

Expensive Optimizationにおける進化アルゴリズムのEvolvabilityについて

On the Evolvability of Evolutionary Algorithms for Expensive Computational Budgets

田邊遼司 福永 Alex
Ryoji Tanabe Alex Fukunaga

東京大学大学院総合文化研究科
Graduate School of Arts and Sciences, The University of Tokyo

In some real-world problems, a computational time for evaluating the fitness of a solution is expensive. Thus, in recent years, there has been much research on such Expensive Optimization Problems (EOP) in the Evolutionary Computation community. The benchmark problems for evaluating the search performance of evolutionary algorithms for EOP have recently been proposed. However, to the best of our knowledge, there has been little work on the characteristic analysis for the EOP benchmarks. In this paper, in order to analyze the characteristic of the EOP benchmarks, we investigate evolvability of IPOP-CMA-ES on single/multi-funnel functions from the BBOB benchmarks for different computational budget scenarios. We used Fitness Distance Correlation (FDC) as an evolvability metric and the experimental results show that searchable spaces of IPOP-CMA-ES in some multi-funnel functions significantly differ depending on the maximum number of fitness evaluations.

注: 本原稿は我々の先行論文 [Tanabe 15] を本大会用に拡張したものである。

1. はじめに

進化計算コミュニティでは、実数値最適化問題に対する Evolutionary Algorithm (EA) の性能をベンチマーク関数を用いて評価する場合、最大評価回数を比較的大きく設ける慣習がある。例えば、一般的に使用されている CEC2005 benchmarks [Suganthan 05] では最大評価回数 $= 10^4 \times D$ である。ここで、 D は対象問題の次元数である。しかし、実問題においては解の評価にシミュレーションなどを使用するため、一つの解を評価するのに 10 分以上の計算時間がかかるような expensive optimization problem が存在する [Jin 05]。このような実問題において、最大評価回数を数万回に設定することは現実的に困難であり、高々 1,000 回程度しか設けることができない。近年では、このような背景から expensive optimization problem に対する研究が進化計算コミュニティでは増加している。

Expensive optimization problem に対する EA の性能評価のためのベンチマークセットがいくつか提案されている。例えば、expensive scenario の BBOB benchmarks [Hansen 09], CEC2014 expensive optimization benchmarks [Liu 14] などがある。しかし、その多くが通常の最適化問題におけるベンチマーク関数を使用し、単に最大評価回数を数千以内に制限しているだけである。例えば、[Hansen 09] では、従来の BBOB benchmarks の 24 個の関数を使用し、最大評価回数 $= 10^2 \times D$ としている。しかし、使用できる評価回数によって EA が到達可能な探索領域は異なる。そのため、最大評価回数が厳しく制限される状況 (例えば、最大評価回数 $= 10^2 \times D$) と、最大評価回数が比較的多く与えられる状況 (例えば、最大評価回数 $= 10^4 \times D$) では、EA が探索する空間が大きく異なる可能性がある。

本論文では、異なる最大評価回数における single-funnel 関数と multi-funnel 関数の EA の探索空間の比較を行う。こ

で、single-funnel 関数は大谷構造とも呼ばれる、微視的には無数の局所解が存在するものの、大域的には大きな 1 つの谷を有する [Boese 94]。一方、multi-funnel 関数は大域的に複数の谷を有する騙し構造を持つ関数であり、EA にとっては探索が困難であることが知られている [Lunacek 08]。multi-funnel 構造を有する実問題は存在し得るため、BBOB benchmarks を始めとする多くのベンチマークセットにはいくつかの multi-funnel 関数が含まれている。しかし、先に述べたように使用可能な評価回数によって EA の探索空間は異なるため、expensive optimization problem の設定においても multi-funnel 関数の特徴の影響を EA は受けない可能性がある。このことは誤評価の原因になり得るため、異なる最大評価回数における関数の性質を解析することは、今後の expensive optimization problem に対する EA の評価のために重要である。

