



**HAL**  
open science

# A Block-Coordinate Descent EMO Algorithm: Theoretical and Empirical Analysis

Benjamin Doerr, Joshua Knowles, Aneta Neumann, Frank Neumann

► **To cite this version:**

Benjamin Doerr, Joshua Knowles, Aneta Neumann, Frank Neumann. A Block-Coordinate Descent EMO Algorithm: Theoretical and Empirical Analysis. GECCO 2024, 2024, Melbourne, Australia. 10.1145/3638529.3654169 . hal-04649090

**HAL Id: hal-04649090**

**<https://hal.science/hal-04649090>**

Submitted on 16 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License



# A Block-Coordinate Descent EMO Algorithm: Theoretical and Empirical Analysis

Benjamin Doerr

Laboratoire d'Informatique (LIX),  
CNRS, École Polytechnique  
Institut Polytechnique de Paris, Palaiseau, France

Joshua Knowles

Schlumberger Cambridge Research  
SLB  
Cambridgeshire, United Kingdom

Aneta Neumann

Optimisation and Logistics,  
School of Computer and Mathematical Sciences,  
The University of Adelaide, Adelaide, Australia

Frank Neumann

Optimisation and Logistics,  
School of Computer and Mathematical Sciences,  
The University of Adelaide, Adelaide, Australia

## ABSTRACT

We consider whether conditions exist under which block-coordinate descent is asymptotically efficient in evolutionary multi-objective optimization, addressing an open problem. Block-coordinate descent, where an optimization problem is decomposed into  $k$  blocks of decision variables and each of the blocks is optimized (with the others fixed) in a sequence, is a technique used in some large-scale optimization problems such as airline scheduling, however its use in multi-objective optimization is less studied. We propose a block-coordinate version of GSEMO and compare its running time to the standard GSEMO algorithm. Theoretical and empirical results on a bi-objective test function, a variant of LOTZ, serve to demonstrate the existence of cases where block-coordinate descent is faster. The result may yield wider insights into this class of algorithms.

## CCS CONCEPTS

• Theory of computation → Evolutionary algorithms.

## KEYWORDS

block coordinate decent, evolutionary multi-objective optimization, runtime analysis, theory

## ACM Reference Format:

Benjamin Doerr, Joshua Knowles, Aneta Neumann, and Frank Neumann. 2024. A Block-Coordinate Descent EMO Algorithm: Theoretical and Empirical Analysis. In *Genetic and Evolutionary Computation Conference (GECCO '24)*, July 14–18, 2024, Melbourne, VIC, Australia. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3638529.3654169>

## 1 INTRODUCTION

Evolutionary algorithms have been widely used to tackle multi-objective optimization problems [8, 9, 11]. This includes the use of popular algorithms such as NSGA-II [12], SMS-EMOA [3] and SPEA2 [44] for problems with two or three objectives, and approaches such as NSGA-III [13] and MOEA/D [40] for many-objective

problems. In terms of theoretical understanding, the area of rigorous runtime analysis [1, 21, 32, 34, 43] has provided a sound understanding of different algorithmic components [28] and problem formulations [24, 36], and more recently the overall performance of popular algorithms such as NSGA-II [7, 41], NSGA-III [35, 39], MOEA/D [15, 29, 30], and SMS-EMOA [4, 42].

In this paper, we explore the use of a technique called *block coordinate descent* for tackling large-scale problems in the context of evolutionary multi-objective optimization. Block-coordinate descent is a technique that has been shown to be particularly useful in the context of classical optimization when the problem can be structured according to different subproblems. While this technique allows for significant speed-ups and better overall results in many situations, it has not been explored until quite recently in the context of classical multi-objective programming [2, 10, 31, 38] or evolutionary multi-objective optimization [10].

We provide a set-up and demonstrate, through a class of example functions, where the block coordinate evolutionary multi-objective approach has a significant benefit over the standard approach. This benchmark function is a block-wise variant that incorporates the classical LEADINGONES problem and building on the single-objective benchmark function studied in [22, 25]. Our multi-objective problem consists of  $k$  blocks and has a Pareto front of size  $2^k$ . We examine the time until all such trade-offs have been obtained. The problem is structured in such a way that blocks with an index  $i$  contribute more to the two objective functions than the sum of all blocks with indices  $j > i$ . We show that the block-coordinate approach allows for parallel optimization of the different blocks. In contrast, a standard approach tends to destroy the beneficial structure in lower priority blocks during the run when achieving improvements in higher priority blocks.

In order to show this, we analyze the classical GSEMO algorithm and the block-coordinate variant BC-GSEMO introduced here on the benchmark functions described. We use rigorous runtime analysis and provide a lower bound for GSEMO and an upper bound for BC-GSEMO, which shows that BC-GSEMO is faster in optimizing the problem. The reason for this behavior is that GSEMO, which tackles the whole problem at once, tends to destroy components in different blocks that might later be optimized when the population size of GSEMO has increased significantly. Compared to this, BC-GSEMO does not encounter this problem and optimizes



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

GECCO '24, July 14–18, 2024, Melbourne, VIC, Australia

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0494-9/24/07.

<https://doi.org/10.1145/3638529.3654169>

the different blocks in parallel while not destroying the solution structure of other blocks.

We also confirm from an experimental perspective the usefulness of the block-coordinate approach in the context of evolutionary multi-objective optimization. In order to show that the theoretically observed behavior already occurs for relatively small problem sizes, we carry out experimental investigations for a small number of up to  $k = 4$  blocks. The experiments show a significantly different scaling behavior of GSEMO and BC-GSEMO for all settings investigated.

The paper is structured as follows. In Section 2, we present the basics and introduce our block-coordinate descent approach incorporated into the GSEMO algorithm. Section 3 introduces our class of benchmark problems and points out important structural properties. We carry out the runtime analysis of GSEMO and BC-GSEMO for the problem introduced in Section 4 and complement our theoretical findings with an experimental study in Section 5. Finally, we finish with some discussion and concluding remarks.

## 2 PRELIMINARIES AND BLOCK COORDINATE GSEMO

We consider the classical bi-objective task of maximizing two conflicting objective functions given the search space  $S = \{0, 1\}^n$ . The fitness of a search point  $x \in S$  is given as  $f = (f_1, f_2): S \rightarrow \mathbb{R}^2$ . A solution  $x$  (weakly) dominates a solution  $y$  (denoted as  $x \geq y$ ) iff  $f_i(x) \geq f_i(y)$ ,  $1 \leq i \leq 2$ . A solution  $x$  strongly dominates a solution  $y$  (denoted as  $x > y$ ) iff  $x \geq y$  and  $f_i(x) > f_i(y)$  for at least one  $i \in \{1, 2\}$  holds. A search point  $x$  is called Pareto optimal if there is no other search point  $y$  with  $y > x$ . The notion of dominance translates from the search points to their objective vectors. The set of all Pareto optimal search points  $X^* \subseteq 2^S$  is called the Pareto set and the set of corresponding objective vectors is called the Pareto front  $F$ . The classical goal in multi-objective optimization is to compute for each objective vector of  $q \in F$  a solution  $x \in X^*$  with  $f(x) = q$ .

The GSEMO algorithm (see Algorithm 1) is a simple evolutionary multi-objective algorithm that has been frequently studied in the area of runtime analysis [5, 14, 20, 23, 24, 26, 27]. It starts with a population consisting of a single solution chosen uniformly at random. In each iteration, a solution  $x$  is chosen uniformly at random from the population  $P$ , and an offspring  $y$  is produced from  $x$  by mutation. Mutation is usually standard bit mutation which flips each bit independently with probability  $1/n$ . The offspring  $y$  is accepted and added to the population if it is not strongly dominated by other search points in  $P$ . If  $y$  is added, then all other search points in  $P$  that are weakly dominated by  $y$  are removed from  $P$ .

