Stochastic modelling of Genetic Algorithms

David Reynolds*, Jagannathan Gomatam

Department of Mathematics, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, UK

Abstract

This paper presents stochastic models for two classes of Genetic Algorithms. We present important distinctions throughout between classes of Genetic Algorithms which sample with and without replacement, in terms of their search dynamics. For both classes of algorithm, we derive sufficient conditions for convergence, and analyse special cases of Genetic Algorithm optimisation. We also derive a long-run measure of crossover bias for optimisation via Genetic Algorithms, which has practical implications with respect to the choice of crossover operators. For a class of Genetic Algorithms, we provide theoretical underpinning of a class of empirically derived results, by proving that the algorithms degenerate to randomised, cost-independent search as mutation probabilities increase. For an alternative class of Genetic Algorithms, we show that degeneration accompanies excessive crossover rates. In formulating the models, important definitions are introduced which capture in simple form the probabilistic properties of the genetic operators, which provides models which are independent of solution encoding schemes.

1. Introduction

Genetic Algorithms (GAs) are a set of heuristic search algorithms which have been applied with success to a wide variety of combinatorial optimisation problems, including many that are NP-hard [3,10]. Theoretical investigations into the reasons for this success are divided mainly into two camps; those who formulate and analyse deterministic mathematical models for GAs [2, 10], and those who investigate the application of stochastic models based on the theory of Markov Chains [4, 5, 7–9, 15, 17, 18]. In this paper we shall present a number of stochastic models for GAs.

---

* Corresponding author. E-mail: dre@gcal.ac.uk.

We begin by giving a brief introduction to GAs in Section 2, making important distinctions between large classes of GAs, those which create new populations of solutions from old ones by sampling with replacement, and those which sample solutions without replacement. In Section 3 we present a brief introduction to the theory of homogeneous Markov Chains [4, 5, 15, 18], and then proceed to summarise current stochastic models for GAs [7–9, 17]. The work reported in [7] was previously unknown to the authors; in this paper we generalise and extend the results reported there. In Sections 4 and 5 we introduce new definitions for describing the probabilistic properties of the GA operators and use these definitions to formulate representation-independent stochastic models for the two classes of GAs mentioned. The models are then used to investigate various properties of GAs. We analyse special cases of optimisation via GAs, and show that, given certain choices of mutation and crossover rates, both GAs reduce to a search of the solutions where no role is played by the cost function, despite the presence of a selection operator. For these cases, we show that both GAs converge in one generation. We present bounds for the expected numbers of optimal and other solutions within the population in the long run which is based on the probability of creation of solutions under crossover, for the class of GAs which operate on the basis of sampling with replacement. Finally, we discuss the potential for further analysis of the models contained in this paper with respect to obtaining the long-run probability distribution of a generalised GA [7] and for analysing their rates of convergence towards this distribution.

2. Definitions of GAs

The objective of this section is to formulate definitions for GAs, together with various assumptions and concepts which are necessary for the formulation of mathematical models to describe their behaviour. We stipulate that our optimisation problem has $k > 0$ candidate solutions, where k is an integer. We assume that we have assigned a bijective mapping from the set of solutions to the integers $\{1, \ldots, k\}$, and that to each solution i $(1 \leq i \leq k)$ there corresponds a well-defined cost (or fitness), f_i, which is restricted by $0 < f_i < \infty$. Thus, to the set of solutions $s = \{s_1, s_2, \ldots, s_k\}$ there corresponds a set of fitness values $f = \{f_1, \ldots, f_k\}$. For the purposes of formulating and developing the models, we merely assume that we have assigned a natural index labelling system to the solutions, e.g., in a binary string encoding [10] each solution would be represented in our analysis as the base 10 version of its binary representation; thus, we do not assume any particular encoding scheme for solutions.

A GA for solving finite and discrete maximisation problems of the type mentioned is characterised by possessing the following features [3, 10]:

   (i) a scheme for encoding solutions to the optimisation problem to be solved (chromosomes);
  (ii) an evaluation function that rates each solution, assigning a positive cost, or fitness to each one;

(iii) an initialisation procedure for generating an initial population of size N of candidate solutions;
(iv) a set of operators used to manipulate the genetic composition of the population between generations;
 (v) a set of parameter values, such as stopping_criteria, etc.

This algorithm, together with the genetic operators, has been implemented in many different ways [6]. For example, there are many different representation schemes [6] for candidate solutions to optimisation problems, the most common being binary strings of fixed length L. Also, the fitness evaluation/selection, mutation and crossover operators have all been implemented in different ways [6], including the recent introduction of a time-varying mutation operator [7]. Further, the order in which these operators are applied in order to generate new populations has also been varied [6]. However, we can divide into two classes the majority of existing GAs; those algorithms which form new populations of solutions by sampling solutions with replacement from the current population and then implementing the genetic operators, and those which sample solutions without replacement and then implement the genetic operators. A pseudo-code description for GAs which use sampling with replacement is:

> GA/WI(N, s, f, *fitness selection*, *mutation*, *crossover*)
> begin
>   Initialise with population of size N at generation $t = 0$
>   REPEAT
>   WHILE $|Pop(t + 1)| < N$
>   BEGIN
>   Select ($p \geqslant 2$) parent solutions with replacement from $Pop(t)$ by
>   *fitness selection*
>   Carry out *mutation* of the copies of the selected parents
>   Combine the mutant parents to form ($c \geqslant 1$) child solutions by
>   *crossover*
>   Place the child solutions in the new population $Pop(t + 1)$
>   END
> $t := t + 1$
>   UNTIL stopping_criteria_reached
> end

Here and in what follows, *mutation* and *crossover* are pre-defined operators.

We shall consider *fitness selection* above and in what follows to be implemented by the common roulette wheel selection [10]. Further, for GA/WI, we shall consider the selected parent solutions to be replaced immediately after selection into the old population (i.e., into $Pop(t)$), and that we subsequently operate on copies of the parents. For GAs which operate on the basis of sampling without replacement, a pseudo-code description is:

> GA/WO(N s, f, *fitness selection*, *mutation*, *crossover*)
> begin

Initialise with population of size N, N even, at generation $t = 0$
REPEAT
    Select N parent solutions with replacement from $Pop(t)$ by *fitness selection*, to form population $Pop'(t)$
    Carry out *mutation* to each solution in $Pop'(t)$ to form population $Pop''(t)$
    Carry out *crossover* pairwise between solutions in $Pop''(t)$ to form N new child solutions, to form population $Pop'''(t)$
    Set $Pop(t + 1) := Pop'''(t)$
    $t = t + 1$
UNTIL stopping_criteria_reached
end

It is important to note that both algorithms contain a sampling with replacement *fitness selection* operator, although in GA/WO the *mutation* and *crossover* operators are implemented on the basis of sampling from the current population without replacement. According to [9], GAs which contain implementations of the latter two operators without replacement are more common in GA applications. Note that we have not described the above algorithms with respect to the choice of solution representation scheme, of which there are many [6]. The models we formulate will be independent of representation.

3. Current stochastic models

We present a brief introduction to Markov Chains. We then proceed to discuss the current approaches to the stochastic modelling of GAs. The reader is referred to [5,15,18] for further details.

3.1. Homogeneous Markov Chain theory

Definition 3.1. A stochastic process (i.e., a sequence of random variables over time, $X(t)$) is called a Markov process if it exhibits Markov dependence, i.e., if

$$\Pr[X(t) \leqslant x \mid X(t_n) = x_n, \ldots, X(t_0) = x_0] \tag{3.1}$$

$$= \Pr[X(t) \leqslant x \mid X(t_n) = x_n] \tag{3.2}$$

$$= F(x_n, x; t_n), \tag{3.3}$$

where $t > t_n > \cdots > t_0$. A realisation of $X(t)$ over time is called a Markov Chain.

The Markov Chains we shall discuss in this paper are realisations of random variables defined over discrete state-spaces and discrete time. If a Markov process has parameters which do not vary over time, the process is said to be homogeneous, otherwise, inhomogeneous. In the homogeneous case, the probabilities of

transiting from state i to state j between time steps $t_0 = m$ and $t_1 = n$ depend only on $t_1 - t_0$ and are defined to be

$$P_{ij}^{(m,n)} = \Pr[X_n = j \mid X_m = i] . \tag{3.4}$$

A Markov Chain is said to be irreducible if the corresponding one time-step transition probability matrix is irreducible. If the chain has all of its states i, $1 \leqslant i \leqslant S$, s.t. $P_{ii} > 0$, then the chain is said to be aperiodic. For irreducible and aperiodic finite length Markov Chains, the following two theorems hold:

Theorem 3.2 [15, p. 47]. *Let P be the transition probability matrix of an irreducible, aperiodic finite Markov Chain. Then there exists an N s.t. for all $n \geqslant N$, the n-step transition probability matrix P^n has no zero elements.*

Theorem 3.3 [15, p. 88]. *Let P be the transition probability matrix of an irreducible, S-state finite Markov Chain. Then*

$$\lim_{t \to \infty} P^t = \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{bmatrix}, \tag{3.5}$$

where $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_S]$ with $0 < \pi_j < 1$ and $\boldsymbol{\pi} \cdot [1, \ldots, 1]_{1 \times S} = 1$.

