AN INNOVATIVE METHOD TO ATTENUATE GENETIC DRIFT
IN GENETIC ALGORITHM OPTIMIZATIONS: APPLICATIONS TO ANALYTIC
OBJECTIVE FUNCTIONS AND RESIDUAL STATICS CORRECTION

S. Pierini, M. Aleardi, A. Mazzotti
Earth Sciences Department, University of Pisa, Italy

Introduction. The solution of geophysical non-linear inverse problems presents several challenges mainly related to convergence and computational cost. In addition, the performances of local optimization algorithms are strongly dependent on the initial model definition. For this reason, global optimization is often preferred in case of model spaces with complex topology (i.e. many local minima, small gradient of the objective function in a neighbourhood of the global minimum). Genetic Algorithms (GAs) (Holland, 1975) are a class of global optimization methods that have been proven very effective in solving geophysical optimization problems (Sajeva et al., 2016). In GA terminology, an individual, or chromosome, is a solution in the model space, whereas a population represents a set of individuals (i.e. an ensemble of possible solutions). A very simple GA flow starts with the generation of a random population of individuals over which the fitness function (namely the goodness of each solution) is evaluated. The fitness value stochastically contributes to the selection of the best individuals for the reproduction step in which a set of new solutions (offspring) is generated by combinations of parent individuals. The offspring are mutated, the fitness is evaluated, then a new generation is created by replacing some of the parent individuals with the generated offspring. The algorithm iterates until convergence conditions are satisfied. By assuming a population containing an infinite number of individuals, the convergence of the algorithm to the global minimum is guaranteed by the Holland theorem (Holland, 1975). For finite populations, the convergence of GAs is not guaranteed: this characteristic is often called “genetic drift”. A more heuristic description of the genetic drift phenomenon can be given in terms of population behaviour: after some generations, the chromosomes tend to converge in a convex neighbourhood of a minimum of the objective function (not necessary the global minimum), and thus the exploration of other promising portions of the model space is prevented. In the worst case, the population cannot escape from such convex neighbourhood, and a non-optimal solution is provided. Over the last decades, many strategies have been proposed to attenuate the genetic drift effect (Eldos, 2008; Aleardi and Mazzotti, 2017). For example, a possible approach is the so called niched genetic algorithm (NGA), in which the initial random population is divided into multiple subpopulations that are subjected to separate selection and evolution processes.

In this work, we propose an innovative method to attenuate the genetic drift effect that we call “drift avoidance genetic algorithm” (DAGA). The implemented method combines some principles of NGAs and Monte-Carlo algorithm (MCA) with the aim to increase the exploration of the model space and to avoid premature convergence and/or entrapment into local minima.
The DAGA performances are first tested on two analytic objective functions often used to test optimization algorithms. Then, the DAGA approach is applied to a well-known non-linear geophysical optimization problem: residual static correction. The DAGA method is also compared with standard NGA and MCA implementations.

**The DAGA method.** With the aim to increase the exploration capabilities of GAs, we first hybridize NGA and MCA by triggering catastrophic events over the evolving subpopulations (Eldos, 2008). A catastrophe is a random phenomenon, centred around a subpopulation that, basing on stochastic criteria, destroys a certain number of chromosomes in the central subpopulation and, to a lesser extent, in the neighbour subpopulations. After a subpopulation has been hit by a catastrophic event, its number of individuals decreases. For this reason, the destroyed chromosomes are progressively replaced over the following generations by new individuals migrating from the other subpopulations, and by new randomly generated solutions. The centre of the catastrophe is stochastically selected basing on the mean fitness values of the entire ensemble of subpopulations. The intensity of the catastrophic event (the number of chromosomes that will be removed) is randomly determined basing on a gaussian distribution with a user-defined variance and an adaptive mean value, which decreases as the number of generation increases (i.e. as the minimum of the objective function is approached). Note that in case of a catastrophic event the best chromosomes are always preserved to ensure that the optimization does not waste time re-discovering previously promising solutions.

