Cost of Elitist Selection in Discrete Black-Box Optimization

Carola Doerr
CNRS and Université Pierre et Marie Curie, Paris, France

PGMO project “Towards a Complexity Theory for Black-Box Optimization”
(2014 --- 2016)

Anne Auger¹, Dimo Brockhoff², Benjamin Doerr³, Carola Doerr⁴, Nikolaus Hansen¹, Timo Koetzing⁵, Johannes Lengler⁶, Jonathan Rowe⁷

(¹INRIA Saclay, ²INRIA Lille, ³LIX, Ecole Polytechnique, ⁴CNRS and LIP6, UPMC, ⁵HPI Potsdam, ⁶ETH Zürich, ⁷U. of Birmingham)
Goals of the Talk

- Give you an idea of what our project is about
- At the end of the talk, you should have an idea
  - what black-box complexity is
  - how it relates to runtime analysis
  - that both ideas complement each other well
- about what kind of research questions we are working on in the project
Black-Box Optimization

- Black-Box Setting: Goal is to optimize a function $f: S \to \mathbb{R}$

  \[ x \in S \quad \text{and} \quad f(x) \]

- Learn about $f$ only through queries
- Performance measure: worst-case (among all $f \in F$) expected # of queries needed until an optimal search point is evaluated for the first time
- Motivation:
  - Large optimization problems with lots of data,
  - privacy issues,
  - ...

C. Doerr: Cost of Elitist Selection in Discrete Black-Box Optimization
Runtime Analysis vs. Black-Box Complexity

- Objective of black-box complexity: understand how certain algorithmic choices influence the performance of black-box optimizers
- Algorithmic choices:
  - Size of the memory
  - Usage of absolute ("$f(x) = 17$") or only relative ("$f(x) > f(y)$") fitness information
  - ...

How long does Algorithm A need (in expectation) to optimize problem P?

How long does it take the fastest algorithm (in expectation) to optimize problem P?
Example: Memory Restriction

- Motivation: many black-box optimizers do not store full search history
- Example: Evolutionary Algorithms

**Initialization of the population:**
Sample search points $X = \{x^1, \ldots, x^\mu\}$

**Variation:**
Create $\lambda$ offspring by mutating and recombining search points from $X$

**Selection:**
Update population $X$ by selecting $\mu$ individuals

Stop? → Output best search point(s) seen
Memory-Restricted Black-Box Model

- Suggested in [Droste/Jansen/Wegener ToCS 2006]

\[ f \in F \]

Algorithm A

1. (x,f(x))
2. (y,f(y))
... 
\( \mu \). (z,f(z))

- Research question in black-box complexity: how does the performance of memory-restricted algorithms compare to that of unrestricted ones?
The Hamming Distance Problem

- Unknown target vector \( z \in \{0,1\}^n \)
  \[
  z = 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0
  \]
- Goal is to learn \( z \)
- "Fitness" of a search point \( x \) is
  "closeness of \( x \) to \( z \)”, i.e., \( n - H(x, z) \)
  \[
  f_z(x) = 9 - 4 = 5
  \]
- Mastermind game from the 70s
The Hamming Distance Problem

- Unknown target vector \( z \in \{0,1\}^n \) \( z = 1 1 1 1 0 0 0 0 \)
- Goal is to learn \( z \)
- “Fitness” of a search point \( x \) is \( x = 0 1 0 1 1 0 1 0 1 \)
  “closeness of \( x \) to \( z \)”, i.e., \( n - H(x, z) \) \( f_z(x) = 9 - 4 = 5 \)
- How long do you need to solve this problem?
The Hamming Distance Problem

- Unknown target vector $z \in \{0,1\}^n$ \quad $z = 1 1 1 1 0 0 0 0$
- Goal is to learn $z$
- “Fitness” of a search point $x$ is \quad $x = 0 1 0 1 1 0 1 0 1$
  “closeness of $x$ to $z$”, i.e., $n - H(x, z)$ \quad $f_z(x) = 9 - 4 = 5$
- How long do you need to solve this problem?
- ...
- Hopefully, less than $n+1$ queries:
The Hamming Distance Problem

