Copyright © 2009 Tech Science Press CMC, vol.14, no.1, pp.35-60, 2009

Convergence Properties of Genetic Algorithms in a Wide Variety of Noisy Environments

Takehiko Nakama[1]

Abstract: Random noise perturbs objective functions in practical optimization problems, and genetic algorithms (GAs) have been proposed as an effective optimization tool for dealing with noisy objective functions. In this paper, we investigate GAs in a variety of noisy environments where fitness perturbation can occur in any form—for example, fitness evaluations can be concurrently disturbed by additive and multiplicative noise. 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. We compute the one-step transition probabilities of the Markov chain and show that the chain has only one positive recurrent communication class, which is also aperiodic. Based on 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. We also identify a condition that is both necessary and sufficient for GAs to eventually with probability 1 fail to find any globally optimal solution. Furthermore, in all the noisy environments, our analysis shows that the chain has a stationary distribution that is also its steady-state distribution. Based on this property and the transition probabilities of the chain, we examine the number of iterations sufficient to ensure with at least a specified probability that GAs select a globally optimal solution upon termination.

Keywords: Evolutionary computation, genetic algorithms, Markov chain analysis, noisy (uncertain) environments, perturbed fitness functions, additive noise, multiplicative noise, convergence.

1 Introduction and Summary

Objective functions are perturbed by random noise in many practical problems, and genetic algorithms (GAs) have been proposed as an effective optimization

---

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

tool for dealing with noisy objective functions [e.g., Beyer (2000), Arnold (2002), Amaziane, Naji, Ouazar and Cheng (2005), Devroy and Krzyzak (2005), Chen, Subprasom and Ji (2006), Sinha and Ch (2008), Galán Moreno, Sánchez Medina, Álvarez Álvarez and Rubio Royo (2009)]. However, we have yet to fully establish the theoretical foundation of the algorithms in noisy environments. Quite a few studies [e.g., Goldberg and Rudnick (1991), Miller and Goldberg (1996), Nissen and Propach (1998), Beyer (2000), Arnold (2002), Di Pietro, White and Barone (2004)] have examined GAs and more general evolutionary algorithms in noisy environments using either numerical or other theoretical methods, but Markov chain analysis has not been applied to such GAs. [For a survey of the literature, we refer the reader to Beyer (2000) and Jin and Branke (2005).] This is quite contrastive to the noiseless case; Markov chain theory has been effectively used to reveal important properties of GAs applied to noiseless fitness functions [e.g., Vose and Liepins (1991), Nix and Vose (1992), Davis and Principe (1993), Rudolph (1994), Suzuki (1995), Vose (1999)]. In this study, we take a first step towards theoretically analyzing GAs in noisy environments using Markov chain theory.

We will address a wide variety of noisy environments where fitness perturbation can occur in any form. For instance, fitness functions can be concurrently disturbed by additive and multiplicative noise—see various examples in Section 3, which clearly illustrate the generality of the noisy environments considered in this study. To our knowledge, this study is the first to address GAs with such a wide variety of fitness perturbation. Except for Nakama (2008), previous studies that examined evolutionary computation schemes applied to perturbed fitness functions all assumed that a single source of additive noise disturbs fitness functions [e.g., Goldberg and Rudnick (1991), Miller and Goldberg (1996), Nissen and Propach (1998), Beyer (2000), Arnold (2002), Di Pietro, White and Barone (2004), Devroy and Krzyzak (2005)]. Although Nakama's study is the first to theoretically examine the convergence properties of GAs in a multiplicatively noisy environment, multiplicative noise has been considered in a variety of optimization problems, such as control optimization [e.g., Hopkins (1987)], image restoration [e.g., Rudin and Osher (2004)], and portfolio optimization [e.g., Primbs (2007)]. In many optimization problems, objective functions may be perturbed by more than one noise source. For instance, if objective function values are measured by a device that consists of multiple components, then each of the components may disturb objective function evaluations. Similarly, if objective function evaluations are conducted through several stages, then they may be perturbed differently at different stages. Clearly, disturbance by multiple sources of noise is in general stochastically quite different from that by a single source of noise. We will demonstrate that our Markov chain framework allows us to uncover several fundamental properties of GAs in noisy environments with multiple noise sources.

For analytical simplicity, we assume that each noise source takes on finitely many values. However, this may not be an unrealistic assumption since one can reasonably argue that, due to limited accuracy and range of measuring devices, observed noise in practice cannot be continuous and can take on only finitely many values. We do not make any assumptions about the expected values or variances of the noise sources. Thus our analysis is valid even if their expected values are nonzero. In previous studies [e.g., Goldberg and Rudnick (1991), Miller and Goldberg (1996), Nissen and Propach (1998), Beyer (2000), Arnold (2002), Di Pietro, White and Barone (2004)], noise was typically assumed to have mean zero [however, see Devroy and Krzyzak (2005) and Nakama (2008)]. In Section 3, we mathematically characterize the various noisy environments considered in this study.

GAs in our study are assumed to implement the elitist strategy, which is in general necessary for guaranteeing that the algorithms eventually (i.e., as the number of iterations goes to infinity) find a globally optimal solution with probability 1 in noiseless environments [Rudolph (1994)]. It will become clear that the elitist strategy is necessary for maintaining this convergence property of GAs in certain noisy environments. Another important strategy is that GAs reevaluate the fitness of each population member except for that of the chromosome preserved by the elitist strategy every time a new population is formed. This strategy can be described as one of "forgetting techniques" often employed in noisy environments [e.g., Di Pietro, White and Barone (2004)], and it is also essential for guaranteeing that GAs eventually find a globally optimal solution with probability 1 in certain noisy cases. The reassessment of the fitness value also significantly reduces memory requirements compared to the strategy of storing a fitness value for every chromosome that has been included in some population formed since the beginning of the execution. Sections 2 and 3 fully describe the implementation of GAs considered in this study.

In Section 4, we explicitly construct a Markov chain that models GAs in the wide range of noisy environments described in Section 3. Mathematically, this Markov chain is completely determined by its transition kernel, which we compute exactly; see Theorem 4.3. It turns out that these probabilities can be used to examine the number of iterations sufficient to ensure with at least a specified probability that a GA selects a globally optimal solution upon termination. In Section 5, we describe two essential properties of the chain: indecomposability and convergence to stationarity. (Note that indecomposability is not the same as irreducibility, and the chain is *not* irreducible in general.) In Section 6, we demonstrate how the indecomposability of the chain can be used to uncover the convergence properties of GAs in various noisy environments. First we establish a condition (Condition 6.1) that is necessary and sufficient for GAs to eventually find a globally optimal solution with probability 1. This is Theorem 6.2. We also identify a condition (Condition 6.8) that is necessary and sufficient for GAs to eventually fail to find any globally optimal solution. See Theorem 6.9.

Based on the chain's one-step transition probabilities and convergence to stationarity, we analyze how many iterations are sufficient to guarantee with at least a specified probability that GAs in noisy environments select a globally optimal solution upon completion of each iteration. This analysis is described in Section 7.

2 Preliminaries

Let S denote the search space, and let $f : S \to \mathbb{R}$ denote the (noiseless) fitness function. Then the objective of GAs is to find an i belonging to $S^* := \{i \in S \,|\, f(i) = \max_{j \in S} f(j)\}$, which is the set of globally optimal solutions. We assume that, as in traditional GAs, 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 $1, \ldots, 2^L$.

For concreteness, we consider the following simple implementation of GAs:

1. An initial population of M chromosomes is formed, and the fitness of each of the chromosomes is evaluated.
2. A chromosome with the best observed fitness value is selected, and the elitist strategy includes this chromosome in the next population.
3. Selection is performed to form $\frac{M-1}{2}$ pairs of chromosomes (M is assumed to be odd and greater than 1).
4. Crossover is performed on each of the $\frac{M-1}{2}$ pairs formed by selection to generate $M-1$ offspring.
5. Mutation is performed on each of the $M-1$ chromosomes generated by crossover. This completes the formation of the next population.
6. The fitness of each new chromosome is determined. If the new population satisfies a stopping criterion, then the algorithm terminates. Otherwise, steps 2-6 are repeated.

