

Parameterless Gene-pool Optimal Mixing Evolutionary Algorithms

Arkadiy Dushatskiy, Marco Virgolin, Anton Bouter, Dirk Thierens and Peter A. N. Bosman

Abstract—When it comes to solving optimization problems with evolutionary algorithms (EAs) in a reliable and scalable manner, detecting and exploiting linkage information, i.e., dependencies between variables, can be key.

In this article, we present the latest version of, and propose substantial enhancements to, the Gene-pool Optimal Mixing Evolutionary Algorithm (GOMEA): an EA explicitly designed to estimate and exploit linkage information. We begin by performing a large-scale search over several GOMEA design choices, to understand what matters most and obtain a generally best-performing version of the algorithm. Next, we introduce a novel version of GOMEA, called CGOMEA, where linkage-based variation is further improved by filtering solution mating based on conditional dependencies.

We compare our latest version of GOMEA, the newly introduced CGOMEA, and another contending linkage-aware EA DSMGA-II in an extensive experimental evaluation, involving a benchmark set of 9 black-box problems that can only be solved efficiently if their inherent dependency structure is unveiled and exploited. Finally, in an attempt to make EAs more usable and resilient to parameter choices, we investigate the performance of different automatic population management schemes for GOMEA and CGOMEA, *de facto* making the EAs parameterless. Our results show that GOMEA and CGOMEA significantly outperform the original GOMEA and DSMGA-II on most problems, setting a new state of the art for the field.

Index Terms—Model-based Evolutionary Algorithms, Linkage Learning, Optimal Mixing, Estimation-of-Distribution Algorithms, Genetic Algorithms

I. INTRODUCTION

KEY to the success of any optimization algorithm in terms of search effectiveness and efficiency, is the ability to exploit structural features of the problem being solved. To this, Evolutionary Algorithms (EAs) are no exception. For EAs, it is predominantly the variation operators that need to be favorably configured to exploit structural features. One structural feature that is of particular importance is variable dependence. Not only does variable dependence have a direct influence on the inherent difficulty of a problem, not being able to exploit such dependency information may lead to very inefficient optimization performance. If two variables are

Arkadiy Dushatskiy is with the Life Sciences and Health research group of the Centrum Wiskunde and Informatica

Marco Virgolin is with the Life Sciences and Health research group of the Centrum Wiskunde and Informatica

Anton Bouter is with the Life Sciences and Health research group of the Centrum Wiskunde and Informatica

D. Thierens is with the department of Information and Computing of the University of Utrecht

Peter A. N. Bosman is with the Life Sciences and Health research group of the Centrum Wiskunde and Informatica and the department of Software Technology of the Delft University of Technology

completely independent in a problem, this problem can be solved by considering the variables separately. Conversely, if two variables are strongly dependent, joint settings of these variables need to be considered in order to find the optimal solution. In EAs, such dependencies (between the variables that are directly manipulated by the EA, i.e., the genes), are also known as *linkages*. It has been long known that groups of variables that exhibit such strong linkages need to be treated, with high probability, in a joint fashion by the variation operator in order for an EA to be an efficient solver [38], [34]. Especially in the domain of discrete variables that constitute a Cartesian search space, which is also the domain that this article pertains to, many EAs employ a mixing operator that exchanges parts of solutions. Ensuring this mixing operator is linkage-friendly, i.e., has a high probability of exchanging groups of genes that are highly dependent, can make the difference between obtaining efficient, i.e., low-polynomial, and inefficient, i.e., exponential, scale-up of the required running time to solve the problem [38], [34].

The relevance and importance of linkage processing is even more prominent when taking a black-box perspective on optimization. In Black-Box Optimization (BBO), there is very little to no information available on the problem being solved. Metaheuristics, among which EAs, are commonly formulated and studied in this context, with the notion of designing a powerful general problem solver in mind. Certainly, the no-free-lunch theorem assures us, that considering all possible optimization problems, no such solver exists [45]. However, a generally valid assumption can be made that the types of optimization problems we are interested in, are not completely random, but have some sort of exploitable structure. It is the exploitation of this structure however that governs whether or not optimization will proceed effectively and efficiently. This then brings us back to the linkage problem, for it is assumed that the structure of the typical optimization problems we are interested in, is nontrivial, i.e., its variables are not all fully independent. For this reason, we have no guarantee that a simple genetic algorithm with uniform crossover, or any static crossover operator for that matter, will effectively exploit the structure of the problem. Thus, their use comes with the risk of exponential scale-up of the required runtime on problems that are polynomial-time solvable [38]. To avoid this, linkage information needs to be exploited properly. In a BBO setting however, such information isn't readily available, and thus must be determined otherwise, using previously performed solution-quality, i.e., fitness, evaluations. This process is commonly known as *linkage learning*, which is a key concept in this article.

An argument can be made at this point that the added complexity and effort of performing linkage learning is superfluous because a true BBO scenario isn't frequently encountered when solving real-world problems. A need for BBO may still very well surface however, even when efficient local search heuristics are available for a particular problem. It is well-known that combining EAs and local search is highly effective for many problems [22]. The reason for this is that by applying local search to every solution, a second search problem can be seen to exist in the space of local optima of the optimization problem. Running a local search heuristic multiple times, i.e., a random restart heuristic, then can effectively be seen as random search in the space of local optima. This space may be searched more efficiently using an EA, which can be obtained by applying local search to every solution that the EA generates. Even when we understand very well the problem being solved to the point where we can design efficient local search algorithms, the nature of the search space composed of the local optima may still be extremely hard to analyze. In that space then, there is again a need for powerful BBO algorithms.

The linkage problem has already been identified a long time ago, and much work on tackling this problem has previously been done, often presented simultaneously with a new EA, see e.g., [33], [4]. Much of this work has been toward building more complex models that are capable of capturing problem structure in more intricate detail, up to the relatively complex task of estimating entire (factorized) probability distributions, as is done in Estimation-of-Distribution Algorithms (EDAs) [27]. Although ultimately capable of exploiting problem structure properly, the overhead involved with estimating actual probabilities surpasses the need to determine the linkage information that needs to be effectively exploited. Importantly, such overhead becomes more significant on large-scale problems causing scalability issues. This article focuses specifically on the linkage hurdle on the road to powerful BBO algorithms. In particular, we introduce the new version of the Genepool Optimal Mixing Evolutionary Algorithm (GOMEA) that seamlessly integrates the traditionally separate operators of selection and variation in EAs in order to get the most out of available linkage information. Moreover, a generalized model of linkage information allows linkage information to be processed at more than one level, e.g. processing a hierarchy of weak and strong dependencies.

This article extends our previously published work on GOMEA by a more extensive experimental analysis on more optimization problems and larger problem sizes, testing the impact of various possible design choices such as local search operators on GOMEA, and, importantly, we propose a novel variation operator which exploits conditional dependencies between sets of variables and is called Conditional GOM (CGOM). Finally, we demonstrate the practical applicability of GOMEA by designing parameterless modifications of it. The performance of new and old versions of GOMEA are compared is shown in comparison with other EAs, including, the recent version of DSMGA-II [9], and the Parameterless Population Pyramid (P3) [17]. Moreover, this article joins all algorithmic information from our previous work that is needed

to make this article self-contained and represent the current state-of-the-art in the GOMEA research line.

The remainder of this article is organized as follows. In Section II we discuss related work. In Section III we outline the general working scheme of GOMEA and present design options for its most important components in more detail. Also we present GOMEA instances without the population size parameter and describe schemes to run GOMEA in a population size-free fashion. Then, we present our benchmark problems and the design of experiments in Section IV, followed by the results in Section V. The article ends with a discussion in Section VI and conclusions in Section VII.

II. FROM GENETIC ALGORITHMS TO ESTIMATION-OF-DISTRIBUTION ALGORITHMS AND BACK AGAIN

It was already hypothesized by John Holland himself that the simple Genetic Algorithm succeeds at optimization if it can proliferate important building blocks [23], i.e., partially defined solutions for which it holds that, when averaging over all the solutions that it is part of in a population, the fitness is better than the average fitness of the population.

Key is the proper mixing of these partial solutions, which means disrupting them as little as possible (i.e., copying them entirely from one solution to the next) and not copying other parts of the solution that they are (semi)-independent from. If these important partial solutions have a large probability of being destroyed during variation, for instance by using uniform crossover, the population size that is required to find the optimum may grow exponentially with the problem size. Conversely, polynomial population size growth can be achieved if the partial solutions are properly mixed. A well-known example of this is represented by the sum of additively decomposable, non-overlapping, deceptive trap functions [12].

Since then, there has been a dedicated research line in the field of evolutionary computation to design variation operators that are capable of automatically detecting the presence of important building blocks, and of reconfiguring the way in which variation proceeds to ensure that building blocks are mixed well and disrupted as little as possible. The first family of EAs along this line was the messy genetic algorithm family [25]. Algorithms in this family allowed genes to be re-ordered by explicitly encoding their location. Although eventual algorithms were able to avoid exponential scale-up, the overhead of re-ordering genes was still substantial and lacked explanatory statistical underpinning.

For this reason, researchers started looking into probabilistic approaches that were capable of explicitly computing dependencies between problem variables by estimating probability distributions over them. The population can be seen as a database that represents the type of solutions that are desired, and, over time, through selection, gets pushed toward the optimum. Selecting the better solutions makes dependencies stand out, since on average the solutions that contain important building blocks will have a better fitness than those solutions that do not. By estimating a probability distribution from the population, these dependencies can be explicitly modeled in a probabilistic fashion. Moreover, by sampling new

solutions from the estimated distribution, these dependencies are respected. Because the process of sampling generates a new database that has the same statistical properties as the original database (to the extent to which these properties were modeled in the probability distribution), this approach can be considered as mixing solutions at a population level rather than at the two-parent level as was before typically reminiscent of genetic algorithms. This type of algorithm is known as the Estimation-of-Distribution Algorithm (EDA) [27], [29]. Effectively and efficiently estimating probability distributions that capture higher-order dependencies was still key however to avoid exponential scale-up on problems with non-trivial dependency structure. Initial attempts that used either univariately factorized probability distributions (e.g., pBIL [1], and cGA [20]) that modeled every variable to be independent from every other variable, or bivariately factorized distributions that considered variable dependencies of at most order two (e.g., MIMIC [11], COMIT [2], and BMDA [35]) still fail to obtain polynomial scale-up on the additively decomposable deceptive trap functions. Low-order polynomial scale-up is only obtained by EDAs that model higher-order dependencies (e.g., ECGA [21], LFDA [31], and BOA [34]).

The most advanced EDA in this line is commonly accepted to be the hierarchical Bayesian Optimization Algorithm (hBOA) [33]. Both its predecessor BOA and hBOA itself estimate a Bayesian network every generation using a greedy learning procedure. hBOA however has the ability to store the parameters in this network more efficiently by storing only those combinations of values for dependent variables that actually appear in the population, thereby preventing the need to generate huge probability tables that require the explicit enumeration of *all* possible value combinations of a set of dependent variables. This, combined with a mechanism (restricted tournament selection) that promotes population diversity, allowed hBOA to be the only EDA capable of solving problems with hierarchical dependency structures while requiring only low-order polynomial-time scale-up of the population size and number of function evaluations. Although to a large extent now satisfactorily solving the linkage problem and providing a solid, statistically sound basis for doing so, the overhead required by hBOA is still substantial, requiring asymptotically $\mathcal{O}(n\ell^3)$ time per generation where ℓ is the number of problem variables and n the population size. Moreover, the number of generations required to solve a problem is typically in the same order as a properly configured GA, which is typically in the order of $\mathcal{O}(\sqrt{\ell})$ [32].

