

Negative Drift in Populations

Per Kristian Lehre

Technical University of Denmark, 2800 Lyngby, Denmark

Abstract. An important step in gaining a better understanding of the stochastic dynamics of evolving populations, is the development of appropriate analytical tools. We present a new drift theorem for populations that allows properties of their long-term behaviour, e.g. the runtime of evolutionary algorithms, to be derived from simple conditions on the one-step behaviour of their variation operators and selection mechanisms.

1 Introduction

Drift analysis is one of the primary mathematical techniques used to estimate the runtime of evolutionary algorithms (EAs) and other randomised search heuristics. Based on Hajek [4], drift analysis was introduced to evolutionary computation by He and Yao [5]. The dynamics of the EA on a fitness function is aggregated into a real-valued stochastic process X_0, X_1, \dots , by mapping each element of the state space Ω of the EA to a real value using a potential function $g : \Omega \rightarrow \mathbb{R}$. If the potential function is appropriately defined, drift theorems allow properties about the long-term behaviour of the EA to be deduced from conditions on the mean one-step drift of the process, defined as $\mathbf{E}[\Delta_t]$, where $\Delta_t := X_{t+1} - X_t$. While this approach has proven effective to analyse search heuristics that maintain only a single search point, e.g. simulated annealing [10], and the (1+1) EA [2], there are few applications of drift analysis on search heuristics that maintain several search points, e.g. population-based EAs. Even for simple fitness functions, highly non-trivial potential functions seem necessary to aggregate the state of the population [9]. Other approaches may be needed to analyse population-based search heuristics in more complex scenarios.

We introduce a new drift theorem particularly aimed at analysing population-based search heuristics. An essential feature of the theorem is that the effects of the variation operator and the effects of the selection mechanism are decoupled in the conditions of the theorem, thus alleviating the state aggregation problem. The result applies to any population-based process of the type described in Algorithm 1 below. The algorithm keeps a vector $P_{t \geq 0} \in \Omega^\lambda$ of λ search points. The vector will be referred to as a *population*, and its elements as *individuals*. Given a population P_t , the next population P_{t+1} is generated by sampling and perturbing, according to the variation operator p_{mut} , λ individuals in P_t . An iteration of the of outer loop is called a *generation*. Variation operators are represented as transition matrices $p_{\text{mut}} : \Omega \times \Omega \rightarrow [0, 1]$, where $p_{\text{mut}}(x, y)$ represents the probability of perturbing an individual x into an individual y . For a given transition matrix p_{mut} , we associate a Markov process $X_{t \geq 0}$, where for all $t > 0$, the

1 Population Selection-Variation Algorithm

Require: Finite state space Ω , transition matrix p_{mut} , and $P_0 \in \Omega^\lambda$.

- 1: **for** $t = 0, 1, 2, \dots$ until termination condition **do**
 - 2: **for** $i = 1$ to λ **do**
 - 3: Choose a parent index $I_t(i) \in \{1, \dots, \lambda\}$, and set $x := P_t(I_t(i))$.
 - 4: Sample x' according to $p_{\text{mut}}(x)$, and set $P_{t+1}(i) := x'$.
 - 5: **end for**
 - 6: **end for**
-

state transition probability is given by $\Pr[X_{t+1} = x' \mid X_t = x] := p_{\text{mut}}(x, x')$, i.e. $X_{t \geq 0}$ corresponds to a random walk of a single individual in Ω according the variation operator. The selection of individuals are specified by the vector $I_t \in \{1, 2, \dots, \lambda\}^\lambda$ of indices, where the i -th element $I_t(i)$ represents the “parent” of the i -th individual. The selection mechanism is unspecified, but will typically depend on an objective function on Ω . The sequence of index vectors $I_{t \geq 0}$ is associated with a stochastic process $R_{t \geq 0}$ on $\{0, 1, \dots, \lambda\}^\lambda$, defined as $R_t(i) := \sum_{j=1}^\lambda [I_t(j) = i]$ for all $i, 1 \leq i \leq \lambda$. The i -th element $R_t(i)$ represents the number of times the individual with index i was selected during generation t . The expectation $\mathbf{E}[R_t(i)]$ is called the *reproductive rate* of the i -th individual in generation t .