GAs are assumed to implement the elitist strategy. This strategy guarantees that the best candidate solution in the current population, which is a chromosome with the best *observed* fitness, is preserved in the next population. In order to monotonically improve the best observed fitness over time, the preserved chromosome retains its observed fitness (see Example 3.9). It is important to note that in noisy environments, this evolutionary strategy may fail to monotonically improve (noiseless) fitness, which cannot be observed directly in noisy environments.

Another important strategy is that in step 6, GAs (re)evaluate the fitness of each population member except for that of the chromosome preserved by the elitist strategy every time a new population is formed. It will become clear that this reevaluation of fitness is also essential for ensuring that GAs eventually find at least one globally optimal solution with probability 1 in certain noisy environments.

When GAs terminate, note that they select a chromosome with the best *observed* fitness among the chromosomes in the last population as a (candidate for a) globally optimal solution. Since the observed fitness may not be the same as the (noiseless) fitness 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 best observed fitness. This observation is important for properly characterizing GAs in noisy environments.

In Section 3, we specify the details of the genetic operations mentioned above and mathematically describe the various noisy environments considered in this study. We construct a Markov chain that models these GAs and derive the one-step transition probabilities of the chain in Section 4.

3 Mathematical Descriptions of Noisy Environments and Genetic Operations

We will closely follow and extend the notation firmly established by the previous studies that conducted Markov chain analysis of GAs in noiseless environments [e.g., Vose and Liepins (1991), Nix and Vose (1992), Davis and Principe (1993), Rudolph (1994), Suzuki (1995), Vose (1999)]. Our notation inevitably becomes complicated due to the generality of the noisy environments considered in this study (see Examples 3.2–3.6). We will use simple examples to explain our notation.

First, we select M chromosomes from S *with replacement* to form an initial population $\mathscr{P}_0$. The population generated during the k-th iteration (the k-th population) will be denoted by $\mathscr{P}_k$. Note that $\mathscr{P}_k$ is a multiset of chromosomes for each k. Let $m(i, \mathscr{P}_k)$ denote the number of instances of chromosome i included in the k-th population $\mathscr{P}_k$. Let $i(j, \mathscr{P}_k)$ represent the j-th instance of chromosome i in $\mathscr{P}_k$ [thus $1 \le j \le m(i, \mathscr{P}_k)$]. We need this notation because we must distinguish all the elements in the multiset $\mathscr{P}_k$ in order to mathematically characterize the noisy environments considered in this study and to define the states of the Markov chain constructed in Section 4.

Example 3.1. Consider a search space $S = \{1,2,3,4,5,6,7,8\}$ (thus $L = 3$ for this search space). Suppose that the k-th population $\mathscr{P}_k$ consists of two instances of chromosome 1, one instance of chromosome 6, and two instances of chromosome 8 (hence $M = |\mathscr{P}_k| = 5$). Then $\mathscr{P}_k = \{1,1,6,8,8\}$. When the two instances of chromosome 1 in $\mathscr{P}_k$ must be distinguished, the first instance of 1 will be denoted by $1(1,\mathscr{P}_k)$, and the second instance of 1 will be denoted by $1(2,\mathscr{P}_k)$. When it is necessary to distinguish all the elements in $\mathscr{P}_k$, we write

$$\mathscr{P}_k = \{1(1,\mathscr{P}_k), 1(2,\mathscr{P}_k), 6(1,\mathscr{P}_k), 8(1,\mathscr{P}_k), 8(2,\mathscr{P}_k)\}.$$

We have $m(1,\mathscr{P}_k) = 2$ since there are two instances of chromosome 1 in this population. We also have $m(6,\mathscr{P}_k) = 1$, $m(8,\mathscr{P}_k) = 2$, and $m(i,\mathscr{P}_k) = 0$ for i = 2, 3, 4, 5, and 7. We will refer to this simple example several times to explain our notation.

As mentioned in Section 1, our investigation addresses GAs with any form of fitness perturbation. Let $F(i(j,\mathscr{P}_k))$ denote the (noisy) observed fitness of chromosome $i(j,\mathscr{P}_k)$ (the j-th instance of chromosome i in $\mathscr{P}_k$).

Example 3.2 (additive noise)**.** Suppose that random noise is added upon each fitness evaluation. Then the observed fitness $F(i(j,\mathscr{P}_k))$ of chromosome $i(j,\mathscr{P}_k)$ can be written as

$$F(i(j,\mathscr{P}_k)) = f(i) + X_{i(j,\mathscr{P}_k)}, \tag{1}$$

where $X_{i(j,\mathscr{P}_k)}$ represents the random noise that additively perturbs the fitness $f(i)$ of $i(j,\mathscr{P}_k)$. Note that $f(i(j,\mathscr{P}_k)) = f(i)$ because, as described earlier, $i(j,\mathscr{P}_k)$ denotes the j-th instance of chromosome i in $\mathscr{P}_k$. Except for Nakama (2008), previous theoretical or numerical studies that examined evolutionary computation schemes with perturbed fitness functions assumed that fitness disturbance is caused by a single source of additive noise as in this example [e.g., Goldberg and Rudnick (1991), Miller and Goldberg (1996), Nissen and Propach (1998), Beyer (2000), Arnold (2002), Di Pietro, White and Barone (2004), Devroy and Krzyzak (2005); however, see Nakama (2008)]. Example 6.3 and Corollary 6.4 specifically address the convergence properties of GAs in this additively noisy environment.

Example 3.3 (multiplicative noise)**.** Suppose that random noise multiplicatively disturbs each fitness evaluation. Then the observed fitness $F(i(j,\mathscr{P}_k))$ of chromosome $i(j,\mathscr{P}_k)$ can be written as

$$F(i(j,\mathscr{P}_k)) = X_{i(j,\mathscr{P}_k)} f(i), \tag{2}$$

where $X_{i(j,\mathscr{P}_k)}$ represents the random multiplicative noise that perturbs the fitness $f(i)$ of $i(j,\mathscr{P}_k)$. Here if we do not assume that the multiplicative noise $X_{i(j,\mathscr{P}_k)}$ is positive with probability 1 or that each fitness value $f(i)$ is positive, then we cannot simply take the logarithm of (2). As described in Section 1, multiplicative noise has been considered in a variety of practical optimization problems. Examples 6.5 and 6.10 specifically address the convergence properties of GAs in this multiplicatively noisy environment.

Example 3.4 (concurrent perturbation by additive and multiplicative noise)**.** Suppose that each fitness evaluation is disturbed concurrently by additive and multiplicative noise. Then the observed fitness $F(i(j,\mathscr{P}_k))$ of chromosome $i(j,\mathscr{P}_k)$ can be written as

$$F(i(j,\mathscr{P}_k)) = X_{i(j,\mathscr{P}_k)} f(i) + Y_{i(j,\mathscr{P}_k)}, \tag{3}$$

where the random variables $X_{i(j,\mathscr{P}_k)}$ and $Y_{i(j,\mathscr{P}_k)}$ represent multiplicative noise and additive noise, respectively, that perturb the fitness $f(i)$ of $i(j,\mathscr{P}_k)$. As mentioned in Section 1, objective functions may be perturbed by more than one noise source in many optimization problems, and disturbance by multiple sources of noise is in general stochastically quite different from that by a single source of noise. Examples 6.6 and 6.11 specifically address the convergence properties of GAs in this noisy environment.

Example 3.5 (another form of concurrent perturbation by additive and multiplicative noise)**.** The following is another form of concurrent perturbation by additive and multiplicative noise:

$$F(i(j,\mathscr{P}_k)) = X_{i(j,\mathscr{P}_k)} [f(i) + Y_{i(j,\mathscr{P}_k)}], \tag{4}$$

where the random variables $X_{i(j,\mathscr{P}_k)}$ and $Y_{i(j,\mathscr{P}_k)}$ again represent multiplicative noise and additive noise that disturb the fitness $f(i)$ of $i(j,\mathscr{P}_k)$. In general, this fitness perturbation is stochastically quite different from that shown in Example 3.4. See Example 6.6 for the convergence properties of GAs in this noisy environment.

Example 3.6 (exponential noise)**.** Our analysis is valid even if noise is neither additive nor multiplicative, and fitness perturbation can be of the following form:

$$F(i(j,\mathscr{P}_k)) = [f(i)]^{X_{i(j,\mathscr{P}_k)}}. \tag{5}$$

Here the random noise $X_{i(j,\mathscr{P}_k)}$ "exponentially" disturbs the fitness evaluation. Theorems 6.2 and 6.9, which provide essential characterizations of GAs in noisy environments, even apply to this rather unusual exponentially noisy environment.

