ELSEVIER

Theoretical Computer Science 200 (1998) 101–134

Theoretical Computer Science

Fundamental Study

Linear analysis of genetic algorithms

Lothar M. Schmitt*, Chrystopher L. Nehaniv, Robert H. Fujii

School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu City, Fukushima 965-80, Japan

Received March 1997; revised November 1997
Communicated by M. Ito

Abstract

We represent simple and fitness-scaled genetic algorithms by Markov chains on probability distributions over the set of all possible populations of a fixed finite size. Analysis of this formulation yields new insight into the geometric properties of the three phase mutation, crossover, and fitness selection of a genetic algorithm by representing them as stochastic matrices acting on the state space. This indicates new methods using mutation and crossover as the proposal scheme for simulated annealing. We show by explicit estimates that for small mutation rates a genetic algorithm asymptotically spends most of its time in uniform populations regardless of crossover rate. The simple genetic algorithm converges in the following sense: there exists a fully positive limit probability distribution over populations. This distribution is independent of the choice of initial population. We establish strong ergodicity of the underlying inhomogeneous Markov chain for genetic algorithms that use any of a large class of fitness scaling methods including linear fitness scaling, sigma-truncation, and power law scaling. Our analysis even allows for variation in mutation and crossover rates according to a pre-determined schedule, where the mutation rate stays bounded away from zero. We show that the limit probability distribution of such a process is fully positive at all populations of uniform fitness. Consequently, genetic algorithms that use the above fitness scalings do *not* converge to a population containing only optimal members. This answers a question of G. Rudolph (IEEE Trans. on Neural Networks 5 (1994) 96–101). For a large set of fitness scaling methods, the limit distribution depends on the pre-order induced by the fitness function f, i.e. $c \geqslant c' \iff f(c) \geqslant f(c')$ on possible creatures c and c', and not on the particular values assumed by the fitness function.

Keywords: Stochastic optimization; Fitness-scaled genetic algorithms; fitness-rank dependence; Markov chain model; Spectral analysis of stochastic matrices

* Corresponding author. E-mail: lothar@u-aizu.ac.jp.

0304-3975/98/$19.00
PII S0304-3975(98)00004-8

Contents

1. Introduction

Holland [13] introduced genetic algorithms as a search and optimization method based upon adaptation principles of nature. One way to look at genetic algorithms is to see them as function optimizers, cf. [3]. Given are a finite collection C of creatures in a model "world" and a function $f : C \to \mathbb{R}^+$. The task is to find an element $c \in C$ such that $f(c)$ is maximal. Usually, the number of elements in C is very large prohibiting an exhaustive search. Genetic algorithms provide a probabilistic way to conduct a blind search in C for arbitrary f given a suitable encoding of creatures into strings of symbols. A genetic algorithm comprises three operations: mutation, crossover and fitness selection. These are applied cyclically and iteratively to fixed-size finite populations consisting of elements of C. Mutation and crossover model analogous phenomena for DNA strings, while fitness selection models reproductive success of adapted organisms.

Our model for genetic algorithms is a Markov chain model. Previously, such models for genetic algorithms without fitness scaling have been developed in [5, 6, 11, 14, 23, 25, 31, 33]. In the work of Davis and Principe [5, 6] the main point of consideration is whether or not annealing the mutation rate to zero in the simple genetic algorithm implies convergence to global optima. It does not. However, strong ergodicity of the resulting non-stationary Markov chain is established. Another comprehensive model for the simple genetic algorithm based upon Markov chain analysis can be found in the work of Vose, Liepins, and Nix [23, 31, 33]. The main advantage of their model is that the fitness selection is modelled as a diagonal matrix acting on a suitable state space. The price paid for such an approach is that the operator describing mutation and crossover combined is given by a vector valued quadratic form.

The starting point of our investigation is the formulation of a much more convenient description of the three components of genetic algorithm. Our model is based on the

fact that populations are represented in the computer as bit-strings, and it uses as its state space probability distributions in the free vector space over all such bit-string populations. In contrast to other methods, this approach can easily be used to model the effects of spatial structure on an evolving population. By separate description of the three phases mutation, crossover and fitness selection of a genetic algorithm, we can apply spectral theory and other techniques to the matrices that arise. We obtain explicit bounds for the eigenvalues of the mutation and crossover operators, descriptions of the fixed points they contract towards, and characterizations of their invariant subspaces. This characterizes mutation and crossover as procedures with geometric rates of convergence towards uniformly distributed probability distributions over all populations in associated invariant subspaces. In particular, Proposition 10 shows how crossover assists mutation in the averaging process. We remark that we have discovered a new link between the crossover operation and representations of the group of permutations of a finite set. Our analysis also shows by explicit probabilistic estimates how scaled or unscaled fitness selection drives the algorithm towards uniform populations i.e. populations that contain only multiple copies of one creature. This offers an approach to genetic drift in finite populations, a phenomenon well-known to population geneticists (e.g. [22, 26]). The above-mentioned probabilistic estimates for fitness selection and a corresponding similar analysis for crossover combined with mutation are used in Theorem 15 to show convergence of the simple genetic algorithm in the following sense: there exists a fully positive limit probability distribution over populations, independent of the choice of initial population. We give explicit bounds on the combined probability for non-uniform populations in the limit distribution. In fact, Theorem 15 shows that for small mutation rates a simple genetic algorithm asymptotically spends most of its time in uniform populations. This sheds some light on a discussion of "punctuated equilibrium" by Vose [32], who states that a simple genetic algorithm is near local optima most of the time, visiting every state infinitely often. We discuss how to use crossover-mutation in a new way as a proposal scheme for a new, convergent, genetic variant of the simulated annealing optimization method.

Let us come back for a moment to other models for genetic algorithms. In contrast to our approach, [5, 6, 23, 31–33] model populations as unordered multi-sets. As a consequence, these models of genetic algorithms can be obtained as projections of our model induced simply by forgetting the order on populations. Our model frees the initial description from clumsy, combinatorial coefficients which may hinder a subsequent detailed analysis. In particular, the bilinear approach of [32] to crossover is replaced in our model by an appropriate tensor product construction. The linear model of simple genetic algorithms we discovered was found also independently by Rudolph [25], who proves a part of our Theorem 15 and analyses a convergent variant of the simple genetic algorithm by extending all linear operators to keep track of a best-so-far individual seen by the algorithm. Our analysis does not treat recording the best-so-far individual, but otherwise extends Rudolph's findings for simple genetic algorithms. In particular, we obtain estimates for the coefficient of ergodicity for simple genetic algorithms, if multiple-bit mutation is considered.

The final part of our analysis contains the most striking new results, concerning the properties of fitness-scaled and variation-scheduled genetic algorithms. By "variation-scheduled", we mean that besides possible fitness scaling, the mutation as well as the crossover rates are changed according to a schedule fixed in advance. We treat most standard methods of fitness scaling such as linear fitness scaling, sigma-truncation and power law scaling. Convergence of fitness-scaled genetic algorithms was posted as an open question by Rudolph [25]. Our approach includes a detailed analysis on convergence of fitness-scaled genetic algorithms which answers his question in the negative for convergence to an optimal individual or even a population containing one, and in the positive in the sense of convergence to a unique probability distribution over all populations. In fact, we show that the limit probability distribution of such processes is fully positive at populations of uniform fitness. Moreover, these results hold even when crossover and mutation rates are varied according to a fixed schedule (see Section 3 below for the precise formulation). Quite surprisingly and quite strikingly, for a large set of fitness scaling methods the limit distribution is independent of the particular method of scaling but depends rather on the pre-order on the set of creatures induced by the fitness function and not the particular values of the fitness function.

We establish strong ergodicity of the underlying inhomogeneous Markov chain for genetic algorithms that use multiple-bit mutation and the above-mentioned fitness scaling and variation scheduling methods. In order to establish this result, we have included a customized version of a result by Gidas [7, Theorem 1.1] with significantly simplified proof.

1.1. Notation and preliminaries

Throughout this paper the following notation will be used:

Scalars: $\mathbb{N}, \mathbb{R}, \mathbb{R}^+$, and $\mathbb{C}$ will stand for the *strictly positive integers*, the *real numbers*, the *positive real numbers* including 0, and the *complex numbers*, respectively. Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. If $t \in \mathbb{R}$, then let $t^+ = \frac{1}{2}(t + |t|)$. If $r \in \mathbb{R}$, then $\lfloor r \rfloor$ (the *floor* of r) is the largest integer not strictly larger than r.

Vectors: We shall interpret bit vectors of length ℓ as corners of the positive unit cube in $\mathbb{R}^n$ in the obvious way. If not indicated otherwise, any given finite dimensional complex vector space V with a fixed, ordered basis P of length n will be identified canonically with $\mathbb{C}^n$, having its standard basis and standard inner product. If W is a subspace of V generated by a subset P' of P, then e_W is, by definition, the vector in W such that $\langle e_W, p\rangle = 1$, for every $p \in P'$. In particular, we set $e = e_{\mathbb{C}^n}$. The vector e is an invariant vector for any row stochastic matrix acting on the left. Let $\mathbf{P}_e$ denote the orthogonal projection onto the subspace generated by e.

If $\mathbf{X} : V \to V$ is a linear map and $p, q \in P$, then $(\mathbf{X})_{p,q} = \langle \mathbf{X}p, q\rangle$ will denote the (p,q) entry in the matrix representing $\mathbf{X}$ with respect to basis P.

If V_k are vector spaces with bases P_k, $k = 1,2$, then the tensor space $V_1 \otimes V_2$ is the free vector space over the basis $P_1 \times P_2$. Elements in the basis $P_1 \times P_2$ are denoted

by $p_1 \otimes p_2$, $p_k \in P_k$. If $v_k = \sum_{p \in P_k} \lambda_p^{(k)} p$, then

$$v_1 \otimes v_2 = \sum \lambda_{p_1}^{(1)} \lambda_{p_2}^{(2)} (p_1 \otimes p_2).$$

Norms and the Hamming Metric: We shall equip $\mathbb{C}^n$ will the usual ℓ^r-norms, $1 \leqslant r \leqslant \infty$ (see [29, p. 4]). Namely, for $\xi = (\xi_1, \ldots, \xi_n)$, the r-norm of ξ is

$$\|\xi\|_r = \left(\sum_{i=1}^{n} |\xi_i|^r \right)^{1/r}.$$

In particular, we shall use the ℓ^1- or *Hamming norm* ($r = 1$) and the *Euclidean norm* ($r = 2$). The *Hamming metric* on $\mathbb{C}^n$ induced by the Hamming norm is denoted by Δ.

Matrices: $\mathbb{M}_n$ will denote the $n \times n$ matrices with entries in $\mathbb{C}$. The *spectrum* of a matrix $\mathbf{X} = (X_{i,j})$ will be denoted by $sp(\mathbf{X})$. A matrix is *self-adjoint* (or *Hermitian*) if $X_{ij} = \overline{X_{ji}}$ for all $1 \leqslant i, j \leqslant n$. A matrix will be called *C^*-positive*, if it is the square of a self-adjoint matrix (see Corollary B.3 in Appendix B.1). A matrix will be called *fully positive*, if its entries are all non-zero positive numbers. A matrix is called [*column*] *stochastic*, if its entries are non-negative reals and its columns all sum to 1. A matrix is called *row stochastic*, if its transpose is column stochastic. If $\mathbf{X}$ is a stochastic matrix such that $\mathbf{X}^k$ is fully positive for some $k \in \mathbb{N}$, then 1 is a simple root of the characteristic polynomial of $\mathbf{X}$, and $\mathbf{X}$ has a fully positive, uniquely determined fixed point of Hamming norm 1. If such an $\mathbf{X}$ has at least one strictly positive diagonal element, then 1 is the only eigenvalue of modulus 1. For a proof of these facts, see [29, Proposition I.6.2, Proposition I.6.3, Theorem I.6.5, and p. 23, Corollary 2]. $\mathbb{1}_n$ will denote the *identity matrix* in $\mathbb{M}_n$. If $\mathbf{X} \in \mathbb{M}_n$ and $\mathbf{Y} \in \mathbb{M}_m$, then we set

$$\mathbf{X} \oplus \mathbf{Y} = \begin{pmatrix} \mathbf{X} & 0 \\ 0 & \mathbf{Y} \end{pmatrix}, \text{ and } \mathbf{f} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{M}_2.$$

In addition, the tensor product $\mathbf{X} \otimes \mathbf{Y}$ of matrices is defined by action on the basis of the tensor space via

$$(\mathbf{X} \otimes \mathbf{Y})(e_i \otimes f_j) = \mathbf{X}e_i \otimes \mathbf{Y}f_j,$$

where $e_i \in \mathbb{C}^n$ $(i = 1, \ldots, n)$ and $f_j \in \mathbb{C}^m$ $(j = 1, \ldots, m)$ are standard bases.

If $\mathbf{X}, \mathbf{X}' \in \mathbb{M}_n$ commute, then by [27, Theorem 11.23], $sp(\mathbf{X} + \mathbf{X}') \subseteq sp(\mathbf{X}) + sp(\mathbf{X}')$ and $sp(\mathbf{X}\mathbf{X}') \subseteq sp(\mathbf{X}) \cdot sp(\mathbf{X}')$, where sum (resp. product) of two sets of numbers means all possible sums (resp. products) obtained by taking one term from each of the sets. As a consequence, the matrix $\mathbf{X} \otimes \mathbb{1}_m + \mathbb{1}_n \otimes \mathbf{Y}$ has spectrum contained in $sp(\mathbf{X}) + sp(\mathbf{Y})$. Furthermore, if v and w are eigenvectors to eigenvalues λ and η of $\mathbf{X}$ and $\mathbf{Y}$, respectively, then $v \otimes w$ is an eigenvector to eigenvalue $\lambda + \eta$, hence equality holds:

$$sp(\mathbf{X} \otimes \mathbb{1}_m + \mathbb{1}_n \otimes \mathbf{Y}) = sp(\mathbf{X}) + sp(\mathbf{Y}).$$

Similarly, $sp(\mathbf{X} \otimes \mathbf{Y}) = sp(\mathbf{X}) \cdot sp(\mathbf{Y})$.

We shall denote the norm of $\mathbf{X} \in \mathbb{M}_n$ induced by the ℓ^r-norm on $\mathbb{C}^n$ as $\|\mathbf{X}\|_r$. That is,

$$\|\mathbf{X}\|_r = \sup_{v \neq 0} \frac{\|\mathbf{X}v\|_r}{\|v\|_r}.$$

If $\mathbf{X} \in \mathbb{M}_n$ is column stochastic, then $\|\mathbf{X}\|_1 = \|\mathbf{X}^*\|_\infty = 1$ as an easy calculation shows (or see [29, p. 5]).

Coefficients of Ergodicity: If $\mathbf{X} = (X_{ij}) \in \mathbb{M}_n$ is column stochastic, then let the *coefficient of ergodicity with respect to the r-norm*, is defined to be (see [28, Lemma 4.2, p. 138]):

$$\tau_r(\mathbf{X}) = \max\{\|\mathbf{X}v\|_r : v \in \mathbb{R}^n, v \perp e \text{ and } \|v\|_r = 1\}.$$

A useful fact from [28, p. 137] gives:

$$1 - \tau_1(\mathbf{X}) = \min_{1 \leqslant k_1, k_2 \leqslant n} \left\{ \sum_{i=1}^{n} \min(X_{i,k_1}, X_{i,k_2}) \right\} \geqslant \sum_{i=1}^{n} \min_{1 \leqslant k \leqslant n} X_{i,k}.$$

1.2. The idea of genetic algorithms and its linear analysis

The genomes of a fixed-size population of "creatures" are modelled as binary strings. The strings may represent candidate solutions for a *fitness function* (or *objective function*) f to be optimized (minimized or maximized). Initially, the population is selected randomly and for most functions f will do very poorly. A digital version of evolution is applied in which genetic variation is introduced and fitness-proportional selection is applied stochastically to the population to obtained a new population that is (hopefully) better at optimizing the fitness function.

