Theoretical Computer Science 801 (2020) 110–142

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Working principles of binary differential evolution ☆

Benjamin Doerr [a,*], Weijie Zheng [b,*,1]

[a] Laboratoire d'Informatique (LIX), CNRS, École Polytechnique, Institute Polytechnique de Paris, Palaiseau, France
[b] Shenzhen Key Laboratory of Computational Intelligence, University Key Laboratory of Evolving Intelligent Systems of Guangdong Province, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 31 December 2018
Received in revised form 15 August 2019
Accepted 19 August 2019
Available online 26 August 2019
Communicated by C. Witt

Keywords:
Differential evolution
Running time analysis
Working principles of evolutionary computing

ABSTRACT

We conduct a first fundamental analysis of the working principles of binary differential evolution (BDE), an optimization heuristic for binary decision variables that was derived by Gong and Tuson (2007) from the very successful classic differential evolution (DE) for continuous optimization. We show that unlike most other optimization paradigms, it is stable in the sense that neutral bit values are sampled with probability close to 1/2 for a long time. This is generally a desirable property, however, it makes it harder to find the optima for decision variables with small influence on the objective function. This can result in an optimization time exponential in the dimension when optimizing simple symmetric functions like OneMax. On the positive side, BDE quickly detects and optimizes the most important decision variables. For example, dominant bits converge to the optimal value in time logarithmic in the population size. This enables BDE to optimize the most important bits very fast. Overall, our results indicate that BDE is an interesting optimization paradigm having characteristics significantly different from classic evolutionary algorithms or estimation-of-distribution algorithms (EDAs).

On the technical side, we observe that the strong stochastic dependencies in the random experiment describing a run of BDE prevent us from proving all desired results with the mathematical rigor that was successfully used in the analysis of other evolutionary algorithms. Inspired by mean-field approaches in statistical physics we propose a more independent variant of BDE, show experimentally its similarity to BDE, and prove some statements rigorously only for the independent variant. Such a semi-rigorous approach might be interesting for other problems in evolutionary computation where purely mathematical methods failed so far.

☆ This is a significantly extended version of the 8-page conference paper [65]. All authors of this version have contributed equally. The authors are given in alphabetical order as common in theoretical computer science. In this version, we followed the suggestion of the reviewers of [65] and removed the results on the stability of two classic EDAs (Section 4.3 of [65] resp. Section 4.4 in the extended version [26]), which were given to demonstrate the difference from binary differential evolution. As suggested by the reviewers, we plan to publish these results in a separate paper.

* Corresponding authors.
E-mail address: zhengwj13@tsinghua.org.cn (W. Zheng).

1 This work was partially done while Weijie Zheng was a PhD student at the Department of Computer Science and Technology, Tsinghua University, Beijing, China, and partially during a research stay at École Polytechnique's computer science lab (LIX), Palaiseau, France.

https://doi.org/10.1016/j.tcs.2019.08.025

1. Introduction

The family of differential evolution (DE) heuristics, first proposed by Storn and Price in 1995 [57], has become one of the most successful branches of evolutionary computation in continuous optimization and has been applied with great success to many real world problems, see, e.g., the survey [9].

However, compared to the abundance of results in continuous optimization, DE for discrete search spaces is much less understood. The difficulties start with how to implement the inherently continuous working principles of DE in discrete search spaces. One approach is to embed the discrete optimization problem into a continuous setting and then utilize continuous DE. For instance, Pamparà, Engelbrecht, and Franken [51] employ angle modulation to generate binary strings from floating-point individuals. Engelbrecht and Pamparà [30] further use the sigmoid value of the individual as the probability to generate the bit value, and also propose a normalization mapping.

Much less effort has been put into the design of truly discrete DE algorithms. Historically the first to do so, to the best of our knowledge, are Gong and Tuson [33]. They apply the rigorous forma analysis method to derive in a generic way a DE variant for binary search spaces. Moraglio and Togelius as well as Moraglio, Togelius and Silva [47,48] define discrete versions of DE via another generic approach, namely by requiring that certain geometric properties of the operators should be maintained. They demonstrate the usefulness of this approach not only for binary representations, but also for permutations and vectors of permutations. Recently, Santucci, Baioletti and Milani [55] propose another differential mutation for permutation.