In this study, we will construct a Markov chain that explicitly captures the evolution of GAs with any form of fitness disturbance, including the various cases shown in Examples 3.2–3.6. We let D random variables $X^{(1)}_{i(j,\mathscr{P}_k)}, X^{(2)}_{i(j,\mathscr{P}_k)}, \ldots, X^{(D)}_{i(j,\mathscr{P}_k)}$ represent D noise sources that perturb the fitness $f(i)$ of chromosome $i(j,\mathscr{P}_k)$. Thus, as shown in Examples 3.2–3.6, the observed fitness $F(i(j,\mathscr{P}_k))$ of chromosome $i(j,\mathscr{P}_k)$ is a function not only of $f(i)$ but also of the D random variables.

For analytical simplicity, we assume that each of the D noise sources $X^{(1)}_{i(j,\mathscr{P}_k)}, X^{(2)}_{i(j,\mathscr{P}_k)}, \ldots, X^{(D)}_{i(j,\mathscr{P}_k)}$ takes on finitely many values. Thus for each d and $i(j,\mathscr{P}_k)$, we define the distribution of $X^{(d)}_{i(j,\mathscr{P}_k)}$ by

$$X^{(d)}_{i(j,\mathscr{P}_k)} = x^{(d)}_n \text{ with probability } p^{(d)}_n, \quad 1 \le n \le N_d, \tag{6}$$

where N_d represents the number of distinct possible values of the d-th noise source $X^{(d)}_{i(j,\mathscr{P}_k)}$ ($N_d < \infty$ for each d). Here we do not assume that these noise sources are identically distributed. Thus they can have different numbers of possible values. For each d, we let $\mathfrak{X}^{(d)}$ denote the set of possible values of the d-th noise source:

$$\mathfrak{X}^{(d)} = \{x^{(d)}_1, x^{(d)}_2, \ldots, x^{(d)}_{N_d}\}. \tag{7}$$

For analytical simplicity, we assume that these D noise sources are independent. We further assume that for each d, the random variables $X^{(d)}_{i(j,\mathscr{P}_k)}$ are independent and identically distributed for all i, j, and $\mathscr{P}_k$: The d-th noise source is identically distributed for each fitness evaluation and is independent of the d-th noise source (and all the other noise sources) in any other instance of fitness evaluation.

Example 3.7. Consider the following two sources of noise: For each i, j, and $\mathscr{P}_k$,

$$X^{(1)}_{i(j,\mathscr{P}_k)} = \begin{cases} x^{(1)}_1 & \text{with probability } p^{(1)}_1 \\ x^{(1)}_2 & \text{with probability } p^{(1)}_2 = 1 - p^{(1)}_1, \end{cases}$$

and

$$X^{(2)}_{i(j,\mathscr{P}_k)} = \begin{cases} x^{(2)}_1 & \text{with probability } p^{(2)}_1 \\ x^{(2)}_2 & \text{with probability } p^{(2)}_2 \\ x^{(2)}_3 & \text{with probability } p^{(2)}_3 = 1 - p^{(2)}_1 - p^{(2)}_2. \end{cases}$$

We have $N_1 = 2$ and $N_2 = 3$ because $X^{(1)}_{i(j,\mathscr{P}_k)}$ and $X^{(2)}_{i(j,\mathscr{P}_k)}$ take on two and three values, respectively. For each i, j, and $\mathscr{P}_k$, we assume that $X^{(1)}_{i(j,\mathscr{P}_k)}$ and $X^{(2)}_{i(j,\mathscr{P}_k)}$

are independent. Also, we assume that $X^{(1)}_{i(j,\mathscr{P}_k)}$ are independent and identically distributed for all i, j, and $\mathscr{P}_k$, as are $X^{(2)}_{i(j,\mathscr{P}_k)}$.

Let $F^*_{\mathscr{P}_k}$ denote the highest observed fitness achieved by the k-th population $\mathscr{P}_k$, and let $i^*(j^*,\mathscr{P}_k)$ represent an instance of a chromosome in $\mathscr{P}_k$ whose observed fitness equals $F^*_{\mathscr{P}_k}$ (it is the j^*-th instance of chromosome i^* in $\mathscr{P}_k$). Then we have

$$F(i^*(j^*,\mathscr{P}_k)) = F^*_{\mathscr{P}_k} \geq F(i(j,\mathscr{P}_k)) \quad \forall\, i(j,\mathscr{P}_k) \in \mathscr{P}_k. \tag{8}$$

The elitist strategy guarantees the inclusion of $i^*(j^*,\mathscr{P}_k)$ in the next population $\mathscr{P}_{k+1}$. [If $F^*_{\mathscr{P}_k}$ is achieved by more than one chromosome of $\mathscr{P}_k$, then break a tie by selecting one of them uniformly at random to determine $i^*(j^*,\mathscr{P}_k)$.]

Example 3.8. Consider the k-th population $\mathscr{P}_k$ described in Example 3.1. Suppose that the observed fitness values of the five chromosomes in $\mathscr{P}_k$ are as follows:

$$F(1(1,\mathscr{P}_k)) = 8.9, \qquad F(1(2,\mathscr{P}_k)) = 2.5, \qquad F(6(1,\mathscr{P}_k)) = 4.8,$$
$$F(8(1,\mathscr{P}_k)) = 6.6, \qquad F(8(2,\mathscr{P}_k)) = 7.8.$$

Then we have $F^*_{\mathscr{P}_k} = 8.9$ and $i^*(j^*,\mathscr{P}_k) = 1(1,\mathscr{P}_k)$. Thus the elitist strategy guarantees the inclusion of $1(1,\mathscr{P}_k)$ in the next population $\mathscr{P}_{k+1}$.

In order to generate a new population, pairs of chromosomes are formed by selection. In this study, we consider forming $\frac{M-1}{2}$ pairs for concreteness (thus M is assumed to be odd and greater than 1). Selection is affected by fitness perturbation, since it operates on noisy observed fitness values, which affect selection probabilities. Our Markov chain analysis of GAs in noisy environments is valid for any selection scheme—for example, proportional selection, ranking selection, and tournament selection [e.g., Miller and Goldberg (1996), Vose (1999), Leung, Duan, Xu and Wong (2001)].

Crossover and mutation in noisy environments do not differ from those in the noiseless case. Crossover is performed on each of the $\frac{M-1}{2}$ pairs formed by selection in order to generate $M-1$ offspring. Our analysis holds for any crossover scheme. Finally, mutation is performed on each of the $M-1$ offspring. For concreteness, we assume that mutation inverts each bit of an individual chromosome with some probability μ that satisfies $0 < \mu < 1$. The elitist strategy does not allow mutation to alter $i^*(j^*,\mathscr{P}_k)$, which is the chromosome preserved by the elitist strategy.

A new population $\mathscr{P}_{k+1}$ emerges upon completion of selection, crossover, and mutation on $\mathscr{P}_k$, and GAs must determine the fitness of each chromosome in $\mathscr{P}_{k+1}$.