2 Negative Drift Theorem For Populations

Theorem 1. *Given Alg. 1 with positive transition matrix p_{mut} over a finite state space Ω , and a function $g : \Omega \rightarrow \mathbb{N}^+$. Pick two positive integers $a(n)$ and $b(n)$ such that $d(n) := b(n) - a(n) \geq 0$. Let $T(n)$ denote the earliest point in time $t \geq 0$ such that $g(P_t(j)) \leq a(n)$ holds for some $j, 1 \leq j \leq \lambda$. If there are $D(n) \geq 1$, and $\kappa(n) > 0$, and constants $\delta, \delta_2, \delta_3 > 0, \alpha_0 \geq 1$, such that for all $t \geq 0$ and integers i, j, k , and l where $a(n) \leq i \leq b(n)$ and $1 \leq l + k \leq j$, it holds*

1. $\mathbf{E}[R_t(i) \mid a(n) < g(P_t(i)) < b(n)] \leq \alpha_0$ for all $i, 1 \leq i \leq \lambda$,
2. $\mathbf{E}\left[e^{-\kappa(n)\Delta_t(i)} \mid a(n) < g(X_t) < b(n)\right] < 1/(\alpha_0(1 + \delta))$
3. $\mathbf{E}\left[e^{-\kappa(n)(g(X_{t+1}) - b(n))} \mid g(X_t) \geq b(n)\right] < D(n)$
4. $\Pr[\Delta_t(i) = -l \wedge \Delta_{t+1}(i - l) = -k] \leq e^{\kappa(n)d(n)(1 - \delta_2)} \Pr[\Delta_t(i) = -l - k]$
5. $\Pr[\Delta_t(i) = -j] \leq \Pr[\Delta_t(i - k) = -l] \cdot \delta_3$

where

- $R_t(i) := \sum_{j=1}^\lambda [I_t(j) = i]$,
- $X_{t \geq 0}$ is the Markov process on Ω associated with p_{mut} , and
- $\Delta_t(i) := (g(X_{t+1}) - g(X_t) \mid g(X_t) = i)$,

then for all time bounds $L(n) > 0$,

$$\Pr[T(n) \leq L(n) \mid g(P_0) \geq b(n)] = O\left(\lambda L(n)^2 D(n) d(n) e^{-\kappa(n)d(n)\delta_2}\right).$$

The result can be described informally as follows. The theorem assumes a potential function g over a search space Ω , and a goal potential $a(n) \geq 0$. Any search point $x \in \Omega$ with potential $g(x) \leq a(n)$ will be called a *solution*. If the algorithm satisfies the five conditions, then the theorem provides an upper bound on the probability that a solution has been found within a chosen number $L(n)$ of generations. The conditions are w.r.t. the reproductive rate α_0 and the random walk X_t . The first two conditions mean that if the potential of the random walk is close to $a(n)$, then the random walk should have a negative drift towards higher potential values. The requirement on the negative drift is proportional to the reproductive rate. The larger the reproductive rate, the larger negative drift is required for the theorem to hold. The third condition is a milder requirement on the negative drift when the random walk is far from the goal potential. The fourth condition limits the advantage of reducing the potential by a given value during two, instead of one step. The last condition states that the probability of reducing the potential by much, should not be much larger than reducing the potential a little. Note that the drift conditions 2-5 are w.r.t. a random walk of a single individual, and not w.r.t. the population. Hence, these conditions can be verified independently of the selection mechanism and fitness function. For economy of use, it is therefore helpful to derive special versions of the theorem for specific settings of Ω, p_{mut} and g , as will be illustrated in Section 3.