### BC-GSEMO

Throughout the paper we consider a problem where we decompose the search points  $x = (x_{B_1}, \dots, x_{B_k})$  into  $k$  blocks  $x_{B_1}, \dots, x_{B_k}$ . We assume that each block consists of  $\ell$  bits and therefore have  $n = k \cdot \ell$ . The BC-GSEMO algorithm is similar to GSEMO. However, it applies mutation only to a chosen block. As each block has length  $\ell$ , mutation flips each bit in the chosen block with probability  $1/\ell$ . Blocks are mutated consecutively by applying the mutation operator to any given block  $t_{\text{epoch}}$  times before moving on to the next one (see inner repeat-loop). The parameter  $t_{\text{epoch}}$  is assuming to be a value

---

### Algorithm 1: Global simple evolutionary multi-objective optimizer (GSEMO)

---

```

1 Initialize  $x \in \{0, 1\}^n$  uniformly at random;
2  $P \leftarrow \{x\}$ ;
3 repeat
4   Choose  $x \in P$  uniformly at random;
5   Create  $y$  from  $x$  by mutation;
6   if  $\nexists w \in P : w > y$  then
7      $P \leftarrow (P \setminus \{z \in P \mid y \geq z\}) \cup \{y\}$ ;
8 until stop;
```

---



---

### Algorithm 2: Block coordinate GSEMO (BC-GSEMO)

---

```

1 Initialize  $x \in \{0, 1\}^n$  uniformly at random;
2  $P \leftarrow \{x\}$ ;
3  $i \leftarrow 0$ ;
4 repeat
5    $t \leftarrow 0$ ;
6   repeat
7      $t \leftarrow t + 1$ ;
8     Choose  $x \in P$  uniformly at random;
9     Create  $y$  from  $x$  by block-coordinate mutation on
       block  $x_{i+1}$ ;
10    if  $\nexists w \in P : w > y$  then
11       $P \leftarrow (P \setminus \{z \in P \mid y \geq z\}) \cup \{y\}$ ;
12    until  $t \geq t_{\text{epoch}}$ ;
13     $i \leftarrow (i + 1) \bmod k$ ;
14 until stop;
```

---

significantly smaller than the overall runtime which implies that each block is mutated roughly equally often. It should be noted that this scheme of applying mutation to each block a  $k$ -th fraction of the time and if so flipping each bit with probability  $1/\ell$  matches the mutation probability of  $1/n$  for GSEMO in terms of how often a particular bit is flipped.

When analyzing GSEMO and BC-GSEMO with respect to their runtime behavior, we measure time in terms of the expected number of fitness evaluations to reach a given goal. The expected optimization time refers to the expected number of fitness evaluations until the particular algorithm has obtained an optimal population, i.e. a population containing for each Pareto optimal objective vector a corresponding solution, for the first time.

## 3 BENCHMARK FUNCTION AND PROPERTIES

We now define the benchmark problem we shall regard in this work, and for which we shall show that a block-coordinate approach is superior to a classic multi-objective approach.

Let  $n \in \mathbb{N}$  and  $k$  be a divisor of  $n$ . Let  $\ell := n/k$ . For a bit-string  $x \in \{0, 1\}^n$ , we write  $x = (x_{B_1}, \dots, x_{B_k})$ , where  $x_{B_i} = (x_{(i-1)\ell+1}, \dots, x_{i\ell})$  is the  $i$ th block of  $x$ .

For all  $z \in \{0, 1\}^\ell$ , we define

$$LO_z(x_{B_i}) = \sum_{i=1}^{\ell} \prod_{j=1}^i (z_i = x_j)$$

as the number of leading positions in which  $x_{B_i}$  agrees with  $z$ . We note that this function is isomorphic to the classic LeadingOnes problem, first proposed in [37] and then intensively studied in the theory of evolutionary algorithms, see, e.g., [16]. We shall be interested in two particular target strings. Let  $r \leq \ell$ . Let  $z^1 = 1^\ell$  and  $z^2 = 1^{\ell-r} 0^r$ .

We are now ready to define the bi-objective problem  $f_{nkr}$  on the search space  $\{0, 1\}^n$ . To ease the notation, we suppress the indices  $n$ ,  $k$ , and  $r$  and simply write  $f$  whenever there is no risk for confusion. We define  $f = (f_1, f_2): \{0, 1\}^n \rightarrow \mathbb{R}^2$  via

$$f_1(x) = \sum_{b=1}^k \left( (\ell+1)^{2(k-b)+1} LO_{z^1}(x_{B_b}) + (\ell+1)^{2(k-b)} LO_{z^2}(x_{B_b}) \right),$$

$$f_2(x) = \sum_{b=1}^k \left( (\ell+1)^{2(k-b)+1} LO_{z^2}(x_{B_b}) + (\ell+1)^{2(k-b)} LO_{z^1}(x_{B_b}) \right).$$

To analyze these functions, let us define the *base functions*

$$f_1^B(x_B) = (\ell+1) LO_{z^1}(x_B) + LO_{z^2}(x_B),$$

$$f_2^B(x_B) = (\ell+1) LO_{z^2}(x_B) + LO_{z^1}(x_B),$$

for all  $x_B \in \{0, 1\}^\ell$ . Then for both objectives  $m \in [1..2]$ , we can write

$$f_m(x) = \sum_{b=1}^k (\ell+1)^{2(k-b)} f_m^b(x_{B_b}). \quad (1)$$

We collect some elementary properties of  $f$  and its ingredients. Right from the definition, we obtain the following observation on how different bitstrings compare under the  $f_m$ ,  $m = 1, 2$ .

LEMMA 1. *Let  $x_B, y_B \in \{0, 1\}^\ell$  with  $x_B \neq y_B$ . Let  $i \in [1..\ell]$  be the first bit in which  $x_B$  and  $y_B$  differ, and let  $(x_B)_i = 1$  (and thus  $(y_B)_i = 0$ ). Then the following is true.*

- (1) *If  $i < \ell - r + 1$ , then  $f_1^B(x_B) > f_1^B(y_B)$  and  $f_2^B(x_B) > f_2^B(y_B)$ .*
- (2) *If  $i = \ell - r + 1$ , then  $f_1^B(x_B) > f_1^B(y_B)$  and  $f_2^B(x_B) < f_2^B(y_B)$ .*
- (3) *If  $i > \ell - r + 1$ , then one of the following three cases applies.*
  - (a) *If  $(x_B)_{\ell-r+1} = \dots = (x_B)_{i-1} = 1$  (and similarly for  $y_B$ ), then  $f_1^B(x_B) > f_1^B(y_B)$  and  $f_2^B(x_B) > f_2^B(y_B)$ .*
  - (b) *If  $(x_B)_{\ell-r+1} = \dots = (x_B)_{i-1} = 0$  (and similarly for  $y_B$ ), then  $f_1^B(x_B) < f_1^B(y_B)$  and  $f_2^B(x_B) < f_2^B(y_B)$ .*
  - (c) *Otherwise,  $f_1^B(x_B) = f_1^B(y_B)$  and  $f_2^B(x_B) = f_2^B(y_B)$ .*

*In particular, if  $x_B$  and  $y_B$  have equal value under one objective function, then also under the other.*

The exponentially decreasing coefficients of the  $f_m^b$  in (1) allow to extend, in a suitable way, the previous lemma to the full function  $f_m$ .

LEMMA 2. *Let  $m \in \{1, 2\}$  and  $\bar{m} = 3 - m$  be the other objective. Let  $x, y \in \{0, 1\}^n$  with  $f_m(x) \neq f_m(y)$ . Let  $b$  be minimal such that  $f_m^B(x_{B_b}) \neq f_m^B(y_{B_b})$ . Then*

- (1)  $f_{\bar{m}}^B(x_{B_b}) \neq f_{\bar{m}}^B(y_{B_b})$ ;
- (2)  $b$  is minimal subject to  $f_{\bar{m}}^B(x_{B_b}) \neq f_{\bar{m}}^B(y_{B_b})$ ;

- (3) *the function values of  $x$  and  $y$  compare like the base function values of  $x_{B_b}$  and  $y_{B_b}$ , that is, for all  $m' \in \{1, 2\}$  we have*
  - $f_{m'}(x) < f_{m'}(y)$  if and only if  $f_{m'}^B(x_{B_b}) < f_{m'}^B(y_{B_b})$ , and
  - $f_{m'}(x) > f_{m'}(y)$  if and only if  $f_{m'}^B(x_{B_b}) > f_{m'}^B(y_{B_b})$ .