Another important detail for guaranteeing that GAs eventually find a globally optimal solution with probability 1 in certain noisy environments is that in order to monotonically improve the highest observed fitness over time, we let the elitist strategy preserve not only a chromosome with the highest observed fitness but also its observed fitness. This is illustrated by the following example:

Example 3.9. Consider the k-th population $\mathscr{P}_k$ in Example 3.1 and the additively noisy environment in Example 3.2. Suppose that the observed fitness value of $1(1,\mathscr{P}_k)$, the first instance of chromosome 1, is greater than the other four observed fitness values. Then the elitist strategy guarantees the inclusion of $1(1,\mathscr{P}_k)$ (i.e., chromosome 1) in the next population, $\mathscr{P}_{k+1}$. Suppose that after selection, crossover, and mutation are performed, the other four chromosomes of $\mathscr{P}_{k+1}$ turn out to be three instances of chromosome 3 and one instance of chromosome 6: $\mathscr{P}_{k+1} = \{1,3,3,3,6\}$. Again, in order to mathematically describe the evolution of GAs in noisy environments, we distinguish all the chromosomes in $\mathscr{P}_{k+1}$ by expressing the population as

$$\{1(1,\mathscr{P}_{k+1}),3(1,\mathscr{P}_{k+1}),3(2,\mathscr{P}_{k+1}),3(3,\mathscr{P}_{k+1}),6(1,\mathscr{P}_{k+1})\}.$$

In order to monotonically improve the highest observed fitness over time, we let the elitist strategy preserve the observed fitness of chromosome 1. Thus $F(1(1,\mathscr{P}_{k+1}))$ will be set to $F(1(1,\mathscr{P}_k))$. On the other hand, GAs reevaluate the fitness of each of the other four chromosomes in $\mathscr{P}_{k+1}$. Hence the observed fitness values of the five chromosomes in $\mathscr{P}_{k+1}$ are:

$$\begin{aligned}
&F(1(1,\mathscr{P}_{k+1})) = F(1(1,\mathscr{P}_k)), && F(3(1,\mathscr{P}_{k+1})) = f(3)+X_{3(1,\mathscr{P}_{k+1})},\\
&F(3(2,\mathscr{P}_{k+1})) = f(3)+X_{3(2,\mathscr{P}_{k+1})}, && F(3(3,\mathscr{P}_{k+1})) = f(3)+X_{3(3,\mathscr{P}_{k+1})},\\
&F(6(1,\mathscr{P}_{k+1})) = f(6)+X_{6(1,\mathscr{P}_{k+1})},
\end{aligned}$$

Note that although chromosome 6 [i.e., $6(1,\mathscr{P}_k)$] was in $\mathscr{P}_k$, its observed fitness $F(6(1,\mathscr{P}_k))$ is not necessarily preserved in $\mathscr{P}_{k+1}$.

The reevaluation of fitness illustrated in Example 3.9 can be described as a forgetting technique [e.g., Di Pietro, White and Barone (2004)], and it also reduces memory requirements compared to the strategy of storing a fitness value for every chromosome that has been included in some population formed since the beginning of the execution. It will become clear that these strategies are important for guaranteeing that GAs eventually find a globally optimal solution with probability 1 in certain noisy environments.

4 Construction of a Markov Chain that Models Genetic Algorithms in Noisy Environments

A Markov chain we construct to model GAs in noisy environments is fundamentally different from that considered for the noiseless environment in one essential aspect. In the noiseless case, each state of the Markov chain that models GAs is a population, and the k-th population is the state of the chain at time k [for example, see Davis and Principe (1993), Suzuki (1995), Vose (1999)]. However, this Markov chain does not explicitly capture the evolution of GAs in noisy environments. Instead, we consider a Markov chain, call it (Z_k), whose state space consists of multisets not of chromosomes but of the *ordered* $(D+1)$*-tuples* defined below.

The state of the chain (Z_k) that corresponds to the k-th population $\mathscr{P}_k$ can be derived as follows. As described in Section 3, the observed fitness $F(i(j,P_k))$ of chromosome $i(j,P_k)$ is determined by its noiseless fitness $f(i)$ and by the D noise sources $X^{(1)}_{i(j,\mathscr{P})}$, $X^{(2)}_{i(j,\mathscr{P})}, \ldots,$ $X^{(D)}_{i(j,\mathscr{P})}$ that are observed when the fitness of $i(j,\mathscr{P})$ is evaluated. In order to explicitly express this dependence of the observed fitness on the D noise sources, we pair the chromosome $i(j,P_k)$ with the observed noise sources to form the following ordered $(D+1)$-tuple:

$$\left[i, X^{(1)}_{i(j,\mathscr{P}_k)}, X^{(2)}_{i(j,\mathscr{P}_k)}, \ldots, X^{(D)}_{i(j,\mathscr{P}_k)}\right]. \tag{9}$$

Here j and $\mathscr{P}_k$ are suppressed in the first entry because they are unnecessary. In this manner, we form M ordered $(D+1)$-tuples from the M chromosomes in P_k, and the resulting set of M ordered $(D+1)$-tuples is the state of the chain (Z_k) that corresponds to population P_k [thus each state of this chain is a multiset of the ordered $(D+1)$-tuples]. In order to explicitly capture the evolution of GAs in the various noisy environments described in Section 3 using any Markov chain, we need the last D entries of each ordered $(D+1)$-tuple in (9) because the selection process in noisy environments is based on a function not only of chromosomes but also of the D noise sources.

Example 4.1. Consider the search space S and the population $\mathscr{P}_k$ described in Example 3.1; we have $\mathscr{P}_k = \{1,1,6,8,8\}$. This population leads to the following state $\mathscr{T}$ of the Markov chain (Z_k) for noisy environments:

$$\begin{aligned}\mathscr{T} = \Big\{ &\left[1, X^{(1)}_{1(1,\mathscr{P}_k)}, X^{(2)}_{1(1,\mathscr{P}_k)}, \ldots, X^{(D)}_{1(1,\mathscr{P}_k)}\right], \left[1, X^{(1)}_{1(2,\mathscr{P}_k)}, X^{(2)}_{1(2,\mathscr{P}_k)}, \ldots, X^{(D)}_{1(2,\mathscr{P}_k)}\right], \\ &\left[6, X^{(1)}_{6(1,\mathscr{P}_k)}, X^{(2)}_{6(1,\mathscr{P}_k)}, \ldots, X^{(D)}_{6(1,\mathscr{P}_k)}\right], \left[8, X^{(1)}_{8(1,\mathscr{P}_k)}, X^{(2)}_{8(1,\mathscr{P}_k)}, \ldots, X^{(D)}_{8(1,\mathscr{P}_k)}\right], \\ &\left[8, X^{(1)}_{8(2,\mathscr{P}_k)}, X^{(2)}_{8(2,\mathscr{P}_k)}, \ldots, X^{(D)}_{8(2,\mathscr{P}_k)}\right] \Big\}.\end{aligned}$$

Each state of this particular Markov chain is a set consisting of five ordered $(D+1)$-tuples.

Mathematically, this chain is completely determined by its transition kernel. Hence we first derive the one-step transition probabilities of the chain exactly. 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 $(D+1)$-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 3]. Similarly, we denote by $m(x_n^{(d)},\mathscr{T})$ the number of instances of the value $x_n^{(d)}$ of the d-th noise source contained in the ordered $(D+1)$-tuples of $\mathscr{T} \in \mathfrak{T}$. Note that

$$\sum_{i=1}^{2^L} m(i,\mathscr{T}) = \sum_{n=1}^{N_d} m(x_n^{(d)},\mathscr{T}) = M \tag{10}$$

for each $\mathscr{T} \in \mathfrak{T}$ and d.

Example 4.2. Consider the population $\mathscr{P}_k$ described in Example 3.1 and the two sources of noise described in Example 3.7. We have $\mathscr{P}_k = \{1,1,6,8,8\}$ and $D=2$. The following is a possible state $\mathscr{T} \in \mathfrak{T}$ of the Markov chain (Z_k) that models GAs in this noisy case:

$$\mathscr{T} \;=\; \{[1,x_2^{(1)},x_1^{(2)}],[1,x_1^{(1)},x_3^{(2)}],[6,x_1^{(1)},x_3^{(2)}],[8,x_2^{(1)},x_2^{(2)}],[8,x_2^{(1)},x_3^{(2)}]\}.$$

As described earlier, the first entry of each ordered triple is a chromosome. The second entry is the value of the first noise source observed when the fitness of the chromosome is evaluated, and the third entry is the value of the second noise source. Regarding the chromosomes included in this state, we have $m(1,\mathscr{T})=2$ since two instances of chromosome 1 are contained in the ordered triples of $\mathscr{T}$. Similarly, we have $m(6,\mathscr{T})=1$ and $m(8,\mathscr{T})=2$. Regarding the first noise source, we have $m(x_1^{(1)},\mathscr{T})=2$ and $m(x_2^{(1)},\mathscr{T})=3$. For the second noise source, we have $m(x_1^{(2)},\mathscr{T})=1$, $m(x_2^{(2)},\mathscr{T})=1$, and $m(x_3^{(2)},\mathscr{T})=3$. Note that (10) is satisfied with $M=5$.

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 $\mathscr{T}$. [In Example 4.1, we have $\mathscr{C}(\mathscr{T}) = \mathscr{P}_k$.]