We now explain the proof idea. We focus on the event that an individual x reaches a g -value below $b(n)$, and aim to show that within $L(n)$ generations, all of its ancestors either become extinct, or have drifted back to g -values above $b(n)$. The ancestors will be modelled as a *non-selective family tree*, a concept introduced in Lehre and Yao [7]. The nodes in the family tree correspond to the ancestors of x , where node x is the root node. A path in the family tree is called a *lineage*. We will prune the tree, and only consider the part of the tree that corresponds to individuals with g -values below $b(n)$, i.e. any subtree that is rooted in an individual with g -value above $b(n)$ is removed. The number of times an individual with index i is selected, is given by the random variable $R_i(t)$. Different individuals have different offspring distributions. To simplify the analysis, we consider the family tree as if it had been subject to a modified selection process. Here, the number of times each individual is selected is distributed according to $R_{i^*}(t)$, where i^* is the index of the individual with highest reproductive rate. Thus, each individual will be selected as often as it would have been, had it been the individual with highest reproductive rate. Assuming condition 1, each member of the family tree will on average be selected α_0 times. A consequence of modifying the selection process is that the family tree grows quicker than the real family tree. And as there is no selective differences, each lineage corresponds to a random walk.

The proof consist of two parts. The first part provides an upper bound on the number of different lineages in the family tree. The second part provides an upper bound on the probability that a given lineage of length $L(n)$ will reach a solution. The final result is obtained by combining these two parts using a union bound. We start with the second part, and apply Hajek's drift theorem [4].

Theorem 2 (Hajek [4]). Let X_0, X_1, \dots be the random variables describing a Markov process over the state space Ω , and $g : \Omega \rightarrow \mathbb{R}_0^+$ a function that assigns to each state a non-negative real number. Pick two real numbers $a(n)$ and $b(n)$ which depend on a parameter n such that $0 \leq a(n) < b(n)$ holds. Let the random variable $T(n)$ denote the earliest point in time $t \geq 0$ that satisfies $g(X_t) \leq a(n)$. If there are $\kappa(n) > 0$ and $p(n) > 0$ such that for all $t \geq 0$, the condition

$$\mathbf{E} \left[e^{-\kappa(n)(g(X_{t+1})-g(X_t))} \mid a(n) < g(X_t) < b(n) \right] \leq 1 - 1/p(n)$$

holds, then for all time bounds $L(n) \geq 0$

$$\Pr [T(n) \leq L(n) \mid g(X_0) \geq b(n)] \leq e^{\kappa(n)(a(n)-b(n))} \cdot L(n) \cdot D(n) \cdot p(n),$$

where $D(n) := \max \{1, \mathbf{E} [e^{-\kappa(n)(g(X_{t+1})-b(n))} \mid b(n) \leq g(X_t)]\}$.

This theorem will be applied later in the proof of Theorem 1. However, note that if the conditions in Theorem 1 hold, then the conditions in Theorem 2 hold for the Markov process associated with the transition matrix p_{mut} .

Knowing that a single lineage will not find the optimum within polynomial time, we now estimate the number of different lineages in the family tree. The number of lineages is trivially bounded by the number of family tree members, which can be analysed using *multi-type branching processes*.

Definition 1 (Multi-Type Branching Process [3]). A multi-type branching process with d types is a Markov process Z_0, Z_1, Z_2, \dots which for all $t \geq 0$, is given by $Z_{t+1} := \sum_{j=1}^d \sum_{i=1}^{Z_{tj}} \xi_i^{(j)}$, where for all j , $\xi_i^{(j)} \in \mathbb{N}_0^d$ are i.i.d. random vectors having expectation $\mathbf{E} [\xi^{(j)}] =: (m_{j1}, m_{j2}, \dots, m_{jd})^\top$. The associated matrix $M := (m_{hj})_{d \times d}$ is called the mean matrix of the process.

A multi-type branching process can be thought of as a population of individuals of d types. The vector component Z_{tj} represents the number of individuals of type j in generation t . An individual survives one generation, during which it may produce some offspring. The offspring produced by an individual depends on its type j , and is given by an independent random vector $\xi^{(j)}$, where the vector component $\xi_i^{(j)}$ is the number of offspring of type i . Each entry m_{hj} in the mean matrix represents the expected number of offspring a type h -individual will have of type j -individuals. The expectation of a multi-type branching process can be calculated from its mean matrix by $\mathbf{E} [Z_t]^\top = \mathbf{E} [\mathbf{E} [Z_t \mid Z_{t-1}]]^\top = \mathbf{E} [Z_{t-1}]^\top M = \dots = \mathbf{E} [Z_0]^\top M^t$. Matrix powers M^t of irreducible matrices can be determined using the Perron-Frobenius theorem, where irreducibility is defined as follows.