2. Evolvability

Evolvability は対象問題に対して EA の探索がどれほど容易、又は困難なのかを表す、非常に緩く定義された指標である [Smith 02]。ある EA がなぜ特定の問題において効率的な探索が可能なのかなどを調査するには、evolvability に基づく解析は有用である [García-Nieto 12]。ここで、地形解析手法は対象問題の性質、構造の解析など問題そのものに焦点を当てている [Müller 11, Mersmann 14]。解のサンプリングには random sampling, random walk といった方法を使用し、EA とは独立した結果を得ることができる。一方、evolvability に基づく解析では、EA がサンプリングした解群を解析に使用する。そのため、evolvability は EA から見た景観とも解釈することができ、得られる解析結果は使用する EA に大きく依存する。本論文の目的である異なる最大評価回数における EA の探索空間の解析は、EA と問題が密接に関連しているため、evolvability に基づく解析を使用することにした。

本論文では 2.1 節で説明する Fitness Distance Correlation (FDC, r_{FDC}) [Jones 95] を evolvability の指標として使用した。ここで、evolvability の指標として一般的なものには、fitness clouds [Vérel 03], escape probability [Merz 04] などがある。しかし、これらは対象問題の大域的な探索空間にお

連絡先: 田邊遼司 東京大学大学院総合文化研究科 153-8902 東京都目黒区駒場 3-8-1 rt.ryoji.tanabe@gmail.com

る evolvability を計測するために設計された手法であり、本論文の目的である expensive scenario における evolvability の計測には不適切であったため、使用しなかった。

2.1 Fitness Distance Correlation (FDC)

FDC [Jones 95] は地形解析手法の一つであり、ある解 \mathbf{x} の評価値 $f(\mathbf{x})$ と最適解 \mathbf{x}^* からの距離 $d(\mathbf{x}, \mathbf{x}^*)$ の関係を測定することで、対象問題の大域的な構造を知ることができる。ここで、実数値最適化問題においては距離尺度としてユークリッド距離が一般的に使用される [Müller 11]。

図 1 のように、単に $f(\mathbf{x})$ を縦軸に、 $d(\mathbf{x}, \mathbf{x}^*)$ を横軸にとった 2 次元図で景観を可視化することができるが、より定量的な議論のためには以下の r_{FDC} [Jones 95] を使用する*1:

$$r_{FDC} = \frac{c_{FD}}{s_{FS}}, c_{FD} = \frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})(d_i - \bar{d}) \quad (1)$$

ここで、 n はサンプルされた解群 $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ の数である。 f_i , d_i は i 番目の解 \mathbf{x}_i ($1 \leq i \leq n$) の評価値と最適解 \mathbf{x}^* からの距離を表す。また、 \bar{f} , \bar{d} , 及び s_F , s_D はそれぞれ評価値と距離の平均値と標準偏差である。対象問題の r_{FDC} 値が 1 に近い程、最適解に近づくほど評価値が改善されていく single-funnel 関数であり、EA はこのような問題を容易に解くことができる。一方、 r_{FDC} 値が 0 付近では大域的な構造が弱く、負の場合は騙し構造を持つ multi-funnel 関数であることを意味する。

本論文の興味は [Boese 94, Jones 95, Müller 11] のように対象問題がどのような景観をしているかの調査ではなく、異なる最大評価回数における EA の evolvability を調査することである。そのため、最大評価回数が少ない設定において、多くの場合は到達できない最適解 \mathbf{x}^* を FDC に使用することは evolvability を計測する上で不適切であると考えられる。そのため、本論文では最適解 \mathbf{x}^* ではなく、1 試行において得られた最良解 \mathbf{x}^{bsf} を使用することにした。すなわち、 $d(\mathbf{x}, \mathbf{x}^*)$ を $d(\mathbf{x}, \mathbf{x}^{bsf})$ とした。

