 ICCES, vol.9, no.4, pp.233-242

Using genetic algorithms to find a globally optimal solution in uncertain environments with multiple sources of additive and multiplicative noise

Takéhiko Nakama[1]

Summary

Random noise perturbs objective functions in a variety of practical optimization problems, and genetic algorithms (GAs) have been widely proposed as an effective optimization tool for dealing with noisy objective functions. In this paper, we investigate GAs applied to objective functions that are perturbed by multiple sources of additive and multiplicative noise that each take on finitely many values. We reveal the convergence properties of GAs by constructing and analyzing a Markov chain that explicitly models the evolution of the algorithms in noisy environments. Our analysis shows that this Markov chain is indecomposable; it has only one positive recurrent communication class. Using this property, we establish a condition that is both necessary and sufficient for GAs to eventually (i.e., as the number of iterations goes to infinity) find a globally optimal solution with probability 1. Similarly, we identify a condition that is both necessary and sufficient for the algorithms to eventually with probability 1 fail to find any globally optimal solution. We also discuss how the transition probabilities of the chain can be used to derive an upper bound for the number of iterations sufficient to ensure with certain probability that a GA selects a globally optimal solution upon termination.

Introduction and Summary

Random noise perturbs objective functions in many practical problems, and genetic algorithms (GAs) have been widely proposed as an effective optimization tool for dealing with noisy objective functions (e.g., [**??**], [**??**]). Theoretical studies that examine evolutionary computation schemes applied to perturbed fitness functions typically assume that fitness functions are disturbed by a single source of additive noise (e.g., [**??**], [**??**], [**??**], [**??**], [**??**], [**??**]). In this study, we examine GAs applied to fitness functions perturbed concurrently by multiple sources of noise. In many practical optimization problems, objective functions may be perturbed by more than one noise source. For example, if objective function values are measured by a device that consists of multiple components, then each component may disturb objective function evaluations. Clearly, disturbance by multiple sources of noise is in general stochastically quite different from that by a single source of noise.

In this study, we focus on a noisy environment where fitness functions are perturbed concurrently by multiple sources of additive and multiplicative noise. To our

[1]Department of Applied Mathematics and Statistics, The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218 USA

knowledge, this study is the first to investigate the transition and convergence properties of GAs in this rather complex noisy environment. For analytical tractability, we assume that additive and multiplicative noise sources each take on finitely many values and that they independently disturb fitness function evaluations. However, we do not assume that they are identically distributed, and neither do we make any assumptions about their expected values or variances. We fully characterize this noisy environment in Section 3.

We take a novel approach to investigating GAs in noisy environments; we explicitly construct a Markov chain that models the evolution of GAs applied to perturbed fitness functions and analyze the chain to investigate the transition and convergence properties of the algorithms (see [**??**] and [**??**]). In Section 5, we show that the Markov chain that models GAs with multiple sources of additive and multiplicative noise is indecomposable; it has only one positive recurrent communication class. See Theorem 2. Using this property, we establish a condition (Condition 1) that is both necessary and sufficient for GAs to eventually find at least one globally optimal solution with probability 1. This is Theorem 3. Furthermore, in Theorem 4, we identify a condition (Condition 2) that is both necessary and sufficient for GAs to eventually with probability 1 *fail* to find any globally optimal solution. We will show that, interestingly, both of these essential conditions are completely determined by the fitness function and multiplicative noise; they are unaffected by the additive noise. Theorem 5 states that the chain has a stationary distribution that is also its steady-state distribution, and we will discuss how this property can be used to provide an upper bound for the number of iterations sufficient to ensure with certain probability that a GA selects a globally optimal solution upon termination.

Space limitations on this paper do not allow us to present the complete proofs of these theorems; we will provide them in our full-length paper.

Preliminaries

We assume that, as in traditional GAs, the search space S consists of 2^L binary strings of length L. These 2^L candidate solutions are also referred to as chromosomes. Since the search space is finite, these chromosomes will be labeled by integers $1,\ldots,2^L$. Let f denote the (noiseless) fitness function, and let S^* denote the set of chromosomes that are globally optimal solutions: $S^* := \{i \in S \,|\, f(i) = \max_{j\in S} f(j)\}$. Then the objective of GAs is to find $i \in S^*$.