Definition 2 (Irreducible matrix [11]). A $d \times d$ non-negative matrix M is irreducible if for every pair i, j of its index set, there exists a positive integer t such that $m_{ij}^{(t)} > 0$, where $m_{ij}^{(t)}$ are the elements of the t -th matrix power M^t .

Note that positive matrices are irreducible. The following statement of the Perron-Frobenius theorem is taken from [3].

Theorem 3 (Perron-Frobenius). *If M is an irreducible matrix with non-negative elements, then it has a unique positive eigenvalue ρ , called the Perron root of M , that is greater in absolute value than any other eigenvalue. All elements of the left and right eigenvectors $u = (u_1, \dots, u_d)^\top$ and $v = (v_1, \dots, v_d)^\top$ that correspond to ρ can be chosen positive and such that $\sum_{k=1}^d u_k = 1$ and $\sum_{k=1}^d u_k v_k = 1$. Also, $M^n = \rho^n A + B^n$, where $A = (v_i u_j)_{i,j=1}^d$ and B are matrices where $AB = BA = 0$, and there are constants $\rho_1 \in (0, \rho)$ and $C > 0$ such that none of the elements of the matrix B^n exceeds $C\rho_1^n$.*

Hence, the asymptotics of the matrix power M^t depends primarily on the Perron root. This can be used to bound the expected number of descendants from a single individual of type h .

Lemma 1. *Let Z_0, Z_1, \dots be a multi-type branching process with irreducible mean matrix $M = (m_{ij})_{d \times d}$ and Perron root $\rho < 1$ with corresponding right eigenvector v . The number of descendants L_t of the initial individual after $t > 0$ generations satisfies $\mathbf{E}[L_t | Z_0 = e_h] \leq \frac{\rho}{1-\rho} \cdot \frac{v_h}{v^*}$, where $e_h, 1 \leq h \leq d$, denote the standard basis vectors, and $v^* := \min_{1 \leq i \leq d} v_i$.*

Proof. The proof follows [3, p. 122]. By Theorem 3, matrix M has a unique largest eigenvalue ρ , and all the elements of the corresponding right eigenvector v are positive, implying $v^* > 0$. By using that $v_j \geq v^*$ for all j , we get $\mathbf{E}[L_t | Z_0 = e_h] \leq \frac{1}{v^*} \sum_{r=1}^t \sum_{j=1}^d \mathbf{E}[Z_{rj} v_j | Z_0 = e_h]$. As seen above, the expectation on the right hand side can be expressed as $\mathbf{E}[Z_r | Z_0 = e_h]^\top = \mathbf{E}[Z_0 | Z_0 = e_h]^\top M^r$. Additionally, by taking into account the starting conditions, $Z_{0h} = 1$ and $Z_{0j} = 0$, for all indices $j \neq h$, this simplifies further to $\frac{1}{v^*} \sum_{r=1}^t \sum_{j=1}^d \sum_{i=1}^d \mathbf{E}[Z_{0i} v_j m_{ij}^{(r)} | Z_0 = e_h] = \frac{1}{v^*} \sum_{r=1}^t \sum_{j=1}^d v_j m_{hj}^{(r)}$. Finally, by iterating $M^r v = M^{r-1}(Mv) = \rho M^{r-1}v$, which on coordinate form gives $\sum_{j=1}^d v_j m_{hj}^{(r)} = \rho^r v_h$, one obtains the final bound $\frac{v_h}{v^*} \sum_{r=1}^t \rho^r \leq \frac{\rho}{1-\rho} \cdot \frac{v_h}{v^*}$. \square

We formalise the non-selective family tree as a multi-type branching process. Each family tree member corresponds to an individual in the branching process. We have $d(n) := b(n) - a(n)$ types, and the type of an individual x is given by $g(x) - a(n)$. Each family tree member is selected in expectation α_0 times per generation, so an individual of type i , will in expectation have $\alpha_0 p_{ij}$ offspring of type j , where p_{ij} is the probability that the the variation operator produces an offspring of type j from an individual of type i .