The following theorem shows that irreducible and aperiodic Markov Chains converge towards stationarity at a geometric rate (i.e., possess geometric ergodicity, [15]).

Theorem 3.4 [15, p. 90]. *Let P be the transition matrix of an irreducible, aperiodic S-state Markov Chain. Let the limiting probabilities be defined as in (3.5). Then there exists constants c and r with $c > 0$ and $0 < r < 1$ s.t.*

$$P_{ij} = \pi_j + e_{ij}^{(n)} , \tag{3.6}$$

where

$$\left| e_{ij}^{(n)} \right| \leqslant c r^n . \tag{3.7}$$

When $P > 0$, $c = 1$.

Finally, the following theorem presents the stationary distribution of an irreducible and aperiodic Markov Chain in terms of the corresponding transition probability matrix:

Theorem 3.5 [15, p. 92]. *Given the transition probability matrix P of an aperiodic and irreducible, S-state finite Markov Chain, there exists a unique probability vector $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_S]$ with $\boldsymbol{\pi} \cdot [1, \ldots, 1]_{1 \times S} = 1$, and*

$$\Pi P = \Pi = P\Pi , \tag{3.8}$$

where Π is a matrix where each row vector is identical to π. The probability vector π gives the stationary distribution of the process.

Notice that (3.8) implies $\pi P = \pi$. In the rest of the paper, we shall refer to (3.8) as the invariant equation. We proceed in the next subsection to discuss current approaches to the stochastic modelling of Genetic Algorithms.

3.2. General stochastic models for GAs

Most current models formulated to describe the probabilistic search carried out by GAs are mainly based on quite general matrix properties, and the properties of the binary string encoding scheme. Thus, the genetic operators are cast as irreducible and primitive stochastic matrices, and it can be shown that, for example, due to the properties of these matrices a stationary distribution holds for certain classes of GA [7–9, 17]. However, since these models generally consider the action of the mentioned operators upon the states of the algorithm, i.e., the populations, it is quite difficult to obtain results regarding the actual solutions to the optimisation problem [17]. Thus, the transition probability matrices of many current models are based in the large part upon the irreducibility and primitivity properties of certain genetic operators, and this means that the models are not formulated on the basis of costs of solutions, the probability that a particular solution is created upon *mutation, crossover*, and other algorithm parameters. Thus, little information can be extracted with respect to the search dynamics of the corresponding algorithms. For example, the main stochastic model of [17] is

$$P = C \cdot M \cdot S \tag{3.9}$$

where P is the transition probability matrix of the underlying stochastic process of the Genetic Algorithm, and C, M and S are irreducible matrices which correspond to the three genetic operators acting upon the population states of the algorithm. The stochastic model contained in [8] is essentially an abstract generalisation, which also appears to be over-simplified [17], as is the main model of [9]. Note that in [8, 17] a globally optimal GA is given as one which maintains the currently optimal solution from generation to generation (i.e., elitism [10]). In [17] the conclusion is reached that the global optimality of this algorithm depends essentially on this "algorithmic trick". We verify this conclusion explicitly in this paper by carrying out the analyses of special cases mentioned previously.

3.3. Explicit stochastic models for GAs

In [7], a description is given of the application and analysis of modelling a binary encoded GA as a sequence of random vectors [4,11]. Thus, using these models, it is shown in [7] that a stationary distribution exists for certain classes of GAs, at least for those which implement the genetic operators on the basis of sampling with replacement, i.e., instances of GA/WI under a binary encoding

scheme [10]. The stationary distribution of the GAs described in [7] is given in terms of the characteristic equations of modified transition probability matrices. However, the main work of [7] is concerned with the analysis of time-varying mutation probabilities within Genetic Algorithms, and does not generalise to both classes of Genetic Algorithm under discussion here. In this paper we present and formalise the distinction between algorithms GA/WI and GA/WO by formulating stochastic models for each algorithm which show explicitly their differing search properties; these differing properties have not yet been discussed in the literature. The models are then used to present explicitly the stationary distribution of both of these GAs in certain cases, and to provide practical guidance for the parameterisation of such algorithms.

4. Stochastic model for GA/WI

The objective of this section is to formulate a stochastic model for the first class of algorithms, GA/WI, i.e., those which create new populations of solutions purely on the basis of sampling with replacement from the current population and then operating on copies of the sampled solutions with the mentioned genetic operators.

4.1. K-solution, 1-operator GA/WI(N, s)

We present a stochastic model for GA/WI(N, s), i.e., with no fitness selection, *mutation* or *crossover* operators. We begin by formulating essential definitions.

Definition 4.1. Let $(X_i)_t$ represent the numbers of solution i within the population GA/WI at time t. Then $(X_1, X_2, \ldots, X_k)_t$ represents the numbers of all solutions in the population at time t, where $X_1 + X_2 + \cdots + X_k = N$ and X_i is a positive integer.

Definition 4.2. Let

$$S^{[k]} = \{X = (X_1, X_2, \ldots, X_k): X_1 + X_2 + \cdots + X_k = N,$$
$$X_i \text{ a positive integer}\}$$

represent the set consisting of all valid states of algorithm GA/WI.

Thus, since GA/WI samples solutions with immediate replacement and then implements the genetic operators on copies of the sampled solutions, no distinction is made between the permutations of solutions within the population states of $S^{[k]}$. Since the numbers of each solution within the population is a random variable, the collection $(X_1, X_2, \ldots, X_k) \in S^{[k]}$ is a random vector [11], subject to $X_1 + X_2 + \cdots + X_k = N$. Therefore, the states of GA/WI are random vectors. Essentially, the analysis of the properties of GA/WI with respect to

heuristic search for optimisation reduces, in certain cases, to the analysis of the probability distributions of the given random variables/vectors.

In order to formulate transition probability matrices and corresponding convergence results for GA/WI, we make the following definition of matrix multiplication for general matrices which are indexed by vectors;

Definition 4.3. Let $A_{XY} \in \Re$ and $B_{XY} \in \Re$ be $|S^{[k]}| \times |S^{[k]}|$ matrices, where $X, Y \in S^{[k]}$. Then we define as the IJth element of the matrix product AB the real number

$$\sum_{D \in S^{[k]}} A_{ID} B_{DJ} = \sum_{i=1}^{|S^{[k]}|} A_{ID_i} B_{D_i J} \quad \forall I, J \in S^{[k]}, \tag{4.1}$$

where by summation over D_i is meant the summation over the list of vectors formed under an arbitrary bijective mapping between $\{1, 2, \ldots, |S^{[k]}|\}$ and $S^{[k]}$. Thus D is a dummy vector in (4.1). Only the existence of the (arbitrary) mapping is important, and there are $|S^{[k]}|!$ such mappings [19]. When successive matrix multiplications are carried out for such matrices, it is assumed that the same mapping is used for indexing purposes throughout.

From this general definition, we make the following definition for premultiplication of an $|S^{[k]}| \times |S^{[k]}|$ matrix by a $1 \times |S^{[k]}|$ column vector.

Definition 4.4. Let $q_Y \in \Re$ and $B_{XY} \in \Re$ where B is an $|S^{[k]}| \times |S^{[k]}|$ matrix and q is a $1 \times |S^{[k]}|$ column vector, where $X, Y \in S^{[k]}$. Then we define as the Jth element of the matrix product qB the real number

$$\sum_{D \in S^{[k]}} q_D B_{DJ} = \sum_{i=1}^{|S^{[k]}|} q_{D_i} B_{D_i J} \quad \forall J \in S^{[k]}, \tag{4.2}$$

where by summation over D_i is meant the summation over the list of vectors formed under an arbitrary bijective mapping between $\{1, 2, \ldots, |S^{[k]}|\}$ and $S^{[k]}$, as in Definition 4.3.