A simple genetic algorithm is comprised of three iterated phases: mutation, crossover, and fitness selection. We shall consider two kinds of mutation: one-bit and multiple-bit mutation. We also analyze two types of crossover including (simple) crossover as defined in [10, p. 64]. (See also [19, p. 8], [13, p. 97] and below.) The class of fitness selection operators considered includes a broad range of time varying selection schemes such as linear fitness scaling [10, p. 79] and power law fitness scaling [10, pp. 123–134]. In what follows, we shall develop separate linear representations for the three iterated phases. Developing such representations makes it possible to apply spectral theory to the linear operators that arise. Treating the three phases separately yields insight into the geometry of the various genetic operators so that one can see what each is doing on the underlying space of probability distributions over populations. Combining these insights with some more analysis illuminates the longterm behaviour of genetic algorithms with respect to questions of convergence (in probability distribution space), i.e. ergodicity, and non-convergence to an optimal solution.

1.3. The state space of a genetic algorithm

Let $\ell > 1$ be a fixed integer, and let $C = \mathbb{N}_0 \cap [0, 2^\ell - 1]$. We shall consider C as the set of all possible *creatures* in a given "world" that is modelled. For a given

world in which a finite set of creatures is possible, these creatures can be mapped in a one-to-one fashion to the set C of length ℓ binary strings for sufficiently large ℓ. If, in the process of using such an encoding, a non-admissible creature (i.e., one not in the image of this embedding) is generated by mutation or crossover, then we may assign it a very low fitness value (see below).[1] Furthermore, specific properties one would like to single out in a particular model can be given reserved bits in a binary encoding. Thus a binary encoding suffices for modelling the finite possible set of creatures in their world.[2]

Let $s \in \mathbb{N}$, $s \geqslant 2$, and let $P = C^s$. P is the set of all possible *populations* of size s. The population size s is fixed during the running of a genetic algorithm, and so P is the set of all its possible states. Every population $p \in P$ can be seen as a bit vector of length

$$L = \ell \cdot s.$$

If each such bit vector is identified canonically with a binary encoded positive integer, then this induces a *canonical total order on P*.

Let Π_s denote the group of permutations of the set $\{1,\ldots,s\}$. Then Π_s acts naturally on P by exchanging creatures according to their positions in the population.[3]

Let $p = (c_1, c_2, \ldots, c_s) \in P$, $c_i \in C$, $1 \leqslant i \leqslant s$. We define the *mean vector* [or *gene frequency vector*] of p to be be vector in $\mathbb{R}^\ell$ given by

$$mean(p) = \frac{1}{s}\sum_{i=1}^{n} c_i.$$

If $mean(p)$ is an element of the interior of the positive unit cube in $\mathbb{R}^\ell$, then the population p will be called *rich* (or, alternatively, *polymorphic for all loci*). If $mean(p)$ is an extreme point (corner) of the unit cube in $\mathbb{R}^\ell$, then p will be called *uniform* [*for all loci*].

By analogy to the above, define for $p = (c_1, c_2, \ldots, c_s) \in P$, $c_i \in C$, and s even:

$$Mean(p) = (\tfrac{1}{2}(c_1 + c_2), \ldots, \tfrac{1}{2}(c_{s-1} + c_s)) \in (\mathbb{R}^\ell)^{\frac{1}{2}s}.$$

One can think of $Mean(p)$ as a gene frequency vector for local pairs in p.

---

[1] We observe that in the case of encoding the four DNA nucleotides into binary, this problem does not arise if one encodes each nucleotide as two bits.

[2] The performance of the genetic algorithm generally depends on the particular encoding into the fixed length binary strings. Indeed, in applications, the problem is often to find an appropriate encoding for possible solutions to a given optimization problem. The analysis carried out here begins after such an encoding has been fixed.

[3] Note that our representation distinguishes populations in which the order of creatures is different. This is natural as it corresponds to what actually occurs in computer memory when a genetic algorithm is running. Another advantage of this is that one can naturally introduce *spatial structure* into our model since one may identify position in the population with a location in a world of any topology a researcher would like to model. The study of evolution of spatially structured populations is an area of active interest for both computer scientists [19] and evolutionary ecologists and population geneticists [26]. This represents a substantial advantage of our model over those of [5, 6, 23, 31–33] which do not capture any spatial structure. Moreover, their models may be obtained as simple projections of ours, so there is much to gain and nothing to lose in keeping track of order of individuals in the population.

Let V be the free vector space over the set P with canonical basis $\{p : p \in P\}$. The dimension of V is 2^L, which is very large even for small ℓ and s. Thus in practice, our model (using matrices acting on V) allows for no *direct* representation on a real world machine. Nevertheless, the results of rigorous analysis using our representation apply to genetic algorithms running on real world machines (compare [14, p. 2]).

The action of Π_s induces a canonical representation Π_s by linear transformations on V. We shall identify $\pi \in \Pi_s$ with its (unique) induced transformation on V. Let $\mathbf{P}_{\Pi_s}$ be the projection of V onto the subspace of vectors invariant under the action of Π_s.

The *mean* of populations canonically induces a linear map $mean : V \to \mathbb{C}^\ell$.

The *space over the uniform populations* U is defined as

$$U = span_{\mathbb{C}}\{p : p \in P,\ p \text{ is uniform}\}.$$

For a given $\xi \in \mathbb{R}^\ell$, let V_ξ, the *Hardy–Weinberg space of populations with gene frequency* ξ, be defined as

$$V_\xi = span_{\mathbb{C}}\{p : p \in P,\ mean(p) = \xi\}.$$

Define $\overline{V}_\xi$ analogously, if $\xi = Mean(p)$ for a population p of even size. These alternative definitions will be useful in analyzing the linear geometry of various crossover operators.

In addition, let

$$D = span_{\mathbb{C}}\{e_{V_\xi} : \xi = mean(p),\ \text{for some } p \in P, \xi \in \mathbb{R}^\ell\}.$$

Let $\mathbf{P}_U$, $\mathbf{P}_\xi$, and $\mathbf{P}_D$ be the orthogonal projections of V onto the subspaces U, V_ξ, and D, respectively. Observe that $\mathbf{P}_e\mathbf{P}_D = \mathbf{P}_e$ since $e \in D$.

Elements in

$$S = \{v \in V : \|v\|_1 = 1 \text{ and } \forall p \in P, \langle v, p\rangle \geqslant 0\}.$$

are *probability distributions over all possible states of the genetic algorithm*. The pth component $\langle v, p\rangle$ of $v \in S$ is naturally interpreted as the probability that the genetic algorithm current assumes state $p \in P$. One can naturally identify a population p with the probability distribution v with $\langle v, q\rangle = 0$ for all $p \neq q \in P$ and $\langle v, p\rangle = 1$. In our representation of genetic algorithms, we interpret the actions of the operators *mutation*, *crossover*, and *fitness selection* as column stochastic matrices operating via matrix multiplication on S. Thus, the evolution of the genetic algorithm will be described as a sequence of elements of S, while the algorithm itself is represented by a sequence (of products) of linear, stochastic operators.

2. Geometry of genetic algorithms

2.1. Mutation

One-bit mutation. Let $\mu \in (0, 1/L)$. In one-bit mutation, a mutation in a single bit occurs in the population with probability $L\mu$. If the mutation occurs, then a single bit

is chosen at random with equal probability for all possible positions. The bit at the chosen position is (logically) negated.

One-bit mutation is easy to implement and has a particularly simple mathematical description. It corresponds closely to what in biology is called a "point mutation" in a single individual of a population.

Proposition 1. *One-bit mutation can be described as self-adjoint, stochastic matrix* $\mathbf{M}_\mu$ *acting on* V.

1. *Let* $p,q \in P$. *The coefficients of* $\mathbf{M}_\mu$ *are as follows:*
   - $\langle \mathbf{M}_\mu p, p\rangle = 1 - L\mu$.
   - *If* $\Delta(p,q) = 1$, *then* $\langle \mathbf{M}_\mu p, q\rangle = \mu$.
   - *In any other case,* $\langle \mathbf{M}_\mu p, q\rangle = 0$
2. $\|\mathbf{M}_\mu\|_r = 1$ *for* $1 \leqslant r \leqslant \infty$.
3. $\mathbf{M}_\mu$ *commutes with every permutation operator* $\pi \in \Pi_s$.
4. $\mathbf{M}_\mu^L$ *is fully positive. Consequently,* e *is, up to scalar multiples, the only eigenvector of* $\mathbf{M}_\mu$ *to eigenvalue* 1. *Thus,* $\mathbf{P}_e$ *is the spectral projection of* $\mathbf{M}_\mu$ *to eigenvalue* 1.
5. $sp(\mathbf{M}_\mu) = (1 - L\mu) + \mu \cdot \{-L, -L+2, -L+4, \ldots, L-2, L\}$.
   Consequently, $\mathbf{M}_\mu$ *is invertible if* $\mu < \frac{1}{2L}$.
6. *If* $\mu \leqslant \frac{1}{2L}$, *then* $\mathbf{M}_\mu$ *is* C^**-positive.*
7. $\mathbf{M}_\mu D \subseteq D$. *Consequently,* $\mathbf{M}_\mu \mathbf{P}_D = \mathbf{P}_D \mathbf{M}_\mu \mathbf{P}_D = (\mathbf{M}_\mu \mathbf{P}_D)^* = \mathbf{P}_D \mathbf{M}_\mu$.
8. *If* $\xi = mean(p)$ *for some* $p \in P$, *then* $\mathbf{P}_\xi \mathbf{M}_\mu \mathbf{P}_\xi = (1 - L\mu)\mathbf{P}_\xi$.

Proof. The coefficients of $\mathbf{M}_\mu$ are immediate from the definition of one-bit mutation. Since every bit in the population can be chosen for mutation with equal probability, (3) holds. If $\mathbf{M}_\mu$ has been applied L times, then every bit in any given population may have changed. This shows (4). Let $\mathbf{m}_L \in \mathbb{M}_{2^L}$ be defined through the following identity:

$$\mathbf{M}_\mu = (1 - L\mu) \cdot \mathbb{1}_{2^L} + \mu \cdot \mathbf{m}_L \tag{1}$$

We may prove (2) and (5) by induction over $L \in \mathbb{N}$. For $L = 1$, we have $\mathbf{m}_1 = \mathbf{f}$. Furthermore, it is easy to see that for the canonical order of the basis P:

$$\mathbf{m}_{L+1} = \begin{pmatrix} \mathbf{m}_L & \mathbb{1}_{2^L} \\ \mathbb{1}_{2^L} & \mathbf{m}_L \end{pmatrix} = \mathbf{m}_L \otimes \mathbb{1}_2 + \mathbb{1}_{2^L} \otimes \mathbf{f}. \tag{2}$$

Observe that $\|\mathbf{m}_1\|_r = 1$ for $1 \leqslant r \leqslant \infty$. Thus, $\|\mathbf{m}_L\|_r \leqslant L$. Hence $\|\mathbf{M}_\mu\|_r \leqslant 1$ for $1 \leqslant r \leqslant \infty$. Now item (2) is clear. (5) follows the facts about commuting matrices mentioned in the introduction and Eq. (1). (6) follows from (5) and Corollary B.3 in Appendix B.1.

If $\xi = mean(p)$ for $p \in P$, and $q \in P$ satisfies $\Delta(p,q) = 1$, then $mean(p) \neq mean(q)$. Thus $\mathbf{P}_\xi q = 0$. Hence $\mathbf{P}_\xi \mathbf{M}_\mu p = (1 - L\mu)p$. This shows (8). If $\zeta = mean(q)$, we may assume that $\zeta_1 > \xi_1$ for first components of the *mean*s (and still $\Delta(p,q) = 1$). Then every $q' \in P$ such that $\zeta = mean(q')$ can be produced by $\mathbf{M}_\mu$ from exactly $s\zeta_1$

populations $p' \in P$ with $\xi = mean(p')$. Hence (since p' must be identical with q' except that p' has a '0' at exactly one position where q' has a '1'),

$$\mathbf{P}_\zeta \mathbf{M}_\mu e_{V_\xi} = s\zeta_1 \cdot \sum_{q' \in P,\, mean(q')=\zeta} \mu q' = s\zeta_1 \mu\, e_{V_\zeta}.$$

Varying ζ yields (7). □

Proposition 2. *Let $\mathbf{M}_\mu$ be the doubly stochastic matrix describing one-bit mutation.*

1. *$\mathbf{M}_\mu$ is a contracting map on both $e^\perp$ and S in the Euclidean norm with fixed points 0 resp. $(1/2^L)e$.*[4] *The contracting factor is given by* $\max\{2L\mu - 1, 1 - 2\mu\}$. *In particular, the smallest possible value of the contracting factor is* $1-[2/(L+1)]$ *and is obtained for* $\mu = 1/(L+1)$.
2. *If $\mu \leq 1/(2L)$ and $v \perp e$, then* $(1 - 2L\mu)\|v\|_2 \leq \|\mathbf{M}_\mu v\|_2 \leq (1 - 2\mu)\|v\|_2$.
3. *The coefficients of ergodicity* $\tau_1(\mathbf{M}_\mu) = \tau_\infty(\mathbf{M}_\mu) = 1$.
4. *If v is a probability distribution, then*

$$L\mu + (1 - L\mu - \mu)\|(\mathbb{1} - \mathbf{P}_U)v\|_1 \leq \|(\mathbb{1} - \mathbf{P}_U)\mathbf{M}_\mu v\|_1 \leq L\mu + (1 - L\mu)\|(\mathbb{1} - \mathbf{P}_U)v\|_1.$$

Proof. By Proposition 1.4, $span_{\mathbb{C}}(e)$ is the eigenspace to eigenvalue 1 of $\mathbf{M}_\mu$. It follows from the spectral theorem (Corollary B.2 in Appendix B.1) and Proposition 1.5 that $\mathbf{M}_\mu$ acts as a contracting map on $e^\perp$ with the contracting factor claimed. Also, (2) follows from this. All $v \in S$ are of the form $v = (1/2^L)e + w$ where w is perpendicular to e. This implies that $\mathbf{M}_\mu$ acts as a contracting map on S with fixed point $(1/2^L)e$.

Let p_0 and p_1 be the populations with all bits zero and all bits one, respectively. There is no population that can reach both of these by a single-bit mutation. Therefore $\min\{(\mathbf{M}_\mu)_{q,p_0}, (\mathbf{M}_\mu)_{q,p_1}\}$ is zero for all $q \in P$. Thus, $1 - \tau_1(\mathbf{M}_\mu) = \min_{p,p' \in P}\{\sum_{q \in P} \min\{(\mathbf{M}_\mu)_{q,p}, (\mathbf{M}_\mu)_{q,p'}\}\} = 0$.

For the above p_0, define $v \in V$ by $\langle v, p_0\rangle = 1$, $\langle v, p\rangle = 1$ if $\Delta(p, p_0) = 1$, and $\langle v, q\rangle = -1$ for exactly $L+1$ populations q with $\Delta(q, p_0) > 1$. Now $v \perp e$, $\|v\|_\infty = 1$, and $\mathbf{M}_\mu v$ has p_0th coordinate equal to $1 - L\mu + L\mu = 1$. Hence, $\tau_\infty(\mathbf{M}_\mu) \geq 1$. We have $\|\mathbf{M}_\mu\|_\infty \leq 1$ by the discussion in the notation section of this paper or by [29, p. 5]. Thus, (3) is established.