Definition 3 (Mean Matrix of EA). *Given Algorithm 1 with reproductive rate α_0 , and two integers $0 \leq a(n) < b(n), d(n) := b(n) - a(n)$. The associated $d(n) \times d(n)$ mean matrix is defined as $m_{ij} := \alpha_0(p_{ij} + 1/d(n)^2)$ if $j > i$, and $m_{ij} := \alpha_0 p_{ij}$ if $j \leq i$, where $p_{ij} := \Pr[\Delta_t(i + a(n)) = j - i]$.*

The extra term $\alpha_0/d(n)^2$ in the mean matrix is added for technical reasons, and will lead to overestimation of the survival probability. If the Markov chain p_{mut} limited to the set of states x in Ω where $g(x) \leq b(n)$, is irreducible, then the mean matrix is irreducible. To apply Lemma 1, we consider next the Perron root and the associated eigenvector.

Lemma 2. *If Algorithm 1 satisfies the conditions in Theorem 1, then the associated mean matrix M has Perron root $\rho(M) \leq 1/(1 + \delta)$ for a constant $\delta > 0$.*

Proof. The Frobenius bound for the Perron root of a matrix M states that $\rho(M) \leq \max_i \sum_j m_{ij}$ [6]. However, when applied directly to our matrix, this bound is insufficient for our purposes. Instead, we can consider the transformation $SM S^{-1}$, for the invertible matrix $S := \text{diag}(e^\kappa, e^{2\kappa}, \dots, e^{d(n)\kappa})$. Note that $\det(SAS^{-1}) = \det(S)$ for any $d(n) \times d(n)$ matrix A . So if ρ is an eigenvalue of M , then $0 = \det(M - \rho I) = \det(S(M - \rho I)S^{-1}) = \det(SMS^{-1} - \rho I)$, and ρ must also be an eigenvalue of $SM S^{-1}$. It follows that $\rho(M) = \rho(SM S^{-1})$. By using the Frobenius bound along the rows of matrix $SM S^{-1}$, which has elements $(SM S^{-1})_{ij} = m_{ij}e^{-\kappa(j-i)}$, we can bound $\rho(M)$ for large $d(n)$ by

$$\begin{aligned} \rho(SM S^{-1}) &\leq \max_{1 \leq i \leq d(n)} \alpha_0 \sum_{j=1}^{d(n)} p_{ij} e^{-\kappa(j-i)} + \sum_{j=i+1}^{d(n)} \frac{\alpha_0}{e^{\kappa(j-i)} d(n)^2} \\ &\leq \max_{1 \leq i \leq d(n)} \alpha_0 \sum_{j=-\infty}^{\infty} \Pr[\Delta_t(i + a(n)) = j - i] e^{-\kappa(j-i)} + \frac{\alpha_0}{d(n)} \\ &= \max_{1 \leq i \leq d(n)} \alpha_0 \mathbf{E} \left[e^{-\kappa \Delta_t(i + a(n))} \right] + \frac{\alpha_0}{d(n)} \leq \frac{1}{1 + \delta/2}. \end{aligned}$$

□

Lemma 3. *Let M be the mean matrix associated with Algorithm 1, v the right eigenvector corresponding to the Perron root of M , and $v_* := \min_i v_i$ the minimal component of this eigenvector. If the conditions of Theorem 1 are satisfied, then it holds for all indices $h, 1 \leq h \leq d(n)$, that $p_{d(n)h} \frac{v_h}{v_*} \leq \delta_3 e^{\kappa d(n)(1-\delta_2)}$.*