This notation, which enables us to ignore the particular mapping used and retain vector indexing notation for the elements of the transition matrices which are to follow, is useful in terms of estimating marginal probability distributions from transition matrices [15], as will become clear. It is usual when formulating stochastic models for GAs to show the existence of the bijective mappings described in the previous definitions, and then use these to index the states of the GA [7–9]. However, in this paper we do not require this restriction, preferring to retain the identity of solutions within the states of GA/WI; this then gives useful information on the long run search properties of the GAs with respect to *both* solutions and states.

A result is given in [7, 17] concerning the number of states of a GA, in terms of

the chosen representation scheme, i.e., fixed length binary strings. We now present the following general result.

Proposition 4.5. *GA/WI with population size N, where the order in which the population is stored is irrelevant (sampling with replacement GA), operating on a solution space of size $|s| = k$ searches a state-space of size $|S^{[k]}|$, where*

$$|S^{[k]}| = \binom{N + k - 1}{k - 1}. \tag{4.3}$$

Proof. A population configuration $(X_1, X_2, \ldots, X_k)$ is a valid state of GA/WI if and only if

$$X_1 + X_2 + \cdots + X_k = N \tag{4.4}$$

by definition of $X \in S^{[k]}$. Thus the total number of valid states is given by the total number of distinct solutions of this equation in non-negative integers, which is a well-known problem from combinatorics, with solution as stated [19]. □

Therefore, the number of states of GA/WI is a function of the solution space and the population size alone, regardless of the chosen solution encoding scheme. It will be shown later that a fundamental difference between GA/WI and GA/WO is the number of states of each algorithm; GAs which contain operators which do not sample with replacement (i.e., GA/WO-type algorithms) will be shown to have more states than those which do not contain such operators.

With the necessary definitions of the underlying random variables (solutions) and random vectors (states) of the following stochastic model in place, we now proceed to calculate the conditional probability distributions of these entities; for GA/WI(N, s), with random sampling of solutions, the probability that solution i is sampled from population $X = (X_1, X_2, \ldots, X_k) \in S^{[k]}$ is X_i/N where $1 \leqslant i \leqslant k$. Therefore, the probability of $Y = (Y_1, Y_2, \ldots, Y_k) \in S^{[k]}$ instances of solutions $(1, \ldots, k)$ at time $t + 1$ given $X = (X_1, X_2, \ldots, X_k)$ instances at time t is

$$\Pr\{(Y_1, Y_2, \ldots, Y_k)_{t+1} \,|\, (X_1, X_2, \ldots, X_k)_t\}$$

$$= P_{XY} = \frac{N!}{Y_1! Y_2! \cdots Y_k!} \left(\frac{X_1}{N}\right)^{Y_1} \left(\frac{X_2}{N}\right)^{Y_2} \cdots \left(\frac{X_k}{N}\right)^{Y_k}. \tag{4.5}$$

This equation gives the conditional probability distribution of the random vector $(Y_1, Y_2, \ldots, Y_k)$ at time $t + 1$. Note that the state-space of the Markov process defined by (4.5) is the same as the space defined in Definition 4.2, i.e., it consists of all distinct vectors $(X_1, X_2, \ldots, X_k) \in S^{[k]}$. Each solution has a one-step conditional probability distribution given by [11]

$$\Pr\{(Y_i)_{t+1} \,|\, (X_1, X_2, \ldots, X_k)_t\} = \sum_{Y_1} \sum_{Y_2} \cdots \sum_{Y_{i-1}} \sum_{Y_{i+1}} \cdots \sum_{Y_k} P_{XY}, \tag{4.6}$$

where $X, Y \in S^{[k]}$. Substituting (4.5) into (4.6) shows that the numbers of solution

i, $1 \leq i \leq k$, within the population at time t, $(X_i)_t$ is a random variable with a binomial distribution. In [7] certain basic properties of the model (4.5) are investigated for the case of binary representation schemes.

4.2. Incorporating fitness-based selection: GA/WI(N, s, f, roulette wheel)

We consider the standard roulette-wheel-based selection procedure [3,10]. Since each solution has a strictly positive cost, or fitness value, i.e., to $(X_1, X_2, \ldots, X_k) \in S^{[k]}$ there corresponds a fitness vector, $f = (f_1, \ldots, f_k)$ with $f_i > 0$, $1 \leq i \leq k$, if we select from the population at time t (with replacement) on the basis of roulette-wheel fitnesses then we have that the probability that solution i is sampled from population $X \in S^{[k]}$ is

$$r_{xi} = \frac{X_i f_i}{\sum\limits_{j=1}^{k} X_j f_j} , \tag{4.7}$$

so that

$$\Pr\{(Y_1, Y_2, \ldots, Y_k)_{t+1} | (X_1, X_2, \ldots, X_k)_t\} = \frac{N!}{\prod\limits_{i=1}^{k} Y_i!} \prod\limits_{i=1}^{k} r_{xi}^{Y_i} , \tag{4.8}$$

where by r_{xi} is meant the relative fitness of solution i conditional upon population X. Since the fitness of a solution is assumed fixed, we have again that this model defines a time-homogeneous first-order Markov process, over the states defined as before by $X = (X_1, X_2, \ldots, X_k) \in S^{[k]}$. This model is also discussed in [7] for the special case of binary string representation of solutions [10]. In the next section, we discuss the problem of introducing the GA *mutation* and *crossover* operators into this model, in a form which leaves the model representation-independent.

4.3. Incorporating mutation and crossover: GA/WI(N, s, f, roulette wheel, mutation, crossover)

Since our models are to be representation-independent, we make the following definitions, which capture in simple form the stochastic properties of the genetic operators *mutation* and *crossover*.

Definition 4.6. Let η be a row-stochastic mutation matrix, i.e., η_{ij} stores the probability that solution i transforms to solution j under *mutation*, i.e.,

$$\sum_{j=1}^{k} \eta_{ij} = 1 . \tag{4.9}$$

For example, for a binary representation with strings of fixed length L and using the fixed mutation probabilities $p_m \leq 1/2$, used in [6, 10], we have that

$$\eta_{ij} = p_m^{H(i,j)} (1 - p_m)^{L - H(i,j)}, \quad \eta_{ij} = \eta_{ji} > 0 \tag{4.10}$$

[17], where $H(i, j)$ represents the Hamming distance between the binary string encodings of solutions i and j [17]. To both incorporate *crossover* and to remain representation-independent, we define a crossover probability matrix, which gives the probability that a fixed number of (ordered) parent solutions will generate any fixed number of (ordered) child solutions under the action of some *crossover* operation. Thus, this matrix is indexed by ordered pairs, triples, etc. The "minimal" crossover probability matrix is defined below.

Definition 4.7. Let $C_{ij,m}$, $j \geq i$, be a row-stochastic crossover probability matrix, i.e.,

$$\sum_{m=1}^{k} C_{ij,m} = 1 \quad \forall 1 \leq i \leq j \leq k, \tag{4.11}$$

that stores the probability with which solutions i, j, will produce solution m under the action of *crossover* $(1 \leq i \leq j \leq k, 1 \leq m \leq k)$. Moreover, we define $C_{ij,i} > 0$ and $C_{ij,j} > 0$ (non-zero probability of parental retention). Note that the rows of C are indexed by ordered pairs of parent solutions, the columns by single child solutions.

An example is given in Appendix A of a crossover probability matrix for a binary string encoding of solutions, where the string length is 2. Note that we force the parent solutions to be gathered in an ordered pair, which forces uniqueness of reference.

With this definition in place, it is now possible to see that this particular *crossover* is a 2-parent 1-child crossover operation, as implemented in, for example, [14], i.e., 2-parent solutions will always generate only one child (which may be one of the parents). It was noticed that in fact it is quite possible to define an X-parent Y-child crossover operation [6], where $X \geq 2$ and $Y \geq 1$. Therefore, it should be pointed out that descriptions of *crossover* in the literature are mere instances of an X-parent Y-child crossover operation, i.e., the C matrix can be easily extended to triples, etc., of both parents and children; the essential requirement is that the pairs, triples, must be distinct so that the solution production probabilities under *crossover* are not over/underestimated within the C matrix. Since our definition is the simplest possible, we shall develop a Markov Chain model for GA/WI which uses this definition. As usual, the previous two definitions apply to the given genetic operators regardless of solution encoding scheme.

