スケールフリーの呪い
—多スタート法による大規模組合せ最適化の困難さ—

01606724　東京都立大学　*増山博之　MASUYAMA Hiroyuki
01308414　関西大学　檀寛成　DAN Hiroshige
01014954　大阪大学　梅谷俊治　UMETANI Shunji

本研究の目的は, 多スタート法を用いた大規模組合せ最適化の困難さについて理論的に解析し, 理解を深めることである. 多スタート法は,「多様化機構」をもつメタヒューリスティクスであり, 異なる初期解から局所探索法を繰り返し再スタートさせる. 多スタート法は, 多様化機構の種類によって, (i) ランダム多スタート局所探索法や GRASP などのランダム多スタート (RMS) 法と, (ii) 反復局所探索法や可変近傍探索法などのランダム摂動 (RP) 法に大別される. RMS 法は, ランダム化構築法によって初期解を生成するのに対し, RP 法は得られた「良質解」を摂動させて初期解を生成する. 一般に, 多スタート法は実装が容易で比較的良好な性能を達成することから, 目的特化型のヒューリスティクスを設計する際のベンチマークとなる.

本研究の目的を果たすため, 標準的な多様化機構をもつ RMS 法に着目し, 大規模組合せ最適化における最適値と最良実験値のギャップ (誤差) を詰める困難さを解析する. 厳密に言えば, RMS 法で生成される実験解の目的関数値は独立同一分布に従う保証はないが, その仕様から, 偏りの少ない良質解を生成している可能性が高く, 実験解の集合は良質解の分布に関する有益な情報を含むと期待できる.

さて, 本研究に関連するテーマとして, 最適値の推定がある. 最適値の推定法は 2 つに分類できる. 一つは, 緩和問題を利用して最適値の下界 (あるいは上界) を求める「緩和法」である. もう一つは, 極値理論などの統計的手法を用いて最適値の点推定や区間推定を行うものである [1]. ちなみに, 極値理論による最適値の推定に関する先駆的な研究は, Golden [2] によってなされている.

しかし仮に, 最適値からの最良実験値のギャップがわかったとしても, そのギャップを詰めるための計算時間の見積りは容易ではない. 多くの場合, 計算の初期段階ではギャップは急速に改善するが, 比較的早くギャップがほとんど減少しない定常的な段階に到達する. 図 1a は, そのような状況の典型例を示している. 図 1a が示すような状況では, 推定最適値からの最良実験値のギャップが小さくても, そのギャップを縮めるための計算コストは膨大である.

また, 図 1b は, 最適値からの最良実験値の相対ギャップがスケールフリー性をもつ, すなわち, 冪的に減少することを示している. つまり, 最良実験値の相対ギャップを半減させるのに必要な反復回数は, 完了した反復回数に漸近的に比例して増加する. これは「ゴールに近づこうとすると遠ざかる」現象であり, スケールフリー性に起因する「呪い」とも言うべき「大規模組合せ最適化の困難さ」の現れの一つである. 以下では, この現象を「スケールフリーの呪い」という.

ここで「スケールフリーの呪い」の理論的根拠となる「期待相対ギャップの冪乗則公式」を紹介する (導出に極値理論を用いるので最大化問題を考える). $x^* > 0$ を最適値とし, 最良実験値が x であるという条件のもとで RMS の追加 n 回の反復によって達成される最良実験値を Z_n とおき,

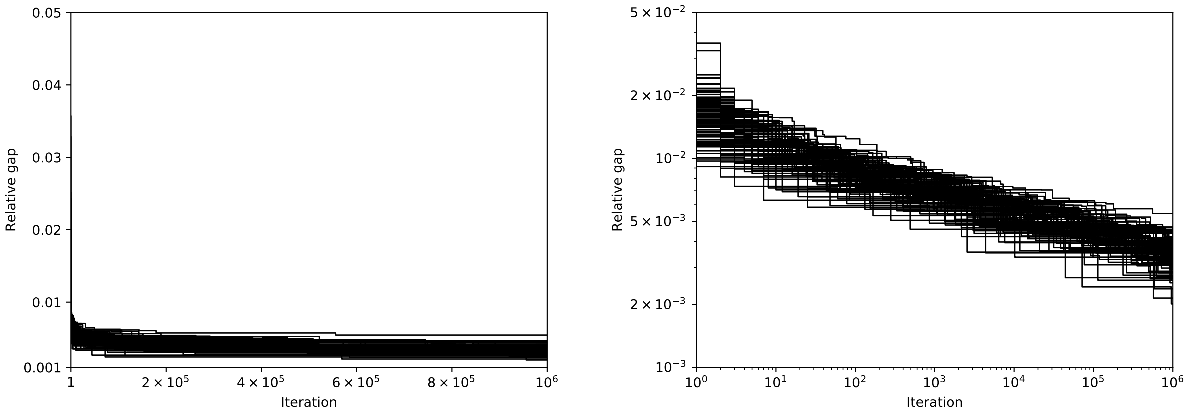
$$\Delta_n(x) := \left|\frac{x^* - \max(Z_n, x)}{x^*}\right|$$