PROOF. The first two claims follow immediately from the last clause of Lemma 1. For the third claim, note that for any  $z \in \{0, 1\}^n$  the contribution of the all blocks  $B_c$ ,  $c > b$ , to  $f_{m'}$  is at most

$$\sum_{c=b+1}^k (\ell+1)^{2(k-c)} f_{m'}^B(z_{B_c}) \leq \sum_{c=b+1}^k (\ell+1)^{2(k-c)} ((\ell+1)\ell + \ell)$$

$$= \ell \sum_{i=0}^{2(k-b-1)+1} (\ell+1)^i < (\ell+1)^{2(k-b)}.$$

In contrast, any difference between  $f_{m'}^B(x_{B_b})$  and  $f_{m'}^B(y_{B_b})$  immediately leads to a contribution to  $f_{m'}$  that differs by at least  $2^{2(k-b)}$ , thus outnumbering any possibly different contributions of the later blocks. By assumption and the first two claims, the earlier blocks contribute the same to  $f_{m'}$ . This proves the third claim.  $\square$

LEMMA 3. *Let  $P \subseteq \{0, 1\}^n$  be a set of pair-wise non-dominating solutions. Then  $|P| \leq 2^k$ .*

PROOF. By induction, we prove, for all  $i \in [0..k]$ , the following assertion.

- (A<sub>i</sub>) *Let  $P \subseteq \{0, 1\}^n$  be any set of pair-wise non-dominating solution such that for all  $b \in [1..k-i]$ , we have that  $f_1^B(x_{B_b}) = f_1^B(y_{B_b})$  for all  $x, y \in P$ . Then  $|P| \leq 2^i$ .*

We note that  $A_k$  is exactly the claim of the lemma, hence proving (A<sub>i</sub>) for all  $i \in [0..k]$  also proves the lemma.

Clearly, we have (A<sub>0</sub>). In this case, the assumptions of (A<sub>i</sub>) imply that all members of  $P$  have the same  $f_1$  value. Since  $P$  is pair-wise non-dominating, this implies  $|P| = 1 = 2^0$ .

Hence assume that for some  $i \in [1..k]$  we have shown already (A<sub>i-1</sub>). We use this to show (A<sub>i</sub>). We regard the set  $V = \{f_1^B(x_{B_{k-i}}) \mid x \in P\}$  of different objective values in the  $(k-i)$ -th block. If  $|V| = 1$ , then  $P$  satisfies the assumptions of (A<sub>i-1</sub>) and we immediately conclude  $|P| \leq 2^{i-1} \leq 2^i$ . Similarly, if  $|V| = 2$ , say  $V = \{v_1, v_2\}$ , then the sets  $P_j = \{x \in P \mid f_1^B(x_{B_{k-i}}) = v_j\}$ ,  $j = 1, 2$ , satisfy the assumptions of (A<sub>i-1</sub>), we conclude  $|P_j| \leq 2^{i-1}$ , and thus  $|P| = |P_1| + |P_2| \leq 2^i$ .

We conclude the proof by showing that  $|V| \geq 3$  contradicts our assumption that  $P$  is pair-wise non-dominating. Hence assume  $|V| \geq 3$ . Assume first that there is an  $x \in P$  such that  $LO(x_{B_{k-i}}) < \ell - k$ . Then all other  $y \in P$  satisfy  $LO(y_{B_{k-i}}) = LO(x_{B_{k-i}})$  as otherwise the first case of Lemma 1 would apply and show a contradiction to our assumption that  $P$  is pairwise non-dominating. However, if all  $y \in P$  have the same value  $LO(y_{B_{k-i}})$  that is less than  $\ell - k$ , then by definition of  $f_1^B$  all these  $y$  have the same  $f_1^B$ -value, contradicting our assumption  $|V| \geq 3$ . Consequently, we have  $LO(x_{B_{k-i}}) \geq \ell - k$  for all  $x \in P$ . Since  $|V| \geq 3$ , by the pigeon-hole principle, we now have two different values  $u, v \in V$  and  $x, y \in P$  such that  $f_1^B(x_{B_{k-i}}) = u$ ,  $f_1^B(y_{B_{k-i}}) = v$ , and  $(x_{B_{k-i}})_{\ell-r+1} = (y_{B_{k-i}})_{\ell-r+1}$ . Since  $f_1^B(x_{B_{k-i}}) = u \neq v = f_1^B(y_{B_{k-i}})$ , the bit strings  $x_{B_{k-i}}$  and  $y_{B_{k-i}}$  differ in some position higher than  $\ell - r + 1$ . We are thus in the

third case of Lemma 1. All three subcases of this case lead to a situation in which  $(f_1^B(x_{B_{k-i}}, f_2^B(x_{B_{k-i}}))$  and  $(f_1^B(y_{B_{k-i}}, f_2^B(y_{B_{k-i}}))$  are not non-dominating. By Lemma 2, this property extends to  $x$  and  $y$ , contradicting our assumption on  $P$ . This terminates the proof.  $\square$

## 4 RUNTIME ANALYSIS

In this section, we prove our mathematical runtime results. We start with a lower bound for the classic GSEMO and continue with a (smaller) upper bound for the BC-GSEMO.

### 4.1 Lower Bound for the GSEMO

We prove the following lower bound on the runtime of the GSEMO.

**THEOREM 1.** *Let  $n \in \mathbb{N}$  and  $k = o(\sqrt{n/\ln(n)})$  such that  $k$  divides  $n$ . Write  $\ell = n/k$ . Let  $r = o(\ell)$  and  $r \geq \max\{8 + 2k/\log_2(n), \gamma \ln(n)\}$  for some sufficiently large constant  $\gamma$ . Then the expected optimization time of GSEMO on  $f_{nr}$  is  $\Omega(2^{k n \ell})$ .*

As for many lower bounds proofs, the proof of this result is quite technical. Our main argument shall be that at some time, we have a population consisting of  $\Omega(2^k)$  individuals such that least one of them misses at a constant fraction of ones in the cooperative segment of the last block. Finding these  $\Omega(\ell)$  ones takes  $\Omega(2^{k n \ell})$  time, since a particular bit is flipped with probability  $1/n$  and this this particular parent is chosen for mutation only with probability  $O(2^{-k})$  (uniform choice among the  $\Omega(2^k)$  individuals). The main work is showing that such an individual exists, which means that the later bits are not optimized to a large extend while the earlier bits are optimized. Also, we need to take care that not too many bits are optimized in one go by simply generating such an offspring from a different, already more optimized parent.

In the following, we shall always assume that the assumptions of Theorem 1 hold. Being an asymptotic statement, we can and will always assume that  $n$  is sufficiently large. We use the elementary notation from the combinatorics of words, e.g., that  $1^i$  denotes the bit string  $(1, \dots, 1) \in \{0, 1\}^i$ , or that for  $x \in \{0, 1\}^i$  and  $y \in \{0, 1\}^j$  the expression  $xy$  denotes the bit string  $(x_1, \dots, x_i, y_1, \dots, y_j) \in \{0, 1\}^{i+j}$ . We say that some individual  $x \in \{0, 1\}^n$  is of *type*  $t \in \{0, 1\}^i$ ,  $i \leq n$ , to express that  $x$  agrees with  $t$  on its first  $i$  elements, that is,  $(x_1, \dots, x_i) = t$ .

We say that a block  $x_B$  of some individual  $x$  is *regular* if it has at most  $d = \lceil 0.7\ell \rceil$  leading ones and all bits with index higher than  $d + 1$  in this block are independently distributed with probability of being one at most  $q = 0.85$ .