With these definitions of *mutation* and *crossover* in place, we now proceed to calculate the probability that solution i, $1 \leq i \leq k$, is produced given $X = (X_1, X_2, \ldots, X_k) \in S^{[k]}$, under the action of roulette-wheel fitness selection of two parents (with replacement), then *mutation* of one parent and then *crossover* between the mutant and non-mutant to produce solution i; this is clearly given by, in terms of the unions and intersections of probabilistic events,

$$\mathrm{Pr}\left\{ \bigcup_{j \in s} \bigcup_{d \in s} \left(j_{\mathrm{fs}} \cap d_{\mathrm{fs}} \cap \left\{ \bigcup_{b \in s} j \underset{\eta}{\to} b \cap b, d \underset{C}{\to} i \right\} \right) \right\} \tag{4.12}$$

$$= \mathrm{Pr}\left\{ \bigcup_{j \in s} j_{\mathrm{fs}} \cap \bigcup_{b \in s} j \underset{\eta}{\to} b \cap \bigcup_{d \in s} d_{\mathrm{fs}} \cap d, b \underset{C}{\to} i \right\} \tag{4.13}$$

(where subscript "fs" refers to fitness selection) which has probability

$$p_{xi} = \sum_{j=1}^{k} \frac{X_j f_j}{X \cdot f} \sum_{b=1}^{k} \eta_{jb} \left[\sum_{d=1}^{b} \frac{X_d f_d}{X \cdot f} C_{db,i} + \sum_{d=b+1}^{k} \frac{X_d f_d}{X \cdot f} C_{bd,i} \right], \tag{4.14}$$

where by p_{xi} is meant the probability of generation of solution i by GA/WI, conditional upon population X, where $1 \leqslant i \leqslant k$. This gives the stochastic model for GA/WI, with the given operators, as

$$P_{XY} = \frac{N!}{Y_1! Y_2! \cdots Y_k!} \prod_{i=1}^{k} (p_{xi})^{Y_i}, \tag{4.15}$$

where $Y \in S^{[k]}$. Note that the independence of parent selection gives (4.13) from (4.12). The row-stochasticity of P follows from that of the given p_{xi} terms (which follows by the previous two definitions) and the properties of the multinomial distribution [12]. With *mutation* of both parents, the stochastic model would be obtained in a similar manner, for $X, Y \in S^{[k]}$,

$$P_{XY} = \frac{N!}{Y_1! Y_2! \cdots Y_k!} \prod_{i=1}^{k} (p_{xi})^{Y_i}, \tag{4.16}$$

$$p_{xi} = \sum_{j=1}^{k} \frac{X_j f_j}{X \cdot f} \sum_{d=1}^{k} \frac{X_d f_d}{X \cdot f} \sum_{b=1}^{k} \eta_{jb} \sum_{e=1}^{k} \eta_{de} C_{eb,i}^*,$$

$$C_{eb,i}^* = \begin{cases} C_{eb,i}, & e \leqslant b, \\ C_{be,i}, & e > b. \end{cases}$$

4.4. Analysis of the models for
GA/WI(N, s, f, roulette wheel, mutation, crossover)

We proceed in this section to present three results relating to the search properties of GA/WI. The first result presents sufficient conditions for GA/WI to converge to a stationary distribution; the second result presents a special case of GA/WI search which has important implications for Genetic Algorithm parameterisation and representation issues, and the third result presents a long-run approximation result for GA/WI with respect to the implementation of *crossover*. We first show the usefulness of the independence from solution encoding schemes of the models by presenting the following for model (4.15). A convergence theorem for (4.16) may be obtained similarly.

4.4.1. Parameter choice for convergence

Theorem 4.8. *GA/WI with $\eta > 0$, and C as defined, converges in limit towards a stationary probability distribution q where q_X is strictly positive for all $X \in S^{[k]}$, i.e.,*

$$\lim_{t \to \infty} P^t = \begin{bmatrix} q \\ q \\ \vdots \\ q \end{bmatrix}. \tag{4.17}$$

Also, convergence takes place at least geometrically, i.e., there exists $0 \leqslant r < 1$, s.t.

$$P_{XY}^{(t)} = q_Y + e_{XY}^{(t)} \tag{4.18}$$

for all states $X, Y \in S^{[k]}$, where

$$|e_{XY}^{(t)}| \leqslant r^t. \tag{4.19}$$

Finally, q is independent of the initial choice of population $X^{(0)}$.

Proof. Suppose we are given arbitrary states $X, Y \in S^{[k]}$ where we transit from X to Y in one time step according to the appropriate probability in P given by (4.15). Then, there are two possible cases;

   Case 1: $X_I = 0$, *some* $1 \leqslant I \leqslant k$. Then there exists at least one component X_j in X, with corresponding relative fitness r_{xj} greater than zero, since the population is of fixed size N. Thus,

$$p_{xI} = r_{xj}\eta_{jI}r_{xj}C_{jI,I} + \cdots > 0 \quad (j \leqslant I),$$
$$p_{xI} = r_{xj}\eta_{jI}r_{xj}C_{Ij,I} + \cdots > 0 \quad (j > I). \tag{4.20}$$

   Case 2: $X_I > 0$, $1 \leqslant I \leqslant k$. Then

$$p_{xI} = r_{xI}\eta_{II}r_{xI}C_{II,I} + \cdots > 0. \tag{4.21}$$

Therefore, for arbitrary states $X, Y \in S^{[k]}$ the entry in P consists of a product of strictly positive numbers, since I was also arbitrary, which implies that P is strictly positive. The result follows by application of Theorems 3.2 and 3.3. $\square$

Note that *mutation* is often implemented for binary-encoded GAs in such a way as to force $\eta > 0$, as mentioned previously [7, 17]. In the next section we show that it is possible to parameterise GA/WI in a way which completely removes bias towards solutions based on cost.

4.4.2. Special case

Proposition 4.9. *GA/WI has a random-search-type algorithm over the solution space as a special case, i.e., there exists at least one choice for C, s.t. the underlying stochastic process of GA/WI is a Markov Chain in which any given $k - 1$ solutions*

can be represented by independent and identically distributed random variables. Further, the Markov Chain under this choice of C converges to stationarity immediately, i.e., in one generation.

Proof. Set $C_{bd,t} = 1/k \ \forall 1 \leq b \leq d \leq k, \ 1 \leq t \leq k$. Then,

$$p_{xi} = 1/k \quad \forall 1 \leq i \leq k, \quad \forall X \in S^{[k]} \tag{4.22}$$

by substitution in (4.14) so that

$$P_{XY} = \frac{N!}{Y_1! Y_2! \cdots Y_k!} \frac{1}{k^N}, \tag{4.23}$$

where $Y \in S^{[k]}$. Since the stationary distribution of a homogeneous stochastic process, if it exists, is unique, we can "guess"

$$q_X = \frac{N!}{X_1! X_2! \cdots X_k!} \frac{1}{k^N}, \tag{4.24}$$

and check by substitution in (3.8)

$$\sum_{X \in S^{[k]}} q_X P_{XY} = \sum_{X \in S^{[k]}} \left[\frac{N!}{X_1! X_2! \cdots X_k!} \frac{1}{k^N} \frac{N!}{Y_1! Y_2! \cdots Y_k!} \frac{1}{k^N} \right] = P_{XY} \tag{4.25}$$

and

$$\sum_{X \in S^{[k]}} q_X = 1, \quad q_X > 0 \quad \forall X, \tag{4.26}$$

where $X \in S^{[k]}$. Thus the chosen q is indeed the stationary distribution of GA/WI with $C = [1/k]$. To show that the GA converges in one generation, we observe that

$$(P^2)_{XY} = \sum_{M \in S^{[k]}} \frac{N!}{M_1! M_2! \cdots M_k!} \frac{1}{k^N} \frac{N!}{Y_1! Y_2! \cdots Y_k!} \frac{1}{k^N} = P_{XY}, \tag{4.27}$$

and so P is its own limiting matrix, by repeated substitution. To show that the random variables X_i, $1 \leq i \leq k$, have the same marginal distributions, observe that each of these random variables has a marginal distribution at stationarity which is binomially distributed with probability parameter $1/k$, i.e., at stationarity, the unconditional probability of obtaining j of solution i is given by

$$\Pr[X_i = j] = \binom{N}{j} \left(\frac{1}{k}\right)^j \left(1 - \frac{1}{k}\right)^{N-j}, \quad t = 1, 2, \ldots \tag{4.28}$$

(by use of (4.6)) and this holds for all $1 \leq i \leq k$. Notice that, since $X_1 + X_2 + \cdots + X_k = N$, $\forall X \in S^{[k]}$, only $k - 1$ of the given random variables are independent [11], as claimed, and this completes the proof. □

Again, an analogous result holds for model (4.16), for GA/WI with *mutation* of both parents. We describe the algorithm with the previous *crossover* operator

choice as "random search" since the random variables X_i, $1 \le i \le k$, are all identically distributed at all times during the search (since the algorithm converges immediately), i.e., there is no selection bias towards any solution based on cost or otherwise. In fact, there are other choices for C (and η) which provide this result also; we sought only to prove an existence result for the special case. The above result has important implications for the parameterisation and solution encoding schemes of Genetic Algorithms with respect to the implementation of *crossover*. Essentially, GA/WI-type Genetic Algorithms which do not relate the probability of child production under *crossover* to any of the properties of the given parents are degenerate algorithms with respect to the search for optimal solutions; thus, we strongly conjecture that a fundamental reason for the observed success of binary-encoded Genetic Algorithms depends to a certain extent on the properties of the binary encoding scheme. In practical terms, the binary encoding scheme for solutions automatically prevents the degeneration of GA/WI-type GAs to (multidimensional) random search algorithms. In the next section, we extract information from the model (4.15) which also provides a measure of crossover bias with respect to the choice of *crossover* operators.