Proof. Minc's bound for the principal ratio of positive matrices [8] can be generalised as follows, $\frac{v_h}{v_*} = \max_k \frac{v_h}{v_k} = \max_k \frac{\rho v_h}{\rho v_k} = \max_k \frac{\sum_j m_{hj} v_j}{\sum_j m_{kj} v_j} \leq \max_{k,j} \frac{m_{hj}}{m_{kj}}$. It now suffices to bound the ratio $p_{d(n)h} m_{hj} / m_{kj}$ for all values of h, j and k . In the case where $k \leq j$, we have $p_{d(n)h} m_{hj} / m_{kj} \leq m_{hj} d(n)^2 \leq 2d(n)^2$. In the case where $h \leq j < k$, condition 5 implies that $p_{d(n)h} \leq \delta_3 p_{kj}$, so $p_{d(n)h} m_{hj} / m_{kj} \leq m_{hj} \delta_3 \leq 2\delta_3$. Finally, when $j < k$ and $j < h$, condition 4 and 5 imply that

$$\frac{p_{d(n)h} m_{hj}}{m_{kj}} = \frac{p_{d(n)h} p_{hj}}{p_{kj}} \leq \frac{\delta_3 p_{d(n)h} p_{hj}}{p_{d(n)j}} \leq \delta_3 e^{\kappa(n)d(n)(1-\delta_2)} \quad \square$$

Proof (of Theorem 1). Consider the event that an individual x with g -value higher than $b(n)$ obtains an offspring with g -value less than $b(n)$. We model the non-selective family tree corresponding to individual x pruned to g -values in the interval $b(n)$ to $a(n)$, as a multi-type branching process with $d(n)$ types and mean matrix given by Definition 3. Any search point with g -value less than $a(n)$ is called a *solution*. To bound the probability q that a given lineage of length at most $L(n)$ reaches a solution, we apply Theorem 2 with the parameter $p(n) := \alpha_0(1 + \delta)/(\alpha_0(1 + \delta) - 1)$, yielding $q = O(L(n)D(n)e^{-\kappa(n)d(n)})$. If

there are k lineages in the family tree, then by union bound, the probability that any lineage reaches a solution is less than kq . The number of lineages is less than the number L of family tree members. Hence, the probability that the family tree contains a solution is less than $\sum_{k=1}^{\infty} \Pr[L = k] kq = q\mathbf{E}[L] \leq q \sum_{h=1}^{d(n)} p_{d(n)h} \mathbf{E}[L | Z_0 = e_h]$. By applying Lemmas 1, 2, and 3, this is no more than $q \sum_{h=1}^{d(n)} \frac{\rho}{1-\rho} \frac{p_{d(n)h} v_h}{v^*} = O(qd(n)e^{\kappa(n)d(n)(1-\delta_2)})$. Finally, by noting that there can be no more than $\lambda L(n)$ such family trees within $L(n)$ generations, the probability that any individual within the first $L(n)$ generations is a solution, is by union bound $O(\lambda L(n)^2 D(n) d(n) e^{-\kappa(n)d(n)\delta_2})$. \square

3 Applications to Evolutionary Algorithms

We now derive a simplified variant of Theorem 1 tailored to runtime analysis of EAs on Pseudo-boolean functions. We consider bitwise mutation $p_{\text{bit}}(x, y) = \left(\frac{\chi}{n}\right)^{H(x,y)} \left(1 - \frac{\chi}{n}\right)^{n-H(x,y)}$, where H denotes Hamming-distance, and the constant parameter $\chi > 0$ determines the mutation rate. The potential function g is defined as the Hamming distance to some search point $x^* \in \{0, 1\}^n$.

Theorem 4. *Given Algorithm 1 on $\Omega = \{0, 1\}^n$ with transition matrix p_{bit} , mutation rate χ , and population size $\lambda = \text{poly}(n)$. Let $a(n)$ and $b(n)$ be positive integers s.t. $b(n) \leq n/\chi$ and $d(n) := b(n) - a(n) = \omega(\ln n)$. For an $x^* \in \{0, 1\}^n$, let $T(n)$ be the smallest $t \geq 0$, s.t. $H(P_t(j), x^*) \leq a(n)$ for some $j, 1 \leq j \leq \lambda$. Let $R_t(i) := \sum_{j=1}^{\lambda} [I_t(j) = i]$. If there are constants $\alpha_0 \geq 1$ and $\delta > 0$ s.t.*