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

Theorem 4.3. *Let* (Z_k) *denote the Markov chain with state space* $\mathfrak{T}$ *that models GAs in noisy environments where fitness functions are perturbed by the D sources of noise described in Section 3. Let* $\mathscr{T}$ *and* $\mathscr{T}'$ *denote states in* $\mathfrak{T}$*, and let* $i^*(\mathscr{T})$ *denote a chromosome in an ordered* $(D+1)$*-tuple of* $\mathscr{T}$ *that has the highest observed fitness value. If the observed fitness of* $i^*(\mathscr{T}')$ *is greater than or equal to that of* $i^*(\mathscr{T})$*, then for each k,*

$$P\{Z_{k+1}=\mathscr{T}'|Z_k=\mathscr{T}\}=P\{\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}')|Z_k=\mathscr{T}\}$$
$$\times\prod_{d=1}^{D}(M-1)!\prod_{n=1}^{N_d}\frac{1}{\widetilde{m}(x_n^{(d)},\mathscr{T}')!}(p_n^{(d)})^{\widetilde{m}(x_n^{(d)},\mathscr{T}')}, \tag{11}$$

where for each d,

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

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

$$P\{Z_{k+1}=\mathscr{T}'|Z_k=\mathscr{T}\}=0. \tag{12}$$

Proof of Theorem 4.3. Clearly, the elitist strategy guarantees (12). Suppose that the observed fitness of $i^*(\mathscr{T}')$ is greater than or equal to that of $i^*(\mathscr{T})$. Then for each k, we have

$$\begin{aligned}P\{Z_{k+1}=\mathscr{T}'|Z_k=\mathscr{T}\}&=P\{Z_{k+1}=\mathscr{T}',\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}')|Z_k=\mathscr{T}\}\\ &=P\{\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}')|Z_k=\mathscr{T}\}P\{Z_{k+1}=\mathscr{T}'|\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}'),Z_k=\mathscr{T}\}\\ &=P\{\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}')|Z_k=\mathscr{T}\}P\{Z_{k+1}=\mathscr{T}'|\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}')\}. \end{aligned}\tag{13}$$

Here

$$\begin{aligned}&P\{Z_{k+1}=\mathscr{T}'|\mathscr{C}(Z_{k+1})=\mathscr{C}(\mathscr{T}')\}\\ &\quad= P\{\text{each noise value } x_n^{(d)} \text{ appears } \widetilde{m}(x_n^{(d)},\mathscr{T}') \text{ times in } \mathscr{T}'\}\\ &\quad= \prod_{d=1}^{D}(M-1)!\prod_{n=1}^{N_d}\frac{1}{\widetilde{m}(x_n^{(d)},\mathscr{T}')!}(p_n^{(d)})^{\widetilde{m}(x_n^{(d)},\mathscr{T}')},\end{aligned}\tag{14}$$

because for each d, the random variables $\widetilde{m}(x_n^{(d)},\mathscr{T}')$ are distributed multinomially with parameters $M-1$ and $p_n^{(d)}$ $(n=1,\ldots,N_d)$. Note that $\widetilde{m}$ is used instead of m due to the elitist strategy. Combining (13)–(14), we obtain (11). □

The conditional probability $P\{\mathscr{C}(Z_{k+1}) = \mathscr{C}(\mathscr{T}')|Z_k = \mathscr{T}\}$ in (11) depends on the selection, crossover, and mutation schemes employed by GAs. For example, suppose that proportional selection is used and that crossover is not performed. As described in Section 3, we assume that mutation inverts each bit of an individual chromosome with some probability μ $(0 < \mu < 1)$ independently of other bits. Then we have

$$P\{\mathscr{C}(Z_{k+1}) = \mathscr{C}(\mathscr{T}')|Z_k = \mathscr{T}\} = (M-1)!\prod_{i=1}^{2^L} \frac{1}{\widetilde{m}(i,\mathscr{T}')!}\phi(i,\mathscr{T})^{\widetilde{m}(i,\mathscr{T}')},$$

where

$$\widetilde{m}(i,\mathscr{T}') = \begin{cases} m(i,\mathscr{T}') - 1 & \text{if } i = i^*(\mathscr{T}) \\ m(i,\mathscr{T}') & \text{otherwise,} \end{cases}$$

and

$$\phi(i,\mathscr{T}) = \sum_{j_1 \in \mathscr{C}(\mathscr{T})} \mu^{H(i,j_1)}(1-\mu)^{L-H(i,j_1)} \frac{\sum_{j_2=1}^{m(j_1,\mathscr{T})} F(j_1(j_2,\mathscr{C}(\mathscr{T})))}{\sum_{j_3 \in \mathscr{C}(\mathscr{T})} \sum_{j_2=1}^{m(j_3,\mathscr{T})} F(j_3(j_2,\mathscr{C}(\mathscr{T})))},$$

where $H(i, j_1)$ denotes the Hamming distance between chromosomes i and j_1. Note that Theorem 4.3 and the convergence analysis described in the next section hold for any selection scheme and any crossover scheme.

5 Two Essential Properties of the Markov Chain for Noisy Cases

In this section, we present two fundamental properties of the Markov chain (Z_k) constructed in Section 4. In Sections 6–7, these properties of the chain will be effectively used to uncover convergence properties of GAs in the various noisy environments described in Section 3.

First we establish more notation. Since each ordered $(D+1)$-tuple

$$\left[i, X^{(1)}_{i(j,\mathscr{P}_k)}, X^{(2)}_{i(j,\mathscr{P}_k)}, \ldots, X^{(D)}_{i(j,\mathscr{P}_k)}\right]$$

defined at (9) is uniquely associated with an observed fitness value, namely $F(i(j,P_k))$, we will slightly abuse our notation and let

$$F\left(i, X^{(1)}_{i(j,\mathscr{P}_k)}, X^{(2)}_{i(j,\mathscr{P}_k)}, \ldots, X^{(D)}_{i(j,\mathscr{P}_k)}\right) \tag{15}$$

denote the observed (perturbed) fitness of the ordered $(D+1)$-tuple. It will become clear that this notation facilitates the exposition of our Markov chain analysis.

The following subset $\mathfrak{A}$ of the state space $\mathfrak{T}$ of the chain (Z_k) plays an important role for the two properties of the chain described in this section and for the convergence properties described in Section 6.

Definition 5.1. We define $\mathfrak{A}$ to be the set of all states in $\mathfrak{T}$ that each contain at least one ordered $(D+1)$-tuple

$$[i^*, \widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(D)}] \tag{16}$$

such that

$$F(i^*, \widetilde{x}^{(1)}, \ldots, \widetilde{x}^{(D)}) = \max\{F(i, x^{(1)}, \ldots, x^{(D)}) \,|\, i \in S,\ x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D\}.$$

In other words, $\mathfrak{A}$ is the set of all states that each contain at least one ordered $(D+1)$-tuple whose observed fitness achieves the maximum perturbed fitness.

Example 5.2. Consider the additively noisy environment described in Example 3.2: Each fitness evaluation is perturbed by a single source of additive noise. For this noisy case, $\mathfrak{A}$ is the set of all states that each contain at least one ordered pair $[i^*, \widetilde{x}]$ where i^* is a globally optimal solution and $\widetilde{x}$ denotes the maximum of the possible values of the additive noise.

Example 5.3. Consider the multiplicatively noisy environment described in Example 3.3: Each fitness evaluation is perturbed by a single source of multiplicative noise. Suppose that the fitness value of every chromosome is positive and that the multiplicative noise is positive with probability 1. In this case, $\mathfrak{A}$ is the set of all states that each contain at least one ordered pair $[i^*, \widetilde{x}]$ where i^* is a globally optimal solution and $\widetilde{x}$ denotes the maximum of the possible values of the multiplicative noise. On the other hand, if we suppose that the fitness value of every chromosome is positive and that the multiplicative noise is *negative* with probability 1, then $\mathfrak{A}$ is the set of all states that each contain at least one ordered pair $[i^*, \widetilde{x}]$ where i^* is a chromosome that *minimizes* the fitness function, i.e.,

$$i^* \in \{i \in S \,|\, f(i) \le f(j) \ \ \forall\, j \in S\},$$

and $\widetilde{x}$ denotes the *minimum* of the possible values of the multiplicative noise.