4.4.3. Long run approximation result

We now present a general result for GA/WI, where *mutation* of only one parent is carried out (4.15). As before, analogous results exist for the case where *mutation* of both parents is carried out (4.16). To proceed, we formulate definitions of certain parameters of GA/WI, which capture fundamental properties of these operators.

Definition 4.10. Let $\theta_t = \min C_{bd,t}$ $\forall 1 \le b \le d \le k$, $1 \le t \le k$.

Thus, θ_t is the minimum component of the vector which is the tth column of C. Note that θ_t may vary across columns of C (solutions).

Definition 4.11. Let $\varepsilon_t = \max C_{bd,t}$ $\forall 1 \le b \le d \le k$, $1 \le t \le k$.

Thus, ε_t is the maximum component of the vector which is the tth column of C. Note that ε_t may also vary across columns of C (solutions).

Definition 4.12. Let μ_t^{∞} be the long run expected numbers of solution t in the population, i.e., the expected value of random variable X_t within the stationary distribution q where $1 \le t \le k$.

Definition 4.13. Let e be a matrix which rows store the population vectors in the same order in which they index P and q, so that e is of size $|S^{[k]}| \times |s|$. Then e_t is a vector containing X_t of solution t across the populations.

With these definitions, we state and prove the following approximation guarantee for GA/WI with *mutation* of one parent.

Theorem 4.14. *For GA/WI, with mutation of one parent only* (4.15) *and with* $\eta > 0$, *the following holds*:

$$N\theta_t \leqslant \mu_t^\infty \leqslant N\varepsilon_t .$$
(4.29)

Proof. By definition of the expected value of a random variable within a multivariate distribution [11,12],

$$\mu_t^\infty = \sum_{X \in S^{[k]}} q_X \cdot X_t = q \cdot e_t = (qP) \cdot e_t = q \cdot (P \cdot e_t) = q \cdot a ,$$

$$a_X = \left[\sum_{Y \in S^{[k]}} \frac{N!}{Y_1! \cdots Y_k!} \prod_{i=1}^{k} (p_{xi})^{Y_i} Y_t \right]_{1 \times |S^{[k]}|}$$
(4.30)

(using the fact that q is a stationary distribution for P). By the formula for the expected value of the multinomial distribution this gives

$$\mu_t^\infty = N \sum_{X \in S^{[k]}} q_X \left\{ \sum_{j=1}^{k} \frac{X_j f_j}{X \cdot f} \sum_{b=1}^{k} \eta_{jb} \left[\sum_{d=1}^{b} \frac{X_d f_d}{X \cdot f} C_{db,t} + \sum_{d=b+1}^{k} \frac{X_d f_d}{X \cdot f} C_{bd,t} \right] \right\} ,$$
(4.31)

so that the result follows by simple substitution of the given θ_t and ε_t into (4.31).

By substituting the parameters from Proposition 4.9, we see that in these cases the bounds are exact, i.e., they return $\mu_t^\infty = N/k$, as expected. Again, this justifies the use of the term "random search" within that proposition, since in the long run we expect equal numbers of all solutions within the population. Note that an analogous result holds for model (4.16).

Therefore, this long run approximation guarantee can be used to analyse certain cases of *crossover* together with a given population size; it can be seen that the lower bound on μ_t^∞ increases in value as either the population size N becomes larger (for $\theta_t > 0$) or θ_t increases, and the upper bound on μ_t^∞ decreases as either the population size or ε_t decreases; neither of these facts is intuitively obvious from the definition of GA/WI-type algorithms in Section 2. Thus, these bounds highlight the existence of *crossover bias* within GAs, since the bounds given above are independent of the costs of solutions. The existence of this bias is of fundamental importance in terms of optimisation via GAs, since for a given optimisation problem k is generally very large and the costs f are not known beforehand; the existence of crossover bias, for example towards lower cost solutions (depending on the given encoding scheme and choice of crossover operator), for example towards lower costs solutions is generally undesirable. Another important point is that these bounds do not depend on any particular representation scheme, i.e., they may prove useful for the analysis of non-binary encoding schemes, where the choice of *crossover* operator is not immediately apparent [6,10].

Finally, we note that we are justified in calculating the expected values of the

stationary distribution, since, the stochastic process converges in limit (by Theorem 4.8).

5. Explicit stochastic model for GA/WO(N, s, f, roulette wheel, mutation, crossover)

We now proceed to formulate a stochastic model for GA/WO; we proceed at a slightly faster pace than for GA/WI, partly because the resultant model is of a simpler form than for the latter, and partly because some analogous concepts with respect to vector-indexed matrix multiplication etc. need not be made here, for brevity. We present important results for GA/WO-type Genetic Algorithms with respect to the size of the state-space through which they search, sufficient conditions for convergence in limit of these algorithms, and a special case analysis which underpins results observed experimentally, in that excessive mutation rates cause a degradation in performance of the algorithm with respect to the discovery of higher fitness solutions [6,10]. We also present important results which show the relationship between GA/WO- and GA/WI-type Genetic Algorithms; we show that GA/WI and GA/WO possess differing state-spaces, that they possess differing transition probability matrices (and therefore differing search properties); finally, we show that for algorithms such as GA/WO, the *order* in which the solutions are stored within the population fundamentally affects the search dynamics of the algorithm.

5.1. Definitions of GA/WO states/operators

In GA/WO, each of the genetic operators acts upon the population separately (as opposed to GA/WI) so that each in effect is a stochastic process which produces a new population. In particular, for the *mutation* and *crossover* operators, since they are effected to *specific locations* within the population (*mutation* is applied singly to specific locations, *crossover* pairwise) we merely assume that the solutions are stored within the population in list-like fashion, as in the GA implemented in [6]. Thus, we introduce new definitions for the states of GA/WO;

Definition 5.1. Let $i = (i_1, \ldots, i_N)_t$ store the order in which the solutions are contained in the population of GA/WO at time t, where $1 \leqslant i_m \leqslant k$, $1 \leqslant m \leqslant N$.

Thus, every component of the GA/WO state vectors is some solution (whereas every component of the GA/WI state vectors *counts* some solution).

Definition 5.2. Let $S'^{[N]} = \{i = (i_1, \ldots, i_N): 1 \leqslant i_m \leqslant k, 1 \leqslant m \leqslant N\}$. Thus, $S'^{[N]}$ is the set consisting of all valid populations of GA/WO.

These definitions lead to:

Proposition 5.3. *GA/WO with population size N, operating on a solution space of size $|s| = k$ searches a state-space of size $|S'^{[N]}|$, where*

$$|S'^{[N]}| = k^N .$$ (5.1)

Proof. The total number of population states of GA/WO is given by the total number of *distinct* permutations of the population states of GA/WI, i.e., by

$$\sum_{X \in S^{[k]}} \binom{N}{X} ,$$ (5.2)

since, every state vector $X \in S^{[k]}$ of GA/WI will have the bracketed term in (5.2) distinct corresponding permutations, which, within GA/WO, are all distinct states; then the result (5.1) follows from a modification of the well-known result given in [1, p. 49]. □

Thus, we present the following inequality between GA/WI and GA/WO:

Theorem 5.4. *Let $|S^{[k]}|$ be given by (4.3), and $|S'^{[N]}|$ by (5.1). Then $|S'^{[N]}| > |S^{[k]}|$ for $k > 1$ and $N > 1$.*

Proof. Observe that $Nk > k + N - 1$ for $k, N > 1$. Thus,

$$k \cdot 2k \cdots Nk > k(k + 1) \cdots (k + N - 1) ,$$ (5.3)

so that

$$|S'^{[N]}| = k^N > \frac{k(k + 1) \cdots (k + N - 1)}{N!} = |S^{[k]}| ,$$ (5.4)

which completes the proof. □

Note that $k > 1$ corresponds to nontrivial optimisation, and $N > 1$ to a nontrivial Genetic Algorithm. Thus, the number of states of GA/WO is shown here to be strictly larger than that of GA/WI. This fact has been previously ignored within the literature; some authors give the left-hand side of (5.4) as the size of the Genetic Algorithm state-space [17]; others give the right-hand side [7]. Here, we have shown which Genetic Algorithms possess which state-space, with respect to the *implementation* of the genetic operators; the state-space of GA/WO being strictly larger than that of GA/WI.

