A Gray Box Manifesto for Evolutionary Combinatorial Optimization

Darrell Whitley

Colorado State University, Fort Collins, CO, 80523, USA,
whitley@cs.colostate.edu

It is almost an article of faith that most evolutionary algorithms utilize random mutation and random recombination operators. Often the mutation is uniform random and sometimes the recombination operator is also a type of "Uniform Crossover." However when solving classic combinatorial optimization problems, random mutation and recombination operators are often both unnecessary and unproductive.

Another recent trend is to characterize all evolutionary algorithms as "Black Box" optimizers where nothing is known about the objective function. Unfortunately, Black Box optimizers are subject to the restrictions of No Free Lunch theorems. This is especially a concern because recent Focused No Free Lunch proofs hold over finite and tractable sets of functions [9, 8].

Black Box optimizers using random operators are doom to fail in any competition against more intelligent forms of search. Today, all competitive MAXSAT and Graph Coloring heuristic search methods deterministically compute the location of improving moves in constant time. For MAXSAT, this result has been known since 1992 [6]. Algorithms such as GSAT and Walksat (and their modern descendents) do not enumerate bit flip neighborhoods or use random sampling; instead, these algorithms can compute the location of improving moves and can track equal moves while also targeting variables that appear in specific unsatisfied clauses [2]. Again this does **not** involve the enumeration of the bit flip neighborhood; instead, the exact location of improving moves can be determined analytically, on average in O(1) time. Under these conditions, random mutation is hopelessly inefficient. Mathematical proofs now exists which show that these same results hold over all k-bounded pseudo Boolean functions for "next improving move" local search; it is possible to compute which bits can yield an improving move in constant time [10]. For some classes of functions, such as MAXSAT, one can also (almost always) identify the best improving moves in constant time [13].

The requirement that the functions be k-bounded is also not as restrictive as it might at first seem. Just as all SAT problems can be reduced to a MAX-3SAT instance, all problems that have a bit representation can be transformed into a k-bounded pseudo Boolean function [1]. This raises an important challenge to all researchers working in evolutionary computation. Why does the field ignore these advances and continue to use random and blind Black Box operators?

Advances have also made in the realm of recombination operators. In domains such as MAXSAT and NK-Landscapes [7] as well as the Traveling Salesman Problem [5, 12] we can also prove that deterministic forms of recombination can offer new performance guarantees. *Partition Crossover* operators deterministically use problem decomposition to perform intelligent guided recombination. Given q properly chosen crossover points, partition crossover operators are proven to return the best of 2^q possible offspring. If the parents are known to be local optima, all of the offspring are proven to also be locally optimal in largest hyperplane subspace that contains both parents [7]. An example is given in Figure 1. This is a large industrial MAXSAT instance from a recent SAT competition designed to solve an air traffic controller shift scheduling problem (instance atco_enc3_opt1_13_48). Partition crossover decomposes the problem into 1087 subgraphs, and returns the best of 2^{1087} possible offspring in linear time. A more detailed discussion of deterministic operators can be found in the tutorial "Next Generation Genetic Algorithms" [14].

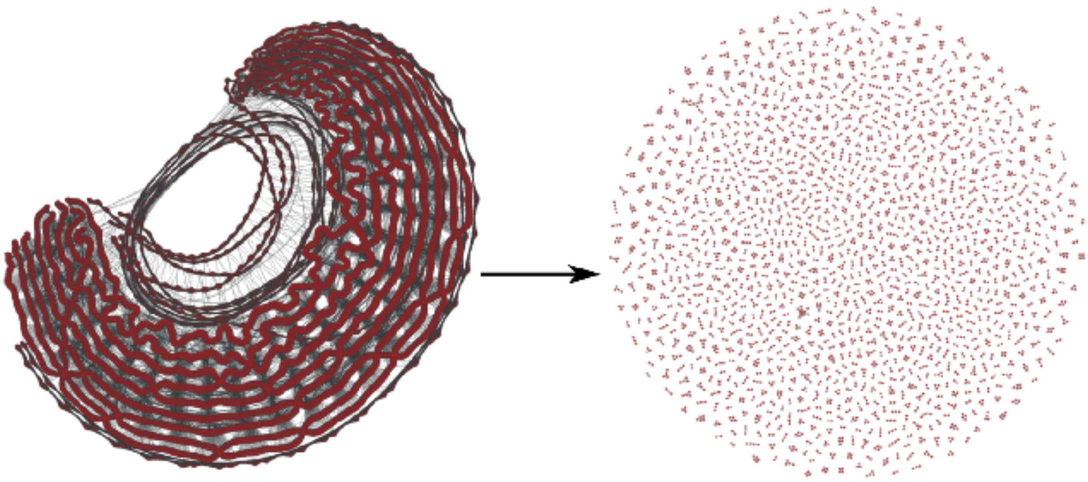

Figure 1: These images represent the variable interactions for a large industrial MAXSAT instance. On the left is the “Variable Interaction Graph” (VIG) where the vertices are variables and the edges represent nonlinear interactions between the variables. On the right, is the “Recombination Graph.” Given two parents that had been improved using local search, all of the variables that have the same assignment have been removed from the VIG to create the recombination graph. This decomposes the variable interaction graph into 1087 linearly independent subgraphs. Partition Crossover, which is a form of intelligent greedy crossover, can now return the best of 2^{1087} offspring in linear time by simply picking the best partial solution from each of the 1087 subgraphs.