3. 実験

3.1 実験設定

本論文では IPOP-CMA-ES [Auger 05] を BBOB benchmarks [Hansen 09] の多峰性関数に適用し、その evolvability を調べた。ここで、IPOP-CMA-ES は CMA-ES [Hansen 01] にリスタート戦略を導入した手法であり、リスタート毎に集団数を 2 倍ずつ増加させることで、探索が経過するにつれてより大域探索能力が重視されていく。現在、IPOP-CMA-ES は実数値最適化問題に対する state-of-the-art な手法の 1 つとして知られている。IPOP-CMA-ES 以外にも DE と PSO といった EA での evolvability を調査したが、FDC 解析では 3 手法とも同様の傾向を示し、また IPOP-CMA-ES が比較手法中最も優れていたため、紙面の都合上 IPOP-CMA-ES の結果のみを報告する。なお、IPOP-CMA-ES のパラメタ設定は文献 [Auger 05] に従った。

Table 1 に BBOB benchmarks の 12 個の多峰性関数の各性質を示す [Mersmann 14]。ここで、本論文での興味の対象は異なる最大評価回数における多峰性関数での EA の探索空間の変化であったため、単峰性関数である $f_1, f_2, f_5 \sim f_{14}$, 及び弱い多峰性の f_8, f_9 は除外した。次元数 D は 5, 10, 20 としたが、いずれの次元数においても同様の傾向を示したので、10 次元の結果のみを報告する。最大評価回数は $10^2 \times D$, 及び $10^4 \times D$

*1 式 (1) は最小化問題を仮定している。

表 1: BBOB benchmarks の多峰性関数 [Hansen 09, Mersmann 14].

f	name	separ.	multi.	gl. structure
f_3	Rastrigin	high	high	strong
f_4	Büche Rastrigin	high	high	strong
f_{15}	Rastrigin	none	high	strong
f_{16}	Weierstrass	none	high	medium
f_{17}	Schaffers F7	none	high	medium
f_{18}	Schaffers F7-III	none	high	medium
f_{19}	Griewank-Rosenbrock	none	high	strong
f_{20}	Schwefel	none	medium	deceptive
f_{21}	Gallagher 101	none	medium	none
f_{22}	Gallagher 21	none	low	none
f_{23}	Katsuura	none	high	none
f_{24}	Bi-Rastrigin	none	high	weak

とした。前者は解評価の計算コストが重いため評価回数が厳しく制限される場合、反対に後者は評価回数を比較的多く与えられる問題を想定した設定である。探索中に得られた最良解 \mathbf{x}^{bsf} の評価値 $f(\mathbf{x}^{bsf})$ と最適解 $f(\mathbf{x}^*)$ との誤差が $1e-8$ 以下になった場合は、探索を打ち切った。また、試行回数は 25 回とした*2。

3.2 実験結果

本節では 2.1 節で説明した FDC を使用し、最大評価回数 $= 10^2 \times D, 10^4 \times D$ における IPOP-CMA-ES の evolvability の比較を行う。図 1 に、IPOP-CMA-ES における $f_{15}, f_{16}, f_{21}, f_{22}, f_{23}, f_{24}$ ($D = 10$) の FDC 図を示す。25 試行中最良の最良解を求めた試行のデータを示している。ここで、 $f_3, f_4, f_{17} \sim f_{20}$ については、 f_{15} と似た傾向を示したので紙面の都合上省略した。また、各試行の r_{FDC} を求め、その中央値を図中に表記している。

図 1(a), (e) から、 f_{15}, f_{23} では最大評価回数 $= 10^2 \times D, 10^4 \times D$ のどちらの場合も似たような景観を示しており、ほぼ同様の r_{FDC} 値を取ることがわかる。このことから、 f_{15}, f_{23} での EA の探索空間は、最大評価回数の設定の影響を受けないと言える。