Theorem 4.3 clearly shows that, starting from any state, the Markov chain (Z_k) reaches any state in $\mathfrak{A}$ with probability 1, and once the chain hits $\mathfrak{A}$, it stays there. Thus the chain has the following property:

Property 5.4 (indecomposability)**.** The Markov chain (Z_k) is indecomposable; the set $\mathfrak{A}$ in Definition 5.1 is the only positive recurrent communication class of the chain.

Note that *indecomposability is not the same as ergodicity, and the chain is not ergodic in general.* In fact, those states which do not contain any ordered $(D+1)$ tuple of the form (16) are transient, so the chain is not irreducible in general.

Notice that each state $\mathscr{A}$ in $\mathfrak{A}$ is aperiodic because, by Theorem 4.3, the chain currently in state $\mathscr{A}$ with positive probability stays in $\mathscr{A}$ after one transition. Thus the chain also has the following property:

Property 5.5 (convergence to stationarity)**.** The Markov chain (Z_k) converges to stationarity: It has a unique stationary distribution that is also its steady-state distribution.

See Chapter 3 in Karlin and Taylor (1975) and Chapter 10 in Karlin and Taylor (1981) for details.

In Sections 6–7, these fundamental properties of the chain will be used to uncover the convergence properties of GAs in various noisy environments.

6 Convergence Analysis

In this section, we demonstrate how Property 5.4 can be used to uncover the convergence properties of GAs in the various noisy environments described in Section 3. Based on the property, we derive two fundamental conditions for characterizing GAs in the noisy environments—one of them is 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, and the other is necessary and sufficient for GAs to *eventually fail* with probability 1.

First, we establish a condition that is necessary and sufficient for GAs in the noisy environments to eventually find a globally optimal solution with probability 1. By Property 5.4, GAs eventually find a globally optimal solution with probability 1 if and only if each state in $\mathfrak{A}$ contains at least one ordered $(D+1)$-tuple that includes a globally optimal solution and achieves the maximum perturbed fitness that is greater than the maximum perturbed fitness of any suboptimal solution. Therefore, it follows that GAs in the noisy environments eventually find a globally optimal solution with probability 1 if and only if the following condition holds:

Condition 6.1. The maximum *perturbed* fitness of globally optimal solutions is greater than that of any suboptimal solution:

$$\begin{aligned}&\sup\{F(i,x^{(1)},\ldots,x^{(D)})\,|\,i\in S^*,x^{(d)}\in\mathfrak{X}^{(d)},\ 1\le d\le D\}\\&\quad>\sup\{F(i,x^{(1)},\ldots,x^{(D)})\,|\,i\in S\setminus S^*,x^{(d)}\in\mathfrak{X}^{(d)},\ 1\le d\le D\}.\end{aligned}\tag{17}$$

Therefore, we have just proved the following theorem:

Theorem 6.2. *Suppose that GAs are executed in the noisy environments described in Section 3. Then Condition 6.1 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.*

Theorem 6.2 shows that we can verify Condition 6.1 in order to determine whether GAs in the noisy environments eventually find a globally optimal solution with probability 1.

Example 6.3. Consider the additively noisy environment described in Example 3.2: Each fitness evaluation is perturbed by a single source of additive noise that takes on finitely many values. Let $\widetilde{x}$ denote the maximum of the possible values of the additive noise. Then for any $i\in S^*$ and $j\in S\setminus S^*$,

$$F(i,\widetilde{x})=f(i)+\widetilde{x}>F(j,x)=f(j)+x,$$

where x denotes any possible value of the additive noise. Thus Condition 6.1 is satisfied in this noisy case. Similarly, Condition 6.1 is satisfied when fitness functions are disturbed by multiple sources of additive noise. Hence we obtain the following corollary:

Corollary 6.4. *In noisy environments where fitness functions are purturbed by additive noise sources that each take on finitely many values, GAs eventually find a globally optimal solution with probability 1.*

Example 6.5. Consider the multiplicatively noisy environment described in Example 3.3: Each fitness evaluation is perturbed by a single source of multiplicative noise that takes on finitely many values. Suppose that the (noiseless) fitness value of each chromosome is positive: $f(i)>0$ for all $i\in S$. Then, excluding the trivial case that $S^*=S$, Condition 6.1 is satisfied if and only if the multiplicative noise is greater than zero with positive probability.

Example 6.6. Consider the noisy environment described in Example 3.4: A single source of multiplicative noise and a single source of additive noise concurrently perturb each fitness evaluation. In this case, one can show that Condition 6.1 is equivalent to

$$\sup\{xf(i) \,|\, i \in S^*,\ x \in \mathfrak{X}\} > \sup\{xf(i) \,|\, i \in S \setminus S^*, x \in \mathfrak{X}\}, \tag{18}$$

where $\mathfrak{X}$ denotes the set of possible values of the single multiplicative noise source. Notice that (18) involves only the fitness function and multiplicative noise; it does not involve the additive noise. Thus for this noisy environment, the fitness function and multiplicative noise completely determine whether GAs eventually find a globally optimal solution with probability 1; the convergence is unaffected by the additive noise. However, note that for the other form of concurrent perturbation by additive and multiplicative noise described in Example 3.5, Condition 6.1 is in general affected by additive noise.

Next, we identify a condition that is both necessary and sufficient for GAs in the noisy environments to *eventually fail* with probability 1. First we define the eventual failure mathematically.

Definition 6.7 (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.

By Property 5.4, GAs eventually fail with probability 1 if and only if each state in $\mathfrak{A}$ contains at least one ordered $(D+1)$-tuple that includes a suboptimal solution and achieves the maximum perturbed fitness that is greater than the maximum perturbed fitness of globally optimal solutions. Therefore, it follows that GAs in the noisy environments eventually fail with probability 1 if and only if the following condition holds:

Condition 6.8. There exists a suboptimal solution whose maximum perturbed fitness is greater than the maximum perturbed fitness of globally optimal solutions:

$$\begin{aligned}&\sup\{F([i,x^{(1)},\ldots,x^{(D)}]) \,|\, i \in S \setminus S^*,\ x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D\}\\ &\qquad > \sup\{F([i,x^{(1)},\ldots,x^{(D)}]) \,|\, i \in S^*,\ x^{(d)} \in \mathfrak{X}^{(d)},\ 1 \le d \le D\}.\end{aligned} \tag{19}$$

Thus we have just proved the following theorem:

Theorem 6.9. *Suppose that GAs are executed in the noisy environments described in Section 3. Then Condition 6.7 is necessary and sufficient for GAs to eventually fail with probability 1.*

Example 6.10. Consider the multiplicatively noisy environment in Example 6.5: A single source of multiplicative noise perturbs each fitness evaluation. Suppose that every fitness value is positive and that $S^* \neq S$. Then Condition 6.7 is satisfied if and only if the multiplicative noise is negative with probability 1. Thus it follows by Theorem 6.9 that GAs eventually fail with probability 1 in this case.

Example 6.11. Consider the noisy environment described in Example 3.4: A single source of multiplicative noise and a single source of additive noise concurrently perturb each fitness evaluation. In this case, Condition 6.7 can be expressed as

$$\sup\{xf(i) \,|\, i \in S \setminus S^*,\ x \in \mathfrak{X}\} > \sup\{xf(i) \,|\, i \in S^*, x \in \mathfrak{X}\}, \tag{20}$$

where $\mathfrak{X}$ denotes the set of possible values of the single multiplicative noise source. Note that (20) depends only on the fitness function and multiplicative noise; it is independent of the additive noise. Hence it follows by Theorem 6.9 that the fitness function and multiplicative noise completely determine whether GAs fail with probability 1 in this noisy environment; the failure is unaffected by the additive noise. For example, GAs in this noisy environment fail with probability 1 if (i) every fitness value is positive, (ii) $S \neq S^*$, and (iii) the multiplicative noise is negative with probability 1 (cf. Example 6.10).