We note that our following proofs use some of the main ideas of the proof of the result that the  $(1 + 1)$  EA optimizes a weighted sum of LEADINGONES blocks in a sequential fashion [22]. Unfortunately, due to the additional challenges imposed by the multi-objective setting, we cannot directly reuse the proofs from there. Differently from there, we need to argue that all later blocks are not optimized, as this ensures that only a single individual of a certain type is in the population; in [22], it was only argued that the next block is not optimized. Also, having a non-trivial population size, we need to take care that a particular individual does not profit from being overwritten by an offspring from a different parent. For this reason, we have to argue that different individuals, should they exist, are

sufficiently different so that it is unlikely that a mutation changes them enough to overwrite another individual. Such problems, naturally, do not appear in the  $(1 + 1)$ -setting regarded in [22]. For these reasons, while reusing some broader lines of the proof, we have essentially to come up with a new proof. On the positive side, our proof arguments could easily be used in the setting of [22], and would there give a proof maybe more accessible than the one given in [22].

**LEMMA 4.** *Consider a run of the GSEMO on our benchmark  $f_{nr}$  with parameters  $n, k$ , and  $r$  as in Theorem 1. Assume that for some iteration  $t_0$  there is a  $b \in [0..k - 1]$  such that the following holds.*

- An individual  $x$  with type  $1^{b\ell}$  has just entered the population.
- All blocks higher than  $b$  are regular.
- If  $b \geq 1$ , then for all  $i \leq b$  an individual of type  $1^{i\ell-r}0^{r/2}$  is contained in the population.

Let  $t_1$  be the first iteration in which an individual  $y$  of type  $1^{(b+1)\ell}$  is generated (and thus enters the population). Then with probability at least  $1 - O(kn^{-2})$ , the following properties hold.

- All blocks higher than  $b + 1$  are regular.
- For all  $i \leq b + 1$  an individual of type  $1^{i\ell-r}0^{r/2}$  is contained in the population.

**PROOF.** We first argue that with probability at least  $1 - O(n^{-2})$ , no parent different from an individual of type  $1^{b\ell}$  will generate an offspring that of this type in the next  $n^2 2^k$  iterations. To this aim, we note that the property we initially assumed, that individuals of type  $1^{i\ell-r}0^{r/2}$  are present, remains true for the remaining run of the algorithm, simply because such individuals can only be replaced by others of this type. From this, we conclude that any individual in the population that is not of type  $1^{b\ell}$  differs from any such individual in at least  $r/2$  of the first  $b\ell$  bits. Consequently, the probability that such a parent generates a  $1^{b\ell}$  offspring in one iteration, is at most  $n^{-r/2} \leq n^{-4} 2^k$  by our assumption on  $r$ . Hence the probability that this happens in the next  $n^2 2^k$  iterations, is at most  $n^{-2}$ . Since, as we will see shortly, the time interval analyzed in this lemma spans at most  $O(\ell n)$  iterations with  $x$  as parent (with probability  $1 - O(n^{-2})$ ), hence  $O(\ell n 2^k)$  with any parent with this probability, we can assume in the remainder that never an individual of type  $1^{b\ell}$  is generated from a parent of a different type.

We start by analyzing the time interval until for the first time an individual with type  $1^{(b+1)\ell-r}$  is in the population. We first analyze a fake version of our algorithm which never adds a second individual of type  $1^{b\ell}$  to the population (hence when such an individual, incomparable with the existing one, is generated, it is just discarded instead of added to the population). We shall later see that this fake algorithm with high probability behaves identical to the true algorithm, so our insights extend to the true algorithm.

In the fake algorithm, up to the point where for the first time an individual of type  $1^{(b+1)\ell-r}$  enters the population,  $x$  is the only individual of type  $1^{b\ell}$  in the population and stays in the population unless replaced by its own offspring. Hence we can analyze the development of  $x$  independent from all other individuals. To ease the language, we shall always denote the individual of type  $1^{b\ell}$  by  $x$ . Let us also count time starting at  $t = 0$  and let us count only iterations in which  $x$  is chosen as parent and in which no leading

one of  $x$  is flipped (as these will surely give an offspring that is not accepted).

We first note that each (such) iteration has a probability of exactly  $\frac{1}{n}$  of increasing the number of leading ones in  $x$ . Hence the expected time to finish this phase is at most  $n(\ell - r) \leq n\ell$  and, by a simple Chernoff bound, this phase ends after at most  $2n\ell$  iterations with probability at least  $1 - \exp(-\Theta(-\ell))$ . Let us assume this in the remainder.

Let us now regard a fixed block  $x_B$  higher than  $b$ . We first note that with probability at least  $(1 - \frac{1}{n})^n = (1 + o(1))\frac{1}{e}$ , the offspring of  $x$  is accepted. Hence with probability at least  $1 - 2n\ell \exp(-\Omega(n)) = 1 - \exp(-\Omega(n))$  in each interval of  $n$  consecutive iterations in this phase, there are at least  $\frac{n}{2e}$  iterations accepting the offspring. Let us also assume this.

Let us partition this phase into consecutive intervals  $I_1, I_2, \dots$  of exactly  $2n$  iterations (plus some leftover of at most  $2n - 1$  iterations). Note that  $L \leq \ell$  by the assumptions made above. With probability at least

$$\begin{aligned} & (1 - (1 - \frac{1}{n})^n)(1 - (1 - \frac{1}{n})^{\ell/2})(1 - \frac{1}{n})^n \\ & \geq (1 - \frac{1}{e})\frac{\ell}{2en}(1 - o(1))\frac{1}{e} \\ & = (1 - o(1))\frac{1}{2e^2}(1 - \frac{1}{e})\frac{\ell}{n}, \end{aligned}$$

in one of the first  $n$  iterations of this interval, let us call it  $t_u$ ,

- (1) we flip the first zero of  $x$  (implying that the offspring is accepted),
- (2) we flip at least one of the first  $\ell/2$  bits of  $x_B$  should they all be one, or we leave one such bit at zero should such a bit exist; let us call the first zero bit in this block after this iteration  $x_u$ ,
- (3) and we do not flip  $x_u$  in the next  $n$  iterations (which still belong to this interval of  $2n$  iterations).

Let us call this event a *reset*, and take note that we have just shown that in each time interval  $I_j$ ,  $j = 1, \dots$ , at most  $\ell$  in number, independently (by definition of the reset), with probability at least  $(1 - o(1))\frac{1}{2e^2}(1 - \frac{1}{e})\frac{\ell}{n}$  a reset happens. Consequently, when  $C$  is a sufficiently large constant, with probability at least  $1 - n^{-2}$  in each consecutive set of  $C\frac{n}{\ell} \log n$  intervals, a reset happens. In terms of iterations, this means that with this probability, in each interval  $[iL + 1..(i+1)L]$  of iterations,  $L := 2C\frac{n^2}{\ell} \ln(n)$ , a reset happens. We study the effect of a single reset and then how these frequent resets influence  $x_B$ .

In case of a reset, all bits higher than  $x_u$  in this block are neutral in iterations  $t_u + 1, \dots, t_u + n$ . Since at least  $\frac{n}{2e}$  of these iterations accept the mutant by our earlier assumption, each of these neutral bits for at least  $\frac{n}{2e}$  times is flipped independently with probability  $\frac{1}{n}$ . By Lemma 3 of [33], this results in each of these bits, independently, being one with probability at most  $\frac{1}{2} + \frac{1}{2}(1 - \frac{2}{n})^{n/2e} \leq \frac{1}{2}(1 + \exp(-1/e)) \leq 0.85 = q$ .

Let us call a block  $R$ -regular if it contains a zero in the first  $R$  bits and if all bits higher than  $R + 1$  are one with probability at most  $q$ , independently. Hence in the notation above, a reset at time  $t_u$  creates a  $u$ -regular block at time  $t_u + n$ , and this is also an  $\frac{\ell}{2}$ -regular block.