In case of genetic drift, the genetic diversity is lost, or in other words, the entire population contains several copies of the same individual. In this case a considerable amount of time is wasted in performing forward modelling for very similar solutions. To tackle this issue, we include in our algorithm an additional strategy that systematically performs random replacements in each generation. Let \( t \) be a user-defined parameter; the algorithm identifies \( t \) couples of similar individuals for each subpopulation and one of the individuals for each couple is replaced with a randomly generated solution. Similarly to the catastrophic event, we again preserve the best individuals. Thanks to this operation, the genetic diversity within each population is preserved and the exploration of the model space is maximized.

The previous considerations make it clear that the DAGA approach is mainly aimed at increasing the exploration capability of a standard niched genetic algorithm and at efficiently exploring the most promising portions of the model space. Obviously in applying the DAGA method a good compromise must be found between exploration and exploitation of the algorithm. In practical applications, this translates in finding (usually by a trial and error procedure) an optimal set of user-defined parameters for the problem at hand. Indeed, a too strong exploration will heavily increase the computational time and slow down the convergence, while a too strong exploitation will result in premature convergence toward sub-optimal solutions. As a final remark, note that the inclusion of several Monte Carlo principles into the genetic algorithm optimization kernel, results in an increased computational effort. For this reason, an accurate code optimization is crucial to ensure the applicability of the method.

**Analytic test functions.** The performances of the DAGA method are first evaluated on the analytic Rastrigin and Schwefel functions. The DAGA method is also compared with standard implementations of NGA and MCA. In particular, for each method and for each objective function, we perform a set of 50 tests, and for each generation we progressively count the number of tests that attain convergence. The convergence criterion is based on a Euclidean measure of distance from the global minimum (Sajeva et al., 2017). To allow for a meaningful comparison, the total number of evaluated models is maintained constant for each considered method.

The Rastrigin function is a linear combination of a paraboloid and a harmonic function:

\[
f(x_1, \ldots, x_n) = 10n + \sum_{i=1}^{n} \left[ x_i^2 - 10 \cos(2\pi x_i) \right]
\]

where \( n \) is the dimension of the model space. Fig. 1a shows a 2-D Rastrigin function computed within \([-5, 5]^n\). The global minimum is located at \([0, \ldots, 0]\) and is surrounded by regularly
distributed local minima. In this case the algorithms are tested on a 4-dimensional Rastrigin function, and considering (for DAGA and NGA) a population of 26 individuals divided into 2 subpopulations and a maximum number of 250 generations. The results (Fig. 1b) are represented as incremental curves showing, for each generation, the cumulative sum of tests that attained convergence. In this case we observe that, as expected, the MCA is not able to efficiently explore the model space and converges in only 7 out of 50 tests, whereas NGA and DAGA show very similar performances. In particular, this example confirms that GAs are very efficient in finding the global minimum in case of regularly distributed minima (Sajeva et al., 2017). For this example, considering a serial Matlab code running on an intel i3@1.70GHz, the average convergence time for SGA and DAGA are 0.21 s and 0.49 s, respectively.

Sajeva et al. (2017) demonstrated that GAs severely suffer in case of objective functions with irregularly distributed minima. For this reason, we now compare the different algorithms on the Schwefel function:

\[
f(x_1, \ldots, x_n) = 418,982.9n - \sum_{i=1}^{n} x_i \sin(\sqrt{|x_i|})
\]  

For this example, considering a serial Matlab code running on an intel i3@1.70GHz, the average convergence time for SGA and DAGA are 0.21 s and 0.49 s, respectively.

Fig. 1 - a) 2-D plot of the Rastrigin function. b) Comparison of the performances of the different algorithms. Note the similar performances of NGA and DAGA.