If p is uniform, then it is mapped with probability $L\mu$ to a non-uniform population. If p is non-uniform, then changing any bit might keep it non-uniform, and it makes it uniform with probability at most μ. Thus,

$$\begin{aligned} L\mu\|\mathbf{P}_U v\|_1 + (1 - \mu)\|(\mathbb{1} - \mathbf{P}_U)v\|_1 &\leq \|(\mathbb{1} - \mathbf{P}_U)\mathbf{M}_\mu v\|_1 \\ &\leq L\mu\|\mathbf{P}_U v\|_1 + \|(\mathbb{1} - \mathbf{P}_U)v\|_1. \end{aligned}$$ □

Multiple-bit mutation. Let $\mu \in (0, \frac{1}{2}]$. In multiple-bit mutation, a lottery is played independently at each bit in a population $p \in P$ to decide whether to change it or not. The probability for each change is μ.

[4] In fact, this is a fixed point of every doubly stochastic matrix.

Multiple-bit mutation is also easy to implement and has a simple mathematical description. Obviously, one-bit mutation is a first order approximation of multiple-bit mutation if μ is small. Multiple-bit mutation is the standard operator used in implementations of simple genetic algorithms, while one-bit mutations fits better with the philosophy of "small neighborhoods" used in simulated annealing (cf. [1, 5, 6, 20, 21]).

Proposition 3. *Multiple-bit mutation can be described as a C^*-positive, stochastic matrix $\mathbf{M}_\mu$ acting on V.*

1. *Let $p,q \in P$. The coefficients of $\mathbf{M}_\mu$ are determined as follows:*

$$\langle \mathbf{M}_\mu p, q\rangle = \mu^{\Delta(p,q)}(1-\mu)^{L-\Delta(p,q)} > 0.$$

   Thus e is, up to scalar multiples, the only eigenvector of $\mathbf{M}_\mu$ to eigenvalue 1. *Consequently, $\mathbf{P}_e$ is the spectral projection of $\mathbf{M}_\mu$ to eigenvalue* 1.
2. *$\mathbf{M}_\mu$ commutes with every $\pi \in \Pi_s$.*
3. *$\|\mathbf{M}_\mu\|_r = 1$ for $1 \leqslant r \leqslant \infty$.*
4. *$sp(\mathbf{M}_\mu) = \{(1-2\mu)^k : 0 \leqslant k \leqslant L\}$. Consequently, if $\mu \neq \frac{1}{2}$ then $\mathbf{M}_\mu$ is invertible.*
5. *$\mathbf{M}_\mu D \subseteq D$. Consequently, $\mathbf{M}_\mu \mathbf{P}_D = \mathbf{P}_D \mathbf{M}_\mu \mathbf{P}_D = (\mathbf{M}_\mu \mathbf{P}_D)^* = \mathbf{P}_D \mathbf{M}_\mu$.*
6. *If $\xi = mean(p)$ for some $p \in P$, then $\mathbf{P}_\xi \mathbf{M}_\mu e_{V_\xi} = t_\xi e_{V_\xi}$ for*

$$t_\xi = \sum_{q \in V_\xi} \mu^{\Delta(p,q)}(1-\mu)^{L-\Delta(p,q)}.$$

Proof. The coefficients of $\mathbf{M}_\mu$ and (2) are immediate from the definition of multiple-bit mutation. Depending upon $L \in \mathbb{N}$, denote $\mathbf{M}_\mu$ as $\mathbf{M}_L$ for the moment. We have the following identities for the canonical order of populations $p \in P$:

$$\mathbf{M}_1 = \begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix},$$

$$\mathbf{M}_{L+1} = \begin{pmatrix} (1-\mu)\mathbf{M}_L & \mu\mathbf{M}_L \\ \mu\mathbf{M}_L & (1-\mu)\mathbf{M}_L \end{pmatrix} = \mathbf{M}_L \otimes \mathbf{M}_1 = (\mathbf{M}_L \otimes \mathbb{1}_2)(\mathbb{1}_{2^L} \otimes \mathbf{M}_1).$$

It is easy to check that $\|\mathbf{M}_1\|_r \leqslant 1$ for $1 \leqslant r \leqslant \infty$. Thus $\|\mathbf{M}_L\|_r \leqslant 1$. The above expression for $\mathbf{M}_{L+1}$ is a product of commuting matrices, hence

$$sp(\mathbf{M}_{L+1}) \subseteq sp(\mathbf{M}_L) \cdot \{1, 1-2\mu\}.$$

Now, (4) follows as before. Let $\mathbf{m}_L$ be as in the proof of Proposition 1. By Proposition 1.7, we have $\mathbf{m}_L D \subseteq D$. The matrix $\frac{1}{L}\mathbf{m}_L$ describes changing exactly one bit in a population with equal probability. Thus, the matrix describing changing exactly two bits is a linear combination of $\mathbb{1}$ and $\mathbf{m}_L^2$. Continuing this argument by induction yields (5). If $\xi = mean(p)$ for some $p \in P$, then

$$\langle \mathbf{P}_\xi \mathbf{M}_\mu e_{V_\xi}, p\rangle = \langle e_{V_\xi}, \mathbf{M}_\mu p\rangle = t_\xi \langle e_{V_\xi}, p\rangle,$$

with t_ξ as above. □

Note that the value $\mu = \frac{1}{2}$ yields $\mathbf{M}_{\frac{1}{2}} = \mathbf{P}_e$. Applied once, this corresponds to a randomly determined restart of the algorithm, while fixing $\mu = \frac{1}{2}$ leads to uniform random search.

Proposition 4. *Let $\mathbf{M}_\mu$ denote the doubly stochastic matrix describing multiple-bit mutation.*

1. *$\mathbf{M}_\mu$ is a contracting map on both $e^\perp$ and S in the Euclidean norm with contracting factor $1-2\mu$. The probability distribution $(1/2^L)e$ is the fixed point of $\mathbf{M}_\mu$ in S.*
2. *If $v \perp e$ then $(1-2\mu)^L\|v\|_2 \leqslant \|\mathbf{M}_\mu v\|_2 \leqslant (1-2\mu)\|v\|_2$.*
3. *The coefficients of ergodicity satisfy $\tau_1(\mathbf{M}_\mu), \tau_\infty(\mathbf{M}_\mu) \leqslant 1-(2\mu)^L$. Consequently, $\mathbf{M}_\mu$ is a contracting map both on $e^\perp$ and S in the ∞-norm and the Hamming norm with contracting factor bounded above by $1-(2\mu)^L$.*
4. *If $v \in S$, $\gamma = (1-(\mu^s+(1-\mu)^s)^\ell)\|\mathbf{P}_U v\|_1$ and $h = \lfloor s/2 \rfloor$, then we have the following estimates:*

$$\gamma + (1-\mu(1-\mu)(\mu^{s-2}+(1-\mu)^{s-2})(\mu^s+(1-\mu)^s)^\ell)\|(\mathbb{1}-\mathbf{P}_U)v\|_1$$
$$\leqslant \|(\mathbb{1}-\mathbf{P}_U)\mathbf{M}_\mu v\|_1 \leqslant \gamma + (1-((1-\mu)^h\mu^{s-h}+\mu^h(1-\mu)^{s-h})^\ell)\|(\mathbb{1}-\mathbf{P}_U)v\|_1.$$

Proof. The first two statements are obtained in a similar fashion as the corresponding statements of Proposition 1.

The inequality $\tau_1(\mathbf{M}_\mu) \leqslant 1-(2\mu)^L$ follows from

$$1-\tau_1(\mathbf{M}_\mu) \geqslant \sum_{p\in P} \min_{q\in P}\{(\mathbf{M}_\mu)_{p,q}\} = \sum_{p\in P} \mu^L = 2^L\mu^L.$$

Let $v \perp e$, that is, $\sum_{p\in P} v_p = 0$. Let v be a vector with $\|v\|_\infty = 1$ with max $\|\mathbf{M}_\mu v\|_\infty$ attained. We may assume $\mathbf{M}_\mu v$ has a maximum modulus qth component which is strictly positive:

$$\begin{aligned}
\tau_\infty(\mathbf{M}_\mu) = (\mathbf{M}_\mu v)_q &= \sum_{p\in P} (\mathbf{M}_\mu)_{q,p}\, v_p \\
&= \sum_{p\in P} ((\mathbf{M}_\mu)_{q,p} - \mu^L)\, v_p \quad \text{since } \sum_{p\in P} v_p = 0 \\
&\leqslant \sum_{p\in P} ((\mathbf{M}_\mu)_{q,p} - \mu^L) \quad \text{since } (\mathbf{M}_\mu)_{q,p} \geqslant \mu^L \text{ and } |v_p| \leqslant 1 \\
&= 1 - 2^L\mu^L \quad \text{since } \mathbf{M}_\mu \text{ is row stochastic.}
\end{aligned}$$

[We remark that if $\mathbf{X} \in \mathbb{M}_n$ is a doubly stochastic matrix a similar argument shows $\tau_\infty(\mathbf{X}) \leqslant 1-n\gamma$, where γ is the smallest value among the entries of $\mathbf{X}$.]

Let $p \in P$ be uniform. In order to produce a uniform population from p one selects k bits to be changed in the first creature of p and then has to change $s \cdot k$ bits in p. Thus, the probability of producing a uniform population from p is given by

$$\sum_{k=0}^{\ell} \binom{\ell}{k} \mu^{sk}(1-\mu)^{s(\ell-k)} = (\mu^s+(1-\mu)^s)^\ell.$$

Let $p \in P$ be non-uniform. Suppose that exactly k of the s creatures in p have 0 as their first bit. One can then either change k bits to 1 or change $s-k$ bits 0 in order to make the first bit in every creature agree. Consider the function

$$\rho : k \mapsto \mu^k(1-\mu)^{s-k} + (1-\mu)^k\mu^{s-k}.$$

By symmetry, ρ is minimal at $h = \lfloor s/2 \rfloor$ and maximal at $k = 0, s$. Thus the probability of generating a uniform population from p is greater than

$$(\mu^h(1-\mu)^{s-h} + (1-\mu)^h\mu^{s-h})^\ell.$$

Let $p \in P$ be such that $\Delta(p,q) = 1$ for some uniform $q \in P$. In this situation, the probability of producing a uniform population (not necessarily q) from p is given by

$$(\mu^{s-1}(1-\mu) + \mu(1-\mu)^{s-1})(\mu^s + (1-\mu)^s)^{\ell-1}.$$

The discussion of ρ shows that this product consists of the greatest possible factors. □

Both mutation operators contract every probability distribution towards $(1/2^L)e$. However, the contractions stays controlled in the sense that for sufficiently small μ the length of a vector perpendicular to e cannot shrink too much (see Propositions 2.2 and 4.2).

2.2. Crossover

Elementary crossover and simple crossover operations. Let $p = (c_1, \ldots, c_s) \in P$, $c_i \in C = \{0,1\}^\ell$. Let $1 \leqslant i < j \leqslant n$ be indices of two creatures

$$c_i = (a_1, \ldots, a_\ell), \qquad c_j = (b_1, \ldots, b_\ell),$$

and $1 \leqslant k \leqslant \ell$ be a *crossover point*. Then, an *elementary crossover operation* $\mathbf{C}(i,j;k)$ on p consists of replacing c_i and c_j in p by offspring

$$c_i' = (b_1, \ldots, b_k, a_{k+1}, \ldots, a_\ell), \qquad c_j' = (a_1, \ldots, a_k, b_{k+1}, \ldots, b_\ell),$$

respectively. No other change occurs. We also write $\mathbf{C}(i,j;k)$ for the associated stochastic matrix. An elementary crossover exchanges the heads of two individuals ("parents") in the population to obtain new individuals ("offspring"). If $k = \ell$, this amounts to exchanging the parents.[5]

The potential to model spatial effects on recombination presents itself in the possibility of constructing various crossover operations as stochastic combinations of elementary crossover components. One may easily modify the definition of the crossover operators given below so that crossover occurs between parents at different pairs of

[5] The case $k = \ell$ has been included for mathematical convenience, although such a crossover point can never introduce a new type of individual. Alternatively, this case could be disallowed and our subsequent analysis repeated with slightly modified but less natural replacements for the Hardy–Weinberg spaces, keeping track of the last bits of individuals in a population (since then crossover could never result in a switch of parents' last bits).

locations with non-constant likelihood varying in an arbitrary manner according to their locations.

Let s be even. Let $K = (k_1, \ldots, k_{s/2})$ be a vector of crossover points, and $\chi \in [0,1]$. Then a *simple crossover operation* $\mathbf{C}(K,\chi)$ is given by

$$\mathbf{C}(K,\chi) = \prod_{i=1}^{s/2} ((1-\chi)\mathbb{1} + \chi \mathbf{C}(2i-1, 2i, k_i)).$$

χ will be called the *crossover rate*.

Observe that for $\chi < 1$, $\mathbf{C}(K,\chi)$ is of the form

$$\mathbf{C}(K,\chi) = \rho\mathbb{1} + (1-\rho)\cdot \mathbf{S},$$

where $\mathbf{S}$ is a stochastic matrix, $\rho = (1-\chi)^{s/2}$. Thus, its spectrum is contained in $\rho + (1-\rho)D$, where D is the closed unit disk in $\mathbb{C}$.

Lemma 5. 1. *An elementary crossover operation determines a self-adjoint, unitary matrix* $\mathbf{C}(i,j;k)$ *acting on* V *which, up to rearrangement of the basis of* V, *equals some* $\mathbf{f} \oplus \ldots \oplus \mathbf{f} \oplus \mathbb{1}$. *The previous statement holds also for a simple crossover operation* $\mathbf{C}(K,1)$ *with crossover rate* $\chi = 1$.

2. $\mathbf{C}(i,j;k)$ *and* $\mathbf{C}(K,\chi)$ *are isometries with respect to the r-norms for* $1 \leqslant r \leqslant \infty$. *In particular,* $\|\mathbf{C}(K,\chi)\|_r = 1$ *for* $1 \leqslant r \leqslant \infty$.

3. *All* $\mathbf{C}(2i-1,2i;k)$ *and thus all* $\mathbf{C}(K,\chi)$ *commute. Consequently,*

$$sp(\mathbf{C}(K,\chi)) \subseteq \{(1-2\chi)^n \mid 0 \leqslant n \leqslant \tfrac{1}{2}s\}.$$

4. *If* $\chi \neq \frac{1}{2}$, *then* $\mathbf{C}(K,\chi)$ *is invertible. If* $\chi \leqslant \frac{1}{2}$, *then* $\mathbf{C}(K,\chi)$ *is* C^**-positive.*

5. *If* $p \in P$, *then we have* $\langle \mathbf{C}(K,\chi)p, p\rangle \geqslant (1-\chi)^{\frac{s}{2}}$.

6. *Let* $p \in P$. $\mathbf{C}(i,j;k)$ *and* $\mathbf{C}(K,\chi)$ *keep mean*(p) *invariant. Thus* $\mathbf{C}(i,j;k)$ *and* $\mathbf{C}(K,\chi)$ *decompose into block diagonal matrices with one block for each Hardy-Weinberg space* V_ξ, $\xi = mean(p)$. *Also* $\mathbf{C}(i,j;k)$ *and* $\mathbf{C}(K,\chi)$ *act as the identity on* U, *the subspace generate by uniform populations.*

7. *Consequently, we have:* $\mathbf{C}(i,j;k)\mathbf{P}_U = \mathbf{P}_U = \mathbf{P}_U\mathbf{C}(i,j;k)$.

8. *Let* $p \in P$. $\mathbf{C}(K,\chi)$ *will keep Mean*(p) *invariant. Thus* $\mathbf{C}(K,\chi)$ *decomposes into block diagonal matrices with one block for each* $\overline{V}_\zeta$, $\zeta = Mean(p)$ *for some population* p.