1. $\mathbf{E}[R_t(i) | a(n) < H(P_t(i), x^*) < b(n)] \leq \alpha_0$, for all $i, 1 \leq i \leq \lambda$,
2. $\psi := \ln(\alpha_0)/\chi + \delta < 1$, and
3. $\frac{b(n)}{n} < \min \left\{ \frac{1}{5}, \frac{1}{2} - \frac{1}{2} \sqrt{\psi(2-\psi)} \right\}$,

then there exists a constant $c > 0$ such that $\Pr[T(n) \leq e^{cd(n)}] \leq e^{-\Omega(d(n))}$.

Proof. We apply Theorem 1 over the interval $[a(n), b(n)]$, where the distance function is defined as $g(x) := H(x, x^*)$. W.l.o.g., we assume that $x^* = 1^n$. The first condition holds immediately. For the second and third conditions, note that $\mathbf{E}[e^{-\kappa \Delta_t(i)}] = M_{\Delta_t(i)}(-\kappa)$, where $M_{\Delta_t(i)}$ is the moment-generating function (m.g.f.) for the drift $\Delta_t(i) := (g(X_{t+1}) - g(X_t) | g(X_t) = i)$. The drift can be expressed as the sum of two, independent random variables $\Delta_t(i) = \Delta_t^+(i) - \Delta_t^-(i)$, where $\Delta_t^+(i)$ is the number of 1-bits that are flipped into 0-bits, and $\Delta_t^-(i)$ is the number of 0-bits that are flipped into 1-bits. These variables are binomially distributed. A binomially distributed random variable X with parameters n and p has m.g.f. $M_X(t) = (1 - p + pe^t)^n$. Setting $\kappa = \ln(2 + \varepsilon + \delta_2)$, where ε and δ_2 are constants that will be determined later, the positive drift component is

$$M_{\Delta_t^+(i)}(-\kappa) = \left(1 - \frac{\chi}{n} + \frac{\chi}{n(2 + \varepsilon + \delta_2)}\right)^{n-i} \leq \exp\left(-\frac{(1 - \frac{i}{n})(1 + \varepsilon + \delta_2)\chi}{(2 + \varepsilon + \delta_2)}\right),$$

and the negative drift component is

$$M_{\Delta_t^-(i)}(\kappa) = \left(1 - \frac{\chi}{n} + \frac{\chi(2 + \varepsilon + \delta_2)}{n}\right)^i \leq \exp\left(\frac{i(1 + \varepsilon + \delta_2)\chi}{n}\right).$$

The random variables $\Delta_t^+(i)$ and $\Delta_t^-(i)$ are independent, so the m.g.f. of the drift is given by the product $M_{\Delta_t(i)}(t) = M_{\Delta_t^+(i)}(t)M_{\Delta_t^-(i)}(-t)$. Hence, by taking into account that $i < b(n)$, we get

$$M_{\Delta_t(i)}(t) \leq \exp\left(\frac{(1 + \varepsilon + \delta_2)\chi}{2 + \varepsilon + \delta_2} \left(\frac{b(n)(3 + \varepsilon + \delta_2)}{n} - 1\right)\right).$$

This bound holds for all $\varepsilon > 0$. We would like to maximise $b(n)$ with respect to ε , subject to condition 2 of Theorem 1, i.e. for the constant $\delta' := e^{\delta\chi} - 1$, we would like to maximise the right hand side of the inequality

$$b(n) \leq \frac{n}{3 + \varepsilon + \delta_2} \left(1 - \frac{(2 + \varepsilon + \delta_2) \ln(\alpha_0(1 + \delta'))}{(1 + \varepsilon + \delta_2)\chi}\right).$$