We consider executing GAs with the elitist strategy. This guarantees that the best candidate solution in the current population, which is a chromosome with the highest *observed* fitness value, is included in the next population. Hence it is important to realize that the elitist strategy guarantees the monotonic improvement of (noisy) *observed* fitness but may fail to monotonically improve (noiseless) fitness. We also assume that GAs reevaluate the fitness value of each population member except for that of the chromosome preserved by the elitist strategy every time a new population is formed. These two strategies turn out to be essential for ensuring that GAs eventually find a globally optimal solution in noisy environments.

When GAs terminate, note that they select a chromosome that has the highest *observed* fitness value among the chromosomes in the last population as a (candidate for a) globally optimal solution. Since the observed fitness value may not be the same as the (noiseless) fitness value of the chromosome due to noise, GAs may not choose a globally optimal solution even if it is included in the last population; in order for GAs to correctly identify a globally optimal solution contained in the last population, the chromosome must have the highest observed fitness value. This observation is important for properly characterizing GAs in noisy environments.

Mathematical Details of Noisy Fitness and Genetic Operations

We will follow and extend the notation developed by Nakama [**??**], who established a Markov chain framework for analyzing GAs in noisy environments. First, we select M chromosomes from S to form an initial population $\mathscr{P}_0$. The population generated during the k-th iteration (the k-th population) will be denoted by $\mathscr{P}_k$, and the number of instances of chromosome i included in the k-th population $\mathscr{P}_k$ will be denoted by $m(i, \mathscr{P}_k)$. We let $i(j, \mathscr{P}_k)$ represent the j-th instance of chromosome i in $\mathscr{P}_k$ (thus $1 \leq j \leq m(i, \mathscr{P}_k)$). We need this notation because it is necessary to distinguish all the elements of the multiset $\mathscr{P}_k$ in order to precisely characterize the mathematical properties of noise considered in this study and to define the states of the Markov chain we construct in Section 4.

Using this notation, we can mathematically describe the noisy environment considered in this study. At each iteration, we evaluate the fitness value of each chromosome in the population. We suppose that multiple sources of additive and multiplicative noise concurrently disturb each fitness function evaluation. Let $F(i(j, \mathscr{P}_k))$ denote the resulting noisy observed fitness of chromosome $i(j, \mathscr{P}_k)$ (the j-th instance of chromosome i in $\mathscr{P}_k$). Since the noiseless fitness $f(i)$ of this chromosome is perturbed by multiple sources of additive and multiplicative noise, its observed fitness $F(i(j, \mathscr{P}_k))$ can be written as

$$F(i(j, \mathscr{P}_k)) = \left(\sum_{d=1}^{D_X} X_{i(j,\mathscr{P}_k)}^{(d)} \right) f(i) + \sum_{d=1}^{D_Y} Y_{i(j,\mathscr{P}_k)}^{(d)}, \tag{1}$$

where the D_X random variables $X_{i(j,\mathscr{P}_k)}^{(1)}, \ldots, X_{i(j,\mathscr{P}_k)}^{(D_X)}$ and the D_Y random variables $Y_{i(j,\mathscr{P}_k)}^{(1)}, \ldots, Y_{i(j,\mathscr{P}_k)}^{(D_Y)}$ represent the multiplicative and additive noise sources, respectively, that concurrently perturb the fitness value $f(i)$ of $i(j, \mathscr{P}_k)$. In many practical optimization problems, objective functions may be perturbed by more than one noise source, and we will consider this rather complex noisy environment.

For analytical simplicity, we assume that each of these $D_X + D_Y$ noise sources is discrete and takes on finitely many values. Thus for each $i(j,\mathscr{P}_k)$ and d, we define the distribution of the d-th multiplicative noise source $X^{(d)}_{i(j,\mathscr{P}_k)}$ by

$$X^{(d)}_{i(j,\mathscr{P}_k)} = x^{(d)}_n \text{ with probability } p^{(d)}_n, \;\; 1 \le n \le N^{(d)}_X, \tag{2}$$