Thus we can determine whether GAs in noisy environments eventually find a globally optimal solution or fail with probability 1 by simply identifying the positive recurrent communication class $\mathfrak{A}$ of the chain (Z_k) that models the algorithms. Notice that if GAs do not implement the elitist strategy, then in general $\mathfrak{A}$ may include those states which do not contain any globally optimal solution whose observed fitness is greater than that of any suboptimal solution. Thus the Markov chain analysis clearly shows that without the elitist strategy, Condition 6.1 is not sufficient for GAs in the noisy environments to find a globally optimal solution with probability 1. Another important point is that if GAs do not reevaluate the fitness of each population member (except for that of the chromosome preserved by the elitist strategy) every time a new population is formed, then $\mathfrak{A}$ may not be positive recurrent; in fact, the chain may with positive probability never hit $\mathfrak{A}$. As a result, Condition 6.1 is not sufficient for GAs to find a globally optimal solution with probability 1 without the fitness reevaluation.

Note that Corollary 6.4 and Example 6.10 reveal a fundamental difference between the additive and multiplicative cases; although GAs eventually find a globally optimal solution with probability 1 in the additive case, they may eventually fail with probability 1 in the multiplicative case.

7 Convergence Rate Analysis

In this section, we demonstrate how Property 5.5, which ensures the convergence of the Markov chain (Z_k) to stationarity, and Theorem 4.3, which reveals the one-step transition probabilities of the chain, can be used to analyze how many iterations are sufficient to ensure with at least a specified probability that GAs in noisy environments select a globally optimal solution upon termination.

Let K denote the transition kernel of the chain. The stationary distribution of the chain will be denoted π. Property 5.4 states that the chain has only one positive recurrent communication class, which is $\mathfrak{A}$ in Definition 5.1. Thus

$$\pi(\mathscr{T}) > 0 \quad \forall\, \mathscr{T} \in \mathfrak{A}, \tag{21}$$

and

$$\pi(\mathscr{T}) = 0 \quad \forall\, \mathscr{T} \in \mathfrak{T} \setminus \mathfrak{A}. \tag{22}$$

Let $\pi^{(k)}$ denote the distribution of the chain (Z_k) at time k. Then by Property 5.5 and (21)–(22),

$$\lim_{k\to\infty} \sum_{\mathscr{T}\in\mathfrak{A}} \pi^{(k)}(\mathscr{T}) = 1. \tag{23}$$

Now if Condition 6.1 holds, then the sum $\sum_{\mathscr{T}\in\mathfrak{A}} \pi^{(k)}(\mathscr{T})$ represents the probability that the k-th population contains at least one globally optimal solution whose observed fitness equals its maximum perturbed fitness, which is greater than the maximum perturbed fitness of any suboptimal solution. Hence if Condition 6.1 holds, then we can analyze the convergence rate of (23) in order to determine how many iterations are sufficient to guarantee with certain probability that a GA selects a globally optimal solution when it is terminated.

In the noiseless case, Suzuki (1995) conducted the spectral analysis of the transition kernel to examine the convergence rate, and the same analysis can be conducted for noisy cases. By properly arranging rows and columns of K, we can show that K is a block lower triangular matrix whose first $|\mathfrak{A}|$ rows and first $|\mathfrak{A}|$ columns correspond to the states in $\mathfrak{A}$. Let $\widetilde{K}$ denote the submatrix of K by eliminating these $|\mathfrak{A}|$ rows and columns. Then each entry of $\widetilde{K}^k$ approaches zero as k goes to infinity (the modulus of each eigenvalue of $\widetilde{K}$ is strictly less than 1), and the convergence rate of (23) is determined by how fast each entry of $\widetilde{K}^k$ goes to zero. Let $\sigma(\widetilde{K})$ denote the spectral of $\widetilde{K}$. We have the following result:

Theorem 7.1. *There exists a constant $C < \infty$ such that for all k,*

$$\sum_{\mathscr{T} \in \mathfrak{A}} \pi^{(k)}(\mathscr{T}) \geq 1 - C\lambda^{*k}, \tag{24}$$

where $\lambda^* := \max\{|\lambda| : \lambda \in \sigma(\widetilde{K})\} < 1$.

Notice that this theorem is an improvement over Theorem 2 in Suzuki (1995). It is based on basic facts in matrix theory [for example, see Horn and Johnson (1985)] and can be proved analogously to Theorem 2 in Suzuki (1995). Thus the proof will be omitted.

Theorem 7.1 can be used to determine how many iterations are sufficient to ensure with at least a specified probability, say p, that a GA selects a globally optimal solution upon termination. Once we identify the constants C and λ^*, we find k such that $1 - C\lambda^{*k} \geq p$, from which we obtain $k \geq \frac{\ln(1-p) - \ln C}{\ln \lambda^*}$.

We also provide numerical examples that illustrate how the one-step transition probabilities shown in Theorem 4.3 can be used to analyze the exact probability that a GA is guaranteed to select a globally optimal solution upon termination. Consider the noisy environment in Example 3.2: A single source of additive noise perturbs each fitness evaluation. By Corollary 6.4, GAs eventually find a globally optimal solution with probability 1 in this noisy environment. For each k, we exactly compute

$$\sum_{\mathscr{T} \in \mathfrak{A}} \pi^{(k)}(\mathscr{T}), \tag{25}$$

which represents the probability that the k-th population contains at least one globally optimal solution whose observed fitness equals its maximum perturbed fitness. Hence the probability that GAs select a globally optimal solution upon completing the kth iteration is at least the sum (25). For each $i(j, \mathscr{P}_k)$, the distribution of additive noise $X_{i(j,\mathscr{P}_k)}$ is defined by

$$X_{i(j,\mathscr{P}_k)} = \begin{cases} 0 & \text{with probability } p \\ x & \text{with probability } 1-p. \end{cases}$$

We examine the exact degrees to which p and x affect the probability (25) for each k. We consider $S = \{1, 2, 3, 4\}$ and set the population size M to 3. For each $i \in S$, its noiseless fitness $f(i)$ will be defined by $f(i) = i$. Thus chromosome 4 is the globally optimal solution. We will execute a GA with proportional selection and without crossover. The fixed mutation rate μ will be set to .1. The initial state of the chain will be set to $\{(1,0),(1,0),(1,0)\}$. Thus the initial population $\mathscr{P}_0$ of the GA is set to $\{1,1,1\}$, and the probability (25) is 0 for $k = 0$.

First we set p to .5. Figure 1 plots the exact probability (25) against the number of iterations [k in (25)] for six different values of x: .5, 5.5, 10.5,...,25.5. Darker grays are used for larger values of x. The figure shows that at each iteration, the probability (25) decreases monotonically with the noise value x and that for each value of x, the probability increases monotonically with the number of iterations. (Corollary 6.4 guarantees that the probability approaches 1 as the number of iterations goes to infinity.)

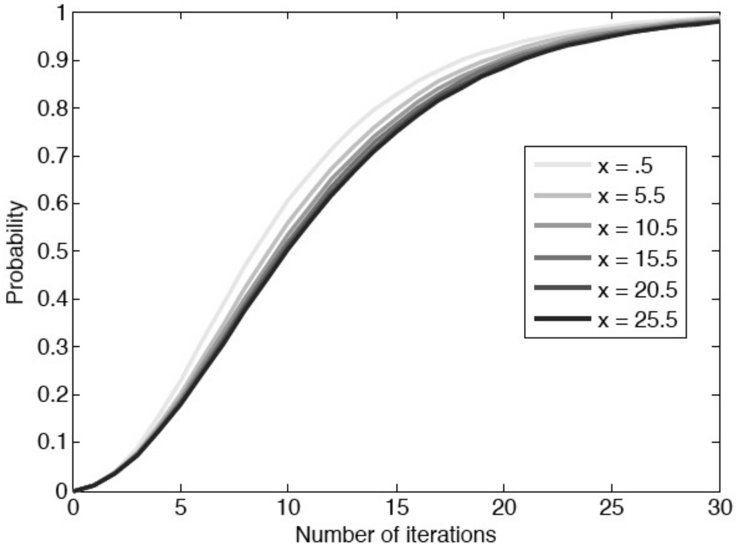

Figure 1: Probability (25) upon completion of each iteration. Additive noise equals 0 with probability .5 and x with probability .5. Darker grays are used for larger values of x.

Figure 2 plots the exact probability (25) against the number of iterations for $p = .999$. Again, darker grays are used for larger values of x, but the six curves appear to overlap since the effect of x is small in this case. Comparing Figures 1 and 2, we can see that the probability converges to 1 much more slowly when $p = .999$ compared to when $p = .5$. Thus, using the entries of the kernel of the chain, we can quantify the exact degrees to which various factors such as the distribution of noise and its possible values, the population size, the mutation rate, and the noiseless fitness value of each chromosome affect the exact probability that GAs are guaranteed to select a globally optimal solution at each iteration.