To the best of our knowledge, apart from the axiomatic definitions of the different binary DE algorithms, there are no theoretical analyses of these methods so far. This contrasts the increasing theoretical understandings on other evolutionary algorithms like simple mutation-based algorithms [28], the compact Genetic Algorithm (cGA) [27], ant colony optimizers [35,50], the univariate marginal distribution algorithm (UMDA) [7], the population-based incremental learning (PBIL) algorithm [37], and the cross-entropy (CE) algorithm [64]. The lack of theoretical work on binary DE could be caused by the relatively complicated dependences in the stochastic process of a run of a DE heuristic. There are two types of the stochastic dependencies in DE, one from reusing the same individuals when generating the mutant, and the other from the selection operator. As we shall see in this work, these dependencies lead to difficulties not seen in the analysis of the other evolutionary algorithms, which often treat the different bit positions independently (apart from the fitness-based selection).

Our results: Since a theoretical understanding of an evolutionary algorithm can be very useful for its future use, this paper conducts a first fundamental analysis of the working principles of the binary differential evolution (BDE) algorithm proposed by Gong and Tuson [33]. We concentrate on this BDE, since it is the historically first and because we feel that its derivation via forma analysis makes it most likely that it inherits the true nature of DE from the continuous world. However, we expect that our results in a similar manner hold for other variants of BDE.

We show that the stochastic dependencies discussed above lead to a behavior significantly different from what is observed with many other nature-inspired optimization heuristics, in particular those, for which a solid theoretical understanding exists. For example, many heuristics have the property that at any time any point of the search space can be generated (possibly with a small probability). For BDE, this is substantially different. We show that from the random initial population, only an exponentially small fraction of all individuals can be generated in one iteration (see Theorem 1). In a similar vein, we present an objective function f and a population P such that BDE from this population with probability 1 never finds the optimum of f. Here P can be chosen exponentially large in the dimension and for each bit position each value may occur exponentially often (Theorem 3).

Unlike most other optimization paradigms for bit-string representations, we show that BDE is stable in the sense of Friedrich et al. [31], that is, neutral bit values are sampled with probability close to $\frac{1}{2}$ for a long time. We prove that BDE is stable when optimizing the Needle function, in which all bits are neutral before the optimum is found. Here, precisely, we show that for a time exponential in the population size all bit values are sampled with frequencies in $[\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon]$, where $\epsilon > 0$ can be any small constant (Theorem 8).

The inherent dependencies in BDE prevent us from mathematically extending this stability result to arbitrary neutral bits. Therefore, similar to the mean-field approach in statistical physics, we analyze a simpler but similar model called iBDE in which each bit position is treated independently when generating the mutant. We experimentally show the similarity of the behavior between BDE and iBDE in neutral bits and theoretically show the stability of iBDE (Theorem 11), which is quite different from the general EDAs that are unstable [31].

As a second positive property, we show that BDE can quickly detect and optimize the most important decision variables. For instance, we prove rigorously that a dominant bit converges to the optimal value in time logarithmic in the population size (Theorem 14). We theoretically discuss the runtime of BDE for the LeadingOnes function under the assumption that the frequency of the ones in the population does not drop below a small constant fraction for a sufficiently long time. In this case, BDE finds the optimum of the D-dimensional LeadingOnes function in $O(D)$ iterations (Theorem 17). Similar to the discussion for neutral bits, we mathematically verify that this assumption holds for iBDE and experimentally show the similarity of BDE and iBDE in this respect. Analogous results hold for the optimization of the BinaryValue function (Theorem 19 and Lemma 20).

Although stability is generally a desirable property (see [31,19] for examples how stable EDAs can outperform common EDAs, which are all unstable), stability can make it hard to find the optimal values of decision variables with small influence