where $N^{(d)}_X$ represents the number of distinct possible values of $X^{(d)}_{i(j,\mathscr{P}_k)}$. Similarly, for each $i(j,\mathscr{P}_k)$ and d, we define the distribution of the d-th additive noise source $Y^{(d)}_{i(j,\mathscr{P}_k)}$ by

$$Y^{(d)}_{i(j,\mathscr{P}_k)} = y^{(d)}_n \text{ with probability } q^{(d)}_n, \;\; 1 \le n \le N^{(d)}_Y,$$

where $N^{(d)}_Y$ represents the number of distinct possible values of $Y^{(d)}_{i(j,\mathscr{P}_k)}$. We will also assume that these random variables are mutually independent. However, we do not assume that the $D_X + D_Y$ noise sources are identically distributed, and neither do we make any assumptions about their expected values or variances—our analysis is valid even if their expectations are nonzero, for example.

Let $i^*(j^*,\mathscr{P}_k)$ represent an instance of a chromosome in $\mathscr{P}_k$ that has the highest observed fitness value in $\mathscr{P}_k$. Then the elitist strategy guarantees the inclusion of $i^*(j^*,\mathscr{P}_k)$ in the next population $\mathscr{P}_{k+1}$. Hence GAs perform selection, crossover, and mutation to determine the other $M-1$ members of the next population. Selection is performed to form $\frac{M-1}{2}$ pairs (thus M is assumed to be odd), and crossover is performed to generate $M-1$ chromosomes from these pairs. Our analysis is valid for any selection scheme and any crossover scheme. Finally, mutation completes the formation of the new population $\mathscr{P}_{k+1}$. The elitist strategy prevents this genetic operation from altering $i^*(j^*,\mathscr{P}_k)$. For each of the other $M-1$ chromosomes, we suppose that mutation inverts each bit with some predetermined probability μ. We assume $0 < \mu < 1$.

GAs evaluate the fitness of each chromosome in $\mathscr{P}_{k+1}$ except for the chromosome $i^*(j^*,\mathscr{P}_k)$ included in this new population by the elitist strategy. These steps are repeated until a stopping criterion is satisfied.

Framework of Markov Chain Analysis

For a Markov chain constructed to model GAs in the noiseless case, the state space is typically the set of all possible distinct populations that can be formed from S. However, this Markov chain fails to explicitly capture the evolution of GAs in noisy environments. Instead, we construct a Markov chain, call it (Z_k), whose state space consists of multisets not of chromosomes but of the *ordered* $(D_X + D_Y + 1)$-*tuples* defined below.

For each iteration of GAs in the noisy environment described in Section 3, the state of the chain (Z_k) that correspond to a population $\mathscr{P}$ can be derived as follows. We form M ordered $(D_X + D_Y + 1)$-tuples from the M chromosomes in $\mathscr{P}$ by pairing each chromosome $i(j,\mathscr{P})$ with the values of the multiplicative and additive noise sources that are each observed when the fitness of $i(j,\mathscr{P})$ is evaluated [see (1)]. We denote the resulting ordered $(D_X + D_Y + 1)$-tuple by

$$\left[i, X^{(1)}_{i(j,\mathscr{P})}, \ldots, X^{(D_X)}_{i(j,\mathscr{P})}, Y^{(1)}_{i(j,\mathscr{P})}, \ldots, Y^{(D_Y)}_{i(j,\mathscr{P})}\right], \tag{3}$$

where j and $\mathscr{P}$ are suppressed in the first entry because they are unnecessary. The resulting M ordered $(D_X + D_Y + 1)$-tuples compose a state of the Markov chain for the noisy case. Thus each state of the chain is a multiset of the ordered $(D_X + D_Y + 1)$-tuples. It is important to recognize that in order to explicitly model the evolution of GAs in the noisy environment using any Markov chain, we need the last $(D_X + D_Y)$ entries of each ordered $(D_X + D_Y + 1)$-tuple in (3).