Although a solid approach to tackling the linkage problem, estimating entire probability distributions comes with the necessity to estimate not only a dependency structure, but also to estimate parameters (e.g., actual probabilities). Moreover, in order to decide what underlying dependency structure is a good one, i.e., not missing key dependencies and not overly complex, quality-of-fit measures need to be computed that decide when to stop the greedy learning approach that iteratively increases the complexity of the underlying dependency structure. These aspects are not necessarily important for tackling the linkage problem, because for that it would suffice to know which variables are (strongly) dependent on which

other variables. The joint probabilities of entire building blocks do not explicitly need to be computed, as they are stored implicitly in the population. Mixing the information stored in the population therefore automatically follows these probabilities. It is on these foundations that the GOMEA framework is based. GOMEA was first introduced in 2011 [4], posed as a broadened scope of the idea behind the original Linkage Tree Genetic Algorithm (LTGA) introduced in 2010 [39]. LTGA was one of the first algorithms to depart from the EDA principle of estimating entire probability distributions, and thus essentially going back to the notion of genetic algorithm, but still using similar statistical concepts as used in EDAs to detect dependencies. Ultimately this lead to a model-building complexity of an order of magnitude faster ($\mathcal{O}(n\ell^2)$) than hBOA, while being able to capture and exploit both low-order dependencies as well as high-order dependencies at the same time. Moreover, LTGA requires only a handful of generations to find the optimal solution due to much more extensive model exploitation during variation, further reducing the overall required model-building complexity. As later versions of LTGA, including the one presented in this article, are seen as instances of the GOMEA framework, details will be described in subsequent sections. Besides LTGA which we will from now refer to as LT-GOMEA, other non-EDA algorithms that build models to model and exploit linkage information have been proposed of late [24], [9], [17]. These algorithms that can more generally be described as model-based EAs have also been successful at outperforming hBOA.

III. THE GOMEA FAMILY OF EAS

The family of Gene-pool Optimal Mixing Evolutionary Algorithms (GOMEAs) has been proved to show impressive performance on benchmarks and, importantly, real-world problems. For instance, Real-Valued Multi-Objective GOMEA (RV-MO-GOMEA) [7] is now used for brachytherapy treatment planning optimization. This application received a Silver Humies award [18] which highlights its practical value and outstanding, better-than-human performance. Another example is adaptation of GOMEA for Genetic Programming (GP-GOMEA) [42]. Beside showing better performance than alternative GP algorithms on classical machine learning benchmarks, GP-GOMEA has been also successfully applied to a real-world medical problem, namely, a radiotherapy dose reconstruction [41], [44]. This application was noted with a Silver Humies award in 2021. These two examples show the potential of the GOMEA family of algorithms.

The family of GOMEAs is actually a subset of the OMEA family [4]. Another subset is the Recombinative OMEA (ROMEAs) family whereby mixing of solutions occurs only between 2 parent solutions rather than between all solutions in the population as is the case for GOMEA. When tested on various problems however, GOMEA was found to have the best performance as long as the models capturing linkage information were adequate [4]. For this reason, we focus particularly on GOMEA here.

The main idea behind the OMEA framework is that linkages are identified using sets of variable indices (see Section III-A),

which we shall also call linkage sets. These individual linkage sets are then explicitly exploited rather than implicitly as is the case in classical GAs and EDAs. In the latter, entire solutions are generated and subsequently evaluated. The main idea of OMEA however is to take values only for a linkage subset from a donor solution, and try these values out in another solution to see if it improves. It is this direct notion of acceptance that makes the success of the mixing operation independent of the effect of all other mixing events that may happen when constructing an entire new solution first. Because this makes each mixing event an optimal decision, unhampered by potential collateral noise, and because when all linkage sets are correctly identified, mixing essentially does not make any mistakes this way (unless unhelpful donor solutions are selected), this approach to variation was called Optimal Mixing (OM).

GOMEAs are a subclass of the general class of EAs and as such are a form of population-based search. The most traditional approach to population management is to have a population of a fixed size. We will discuss what GOMEA looks like with this approach as well as with different approaches to population management that no longer require the specification of a value for the population size parameter. The latter is especially of high practical value.

In the remainder of this section we provide more details on the various components of GOMEA.

A. Family Of Subsets (FOS) as a Linkage Model

The GOMEA class of EAs focuses on modeling linkage by explicitly identifying sets of variables to be treated jointly in the variation process. Moreover, such linkage sets are allowed to overlap. Specifically, any subset of the set of all variables may be identified within the linkage model. This may be defined as follows. Let $\mathcal{L} = \{0, 1, \dots, \ell - 1\}$ be the set of ℓ unique identifiers of variables that the EA processes, then the linkage model in GOMEA is a subset of the powerset of \mathcal{L} . Such a set is commonly called a family-of-subsets in mathematics. We therefore call the linkage model in GOMEA the family-of-subsets, or FOS, model, and denote it by \mathcal{F} , i.e:

$$\mathcal{F} \subseteq \wp(\mathcal{L}) \quad (1)$$

1) *Linkage Tree (LT) Model:* Though different ways to configure a FOS model by learning linkage from the population were introduced, we focus here on a so-called Linkage Tree (LT) model which demonstrated efficiency in solving various combinatorial optimization problems [39]. An LT is a binary tree with $2\ell - 1$ vertices. LT leaves are singletons of problem variables, the root of a LT is the set of all problem variables L , and all other vertices are variables subsets F^i which are unions of disjoint subsets of children k, j of vertex i : $F^i = F^j \cup F^k$, $F^j \cap F^k = \emptyset$.

2) *Similarity Measures:* An LT can be built in a bottom-up fashion using hierarchical clustering [26]: starting from singletons, the most similar subsets of variables are merged until a subset containing all variables is obtained (a tree root). A similarity between two subsets of variables F^i, F^j is defined as average similarity measure of all pairs of variables (X, Y)

where $X \in F^i, Y \in F^j$. Different similarity measures can be used [5]. Here, we consider two of them which are most commonly used (e.g., in [30], [14], [17]), namely, standard Mutual Information (MI) and Normalized Mutual Information (NMI). For two variables X, Y MI and NMI are defined as

$$\text{MI}(X, Y) = H(X) + H(Y) - H(X \cup Y)$$

$$\text{NMI}(X, Y) = \text{MI}(X, Y)/H(X \cup Y)$$

where $H(X)$ is information entropy, defined as

$$H(X) = \sum_{x \in \Omega_X} -P(X = x) \log(P(X = x))$$

3) *Linkage Tree Filtering:* It was shown in [6] that a full LT model (with $2\ell - 1$ vertices) may have redundant subsets which can be filtered out to increase mixing efficiency. Here we consider one particular case of filtering which was successfully applied in [17]. When two subsets F^j and F^k are merged into a subset F^i , it may happen, that the similarity between them is maximal (one in case of MI or NMI), which means that in a population, values of variables from one subset can perfectly predict values of variables from another subset. We suppose that there is no merit in using these subsets in mixing separately as it may disrupt this pattern and use additional unnecessary evaluations. Thus, keeping subsets F^j and F^k in a FOS is not reasonable and it is sufficient to keep only the parent subset F^i . In practice, to deal with possible numerical errors in similarity measure calculation, the filtering rule is invoked if the similarity measure value is above $1 - \varepsilon$ threshold (we use $\varepsilon = 10^{-6}$). Let $S(X, Y)$ be the similarity measure. After the filtering rule is applied, the subsets of an LT model satisfy the description:

$$\begin{aligned} \forall F^i, F^j, F^k \in \mathcal{F} \text{ s.t. } F^i &= F^j \cup F^k, \\ S(F^j, F^k) &\leq 1 - \varepsilon \end{aligned} \quad (2)$$

B. Gene-pool Optimal Mixing

Variation in GOMEA is guided by the contents of the FOS model in order to prevent disrupting the linkage information it represents. To do so, an operator called *Gene-pool Optimal Mixing (GOM)* is used that integrates selection and variation and has many similarities with greedy search algorithms. The GOM operator is described in pseudo-code in Algorithm 1.

GOM is applied to a single solution and outputs a single solution that is never worse than the input solution. To improve a solution, GOM loops over the contents of the FOS model. We consider two ways of iterating over FOS elements: in random order [39] and ascending order of subsets size ($|F^i|$) [17]. For each linkage subset F^i , GOM attempts to overwrite the values of the variables in F^i of the solution in consideration, with values from a donor solution that is chosen at random from the population. If this overwriting action does not cause the fitness of the solution to become worse, the copy action is accepted. Otherwise, the donor material is rejected and the action is undone. To allow traversing of fitness plateaus, changes that lead to the same fitness are also accepted if the solution is not the elitist one. Note that a FOS subset containing all variables,

i.e., the root of the LT model, is not used in GOM, as it implies replacing an entire solution rather than changing only a part of it.

1) *Exhaustive Donor Search*: When population diversity becomes low, it is likely that a randomly selected donor has the same genes F^i as the current solution, therefore, no new genotype is obtained. To deal with this situation, we can continue trying different donors until one is found in which genes F^i are different from the current solution is found. This modification is called *Exhaustive Donor Search (EDS)*, following [17].

2) *Forced Improvements (FI)*: If no subset F^i leads to changes in the solution undergoing GOM, the so-called *forced improvements (FI)* phase can (optionally) start. Originally, the FI was proposed in [4] to deal with convergence issues in MAXCUT. Namely, it can happen that the population starts to drift in fitness plateaus, i.e., solutions keep changing without improving. This lack of convergence makes it unlikely for further improvements to be discovered. Therefore, FI is specifically designed to steer the search towards converging to the (or one, if there is many) elitist solution. The FI phase works like the normal GOM phase, except for the fact that the donor solution is always set to be the elitist solution. Moreover, to further ensure convergence, changes that lead to equal fitness are now rejected (one can no longer drift in fitness plateaus). Only if the solution strictly improves in fitness, the overwrite action is accepted. To prevent the FI phase to reduce diversity too fast, the FI phase is stopped as soon as an improvement happens. Finally, if a solution could not be improved in the FI phase, it is overwritten by the elitist solution.

C. Conditional Gene-pool Optimal Mixing (CGOM)

By design, the GOM operator copies genes from a donor solution independently for each FOS element. Therefore, dependencies between FOS elements are not taken into account, i.e., when GOM is applied to a FOS element, any (weak) dependencies of variables inside the FOS element to variables outside the FOS element are not considered which might lead to suboptimal linkage usage because it may well be that although interactions between variables are of low order, they may still not be defined in terms of mutual exclusive subsets. I.e., consider the NK-landscapes with random subsets of variables for the subfunctions To alleviate this limitation, we consider a new gene-pool optimal mixing operator - the *Conditional GOM (CGOM)*. CGOM is closely related and inspired by recently introduced conditional linkage models for the real-valued GOMEA [8]. However, in RV-GOMEA conditional dependencies were not considered together with a hierarchical model like the LT which we do have for the first time.