の期待値 $\mathsf{E}[\Delta_n(x)]$ を最良実験値の期待相対ギャップとよぶ. そして, 次式が本研究の主要な解析結果の一つ, 期待相対ギャップの冪乗則公式である [3].

$$\mathsf{E}[\Delta_n(x)] = \frac{\Gamma(-\xi)}{|x^*|} L(n) n^{\xi}, \quad x < x^* < \infty. \quad (1)$$

ここで, $\xi < 0$ は極値指数, $\Gamma(\cdot)$ はガンマ関数, $L(\cdot)$ は適当な緩慢変動関数である.

先に述べた「スケールフリーの呪い」は, この冪乗則公式 (1) が理論的な根拠となっている. 実際,



(a) Normal axes　　(b) Logarithmic axes

図 1: 100 個のランダムな巡回セールスマン問題 (TSP) に RMS 法を適用したときの最良実験値の相対ギャップの推移（ランダム化した最近近傍法で初期解を生成し Lin-Kernighan (LK) 法を実行)

式 (1) より, 次が成り立つ.

$$\lim_{t\to\infty} \frac{\mathsf{E}[\Delta_{n+h(n)}(x)]}{\mathsf{E}[\Delta_n(x)]} = \frac{1}{2}, \quad x < x^*. \tag{2}$$

ただし, $h(n) = (2^{-1/\xi} - 1)n$ とする. 式 (2) は, 期待相対ギャップの半減期が, 完了した反復回数に漸近的に比例して増加することを示しており,「ゴールに近づこうとすると遠ざかる」現象を表している.

冪乗則公式 (1) からわかることは, RMS 法と比べて指数関数的に性能が加速されたアルゴリズムでなければ,「スケールフリーの呪い」を打破できないということである. 本研究では, TSPLIB から取得した 5 つの TSP インスタンス (brd14051, d15112, d18512, rl11849, usa13509) に対して, 3 つの反復局所探索法 (LK 法と 3 種類の初期解生成法を組み合わせた) を適用したが, 最適値に対する最良実験値の相対的ギャップは高々冪的にしか減衰せず [3] (口頭発表時に紹介),「スケールフリーの呪い」は克服されなかった. 一般に, 反復局所探索法は有力な多スタート法の一つだとされているので,「スケールフリーの呪い」を打破する多スタート法の開発は容易ではないと思われる.

本研究の結果から, より良いヒューリスティクスの開発を目指しても意味がないと結論づけるのは早計である. これまで多くのソルバー開発者が実践してきたように, 合理的な計算時間でより良い解を生成することは現実的に重要である. つまり本研究が示唆するのは,「いかにギャップを詰めるか」よりも「いつ計算を止めるか」が重要だということである. そして, その「いつ」を, コストと利益のトレードオフを考慮して決める必要がある. しかし, このトレードオフの問題は, 最適値と最良実験値のギャップを推定するだけでは解決しない.

本研究の理論的および数値的な結果が示すように, 大規模組合せ最適化においては, 最適値と最良実験値のギャップが小さいにもかかわらず, そのギャップを縮めるための計算量は予想以上に大きくなる. したがって, 多スタート法の停止判断は, 最適値の推定よりも, 追加反復による最良実験値の「期待改善率」のような指標に基づいて行うのが効果的だろう.

参考文献

[1] A. P. Giddings, R. L. Rardin, and R. Uzsoy, Statistical optimum estimation techniques for combinatorial optimization problems: A review and critique, *Journal of Heuristics* 20(3), 329–358, 2014.

[2] B. L. Golden, A statistical approach to the TSP, *Networks* 7(3), 209–225, 1977.

[3] H. Masuyama, H. Dan, and S. Umetani, Curse of scale-freeness: Intractability of large-scale combinatorial optimization with multi-start methods, arXiv:2210.16678, 2022.