D 1.4 Black-box complexity: limits of evolution

The SAGE Consortium

July 1, 2015

<table>
<thead>
<tr>
<th>Project acronym</th>
<th>SAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project title</td>
<td>Speed of Adaptation in Population Genetics and Evolutionary Computation</td>
</tr>
<tr>
<td>Grant agreement</td>
<td>618091</td>
</tr>
<tr>
<td>Call topic</td>
<td>Future and Emerging Technologies</td>
</tr>
<tr>
<td>Project duration</td>
<td>3 years</td>
</tr>
<tr>
<td>Work Package</td>
<td>WP1 Unified model of evolution</td>
</tr>
<tr>
<td>Deliverable</td>
<td>D1.4 Black-box complexity: limits of evolution</td>
</tr>
<tr>
<td>Deliverable due date</td>
<td>Month 18</td>
</tr>
<tr>
<td>Submission date</td>
<td>July 1, 2015</td>
</tr>
<tr>
<td>Dissemination level</td>
<td>PU</td>
</tr>
</tbody>
</table>
Black-box complexity: limits of evolution

1 Introduction

A black-box complexity theory for models of natural of evolution, showing the inherent limitations of speed of adaptation on given landscapes. This will serve as a baseline for future quantitative performance analyses for all evolutionary models, by itself and when comparing performance across different fitness landscapes, asexual vs. sexual reproduction, and structured populations. In the following sections, we describe published results related to black-box complexity and potential avenues for future work.

Publications

The deliverable lead to two publications in major computer science conferences.

   - Link to publicly available copy of paper
   - DOI link to publisher’s copy
     http://dx.doi.org/10.1007/978-3-319-10762-2_88

   - Link to publicly available copy of paper
   - DOI link to publisher’s copy
     http://dx.doi.org/10.1145/2725494.2725504

2 Unbiased Black-Box Complexity of Parallel Search

Background

Evolutionary algorithms are exploited to solve complex optimisation problems where information about the problem is limited and is regarded as black-box. The only way of getting knowledge about the problem is by evaluation potential solutions.

The black-box complexity of search algorithms describes the minimum number of function evaluations that every black-box search algorithm needs to make to optimise a given function. It hence captures limits to the efficiency of search algorithms.

The black-box complexity for evolutionary algorithms was introduced by Droste et al [8]. The only restriction imposed on the algorithm is the amount of information provided about the function to be optimise. The downside of having a general model with so little restriction is that the model becomes unrealistic and the runtime analysis is almost impractical. As a result, the more restricted black-box models were introduced, for instance, the ranking- based black-box models [13, 5], the memory-restricted black-box models [8, 6], unbiased black box [12].

A shortcoming of the above models is that they do not capture the implicit or explicit parallelism at the heart of many common search algorithms. Although, evaluating several individuals may decrease the number of generations algorithm requires to reach an optimal solution, however, this may lead to more function evaluations
since as evolution can only act on information from the previous generation. Existing black-box models are unable to capture this phenomena as they assume all search points being created in sequence.

**Our Contribution**

We provide a parallel black-box model that captures and quantifies the inertia caused by offspring populations of size $\lambda$ and parallel EAs evaluating $\lambda$ search points in parallel.

Unbiased black-box algorithms query new search points based on the past history of function values, using unbiased variation operators. We define a $\lambda$-parallel unbiased black-box algorithm in the same way, with the restriction that in each round $\lambda$ queries are made in parallel. These $\lambda$ queries only have access to the history of evaluations from previous rounds; they cannot access information from queries made in the same round.

We present lower bounds on the black-box complexity for the well known LEADINGONES problem defined by $\sum_{i=1}^{n} \prod_{j=1}^{i} x_j$ and for the general class of functions with a unique optimum, revealing how the number of function evaluations increases with the problem size $n$ and the degree of parallelism $\lambda$. The results complement existing upper bounds [11], allowing us to characterise the realm of linear speedups, where parallelisation is effective.

Our lower bound for functions with a unique optimum is asymptotically tight: we show that for the simple ONEMAX problem; $\text{ONEMAX}(x) := \sum_{i=1}^{n} x_i$, a $(1 + \lambda)$ EA with an adaptive mutation rate is an optimal parallel unbiased black-box algorithm. Adaptive mutation rates decrease the expected running time by a factor of $\ln \ln \lambda$, compared to the $(1 + \lambda)$ EA with the standard mutation rate of $1/n$, contrasting recent results by He, Chen, and Yao [10].

For details of our results, we attached the poster presentation of this paper 1.

3 Black-box Complexity of Parallel Search with Distributed Populations

**Background**

Distributed population is a population whose individuals are partitioned into groups where they evolve on their own, and they communicate or interact with other groups to coordinate their searches and to exchange promising solutions.

Distributed populations have been extensively exploited in evolutionary computation [9][11]. One benefit of using distributed populations is that they are suitable for implementations on parallel hardware such as distributed memory MIMD computers like a cluster or multicore systems. The other benefit is that distributed populations act like an implicit diversity-preserving mechanism. In contrast to a large panmictic population of the same size, good solutions are spread more slowly. This increases exploration of the search space at the expense of exploitation.