CGOM works similar to GOM but takes into account what gene values are being processed to choose suitable donor solutions. Specifically, each FOS subset can be made conditionally dependent on a group of other variables. If the variables contained in a FOS subset F^i are conditionally dependent on variables not in F^i , CGOM takes this into account during

mixing. Suppose some genes have already been considered during mixing, i.e., for the current application of GOM to a given solution, these variables have been subjected to GOM before (they were in a FOS element considered earlier). We store these genes in a set U . When a new FOS element F^i is considered, we compute (explained below) the set of variables V s.t. 1) $V \cap F^i = \emptyset$, 2) all variables in V depend on the variables in F^i (we refer to such set of variables as G_i), and 3) $V \subseteq U$ (i.e., they were considered before). Since variables from V and F^i are interdependent, we enforce that selecting which genes configuration for F^i is considered, should be conditioned on V . This is achieved by considering as donor solutions only those which have the same genes for variables in V as the current solution undergoing CGOM has.

A minimal CGOM working example is shown in Figure 1. The CGOM differences as compared to GOM in terms of pseudocode are highlighted in Algorithm 1.

In the BBO paradigm, we have no a priori information on the dependence structure between variables. However, similar to FOS learning, we can estimate a notion of variable dependence based on the state of the population and the similarity measure (e.g., MI or NMI). Broadly speaking, we say that a FOS element F^i is dependent on a variable X ($X \notin F^i$) if the average pairwise similarity measure between the variables in F^i and X is relatively large compared to the average similarity measure between the variables only in F^i on the one hand and all the variables that do not belong to F^i on the other hand. Particularly, we use a threshold to detect such dependencies: a FOS element F^i is considered to have a dependency with variable j if the average pairwise similarity measure between j and variable in F^i is greater than λM where M is the largest average pairwise similarity score between variables from F^i and variables not belonging to F^i . This dependencies learning procedure is described in pseudocode in the function *learnDependencies* of Algorithm 3. The Hyperparameter λ is tunable. The smaller the value of λ , the larger the number of estimated dependencies. In other words, small λ values will result in high recall (we are unlikely to miss dependencies but might have many false positives), while large λ values will improve precision (we might miss many dependencies but will have few small positives).

D. GOMEA with a Traditional, Single Population

The initial population of n solutions is initialized randomly. After random solutions are generated, a local search algorithm can be applied to efficiently move them to a local optimum.

Local Search: We consider two local search algorithms here: simple *Single-Iteration Hill Climber (SIHC)* and *Exhaustive Hill Climber (EHC)* which is SIHC repeated multiple times until no improvements are found in a single iteration over all variables. Both hill climber variants were shown to be efficient components of advanced EAs, for instance, SIHC was used in [24], and EHC was used in [17]. The pseudocode for considered Hill Climber algorithms is listed in Algorithm 2.

Tournament Selection: Each iteration of the main GOMEA loop starts with linkage model learning. GOMEA does not

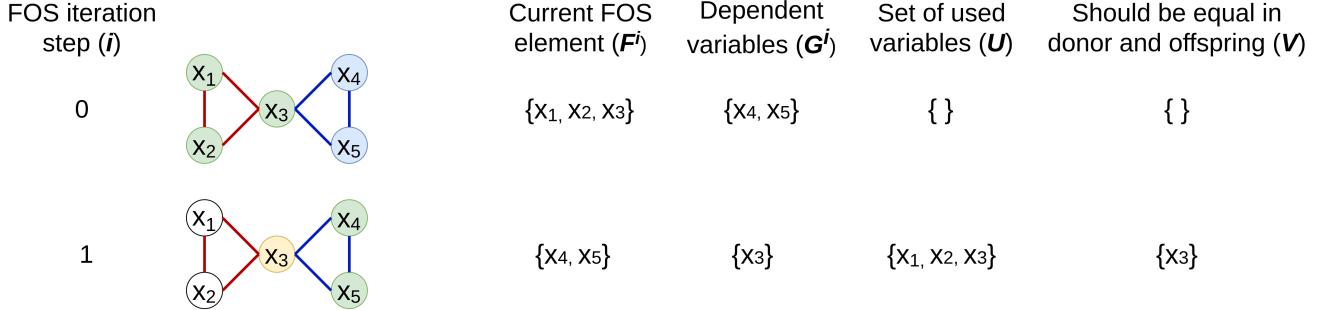


Fig. 1. A minimal CGOM working example. Fitness function f is a function of 5 variables and can be decomposed into two subfunctions: $f(x_1, x_2, x_3, x_4, x_5) = f_1(x_1, x_2, x_3) + f_2(x_3, x_4, x_5)$. This is shown by the different colors of edges. Suppose that after filtering the LT FOS contains two elements: $\{x_1, x_2, x_3\}$ and $\{x_4, x_5\}$. Variables from the current FOS element are colored in green; variables which are dependent with it are colored in blue; variables which are dependent with it and are already used (i.e., are taken into account by CGOM) are colored in yellow. Green colored genes are pooled to an offspring from only those donors which have the yellow colored genes equal to what is found in the current offspring.

have a traditional selection phase because GOM already induces selection, by discarding changes that are detrimental to a solution's fitness. However, we consider the option of using tournament selection to select good solutions upon which to learn the linkage model, as done in [24], [9]. We remark that with this option, the selection is disregarded after the linkage model is learned, i.e., it is not used to override the population.

After the linkage model is learned, the GOM variation operator is applied to every solution in the population to generate n offspring solutions. The population is then completely replaced by the offspring solutions. This main loop runs until the termination criterion is satisfied, which is naturally triggered when the population converges (i.e., all solutions have equal genotypes), but can also include other termination conditions such as a maximum allowed runtime, a maximum number of function evaluations, or a maximum number of generations. The pseudocode of single-population GOMEA is provided in Algorithm 3.

E. Going parameterless: removing the need to set the population size

The population size is a crucial parameter for the success of EAs. With model-based EAs like GOMEA, this is arguably even more so because the linkage model needs sufficient samples to be learned to achieve a sufficient level of accuracy for the linkage to be reliable. However, choosing the right population size is problem-dependent, and highly non-trivial. Methods to scale the population size automatically over time are therefore extremely useful and convenient in practice. In this paper, we consider two well-known population size-free schemes.

First, we consider the Interleaved Multistart Scheme (IMS), which was heavily inspired from the work by Harik and Lobo on parameterless GAs [19]. The IMS has been shown to be easy-to-use and can be naturally applied to almost any EA in various optimization domains [30], [28], [16], [43].

The IMS consists of evolving multiple populations simultaneously, in an interleaved fashion. Its pseudocode with recursive implementation is listed in Algorithm 4. In the

beginning, a single population is initialized, typically of a very small size (e.g., 2). After M_{IMS} generations, a new population is initialized that is larger, and it is advanced by one generation. This larger population will execute its next generation only after the smaller population performs M_{IMS} generations more. When the larger population has performed M_{IMS} generations, an even larger population is initialized, and so on. Our implementation of the IMS uses an initial population of size 2, exponential growth whereby each new population is twice the size of the previous, and $M_{IMS} = 4$, as in [19]. Smaller populations are terminated if they have converged, or their average fitness has become smaller than the average fitness of a larger population. This is because when a larger population has caught up with the smaller population, the latter will likely converge sooner, and can therefore be considered obsolete. Additionally, another convergence criterion such as a maximum allowed number of generations per population can be implemented.

The other schemes that we consider are the Population Pyramid (P3) [17], and its further modification Multiple Insertion Pyramid (P3-MI) [14]. The difference between the two is explained below and pseudocode is provided in Algorithm 5.

P3-MI arranges the population into a pyramidal structure, whereby each level of the pyramid is a set of solutions (duplicates are not stored). When a new population is created and, optionally, a local search is applied, all solutions are added to the bottom level of the pyramid. Then, by using solutions from the current pyramid level as donors, offsprings of the current population are generated. Solutions that are improved by variation are promoted and entered into one level higher in the pyramid. If there is no next level in the pyramid, a new one is created. This process continues until the pyramid's top level is reached or no solutions are improved during a generation. Every generation, Linkage models are learned for each pyramid layer independently. Sizes of populations are determined by a population *growth function*. The growth function takes the iteration number as input, and produces the population size (i.e., the number of solutions added to the bottom level of the pyramid). In [17] different population

Algorithm 1: GOM and the proposed CGOM operators. Highlighted lines are used in CGOM only.

```

Function CGOM( $s, \mathcal{G}, \text{useEDS}, \text{useFI}$ ) :
   $b \leftarrow o \leftarrow s$ 
   $\text{changed} \leftarrow \text{false}$ 
   $U \leftarrow \emptyset$ 
   $\mathcal{F} \leftarrow \text{orderFOS}(\mathcal{F})$ 
  for  $i \in \{0, 1, \dots, |\mathcal{F}| - 1\}$  do
     $\text{donorsList} =$ 
      randomPermutation( $\{\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_{n-1}\}$ )
    for  $j \in \{0, 1, \dots, n - 1\}$  do
       $U \leftarrow U \cup F^i$ 
       $d \leftarrow \text{donorsList}[j]$ 
      if  $\neg \text{checkDonor}(o, d, F^i, \mathcal{G}, U)$  then
         $\quad \text{continue}$ 
       $o_{F^i} \leftarrow d_{F^i}$ 
      if  $o_{F^i} \neq d_{F^i}$  then
        evaluateAndUpdateElitist( $o$ )
        if acceptChange( $o$ ) then
           $\quad b_{F^i} \leftarrow o_{F^i}$ 
           $\quad \text{changed} \leftarrow \text{true}$ 
        else
           $\quad o_{F^i} \leftarrow b_{F^i}$ 
           $\quad \text{break}$ 
      if  $\neg \text{useEDS}$  then
         $\quad \text{break}$ 
    if  $\text{useFI}$  and ( $\neg \text{changed}$  or  $o.NIS > 1 + \log_{10}(n)$ ) then
       $\quad \text{FI}(s);$ 
    if  $o.fitness \leq s.fitness$  then
       $\quad o.NIS \leftarrow o.NIS + 1$ 
    else
       $\quad o.NIS \leftarrow 0$ 
  Function checkDonor( $o, d, F^i, \mathcal{G}, U$ ) :
    /*  $\mathcal{G}(v)$  are variables linked with  $F^i$  */
    for  $p \in \mathcal{G}(v) \cap U$  do
      if  $o.v \neq d.v$  then
         $\quad \text{return false};$ 
    return true;
  Function FI( $s$ ):
     $\text{changed} \leftarrow \text{false}$ 
     $U \leftarrow \emptyset$ 
     $\mathcal{F} \leftarrow \text{orderFOS}(\mathcal{F})$ 
    for  $i \in \{0, 1, \dots, |\mathcal{F}| - 1\}$  do
      if  $\neg \text{checkDonor}(o, \text{elitist}, F^i, \mathcal{G}, U)$  then
         $\quad \text{continue}$ 
       $U \leftarrow U \cup F^i$ 
      if  $o_{F^i} \neq \text{elitist}_{F^i}$  then
         $\quad o_{F^i} \leftarrow \text{elitist}_{F^i}$ 
        evaluateAndUpdateElitist( $o$ )
        if  $o.fitness > b.fitness$  then
           $\quad \text{changed} \leftarrow \text{true}$ 
           $\quad \text{break}$ 
        else
           $\quad o_{F^i} \leftarrow b_{F^i}$ 
      if  $\neg \text{useEDS}$  then
         $\quad \text{break}$ 
      if  $\neg \text{changed}$  then
         $\quad o \leftarrow \text{elitist}$ 