As is evident from the definitions of GA/WI and GA/WO, the single-child crossover probability matrix is insufficient to describe the stochastic properties of the GA/WO *crossover* operator (which produces two children from two parents). Therefore, we introduce:

Definition 5.5. Let $C_{ab,gh}$ be a row-stochastic crossover probability matrix, i.e.,

$$\sum_{g=1}^{k} \sum_{h=g}^{k} C_{ab,gh} = 1 \quad \forall 1 \leq a \leq b \leq k ,$$

that stores the probability with which solutions a, b will produce solutions g, h under the action of *crossover* for $1 \leq a \leq b \leq k$, $1 \leq g \leq h \leq k$. Moreover, we define $C_{ij,ij} > 0$ (since whether or not two solutions are combined is the result of a probabilistic test [6, 10]).

Note that we force both the parent and child solutions to be gathered in an ordered pair, which again forces uniqueness of reference. The 2-parent, 2-child crossover probability matrix has the following useful property.

Proposition 5.6.

$$\sum_{g=1}^{k} \sum_{h=g}^{k} C_{ab,gh} = \sum_{g=1}^{k} \sum_{h=1}^{g} C_{ab,hg} = 1 \quad \forall 1 \leq a \leq b \leq k .$$

Proof. Setting $m = k - g + 1$, $1 \leq g \leq k$, $n = k - h + 1$, $1 \leq h \leq g \leq k$, gives

$$\sum_{g=1}^{k} \sum_{h=1}^{g} C_{ab,hg} = \sum_{m=1}^{k} \sum_{n=m}^{k} C_{ab,(k-n+1)(k-m+1)}$$

$$= \sum_{g=1}^{k} \sum_{h=g}^{k} C_{ab,gh} = 1 \quad \forall 1 \leq a \leq b \leq k ,$$

since the middle sum is merely a rearrangement of the outer two sums, and this completes the proof. □

This then gives the following useful property:

Proposition 5.7.

$$\sum_{g=1}^{k} \sum_{h=g+1}^{k} C_{ab,gh} = \sum_{g=1}^{k} \sum_{h=1}^{g-1} C_{ab,hg} \quad \forall 1 \leq a \leq b \leq k .$$

Proof. Each sum in the above is just the sums written in Proposition 5.6, excluding the term

$$\sum_{g=1}^{k} C_{ab,gg}$$

and this completes the proof. □

With these definitions and results in place, which describe in simple form the stochasticity of the GA/WO *crossover* operator, we formulate a definition which relates the selection operator of GA/WI to that of GA/WO (since both use

sampling with replacement selection operators). It is then used to formulate a transition matrix for the selection operator of GA/WO.

Definition 5.8. Let $(X_{i,b})_t$ store the numbers of solutions b within the population i of GA/WO at time t, i.e.,

$$X_{i,b} = d \Leftrightarrow \exists m_1, \ldots, m_d, \ m_y \neq m_z \quad \forall 1 \leq y \neq z \leq d$$

$$\text{s.t. } i_{m_1} = \cdots = i_{m_d} = b .$$

Thus, $(X_{i,1}, \ldots, X_{i,k})$ is a many-to-one mapping from the states of GA/WO to the states of GA/WI; it counts the total occurrences of solutions i_a, $1 \leq a \leq N$, within $i = (i_1, \ldots, i_N)$ of GA/WO so that each $(i_1, \ldots, i_N) \in S'^{[N]}$ has a unique counting vector $(X_{i,1}, \ldots, X_{i,k})$ but any given $(X_1, X_2, \ldots, X_k) \in S^{[k]}$ in Definition 4.2 can represent

$$\binom{N}{X} \tag{5.5}$$

different $(i_1, \ldots, i_N) \in S'^{[N]}$ vectors [1].

5.2. Stochastic model for GA/WO(N, s, f, roulette wheel, mutation, crossover)

We present in this section a stochastic model for GA/WO; we begin by presenting a stochastic model for the selection operator. An important point to be made is that populations $j \in S'^{[N]}$ which have the same counting vector $Y \in S^{[k]}$ will have the same probability of production from $i \in S'^{[N]}$, because within the GA/WO fitness selection operator the order in which the solutions are selected is unimportant (i.e., their ordering within population i); only the order in which they are *placed* within $j \in S'^{[N]}$ is important, by Definition 5.2, and so different permutations of $j \in S'^{[N]}$ thus have equal probability of occurrence, due to the sampling with replacement roulette-wheel operator (which implies independence between the selection events, as in GA/WI). Therefore, the relative fitness of solution b within population $i \in S'^{[N]}$ is given by

$$r_{xb} = \frac{X_{i,b} f_b}{\sum_{d=1}^{N} f_{i_d}} \quad 1 \leq b \leq k , \tag{5.6}$$

so that

$$F_{ij} = \Pr\{(j_1, j_2, \ldots, j_N) \mid (i_1, i_2, \ldots, i_N)\} = \prod_{b=1}^{k} r_{xb}^{Y_{j,b}} \tag{5.7}$$

is the transition matrix for the selection operator of GA/WO, where $i,j \in S'^{[N]}$ and where $X, Y \in S^{[k]}$ are the counting vectors for populations i and j respectively.

The transition matrix for the GA/WO *mutation* operator is of a simple form;

since *mutation* is carried out independently in list-like manner to the population, it also transforms population i to population, j, $i, j \in S'^{[N]}$,

$$M_{ij} = \Pr\{(j_1, \ldots, j_N) \mid (i_1, \ldots, i_N)\} = \prod_{d=1}^{N} \eta_{i_d j_d} . \tag{5.8}$$

The construction of the transition matrix for the *crossover* operator of GA/WO is non-trivial; therefore we make a definition which relates *crossover* probabilities to the C matrix;

Definition 5.9. Let $C^*_{ab,cd}$ be defined as

$$C^*_{ab,cd} = \begin{cases} C_{ab,cd}, & a \leqslant b, \quad c \leqslant d, \\ C_{ba,cd}, & a > b, \quad c \leqslant d, \\ C_{ab,dc}, & a \leqslant b, \quad c > d, \\ C_{ba,dc}, & a > b, \quad c > d, \end{cases}$$

i.e., $C^*_{ab,cd}$ gives the appropriate value within C for *crossover* between arbitrarily labelled solutions.

Since the *crossover* operator of GA/WO transforms population i to population j, $i, j \in S'^{[N]}$ by the creation of solutions pairwise, the following equation relates the crossover probabilities to the child solutions placed within $j \in S'^{[N]}$,

$$\Pr\{(j_{2d-1}, j_{2d}) = (g, h) \mid (i_{2d-1}, i_{2d}) = (a, b)\}$$
$$= \Pr\{(j_{2d-1}, j_{2d}) = (h, g) \mid (i_{2d-1}, i_{2d}) = (a, b)\} , \tag{5.9}$$

where $1 \leqslant d \leqslant N/2$, i.e., either ordering of the new pair of child solutions within the relevant locations of j is stipulated as equally likely, where each term in the above (5.9) is an ordered pair of solutions. Thus, we define:

Definition 5.10. Let $f(x, y)$ be defined as

$$f(x, y) = \begin{cases} 1, & x = y, \\ \frac{1}{2}, & x \neq y. \end{cases}$$

Note that f is symmetric about $x = y$. Thus, with all necessary definitions in place, observe that

$$\Pr\{(j_{2d-1}, j_{2d}) = (g, h) \mid (i_{2d-1}, i_{2d}) = (a, b)\}$$
$$= \Pr\{ab \xrightarrow[C]{} gh \cap (j_{2d-1}, j_{2d}) = (g, h)\}$$
$$= f(g, h) C^*_{ab,gh} , \tag{5.10}$$

where each term in brackets is an ordered pair. Thus, the general formula for the

conditional probability of obtaining (j_{2d-1}, j_{2d}) from (i_{2d-1}, i_{2d}) under crossover is

$$\Pr\{(j_{2d-1}, j_{2d}) \mid (i_{2d-1}, i_{2d})\} = f(j_{2d-1}, j_{2d}) C^*_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d}}, \tag{5.11}$$