Many empirical studies report that parallel evolutionary algorithms with distributed populations find a better solution quality in a shorter amount of time.

**Our Contribution**

In this paper we introduce a distributed black-box model that additionally captures the way distributed populations communicate across topologies as found in many parallel metaheuristics. The distributed black-box complexity considers an ensemble of $\lambda$ black-box algorithms, called islands in reference to the island model, that exchange information about queried search points along a communication topology.

Here, we illustrate the common topologies: ring, torus, complete graph. Every node represents an island and the communication policy is shown by arrows if directed and by lines if the communication can happen both ways between two islands. The red nodes are the ones who have the best solution so far within the population.
The speed at which information is communicated depends on the communication topology and the communication policy, the strategy of how and when to communicate solutions. Denser topologies generally lead to a faster spread of information and to a faster exploitation. For function classes where exploitation is beneficial, we expect the distributed black-box complexity to be higher for sparse topologies and to decrease with the density of the topology. This intuition is supported by upper bounds on the running time of parallel evolutionary algorithms from [11]; these upper bounds become stronger for denser topologies. We confirm this intuition for unary unbiased black-box algorithms and the class of unimodal functions with a small number (θ(n)) of function values.

For a reasonable value of λ, the number of islands, our lower bound on the distributed black-box complexity of a ring topology is higher than that of torus graphs, and this, in turn, is higher than that of a complete graph, where communication is not restricted. While sparse communication topologies in practice often lead to better diversity and better exploration, our results suggest that they can inhibit exploitation and lead to worse performance on problems where exploitation is crucial.

Results in black-box complexity can provide insights into how given algorithmic characteristics [4, 7] impact performance. For example, it is known that the unbiased black-box complexity with higher-arity variation operators can be strictly lower than the unbiased black-box complexity with unary variation operators [4, 7]. Hence, allowing variation operators that combine more than one previously seen search point can lead to more efficient search. This insight about the impact of the arity of variation operators on performance is difficult to obtain without black-box complexity.

After presenting general analysis of communication topologies we study the random short path problem. Let π be a permutation of [n] chosen uniformly at random. Define a path $P = (p_0, p_1, \ldots, p_{n/2})$ such that $p_0 := 1^n$ and $p_i := p_{i-1} \oplus 0^\pi_{i-1} 10^{n-\pi_i}$. The function $RSP(x)$ is then defined as

$$RSP(x) := \begin{cases} n + i & \text{if } x = p_i \\ |x|_1 & \text{otherwise.} \end{cases}$$

The summary of results is presented here:

<table>
<thead>
<tr>
<th>Topology</th>
<th>Upper bounds</th>
<th>Lower bounds</th>
<th>Cut-off point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ring</td>
<td>$O(\lambda n^{7/3} + n^2)$</td>
<td>$\Omega(\lambda n + \lambda^{5/3} + n^2)$</td>
<td>$\lambda^* = \Theta(n^{1/2})$</td>
</tr>
<tr>
<td>Grid/Torus</td>
<td>$O(\lambda n^{11/4} + n^2)$</td>
<td>$\Omega(\lambda n + \lambda^{5/3}/n^{3/2} + n^2)$</td>
<td>$\lambda^* = \Theta(n^{2/3})$</td>
</tr>
<tr>
<td>Complete/Kλ</td>
<td>$O(\lambda n + n^2)$</td>
<td>$\Omega(\lambda n + n^{3/2})$</td>
<td>$\lambda^* = \Theta(n)$</td>
</tr>
</tbody>
</table>

Table 1: Asymptotic bounds on the λ-distributed unary unbiased black-box complexity for various topologies on RSP and, more generally, the class $U(\Theta(n))$ of all unimodal functions with Θ(n) function values. The bounds reflect the number of function evaluations (sequential time); the corresponding number of rounds (parallel time) is obtained by dividing the former by λ. The upper bound on grid/torus topologies requires a torus with side lengths $\sqrt{\lambda} \times \sqrt{\lambda}$; the lower bound holds for all side lengths. The cut-off point $\lambda^*$ determines the realm of possible linear speedups for each topology: below the cut-off the parallel (1+1) EA [11] has a linear speedup, while for $\lambda = \omega(\lambda^*)$ no distributed black-box algorithm can achieve a linear speedup.

Our results allow for fundamental conclusions about the impact of communication topologies on performance. Furthermore, for the considered problem class our results enable us to locate a cut-off point for the choice of λ above which the benefit of using distributed populations deteriorates and no linear speedup can be obtained by any distributed black-box algorithm. The amount of communication is restricted by the density of the topology.
of the communication topology. Hence, our results also give some insight into how performance degrades when communication between nodes becomes restricted.