```

Algorithm 2: Single-Iteration and Exhaustive Hill Climber algorithms .

```

Function exhaustiveHillClimber( $s$ ) :
  do
     $s, \text{improved} \leftarrow \text{singleIterationHillClimber}(s)$ 
  while  $\text{improved}$ 

Function singleIterationHillClimber( $s$ ) :
   $\text{improved} = \text{false}$ 
   $f_s \leftarrow \text{evaluate}(s)$ 
  for  $i \in \text{randomPermutation}(\{0, \dots, \ell - 1\})$  do
     $s' \leftarrow \text{copy}(s)$ 
     $s'_i \leftarrow s'_i \oplus 1$ 
     $f_{s'} \leftarrow \text{evaluate}(s')$ 
    if  $f_{s'} > f_s$  then
       $\quad s_i \leftarrow s'_i$ 
       $\quad f_s \leftarrow f_{s'}$ 
       $\quad \text{improved} \leftarrow \text{true}$ 
  return  $s, \text{improved}$ 

```

Algorithm 3: Single population GOMEA. Necessary modifications to use CGOM instead of GOM are highlighted.

```

Function singlePopulationGOMEA( $\text{useHC}, \text{useEDS}, \text{useFI}$ ) :
  while  $\neg \text{terminationCriterionSatisfied}$  do
     $\quad \mathcal{P} \leftarrow \text{doOneGeneration}(\mathcal{P})$ 

Function createPopulation( $n, \text{useHC}$ ) :
  for  $i \in \{0, 1, \dots, n - 1\}$  do
     $\quad \mathcal{P}_i \leftarrow \text{createRandomSolution}()$ 
    if  $\text{useHC}$  then
       $\quad \mathcal{P}_i = \text{hillClimber}(\mathcal{P}_i)$ 
    evaluateAndUpdateElitist( $\mathcal{P}_i$ )

Function doOneGeneration( $\mathcal{P}$ ) :
   $\mathcal{F} \leftarrow \text{learnModel}(\mathcal{P})$ 
   $\mathcal{F} \leftarrow \mathcal{F} \setminus \{0, 1, \dots, \ell - 1\}$ 
   $\mathcal{G} \leftarrow \text{learnDependencies}(\mathcal{F}, \mathcal{P})$ 
  for  $i \in \{0, 1, \dots, n - 1\}$  do
     $\quad O_i \leftarrow \text{GOM}(\mathcal{P}_i, \text{useEDS}, \text{useFI})$ 
    /* GOM usage can be changed to CGOM */
     $\quad O_i \leftarrow \text{CGOM}(\mathcal{P}_i, \mathcal{G}, \text{useEDS}, \text{useFI})$ 
   $\mathcal{P} \leftarrow \mathcal{O}$ 
  return  $\mathcal{P}$ 

Function learnDependencies( $\mathcal{F}, \mathcal{P}$ ) :
   $\mathcal{G} \leftarrow \{\emptyset\}_{i=0}^{|\mathcal{F}| - 1}$ 
   $S \leftarrow \text{calculateSimilarityMatrix}(\mathcal{P})$ 
  for  $i \in \{0, 1, \dots, \mathcal{F} - 1\}$  do
     $R \leftarrow \{0\}_{j=0}^{\ell - 1}$ 
    for  $j \in \{0, 1, \dots, \ell - 1\} \setminus F^i$  do
       $\quad R_j \leftarrow \frac{1}{|F^i|} \sum_{k=0}^{|F^i|-1} S_{j, F_k^i}$ 
     $M \leftarrow \max(R)$ 
    for  $j \in \{0, 1, \dots, \ell - 1\} \setminus F^i$  do
      if  $R_j > \lambda M > 0$  then
         $\quad \mathcal{G}^i \leftarrow \mathcal{G}^i \cup \{j\}$ 
  return  $\mathcal{G}$ 

```

Algorithm 4: Interleaved Multistart Scheme (IMS).

```

Function IMS (useHC, useEDS, useFI) :
    Populations  $\leftarrow \emptyset$ 
    while  $\neg$ terminationCriterionSatisfied do
         $\mathcal{P} \leftarrow \text{createPopulation}(2^{|Populations|+1})$ 
        Populations.append( $\mathcal{P}$ )
        generationalStep(Populations)
    Function checkTermination (Populations, i) :
        if converged(Populationsi) then
            return true
        for j  $\in \{i+1, \dots, |Populations| - 1\} do
            if averageFitness(Populationsj)  $>$  averageFitness(Populationsi) then
                return true;
        return false;
    Function generationalStep (Populations, first, last) :
        for iter  $\leftarrow \{0, 1, \dots, M_{IMS} - 1\} do
            for i  $\in \{first, \dots, last\} do
                if  $\neg$ Populationsi.terminated then
                    Populationsi.terminated  $\leftarrow$  checkTermination(Populations, i)
                if  $\neg$ Populationsi.terminated then
                    doOneGeneration(Populationsi)
            for i  $\in \{first, \dots, last - 1\} do
                generationalStep(Populations, first, i)$$$$ 
```

growth functions were studied. In this work we use a quadratic function (t^2 , where t is the iteration, starting from 1) as a trade-off between speed and number of function evaluations.

The P3 scheme is a special case of P3-MI with a constant growth function with value 1, in other words, one new solution is created and evolved in each iteration.

Algorithm 5: P3-MI population management scheme.

P3 is a special of case P3-MI when population growth function is constant and has value 1 for all iterations.

```

Function P3MI (useHC, useEDS, useFI) :
    iter  $\leftarrow 0$ 
    Pyramid  $\leftarrow [\emptyset]$ 
    while  $\neg$ terminationCriterionSatisfied do
        n  $\leftarrow$  growthFunction(iter)
        createPopulation(n, useHC)
        Pyramid0  $\leftarrow$  Pyramid0  $\cup \mathcal{P}$ 
        solutionsAdded  $\leftarrow$  true
        currentTopLevel  $\leftarrow |\text{Pyramid}| - 1$ 
        L  $\leftarrow 0$ 
        while L  $\leq$  currentTopLevel and solutionsAdded do
             $\mathcal{F} \leftarrow \text{learnModel}(\text{Pyramid}^L)$ 
             $\mathcal{F} \leftarrow \mathcal{F}/\{0, 1, \dots, \ell - 1\}$ 
            for i  $\in \{0, 1, \dots, n - 1\} do
                 $\mathcal{O}_i \leftarrow GOM(\mathcal{P}_i, \text{Pyramid}^L)$ 
                if  $\mathcal{O}_i.\text{fitness} > \mathcal{P}_i.\text{fitness}$  then
                    if L = currentTopLevel then
                         $\text{Pyramid}^{L+1}.\text{append}(\emptyset)$ 
                     $\text{Pyramid}^{L+1} \leftarrow \text{Pyramid}^{L+1} \cup \{\mathcal{O}_i\}$ 
                    solutionsAdded  $\leftarrow$  true
             $\mathcal{P} \leftarrow \mathcal{O}$ 
            L  $\leftarrow L + 1$ 
        iter  $\leftarrow iter + 1$$ 
```

IV. EXPERIMENTS
A. Benchmark problems

We consider various combinatorial optimization problems that are commonly considered to be particularly interesting for benchmarking GAs.

1) *Concatenated deceptive traps*: Concatenated deceptive trap is a well-known benchmark problem that was introduced to show that with disrupting building blocks, it takes exponentially growing resources to solve this problem. The fitness function of this problem is defined as:

$$f_{Trap_K^S}(x) = \sum_{i \in \{0, s, 2s, \dots\}, i < \ell} f_{Trap_K}^{\text{sub}} \left(\sum_{j=0}^{k-1} x_{(i+j)\% \ell} \right)$$

$$f_{Trap_K}^{\text{sub}}(u) = \begin{cases} k & \text{if } u = k \\ k - 1 - u & \text{otherwise} \end{cases}$$

Particularly, we consider trap functions with subfunctions size $k = 5$ and two different values of subfunctions overlap: separable traps with $s = 5$ (further referred to as $Trap_5^5$) and overlapping traps with $s = 4$ ($Trap_5^4$).

2) *Bimodal separable deceptive trap*: The bimodal symmetric concatenated trap functions [13] are interesting because in contrast to standard concatenated trap described above each subfunction has two modes. We consider bimodal symmetric traps of size 6, such that the subfunctions are given by

$$f_{BimodalTrap_K}^{\text{sub}}(u) = \begin{cases} 6 & \text{if } u = 0 \text{ or } u = 6 \\ 0 & \text{if } u = 1 \text{ or } u = 5 \\ 2 & \text{if } u = 2 \text{ or } u = 4 \\ 5 & \text{otherwise} \end{cases}$$

The considered bimodal traps do not have subfunctions overlap.

3) *NK-landscapes*: The NK-landscapes with maximum overlap (also called NK-S1 landscapes) [36] with subfunctions of size $k = 5$ are interesting because of overlapping subfunctions which are different depending on the position in genotype.

$$f_{NK}(x) = \sum_{i=0}^{l-k} f_{NK}^{\text{sub}}(x_{(i,i+1\dots,i+k)})$$

where the values of f_{NK}^{sub} are tabular values, sampled from the uniform distribution in $[0; 1]$ interval independently for different subfunctions positions.

4) *Hierarchical-If-And-Only-If (HIFF)*: The Hierarchical If-And-Only-If (HIFF) function is interesting because includes hierarchically ordered dependencies of exponentially growing sizes that overlap:

$$f_{Hiff}(x) = \sum_{k \in \{1, 2, 4, \dots, \frac{l}{2}, l\}} \sum_{i=0}^{l/k-1} f_{Hiff}^{\text{sub}}(x_{ik\dots(i+1)k-1})$$

$$f_{Hiff}^{\text{sub}}(u) = \begin{cases} 1 & \text{if } \sum_{j=0}^{k-1} u_j = k \text{ or } \sum_{j=0}^{k-1} u_j = 0 \\ 0 & \text{otherwise} \end{cases}$$

5) *MAXCUT*: We consider MAXCUT as a well-known combinatorial optimization problem. Given a weighted undirected graph (V, E) the goal is to find a partition of the vertices in two sets such that the sum of weights of edges running between vertices in different partitions is maximized. The fitness function is therefore defined as:

$$f_{MAXCUT} = \sum_{(i,j) \in E: x_i \neq x_j} w_{ij}$$

where w_{ij} is the weight of edge (i, j) .

We consider two types of MAXCUT instances. The first type is 3D square torus graphs. Each vertex is connected to 4 neighbors, forming a torus. Edges weights are integer values from $[1, 5]$ sampled uniformly. This type of instance is further referred to as MAXCUT Sparse. The second type of instance is dense graphs with randomly selected $\sqrt{\ell}$ neighbors for each vertex. This type of MAXCUT instances further referred to as MAXCUT Dense, and they are known to be NP-hard problem. For MAXCUT Dense, we use edge weights values in $[0, 1000]$ sampled uniformly.

6) *Ising spin-glass*: 2D Ising spin-glass problems have often been considered in the benchmarking of EDAs and other model-based EAs. The spin-glass problem fitness function is defined as

$$f_{spin\text{-}glass}(x) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} x_i x_j J_{ij}$$

where J_{ij} defines an interaction value between two variables, $J_{ij} \in \{-1, 1\}$. In the considered spin-glass instances each variable interacts with up to 4 neighbors in a 2D grid.