We have shown above that (i) after the start of the phase and (ii) after each reset we have at most  $2L$  iterations without a reset

(note that this argument is valid including the last few leftover iterations not fitting into a full interval  $I_j$ ). Consider for the moment any interval of  $2L$  iterations (independent from the reset logic just set up). In each of these iterations, the probability that the number of leading ones in this block increases, is at most  $\frac{1}{n}$ . Hence we expect at most  $2\frac{L}{n} = 4C\frac{n}{\ell} \ln(n)$  such increases. Since bits above the first zero are one with probability at most  $q$ , the number of leading ones gained in such an improvement is  $1 + \text{Geom}(q)$ , where  $\text{Geom}(q)$  denotes a random variable, independent from all others, following a geometric law with success rate  $q$ . Consequently, the expected total gain of the number of leading ones in all these improvements is at most  $(1 + \frac{1}{q})4C\frac{n}{\ell} \ln(n)$  and a Chernoff bound for geometric random variables gives that with probability at least  $1 - \exp(-\Omega(\frac{n}{\ell} \ln(n)))$ , the total gains is bounded by  $(1 + \frac{1}{q})8C\frac{n}{\ell} \ln(n) = O(\frac{n}{\ell} \log n)$ . By assumption, this gain is  $o(\ell)$ , which shows that up to the first reset the block is  $r$ -regular for some  $r = (0.7 + o(1))\ell$ , from then on it is  $r$ -regular for some  $r = (0.5 + o(1))\ell$ . By taking a union bound over the at most  $k$  later block, we extend this insight from one block to all later blocks, with probability at least  $1 - O(kn^{-2})$ .

From what we have just shown, in particular, we observe that we never see that the number of leading ones in a later block reaches  $\ell - r$ , which would be necessary to generate a second individual of type  $1^{b\ell}$  in the population. Hence with probability at least  $1 - O(kn^{-2})$ , also in the true algorithm we never see a second individual of this type, and thus our results above extend to the true algorithm.

We finally analyze the time until we have developed the individual of type  $1^{(b+1)\ell-r}$  into one of type  $1^{(b+1)\ell}$ . We sketch this part as it mostly reuses arguments seen before. The difference is that now we will have two individuals of type  $1^{(b+1)\ell-r}$ , one collecting ones and the other collecting zeroes in the last  $r$  bit positions. Both develop at a comparable speed, increasing the number of correct bits by an expected value of  $\Theta(1/n)$  when counting only the iterations regarding these two individuals. Also generating the second from the first has a probability of  $\Theta(1/n)$ . Since  $r \geq \gamma \ln n$  for some sufficiently large constant  $\gamma$ , the probability that one individual is significantly faster than the other is  $O(n^{-2})$ . Since this phase is short,  $O(nr)$  iterations with probability  $1 - O(n^{-2})$ , also the later blocks with high probability do not gain more than  $o(\ell)$  leading ones.

This proves the claim.  $\square$

**PROOF OF THEOREM 1.** Consider a run of the GSEMO on our benchmark. Trivially, at the start of the run the conditions of Lemma 4 are satisfied for  $b = 0$ . Also, if at some time  $t_0$  the conditions of this lemma are satisfied for some  $b \leq k - 1$ , then the lemma asserts that at some later time  $t_1$ , they are satisfied for  $b + 1$ , apart from a failure probability of  $O(k/n^2)$ . Hence by induction, using a union bound for the failure probabilities, we see that with probability at least  $1 - O(k^2/n^2)$ , at the moment  $t$  when an individual  $y$  of type  $1^{(k-1)\ell}$  enters the population, this individual has the last  $0.3\ell - 1$  bits independently distributed with a probability of at most  $q = 0.85$  of being one. Also, at this time for each  $i \in [1..k - 1]$  the population contains an individual of type  $1^{i\ell-r}0^{r/2}$ . As in the proof of Lemma 4, this implies that the population contains no individual  $z \neq y$  such that  $z$  and  $y$  differ in fewer than  $r/2$  of their first  $n - \ell$  bits.

Let  $S$  be the set of types of length  $(k-1)\ell$  that can be extended to a Pareto optimal individual, that is,

$$S = \{x \in \{0, 1\}^{(k-1)\ell} \mid \forall b \in [1..k-1] : \\ x_{(b-1)\ell+1} = \dots = x_{b\ell-r} = 1 \wedge x_{b\ell-r+1} = \dots = x_{b\ell}\}.$$

We note that the GSEMO treats the different types in  $S$  in a symmetric fashion (see [18] for a formal treatment of symmetry arguments in the analysis of evolutionary algorithms) and that the types appear one after the other in the population (since only one offspring is generated per iteration). Consequently, the position of type  $1^{(k-1)\ell}$  in the order of appearance of all types in  $S$  is uniformly distributed in  $[1..|S|]$ . It thus appears, with probability at least  $1/2$ , on position  $(|S|+1)/2$  or higher. Hence with probability at least  $1/2$ , the population size from this time on is at least  $(|S|+1)/2 \geq 2^{k-2} =: m$ .

Let us condition on this event. We analyze how the individual  $y$  with type  $1^{(k-1)\ell}$  develops into a solution of type  $1^{n-r}$ . We note first that the probability  $Q$  that in one iteration a parent with type different from  $1^{(k-1)\ell}$  generates an offspring of type  $1^{(k-1)\ell}$  is at most  $Q \leq n^{-r/2}$ . Hence the probability that this happens in the next  $T_0 := 0.05\ell mn$  iterations is at most  $QT_0 \leq n^{-r/2} \cdot 0.05\ell 2^k n \leq 2^k n^{2-r/2}$ . Since  $r \geq 8 + 2k/\log_2(n)$ , this is at most  $n^{-2}$ .

Ignoring such events for the moment, we see that the only way towards having a solution of type  $1^{n-r}$  in the population is to select the single individual of type  $1^{(k-1)\ell}$  as parent and mutate it. Apart from the waiting time for selecting the parent, this process up to reaching the type  $1^{k\ell-r}$  is identical to the optimization of the LEADINGONES benchmark, starting with an individuals with  $(k-1)\ell$  leading ones, then  $0.7\ell + 1$  bits which we do not specify further, and then  $0.3\ell - 1$  bits which are, independently, one with probability  $q$  only. We shall not regard for our lower bound the time it takes to optimize the first  $0.7\ell + 1$  bits in the last block. From that point on, the bits are one independently with some probability that is at most  $q$ . Hence the  $T$  time to reach an individual of type  $1^{k\ell-r}$  stochastically dominates the sum of  $0.3\ell - 1 - r$  independent random variables, each being a geometric distribution with success rate  $\frac{1}{mn}$  (with probability  $1 - q$ ) or being constant zero (with the remaining probability of  $q$ ), see [16] for more details on this argument). Hence  $T$  has an expectation of at least  $(0.3\ell - 1 - r)(1 - q)pn = (0.045 - o(1))\ell pn$  and with high probability (use a Chernoff bound for sums of independent geometric random variables, e.g., Theorem 1.10.32 in [17]),  $T$  is at least  $0.002\ell pn$ .

In summary, we see that with probability at least  $1/2 - o(1)$ , and this includes all small failure probabilities encountered along the proof, we need at least  $T_0 = 0.002\ell pn = \Omega(2^k \ell n)$  iterations to generate a solution with type  $1^{n-r}$ , and hence also to generate the Pareto-optimal solution  $1^n$ , which is needed to witness the Pareto front. This implies an  $\Omega(2^k \ell n)$  bound on the expected runtime.  $\square$

## 4.2 Upper Bound for the BC-GSEMO

We now prove an upper bound on the runtime of the BC-GSEMO algorithm which is lower than the lower bound shown for GSEMO in the previous section. The fraction of mutation steps that BC-GSEMO spends on a particular block  $b$  is roughly  $1/k$  if the parameter  $t_{\text{epoch}}$  is small compared to the overall runtime. We assume that the maximal number of consecutive steps  $t_{\text{epoch}}$  that a block is chosen for mutation is small enough such that each block is mutated

frequently enough during the phases considered in the upcoming proofs. For example, any  $t_{\text{epoch}} \leq \sqrt{n}$  meets our requirement. We analyze the algorithm using different phases of  $T$  steps. When doing so, we always assume without loss of generality that  $T/(k \cdot t_{\text{epoch}})$  is a positive integer in order to ease the presentation. This implies that mutation is applied to each block exactly  $T/k$  times.