9. *If* $\mathbf{M}_\mu$ *denotes either one- or multiple-bit mutation, then* $\mathbf{M}_\mu$ *commutes with every* $\mathbf{C}(i,j;k)$ *and thus with every* $\mathbf{C}(K,\chi)$.

Proof. (1), (5), (6) and (8) are obvious. (2) follows from (1). (3) follows from the facts about commuting matrices mentioned in the introduction. (4) follows from (3) and Corollary B.3 in Appendix B.1. (7) follows from (6) and the fact that $\mathbf{P}_U$ is self-adjoint. Finally, we show (9): Likelihoods of all results are the same whether one first (logically) negates bits in a population at random and then to exchanges the positions or first exchanges the positions and then negates at random. □

Lemma 6. *Let p,q be populations in P.*

1. *If $mean(p) = mean(q)$ than at most $\ell \cdot \lfloor s/2 \rfloor$ elementary crossover operations are needed to transform p into q.*
2. *If s is even and $Mean(p) = Mean(q)$, then at most $\ell \cdot s/2$ elementary crossover operations $\mathbf{C}(2i-1,2i;k)$ are needed to transform p to q.*
3. *If p is rich (see Section 1.3) and $c \in C$, then using at most $\ell - 1$ elementary crossover operations $\mathbf{C}(i,j;k)$ successively applied (starting with p) a population containing c is obtained.*

Proof. For a moment, consider $\ell = 1$. If $mean(p) = mean(q)$, then the number of indices i, $1 \leqslant i \leqslant s$ at which p and q differ must be even. A suitably chosen elementary crossover operation will correct two of these positions ($1 \leqslant i,j \leqslant s$) simultaneously in the process of generating p from q. Thus, at most $\lfloor s \rfloor/2$ such elementary crossover operations are needed. In order to transform populations with creatures of length $\ell > 1$ into each other, one can apply the above from last to first bit. (2) and (3) are now immediate. □

The next construct will be used in proving convergence of the probability distribution for genetic algorithms. It faithfully represents the standard implementation of crossover in genetic algorithms: pairing-off adjacent pairs in an (even size) population, for each pair, crossover occurs with probability χ at a randomly selected crossover point.

Simple crossover. Let $\chi \in (0,1)$ be fixed. *Simple crossover* is the application of a simple crossover operation $\mathbf{C}(K,\chi)$ where the crossover vector K is determined randomly such that all crossover points have equal probability. Consequently each K has probability $\ell^{-\frac{s}{2}}$, if crossover occurs.[6]

Proposition 7. *Simple crossover can be described by a self-adjoint, stochastic matrix $\mathbf{C}_\chi$, $\chi \in (0,1)$, acting on V.*

1. *If $p \in P$, then we have $\langle \mathbf{C}_\chi p, p\rangle \geqslant (1-\chi)^{s/2}$.*
2. *$\|\mathbf{C}_\chi\|_r = 1$ for all $1 \leqslant r \leqslant \infty$.*
3. *$\tau_r(\mathbf{C}_\chi) = 1$ for all $1 \leqslant r \leqslant \infty$.*
4. *$\mathbf{C}_\chi$ commutes with $\mathbf{M}_\mu$ (single- or multiple-bit mutation). In particular, $\mathbf{M}_\mu \mathbf{C}_\chi^n$, $n \in \mathbb{N}$, is self-adjoint. If n is even, then $\mathbf{M}_\mu \mathbf{C}_\chi^n$ is C^*-positive (and $\mu L \leqslant \frac{1}{2}$, in the case of one-bit mutation).*
5. *$\mathbf{C}_\chi$ decomposes into a block diagonal matrix with one block $\mathbf{C}_{\chi,\zeta}$ for each $\overline{V}_\zeta$, $\zeta = Mean(p)$ for some $p \in P$.*

[6] Counterexample: For $\chi = 1$, consider the creatures $c_0 = 0\ldots0$ and $c_1 = 10\ldots0$, which differ only in the first bit. Let $p = c_1 c_0 \ldots c_0$ and $p' = c_0 c_1 c_0 \ldots c_0$. It is easy to see that $p - p'$ is an eigenvector of $\mathbf{C}_\chi$, $\chi = 1$, for eigenvalue $\lambda = -1$ since $\mathbf{C}(1,2;k_1)$ must always interchange these two populations (since $k_1 \geqslant 1$ and the other $\mathbf{C}(2i-1,2i;k_i)$ have no effect on p,p'). But as we shall see for $\chi < 1$, we never have -1 as eigenvalue.

6. *For each of the blocks* $\mathbf{C}_{\chi,\zeta}$ *of* $\mathbf{C}_\chi$ *associated with* $\overline{V}_\zeta$, $C_{\chi,\zeta}^{L/2}$ *is fully positive. Consequently,* $e_{\overline{V}_\zeta}$ *is, up to scalar multiples, the only eigenvector to eigenvalue* 1 *for* $\mathbf{C}_\chi$ *in* $\overline{V}_\zeta$. *Furthermore,* $-1 \notin sp(\mathbf{C}_{\chi,\zeta})$. *In addition,* $\mathbf{C}_\chi \mathbf{P}_D = \mathbf{P}_D \mathbf{C}_\chi \mathbf{P}_D = \mathbf{P}_D \mathbf{C}_\chi$.
7. *The spectrum of* $\mathbf{C}_\chi$ *is contained in* $[-1+\beta, 1-\beta] \cup \{1\}$, *where* $\beta = \ell^{-s/2}(1 - |1-2\chi|)$.
8. *Let* $\zeta = Mean(p)$ *for some* $p \in P$, *and* $v \in \overline{V}_\zeta$ *such that* $v \perp e$. *In this situation,* $\|\mathbf{C}_\chi v\|_2 \leqslant (1-\beta)\|v\|_2$.
9. $\mathbf{C}_\chi$ *acts as the identity map on* U. *Thus,* $\langle \mathbf{C}_\chi p, p\rangle = 1$ *if* p *is uniform.*
10. *If* $v \in S$, *then* $\|(\mathbb{1} - \mathbf{P}_U)\mathbf{C}_\chi(\mathbb{1} - \mathbf{P}_U)v\|_1 = \|\mathbf{C}_\chi(\mathbb{1} - \mathbf{P}_U)v\|_1 = \|(\mathbb{1} - \mathbf{P}_U)v\|_1$.

Proof. (1), (2), (4) and (5) follow from the corresponding statements of Lemma 5 and the fact that $\mathbf{C}_\chi$ is a convex combination of the $\mathbf{C}(K,\chi)$. $\mathbf{C}_\chi$ has strictly positive diagonal since the $\mathbf{C}(K,\chi)$ do. Next we prove (6): Lemma 6.2 shows that $C_{\chi,\zeta}^{L/2}$ is fully positive. Now consult the discussion in the notation section of this paper on stochastic matrices. The discussion preceding Lemma 5 implies that -1 is not in the spectrum of $\mathbf{C}_{\chi,\zeta}$. (7) follows from Lemma 5.3, (6) and the facts on the spectrum of commuting matrices listed in the notation section of this paper (the spectrum of a convex combination of commuting matrices is the same combination of their spectra). (8) follows from (7) and Corollary B.2 in Appendix B.1. (9) is obvious. Applying (9) to scalar multiples of $p_0 - p_1$ (the difference of the basis vectors associated with the all-zero population and the all-ones population) implies (3). As for the last statement, we have by Lemma 5.7,

$$(\mathbb{1} - \mathbf{P}_U)\mathbf{C}(K,\chi)(\mathbb{1} - \mathbf{P}_U) = \mathbf{C}(K,\chi)(\mathbb{1} - \mathbf{P}_U)$$

since $\mathbb{1} - \mathbf{P}_U$ is a projection. Thus,

$$\|(\mathbb{1} - \mathbf{P}_U)\mathbf{C}_\chi(\mathbb{1} - \mathbf{P}_U)v\|_1 = \|\mathbf{C}_\chi(\mathbb{1} - \mathbf{P}_U)v\|_1 = \|(\mathbb{1} - \mathbf{P}_U)v\|_1$$

since $\mathbf{C}_\chi$ is stochastic. □

Let $p = (c_1, \ldots, c_s), q = (d_1, \ldots, d_s) \in P$, $c_i, d_i \in C$, $1 \leqslant i < j \leqslant s$. Define the *crossover multiplicity* for populations $p, q \in P$ by

$$m(p,q) = cardinality(\{(i,j;k) : \mathbf{C}(i,j;k)p = q\}).$$

Let c_i and c_j, $1 \leqslant i < j \leqslant s$ be such that

$$c_i = ha \quad c_j = hb,$$

where a, b and h are (possibly empty) bit strings. Assume that h is chosen with maximal length. Let the *crossover multiplicity* for creatures $m_c(c_i, c_j) \in [0, \ell]$ be defined as

$$m_c(c_i, c_j) = length(h).$$

We have:

Case 1: If $p = q$, then

$$m(p, p) = \sum_{j=2}^{s} \sum_{i=1}^{j-1} m_c(c_i, c_j).$$

It is easy to see that $m(p, p) \geqslant \frac{s(s-2)}{4}$.

Case 2: If $p \neq q$ and, for r such that $1 \leqslant r \leqslant s$ with the exception of exactly two indices i and j, we have $c_r = d_r$. In addition,

$$c_i = uhv \quad c_j = u'hv',$$

where $u, u', h, v, v' v'$ are bit strings, $length(u) = length(u')$, h is of maximal length, and

$$d_i = u'hv \quad d_j = uhv'.$$

In this situation,

$$m(p, q) = length(h) + 1.$$

Case 3: In any other situation, $m(p, q) = 0$.

Unrestricted crossover Unrestricted crossover is the application of an elementary crossover over $\mathbf{C}(i, j; k)$ under the following rules: With probability $\chi \in (0, 1]$ a single crossover takes place in the population. If crossover occurs, then there is equal probability for all triples $(i, j; k)$.

If unrestricted crossover is considered, then we shall suppose $s \geqslant 3$. Otherwise, $s = 2$, and unrestricted and simple crossover coincide.

Next, we shall explore a connection of elementary crossover with representations of the group of permutations Π_s of $\{1, \ldots, s\}$. The set $\mathbf{T}_s$ of all transpositions in Π_s has $\frac{1}{2}s(s-1)$ elements. Π_s acts[7] canonically as linear operators on the free vector space W over Π_s: Namely, we have for $\pi, \sigma \in \Pi_s$,

$$\pi(\sigma) = \pi \cdot \sigma,$$

where $\pi(\sigma)$ denotes the result of action by the linear operator $\pi(\cdot)$ on the basis element $\sigma \in W$. Identifying Π_s with the linear operators on W just defined, we let

$$\Gamma_s = \frac{2}{s(s-1)} \sum_{\tau \in \mathbf{T}_s} \tau$$

in this representation. Γ_s is a self-adjoint stochastic matrix. Furthermore, $sp(\Gamma_s) = -sp(\Gamma_s)$. To see this, replace any eigenvector $\sum a_\pi \pi$ by $\sum sgn(\pi) a_\pi \pi$ and check the eigenvalue equation coordinatewise at $\pi \in \Pi_s$. For $s \in \mathbb{N}$, let $\alpha(s) < 1$ denote the second largest eigenvalue of Γ_s. Numerical computations suggest a formula for $\alpha(s)$ given in Appendix A.

[7] This is called the left regular representation; see, e.g., [24, Ch. 7].

Let $1 \leqslant k \leqslant \ell$ be fixed. The operation of crossover induces a group representation ρ_k of Π_s such that the transpositions $(i\,j)$ are mapped to elementary crossover operations $\mathbf{C}(i,j;k)$. Let

$$\mathbf{C}(k) = \frac{2}{s(s-1)} \sum_{j=2}^{s} \sum_{i=1}^{j-1} \mathbf{C}(i,j;k),$$

$$\mathbf{C}_\chi = (1-\chi)\mathbb{1} + \frac{\chi}{\ell} \sum_{k=1}^{\ell} \mathbf{C}(k).$$

Lemma 8. *$sp(\mathbf{C}(k)) \subseteq \{-1,1\} \cup [-\alpha(s), \alpha(s)]$, where $\alpha(s) < 1$ is as defined above.*

Proof. The group representation ρ_k from Π_s to linear operators on V extends canonically to a representation $\hat{\rho_k}$ of the algebra of linear operators on W generated by Π_s (i.e., the group C^*-algebra of Π_s; see, e.g., [24, Ch. 7]). Since invertible elements in the group C^*-algebra stay invertible under $\hat{\rho_k}$, eigenvalues can only disappear. □

Proposition 9. *Unrestricted crossover can be described by the self-adjoint, stochastic matrix $\mathbf{C}_\chi$, $\chi \in (0,1]$, defined above, acting on V.*

1. *If $p,q \in P$ with $p \neq q$, then we have*

$$\langle \mathbf{C}_\chi p, p\rangle = (1-\chi) + \frac{2\chi}{\ell s(s-1)} m(p,p) \geqslant (1-\chi) + \frac{\chi}{2\ell} \frac{s-2}{s-1} > 0$$

$$\langle \mathbf{C}_\chi p, q\rangle = \frac{2\chi}{\ell s(s-1)} m(p,q).$$

2. $\|\mathbf{C}_\chi\|_r = 1$ *for* $1 \leqslant r \leqslant \infty$.
3. $\tau_r(\mathbf{C}_\chi) = 1$ *for* $1 \leqslant r \leqslant \infty$.
4. $\mathbf{C}_\chi$ *commutes with every permutation operator* $\pi \in \Pi_s$ *and with* $\mathbf{M}_\mu$, *where* $\mathbf{M}_\mu$ *represents either one- or multiple-bit mutation. In particular,* $\mathbf{M}_\mu \mathbf{C}_\chi^n$, $n \in \mathbb{N}$, *is self-adjoint. If n is even (and* $\mu L \leqslant \frac{1}{2}$, *in the case of one-bit mutation), then* $\mathbf{M}_\mu \mathbf{C}_\chi^n$ *is C^*-positive.*
5. $\mathbf{C}_\chi$ *decomposes into a block diagonal matrix with one block* $\mathbf{C}_{\chi,\xi}$ *for each of the* V_ξ, $\xi = mean(p)$ *for some* $p \in P$.
6. *For each of the blocks* $\mathbf{C}_{\chi,\xi}$ *of* $\mathbf{C}_\chi$ *associated to* V_ξ, *the matrix* $\mathbf{C}_{\chi,\xi}^{\ell \cdot \lfloor \frac{s}{2} \rfloor}$ *is fully positive. Consequently,* 1 *is the only eigenvalue of modulus* 1 *of* $\mathbf{C}_\chi$. *In addition,* e_{V_ξ} *is, up to scalar multiples, the only eigenvector to eigenvalue* 1 *for* $\mathbf{C}_\chi$ *in each* V_ξ. *Hence,* $\mathbf{C}_\chi \mathbf{P}_D = \mathbf{P}_D \mathbf{C}_\chi \mathbf{P}_D = \mathbf{P}_D \mathbf{C}_\chi$.
7. *The spectrum of* $\mathbf{C}_\chi$ *is contained in*

$$\{1\} \cup \left[1 - 2\chi + \frac{\chi}{\ell}(1-\alpha(s)), 1 - \frac{\chi}{\ell}(1-\alpha(s))\right].$$

8. *Let* $\xi = mean(p)$ *for some* $p \in P$, *and let* $v \in V_\xi$ *such that* $v \perp e$. *Then,* $\|\mathbf{C}v\|_2 \leqslant (1 - \frac{\chi}{\ell}(1-\alpha(s)))\|v\|_2$.
9. $\mathbf{C}_\chi$ *acts as the identity map on U. Thus,* $\langle \mathbf{C}_\chi p, p\rangle = 1$ *if p is uniform.*
10. *If* $v \in S$, *then* $\|(\mathbb{1} - \mathbf{P}_U)\mathbf{C}_\chi(\mathbb{1} - \mathbf{P}_U)v\|_1 = \|(\mathbb{1} - \mathbf{P}_U)v\|_1$.