We analyze the Markov chain (Z_k) to uncover the transition and convergence properties of GAs in the noisy environment. We denote by $\mathfrak{T}$ the state space of (Z_k). Let $m(i,\mathscr{T})$ denote the number of instances of chromosome i included in the ordered $(D_X + D_Y + 1)$-tuples of $\mathscr{T} \in \mathfrak{T}$ (thus $m(i,\mathscr{T})$ is analogous to $m(i,\mathscr{P}_k)$ defined at the beginning of Section). We also denote by $m(x_n^{(d)},\mathscr{T})$ the number of instances of the value $x_n^{(d)}$ of the d-th multiplicative noise source contained in the ordered $(D_X + D_Y + 1)$-tuples of $\mathscr{T} \in \mathfrak{T}$. Similarly, we denote by $m(y_n^{(d)},\mathscr{T})$ the number of instances of the value $y_n^{(d)}$ of the d-th additive noise source contained in the ordered $(D_X + D_Y + 1)$-tuples of $\mathscr{T} \in \mathfrak{T}$. Note that for each $\mathscr{T} \in \mathfrak{T}$ and d,

$$\sum_{i=1}^{2^L} m(i,\mathscr{T}) = \sum_{n=1}^{N_X^{(d)}} m(x_n^{(d)},\mathscr{T}) = \sum_{n=1}^{N_Y^{(d)}} m(y_n^{(d)},\mathscr{T}) = M.$$

In order to simplify expressions for the one-step transition probabilities of the chain (Z_k), we let $\mathscr{C}(\mathscr{T})$ denote the set of chromosomes contained in the ordered pairs of $\mathscr{T} \in \mathfrak{T}$. Thus $\mathscr{C}(\mathscr{T})$ represents the population component of $\mathscr{T}$.

We are now ready to precisely characterize the transitions of the Markov chain (Z_k). The following theorem shows the one-step transition probabilities of (Z_k). It turns out that these transition probabilities can be used to bound the number of iterations sufficient to ensure with certain probability that a GA selects a globally optimal solution upon termination.

Theorem 1 *Let (Z_k) denote the Markov chain with state space $\mathfrak{T}$ that models GAs in the noisy environment. Let $\mathscr{T}$ and $\mathscr{T}'$ denote states in $\mathfrak{T}$, and let $i^*(\mathscr{T})$ denote a chromosome in an ordered $(D_X + D_Y + 1)$-tuple of $\mathscr{T} \in \mathfrak{T}$ that has the highest* *observed fitness value. If the observed fitness value of $i^*(\mathscr{T}')$ is greater than or equal to that of $i^*(\mathscr{T})$, then for each k,*

$$\begin{aligned} P\{Z_{k+1} = \mathscr{T}' | Z_k = \mathscr{T}\} &= P\{\mathscr{C}(Z_{k+1}) = \mathscr{C}(\mathscr{T}') | Z_k = \mathscr{T}\} \\ &\times \prod_{d=1}^{D_X} (M-1)! \prod_{n=1}^{N_X^{(d)}} \frac{1}{\widetilde{m}(x_n^{(d)}, \mathscr{T}')!} (p_n^{(d)})^{\widetilde{m}(x_n^{(d)}, \mathscr{T}')} \\ &\times \prod_{d=1}^{D_Y} (M-1)! \prod_{n=1}^{N_Y^{(d)}} \frac{1}{\widetilde{m}(y_n^{(d)}, \mathscr{T}')!} (q_n^{(d)})^{\widetilde{m}(y_n^{(d)}, \mathscr{T}')}, \end{aligned}$$

where

$$\widetilde{m}(x_n^{(d)}, \mathscr{T}') = \begin{cases} m(x_n^{(d)}, \mathscr{T}') - 1 & \text{if } i^*(\mathscr{T}) \text{ is paired with } x_n^{(d)} \text{ in } \mathscr{T} \text{ (and in } \mathscr{T}') \\ m(x_n^{(d)}, \mathscr{T}') & \text{otherwise,} \end{cases}$$

and

$$\widetilde{m}(y_n^{(d)}, \mathscr{T}') = \begin{cases} m(y_n^{(d)}, \mathscr{T}') - 1 & \text{if } i^*(\mathscr{T}) \text{ is paired with } y_n^{(d)} \text{ in } \mathscr{T} \text{ (and in } \mathscr{T}') \\ m(y_n^{(d)}, \mathscr{T}') & \text{otherwise.} \end{cases}$$

On the other hand, if the observed fitness value of $i^*(\mathscr{T}')$ is less than that of $i^*(\mathscr{T})$, then

$$P\{Z_{k+1} = \mathscr{T}' | Z_k = \mathscr{T}\} = 0.$$

It is straightforward to prove this theorem by realizing that given Z_k, the numbers of instances of multiplicative or additive noise values in Z_{k+1} are each multinomially distributed.