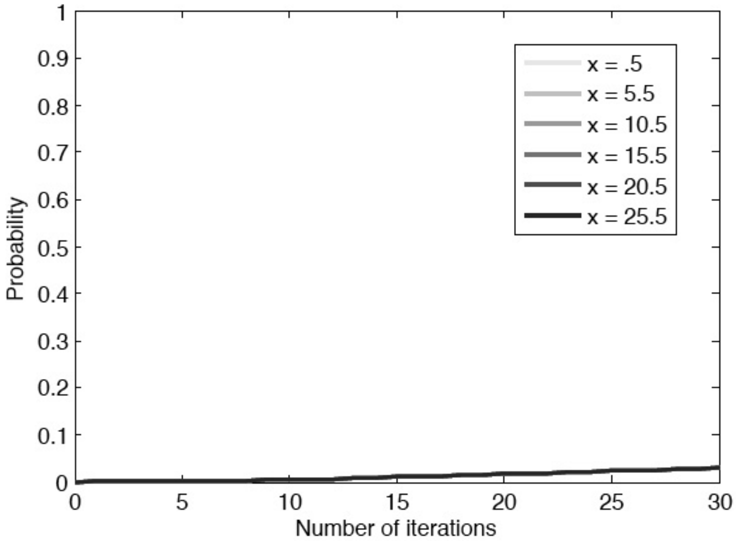

Figure 2: Probability (25) upon completion of each iteration. Additive noise equals 0 with probability .999 and x with probability .001.

8 Discussion

Our study clearly shows that Markov chain theory can be effectively used to mathematically describe GAs in various noisy environments and to reveal their fundamental convergence properties. As described in Section 3, the noisy environments considered in this study are quite general (see, in particular, Examples 3.2–3.6), and we did not make any assumptions about the expectation or variance of each noise source—our analysis is valid even if the noise sources have nonzero means, for instance. Mathematically, the Markov chain constructed in Section 4 completely captures the evolution of GAs in all the noisy environments considered in this study, and by analyzing the chain, we can easily prove the necessity and sufficiency of two conditions, Conditions 6.1 and 6.8, which are essential for characterizing GAs applied to noisy fitness functions.

Our theoretical results probably have significant implications in practice. As mentioned in Section 1, multiplicative noise disturbs objective functions in a variety of optimization problems, and objective functions can often be perturbed by more than one noise source. On the one hand, it is reassuring that GAs are guaranteed to eventually find a globally optimal solution in noisy environments provided Condition 6.1 is satisfied. On the other hand, one must be cautious about using GAs to find a globally optimal solution if it is suspected that this condition is not satisfied. Indeed, one is advised not to simply apply GAs if Condition 6.8 holds.

Mathematically, the Markov chain (Z_k) constructed and analyzed in this study is completely determined by its transition kernel, and we thoroughly described the kernel in Section 4. Our analysis described in Section 7 shows that knowledge about the kernel can be used to investigate the convergence rate of the probability in (23); using the one-step transition probabilities of the chain, we can compute the exact probability that GAs are guaranteed to select a globally optimal solution upon completion of each iteration. Thus, at least for simple cases, we can quantify the exact degrees to which various factors such as the distribution of noise and its possible values, the selection, crossover, and mutation schemes, the population size, and the noiseless fitness value of each chromosome affect the probability. This quantification can be useful for investigating optimal selection, crossover, and mutation schemes in various noisy environments.

Currently, we are applying results on simulated annealing in Markov chain theory to examine how to adaptively optimize the mutation rate in various noisy environments. We also intend to extend our Markov chain analysis to noisy environments with continuous noise. We believe that our Markov-chain-theoretic approach to analyzing GAs in noisy environments will further elucidate essential properties of the algorithms.

Acknowledgement: The author thanks James Allen Fill for helpful comments and discussions.

References

Amaziane, B.; Naji, A.; Ouazar, D.; Cheng, A. H.-D. (2005): Chance-constrained optimization of pumping in coastal aquifers by stochastic boundary element method and genetic algorithm. *CMC: Computers, Materials, and Continua*, vol. 2, pp. 85–96.

Arnold, D. V. (2002): *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, Boston.

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

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

Davis, T. E.; Principe, J. (1993): A Markov chain framework for the simple genetic algorithm. *Evolutionary Computation*, vol. 1, pp. 269–288.

Devroy, L.; Krzyzak, A. (2005): Random search under additive noise. In Dror, M.; L'Ecuyer, P.; Szidarovszky, F.(Eds): *Modeling Uncertainty*, International Series in Operations Research & Management Science, chapter 19, pp. 383–417. Springer, New York.

Di Pietro, A.; White, L.; Barone, L. (2004): Applying evolutionary algorithms to problems with noisy, time-consuming fitness functions. In *Proceedings of the 2004 Congress on Evolutionary Computation*, volume 2, pp. 1254–1261.

Galán Moreno, M. J.; Sánchez Medina, J. J.; Álvarez Álvarez, L.; Rubio Royo, E. (2009): Numerical simulation and natural computing applied to a real world traffic optimization case under stress conditions: "La Almozara" district in Saragossa. *CMES: Computer Modeling in Engineering and Sciences*, vol. 50, pp. 191–226.

Goldberg, D. E.; Rudnick, M. W. (1991): Genetic algorithms and the variance of fitness. *Complex Systems*, vol. 5, pp. 265–278.

Hopkins, W. E. (1987): Optimal stabilization of families of linear stochastic differential equations with jump coefficients and multiplicative noise. *SIAM Journal on Control and Optimization*, vol. 25, pp. 1587–1600.

Horn, R. A.; Johnson, C. R. (1985): *Matrix Analysis*. Cambridge University Press, Cambridge, UK.

Jin, Y.; Branke, J. (2005): Evolutionary optimization in uncertain environments—a survey. *IEEE Transactions on Evolutionary Computation*, vol. 3, pp. 303–317.

Karlin, S.; Taylor, H. M. (1975): *A First Course in Stochastic Processes*. Academic Press, New York.

Karlin, S.; Taylor, H. M. (1981): *A Second Course in Stochastic Processes*. Academic Press, New York.

Leung, K. S.; Duan, Q. H.; Xu, Z. B.; Wong, C. K. (2001): A new model of simulated evolutionary computation—convergence analysis and specifications. *IEEE Transactions on Evolutionary Computation*, vol. 5, pp. 3–16.

Miller, B. L.; Goldberg, D. E. (1996): Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, vol. 4, pp. 113–131.

Nakama, T. (2008): Theoretical analysis of genetic algorithms in noisy environments based on a Markov model. In *Proceedings of the 2008 Genetic and Evolutionary Computation Conference*, pp. 1001–1008.

Nissen, V.; Propach, J. (1998): On the robustness of population-based versus point-based optimization in the presence of noise. *IEEE Transactions on Evolutionary Computation*, vol. 2, pp. 107–119.

Nix, A.; Vose, M. D. (1992): Modeling genetic algorithm with Markov chains. *Annals of Mathematics and Artificial Intelligence*, vol. 5, pp. 27–34.

Primbs, J. A. (2007): Portfolio optimization applications of stochastic receding horizon control. In *Proceedings of the 2007 American Control Conference*, volume 1, pp. 1811–1816.

Rudin, L. I.; Osher, S. (2004): Total variation based image restoration with free local constraints. In *Proceedings of the 2004 International Conference on Image Processing*, volume 1, pp. 31–35.

Rudolph, G. (1994): Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, vol. 5, pp. 96–101.

Sinha, S.; Ch, P. (2008): Optimization of industrial fluid catalytic cracking unit having five lump kinetic scheme using genetic algorithm. *CMES: Computer Modeling in Engineering and Sciences*, vol. 32, pp. 85–102.

Suzuki, J. (1995): A Markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, pp. 655–659.

Vose, M. D. (1999): *The Simple Genetic Algorithm.* MIT Press, Cambridge, Massachusetts.

Vose, M. D.; Liepins, G. E. (1991): Punctuated equilibria in genetic search. *Complex Systems*, vol. 5, pp. 31–44.