We first show that BC-GSEMO quickly obtains for each block a sufficiently large number of leading ones. The key argument is that blocks are optimized in parallel and during this process and that BC-GSEMO works with a population size of 1.

**LEMMA 5.** *Let  $r < \sqrt{\ell \log \ell}$ . Then with probability  $1 - o(1)$ , BC-GSEMO has obtained after  $cn(\ell - \sqrt{\ell \log \ell}) = O(n\ell)$  steps,  $c$  an appropriate constant, a population of size 1 where for the individual  $x \in P$ ,  $LO_{z_1}(x_{B_i}) \in [\ell - 2\sqrt{\ell \log \ell}, \ell - \sqrt{\ell \log \ell}]$ ,  $1 \leq i \leq k$ , holds.*

**PROOF.** We consider the first phase consisting of  $T = cn(\ell - \sqrt{\ell \log \ell})$  steps,  $c$  an appropriate constant, which implies that  $c\ell(\ell - \sqrt{\ell \log \ell})$  mutations are applied to each block  $i$ . We show that each block consists of  $\ell - \Theta(\sqrt{\ell \log \ell})$  leading one bits after these  $T$  steps. If a block  $i$  is selected for mutation, then we have a classical leading ones problem on  $\ell$  bits with mutation probability  $1/\ell$ . We have  $r < \sqrt{\ell \log \ell}$  which implies that as long as no block has at least  $\ell + 1 - \sqrt{\ell \log \ell}$  leading ones, the population size is 1. Each block develops independently of the others and the probability that there is a block  $b$  for which the single solution  $x \in P$  obtained after  $T$  steps does not have the property  $LO_{z_1}(x_{B_i}) \in [\ell - 2\sqrt{\ell \log \ell}, \ell - \sqrt{\ell \log \ell}]$  after  $c\ell(\ell - \sqrt{\ell \log \ell})$  mutations to block  $b$  is  $o(1)$  using Theorem 9 in [19].

Hence, after  $T = cn(\ell - \sqrt{\ell \log \ell})$  steps the population  $P$  contains a single solution  $x$  for which we have  $LO_{z_1}(x_{B_i}) \in [\ell - 2\sqrt{\ell \log \ell}, \ell - 2\sqrt{\ell \log \ell}]$  for all  $i \in \{1, \dots, k\}$  with probability  $1 - o(1)$ .  $\square$

The previous lemma showed that the population does not grow for a significant phase of the optimization process and that the different blocks have obtained already a large number of leading ones. As BC-GSEMO does not destroy the progress made in any of the blocks we can show the following upper bound by taking into account the remaining work that needs to be done on each block and the maximum population size of  $2^k$ .

**THEOREM 2.** *Let  $r < \sqrt{\ell \log \ell}$  and  $k \leq n^{1/3}$ . Then with probability  $1 - o(1)$ , the optimization time of BC-GSEMO on  $f_{nkr}$  is  $O(2^k n \sqrt{\ell \log \ell})$ .*

**PROOF.** We already know that BC-GSEMO obtains with probability  $1 - o(1)$  in time  $O(n\ell)$  a population of size 1 where each block has a leading ones value at least  $\ell - 2\sqrt{\ell \log \ell}$ . We consider the process after this population has been produced for the first time and investigate the remaining process.

In the second phase, the population size is at most  $2^k$  and optimizing the remaining bits to agree with  $z_1$  (or  $z_2$ ) for each solution takes time  $O(2^k nk \sqrt{\ell \log \ell})$  with high probability. We show this by optimizing the blocks in lexicographic order. Assume all blocks  $b < i < k$  have already been fully optimized. Block  $i$  has at least  $\ell - 2\sqrt{\ell \log \ell}$  leading ones in each solution present in the population.

Consider a solution  $x$  with profile  $p_{i-1}(x) = (x_{B_1}, \dots, x_{B_i}) \in \{z_1, z_2\}^{i-1}$  for the first  $i-1$  blocks. Let  $s^1 = \max_{y \in P} \{LO_{z_1}(x_{B_i}) \mid p_{i-1}(y) = p_{i-1}(x)\}$  and  $s^2 = \max_{y \in P} \{LO_{z_2}(x_{B_i}) \mid p_{i-1}(y) = p_{i-1}(x)\}$ . The values of  $s^1$  and  $s^2$  do not decrease and the probability of an increase for each of them is at least  $1/(2^k e \ell)$  if block  $i$  is selected for mutation.

For each value of  $i$ ,  $1 \leq i \leq k$ , we consider a (sub)-phase of  $T_i = cn2^k \sqrt{\ell \log \ell}$  steps,  $c$  an appropriate large enough constant, which starts when the population  $P$  contains an individual for each profile  $\{z_1, z_2\}^{i-1}$  for the first time. A phase of  $T_i$  steps implies that there are  $c\ell 2^k \sqrt{\ell \log \ell}$  mutation steps applied to block  $i$ . The expected number of mutation steps applied to each solution  $x$  with profile  $p_{i-1}(x) \in \{z_1, z_2\}^{i-1}$  and maximal  $s^1$ -value (or  $s^2$ -value) is at least  $c\ell \sqrt{\ell \log \ell}$  and the expected number of increases of  $s^1$  (or  $s^2$ ) is at least  $c' \sqrt{\ell \log \ell}$ ,  $c' \geq 4$  a constant if  $c$  is large enough. The probability that there are less than the sufficient

$$2\sqrt{\ell \log \ell} \leq \frac{c'}{2} \cdot \sqrt{\ell \log \ell}$$

successful increases of  $s^1$  (or  $s^2$ ) is  $2^{-\hat{c}\sqrt{\ell \log \ell}}$  using Chernoff bounds where  $\hat{c} > 1$  if  $c$  is chosen large enough. Assuming  $k \leq n^{1/3}$  which implies  $\ell \geq n^{2/3}$  and using the union bound, the probability that there is a profile  $p_{i-1}(x)$  for which  $s^1 = \ell$  (or  $s^2 = \ell$ ) has not been achieved is at most

$$2n^{1/3} \cdot 2^{-\hat{c}\sqrt{n^{2/3}}} = 2n^{1/3} \cdot 2^{-\hat{c}n^{1/3}} = 2(1-\hat{c})n^{1/3} = 2^{-\Omega(n^{1/3})}$$

Considering all values of  $i$ ,  $1 \leq i \leq k$ , sequentially, the probability that there is a block  $i$  which has not been fully optimized its phase of  $T_i$  steps is  $k \cdot 2^{-\Omega(n^{1/3})} = 2^{-\Omega(n^{1/3})}$ . As there are  $k$  blocks that need to be considered, an optimal population is reached after at most  $\sum_{i=1}^k T_i = cnk2^k \sqrt{\ell \log \ell}$  steps with probability  $1 - 2^{-\Omega(n^{1/3})}$ . Taking into account the failure probability of  $o(1)$  for the first phase given in Lemma 5 completes the proof.  $\square$

## 5 EXPERIMENTAL INVESTIGATIONS

We now complement our theoretical findings with an experimental study. In order to show that a significant difference can already be observed for a moderate number of blocks, we consider  $k = 2, 3, 4$  for values of  $n = 24, 120, 240, 360, 480, 600, 720, 840$ . We consider problems with a value of  $r = 1, 2, 4$  and use  $t_{\text{epoch}} = 1000$  for BC-GSEMO. We obtained similar results as shown in the following when using  $t_{\text{epoch}} = 1, 100$  in BC-GSEMO. The results are shown as the mean number of function evaluations required to reach the optimal population for 30 independent runs. As there is a clear difference in the scaling behaviour of GSEMO and BC-GSEMO for all experimental investigations carried out, we do not report statistical test results. (Note: the standard error on the mean is too small to be visible, so we omit to plot it.)