7) *MAXSAT*: Finally, we consider is MAXSAT problem. Particularly, we consider unweighted MAX-3SAT uniform random instances [9]. MAX-3SAT is NP-hard.

$$f_{MAXSAT}(x) = \sum_{i=0}^{m-1} (\vee_{j=0}^{p_i-1} \Gamma_{ij} x_{f_{ij}})$$

where m is the number of subfunctions (clauses), p_i is subfunction size (in the used instances $\forall i p_i = 3$), f_i determines which variables are contained by the subfunction with index i , and Γ can be either an unary negation operator (turning a binary x to an opposite value) or an identity operator keeping the value of x intact.

B. Sizes of problems

We frame our experiments in terms of scalability, i.e., we record what the effort is (in terms of time and function evaluations) for an EA to find the optimum, for growing problem dimensionality. For experiments where we traditionally adopt a single population, the maximum dimensionality we consider is set to 640 for Trap⁵, Trap⁴, and NK-S1, to 636 for Bimodal Trap, to 1600 for MAXCUT Sparse, to 784 for Spin-glass, to 1024 for HIFF, and to 100 for NP-hard MAXCUT Dense and MAXSAT. In experiments with automatic population sizing schemes, the maximum problem sizes are doubled for all problems except for MAXCUT Sparse and Spin-glass. For the experiments with a single population we need to use smaller

maximal dimensionalities because we included bisection to discover what the optimal population size is, but bisection quickly becomes computationally prohibitive to run for large problems.

C. Finding the best settings for single-population GOMEA

We summarize different possible choices of single-population GOMEA components in Table 1. Since we are interested in eliminating the need to choose parameters, we attempt to define what the best GOMEA variant is across the different benchmark problems.

Table 1. Considered hyperparameters of single-population GOMEA. In bold, the best settings found by the experiment described in Section IV-C.

Hyperparameter	Options
Forced Improvements	on / off
Exhaustive Donor Search	on / off
Hill Climber	SIHC / EHC / off
Linkage Tree and similarity measure	unfiltered, MI / filtered, NMI
FOS ordering	random / ascending subsets size
Tournament Selection (size 2)	on / off

In total, there are 96 ($2^5 * 3$) combinations of hyperparameters. We perform an exhaustive hyperparameter search by running all 96 GOMEA variants on a set of benchmark problems. Here, our goal is to fairly compare all GOMEA variants. In order to do so, for the largest considered size of each problem, we carry out the comparisons among configurations that all have a respective optimal population size. We estimate the optimal population size using the bisection method. The success condition in bisection is solving (i.e., achieving a global optimum) a problem instance in every of 50 consecutive runs. We do not put any hard constraints on runtime. Instead, we bound it by limiting the total number of function evaluations by 10^8 and, additionally, the total number of generations of each population by 200 (the same value as used in [9]) to prevent convergence problems. As it might happen that the smallest population size which allows to solve a problem instance does not require the fewest function evaluations, during the bisection procedure we keep track of the population size which allows to solve a problem instance with the fewest function evaluations. If the population size reaches 10^5 solutions and a problem instance is still not solved, the optimal population search procedure is terminated.

We rank the variants of GOMEA based on the minimal number of evaluations taken to find the optimal solution for each problem. The final ranking of a variant is the average of the rankings across the problems. If a variant is not able to solve one or more problems, it is dropped from the comparison.

The best GOMEA version is further referred to as *GOMEA^{best}*.

D. Adding CGOM operator

Once *GOMEA^{best}* is found, we look into the effect of replacing GOM with the new CGOM operator. Since CGOM

requires a detection threshold λ to be set, we run comparisons with $\lambda \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. We determine the best performing value of λ using the same approach as in Section IV-C. This best performing CGOMEA version is further referred to as $CGOMEA^{best}$.

E. Benchmarking algorithms using the optimal population size

$CGOMEA^{best}$ and $GOMEA^{best}$ are compared against each other, against the best previously published version of GOMEA [39], and against the most recent single-population DSMGA-II version [9]. The optimal population sizes for all algorithms are determined using bisection. To test the statistical significance of performance differences between two algorithms we use the pairwise Mann-Whitney U -test. For each problem we do a separate test which checks the hypothesis that the first algorithm needs fewer evaluations than the second algorithm to solve this problem in its maximum dimensionality.

F. Finding the best settings for parameterless algorithms

Next, we find the best performing parameterless version of GOMEA. The considered options of a parameterless scheme are IMS, P3-MI with quadratic population growth function, and P3. The scheme is seen as another tunable hyperparameter. We combine it with 96 hyperparameter combinations described in Section IV-C and perform a large-scale hyperparameter search, consisting of $96*3=288$ possible algorithm configurations. This best performing parameterless GOMEA version is further referred to as $GOMEA-P3^{best}$.

Once $GOMEA-P3^{best}$ is found, we replace GOM with the new CGOM operator (with λ value which was chosen for $CGOMEA^{best}$, i.e., 0.8). This CGOMEA version is further referred to as $CGOMEA-P3^{best}$.

Table 2. Considered hyperparameters of parameterless GOMEA. In bold, the best settings found by the experiment described in Section IV-F.

Hyperparameter	Options
Forced Improvements	on / off
Exhaustive Donor Search	on / off
Hill Climber	SIHC / EHC / off
Linkage Tree and similarity measure	unfiltered, MI / filtered , NMI
FOS ordering	random / ascending sub-sets size
Tournament Selection (size 2)	on / off
Population scheme	P3 / P3-MI / IMS

Additionally, we add to the experiments the original P3 algorithm and DSMGA-II with IMS [28]. Note that we do not test other population management schemes for DSMGA-II since, to the best of our knowledge, their integration with DSMGA-II have not been studied.

To study the practical applicability of the algorithms, we remove the limit on the number of function evaluations. Instead, in all experiments with parameterless algorithms we set a time limit of 24 hours. This is needed to make experiments computationally feasible as some of the considered algorithms (especially some configurations which use the P3 scheme and

DSMGA-II with IMS) perform in a way that the number of function evaluations is increasing very slowly.

G. Implementation details

All GOMEA variants and the P3 algorithm are implemented in C++¹. The P3² and DSMGA-II³ implementations are the ones used in their corresponding original articles with modified fitness functions to make them identical for all conducted experiments. Compiler settings for all considered algorithms are also identical.

V. RESULTS

A. GOMEA design choices search results

We found that the best performance of single-population GOMEA is achieved when using *Single-Iteration Hill Climber*, *Forced Improvements*, *Exhaustive Donor Search*, *Filtered Linkage Tree* which is build based on *Normalized Mutual Information*, *FOS sorted in ascending elements size order*, and *Tournament Selection with tournament size 2* applied before linkage model learning, as highlighted in Table 1.

With the best hyper-parameter settings, $GOMEA^{best}$ has better performance than the previously published GOMEA version on 7 out of 9 considered problems. These results are shown in Table 3, and scalability plots are presented in Figure 2. As shown in Table 6, these differences in performance are statistically significant at $\alpha = 0.01$. On the MAXSAT and HIFF problems the improvement is approximately of an order of magnitude. Two problems on which performance became worse are Bimodal Concatenated Trap and MAXCUT Dense. We notice that for the Bimodal Trap problem all algorithms of the GOMEA family perform worse than DSMGA-II. We believe that this is due to pairwise mutual information-based dependency learning fails, not optimal mixing itself (i.e., with the right FOS, scalability is excellent). Improving performance for this type of Deceptive Traps is an interesting question for future research. Importantly, the $GOMEA^{best}$ algorithm was able to solve all considered problems with the given constraints while DSMGA-II failed to solve the HIFF, and MAXCUT Dense problems. Therefore, we can say that $GOMEA^{best}$ is an algorithm that can tackle a larger class of non-trivial problems efficiently and it less likely fails to solve a problem. However, we see that on 4 out of 9 problems the performance of DSMGA-II is better (on Trap₅, NK-S1, and HIFF, with statistical significance).

Though the ultimate goal of the conducted hyperparameter search is to find the best performing combination of design choices for GOMEA, it is also interesting to analyze how these choices affect the performance individually. To do so, for each design choice, we study aggregated performance of all algorithms which use this design choice regardless of all other options they use. These results are shown in Figure 4. The most impactful design choices are Hill Climber, and

¹Source code is available on the repository of the first author: <https://github.com/ArkadiyD/BinaryGOMEA> and the website of the last author: <http://www.cwi.nl/~bosman>

²<https://github.com/brianwgoldman/FastEfficientP3/>

³<https://github.com/tianliyu/DSMGA-II-TwoEdge>

Exhaustive Donor Search. Results show that for most problems Exhaustive Donor search is beneficial and substantially improves the performance. Algorithms with Single-Iteration Hill Climber on most problems outperform the ones without it, but Exhaustive Hill Climber is, apparently, too greedy and therefore is inferior to both a more simple Hill Climber and no Hill Climber at all. This is in-line with earlier reported results [3]. Using the Filtered Linkage Tree built with the Normalized Mutual Information measure slightly improves the performance on some problems from the benchmark set, though it worsens the performance on the remaining ones. We see that Forced Improvements, FOS ordering and Tournament Selection do not have a strong effect on the performance. It is noteworthy that the effects of different design choices on the performance of GOMEA on the NK-landscapes are the opposite to their effect on the majority of other problems (e.g., Exhaustive Donor Search, Hill Climber and Filtered LT have worse performance), which suggests that the NK-landscapes problem has some unique properties compared to the other problems in the benchmark set.

B. CGOMEA performance

We take the found best performing GOMEA version ($GOMEA^{best}$) and replace GOM with CGOM. First, we analyze how the performance of CGOM-based GOMEA depends on the threshold parameter λ . Results for single-population CGOMEA with different λ values are provided in Table 4. We see that λ values between 0.6 and 0.9 provide similar performance on most problems though there are some outliers in performance (as on MAXCUT Dense problem with $\lambda = 0.7$) which are caused by the stochastic nature of the bisection procedure. Nevertheless, using the same approach as for selecting the best GOMEA version, we select $\lambda = 0.8$ as the value which provides the best average performance. CGOMEA with tuned λ value is further referred to as $CGOMEA^{best}$. We see that with $\lambda = 0.5$ performance deteriorates as detecting too many spurious dependencies slows down the mixing procedure. Hence, trying smaller values for λ is not necessary.

As shown in Table 3 and in scalability plots in Figure 2, $CGOMEA^{best}$ outperforms $GOMEA^{best}$ on 7 out of 9 considered problems. These differences are statistically significant as shown in Table 6. On $Trap_5^5$, $CGOMEA^{best}$ performs on par with $GOMEA^{best}$. Only on the HIFF problem CGOM performs slightly worse. Moreover, $CGOMEA^{best}$ performs better (with statistical significance) than DSMGA-II on 5 problems, and there are two problems (HIFF and MAXCUT Dense) which CGOMEA managed to solve but DSMGA-II did not. Importantly, CGOMEA is still able to reliably solve all considered problems. CGOMEA's slightly inferior performance on the HIFF problem can be explained by the complicated structure of HIFF: dependencies exist between all pairs of variables, and CGOM tends to include many variables as dependent ones, leading to less efficient variation as the pool of appropriate donors becomes more limited.