- Unknown target vector $z \in \{0,1\}^n \quad z = 1 1 1 1 0 0 0 0$
- Goal is to learn $z$
- “Fitness” of a search point $x$ is $\quad x = 0 1 0 1 1 0 1 0 1$
  - “closeness of $x$ to $z$”, i.e., $n - H(x, z) \quad f_z(x) = 9 - 4 = 5$
- How long do you need to solve this problem?
- ...
- Hopefully, less than $n+1$ queries
- [Erdős & Rényi, 1963]: one can do better: $O(n/\log n)$ strategy
- Their algorithm needs a lot of memory:
The Algorithm by Erdős & Rényi / Chvátal

\[
\begin{array}{cccccccc}
1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\
\hline
1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 4 \\
2 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 5 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
11\ldots1 & 1 & 4 & . & 4 & 5 \\
01\ldots0 & 6 & 5 & . & 5 & 2 \\
\end{array}
\]

\[
\frac{cn}{\log n} \quad 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 4 \\
\]

\[
00\ldots1 & 3 & 4 & . & 4 & 5 \\
\]

C. Doerr: Cost of Elitist Selection in Discrete Black-Box Optimization
The Hamming Distance Problem

- Unknown target vector $z \in \{0,1\}^n$ $z = 11110000$
- Goal is to learn $z$
- “Fitness” of a search point $x$ is $x = 010110101$
  “closeness of $x$ to $z$”, i.e., $n - H(x, z)$ $f_z(x) = 9 - 4 = 5$
- How long do you need to solve this problem?
- ...
- Hopefully, less than $n+1$ queries
- [Erdős & Rényi, 1963]: one can do better: $O(n/\log n)$ strategy
- Their algorithm needs a lot of memory
- Conjecture [Droste/Jansen/Wegener ToCS 2006]: when algorithm can only store one search point and its fitness, $\Omega(n \log n)$ queries are needed to solve this problem
- [Doerr/Winzen ToCS 2014]: falsified conjecture: $O(n/\log n)$ still possible
The Comparison-Based Black-Box Model

- Suggested in [Teytaud/Gelly PPSN 2006] and [Doerr/Winzen Algorithmica 2014]
- Motivation: many black-box optimizers do not use *absolute* but rather *relative* fitness values (e.g., elitist selection, truncation selection,...)

\[ f \in F \]

Algorithm A

\[ x \]

\[ y \]

\[ f(y) > f(x) \]

Black-Box = “Oracle”
The Comparison-Based Black-Box Model

- Suggested in [Teytaud/Gelly PPSN 2006] and [Doerr/Winzen Algorithmica 2014]
- Motivation: many black-box optimizers do not use absolute but rather relative fitness values (e.g., elitist selection, truncation selection,...)
- In the combined comparison-based and memory-restricted model (memory = 1 previous query), $\Theta(n)$ queries are needed to solve the Hamming distance problem
  - in this model, the simple $n+1$ strategy is asymptotically optimal!
  (implementing it in a way that uses only memory 1 is not so trivial)
The Elitist Black-Box Model

- Suggested in [Doerr/Lengler GECCO 2015]
- Motivation: (on top of being memory-restricted and comparison-based) many black-box optimizers continue search in the most promising regions ("greedy" behavior)

\[ f \in F \]

- How does this influence the performance?
Results for the Elitist Black-Box Model

- Hamming-distance problem:
  - algorithms that with high probability solve any instance in $O(n)$ queries
  - we do not know any algorithm that needs $O(n)$ queries in expectation
    (rather philosophical question as $O(n)$ can be achieved by using restarts)
- For many other problems, elitist selection can cause huge performance gaps, e.g., $O(n^2)$ for non-elitist strategies but $\Omega(2^n)$ for elitist ones
Our Project

- Analysis and development of black-box models
  → understand influence of algorithmic choices on performance
- Development of mathematical tools supporting the analysis
- Promotion of black-box complexity as research topic

- New project (2016-17): Parameter Optimization via Drift Analysis