$1 \leq d \leq N/2$. Denoting by Q the transition matrix for the given *crossover* operator, we thus have

$$
\begin{aligned}
Q_{ij} &= \Pr\{(j_1, \ldots, j_N) \mid (i_1, \ldots, i_N)\} \\
&= \Pr\left\{ \bigcap_{d=1}^{N/2} [(i_{2d-1}, i_{2d}) \xrightarrow{C} (j_{2d-1}, j_{2d})] \right\} \\
&= \prod_{d=1}^{N/2} [f(j_{2d-1}, j_{2d}) C^*_{i_{2-1}i_{2d}, j_{2d-1}j_{2d}}], \tag{5.12}
\end{aligned}
$$

since the pairwise crossover events are independent of one another. To complete the construction of Q, we state and prove the following:

Proposition 5.11. *Let Q be defined as in* (5.12). *Then Q is a row-stochastic matrix, i.e.*,

$$\sum_{j \in S'^{[N]}} Q_{ij} = 1 \quad \forall i \in S'^{[N]}.$$

Proof. Observe that

$$
\begin{aligned}
\sum_{j \in S'^{[N]}} Q_{ij} &= \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \cdots \sum_{j_N=1}^{k} \left[\prod_{d=1}^{N/2} [f(j_{2d-1}, j_{2d}) C^*_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d}}] \right] \\
&= \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} f(j_1, j_2) C^*_{i_1 i_2, j_1 j_2} \cdots \sum_{j_{N-1}=1}^{k} \sum_{j_N=1}^{k} f(j_{N-1}, j_N) C^*_{i_{N-1}i_N, j_{N-1}j_N},
\end{aligned}
\tag{5.13}
$$

and so the general term is

$$
\begin{aligned}
\sum_{j_{2d-1}=1}^{k} \sum_{j_{2d}=1}^{k} & f(j_{2d-1}, j_{2d}) C^*_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d}} \\
&= \frac{1}{2} \sum_{j_{2d-1}=1}^{k} \sum_{j_{2d}=j_{2d-1}+1}^{k} C_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d}} + \frac{1}{2} \sum_{j_{2d-1}=1}^{k} \sum_{j_{2d}=1}^{j_{2d-1}-1} C_{i_{2d-1}i_{2d}, j_{2d}j_{2d-1}} \\
&\quad + \sum_{j_{2d-1}=1}^{k} C_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d-1}}
\end{aligned}
$$

(via the properties of $f(x, y)$ and assuming $i_{2d-1} \leq i_{2d}$ without loss of generality, since the sum is over j components), where $1 \leq d \leq N/2$. By Proposition 5.7, the first two sums are equal, and, added to the third sum, this gives

$$\sum_{j_{2d-1}=1}^{k} \sum_{j_{2d}=1}^{k} f(j_{2d-1}, j_{2d}) C^{*}_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d}}$$

$$= \sum_{j_{2d-1}=1}^{k} \sum_{j_{2d}=j_{2d-1}+1}^{k} C_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d}} + \sum_{j_{2d-1}=1}^{k} C_{i_{2d-1}i_{2d}, j_{2d-1}j_{2d-1}} = 1$$

by Definition 5.5. Since d was arbitrarily chosen, (5.12) is shown to be a finite product of 1's, and so Q is indeed row-stochastic. $\square$

Note that F is easily shown to be row-stochastic using the properties of the multinomial distribution, and the above analysis for Q can be repeated in a simpler form for M, to show that it is also row-stochastic. Thus, the overall stochastic model for GA/WO is

$$P_{ij} = (F \cdot M \cdot Q)_{ij} = \sum_{a \in S'^{[N]}} \left(\sum_{b \in S'^{[N]}} (F_{ib} M_{ba}) Q_{aj} \right) \quad \forall i, j \in S'^{[N]}, \tag{5.14}$$

where F, M and Q are row-stochastic matrices defined in (5.7), (5.8) and (5.12), and vector-indexed matrix multiplication is used throughout. Thus, it can be seen that the transition matrices of GA/WI (i.e., (4.15) and (4.16)) and GA/WO (i.e. (5.14)) differ; generally, the transition matrix of GA/WO is of higher order than that of GA/WI. Since knowledge of the transition matrix of a stochastic process is sufficient to predict the behaviour of that process at all time steps, given any initial probability distribution of the corresponding Markov Chain [15], this implies that GA/WO and GA/WI are quite distinct search algorithms; the models (4.15) and (4.16) and (5.14) capture this distinction in a concise form. Notice also that distinct orderings of the *solutions* with $i \in S'^{[N]}$ are essentially differing states of GA/WO, with corresponding differing transition probabilities within (5.14); thus, the *order* in which the solutions are stored within the populations of GA/WO determines the search dynamics of this algorithm.

5.3. Analysis of model of GA/WO(N, s, f, roulette wheel, mutation, crossover)

In this section we present two fundamental analyses of the model given in (5.14). We show that, if the *mutation* operator is implemented in the usual way [6, 17] that GA/WO also possesses a stationary distribution; we then show that, if each row of the mutation probability matrix is the discrete uniform distribution over the set of solutions $s = \{s_1, s_2, \ldots, s_k\}$, then the search conducted by GA/WO converges immediately to cost-independent search.

5.3.1. Parameter choice for convergence
We proceed to present a similar result to that of Section 4.4.1, in that we derive sufficient conditions for GA/WO to converge in limit towards a stationary distribution. We first formulate definitions essential to the method of proof:

Definition 5.12. Let $f_{\min} = \min\{f_1, \ldots, f_k\} > 0$.

Thus, f_{min} is equal to the cost of the solution(s) with lowest cost.

Definition 5.13. Let $f_{max} = \max\{f_1, \ldots, f_k\} > 0$.

Thus, f_{max} is equal to the cost of the solution(s) with largest cost.

Definition 5.14. Let $\eta_{min} = \min_{i,j} [\eta_{ij}]$.

Thus, η_{min} is the smallest entry in the mutation probability matrix. It is greater than zero when $\eta > 0$.

Definition 5.15. Let $C_{min} = \min_{i \leqslant j} [C_{ij,ij}] > 0$.

Thus, C_{min} is the smallest probability that a pair of parents is retained under the GA/WO *crossover* operator. It is greater than zero by Definition 5.5.

Theorem 5.16. *GA/WO with $\eta > 0$, and C as defined, converges towards a stationary probability distribution, q', where q'_i is strictly positive for all i, where $i is a valid population state of the algorithm, i.e.,*

$$\lim_{t \to \infty} P^t = \begin{bmatrix} q' \\ q' \\ \vdots \\ q' \end{bmatrix}. \tag{5.15}$$

Also, convergence takes place at least geometrically, i.e., there exists $0 < r < 1$ s.t.

$$P_{ij}^{(t)} = q'_j + e_{ij}^{(t)}, \tag{5.16}$$

where

$$|e_{ij}^{(t)}| \leqslant r^t. \tag{5.17}$$

Finally, q' is independent of the initial choice of population $i_{(0)}$.

Proof. Observe that

$$P_{ij} = (F \cdot M \cdot Q)_{ij} \geqslant F_{ii} M_{ij} Q_{jj} \quad \forall i, j \in S'^{[N]}, \tag{5.18}$$

where vector-indexed multiplication is carried out in the general sense of Definitions 4.3 and 4.4. Now

$$F_{ii} = \prod_{b=1}^{k} r_{xb}^{X_{i,b}},$$

so that $r_{xb} \geqslant f_{min}/(N f_{max})$ where at least one component of i has value b, $\forall 1 \leqslant b \leqslant k$ and $\forall X \in S^{[k]}$, since the numerator of the given fraction is smaller in fitness than all other solutions, and the denominator is the cost of the population with maximum fitness; thus,

$$F_{ii} = \prod_{b=1}^{k} r_{xb}^{X_{i,b}} \geq \left(\frac{f_{\min}}{Nf_{\max}}\right)^{N} > 0 \quad \forall i \in S'^{[N]} , \tag{5.19}$$

$$M_{ij} \geq (\eta_{\min})^{N} > 0 \quad \forall i, j \in S'^{[N]} \tag{5.20}$$

(since $\eta > 0$)

$$Q_{jj} \geq \left(\frac{C_{\min}}{2}\right)^{N/2} > 0 \quad \forall j \in S'^{[N]} \tag{5.21}$$

(by definition of Q), which together imply that the product given on the right-hand side of (5.18) is strictly positive, $\forall i, j \in S'^{[N]}$. Since P is then irreducible and aperiodic, the result follows by use of Theorems 3.3 and 3.4.