The scalability of single-population algorithms in terms of wall-clock time required to find an optimum is shown in Figure 3. CGOMEA and GOMEA scale similarly on all problems

Table 3. Results of single-population EAs. Best population sizes are found with bisection. Ranking per problem shown through color gradient from green (best, i.e., the fewest median number of function evaluations) to red (worst, i.e., largest median number of function evaluations or problem instance not solved in all 50 runs). All results are divided by 10^5 . Legend: G = GOMEA; D-II = DSMGA-II; $G^B = GOMEA^{best}$; $CG^B = CGOMEA^{best}$.

Problem	ℓ	G	D-II	G^B	CG^B
$Trap_5^5$	640	4.02	1.30	1.16	1.16
$Trap_5^4$	640	13.23	2.55	4.60	3.41
Bimodal Trap	636	34.87	3.60	114.36	99.69
<i>NK-S1</i>	640	62.65	66.72	51.36	37.03
<i>HIFF</i>	1024	7.81	N/A	0.92	1.20
<i>MAXCUT Sparse</i>	1600	7.31	80.03	3.87	3.35
<i>MAXCUT Dense</i>	100	0.49	N/A	0.51	0.37
<i>Spin-glass</i>	784	88.47	11.92	13.74	8.26
<i>MAXSAT</i>	100	69.63	3.97	5.20	2.95

Table 4. Results of single-population CGOMEA with different threshold values λ . Best population sizes are found with bisection. Ranking per problem shown through color gradient from green (best, i.e., the fewest median number of function evaluations) to red (worst, i.e., largest median number of function evaluations or problem instance not solved in all 50 runs). All results are divided by 10^5 .

Problem	ℓ	0.5	0.6	0.7	0.8	0.9
$Trap_5^5$	640	1.16	1.16	1.16	1.16	1.16
$Trap_5^4$	640	5.89	3.24	5.37	3.41	3.50
Bimodal Trap	636	120.73	106.39	100.11	99.69	104.63
<i>NK-S1</i>	640	31.96	28.77	34.18	37.03	46.86
<i>HIFF</i>	1024	1.35	1.10	1.15	1.20	0.91
<i>MAXCUT Sparse</i>	1600	3.84	3.90	3.27	3.35	3.46
<i>MAXCUT Dense</i>	100	0.62	0.53	0.62	0.37	0.58
<i>Spin-glass</i>	784	16.49	11.98	12.45	8.26	10.78
<i>MAXSAT</i>	100	4.08	4.89	3.44	2.95	3.25

which is better than DSMGA-II, especially on $Trap_5^5$, Bimodal Trap, NK-S1, HIFF, and MAXCUT Sparse. Only on Bimodal Trap CGOMEA is substantially slower than GOMEA though it requires fewer function evaluations. This can be explained by the more careful donor selection as done in CGOMEA. On the NP-hard MAXSAT problem scalability deviates from polynomial as expected, though on the NP-hard MAXCUT Dense problem it cannot be seen for the considered problem sizes.

C. Parameterless EAs

Results of experiments with parameterless EAs are presented in Table 5 and in scalability plots in Figure 5. We found that the best performance of a parameterless GOMEA is achieved when GOMEA uses *Single-Iteration Hill Climber*, *Exhaustive Donor Search*, *Filtered Linkage Tree* which is build based on *Normalized Mutual Information*, *randomly shuffled FOS*, and *P3* scheme, as highlighted in Table 2. The obtained parameterless GOMEA version is further referred to as $GOMEA-P3^{best}$. Note, that when P3 and P3-MI schemes do not use tournament selection, it makes them much more time

efficient, as population statistics needed for Linkage Learning can be efficiently updated instead of re-calculated from scratch [17]. Noteworthy, crucial design choices, such as Hill Climber, Exhaustive Donor Search, and Linkage Tree type and information measure are the same in $GOMEA^{best}$ and $GOMEA-P3^{best}$. Less important design choices (Forced Improvements, FOS ordering) differ, which is most likely due to the results stochastic nature. The $GOMEA-P3^{best}$ version, but with GOM replaced by CGOM ($\lambda = 0.8$ corresponding to the best value found in Section IV-C) is further referred to as $CGOMEA-P3^{best}$.

First, we see that DSMGA-II with IMS scheme was capable to solve only three problems in the experimental setup due to its issues with time efficiency. $CGOMEA-P3^{best}$ performs better than $GOMEA-P3^{best}$ on 5 problems out of 9 (on 4 of them with statistical significance). The most substantial differences are on Bimodal Trap, NK-S1, and Trap $_5^5$. Similar to results for single-population algorithms, CGOMEA performs worse than GOMEA on the HIFF problem, and differences between CGOMEA and GOMEA on Trap $_5^5$, Trap $_5^4$, and MAXCUT Sparse are subtle. Compared to the P3 algorithm [17] $CGOMEA-P3^{best}$ performs better on all problems except HIFF (on 6 problems with statistical significance). $GOMEA-P3^{best}$ performs better than P3 on 6 problems (with statistical significance).

We note that only DSMGA-II with IMS was capable of solving all 50 instances of the MAXSAT problem of size 200 within the given time limit. $CGOMEA-P3^{best}$ and P3 solved 49 problem instances, while $GOMEA-P3^{best}$ solved 48. As problem instances significantly vary in complexity, we show the results for the 48 instances that were solved by all algorithms in order to provide a fair comparison.

As shown in Figure 6, P3 versions of GOMEA and CGOMEA scale similarly to P3, though they are slower. Scalability in terms of required time to find an optimum is almost identical for $CGOMEA-P3^{best}$ and $GOMEA-P3^{best}$. Both $CGOMEA-P3^{best}$ and $GOMEA-P3^{best}$ scale better than DSMGA-II IMS on most problems.

VI. DISCUSSION

We implemented the conditional GOM operator using traditional, entropy-based similarity measures to predict dependencies between variables. Especially in early generations, this approach to detecting dependencies can be inaccurate, determining dependencies between variables which are actually independent, and missing some truly existing ones. Potentially, a more accurate approach to learning dependencies can further improve CGOM performance. Moreover, it can be interesting to apply CGOM operator in a Gray-Box Optimization (GBO) scenario when the true dependencies are known. Then, as was done for RV-GOMEA, a Bayesian network could be used rather than the conditional variant of the LT. The latter is advantageous when learning linkage in a BBO setting, but not as accurate and potentially more complex compared to a direct and concise modelling of conditional dependencies. This analogy of the original concept of CGOM is to be studied in future research. However, in that case, it should be compared

Table 5. Results of parameterless EAs. Ranking per problem shown through color gradient from green (best, i.e., fewest median number of function evaluations) to red (worst, i.e., largest median number of function evaluations or problem instance not solved in all 50 runs). For MAXSAT results are shown for 48 instances that were solved by all algorithms. All results are divided by 10^5 . Legend: P3 = P3; D-II = DSMGA-II; GP3 B = $GOMEA-P3^{BEST}$; CGP3 B = $CGOMEA-P3^{BEST}$.

Problem	ℓ	P3	D-II	GP3 B	CGP3 B
<i>Trap</i> $_5^5$	1280	2.72	31.10	1.71	1.61
<i>Trap</i> $_5^4$	1280	6.29	14.08	5.01	3.76
<i>Bimodal Trap</i>	1278	829.32	65.03	379.01	311.83
<i>NK-S1</i>	1280	99.10	N/A	105.58	71.21
<i>HIFF</i>	2048	5.92	N/A	4.67	6.55
<i>MAXCUT Sparse</i>	1600	6.47	93.12	3.50	3.52
<i>MAXCUT Dense</i>	200	1.24	15.97	0.80	0.80
<i>Spin-glass</i>	784	5.84	18.15	7.69	5.65
<i>MAXSAT</i>	200	32.43	113.84	37.57	28.74

to different forms of EAs such as [10] which were designed specifically for GBO.

The considered model-based EAs relied on entropy-based information measures to learn dependencies between variables. We notice that the performance on Bimodal Trap can be potentially improved if alternative linkage learning methods are used, such as fitness-based ones as it is known that alternative methods that use comparisons can find the right structure [37], [15]. In general, the current state-of-the-art results are achieved by entropy-based linkage learning techniques, though, replacing them or combining with other methods is a promising question for future research.

In this article, to determine the best design choices (hyperparameters) for GOMEA and CGOMEA, we assessed performance on a standard benchmark set, and ranked algorithms based on average performance. Though this benchmark set includes well-known combinatorial optimization problems, problems arising in practical tasks may have properties (such as fitness landscape and dependencies structure) which are very different from all common benchmark functions. Though practitioners are interested in having the best performing algorithm for their specific task, we do not have a priori knowledge of those tasks properties. Defining a good and comprehensive benchmark set is an open problem and active field of research [40]. We hypothesize, however that the state-of-the-art benchmark problems in the field of EAs for binary optimization that we used is a decent compromise, in that we expect that obtaining good average performance on these problems is a good predictor of performance on many a priori unknown tasks. Moreover, in a BBO scenario matching a real-world problem with a problem from a benchmark set is a hard, if even solvable, task itself. Therefore, we did not try to specify the best possible GOMEA and CGOMEA versions for each benchmark problem, but keep the focus on the best average performance.

VII. CONCLUSION

In this article we have continued the research line on the GOMEA family of algorithms with important innovations and comparisons of various ideas that have been proposed separately in the last decade since the introduction of GOMEA. First, we did an extensive hyperparameter search and obtained a version of GOMEA which showed significantly better performance than ever published before for GOMEA. Next, we introduced a new variation operator called Conditional Gene-pool Optimal Mixing (CGOM) which utilizes conditional dependencies of linkage model subsets on other variables to generate offspring solutions. GOMEA with CGOM (CGOMEA) outperformed GOMEA and DSMGA-II on most of the 9 considered diverse and non-trivial benchmark problems in a single-population EA experimental setup where we assess scalability of the algorithms of required resources to obtain the optimum. Finally, we searched for the best performing version of GOMEA integrated with various population size-free schemes. We found that CGOMEA with P3 scheme is a robust scalable algorithm which outperforms the competitors in terms of number of function evaluations required to find the global optimum on almost all problems setting a new state-of-the-art performance for most of the benchmark problems and a new GOMEA variant that can serve as a new baseline in model-based evolutionary algorithms for binary search spaces for the next decade.