一方、図 1(b), (c), (d), (f) から、 $f_{16}, f_{21}, f_{22}, f_{24}$ は f_{15}, f_{23} とは異なり、最大評価回数によっては FDC 図が大きく異なっている。大域的な構造が無く、複数の谷を持つ f_{21}, f_{22} では、最大評価回数 $= 10^4 \times D$ においてはこの特徴を IPOP-CMA-ES は捉えることができている。例えば、 f_{21} では 3 つの谷を明瞭に確認することができる。しかし、最大評価回数 $= 10^2 \times D$ においては限られた評価回数のために 1 つの谷しか探索することができず、 f_{21}, f_{22} が持つ大域的特徴を把握できていない。また、 r_{FDC} 値もそれぞれ 0.94, 0.96 と非常に高い値を取る。このことから、使用可能な評価回数が少ない場合、 f_{21}, f_{22} における EA の探索空間は single-funnel 関数と同様であると言える。2 つの谷を有する f_{24} においても同様の傾向が見られる。最大評価回数 $= 10^4 \times D$ においては f_{24} の 2 つの谷を確認することができるが、最大評価回数 $= 10^2 \times D$ では 1 つの谷しか見られず、その景観は single-funnel のように見える。また、最大評価回数 $= 10^2 \times D$ における r_{FDC} 値は 0.82 と、比較的高い値を示す。

*2 BBOB benchmarks では各試行ごとに異なる関数インスタンスを使用する。しかし、インスタンスを毎試行ごとに変えた場合、得られる結果含まれるノイズが大きくなり evolvability の解析が困難となる。これを防ぐため、全ての試行で各関数のインスタンス 1 のみを使用した。