Convergence Analysis

Using the Markov chain (Z_k) constructed in Section 4, we analyze the convergence properties of GAs applied to fitness functions perturbed concurrently by the additive and multiplicative noise sources described in Section 3. First we state one of the most fundamental properties of the chain:

Theorem 2 *The Markov chain (Z_k) is indecomposable: It has only one positive recurrent communication class.*

It is easy to prove this theorem using Theorem 1, and Theorems 3–5 described below can be easily proved using Theorem 2.

We will let $\mathfrak{X}^{(d)}$ denote the set of all possible values of the d-th multiplicative noise source. Hence $|\mathfrak{X}^{(d)}| = N_X^{(d)}$ (see (2)). The following is one of the essential conditions for characterizing the convergence properties of GAs in the noisy environment with multiple sources of additive and multiplicative noise:

Condition 1. The maximum of the multiplicatively perturbed fitness value of globally optimal solutions is greater than that of any suboptimal solution:

$$\max\left\{\left(\sum_{d=1}^{D_X} x^{(d)}\right) f(i) \mid x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D_X, i \in S^*\right\}$$
$$> \max\left\{\left(\sum_{d=1}^{D_X} x^{(d)}\right) f(j) \mid x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D_X,\ j \in S \setminus S^*\right\}.$$

Here define the right-hand side of this inequality to be $-\infty$ for the trivial case that $S^* = S$.

Notice that Condition 1 is completely determined by the values of the fitness function and multiplicative noise; it is independent of the additive noise. We are now ready to state the following theorem:

Theorem 3 *Suppose that GAs are executed in the noisy environment as described in Sections 2–3. Then Condition 1 is necessary and sufficient for GAs to eventually (i.e., as the number of iterations goes to infinity) find at least one globally optimal solution with probability 1.*

This theorem simply follows from Theorem 2; we can prove it by identifying and examining the states of the positive recurrent communication class of the chain (Z_k). Theorem 3 shows that we can verify Condition 1 in order to determine whether GAs in the noisy environment eventually find a globally optimal solution with probability 1. We can also identify a condition that is both necessary and sufficient for GAs to *eventually fail* with probability 1. First we define the eventual failure mathematically.

Definition 1 (Eventual Failure). GAs are said to *eventually fail* if, after some (random) finite number of iterations, they include in each subsequent population at least one suboptimal solution whose observed fitness value is greater than that of any globally optimal solution. In other words, GAs are said to eventually fail if, after sufficiently many iterations, they are guaranteed to never find any globally optimal solution.

The following is another essential condition for characterizing the convergence properties of GAs in the noisy environment with multiple sources of additive and multiplicative noise:

Condition 2. The maximum of the multiplicatively perturbed fitness value of a suboptimal solution is greater than that of globally optimal solutions:

$$\max\left\{\left(\sum_{d=1}^{D_X} x^{(d)}\right) f(i) \mid x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D_X, i \in S \setminus S^*\right\}$$
$$> \max\left\{\left(\sum_{d=1}^{D_X} x^{(d)}\right) f(j) \mid x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D_X,\ j \in S^*\right\}.$$

Notice that Condition 2 is again completely determined by the fitness function and multiplicative noise and is unaffected by the additive noise. We are now in position to state the following theorem:

Theorem 4 *Suppose that GAs are executed in the noisy environment as described in Sections 2–3. Then Condition 2 is necessary and sufficient for GAs to eventually fail with probability 1.*

We can prove this theorem by again identifying and examining the states of the positive recurrent communication class of the Markov chain (Z_k).

The next theorem follows from Theorem 2 and ensures the convergence of the chain (Z_k) to stationarity regardless of whether Condition 1 (or Condition 2) holds or not.