Proof. The probability that a particular $\mathbf{C}(i,j;k)$ is selected is $2/[\ell s(s-1)]$. Now the coefficients of $\mathbf{C}_\chi$ and the estimates in (1) follow from the discussion preceding the definition of unrestricted crossover. For the remainder of the proof, we may suppose, without loss of generality, that $\chi = 1$. $\mathbf{C}_1$ is a convex combination of the $\mathbf{C}(i,j;k)$. This and Lemma 5 imply that $\mathbf{C}_1$ acts as the identity on U, that $\mathbf{C}_1$ commutes with $\mathbf{M}_\mu$, and $\|\mathbf{C}_1\|_r \leqslant 1$. (3) follows as did Proposition 7.3. By symmetry, the fact that $\mathbf{C}_1$ commutes with every $\pi \in \Pi_s$ is immediate.

The block decomposition of $\mathbf{C}_1$ in (5) follows from Lemma 5.6. The fact that all $\mathbf{C}_{\chi,\xi}^{\ell \cdot \lfloor \frac{s}{2} \rfloor}$ have strictly positive coefficients follows from Lemma 6.1. This, (1) – which shows that under our general assumption of $s \geqslant 3$, $\mathbf{C}_\chi$ has strictly positive diagonal – and the discussion of stochastic matrices in the notation section of this paper show (6).

Let k' be such that $1 \leqslant k' < k$. Then,

$$\mathbf{C}(1,2;k)\mathbf{C}(2,3;k') = \mathbf{C}(1,3;k')\mathbf{C}(1,2;k).$$

With this in mind, it is easy to show that all $\mathbf{C}(k)$ commute. The bounds for the spectrum of $\mathbf{C}_\chi$ follow from Lemma 8 and facts mentioned in the introduction on the spectrum of sums of commuting matrices. (8) follows from (7) and the spectral theorem (see Corollary B.2 in Appendix B.1). (10) is obtained as in Proposition 7.10. □

If $v \in S$, let

$$v = \frac{1}{2^L}e + d(v) + o(v),$$

where $d(v) = \mathbf{P}_D(\mathbb{1} - \mathbf{P}_e)v = (\mathbf{P}_D - \mathbf{P}_e)v = (\mathbb{1} - \mathbf{P}_e)\mathbf{P}_D v$. We have $\mathbf{P}_e v = (1/2^L)e$. Thus, $o(v) = (\mathbb{1} - \mathbf{P}_D)v$.

Proposition 10. *Let $\mathbf{C}_\chi$ represent unrestricted crossover and $\mathbf{M}_\mu$ either type of mutation operator. (If one-bit mutation is used, then suppose $L\mu \leqslant \frac{1}{2}$). Let $v \in S$. If $v = (1/2^L)e + d(v) + o(v)$, then one has for $v' = \mathbf{M}_\mu \mathbf{C}_\chi^n v$, $n \in \mathbb{N}$:*

$$d(v') = \mathbf{M}_\mu\, d(v), \text{ and } o(v') = \mathbf{M}_\mu \mathbf{C}_\chi^n\, o(v).$$

In addition,

$$\|d(v')\|_2 \leqslant (1-2\mu)\|d(v)\|_2, \text{ and } \|o(v')\|_2 \leqslant (1-2\mu)\left(1 - \frac{\chi}{\ell}(1-\alpha(s))\right)^n \|o(v)\|_2.$$

Proof. Suppose for the moment that we use one-bit mutation. By Proposition 1.6, $sp(\mathbf{M}_\mu) \subseteq \mathbb{R}^+$. We have $\mathbf{M}_\mu \mathbf{C}_\chi\, d(v) = \mathbf{M}_\mu\, d(v) \in D$ by Proposition 1.7. Proposition 2.1 and $d(v) \perp e$ imply that $\|d(v')\|_2$ is no larger than $(1-2\mu)\|d(v)\|_2$. We have

$$\mathbf{M}_\mu(\mathbb{1} - \mathbf{P}_D) = (\mathbb{1} - \mathbf{P}_D)\mathbf{M}_\mu(\mathbb{1} - \mathbf{P}_D)$$

by Proposition 1.7. This shows that $\mathbf{M}_\mu D^\perp \subseteq D^\perp$. Combining this with Proposition 9.6 yields $\mathbf{M}_\mu \mathbf{C}_\chi^n o(v) \in D^\perp$. Using Propositions 2.2 and 9.8, we get

$$\|\mathbf{M}_\mu \mathbf{C}_\chi^n o(v)\|_2 \leqslant (1-2\mu)\|\mathbf{C}_\chi^n o(v)\|_2 \leqslant (1-2\mu)\left(1-\frac{\chi}{\ell}(1-\alpha(s))\right)^n \|o(v)\|_2.$$

This completes the proof for one-bit mutation. As for multiple-bit mutation, use Propositions 3 and 4. □

A result similar to Proposition 10 holds for simple crossover. Note that iteration of any type of crossover $\mathbf{C}_\chi^n$ includes crossover operations with *multiple crossover points.*

The results of this section yield deeper insights into the usefulness of crossover. Like mutation, crossover $\mathbf{C}_\chi^n$ ($\chi > 0$, $n \in \mathbb{N}$) has an averaging effect, but within the Hardy–Weinberg subspaces V_ξ (resp. $\overline{V}_\xi$) rather than on the entire state space. This averaging spreads out the search within each V_ξ. Unrestricted crossover averages – i.e. contracts to uniform distributions – on larger subspaces than simple crossover (V_ξ vs. $\overline{V}_\xi$). Mutation–crossover can be seen as an enhanced mutation which tends to preserve schemata, cf. [13, Ch. 4]. An application of this is to use mutation–crossover $\mathbf{M}_\mu \mathbf{C}_\chi^n$ as a proposal procedure for a simulated annealing type selection scheme (see Remark on simulated annealing below). In that case, such an annealing algorithm can be considered as a stochastic sequence consisting of operators $\mathbb{1}$ and $\mathbf{M}_\mu \mathbf{C}_\chi^n$.

2.3. Scaled fitness selection

In this section, we shall discuss the standard way to perform *fitness selection* in genetic algorithms. We also cover a wide variety of possible fitness scaling methods – methods of altering the fitness values during the course of the algorithm without ever making any creature more fit than one it was previously less fit than. Thus, one may potentially change the fitness operator in each iteration of the genetic algorithm. Such methods are commonly used in actual computer implementations of genetic algorithms [10, pp. 77–78], [19, pp. 166–170]. To our knowledge, this is the first general mathematical analysis of fitness scaling. The results are rather striking: In the broad class considered, which includes the standard scaling methods, *asymptotically, all fitness scaling methods are equivalent.* Only the (pre-)ordering of individuals is relevant in the longterm, and not the actual fitness values.[8]

Let f be a function $f : C \to \mathbb{R}^+ \setminus \{0\}$ called the (*raw*) *fitness function* on possible individuals. When used as an optimization procedure, the task of a genetic algorithm is to find *some* element $c \in C$ with maximal fitness $f(c)$.

In some of the results we shall develop from now on, we shall suppose that f is injective. This is not much of a disturbance for most cases. One possibility (of many) to overcome non-injective fitness functions in an implementation is to replace

[8] Baker [2] proposed to use only the fitness ranking of individuals in determining fitness selection probabilities. Our results show that asymptotically, Baker's criterion (rank) is the only essential factor in the limit probability distribution of a genetic algorithm with strong fitness scaling.

the fitness value $f(c)$ of each creature c by $f(c)+[\delta/(n+1)]$, where n is the value of c interpreted as a binary number and $\delta > 0$ is a lower bound for $\min\{|f(c)-f(d)| : c,d \in C, f(c) \neq f(d)\}$.

The fitness function f yields a *raw fitness vector* $f(p)$ for populations $p \in P$ by letting

$$f(p) = (f(c_1),\ldots,f(c_s)) \in \mathbb{R}^s,\ p = (c_1,\ldots,c_s) \in P.$$

Scaling sequence Let $\phi_n : (\mathbb{R}^+)^s \to (\mathbb{R}^+)^s$, $n \in \mathbb{N}$, be a sequence of functions.

1. The sequence $(\phi_n)_{n\in\mathbb{N}}$ will be called a *scaling sequence*, if for every $x = (x_1,\ldots,x_s) \in (\mathbb{R}^+)^s$ and $\phi_n(x) = (y_1,\ldots,y_s)$ the following conditions hold:
   (a) If $\pi \in \Pi_s$, then $\pi(\phi_n(x)) = \phi_n(\pi(x))$.
   (b) If $x_1 \leqslant x_2$ then $y_1 \leqslant y_2$.
   (c) If $x_1 < x_2$ then $y_2 = 0$ or $y_1 < y_2$.
   (d) $x = 0 \iff \phi_n(x) = 0$.
2. For $p \in P$, define $f_n(p) = \phi_n(f(p))$. The ith component $f_n(p,i)$ of $f_n(p)$ is defined as the *nth fitness value* of $c = c_i$ where $p = (c_1,\ldots,c_s)$. We write $f_n(p,c)$ for this value.
3. In addition, define for $n \in \mathbb{N}$ and $p \in P$:
   - $\max(n,p) = \max_{1\leqslant i\leqslant s} f_n(p,i)$
   - $\max(n) = \max_{1\leqslant i\leqslant s, p\in P} f_n(p,i)$
   - $\min(n) = \min_{1\leqslant i\leqslant s, p\in P} f_n(p,i)$
   - If $\{f_n(p,i) : 1 \leqslant i \leqslant s\}$ contains more than one element, then set

$$\max_2(n,p) = \max\{f_n(p,i) : f_n(p,i) \neq \max(n,p), 1 \leqslant i \leqslant s\}.$$

Otherwise, set $\max_2(n,p) = 0$. Let

$$\theta_n = \max_{p\in P} 1 - \left(1 + \frac{(s-1)\max_2(n,p)}{\max(n,p)}\right)^{-s}.$$

Under this definition, the *number* of positions with creatures of highest fitness in a fixed population p is the same for all $n \in \mathbb{N}$.

Scaled fitness selection. Let $p = (c_1,\ldots,c_s) \in P$, $n \in \mathbb{N}$. Scaled fitness selection of p is a lottery played for every position j in p, $1 \leqslant j \leqslant s$, in the following way: The creature c_i in position i, $1 \leqslant i \leqslant s$, is chosen to be the new creature at position j in the nth generation with probability

$$\frac{f_n(p,i)}{\sum_{j=1}^{s} f_n(p,j)}.$$

We remark that, as defined here and under standard implementations of genetic algorithms, fitness selection ignores spatial structure within the population. One way to model spatial structure would be to alter the probabilities so that the probability of c_i being copied to position j in the next generation would also vary with of the

distance between i and j, and perhaps other properties, such as the quality of resources at various positions in the "environment".

Proposition 11. *Scaled fitness selection can be described by column stochastic matrices* $\mathbf{F}_n$, $n \in \mathbb{N}$, *acting on* V.

1. *The components of* $\mathbf{F}_n$ *are determined as follows: let* $p = (c_1,\ldots,c_s)$, $q = (d_1,\ldots,d_s) \in P$, $c_i, d_i \in C$, $1 \leqslant i \leqslant s$. *In addition, let* $n(d_i)$ *denote the number of occurrences of* d_i *in* p.
   If $q \subseteq p$ *as sets and thus* $d_i \in p$, $1 \leqslant i \leqslant s$, *then we have*
   $$\langle \mathbf{F}_n p, q\rangle = \prod_{i=1}^{s} \frac{n(d_i) f_n(p,d_i)}{\sum_{j=1}^{s} f_n(p,j)}.$$
   If q *is not contained in* p *as a set, then* $\langle \mathbf{F}_n p, q\rangle = 0$.
2. *In particular,* $\langle \mathbf{F}_n p, p\rangle \geqslant (\min(n)/[s \max(n)])^s$, $p \in P$.
3. *If all creatures in* $p \in P$ *have the same fitness value, then* $\langle \mathbf{F}_n p, p\rangle \geqslant s^{-s}$.
4. *For every permutation* $\pi \in \Pi_s$, *we have* $\pi \mathbf{F}_n = \mathbf{F}_n = \mathbf{F}_n \pi$.
5. $\mathbf{F}_n$ *acts as the identity map on* U. *Thus,* $\langle \mathbf{F}_n p, p\rangle = 1$ *if* p *is uniform.*
6. *If* $v = (\mathbb{1} - \mathbf{P}_U)v \in S$, *then* $\|(\mathbb{1} - \mathbf{P}_U)\mathbf{F}_n v\|_1 \leqslant 1 - s^{-s}$.
7. *The transient states of* $\mathbf{F}_n$ *are exactly the non-uniform populations.*
8. *If* f *is injective and* $v = (\mathbb{1} - \mathbf{P}_U)v \in S$, *then* $\|(\mathbb{1} - \mathbf{P}_U)\mathbf{F}_n v\|_1 \leqslant \theta_n$, *where* θ_n *is as in the definition of scaling sequence.*
9. *Let* $\mathbf{C}_\chi$ *represent either simple or unrestricted crossover. If* $\mathbf{M}_\mu$ *represents one-bit mutation and* $\min(n) > 0$, *then* $(\mathbf{F}_n \mathbf{M}_\mu \mathbf{C}_\chi^k)^L$, $k \in \mathbb{N}$, *is fully positive. Furthermore, the coefficients of* $(\mathbf{F}_n \mathbf{M}_\mu \mathbf{C}_\chi^k)^L$ *are uniformly bounded away from* 0 *if* $\min(n)/\max(n)$ *is,* μ *(considered as a variable) stays uniformly bounded away from* 0, *and, in the case of simple crossover,* χ *stays uniformly bounded away from* 1.
10. *Let* $\mathbf{C}_\chi$ *represent either simple or unrestricted crossover. Let* $n \in \mathbb{N}$. *If* $\mathbf{M}_\mu$ *represents multiple-bit mutation, then all coefficients of* $\mathbf{M}_\mu \mathbf{C}_\chi^k \mathbf{F}_n$, $k \in \mathbb{N}$, *are uniformly bounded away from* 0 *independently of* $k, n \in \mathbb{N}$ *and of* χ.

Proof. (1), (4) and (5) are obvious. (2) and (3) follow from (1). (6) and (8) are obtained as follows: let $n \in \mathbb{N}$ and $p = (c_1,\ldots,c_s) \in P$ be non-uniform such that without loss of generality $c_1, c_2, \ldots, c_m$ have maximal fitness $f_n(p,c_1)$ in p and all other creatures have lower fitness. The probability to select c_1 for a specific position is at least

$$\frac{\gamma \max(n,p)}{m \max(n,p) + (s-m)\max_2(n,p)} \qquad (*)$$

where $\gamma = 1$ in the proof of (6) and $\gamma = m$ in the proof of (8). Thus, a lower estimate for the combined probability to generate uniform populations is s^{-s} in the proof of (6) and $1 - \theta_n$ in the proof of (8) with θ_n as defined above. (7) follows from (5) and (6). (9) follows the fact that $\mathbf{C}_\chi$ and, by (2), $\mathbf{F}_n$ have strictly positive diagonals and $\mathbf{M}_\mu^L$ is fully positive by Proposition 1.4. (10) follows from the fact that each entry in $\mathbf{M}_\mu \mathbf{S}$, where $\mathbf{S}$ is any column stochastic matrix, has all entries at least μ^L. □

By Proposition 11.7, $\mathbf{F}_n$ is a map that pulls towards the subspace U of uniform populations, in contrast to $\mathbf{M}_\mu$ which pulls in direction e away from U. This interplay will be made precise in Lemma 13 below.