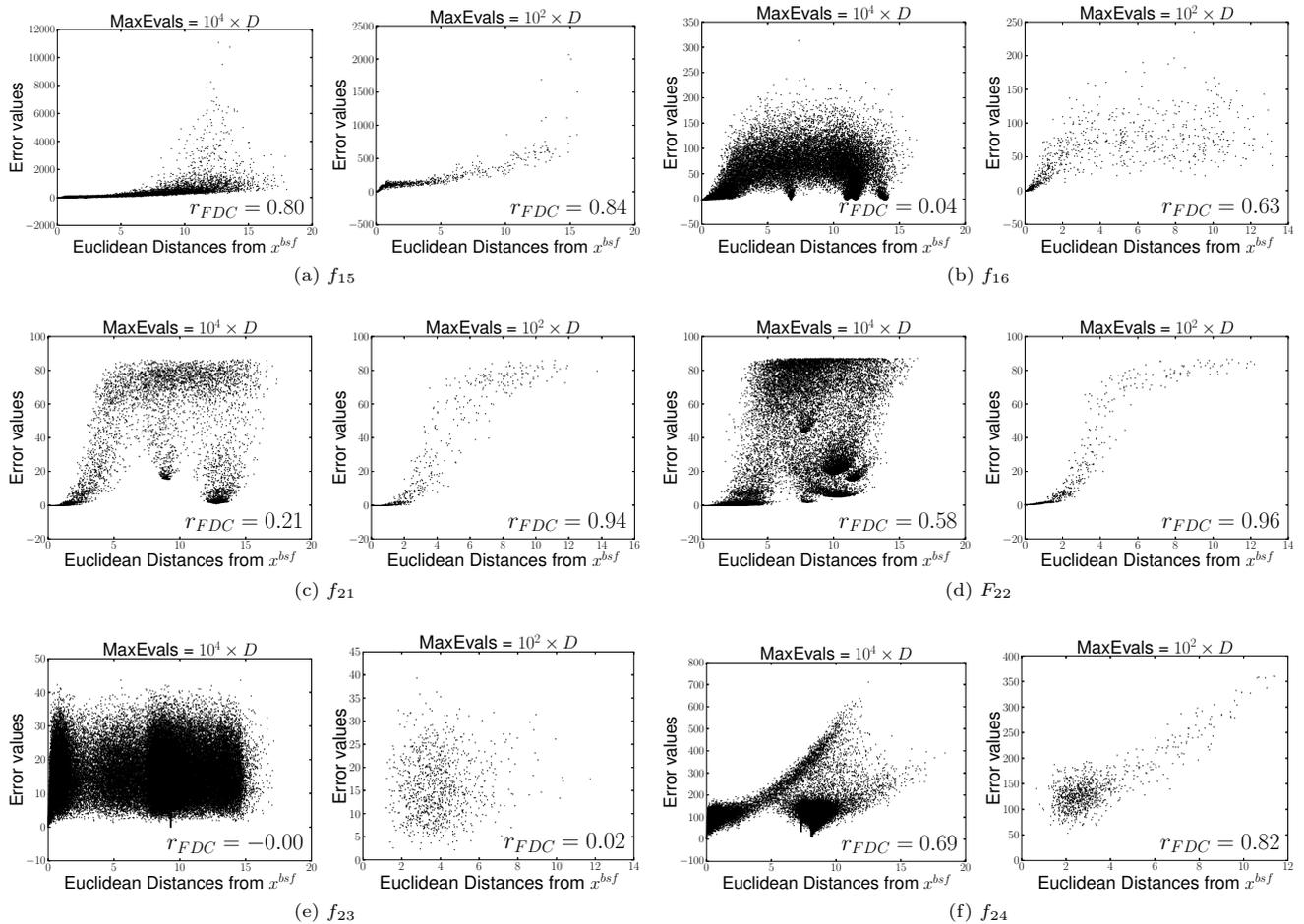


図 1: 図 (a) – (f) は, BBOB benchmarks [Hansen 09] の $f_{15}, f_{16}, f_{21}, f_{22}, f_{23}, f_{24}$ ($D = 10$) における FDC 図である. 25 試行において最も良い最良解を求めた試行における, IPOP-CMA-ES によって生成された解と対応する誤差値を示している. 横軸は \mathbf{x} と最良解 \mathbf{x}^{bsf} のユークリッド距離, 縦軸は \mathbf{x} の誤差値を示す. 左, 及び右の図はそれぞれ最大評価回数 $= 10^4 \times D$, 及び最大評価回数 $= 10^2 \times D$ における結果を表す. また, 図中の r_{FDC} 値は 25 試行における r_{FDC} 値の中央値である.

4. 考察

4.1 FDC に基づく evolvability の解析

本節では 3.2 節の実験結果について議論する. 実験では, 最大評価回数が $10^2 \times D$ と $10^4 \times D$ では関数によっては r_{FDC} 値が大きく異なる場合があった. 特に興味深いのが, 本来弱い大域的構造を有する $f_{16}, f_{21}, f_{22}, f_{24}$ が, 最大評価回数 $= 10^4 \times D$ においてこそ低い r_{FDC} 値を示すものの, 最大評価回数 $= 10^2 \times D$ では高い r_{FDC} 値を示していたことである. これは, 使用可能な評価回数が十分に与えられた場合は対象問題の構造を把握することが可能な一方, 限られた評価回数しか与えられない場合は探索領域を十分に探索することができなかつたため, 対象問題の大域的構造の影響を受けなかつたと考えられる. この考察は図 1 に示す, 各関数の FDC 図, 及び r_{FDC} 値からも明らかである.

García-Nieto と Alba は, FDC [Jones 95] を使用して, 最大評価回数 $= 10^4 \times D$ における PSO の性能調査をしている [García-Nieto 12]. 実験結果から, 例えば single/multi funnel を有する関数では r_{FDC} 値が高く/低くなるなど, 適切なパラメタ設定を使用した PSO アルゴリズムほど対象問題の大域的な景観を的確に捉えることができると結論づけている. 本論文

における最大評価回数 $= 10^4 \times D$ での IPOP-CMA-ES の実験結果も, これと同様の傾向を示している. しかし, 本論文の最大評価回数 $= 10^2 \times D$ における実験結果は, 先に述べた通り, 大きく異なる結果となった. これは, 使用可能な評価回数が乏しい場合は探索空間を広範囲に探索することが出来ず, 局所的な範囲しか探索することが出来なかつたためだと考えられる. そのため, FDC に基づく evolvability の解析結果は, 使用可能な評価回数の設定に大きく依存するといえる.