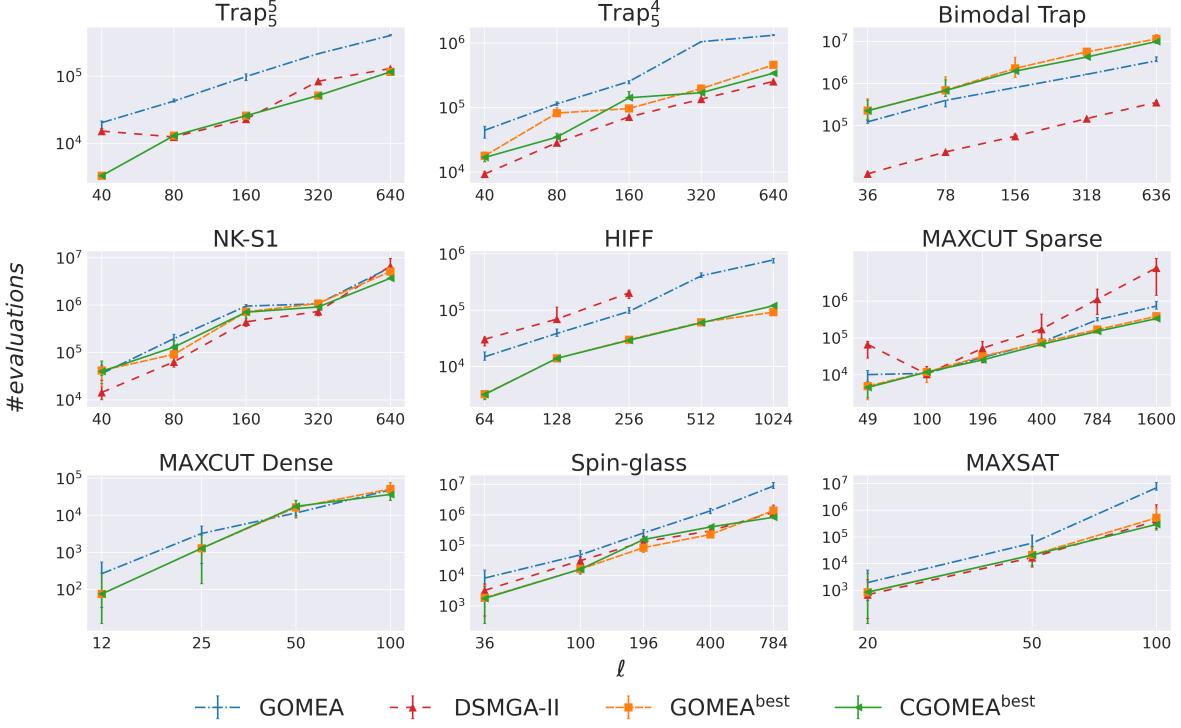


Fig. 2. Scalability of single-population EAs in terms of number of function evaluations required to find an optimum. Points show median values of 50 runs. Bars show 3rd and 48th order statistics (92% confidence interval). If an algorithm fails to find the global optimum of a problem instance in all 50 runs, the corresponding point is not shown. GOMEA refers to the previously published version [4].

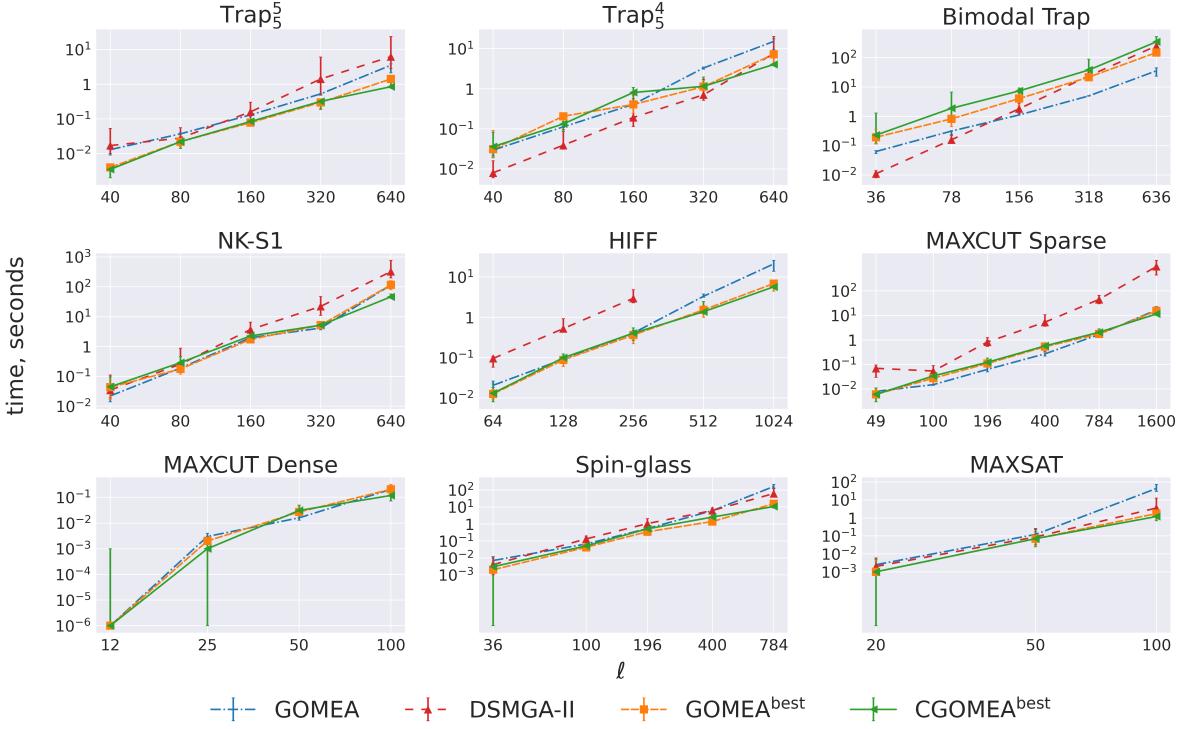


Fig. 3. Scalability of single-population EAs in terms of wall clock time required to find an optimum. Points show median values of 50 runs. Bars show 3rd and 48th order statistics (92% confidence interval). If an algorithm fails to find the global optimum of a problem instance in all 50 runs, the corresponding point is not shown.



Fig. 4. Effects of different design choices in GOMEA. Number of evaluations for each design choice for each problem are aggregated values of all possible GOMEA modifications with this design choice. Population sizes are found with bisection.

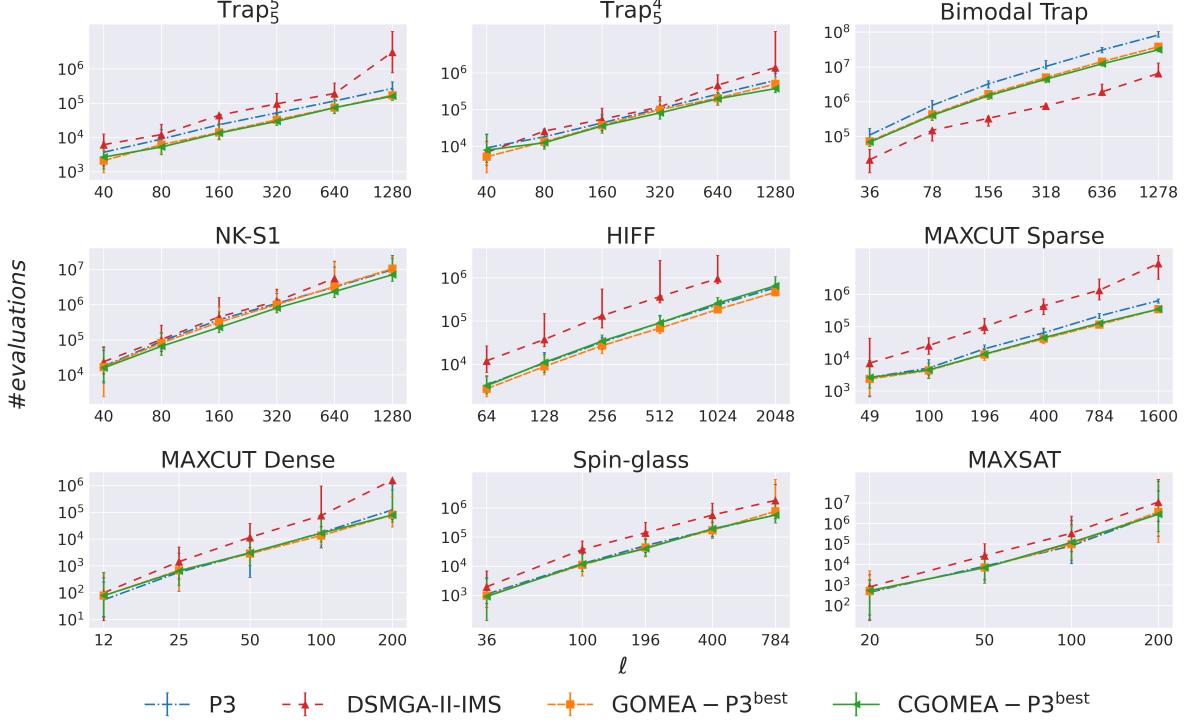


Fig. 5. Scalability of parameterless EAs in terms of number of function evaluations required to find an optimum. Points show median values of 50 runs (48 runs for MAXSAT problem). Bars show 3rd and 48th (46th for MAXSAT problem) order statistics (92% confidence interval). If an algorithm fails to find the global optimum of a problem instance in all 50 runs (48 runs for MAXSAT problem), the corresponding point is not shown.

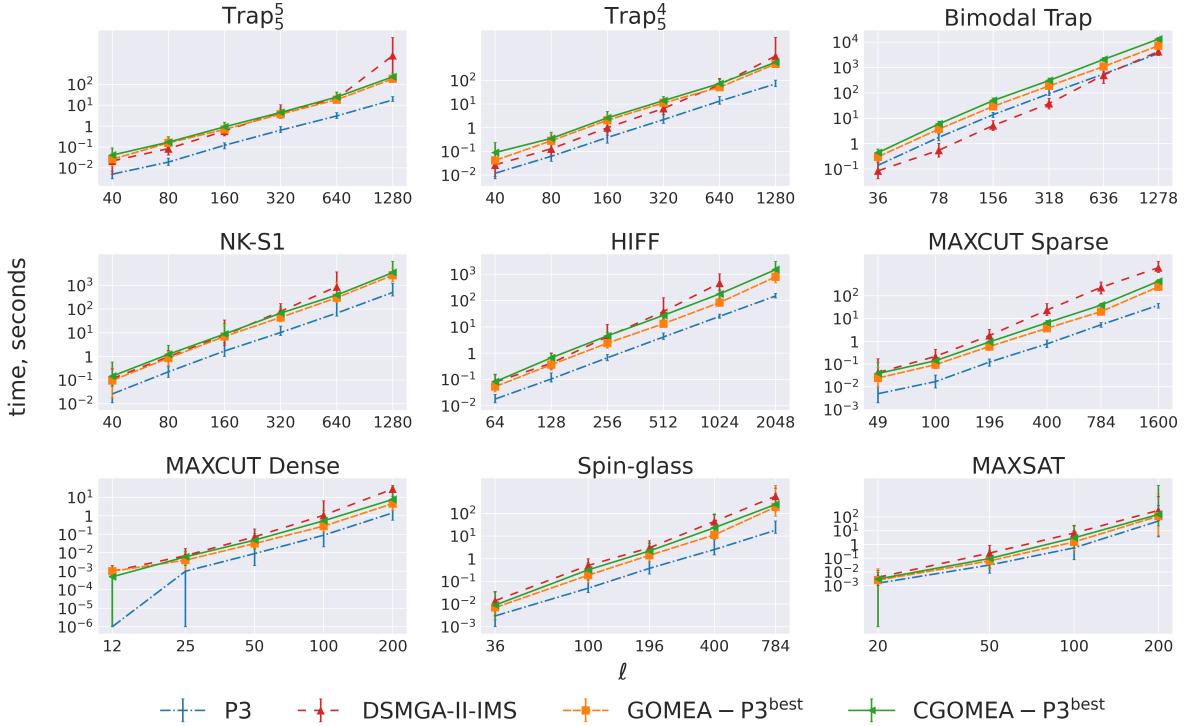


Fig. 6. Scalability of parameterless EAs in terms of wall clock time required to find an optimum. Points show median values of 50 runs (48 runs for MAXSAT problem). Bars show 3rd and 48th (46th for MAXSAT problem) order statistics (92% confidence interval). If an algorithm fails to find the global optimum of a problem instance in all 50 runs (48 runs for MAXSAT problem), the corresponding point is not shown.