Theorem 5 *The Markov chain (Z_k) has a unique stationary distribution that is also its steady-state distribution: There exists a unique distribution π on $\mathfrak{T}$ such that $\pi = \pi K$, where K is the $|\mathfrak{T}| \times |\mathfrak{T}|$ transition kernel of (Z_k), and for any states $\mathscr{T}$ and $\mathscr{T}'$ in $\mathfrak{T}$, we have $\pi(\mathscr{T}) = \lim_{k\to\infty} P\{Z_k = \mathscr{T} | Z_0 = \mathscr{T}'\}$.*

Let $\mathfrak{A}$ denote the positive recurrent communication class of the chain. Then the stationary distribution π in Theorem 5 satisfies $\pi(\mathscr{T}) > 0$ for each $\mathscr{T} \in \mathfrak{A}$ and $\pi(\mathscr{T}) = 0$ for each $\mathscr{T} \in \mathfrak{T} \setminus \mathfrak{A}$. Therefore, if we let $\pi^{(k)}$ denote the distribution of the chain (Z_k) at time k, then it follows from Theorem 5 that $\lim_{k\to\infty} \sum_{\mathscr{T}\in\mathfrak{A}} \pi^{(k)}(\mathscr{T}) = 1$. Hence if Condition 1 holds, then we need to analyze the rate of this convergence in order to determine how many iterations are sufficient to guarantee with certain probability that a GA selects a globally optimal solution upon termination. It turns out that the transition probabilities of the chain (Z_k) shown in Theorem 1 can be used to determine the convergence rate. Unfortunately, space limitations on this paper force us to omit this analysis here; we will describe it in our full-length paper.

Discussion

We believe that this study is the first to rigorously examine the transition and convergence properties of GAs applied to fitness functions perturbed concurrently by multiple sources of additive and multiplicative noise. Our novel Markov chain analysis successfully shows that if the noise sources each take on finitely many values, then Condition 1 is both necessary and sufficient for GAs in the noisy environment to eventually (i.e., as the number of iterations goes to infinity) find at least one globally optimal solution with probability 1. Furthermore, our analysis proves that Condition 2 is both necessary and sufficient for the GAs to eventually with probability 1 fail to find any globally optimal solution. As described in Section 5, both Conditions 1 and 2 are completely determined by the fitness function and multiplicative noise and are unaffected by the additive noise. The noisy environment considered in this study is relatively general—for instance, we did not make any assumptions about the expected values or variances of the noise sources.

We are currently extending our Markov chain analysis to other noisy environments. For example, we are investigating the transition and convergence properties of GAs with fitness functions perturbed by more general discrete noise or by continuous noise. We believe that our Markov-chain-theoretic approach to analyzing GAs in noisy environments will further elucidate essential properties of the algorithms.

*

References

[1] D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, Boston, 2002.

[2] H. G. Beyer. Evolutionary algorithms in noisy environments: theoretical issues and guidelines for practice. *Computer Methods in Applied Mechanics and Engineering*, 186:239–267, 2000.

[3] A. Chen, K. Subprasom, and Z. Ji. A simulation-based multi-objective genetic algorithm (SMOGA) procedure for BOT network design problem. *Optimization and Engineering*, 7:225–247, 2006.

[4] A. Di Pietro, L. White, and L. Barone. Applying evolutionary algorithms to problems with noisy, time-consuming fitness functions. In *Proceedings of the 2004 Congress on Evolutionary Computation*, volume 2, pages 1254– 1261, 2004.

[5] D. E. Goldberg and M. W. Rudnick. Genetic algorithms and the variance of fitness. *Complex Systems*, 5:265–278, 1991.

[6] Y. Jin and J. Branke. Evolutionary optimization in uncertain environments—a survey. *IEEE Transactions on Evolutionary Computation*, 3:303–317, 2005.

[7] T. Nakama. Markov chain analysis of genetic algorithms applied to fitness functions perturbed by multiple sources of additive noise. *Studies in Computational Intelligence*, 149:123–136, 2008.

[8] T. Nakama. Theoretical analysis of genetic algorithms in noisy environments based on a markov model. In *Proceedings of the 2008 Genetic and Evolutionary Computation Conference*, pages 1001– 1008, 2008.

[9] V. Nissen and J. Propach. On the robustness of population-based versus point-based optimization in the presence of noise. *IEEE Transactions on Evolutionary Computation*, 2:107–119, 1998.