4.2 Expensive Optimization Problem のためのベンチマーク関数の必要性について

3.2 節にて示したように, IPOP-CMA-ES は最大評価回数 $= 10^2 \times D$ においては multi-funnel の影響を受けずに探索をしている場合が, 関数によっては見られた. この結果は, 現在のベンチマークセットでは expensive optimization problem に対する手法を正しく評価できない可能性を示唆している. 例えば, multi-funnel における EA の性能評価のために設計された multi-funnel 関数 [Lunacek 08] は, 潤沢に解評価が可能な場合は図 1 から明らかのように最適解を求めるために EA は複数の谷の影響を受けるが, 評価回数が限られる状況においてはこの限りでない. そのため, multi-funnel 関数における EA の性能評価をしているつもりだが, 実質的には単なる single-funnel

関数にて評価をしているのと変わらないということも起こり得る。その結果、手法 A と手法 B を少ない最大評価回数における single-funnel 関数と multi-funnel 関数にて評価をした場合、「手法 A は手法 B と比べ、対象問題の大域的な構造に依らずに良好な探索性能を示す」などと間違った結論を導く恐れがある。

より正しい性能評価のためには、次のような方針が考えられる:

1. f_{21}, f_{22}, f_{24} などの設計者の意図とは異なる evolvability が示される関数を、ベンチマークセットの構成関数から除外する。
2. expensive optimization problem 専用の関数を新たに設計する。

(1) については、設計者の意図とは異なる evolvability が示される関数をベンチマークセットから除外することで誤評価の可能性を軽減できる。これは誤評価を防ぐためには最も簡単な方法である。(2) については、関数生成器のパラメタ設定を調整することで、ある程度は実現可能である。例えば、 f_{24} は Lunacek らの multi-funnel function generator [Lunacek 08], f_{21}, f_{22} は Gallagher と Yuan の MSG function generator [Gallagher 06] を使用している。これらの生成器は問題の景観を制御するための複数のパラメタを有する。つまり、制御パラメタを適切に調整できれば、少ない最大評価回数においても multi-funnel の存在が EA の探索に反映されるようになると思われる。これを確かめることが今後の課題として残される。

5. まとめ

近年では expensive optimization problem に対する EA の性能評価のために、様々なベンチマークセットがいくつか提案されている。しかし、その多くが通常の最適化問題におけるベンチマーク関数を使用し、単に最大評価回数を数千以内に制限しているだけである。使用できる評価回数によって進化アルゴリズムが到達可能な探索領域は異なるため、実際に探索する領域が通常の最大評価回数の設定時とは大きく異なる可能性がある。そのため、これらのベンチマークセットが expensive optimization problem に対する EA の性能評価に適切なものかは不明である。

本論文では、異なる最大評価回数における single-funnel 関数と multi-funnel 関数の EA の探索空間を比較した。EA には IPOPOP-CMA-ES 用いて、BBOB benchmarks の多峰性関数を使用した。Evolvability の解析には Fitness Distance Correlation (FDC) [Jones 95] を用いた。実験結果から、 f_{21}, f_{22}, f_{24} のような multi-funnel 関数は、最大評価回数 $10^2 \times D$ と $10^4 \times D$ では EA の探索空間が大きく異なる場合があることがわかった。これらの関数では、最大評価回数 = $10^4 \times D$ においてこそ低い r_{FDC} 値を示すものの、最大評価回数 = $10^2 \times D$ においては高い r_{FDC} 値を示すことが、実験では見られた。これは、比較的多くの最大評価回数が与えられた場合は対象問題の構造を把握することができている一方、限られた評価回数しか与えられない場合は探索領域を十分に探索することができなかったため、対象問題の大域的構造の影響を受けなかったのだと考えられる。