Fitness Scaling Let $\phi_n : (\mathbb{R}^+)^s \to (\mathbb{R}^+)^s$, $n \in \mathbb{N}$, be a scaling sequence.
(1) $(\phi_n)_{n\in\mathbb{N}}$ will be called a *fitness scaling*, if $\mathbf{F}_\infty = \lim_{n\to\infty} \mathbf{F}_n$ exists.
(2) $(\phi_n)_{n\in\mathbb{N}}$ will be called a *strong fitness scaling*, if, for every $p \in P$,

$$\lim_{n\to\infty} \frac{\max_2(n,p)}{\max(n,p)} = 0.$$

Note for a strong fitness scaling: $\theta_n \to 0$ as $n \to \infty$.

Linear fitness scaling as defined in [10, pp. 77–79] comprises a fitness scaling[9], as does taking all ϕ_n to be the identity.

Another example of a fitness scaling is given by *sigma-truncation* (cf. [10, p. 124]) for populations that are non-uniform in fitness, set

$$\phi_n(r_1,\ldots,r_s) = ((r_1 - \bar{r} + c\sigma)^+,\ldots,(r_s - \bar{r} + c\sigma)^+), \quad r_i \in \mathbb{R}^+, 1 \leqslant i \leqslant s,$$

where $\bar{r}$ and σ denote the average and standard deviation of the r_i, and $c \geqslant 0$. For uniform populations, set $\phi_n(r_1,\ldots,r_s) = (1,\ldots,1)$.

Rank selection introduced by Baker ([2], [19, pp. 169–170]) is a fitness scaling method in which absolute differences in fitness are ignored and only the fitness (pre-) ordering matters: For $(r_1,\ldots,r_s)$, $r_i \in \mathbb{R}^+$, $1 \leqslant i \leqslant s$, let $rank(r_i) = cardinality\{j : r_i \geqslant r_j\}$.

$$\phi_n(r_1,\ldots,r_s) = (rank(r_1),\ldots,rank(r_s)).$$

Another example of a fitness scaling is given by *power law scaling* as defined in [10, p. 124]:

$$\phi_n(r_1,\ldots,r_s) = (r_1^{t_n},\ldots,r_s^{t_n}), \quad r_i \in \mathbb{R}^+, 1 \leqslant i \leqslant s,$$

where $t_n \in \mathbb{R}^+$, $n \in \mathbb{N}$ is an increasing sequence with $t_1 \geqslant 1$. This can be seen as an analogue to a cooling schedule in simulated annealing. See e.g. [1, p. 42] or [21, p. 749]. It has been used in [8]. Power scaling is a strong fitness scaling if and only if the t_n increase without bound.

A slight conceptual variation of this strong fitness scaling is so-called *Boltzmann selection* (*e.g.* [19, pp. 168–169]) also inspired by simulated annealing:

$$\phi_n(r_1,\ldots,r_s) = (e^{r_1/T_n},\ldots,e^{r_s/T_n}) \quad r_i \in \mathbb{R}^+, 1 \leqslant i \leqslant s,$$

where $T_n > 0$ is "temperature" and $\lim_{n\to\infty} T_n = 0$.

Proposition 12. 1. *If $(\phi_n)_{n\in\mathbb{N}}$ is a strong fitness scaling, then it is a fitness scaling.*
Furthermore, the components of $\mathbf{F}_\infty$ are determined as follows: Let $p = (c_1,\ldots,c_s)$,

[9] Note that in the implementation of the procedure `prescale` given by [10, p. 79] one has to include an additional case "`umax` equals `uavg`". In that case, `prescale` should do nothing.

$q = (d_1, \ldots, d_s)$, $c_i, d_i \in C$, $1 \leqslant i \leqslant s$, *and denote by* $n(d_i)$ *the number of occurrences of* d_i *in* p.

- *If* $q \subseteq p$ *as sets and all creatures in* q *have maximal fitness in* p, *then we have*

$$\langle \mathbf{F}_\infty p, q \rangle = \prod_{i=1}^{s} \frac{n(d_i)}{m_p},$$

where m_p *is the number of creatures with maximal fitness in* p.

- *Otherwise,* $\langle \mathbf{F}_\infty p, q \rangle = 0$.

$\mathbf{F}_\infty$ *depends solely on the pre-order on* C *induced by fitness, i.e.* $c \geqslant c' \iff f(c) \geqslant f(c')$, $c, c' \in C$. *In particular, the coefficients of* $\mathbf{F}_\infty$ *are independent of the actual fitness* values *and any particular* method *of fitness scaling.*

Suppose now that $(\phi_n)_{n \in \mathbb{N}}$ *is a fitness scaling, but not necessarily a strong fitness scaling*:

2. *If all creatures in* $p \in P$ *have the same fitness value, then* $\langle \mathbf{F}_\infty p, p \rangle \geqslant s^{-s}$.
3. *If* $p \in P$ *is uniform, then* $\langle \mathbf{F}_\infty p, p \rangle = 1$. *Hence,* $\mathbf{F}_\infty$ *acts as the identity on the subspace* U *generated by uniform populations.*
4. *If* $v = (\mathbb{1} - \mathbf{P}_U)v \in S$, *then* $\|(\mathbb{1} - \mathbf{P}_U)\mathbf{F}_\infty v\|_1 \leqslant 1 - s^{-s}$.
5. *The transient states of* $\mathbf{F}_\infty$ *are exactly the non-uniform populations.*
6. *Let* $\mathbf{C}_\chi$ *describe either simple or unrestricted crossover. If* $\mathbf{M}_\mu$ *represents multiple-bit mutation, then* $\mathbf{M}_\mu \mathbf{C}_\chi^n \mathbf{F}_\infty$, $n \in \mathbb{N}$, *is fully positive.*

Proof. These assertions follow by continuity from Proposition 11. □

Remark on simulated annealing. Let us briefly discuss connecting mutation and crossover with the selection scheme of simulated annealing and the Metropolis algorithm (cf. [1, 4, 20]) as proposed in [9, 18].

The fitness fitness f induces an *average fitness* $\bar{f}$ on populations given by

$$\bar{f}(p) = \frac{1}{s} \sum_{i=1}^{s} f(c_i), \quad p = (c_1, \ldots, c_s) \in P.$$

If $\bar{f}$ obtains a global maximum on populations at p, then p contains only creatures of globally maximal fitness.

Let $\mathbf{M}_\mu$ describe either model for mutation. And, let $\mathbf{C}_\chi$ describe either model for crossover. Let $n \in \mathbb{N}$ be fixed. In accordance with [1, p. 36] and [21, p. 752], we define the *generator matrix* $\mathbf{g}(\mu, \chi)$ of the Metropolis selection scheme to be $\mathbf{g}(\mu, \chi) = \mathbf{M}_\mu \mathbf{C}_\chi^n$. For one-bit mutation, by Propositions 1.4, 7.1 and 9.1, the underlying graph of the inhomogeneous Markov chain associated with the Metropolis selection scheme for fitness selection is connected. For multiple-bit mutation, connectivity follows from Proposition 3.4. By Propositions 7.4 and 9.4, $\mathbf{g}$ is symmetric in the sense of [21, p. 753]. This shows that the hypotheses for a generator matrix as specified in [21] hold. In regard to definition and assured convergence of a simulated annealing type algorithm using our $\mathbf{g}$ as a generator matrix and $-\bar{f}$ as a cost function, the reader may now consult [21, p. 752, Theorems 5.1 and 5.2].

3. Strong ergodicity of genetic algorithms

This section considers product operators representing the genetic algorithm comprised from the three basic genetic operators which have been studied above separately.

Genetic algorithm. Let $\mathbf{M}_\mu$ represent either one- or multiple-bit mutation and $\mathbf{C}_\chi$ either simple or unrestricted crossover. Let $k \in \mathbb{N}$ be fixed. Let $\mathbf{F}_n$ represented scaled fitness selection for a fitness scaling $\{\phi_n\}_{n\in\mathbb{N}}$ as defined above. We define the nth *step* $\mathbf{S}_{\mu,\chi_n}$ in a genetic algorithm as the matrix product:

$$\mathbf{S}_{\mu_n,\chi_n} = \mathbf{F}_n \cdot \mathbf{C}_{\chi_n}^k \cdot \mathbf{M}_{\mu_n}$$

for any admissible choice of parameters $\mu_n, \chi_n \in [0,1]$ permitted in the definitions of the various operators.

A genetic algorithm consists of a sequence of applications of matrices $\mathbf{S}_{\mu_n,\chi_n}$. Thus, application of the first n steps of a genetic algorithm is completely described by the "backwards" product

$$\mathbf{G}_n = \prod_{i=n}^{1} \mathbf{S}_{\mu_i,\chi_i} = \mathbf{S}_{\mu_n,\chi_n} \cdots \mathbf{S}_{\mu_1,\chi_1}.$$

If ϕ_n is the identity map id, $\mu_n = \mu$, $\chi_n = \chi$ for all $n \in \mathbb{N}$, then we may write $\mathbf{S} = \mathbf{S}_{\mu,\chi} = \mathbf{S}_{\mu_n,\chi_n}$, $\mathbf{G}_n = \mathbf{S}^n$, and the genetic algorithm is called *simple*.

3.1. Strong ergodicity of simple genetic algorithms

Next, we study the loss of non-uniformity during the course of a genetic algorithm by giving explicit bounds. Loss of diversity in finite populations as a result of "sampling error" is what evolutionary and population biologists refer to as "genetic drift" (e.g. [26, Ch. 5], [22, Chs. 2, 8]).

Lemma 13. *Let* $\mathbf{S} = \mathbf{F}_n\mathbf{C}_\chi^k\mathbf{M}_\mu$ *be a step in a genetic algorithm as in the above definition. Let* $v \in S$ *be a probability distribution over populations. If* f *is injective, then let* θ_n *be as in the definition of fitness scaling, otherwise let* $\theta_n = 1 - s^{-s}$ *for all* $n \in \mathbb{N}$.

1. *If* $\mathbf{M}_\mu$ *stands for one-bit mutation, then we have*

$$\|(\mathbb{1} - \mathbf{P}_U)\mathbf{S}v\|_1 \leqslant \theta_n(L\mu + (1 - L\mu)\|(\mathbb{1} - \mathbf{P}_U)v\|_1).$$

2. *If* $\mathbf{M}_\mu$ *stands for one-bit mutation, then we have for* $m \in \mathbb{N}$

$$\|(\mathbb{1} - \mathbf{P}_U)\mathbf{S}^m v\|_1 \leqslant \theta_n L\mu(1 - \theta_n(1 - L\mu))^{-1} + \theta_n^m(1 - L\mu)^m\|(\mathbb{1} - \mathbf{P}_U)v\|_1.$$

3. *If* $\mathbf{M}_\mu$ *stands for multiple-bit mutation, then we have*

$$\|(\mathbb{1} - \mathbf{P}_U)\mathbf{S}v\|_1 \leqslant \theta_n(1 - (\mu^s + (1 - \mu)^s)^\ell + \beta\|(\mathbb{1} - \mathbf{P}_U)v\|_1),$$

where $\beta = (\mu^s + (1 - \mu)^s)^\ell - ((1 - \mu)^h\mu^{s-h} + \mu^h(1 - \mu)^{s-h})^\ell$, $h = \lfloor s/2 \rfloor$.

4. *If* $\mathbf{M}_\mu$ *stands for multiple-bit mutation, then we have for* $m \in \mathbb{N}$

$$\|(\mathbb{1} - \mathbf{P}_U)\mathbf{S}^m v\|_1 \leqslant \theta_n \frac{1 - (\mu^s + (1-\mu)^s)^\ell}{1 - \theta_n \beta} + \theta_n^m \beta^m \|(\mathbb{1} - \mathbf{P}_U)v\|_1.$$

Proof. For a moment, consider one-bit mutation. Let $\mathbf{T} = \mathbf{F}_n \mathbf{C}_\chi^k$, and $w = (\mathbb{1} - \mathbf{P}_U)v$. We have by Propositions 7.9, 9.9, and 11.5,

$$\mathbf{T}v = \mathbf{P}_U v + \mathbf{T}w = \mathbf{P}_U v + \mathbf{P}_U \mathbf{T}w + (\mathbb{1} - \mathbf{P}_U)\mathbf{T}w.$$

Hence, $(\mathbb{1} - \mathbf{P}_U)\mathbf{T}v = (\mathbb{1} - \mathbf{P}_U)\mathbf{T}w$. It follows from Proposition 7.9, 7.2 for simple crossover resp. Proposition 9.9, 9.2 for unrestricted crossover, and Proposition 11.6 resp. 11.8 that

$$\|(\mathbb{1} - \mathbf{P}_U)\mathbf{T}v\|_1 = \|(\mathbb{1} - \mathbf{P}_U)\mathbf{T}w\|_1 \leqslant \theta_n \|\mathbf{C}_\chi^k (\mathbb{1} - \mathbf{P}_U)w\|_1 \leqslant \theta_n \|w\|_1.$$

Thus, Proposition 9.10 and Proposition 2.4 imply that

$$\begin{aligned}\|(\mathbb{1} - \mathbf{P}_U)\mathbf{S}v\|_1 &= \|(1 - \mathbf{P}_U)\mathbf{T}\mathbf{M}_\mu v\|_1 \leqslant \theta_n \|(\mathbb{1} - \mathbf{P}_U)\mathbf{M}_\mu v\|_1 \\ &\leqslant \theta_n (L\mu + (1 - L\mu)\|w\|_1).\end{aligned}$$

(2) follows from (1) by iteration and a geometric series estimate. (3) and (4) are obtained similarly using Proposition 4.4. □

Lemma 14. *If* $\mathbf{T}, \mathbf{M} \in \mathbb{M}_n$ *are column stochastic, then* $\tau_1(\mathbf{TM}) \leqslant \tau_1(\mathbf{M})$.

Proof. If $v \in e^\perp$, then $\|\mathbf{TM}v\|_1 \leqslant \|\mathbf{M}v\|_1 \leqslant \tau_1(\mathbf{M}) \cdot \|v\|_1$. □

The estimates in the next result show that for small mutation rates, a simple genetic algorithm asymptotically spends most of its time in uniform populations regardless of crossover rate. This contributes to the understanding of "punctuated equilibrium" type properties of simple genetic algorithms, which have been treated by Vose and Liepins [32, p. 64], [33].

Theorem 15. *Let* $\mathbf{S} = \mathbf{S}_{\mu,\chi}$ *with* μ *and* χ *fixed represent each step of a simple genetic algorithm as in the preceding definition. If one-bit mutation is used, then suppose* $\min(n) = \min(1) > 0$, $n \in \mathbb{N}$. *If* f *is injective, then let* $\theta = \theta_1$ *be defined as in the definition of fitness scaling; otherwise, let* $\theta = 1 - s^{-s}$. *Then*

1. $\mathbf{S}$ *has a uniquely determined, fully positive fixed point probability distribution over populations* $v_\infty \in S \cap \mathbf{P}_{\Pi_s} V$. *Furthermore,* 1 *is the only eigenvalue of* $\mathbf{S}$ *in the unit circle and is a simple root of the characteristic polynomial of* $\mathbf{S}$.
2. *There exist fixed constants* $r \in [0,1)$, $K \in \mathbb{R}^+$ *such that*

$$\|\mathbf{S}^k v_0 - v_\infty\|_1 \leqslant K r^k, \quad k \in \mathbb{N},$$

for every initial probability distribution $v_0 \in S$.