5.3.2. Special case

In the following special case of GA/WO, we show that both algorithms described in this paper contain similar degenerate cases; thus, the following result shows that, in the case where GA/WI is parameterised as in Proposition 4.9, and GA/WO as below, that both algorithms have identical search properties with respect to the absence of bias towards cost; however, the choice of which operator to parameterise is different.

Theorem 5.17. *Let* $\eta = [1/k]_{k \times k}$. *Then the resultant stationary distribution for GA/WO is*

$$q_{j} = \frac{1}{k^{N}} \sum_{a \in S'^{[N]}} Q_{aj} , \quad j \in S'^{[N]} . \tag{5.22}$$

Proof. By substitution of η into (5.8), it follows that

$$M_{ij} = \frac{1}{k^{N}} \quad \forall i, j \in S'^{[N]} . \tag{5.23}$$

Thus, by substitution into (5.14),

$$P_{ij} = \frac{1}{k^{N}} \sum_{a \in S'^{[N]}} Q_{aj} \quad \forall i \in S'^{[N]} , \tag{5.24}$$

$\forall i, j \in S'^{[N]}$ since F is row-stochastic. Substitution of (5.22) into the invariant equation (3.8) with P given by (5.24) obtains the result, observing by definition of Q that

$$q_{j} \geq \frac{1}{k^{N}} Q_{jj} > 0 \quad \forall j \in S'^{[N]} . \tag{5.25}$$

The corresponding result for immediate convergence follows; essentially, this is because P in (5.24) has identical rows, as does P of Proposition 4.9. □

Within a relaxed theoretical framework of discussion, we see the surprising result that GA/WI degenerates immediately to "fitness-free" search for "com-

plete" *crossover* regardless of *mutation*, and GA/WO degenerates immediately to "fitness-free" search for "complete" *mutation* regardless of *crossover*. It is worthy of note that q above in (5.22) is discrete uniform $(= [1/k^N]$ over $S'^{[N]})$ in the case where no crossover actually takes place (i.e., crossover parameter $p_c = 0$ in [10]), since Q in this case is the $|S'^{[N]}| \times |S'^{[N]}|$ identity matrix I. Again, as for GA/WI, there may be other parameterizations for GA/WO which produce cost-independent search; as in the case for the former, we sought to prove an existence result here.

The extension of the approximation guarantee (which measures crossover bias) for GA/WI, given in Section 4.4.3 to GA/WO-type algorithms is currently being developed by the authors, and will be reported in a future publication; it should be pointed out that the distributions given in the rows of the matrix defined by (5.14) are non-standard probability distributions, and thus introduce into this problem an element of intractability.

6. Conclusions and future directions

In this paper we have described a number of models for GAs, which were then used to analyse the convergence properties of the given algorithms. We defined, in terms of stochastic matrices, genetic mutation and crossover operators which captured their essential stochastic properties in such a way that the resultant models were independent of the representation of solutions, yet yielded sufficient conditions for convergence. We made an important distinction between GAs which sample solutions with and without replacement from populations in terms of the size of the state-space of both algorithms. We presented, via definitions of genetic operators, stochastic models for the two classes of algorithm. We then analysed particular cases of genetic search, showing sufficient conditions for both classes for degeneracy to randomised search, and presented an approximation guarantee for optimisation via sampling with replacement GAs; this guarantee indicates the existence of bias within the GA crossover operator towards given solutions.

Future important theoretical work will concentrate on the presentation of the stationary distributions of GAs with parameters other than those used in our special case analyses, for both classes of algorithm. The extension of the estimation of the first moments of the stationary distributions of GA/WI-type GAs to higher moments represents important theoretical and practical work; the estimation of these moments is significant, due to the multivariate nature of the stationary distribution of the algorithms, and provides guidance with respect to algorithm parameterisation, as shown in Sections 4.4.2 and 4.4.3. It may also be possible to estimate these moments using particular cost functions, mutation and crossover probability matrices. Other important work concerns the analysis of the phenomenon known as "premature convergence of Genetic Algorithms" [3, 6, 10], together with the rate of convergence of Genetic Algorithms towards their stationary distributions [18], by use of these models; this is currently being investigated by the authors, and will be reported in a future publication. These

tasks represent important problems with respect to the theoretical and practical aspects of optimisation via Genetic Algorithms.

Acknowledgements

Appendix A. Example of crossover probability matrix, binary string encoding scheme of length 2

Example A.1 (Binary string uniform crossover [3, 16]). This can be implemented by simultaneously moving along both parents and selecting a bit from either parent with probability $1/2$. Thus, in this case, for strings of length 2, the crossover probability matrix is given in Table A.1.

Table A.1
Uniform crossover probability matrix for binary string ($L = 2$) GA

$C_{ij,m}$	00	01	10	11
00 00	1	0	0	0
00 01	0.5	0.5	0	0
00 10	0.5	0	0.5	0
00 11	0.25	0.25	0.25	0.25
01 01	0	1	0	0
01 10	0.25	0.25	0.25	0.25
01 11	0	0.5	0	0.5
10 10	0	0	1	0
10 11	0	0	0.5	0.5
11 11	0	0	0	1

References

[1] I. Anderson, *A First Course in Combinatorial Mathematics* (OUP, Oxford, 1985).
[2] D.L. Battle and M.D. Vose, Isomorphisms of genetic algorithms, *Artif. Intell.* **60** (1993) 155–165.
[3] D. Beasley, D.R. Bull and R.R. Martin, An overview of Genetic Algorithms: Part 2, research topics, *Univ. Comput.* **15** (1993) 170–181.
[4] R.N. Bhattacharya and E.C. Waymire, *Stochastic Processes with Applications* (Wiley, New York, 1990).
[5] E. Cinlar, *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

[6] L. Davis, What is a Genetic Algorithm, in: L. Davis, ed., *Handbook of Genetic Algorithms* (Van Nostrand Reinhold, New York, 1991) Chapter 1.

[7] T.E. Davis and J.C. Principe, A simulated annealing like convergence theory for the simple genetic algorithm, in: R.K. Belew and L.B. Booker, eds., *Proceedings of the Fourth Conference on Genetic Algorithms*, San Mateo, CA (Morgan Kaufmann, San Mateo, CA, 1991) 174–181.

[8] A.E. Eiben, E.H.L. Aarts and K.M. Van Hee, Global convergence of genetic algorithms: a Markov chain analysis, in: H.P. Schwefel and R. Manner, eds., *Parallel Problem Solving from Nature* (Springer, Berlin, 1991) 4–12.

[9] D.E. Goldberg, Finite Markov chain analysis of Genetic Algorithms, in: J.J. Grefenstette, ed., *Proceedings of the Second International Conference on Genetic Algorithms* (Lawrence Erlbaum, Hillsdale, NJ, 1987) 1–8.

[10] D.E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning* (Addison-Wesley, Reading, MA, 1989).

[11] A.F. Karr, *Probability* (Springer-Verlag, New York, 1993).

[12] W. Mendenhall, D.D. Wackerly and R.L. Scheaffer, *Mathematical Statistics with Applications* (Wiley, New York, 1993).

[13] H. Minc, *Nonnegative Matrices* (Wiley, New York, 1988).

[14] D.J. Montana, Automated parameter tuning for interpretation of synthetic images, in: L. Davis, ed., *Handbook of Genetic Algorithms* (Van Nostrand Reinhold, New York, 1991).

[15] U. Narayan Bhat, *Elements of Applied Stochastic Processes* (Wiley, New York, 2nd ed., 1984).

[16] J.C. Potts, T.D. Giddens and S.B. Yadav, The development and evaluation of an improved Genetic Algorithm based on migration and artificial selection, *IEEE Trans. Syst. Man Cybern.* **24** (1) (1993) 73–86.

[17] G. Rudolph, Convergence analysis of canonical Genetic Algorithms, *IEEE Trans. Neural Net.* **5** (1) (1994) 96–101.

[18] E. Seneta, *Non-negative Matrices and Markov Chains* (Springer-Verlag, New York, 2nd ed., 1981).

[19] J.H. van Lint and R.M. Wilson, *A Course in Combinatorics* (CUP, Cambridge, 1993).