There are several possible avenues for future work. One open question is how the choice of communication policy affects the black-box complexity. Another open problem is the relation between that amount of information communicated between islands and the distributed black-box complexity.
Bibliography


Motivation and Definitions

- **Aim**: A Complexity Theory for Search Heuristics
- inherent difficulty of a problem for every search heuristic
- complexity theory requires a model of search heuristics that captures their fundamental computational constraints

- **Black-Box Models**
  - access to function constrained to black-box access
  - black-box algorithms query function value of search points
  - unary unbiased black box model $\mu\text{BBC}$ (Lehre&Witt'12):
    - unary variation operators (one parent) with no search bias

- **Black-Box Complexity**
  - minimum worst-case number of function evaluations needed by every black-box algorithm
  - captures limits to the efficiency of all black-box algorithms, prevents algorithm designers from wasting effort on trying to achieve impossible performance.

Research Questions

- **Algorithms must fix $\lambda$ queries in advance**
  - Parallel black-box complexity (#evaluations) grows with $\lambda$:
    - $1$-$\mu\text{BBC}(F_n) \leq 2$-$\mu\text{BBC}(F_n) \leq 3$-$\mu\text{BBC}(F_n) \leq \ldots$

- **Questions**:
  - How does the $\lambda$-parallel black-box complexity scale with $\lambda, n$?
  - Does this growth apply to every $\lambda$-parallel black-box algorithm?
  - For what $\lambda$ does the runtime start to deteriorate?
  - How to design optimal $\lambda$-parallel algorithms?

Parallel Black-Box Complexity Model

$\lambda$-Parallel Unary Unbiased Black-Box Algorithms:

- $\lambda$-$\mu\text{BBC}(F_n) = \Theta(\frac{\log n}{\lambda}) + n \log n$
- $\lambda$-$\text{upBBC}(F_n) = \Theta(\frac{n}{\log(n)} + \frac{n}{\lambda})$
- $\lambda$-$\text{upBB}(F_n) = \Theta(\frac{n}{\log(n)} + \frac{n}{\lambda})$
- $\lambda$-$\text{upBB}(F_n) = \Theta(\frac{n}{\log(n)} + \frac{n}{\lambda})$

Cut-off point:

- $\lambda^* = \sup\{\lambda \mid \lambda$-$\text{upBB}(F_n) \leq O(\mu\text{BBC}(F_n))\}$

Speedups (w.r.t. #iterations) over best sequential algorithm:

- $\lambda \leq \lambda^*$: no asymptotic increase of #evaluations, linear speedups possible.
- $\lambda = \omega(\lambda^*)$: linear speedups impossible for every $\lambda$-parallel black-box algorithm.

Link to Parallel Times and Speedups

- $\lambda^* - \parallel \text{parallel time, number of iterations}$
- $\lambda^* - \parallel \text{parallel time, number of iterations}$

Bounds on the $\lambda$-Parallel Unary Unbiased Black-Box Complexity

<table>
<thead>
<tr>
<th>Definition</th>
<th>Leading Ones</th>
<th>OneMax</th>
<th>Functions with unique Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$-$\mu\text{BBC}$ lower bound</td>
<td>$\Omega(\frac{\log(n)}{\lambda}) + n^2$</td>
<td>$\Theta(\frac{\lambda}{\log(n)} + n \log n)$</td>
<td>$\Theta(\frac{\lambda}{\log(n)} + n \log n)$</td>
</tr>
<tr>
<td>$\lambda$-$\mu\text{BBC}$ lower bound</td>
<td>$\Omega(\frac{\log(n)}{\lambda}) + \frac{n^2}{\lambda}$</td>
<td>$\Theta(\frac{\lambda}{\log(n)} + n \log n)$</td>
<td>$\Theta(\frac{\lambda}{\log(n)} + n \log n)$</td>
</tr>
<tr>
<td>Cut off point $\lambda^*$</td>
<td>$n$</td>
<td>$\log(n)$</td>
<td>$\frac{n}{\log(n)} - \frac{n \log(n)}{\lambda}$</td>
</tr>
<tr>
<td>$\lambda$-$\text{upBB}$ upper bound</td>
<td>$O(\lambda n + n^2)$</td>
<td>adaptive $(1+\lambda)$ EA</td>
<td>$\Omega(\frac{\log(n)}{\lambda}) + \frac{n^2}{\lambda}$</td>
</tr>
<tr>
<td>Best known $\lambda$-parallel Algo</td>
<td>$(1+\lambda)$ EA</td>
<td>$\Theta(\frac{\lambda}{\log(n)} + n \log n)$</td>
<td>$\frac{n}{\log(n)} - \frac{n \log(n)}{\lambda}$</td>
</tr>
</tbody>
</table>

1. adaptive mutation rate, $p = \max\left\{\frac{2\lambda}{\lambda^2(n) + \lambda^2}, \frac{1}{2}\right\}$, $\lambda^2(n)$ the number of zeros

Conclusion:

- New black-box model for parallel search as in offspring populations, island models, cellular EAs, etc.
- Lower limits to how quickly one can optimise a function with parallel resources
- Linear speedups w.r.t. the number of rounds only possible up to cut-off point $\lambda^*$
- Led to design of optimal $\lambda$-parallel black-box algorithm for OneMax.

Figure 1: Poster of [2] presented at PPSN 2014.