3. *In the case of multiple-bit mutation, we have* $\tau_1(\mathbf{S}) \leqslant 1 - (2\mu)^L$.

4. *If one-bit mutation is considered, then* v_∞ *satisfies*

$$\|(\mathbb{1} - \mathbf{P}_U)v_\infty\|_1 \leqslant \frac{\theta L\mu}{1 - \theta(1 - L\mu)}.$$

5. *If multiple-bit mutation is considered, then* v_∞ *satisfies*

$$\|(\mathbb{1} - \mathbf{P}_U)v_\infty\|_1 \leqslant \frac{\theta(1 - (\mu^s + (1-\mu)^s)^\ell}{1 - \theta\beta},$$

where $\beta = (\mu^s + (1-\mu)^s)^\ell + ((1-\mu)^h\mu^{s-h} + \mu^h(1-\mu)^{s-h})^\ell$, $h = \lfloor s/2 \rfloor$.

Proof. Propositions 11.9 and 11.10 assure that the some power of $\mathbf{S}$ is a fully positive matrix. The fact that $v_\infty \in S \cap \mathbf{P}_{\Pi_s} V$ follows from Proposition 11.4. The remaining statements in (1) follow now from the discussion of fully positive matrices in the notation section of this paper. The geometric convergence in (2) follows from (1) and consideration of the Jordan decomposition of $\mathbf{S}$. To establish (3), use Lemma 14 and Proposition 4.3. The estimates (4) and (5) follow from Lemma 13.2 and 13.4, respectively. □

Theorem 15.1 contains also results found independently by Rudolph [25, Theorem 4]. Rudolph also shows the convergence to a global optimum for simple genetic algorithms with modified genetic operators which always record a creature with highest fitness value encountered so far.

We remark that as a trivial consequence of Theorem 15, one obtains convergence to a limit distribution over populations as *multi-sets*, i.e. as sets with multiplicity but no order on their elements. Thus, the probability distributions over multi-sets as in the formalization of simple genetic algorithms described in [5, 6, 23, 31–33] converge to a *strictly positive* distribution over multi-set populations. Furthermore, the limit is independent of the initial distribution (and hence of the any initial choice of population).

3.2. Strong ergodicity under fitness scaling

In view of Proposition 11.10, the following technical result on ergodicity naturally arises.

Theorem 16. *Let* $\mathbf{T}_n$, $n \in \mathbb{N}$, *be a sequence of fully positive, column stochastic matrices such that* $\mathbf{T}_\infty = \lim_{n\to\infty} \mathbf{T}_n$ *exists and is fully positive. Let* v_n *resp.* v_∞ *be the uniquely determined fully positive probability distributions belonging to eigenvalue* 1 *of* $\mathbf{T}_n$ *resp.* $\mathbf{T}_\infty$. *Let* $\mathbf{Q}_\infty$ *be the projection whose columns coincide with* v_∞. *Let* w_n *be the uniquely determined, fully positive probability distribution belonging to eigenvalue* 1 *of*

$$\mathbf{H}_n = \prod_{i=n}^{1} \mathbf{T}_i.$$

In this situation,

$$\mathbf{Q}_\infty = \lim_{n\to\infty} \mathbf{H}_n \text{ and } v_\infty = \lim_{n\to\infty} v_n = \lim_{n\to\infty} w_n.$$

Before we turn to the proof of Theorem 16, we note that according to [29, p. 41, Corollary], for a (column) stochastic matrix $\mathbf{T} \in \mathbb{M}_n$ with smallest entry t, each eigenvalue $\lambda \neq 1$ of $\mathbf{T}$ satisfies $|\lambda| \leqslant 1 - nt$.

Proof of Theorem 16. Since $\mathbf{T}_\infty = \lim_{n\to\infty} \mathbf{T}_n$ we conclude, by the discussion of coefficients of ergodicity in the notation section of this paper and the above, that there exists an $\varepsilon > 0$ such that $\tau_1(\mathbf{T}_n) \leqslant 1 - \varepsilon$ for all $n \in \mathbb{N}$, and the modulus of the second largest eigenvalue is less than $1 - \varepsilon$ for all $n \in \mathbb{N}$. Using the symbolic calculus for matrix algebras (see Appendix B.2 or [27, Sections 10.21–10.33]), we can compute projections $\mathbf{P}_n$, $n \in \mathbb{N} \cup \{\infty\}$, onto the one-dimensional eigenspaces of $\mathbf{T}_n$ to eigenvalue 1 by

$$\mathbf{P}_n = \frac{1}{2\pi i} \int_\Gamma (\zeta \mathbb{1} - \mathbf{T}_n)^{-1} \, \mathrm{d}\zeta,$$

where Γ is the circle in the complex plane around 1 of radius $\varepsilon/2$. This expression is continuous in $\mathbf{T}_n$. Hence, $\lim_{n\to\infty} \mathbf{P}_n = \mathbf{P}_\infty \neq \mathbf{0}$. Consequently, by summing over the absolute values of the columns of each individual $\mathbf{P}_n$, one derives $v_\infty = \lim_{n\to\infty} v_n$. Thus, the conditions of [15, Theorem V.4.4] are satisfied (with corresponding $D = \sum_{k=0}^{\infty} (1-\varepsilon)^k = \frac{1}{\varepsilon}$). By the proof of [15, Theorem V.4.4], we have $\mathbf{Q}_\infty = \lim_{n\to\infty} \mathbf{H}_n$. Observing that $\mathbf{Q}_\infty$ is fully positive, we get $v_\infty = \lim_{n\to\infty} w_n$ by applying the argument with the integral once more to $\mathbf{H}_n$ instead of $\mathbf{T}_n$. □

Theorem 16 is a special case of [7, Theorem 1.1]. The latter result is stated in [7] and then a continuous-time analogue is proved. We have included the simple proof given above for convenience.

Variation schedule. Let $\mu_n \in (0, \frac{1}{2}]$ and $\chi_n \in (0, 1]$, $n \in \mathbb{N}$ be such that $0 < \mu_\infty = \lim_{n\to\infty} \mu_n$ exists, and $\chi_\infty = \lim_{n\to\infty} \chi_n$ exists. Suppose, in addition, $\mu_\infty, \mu_n < 1/L$ if one-bit mutation is used, and $\chi_\infty, \chi_n < 1$ if simple crossover is used. In this situation, the sequence of pairs (μ_n, χ_n) will be called a *variation schedule* for the genetic algorithm. (Note again that μ_∞ is greater than 0.)

Theorem 17. *Let $\mathbf{M}_\mu$ describe multiple-bit mutation and $\mathbf{C}_\chi$ describe either model for crossover. Let $r \in \mathbb{N}$. Let $(\mu_n, \chi_n)_{n\in\mathbb{N}}$ be a variation schedule and $(\phi_n)_{n\in\mathbb{N}}$ be a fitness scaling sequence, and $\mathbf{F}_n$ describe scaled fitness selection according to this scaling. If f is injective, then let $\theta = \liminf_{n\to\infty} \theta_n$, where θ_n is as in the definition of fitness scaling. Otherwise, let $\theta = 1 - s^{-s}$. Let $\mathbf{G}_n$ represent the first n steps of a genetic algorithm as defined in the beginning of Section 3 using the variation schedule $(\mu_n, \chi_n)_{n\in\mathbb{N}}$.*

In this situation,

$$v_\infty = \lim_{n\to\infty} \mathbf{G}_n v_0 = \lim_{n\to\infty} (\mathbf{F}_\infty \cdot \mathbf{M}_{\mu_\infty} \cdot \mathbf{C}_{\chi_\infty}^k)^n v_0$$

exists and is independent of the choice of $v_0 \in S$. *Furthermore,*

$$\|(1 - \mathbf{P}_U)v_\infty\|_1 \leqslant \frac{\theta(1 - (\mu_\infty^s + (1 - \mu_\infty)^s)^\ell)}{1 - \theta\beta},$$

where $\beta = (\mu_\infty^s + (1 - \mu_\infty)^s)^\ell - ((1 - \mu_\infty)^h \mu_\infty^{s-h} + \mu_\infty^h (1 - \mu_\infty)^{s-h})^\ell$, $h = \lfloor s/2 \rfloor$.

The coefficients $\langle v_\infty, p\rangle$ *of the limit probability distribution are strictly positive for every population* $p \in P$ *of uniform fitness. In the case of a non-constant fitness function, the genetic algorithm does not converge to a population consisting solely of creatures with maximal fitness value.*

Proof. Using Proposition 11.10 and Theorem 16, we obtain for every $v \in S$.

$$\begin{aligned}
v_\infty &= \lim_{n\to\infty} \mathbf{G}_n v \\
&= \lim_{n\to\infty} \prod_{i=n}^{1} \mathbf{F}_i \mathbf{C}_{\chi_i}^k \mathbf{M}_{\mu_i} v \\
&= \lim_{n\to\infty} \mathbf{F}_n \left(\prod_{i=n}^{2} \mathbf{C}_{\chi_i}^k \mathbf{M}_{\mu_i} \mathbf{F}_{i-1} \right) \mathbf{C}_{\chi_1}^k \mathbf{M}_{\mu_1} v \\
&= \mathbf{F}_\infty \lim_{n\to\infty} (\mathbf{C}_{\chi_\infty}^k \mathbf{M}_{\mu_\infty} \mathbf{F}_\infty)^n \mathbf{C}_{\chi_1}^k \mathbf{M}_{\mu_1} v \\
&= \mathbf{F}_\infty w \\
&= \mathbf{F}_\infty \lim_{n\to\infty} (\mathbf{C}_{\chi_\infty}^k \mathbf{M}_{\mu_\infty} \mathbf{F}_\infty)^n \mathbf{C}_{\chi_\infty}^k \mathbf{M}_{\mu_\infty} v \\
&= \lim_{n\to\infty} (\mathbf{F}_\infty \mathbf{M}_{\mu_\infty} \mathbf{C}_{\chi_\infty}^k)^n v,
\end{aligned}$$

where $w \in S$ is the uniquely determined fixed point of $\mathbf{C}_{\chi_\infty}^k \mathbf{M}_{\mu_\infty} \mathbf{F}_\infty$ in S. Since w is fully positive, we obtain by Proposition 12.2 that $\langle v_\infty, p\rangle = \langle \mathbf{F}_\infty w, p\rangle > 0$ for every population $p \in P$ of uniform fitness. Finally, we have by Lemma 13.4:

$$\begin{aligned}
\|(\mathbb{1} - \mathbf{P}_U)v_\infty\|_1 &= \|(\mathbb{1} - \mathbf{P}_U)(\mathbf{F}_\infty \mathbf{M}_{\mu_\infty} \mathbf{C}_{\chi_\infty}^k)^m v_\infty\|_1 \qquad \text{for all } m \in \mathbb{N} \\
&= \lim_{n\to\infty} \|(\mathbb{1} - \mathbf{P}_U)(\mathbf{F}_n \mathbf{M}_{\mu_\infty} \mathbf{C}_{\chi_\infty}^k)^m v_\infty\|_1 \\
&\leqslant \frac{\theta(1 - (\mu_\infty^s + (1 - \mu_\infty)^s)^\ell)}{1 - \theta\beta} + \theta^m \beta^m \|(\mathbb{1} - \mathbf{P}_U)v_\infty\|_1,
\end{aligned}$$

using the fact that a subsequence of the θ_n converges to θ. □

The study in [11] suggests that in general v_∞ may depend very much upon the fitness function f. However, there is *no* such dependence on any particular scaling method if a strong fitness scaling is used. This is shown in the next result.

Theorem 18. *Consider the situation of Theorem* 17. *Suppose in addition, that the* $(\phi_n)_{n\in\mathbb{N}}$ *is a strong fitness scaling.*

1. *The limit distribution* v_∞ *of* $\mathbf{F}_\infty \mathbf{M}_{\mu_\infty} \mathbf{C}_{\chi_\infty}^k$ *which equals the limit distribution of the genetic algorithm is* independent *from any particular method of strong fitness scaling.*

2. $\langle v_\infty, p\rangle = 0$ *for every population* $p \in P$ *which is not of uniform fitness.*
3. *If* f *is injective, then* $\langle v_\infty, p\rangle = 0$ *for every non-uniform population* $p \in P$.

Proof. (1) follows from Proposition 12.1 and Theorem 17. (2) follows from Proposition 12.2. (3) follows from (2). □

Theorems 17 and 18 show the *asymptotic* failure of genetic algorithms, even with fitness scaling and clever schedules for mutation and crossover rate, as optimizers. The only assumptions that were made to obtain these results were that the mutation and crossover rates converge as in the definition of variation schedule. This leaves open the case in which the mutation rate converges to zero (an analog of simulated annealing). This question [10] has been addressed and answered for simple genetic in the negative by Davis and Principe [5, 6].

The result on asymptotic failure of genetic algorithms proved in Theorem 17 allowed for arbitrary fitness functions and scaling methods. However, it is true that genetic algorithms have performed well in applications for many fitness functions [10, 19]. An interesting research direction is to characterize those fitness functions for which genetic algorithms succeed as optimizers in the short-term (although our results describe long-term behaviour).

Appendix A. Computations of spectra

The following *Mathematica*^{TM} program computes the spectrum of the matrix Γ_s of Section 2. First, we generate all permutations of the set $\{1,\ldots,s\}$.

```
perm[s_] := Permutation[Range[s]]
```

Next, we define the function d on pairs of permutations which is 1 if the permutations differs by a transposition and 0 otherwise. With the help of d, we define matrix which equals $\frac{1}{2}s(s-1)\Gamma_s$.

```
d[p1_,p2_] := If[(Map[Abs,Map[Sign,p1-p2]]//.List->Plus)==2,1,0]
matrix[s_] := (p=perm[s];Table[d[p[[k]],p[[n]],k,s!,n,s!])
```

Finally, we compute the spectrum of $\frac{1}{2}s(s-1)\Gamma_s$.

```
sp[s_] := Union[Eigenvalues[N[matrix[s]]]]
```

[10] The case of mutation converging to zero in a variation schedule is claimed to be solved in [30]. However, the 'proof' is flawed: first of all the fixed point distribution $q_p[s]$ for the ergodic Markov chain of a single step of the scaled algorithm is not a function of the parameter F introduced in [30, p. 55] (keeping mutation and crossover rates fixed). Consequently, there is no $\partial q_p[s]/\partial F$ and no way of estimating it. Even in the case of power law scaling with sequence of exponents $m(t)$, $t \in \mathbb{N}$, and interpretation of $F = F(t)$ as a parameter converging to zero as $t \to \infty$ such that $m(t) = \log_{F(0)}(F(t))$, it is by far not clear how $\partial q_p[s]/\partial F$ is kept bounded, as claimed without proof in [30, p. 67].

We obtain (omitting trailing numerical zeroes):

```
sp[2] = {-1, 1}
sp[3] = {-3, 0, 3}
sp[4] = {-6, -2, 0, 2, 6}
sp[5] = {-10, -5, -2, 0, 2, 5, 10}
sp[6] = {-15, -9, -5, -3, 0, 3, 5, 9, 15}
```

This suggests that the second largest eigenvalue of $\frac{1}{2}s(s-1)\Gamma_s$ is given by $\frac{1}{2}s(s-3)$.

Appendix B. Functional calculus for matrices

An anonymous referee has encouraged us to include a brief account of results on the spectral theorem and the analytic functional calculus for matrices. In the formulation we have needed in this paper, these results can be found (in much greater generality) with complete proofs in Rudin's book [27]. What we shall list in this appendix are some versions of results in [27] for finite dimensional spaces including indication of proof. While the spectral theorem for self-adjoint (or normal) matrices can be found in standard texts on linear algebra (see [16, p. 268] for a version on real symmetric matrices and [12, p. 337]), a similar "finite dimensional" exposition for the analytic functional calculus is, to our knowledge, not in the literature.