Table 6. Statistical significance testing of performance difference in pairs of single-population EAs. Reported p -values are obtained by pairwise Mann-Whitney U -test. Statistically significant results at $\alpha = 0.01$ are highlighted. Cells corresponding to cases when an algorithm solves a problem instance and its competitor does not, are not highlighted. Legend: $G = \text{GOMEA}$; $D\text{-II} = \text{DSMGA-II}$; $G^B = \text{GOMEA-BEST}$; $CG^B = \text{CGOMEA-BEST}$.

Problem	ℓ	G^B is better than G	G^B is better than $D\text{-II}$	CG^B is better than G^B	CG^B is better than $D\text{-II}$
Trap_5^5	640	0.000	0.000	0.503	0.000
Trap_5^4	640	0.000	1.000	0.000	1.000
Bimodal Trap	636	1.000	1.000	0.000	1.000
NK-SI	640	0.000	0.000	0.000	0.000
HIFF	1024	0.000	N/A	1.000	N/A
MAXCUT Sparse	1600	0.000	0.000	0.000	0.000
MAXCUT Dense	100	0.740	N/A	0.000	N/A
Spin-glass	784	0.000	1.000	0.000	0.000
MAXSAT	100	0.000	0.960	0.000	0.000

Table 7. Statistical significance testing of performance difference in pairs of the best performing parameterless EAs. Reported p -values are obtained by pairwise Mann-Whitney U -test. Statistically significant results at $\alpha = 0.01$ are highlighted. Legend: $P3 = P3$; $D\text{-II} = \text{DSMGA-II}$; $GP3^B = \text{GOMEA-P3}^{\text{BEST}}$; $CGP3^B = \text{CGOMEA-P3}^{\text{BEST}}$.

Problem	ℓ	$CGP3^B$ is better than $GP3^B$	$CGP3^B$ is better than $P3$	$GP3^B$ is better than $P3$
Trap_5^5	1280	0.213	0.000	0.000
Trap_5^4	1280	0.000	0.000	0.000
Bimodal Trap	1278	0.000	0.000	0.000
NK-SI	1280	0.000	0.000	0.662
HIFF	2048	1.000	0.997	0.000
MAXCUT Sparse	1600	0.537	0.000	0.000
MAXCUT Dense	200	0.769	0.003	0.001
Spin-glass	784	0.007	0.227	0.985
MAXSAT	200	0.608	0.501	0.507

REFERENCES

- [1] Shumeet Baluja and Rich Caruana. Removing the genetics from the standard genetic algorithm. In *Machine Learning Proceedings 1995*, pages 38–46. Elsevier, 1995.
- [2] Shumeet Baluja and Scott Davies. Combining multiple optimization runs with optimal dependency trees. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1997.
- [3] Peter A.N. Bosman and Dirk Thierens. The roles of local search, model building and optimal mixing in evolutionary algorithms from a bbo perspective. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pages 663–670, 2011.
- [4] Peter A.N. Bosman and Dirk Thierens. Linkage neighbors, optimal mixing and forced improvements in genetic algorithms. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation*, pages 585–592, 2012.
- [5] Peter A.N. Bosman and Dirk Thierens. On measures to build linkage trees in Itga. In *International Conference on Parallel Problem Solving from Nature*, pages 276–285. Springer, 2012.
- [6] Peter A.N. Bosman and Dirk Thierens. More concise and robust linkage learning by filtering and combining linkage hierarchies. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 359–366, 2013.
- [7] Anton Bouter, Tanja Alderliesten, Bradley R. Pieters, Arjan Bel, Yury Niatsetski, and Peter A.N. Bosman. Gpu-accelerated bi-objective treatment planning for prostate high-dose-rate brachytherapy. *Medical physics*, 46(9):3776–3787, 2019.
- [8] Anton Bouter, Stefanus C. Maree, Tanja Alderliesten, and Peter A.N. Bosman. Leveraging conditional linkage models in gray-box optimization with the real-valued gene-pool optimal mixing evolutionary algorithm. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pages 603–611, 2020.
- [9] Ping-Lin Chen, Chun-Jen Peng, Chang-Yi Lu, and Tian-Li Yu. Two-edge graphical linkage model for dsmga-ii. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 745–752, 2017.
- [10] Francisco Chicano, Darrell Whitley, Gabriela Ochoa, and Renato Tinós. Optimizing one million variable nk landscapes by hybridizing deterministic recombination and local search. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 753–760, 2017.
- [11] Jeremy S. De Bonet, Charles Lee Isbell Jr., and Paul A. Viola. Mimic: Finding optima by estimating probability densities. In *Advances in neural information processing systems*, pages 424–430, 1997.
- [12] Kalyanmoy Deb and David E. Goldberg. Analyzing deception in trap functions. In *Foundations of genetic algorithms*, volume 2, pages 93–108. Elsevier, 1993.
- [13] Kalyanmoy Deb, Jeffrey Horn, and David E. Goldberg. Multimodal deceptive functions. *Complex Systems*, 7(2):131–154, 1993.
- [14] Willem den Besten, Dirk Thierens, and Peter A.N. Bosman. The multiple insertion pyramid: A fast parameter-less population scheme. In *International Conference on Parallel Problem Solving from Nature*, pages 48–58. Springer, 2016.
- [15] Arkadiy Dushatskiy, Tanja Alderliesten, and Peter A.N. Bosman. A novel approach to designing surrogate-assisted genetic algorithms by combining efficient learning of walsh coefficients and dependencies. *ACM Trans. Evol. Learn. Optim.*, 1(2), July 2021.
- [16] Arkadiy Dushatskiy, Adrienne M. Mendrik, Tanja Alderliesten, and Peter A.N. Bosman. Convolutional neural network surrogate-assisted GOMEA. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 753–761, 2019.
- [17] Brian W Goldman and William F Punch. Fast and efficient black box optimization using the parameter-less population pyramid. *Evolutionary computation*, 23(3):451–479, 2015.
- [18] Erik Goodman. Human-competitive results awards - "humies" 2019 - announces winners at gecco. *SIGEVOlution*, 12(3):3–5, January 2020.
- [19] Georges R. Harik and Fernando G. Lobo. A parameter-less genetic algorithm. In *GECCO*, volume 99, pages 258–267, 1999.
- [20] Georges R. Harik, Fernando G. Lobo, and David E. Goldberg. The compact genetic algorithm. *IEEE transactions on evolutionary computation*, 3(4):287–297, 1999.
- [21] Georges R. Harik, Fernando G. Lobo, and Kumara Sastry. Linkage learning via probabilistic modeling in the extended compact genetic algorithm (ecga). In *Scalable optimization via probabilistic modeling*, pages 39–61. Springer, 2006.
- [22] William Eugene Hart, Natalio Krasnogor, and James E. Smith. Memetic evolutionary algorithms. In *Recent advances in memetic algorithms*, pages 3–27. Springer, 2005.
- [23] John Henry Holland et al. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
- [24] Shih-Huan Hsu and Tian-Li Yu. Optimization by pairwise linkage detection, incremental linkage set, and restricted/back mixing: DSMGA-II. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, pages 519–526, 2015.
- [25] Hillol Kargupta. The gene expression messy genetic algorithm. In *Proceedings of IEEE International Conference on Evolutionary Computation*, pages 814–819. IEEE, 1996.
- [26] Alexander Kraskov and Peter Grassberger. Mic: Mutual information based hierarchical clustering. In *Information theory and statistical learning*, pages 101–123. Springer, 2009.
- [27] Pedro Larrañaga and Jose A. Lozano. *Estimation of distribution algorithms: A new tool for evolutionary computation*, volume 2. Springer Science & Business Media, 2001.
- [28] Yuen-Jen Lin and Tian-Li Yu. Investigation of the exponential population scheme for genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 975–982, 2018.
- [29] Jose A. Lozano, Pedro Larrañaga, Iñaki Inza, and Endika Bengoetxea. *Towards a new evolutionary computation: advances on estimation of distribution algorithms*, volume 192. Springer, 2006.
- [30] Ngoc Hoang Luong, Han La Poutré, and Peter A.N. Bosman. Multi-objective gene-pool optimal mixing evolutionary algorithm with the interleaved multi-start scheme. *Swarm and Evolutionary Computation*, 40:238–254, 2018.
- [31] Heinz Mühlenbein and Thilo Mahnig. Fda-a scalable evolutionary algorithm for the optimization of additively decomposed functions. *Evolutionary computation*, 7(4):353–376, 1999.
- [32] Martin Pelikan. Hierarchical bayesian optimization algorithm. In *Hierarchical Bayesian optimization algorithm*, pages 105–129. Springer, 2005.
- [33] Martin Pelikan and David E. Goldberg. Hierarchical bayesian optimization algorithm. In *Scalable optimization via probabilistic modeling*, pages 63–90. Springer, 2006.
- [34] Martin Pelikan, David E. Goldberg, Erick Cantú-Paz, et al. Boa: The bayesian optimization algorithm. In *Proceedings of the genetic and evolutionary computation conference GECCO-99*, volume 1, pages 525–532. Citeseer, 1999.
- [35] Martin Pelikan and Heinz Mühlenbein. The bivariate marginal distribution algorithm. In *Advances in Soft Computing*, pages 521–535. Springer, 1999.
- [36] Martin Pelikan, Kumara Sastry, David E. Goldberg, Martin V. Butz, and Mark Hauschild. Performance of evolutionary algorithms on nk landscapes with nearest neighbor interactions and tunable overlap. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation (GECCO 2009)*, pages 851–858. ACM, 2009.
- [37] Michał Witold Przewozniczek and Marcin Michał Komarnicki. Empirical linkage learning. *IEEE Transactions on Evolutionary Computation*, 24(6):1097–1111, 2020.
- [38] Dirk Thierens. Scalability problems of simple genetic algorithms. *Evolutionary computation*, 7(4):331–352, 1999.
- [39] Dirk Thierens and Peter A.N. Bosman. Optimal mixing evolutionary algorithms. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 617–624, 2011.
- [40] Koen van der Blom, Timo M. Deist, Tea Tušar, Mariapia Marchi, Yusuke Nojima, Akira Oyama, Vanessa Volz, and Boris Naujoks. Towards realistic optimization benchmarks: a questionnaire on the properties of real-world problems. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, pages 293–294, 2020.
- [41] Marco Virgolin, Tanja Alderliesten, Arjan Bel, Cees Witteveen, and Peter A.N. Bosman. Symbolic regression and feature construction with gp-gomea applied to radiotherapy dose reconstruction of childhood cancer survivors. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1395–1402, 2018.
- [42] Marco Virgolin, Tanja Alderliesten, Cees Witteveen, and Peter A.N. Bosman. Scalable genetic programming by gene-pool optimal mixing and input-space entropy-based building-block learning. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1041–1048, 2017.
- [43] Marco Virgolin, Tanja Alderliesten, Cees Witteveen, and Peter A.N. Bosman. Scalable genetic programming by gene-pool optimal mixing and input-space entropy-based building-block learning. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO '17*, page 1041–1048, New York, NY, USA, 2017. Association for Computing Machinery.

- [44] Marco Virgolin, Ziyuan Wang, Tanja Alderliesten, and Peter A.N. Bosman. Machine learning for the prediction of pseudorealistic pediatric abdominal phantoms for radiation dose reconstruction. *Journal of Medical Imaging*, 7(4):046501, 2020.
- [45] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.