Fig. 2 - a) 2-D plot of the Schwefel function. b) Comparison of the performances of the different algorithms. Note that the NGA is strongly affected by genetic drift, whereas the DAGA algorithm converges for all the tests.
Fig. 2a shows the 2-D Schwefel function computed within [500, 500]^n. Note that this function is characterized by a much more complex topology than the Rastrigin function. In particular, in the Schwefel function the local minima are more irregularly distributed, and important local minima are distant from the non-centred global minimum (located at [420.9687, ..., 420.9687]), or are even located at the opposite edge of the model space. In the following tests, we consider a 4-D space, a population of 100 individuals divided in 4 subpopulations, and a maximum number of generations equal to 800. The final results are represented in Fig. 2b. Note that none of the MCA tests satisfies the convergence criterion, and that for the NGA, only 22 out of 50 tests converge. Differently, the implemented DAGA approach converges for all the tests. From the one hand, these results confirm that the GA approach is strongly affected by genetic drift in case of irregularly distributed minima. From the other hand, Fig. 2b proves that the DAGA approach effectively attenuates the genetic drift and is able to find the global minimum even in case of objective functions with very complex topology. For this test, the average convergence time for SGA and DAGA is 1.76 s and 3.86 s, respectively.

Residual statics correction. The residual statics correction is a highly non-linear optimization problem that is often solved by applying global optimization methods (Rothman, 1985). In the following test, we compare the NGA and DAGA methods on CMP-consistent residual statics correction performed on a synthetic CMP gather. Sajeva et al. (2017) showed that this geophysical optimization problem is characterized by an objective function that shows some similarities to both the Rastrigin and Schwefel functions. For this reason, the GA encounters some problems in finding the global minimum in case of high-dimensional model spaces. In this case we use actual well log information and a convolutional forward modelling to generate the reference CMP gather (without residual statics; Fig. 3a). To simulate residual statics in the data, we apply to each trace in the reference CMP random time shifts uniformly distributed over the range -15/+15 ms (Fig. 3b). In the subsequent optimization process we allow time shifts within the range -25/+25 ms. In this case the model space dimensionality is equal to 40 (i.e. we consider 40 seismic traces), while the total number of individual is 200 divided into two subpopulations. Figs. 3c and 3d show the inversions results for the NGA and DAGA approaches, respectively. We note that the DAGA method returns a final corrected CMP gather (Fig. 3c) very similar to the reference one, whereas the final CMP yielded by the NGA approach shows many misalignments of the reflections and cycle-skipped traces (Fig. 3d). In Fig. 3e, we also observe that the

Fig. 3 - a) The synthetic reference CMP gather, b) the trace-shifted CMP, c) the final DAGA CMP, d) the final NGA CMP. e) Energy curves representing the evolution of the energy of the stack trace over iterations.
energy of the stack trace associated to the DAGA CMP gather is very close to the energy of the stack trace pertaining to the reference CMP. Note that differently from the NGA CMP, the final DAGA CMP could be used as a valid starting model for any local optimization method for a further refinement of the residual statics estimation. Due to the limited number of model evaluations the DAGA resulted to be only 1.11 times slower than NGA.

**Conclusions.** We presented an innovative strategy to attenuate the genetic drift and to increase the exploration of the model space in a GA optimization. In the proposed DAGA approach, a standard NGA is hybridized with principles coming from MCA. The test on analytic objective functions and on residual statics computation, demonstrated that, differently from the NGA, the implemented DAGA approach does not suffer the genetic drift effect. In particular, our tests confirmed that the DAGA approach grants the convergence in case of objective functions with very complex topology, where the standard NGA and MCA fail to converge. Differently, in cases of simpler topologies the NGA and the DAGA algorithms achieved very similar performances. The implemented DAGA method is more computational demanding than the standard NGA and for this reason an accurate code optimization has been performed. After additional code optimization (i.e. parallel implementation), the next step of our research is to apply the DAGA method to 2D acoustic full-waveform inversion that is a highly non-linear geophysical optimization problem with expensive forward modelling.

**References**


