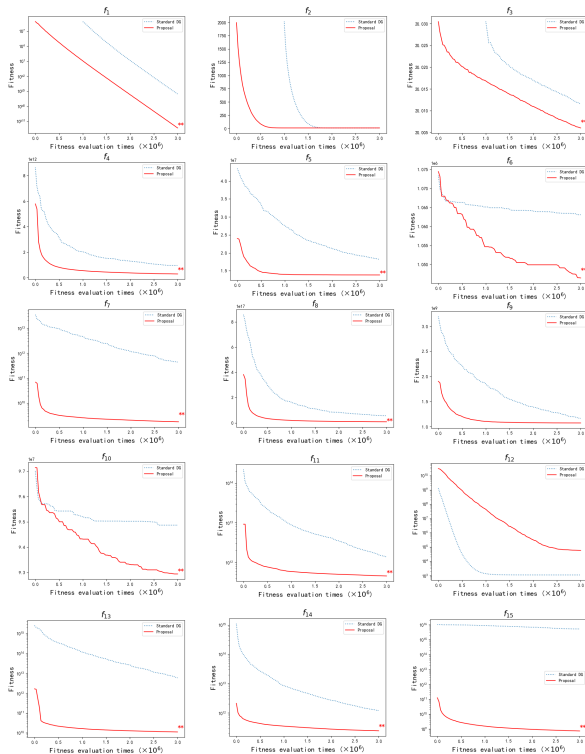


図 4 我々の提案と standard DG の収束曲線

る。 x_1 と x_2 だけの依存関係を判明する場合に、目的関数の値は $(f(s_{123}) + f(s) - (f(s_{12}) + f(s_3)))^2 = 4$ になる。 x_2 と x_3 だけの依存関係を判明する場合に、目的関数の値は $(f(s_{123}) + f(s) - (f(s_{12}) + f(s_3)))^2 = 40000$ になる。強依存関係を判明した場合、目的関数にインパクトは弱依存関係より大きいことが分かる。したがって、自適応ランダムローカルサーチでは、強依存関係を判明する場合、その強依存関係を後代に保留する可能性も高い、同時にいくつかの弱依存関係を無視する。この処理はサブ問題の最適化段階の誤差が増加するけど、探索空間は指数的に減少することができる、特に評価回数制限がある最適化問題に有効と考えられる。

5. 結論

本論文では、変数分割問題を最適化問題と見なし、高次元 LINC-R に基づく目的関数を設計するという新しい分割アルゴリズムを設計した。私たちの研究では、LINC-R はベクトルの足し算形式として理解し、高次元形式を目的関数とする。提案を評価するために、CEC2013 LSGO Suite に standard DG と比較して応用する。実験結果は、私たちの提案のパフォーマンスがほとんどのテスト関数で standard DG より大幅に優れることを示す。また、計算コストも完全分離可能な関数と一部の部分的分離不可関数で standard DG よりも優れる。ということで、高次元 LINC-R は変数分割最適化問題の目的関数とすることは有望な研究トピックである。ただし、自適応ランダムロー

カルサーチの制限により、分割最適化問題に大局的なソリューションを見つけることができない場合が多く、GAなどの最適化手法では、高次元の問題に染色体のコーディングがスパースと長くなって、分割問題に应用することが困難になる。その最適化方法は私たちの将来の研究課題になる。最後、私たちの提案は、大規模問題に直接 LINC-R を使用するよりうまく対応できる。私たちの提案をもっと高次元の問題を解決することも、研究のテーマである。

6. 謝辞

本研究は JSPS 科学研究費（課題番号 JP20K11967）の助成を受けたものである。

参考文献

- [1] M. Potter and K. De Jong, “A cooperative coevolutionary approach to function optimization,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 866 LNCS, pp. 249–257, 1994.
- [2] E. Sayed, D. Essam, and R. Sarker, “Dependency identification technique for large scale optimization problems,” in *2012 IEEE Congress on Evolutionary Computation*, 2012, pp. 1–8.
- [3] M. N. Omidvar, X. Li, and X. Yao, “A review of population-based metaheuristics for large-scale black-box global optimization: Part a,” *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2021.
- [4] X. Li, K. Tang, M. N. Omidvar, Z. Yang, K. Qin, and H. China, “Benchmark functions for the CEC 2013 special session and competition on large-scale global optimization,” *gene*, vol. 7, no. 33, p. 8, 2013.
- [5] M. Omidvar, X. Li, Y. Mei, and X. Yao, “Cooperative co-evolution with differential grouping for large scale optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 378–393, 2014.
- [6] M. Munetomo, “Linkage identification with epistasis measures considering monotonicity conditions,” 2002.
- [7] M. Tezuka, M. Munetomo, and K. Akama, “Linkage identification by nonlinearity check for real-coded genetic algorithms,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3103, pp. 222–233, 2004.
- [8] Z. Yang, K. Tang, and X. Yao, “Large scale evolutionary optimization using cooperative coevolution,” *Information Sciences*, vol. 178, no. 15, pp. 2985–2999, 2008.
- [9] M. Omidvar, X. Li, and X. Yao, “Cooperative co-evolution with delta grouping for large scale non-separable function optimization,” 2010, pp. 1–8.
- [10] Y. Ling, H. Li, and B. Cao, “Cooperative co-evolution with graph-based differential grouping for large scale global optimization,” in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2016.
- [11] Y. Sun, M. Kirley, and S. K. Halgamuge, “A recursive decomposition method for large scale continuous optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 5, pp. 647–661, 2017.