Appendix B.1. On the spectral theorem for self-adjoint matrices

Let $\mathbf{X} \in \mathbb{M}_n$ be a self-adjoint matrix and $\lambda \in sp(\mathbf{X})$. Let $0 \neq \xi \in \mathbb{C}^n$ be an eigenvector of $\mathbf{X}$ pertaining to λ. Then $\lambda \in \mathbb{R}$, since

$$\lambda\langle\xi,\xi\rangle = \langle\lambda\xi,\xi\rangle = \langle\mathbf{X}\xi,\xi\rangle = \langle\xi,\mathbf{X}\xi\rangle = \overline{\lambda}\langle\xi,\xi\rangle.$$

Let $\eta \in \mathbb{C}^n$ be perpendicular to ξ. Then $\mathbf{X}\eta$ is also perpendicular to ξ, since

$$\langle\mathbf{X}\eta,\xi\rangle = \langle\eta,\mathbf{X}\xi\rangle = \lambda\langle\eta,\xi\rangle = 0.$$

Theorem B.1. *There exists an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $\mathbf{X}$.*

The proof of Theorem B.1 works by induction on the dimension of the space: There exists a non-zero eigenvector ξ of $\mathbf{X}$ and $\mathbf{X}$ maps $\xi^{\perp}$ into itself by the above. Now, $\xi^{\perp}$ and the action of $\mathbf{X}$ on $\xi^{\perp}$ can be identified with $\mathbb{C}^{n-1}$ and the action of a self-adjoint matrix on $\mathbb{C}^{n-1}$.

The next result is used repeatedly in this paper:

Corollary B.2. *Let $sp(\mathbf{X}) = \{1, \lambda_k : k = 1,\ldots,k_0\}$ such that $|\lambda_k| < 1$ for $k = 1,\ldots,k_0$. If η is perpendicular to the eigenspace for eigenvalue 1 of $\mathbf{X}$, then*

$$\min_{k=1,\ldots,k_0} |\lambda_k| \cdot \|\eta\|_2 \leqslant \|\mathbf{X}\eta\|_2 \leqslant \max_{k=1,\ldots,k_0} |\lambda_k| \cdot \|\eta\|_2$$

Proof. Let $\eta = \sum_{i=1}^{n} t_i \xi_i$, where $\{\xi_i\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $\mathbf{X}$. Then we have

$$\|\mathbf{X}\eta\|_2 = \|\sum \lambda_{k(i)} t_i \xi_i\|_2 = (\sum \lambda_{k(i)}^2 \cdot |t_i|^2)^{\frac{1}{2}} \leqslant \max_{k=1,\ldots,k_0} |\lambda_k| \cdot \|\eta\|_2.$$

The other half of the inequality follows similarly. □

The following result is also used repeatedly in this paper. It justifies the term C^*-positive. See [27, Theorems 12.32, 12.33].

Corollary B.3. *The following are equivalent for arbitrary* $\mathbf{X} \in \mathbb{M}_n$*:*
1. $\mathbf{X} = \mathbf{Y}^2$ *and* $\mathbf{Y}$ *is self-adjoint.*
2. $\mathbf{X}$ *is self-adjoint and* $sp(\mathbf{X}) \subseteq \mathbb{R}^+$.
3. $\langle \mathbf{X}\xi, \xi \rangle \geqslant 0$ *for all* $\xi \in \mathbb{C}^n$.

Proof. (1) $\Rightarrow$ (3): If (1) holds, then $\langle \mathbf{X}\xi, \xi \rangle = \langle \mathbf{Y}\xi, \mathbf{Y}\xi \rangle \geqslant 0$. (3) $\Rightarrow$ (2): We have $\langle \mathbf{X}\xi, \xi \rangle = \langle \xi, \mathbf{X}^*\xi \rangle = \langle \mathbf{X}^*\xi, \xi \rangle$ which implies $\mathbf{X} = \mathbf{X}^*$ by the polarization identity (extending this expression bilinearly). If λ is an eigenvalue of $\mathbf{X}$ and $\xi \neq 0$ is an associated eigenvector, then $\lambda \langle \xi, \xi \rangle = \langle \mathbf{X}\xi, \xi \rangle \geqslant 0$ which implies $\lambda \geqslant 0$. (2) $\Rightarrow$ (1): If $\mathbf{X}$ is self-adjoint and $sp(\mathbf{X}) \subseteq \mathbb{R}^+$, then using an orthonormal basis of eigenvectors of $\mathbf{X}$, it is easy to define a self-adjoint square root of $\mathbf{X}$. □

Appendix B.2. On the analytic functional calculus for matrices

Let $\mathbf{X} \in \mathbb{M}_n$ and $\mathbf{R}(\eta) = (\eta \mathbb{1}_n - \mathbf{X})^{-1}, \eta \notin sp(\mathbf{X})$. One checks that

$$\mathbf{R}(\eta_1) - \mathbf{R}(\eta_2) = -(\eta_1 - \eta_2)\mathbf{R}(\eta_1) \cdot \mathbf{R}(\eta_2), \qquad (*)$$

which shows that $\mathbf{R}$ is an analytic function outside of $sp(\mathbf{X})$. Now, one fixes disjoint open disks D_λ and circles $\gamma_\lambda \subset D_\lambda$ around each $\lambda \in sp(X)$. Let $U = \bigcup_{\lambda \in sp(X)} D_\lambda$ and $\gamma = \{\gamma_\lambda\}$ be the chain such that $W(\gamma, \lambda) = 1$ where $W(\gamma, \lambda)$ is the winding number of γ with respect to λ (cf. [17, p. 114]).

Proposition B.4. *One has* $\mathbf{X}^k = (1/2\pi \mathrm{i}) \int_\gamma \eta^k \mathbf{R}(\eta) \mathrm{d}\eta$, $k \in \mathbb{N}_0$.

Proof. Let $|\eta| = \|\mathbf{X}\|_1 + 1$. Then we have by a geometric series argument:

$$\eta^k \mathbf{R}(\eta) = \eta^{k-1}(\mathbb{1}_n - \eta^{-1}\mathbf{X})^{-1} = \eta^{k-1} \sum_{m=0}^{\infty} \eta^{-m} \mathbf{X}^m.$$

Let Γ be the circle of radius $\|X\|_1 + 1$ around 0. Then

$$\frac{1}{2\pi \mathrm{i}} \int_\Gamma \eta^k \mathbf{R}(\eta)\, \mathrm{d}\eta = \frac{1}{2\pi \mathrm{i}} \sum_{m=0}^{\infty} \left(\int_\Gamma \eta^{k-m-1}\, \mathrm{d}\eta \right) \mathbf{X}^m = \mathbf{X}^k,$$

since only the integral for $k = m$ is non-zero in the summation. Applying [17, p. 123, Theorem IV.2.2] allows one to replace Γ by γ. □

For any analytic function $f : U \to \mathbb{C}$, define

$$f(\mathbf{X}) = \frac{1}{2\pi \mathrm{i}} \int_{\gamma} f(\eta)\mathbf{R}(\eta)\ \mathrm{d}\eta.$$

Theorem B.5. *Let $f_k : U \to \mathbb{C}$ be analytic functions, $k = 1,2$. Then one has*

$$f_1(\mathbf{X}) \cdot f_2(\mathbf{X}) = (f_1 \cdot f_2)(\mathbf{X}).$$

Proof. Using Cauchy's Theorem [17, p. 122, Theorem IV.2.1] we may pick two disjoint chains $\gamma^{(k)} = \{\gamma_\lambda^{(k)}\}$ as above to define $f_k(\mathbf{X})$ for $k = 1,2$ where the circles $\gamma_\lambda^{(1)}$ lie inside of the corresponding $\gamma_\lambda^{(2)}$. Then we have, using $(\star)$, that

$$\begin{aligned} f_1(\mathbf{X}) \cdot f_2(\mathbf{X}) &= \frac{1}{(2\pi i)^2} \int_{\gamma^{(1)}} \int_{\gamma^{(2)}} f_1(\eta_1) f_2(\eta_2)(\eta_2 - \eta_1)^{-1}(\mathbf{R}(\eta_1) - \mathbf{R}(\eta_2))\,\mathrm{d}\eta_2\,\mathrm{d}\eta_1 \\ &= \frac{1}{(2\pi i)^2} \int_{\gamma^{(1)}} f_1(\eta_1) \left(\int_{\gamma^{(2)}} \frac{f_2(\eta_2)}{\eta_2 - \eta_1}\,\mathrm{d}\eta_2 \right) \mathbf{R}(\eta_1)\,\mathrm{d}\eta_1 - 0 \\ &= \frac{1}{2\pi i} \int_{\gamma^{(1)}} f_1(\eta_1) f_2(\eta_1)\mathbf{R}(\eta_1)\,\mathrm{d}\eta_1 = (f_1 \cdot f_2)(\mathbf{X}), \end{aligned}$$

where Cauchy's formula [17, p. 124, Theorem IV.2.3] has been used twice. Note that $\eta_2 \in \gamma^{(2)}$ lies outside the circles in $\gamma^{(1)}$ yielding 0 for the summand containing $\mathbf{R}(\eta_2)$ in the second-to-last line in the above computation. □

As a consequence of the last theorem, the characteristic functions χ_{D_λ}, $\lambda \in sp(\mathbf{X})$ of the disks D_λ yield projections $\mathbf{P}_\lambda = \chi_{D_\lambda}(\mathbf{X})$, such that $\mathbf{P}_\lambda \cdot \mathbf{P}_{\lambda'} = 0$ if $\lambda \neq \lambda'$. Furthermore, by Proposition B.4, $\sum \mathbf{P}_\lambda = \mathbb{1}_n$.

Suppose now that $0 \neq \xi \in \mathbb{C}^n$ such that $(\lambda \mathbb{1}_n - \mathbf{X})^m \xi = 0$ for some $m \in \mathbb{N}$ and fixed $\lambda \in sp(\mathbf{X})$. We know from the Jordan decomposition of $\mathbf{X}$ that the function $\eta \mapsto \mathbf{R}(\eta)\xi$ has only a removable singularity inside $\gamma_{\lambda'}$ for $\lambda' \in sp(\mathbf{X})$ such that $\lambda' \neq \lambda$. Hence, the Cauchy Theorem implies

$$\mathbf{P}_{\lambda'}(\xi) = \frac{1}{2\pi i} \int_{\gamma_{\lambda'}} \mathbf{R}(\eta)\xi\,\mathrm{d}\eta = 0,$$

which in turn implies $(\mathbb{1}_n - \mathbf{P}_\lambda)\xi = 0$. A dimension counting argument shows that $\mathbf{P}_\lambda$ is a projection onto the kernel of $(\lambda \mathbb{1}_n - \mathbf{X})^n$. These facts are used in the proof of Theorem 16.

References

[1] E.H.L. Aarts, P.J.M. Laarhoven, Simulated annealing: an introduction, Statistica Neerlandica 43 (1989) 31–52.

[2] J.E. Baker, Reducing bias and inefficiency in the selection algorithm, in: J.J. Grefenstette (ed.), Genetic Algorithms and Their Applications: Proc. 2nd Internat. Conf. on Genetic Algorithms, Erlbaum, 1987.

[3] A.D. Bethke, Genetic Algorithms as Function Optimizers, Ph.D. Dissertation, University of Michigan, Dissertation Abstracts International 41 (9), 3503B, University Microfilms No. 8106101 (1981).

[4] K. Binder, Monte Carlo Methods in Statistical Physics, Springer, Berlin, 1978.
[5] T.E. Davis, Toward an extrapolation of the simulated annealing convergence theory onto the simple genetic algorithm, Ph.D. Dissertation, University of Florida, 1991.
[6] T.E. Davis, J.C. Principe, A simulated annealing-like convergence theory for the simple genetic algorithm, Proc. 4th Internat. Conf. on Genetic Algorithms, 1991, pp. 174–181.
[7] B. Gidas, Nonstationary Markov chains and convergence of the annealing algorithm, J. Stat. Phys. 39 (1985) 73–131.
[8] A.M. Gillies, Machine learning procedures for generating image domain feature detectors, Ph.D. Dissertation, University of Michigan, 1985.
[9] D.E. Goldberg, A note on Boltzmann tournament selection for genetic algorithms and population oriented simulated annealing, Complex Systems 4 (1970) 445–460.
[10] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989.
[11] D.E. Goldberg, P. Segrest, Finite Markov chain analysis of genetic algorithms, Genetic Algorithms and their Applications: Proc. 2nd Internat. Conf. on Genetic Algorithms, 1987, pp. 1–8.
[12] W. Greub, Linear Algebra, Springer, Berlin, 1975.
[13] J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Michigan, 1975; Extended new edition, MIT Press, Cambridge, MA, 1992.
[14] J. Horn, Finite Markov chain analysis of genetic algorithms with niching, Illinois Genetic Algorithms Laboratory Report No. 93002, Dept. of General Engineering, University of Illinois at Urbana-Champaign, 1993.
[15] D.L. Isaacson, R.W. Madsen, Markov Chains: Theory and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1961.
[16] S. Lang, Linear Algebra, 2nd Ed., Addison-Wesley, Reading, MA, 1970.
[17] S. Lang, Complex Analysis, Addison-Wesley, Reading, MA, 1977.
[18] S.W. Mahfoud, Finite Markov chain models of an alternative selection strategy for genetic algorithms, Complex Systems 7 (1993) 155–170.
[19] M. Mitchell, An Introduction to Genetic Algorithms, MIT Press, Cambridge, MA, 1996.
[20] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, Equations of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1091.
[21] D. Mitra, F. Romeo, A. Sangiovanni-Vincentelli, Convergence and finite time behaviour of simulated annealing, Adv. Appl. Probab. 18 (1986) 747–771.
[22] J. Maynard Smith, Evolutionary Genetics, Oxford University Press, Oxford, 1989.
[23] A.E. Nix, M.D. Vose, Modeling genetic algorithms with Markov chains, Ann. Math. Artif. Intell. 5 (1992) 79–88.
[24] G.K. Pedersen, C^*-algebras and their automorphism groups, London Mathematical Society Monographs No. 14, Academic Press, New York, 1979.
[25] G. Rudolph, Convergence analysis of canonical genetic algorithms, IEEE Trans. on Neural Networks 5 (1994) 96–101.
[26] J. Roughgarden, Theory of Population Genetics and Evolutionary Ecology, Macmillan, New York, 1976; Reprinted by Prentice-Hall, Englewood Cliffs, NJ, 1996.
[27] W. Rudin, Functional Analysis, McGraw-Hill, New York, 1973.
[28] E. Seneta, Non-negative Matrices and Markov Chains, Springer Series in Statistics, Springer, Berlin, 1981.
[29] H.H. Schaefer, Banach Lattices and Positive Operators, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Springer, Berlin, 1974.
[30] J. Suzuki, A further result on the markov chain model of genetic algorithms and its application to a simulated annealing-like strategy, in: R.K. Belew, M.D. Vose (Eds.), Foundations of Genetic Algorithms 4, Morgan Kaufmann, Los Altos, CA, 1997, pp. 53–72.
[31] M.D. Vose, Formalizing genetic algorithms, Proc. IEEE Workshop on Genetic Algorithms, Neural Networks and Simulated Annealing Applied to Problems in Signal and Image Processing, May 1990, Glasgow, UK, 1990.
[32] M.D. Vose, Modeling simple genetic algorithm, in: G. Rawlins (Ed.), Foundations of Genetic Algorithms Morgan Kaufmann, Los Altos, CA, 1991, pp. 94–101.
[33] M.D. Vose, G.E. Liepins, Punctuated equilibria in genetic search, Complex Systems 5 (1991) 31–44.