本論文で得られた実験結果は、現在の expensive optimization problem 向けのベンチマークセットは、設計者の意図通りの大域的構造を最大評価回数が少ない場合は EA の探索空間に反映できないために、EA の性能を正しく評価できていない

可能性を示唆している。これを解決するために、今後の課題として 4.2 節で述べた適切な expensive optimization problem におけるベンチマークセットの設計が今後の課題として残される。

参考文献

- [Auger 05] Auger, A. and Hansen, N.: A restart CMA evolution strategy with increasing population size, in *IEEE CEC*, pp. 1769–1776 (2005)
- [Boese 94] Boese, K. D., Kahng, A. B., and Muddu, S.: A new adaptive multi-start technique for combinatorial global optimizations, *Oper. Res. Lett.*, Vol. 16, No. 2, pp. 101–113 (1994)
- [Gallagher 06] Gallagher, M. and Yuan, B.: A General-Purpose Tunable Landscape Generator, *IEEE Trans. Evol. Comput.*, Vol. 10, No. 5, pp. 590–603 (2006)
- [García-Nieto 12] García-Nieto, J. and Alba, E.: Why six informants is optimal in PSO, in *GECCO*, pp. 25–32 (2012)
- [Hansen 01] Hansen, N. and Ostermeier, A.: Completely Derandomized Self-Adaptation in Evolution Strategies, *Evol. Comput.*, Vol. 9, No. 2, pp. 159–195 (2001)
- [Hansen 09] Hansen, N., Finck, S., Ros, R., and Auger, A.: Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions, Technical Report RR-6829, INRIA (2009), Updated February 2010
- [Jin 05] Jin, Y.: A comprehensive survey of fitness approximation in evolutionary computation, *Soft Comput.*, Vol. 9, No. 1, pp. 3–12 (2005)
- [Jones 95] Jones, T. and Forrest, S.: Fitness Distance Correlation as a Measure of Problem Difficulty for Genetic Algorithms, in *ICGA*, pp. 184–192 (1995)
- [Liu 14] Liu, B., Chen, Q., Zhang, Q., J.Liang, J., Suganthan, P. N., and Qu, B. Y.: Problem Definitions and Evaluation Criteria for Computational Expensive Optimization, *Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China and Nanyang Technological University, Singapore, Tech. Rep* (2014)
- [Lunacek 08] Lunacek, M., Whitley, D., and Sutton, A.: The Impact of Global Structure on Search, in *PPSN*, pp. 498–507 (2008)
- [Mersmann 14] Mersmann, O., Preuss, M., Trautmann, H., Bischl, B., and Weihs, C.: Analyzing the bboob results by means of benchmarking concepts (2014), in press
- [Merz 04] Merz, P.: Advanced Fitness Landscape Analysis and the Performance of Memetic Algorithms, *Evolutionary Computation*, Vol. 12, No. 3, pp. 303–325 (2004)
- [Müller 11] Müller, C. L. and Sbalzarini, I. F.: Global Characterization of the CEC 2005 Fitness Landscapes Using Fitness-Distance Analysis, in *EvoApplications*, pp. 294–303 (2011)
- [Smith 02] Smith, T., Husbands, P., Layzell, P. J., and O’Shea, M.: Fitness Landscapes and Evolvability, *Evolutionary Computation*, Vol. 10, No. 1, pp. 1–34 (2002)
- [Suganthan 05] Suganthan, P. N., Hansen, N., Liang, J. J., Deb, K., Chen, Y. P., Auger, A., and Tiwari, S.: Problem Definitions and Evaluation Criteria for the CEC 2005 Special Session on Real-Parameter Optimization, Technical report, Nanyang Technological Univ. (2005)
- [Tanabe 15] Tanabe, R.: On the Evolvability for Multimodal Functions for Expensive Computational Budgets, in *GECCO (Companion)* (2015), accepted
- [Vérel 03] Vérel, S., Collard, P., and Clergue, M.: Where are bottlenecks in NK fitness landscapes?, in *IEEE CEC*, pp. 273–280 (2003)