I would argue that the only way for evolutionary algorithms to be competitive on many classic NP Hard combinatorial optimization problems is to abandon “Black Box” optimization and adopt more intelligent search methods. The result is a form of Gray Box optimization where knowledge about problem structure is actively and explicitly exploited [11]. And in many cases, there is an enormous amount of problem structure that is readily available, particularly when working with problems in combinatorial optimization. This line of research can also be described as a form of “Landscape Aware Heuristic Search” [3]. There have been several workshops on Landscape Aware Heuristic Search at the recent GECCO conferences.

Theory also needs to catch up. The (1+1)ES and Holland’s Simple Genetic Algorithm are more than fifty years old and yet most theory is still applied to these ancient algorithms in a Black Box scenario. Runtime analysis is also overly focused on test problems that have polynomial complexity. But this isn’t surprising: performing runtime analysis on NP-Hard problems means that one must deal with exponential runtimes in the worst case unless N=NP. Furthermore, assuming that runtime results on simple linear problems like ONEMAX can tell us anytime about how to apply evolutionary algorithms to nonlinear NP-hard problems is like a physicist assuming that atoms are just like ping pong balls. Finally, under a Gray Box scenario, simple test problems (such as ONEMAX, Trap functions, Leading Ones, and JUMP functions) become trivial to solve in linear time because the problems are separable and/or the order 1 hyperplanes all exactly point to the global optimum [11].

One might argue that “Evolutionary Algorithms” are based on natural selection and natural variation, and that random mutations and random recombination drive natural evolution. But this argument does not hold up to closer examination. In the mini-review article published in the *Journal of Bacteriology* in 2000, Barbara Wright describes DNA “Hot-Spots” where different parts of the DNA can have very different mutation rates. She also discusses mechanisms that can can cause mutation rates to vary. At the most basic level, mutations are more frequent on the non-transcripted parts of DNA. Wright also states there exists “an impressive array of circumstances that enhance background mutation rates in response to environmental stress ....” and that this can target specific genes [15]. Wright suggests several ways in which natural evolution is also “landscape aware.” In a paper in *Nature* in 2012, Martincorena et al. showed

that mutation rates across genes in *E. Coli* can vary by an order of magnitude and that the variation is not random: the mutation rate was lower in highly expressed genes and those undergoing strong selection [4]. They go on to suggest that the mutation rate maybe adapted to reduce the risk of deleterious mutations.
For anyone interested in solving real world optimization problems, Gray Box optimization offers too many advantages to be ignored. And even natural evolution displays features that appear to be “Landscape Aware.” This is good news. And it also open the door to new forms of theory that actively model problem structure in a useful fashion.

References

[1] E. Boros and P.L. Hammer. Pseudo-Boolean Optimization. *Discrete applied mathematics*, 123(1):155–225, 2002.
[2] H.H. Hoos and Th. Stützle. *Stochastic Local Search: Foundations and Applications*. Morgan Kaufman, 2004.
[3] A. Liefooghe, B. Derbel, S. Verel, H. Aguirre, K. Tanaka, EVOCOP, 2017, Towards landscape-aware automatic algorithm configuration: preliminary experiments on neutral and rugged landscapes. *EVOCOP*, Springer LNCS 10197: 215-232. 2017.
[4] I. Martincorena, A. Seshasayee and N. Luscombe Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature*, 485: 95-98, 2012.
[5] A. Möbius, B. Freisleben, P. Merz, and M. Schreiber. Combinatorial Optimization by Iterative Partial Transcription. *Physical Review E*, 59(4):4667–4674, 1999.
[6] B. Selman, H. Levesque, and D. Mitchell. A New Method for Solving Hard Satisfiability Problems. In *The National Conference on Artificial Intelligence (AAAI)*, pages 44–446, San Jose, CA, 1992.
[7] R. Tinós, D. Whitley, and F. Chicano. Partition Crossover for Pseudo-Boolean Optimization. In *Foundations of Genetic Algorithms*, *(FOGA-15)*, pages 137–149, 2015.
[8] E. Duéñez-Guzmán and M. Vose. No Free Lunch and Benchmarks. *Evolutionary Computation*, 21(2):293-312, 2016.
[9] D. Whitley and J. Rowe. Focused No Free Lunch Theorems. In *Genetic and Evolutionary Computation Conference (GECCO)*, pages 811–818. ACM, 2012.
[10] D. Whitley and W. Chen. Constant Time Steepest Descent Local Search with Lookahead for NK-Landscapes and MAX-kSAT. In *Genetic and Evolutionary Computation Conference (GECCO)*, pages 1357–1364. ACM, 2012.
[11] D. Whitley, F. Chicano, and B. Goldman. Gray Box Optimization for Mk Landscapes (NK Landscapes and MAX-kSAT). *Evolutionary Computation*, 2016.
[12] D. Whitley, D. Hains, and A. Howe. A Hybrid Genetic Algorithm for the Traveling Salesman Problem Using Generalized Partition Crossover. In *Parallel Problem Solving from Nature (PPSN)*, pages 566–575. Springer, 2010.
[13] D. Whitley, A. Howe, and D. Hains. Greedy or Not? Best Improving versus First Improving Stochastic Local Search for MAXSAT. In *The National Conference on Artificial Intelligence (AAAI)*, pages 940–946, 2013.
[14] D. Whitley. Next Generation Genetic Algorithms. *The Handbook of Metaheurisics*, 2019.
[15] B. Wright A Biochemical Mechanism for Nonrandom Mutations and Evolution. *Journal of Bacteriology*, 182 (11) 2993-3001, 2000.