The results for  $k = 2$  and  $t_{\text{epoch}} = 1000$  are shown in Figure 1. It can be observed that BC-GSEMO needs much fewer function values than GSEMO to compute the optimal population. Looking at the results of GSEMO for different values of  $r$ , we can see that a larger value of  $r$  results in a larger runtime. Increasing  $r$  from 1 to 2 for GSEMO, this increases the runtime significantly whereas the increase is rather minor when further increasing  $r$  from 2 to 4. This observed behaviour is expected from the problem definition and is

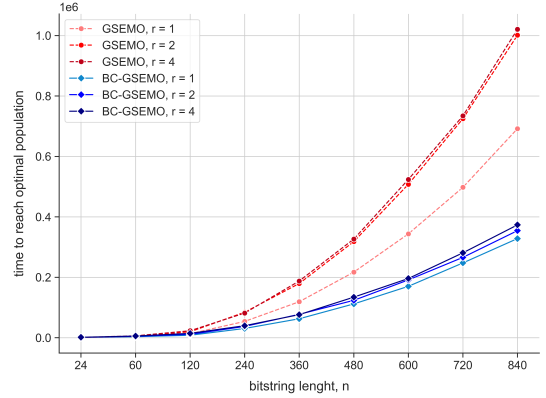


Figure 1: Mean time to reach the optimal population by GSEMO and BC-GSEMO (with  $t_{\text{epoch}} = 1000$ ) for  $k = 2$  dependent on the bitstring length  $n$  for  $r = 1, 2, 4$ .

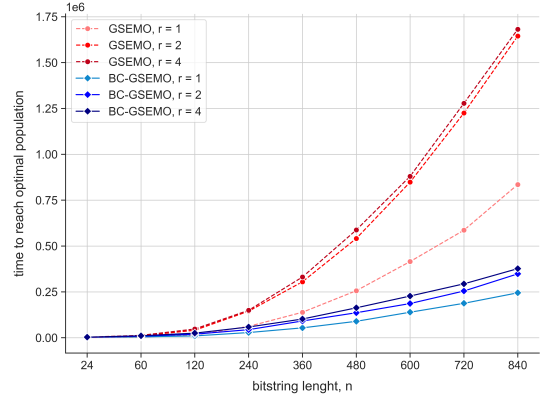
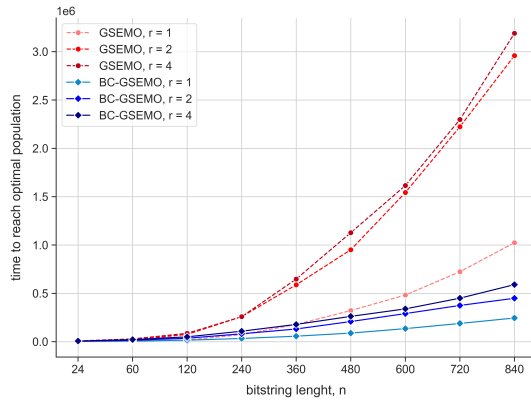


Figure 2: Mean time to reach the optimal population by GSEMO and BC-GSEMO (with  $t_{\text{epoch}} = 1000$ ) for  $k = 3$  dependent on the bitstring length  $n$  for  $r = 1, 2, 4$ .

also in line with our investigations for the lower bound of GSEMO, where we required that  $r$  grows with  $k$ . The increase in runtime with increasing value of  $r$  can also be observed for BC-GSEMO.

Furthermore, the results for  $k = 3, 4$  blocks and  $t_{\text{epoch}} = 1000$  are shown in Figure 2 and Figure 3, respectively. We can observe a similar behavior in terms of runtime for both algorithms. BC-GSEMO is consistently superior to GSEMO in terms of runtime, demonstrating its efficiency across various settings. Again the runtime for GSEMO significantly increases when increasing the problem parameter  $r$  from 1 to 2 whereas this difference is not that larger when increasing  $r$  from 2 to 4. For BC-GSEMO we can observe an increase in runtime when increasing  $r$  which is not as drastic as for GSEMO. This increase is again larger when increasing  $r$  from 1 to 2 than when increasing it further from 2 to 4. Comparing the results for different values of  $k$  across all three figures, we see that the runtimes are clearly increasing with  $k$  across all other problem parameter settings with the increase being significantly larger for GSEMO than for BC-GSEMO.





**Figure 3: Mean time to reach the optimal population by GSEMO and BC-GSEMO (with  $t_{\text{epoch}} = 1000$ ) for  $k = 4$  dependent on the bitstring length  $n$  for  $r = 1, 2, 4$ .**

## 6 DISCUSSION AND CONCLUSIONS

Through an example problem, we have shown that incorporating block coordinate descent into evolutionary multi-objective algorithms is provably useful. Our result is based on the insight that block-coordinate descent allows for the parallel optimization of the underlying blocks without destroying other components. In practical problems, we may need methods for identifying the blocks automatically if the problem has no obvious natural decomposition. Estimation of distribution algorithms (EDAs) that use factorised distributions are candidates for identifying blocks. A particularly promising candidate is Deep Optimization [6], an EDA that identifies and constructs latent variables hierarchically, resulting in lower and lower dimensional spaces to search in. Such latent variables could be interpreted as blocks. Empirically, Deep Optimization exhibits a favourable scaling behaviour on multiple knapsack and other problems due to its ability to search by combining modules without repeatedly destroying them.

Of course, there remains a gap between the simple function and algorithms we have fully analysed here, and more sophisticated, practical algorithms like Deep Optimization which presently remain out of reach of theoretical running time analysis. Nevertheless, our result, as well as providing an existence proof for the efficiency of block-coordinate methods in a multi-objective setting, also suggests a lens by which to understand such algorithms, and points the way to uncovering more about what may make these kind of algorithms efficient in practice.

## ACKNOWLEDGMENTS

The research has been initiated during Dagstuhl seminar 23361, “Multiobjective Optimization on a Budget”. Special thanks are due to the Working Group “Decomposition”, which in addition to the authors included Kerstin Dächert, Andrea Raith, Anita Schöbel, and Margaret M. Wiecek. Benjamin Doerr’s research benefited from the support of the FMJH Program Gaspard Monge for optimization and operations research and their interactions with data science. Frank Neumann has been supported by the Australian Research Council through grant FT200100536.

## REFERENCES

- [1] Anne Auger and Benjamin Doerr (Eds.). 2011. *Theory of Randomized Search Heuristics*. World Scientific Publishing.
- [2] Amir Beck and Luba Tretushvili. 2013. On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization* 23, 4 (2013), 2037–2060.
- [3] Nicola Beume, Boris Naujoks, and Michael Emmerich. 2007. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research* 181 (2007), 1653–1669.
- [4] Chao Bian, Yawen Zhou, Miqing Li, and Chao Qian. 2023. Stochastic population update can provably be helpful in multi-objective evolutionary algorithms. In *International Joint Conference on Artificial Intelligence, IJCAI 2023*. ijcai.org, 5513–5521.
- [5] Dimo Brockhoff, Tobias Friedrich, Nils Hebbinghaus, Christian Klein, Frank Neumann, and Eckart Zitzler. 2009. On the effects of adding objectives to plateau functions. *IEEE Transactions on Evolutionary Computation* 13, 3 (2009), 591–603.
- [6] Jamie Caldwell, Joshua Knowles, Christoph Thies, Filip Kubacki, and Richard Watson. 2022. Deep optimisation: Transitioning the scale of evolutionary search by inducing and searching in deep representations. *SN Computer Science* 3, 3 (2022), 253.
- [7] Sacha Cerf, Benjamin Doerr, Benjamin Hebras, Jakob Kahane, and Simon Wietheger. 2023. The first proven performance guarantees for the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) on a combinatorial optimization problem. In *International Joint Conference on Artificial Intelligence, IJCAI 2023*. ijcai.org, 5522–5530.
- [8] Carlos A Coello. 2000. An updated survey of GA-based multiobjective optimization techniques. *ACM Computing Surveys (CSUR)* 32, 2 (2000), 109–143.
- [9] David W Corne, Kalyanmoy Deb, Peter J Fleming, and Joshua D Knowles. 2003. The good of the many outweighs the good of the one: Evolutionary multi-objective optimization. *IEEE Connections Newsletter* 1, 1 (2003), 9–13.
- [10] Kerstin Dächert, Benjamin Doerr, Joshua D Knowles, Aneta Neumann, Frank Neumann, Andrea Raith, Anita Schöbel, and Margaret M Wiecek. 2024. Modeling and decomposition – biobjective block-coordinate descent. *Dagstuhl Reports* 13, 9 (2024), 45–55.
- [11] Kalyanmoy Deb. 2015. Multi-objective evolutionary algorithms. In *Springer Handbook of Computational Intelligence*. Springer, 995–1015.
- [12] Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2 (2002), 182–197.
- [13] Kalyanmoy Deb and Himanshu Jain. 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation* 18, 4 (2014), 577–601.
- [14] Matthieu Dinot, Benjamin Doerr, Ulysse Hennebel, and Sebastian Will. 2023. Runtime analyses of multi-objective evolutionary algorithms in the presence of noise. In *International Joint Conference on Artificial Intelligence, IJCAI 2023*. ijcai.org, 5549–5557.
- [15] Anh Viet Do, Aneta Neumann, Frank Neumann, and Andrew M. Sutton. 2023. Rigorous runtime analysis of MOEA/D for solving multi-objective minimum weight base problems. In *Advances in Neural Information Processing Systems, NeurIPS 2023*. <https://openreview.net/forum?id=ORmVvN94B9>
- [16] Benjamin Doerr. 2019. Analyzing randomized search heuristics via stochastic domination. *Theoretical Computer Science* 773 (2019), 115–137.
- [17] Benjamin Doerr. 2020. Probabilistic tools for the analysis of randomized optimization heuristics. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, Benjamin Doerr and Frank Neumann (Eds.). Springer, 1–87.
- [18] Benjamin Doerr. 2021. Runtime analysis of evolutionary algorithms via symmetry arguments. *Information Processing Letters* 166 (2021), 106064.
- [19] Benjamin Doerr, Thomas Jansen, Carsten Witt, and Christine Zarges. 2013. A method to derive fixed budget results from expected optimisation times. In *Genetic and Evolutionary Computation Conference, GECCO 2013*. ACM, 1581–1588.
- [20] Benjamin Doerr, Bojana Kodric, and Marco Voigt. 2013. Lower bounds for the runtime of a global multi-objective evolutionary algorithm. In *IEEE Congress on Evolutionary Computation, CEC 2013*. IEEE, 432–439.
- [21] Benjamin Doerr and Frank Neumann (Eds.). 2020. *Theory of Evolutionary Computation—Recent Developments in Discrete Optimization*. Springer.
- [22] Benjamin Doerr, Dirk Sudholt, and Carsten Witt. 2013. When do evolutionary algorithms optimize separable functions in parallel?. In *Foundations of Genetic Algorithms, FOGA 2013*. ACM, 48–59.
- [23] Benjamin Doerr and Weijie Zheng. 2021. Theoretical analyses of multi-objective evolutionary algorithms on multi-modal objectives. In *Conference on Artificial Intelligence, AAAI 2021*. AAAI Press, 12293–12301.
- [24] Tobias Friedrich, Jun He, Nils Hebbinghaus, Frank Neumann, and Carsten Witt. 2010. Approximating covering problems by randomized search heuristics using multi-objective models. *Evolutionary Computation* 18 (2010), 617–633.

- [25] Tobias Friedrich, Timo Kötzing, Aneta Neumann, Frank Neumann, and Aishwarya Radhakrishnan. 2023. Analysis of (1+1) EA on LeadingOnes with constraints. In *Genetic and Evolutionary Computation Conference, GECCO 2023*. ACM, 1584–1592.
- [26] Oliver Giel. 2003. Expected runtimes of a simple multi-objective evolutionary algorithm. In *IEEE Congress on Evolutionary Computation, CEC 2003*. IEEE, 1918–1925.
- [27] Oliver Giel and Per Kristian Lehre. 2006. On the effect of populations in evolutionary multi-objective optimization. In *Genetic and Evolutionary Computation Conference, GECCO 2006*. ACM, 651–658.
- [28] Christian Horoba and Frank Neumann. 2010. Approximating Pareto-optimal sets using diversity strategies in evolutionary multi-objective optimization. In *Advances in multi-objective nature inspired computing*. Studies in Computational Intelligence, Vol. 272. Springer, 23–44.
- [29] Zhengxin Huang, Yuren Zhou, Zefeng Chen, and Xiaoyu He. 2019. Running time analysis of MOEA/D with crossover on discrete optimization problem. In *Conference on Artificial Intelligence, AAAI 2019*. AAAI Press, 2296–2303.
- [30] Zhengxin Huang, Yuren Zhou, Chuan Luo, and Qingwei Lin. 2021. A runtime analysis of typical decomposition approaches in MOEA/D framework for many-objective optimization problems. In *International Joint Conference on Artificial Intelligence, IJCAI 2021*. ijcai.org, 1682–1688.
- [31] Sven Jäger and Anita Schöbel. 2020. The blockwise coordinate descent method for integer programs. *Mathematical Methods of Operations Research* 91, 2 (2020), 357–381.
- [32] Thomas Jansen. 2013. *Analyzing Evolutionary Algorithms – The Computer Science Perspective*. Springer.
- [33] Jörg Lässig and Dirk Sudholt. 2010. The benefit of migration in parallel evolutionary algorithms. In *Genetic and Evolutionary Computation Conference, GECCO 2010*. ACM, 1105–1112.
- [34] Frank Neumann and Carsten Witt. 2010. *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer.
- [35] Andre Opris, Duc Cuong Dang, Frank Neumann, and Dirk Sudholt. 2024. Runtime analyses of NSGA-III on many-objective problems. In *Genetic and Evolutionary Computation Conference, GECCO 2024*. ACM. To appear.
- [36] Vahid Roostapour, Aneta Neumann, Frank Neumann, and Tobias Friedrich. 2022. Pareto optimization for subset selection with dynamic cost constraints. *Artificial Intelligence* 302 (2022), 103597.
- [37] Günter Rudolph. 1997. *Convergence Properties of Evolutionary Algorithms*. Verlag Dr. Kováč.
- [38] Anita Schöbel. 2017. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies* 74 (2017), 348–365.
- [39] Simon Wietheger and Benjamin Doerr. 2023. A mathematical runtime analysis of the Non-dominated Sorting Genetic Algorithm III (NSGA-III). In *International Joint Conference on Artificial Intelligence, IJCAI 2023*. ijcai.org, 5657–5665.
- [40] Qingfu Zhang and Hui Li. 2007. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation* 11, 6 (2007), 712–731.
- [41] Weijie Zheng and Benjamin Doerr. 2023. Mathematical runtime analysis for the non-dominated sorting genetic algorithm II (NSGA-II). *Artificial Intelligence* 325 (2023), 104016.
- [42] Weijie Zheng and Benjamin Doerr. 2024. Runtime analysis of the SMS-EMOA for many-objective optimization. In *Conference on Artificial Intelligence, AAAI 2024*. AAAI Press.
- [43] Zhi-Hua Zhou, Yang Yu, and Chao Qian. 2019. *Evolutionary Learning: Advances in Theories and Algorithms*. Springer.
- [44] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK-report* 103 (2001).