We choose $\varepsilon = \frac{\sqrt{\psi(2-\psi)+(2+\delta_2)\psi-1-\delta_2}}{1-\psi}$. By adjusting parameter δ , one can ensure that $1/5 < \psi < 1$, and hence that $\varepsilon > 0$ for an appropriate choice of δ_2 . This choice of ε gives $\frac{b(n)}{n} < \min\{\frac{1}{5}, \frac{1}{2} - \frac{1}{2}\sqrt{\psi(2-\psi)}\}$. The second condition of Theorem 1 is therefore satisfied for the parameter δ' . For the *third condition*, it is sufficient to use the upper bound $\mathbf{E}[e^{-\kappa(n)(g(X_{t+1})-b(n))} | g(X_t) > b(n)] \leq \mathbf{E}[e^{-\kappa\Delta_t(i)} | g(X_t) > b(n)]$ and observe that the upper bound on the m.g.f. of the drift is bounded from above by a constant $D(n) = O(1)$ for all i . For the *fourth condition*, it holds for all h, j, k where $1 \leq k < j < h \leq d(n)$ that the ratio $p_{hj}p_{jk}/p_{hk}$ is no more than

$$\frac{\binom{h+a(n)}{h-j} \left(\frac{\chi}{n}\right)^{h-j} \binom{j+a(n)}{j-k} \left(\frac{\chi}{n}\right)^{j-k}}{\binom{h+a(n)}{h-k} \left(\frac{\chi}{n}\right)^{h-k} \left(1 - \frac{\chi}{n}\right)^{n-h+k}} \leq e^\chi \binom{h-k}{h-j} \leq e^\chi 2^{h-k} \leq e^{\kappa(n)(1-\delta'_2)d(n)},$$

where the final inequality holds for some constant δ'_2 because $h-k \leq d(n)$, and $\kappa(n) > \ln 2$. Thus the fourth condition is satisfied. Finally, for the *fifth condition*, it holds for all integers i, j, k and l where $1 \leq l \leq k \leq j < i \leq d(n)$ that

$$\begin{aligned} \frac{p_{il}}{p_{jk}} &\leq \frac{\binom{i+a(n)}{i-l} \left(\frac{\chi}{n}\right)^{i-l}}{\binom{j+a(n)}{j-k} \left(\frac{\chi}{n}\right)^{j-k} \left(1 - \frac{\chi}{n}\right)^{n-(j-k)}} \leq \frac{\binom{i+a(n)}{i-j} \binom{j+a(n)}{j-k} \binom{k+a(n)}{k-l} \left(\frac{\chi}{n}\right)^{i-l}}{\binom{j+a(n)}{j-k} \left(\frac{\chi}{n}\right)^{j-k} e^{-\chi}} \\ &= \binom{i+a(n)}{i-j} \left(\frac{\chi}{n}\right)^{i-j} \binom{k+a(n)}{k-l} \left(\frac{\chi}{n}\right)^{k-l} e^\chi \leq e^\chi \end{aligned}$$

where the last inequality holds because $\binom{m}{k} \left(\frac{\chi}{n}\right)^k \leq \left(\frac{m\chi}{n}\right)^k \leq 1$ for any $m \leq n/\chi$ and $k \leq m$. Because of the conditions $d(n) = \omega(\ln n)$ and $\lambda = \text{poly}(n)$, we have $\lambda D(n)d(n) \leq e^{c'd(n)}$ for any constant $c' > 0$, when n is sufficiently large. By setting $L(n) = e^{cd(n)}$ and choosing c and c' sufficiently small, the theorem now follows from Theorem 1. \square

| Selection mechanism | Parameter settings |
|-----------------------------|---------------------|
| Linear ranking selection | $\eta < e^x$ |
| k -tournament selection | $k < e^x$ |
| (μ, λ) -selection | $\lambda < \mu e^x$ |
| Any in cellular EAs | $\Delta(G) < e^x$ |

Table 1. Parameter settings where non-elitist EAs with bit-wise mutation rate χ/n are ineffective (cf. Corollary 1).

Finally, we give some example applications. We first consider JUMP_m , which for any $m, 1 \leq m < n$, is defined as $\text{JUMP}_m(x) = |x|_1$ if $|x|_1 < n - m$ or $|x|_1 = n$, and $\text{JUMP}_m(x) = 0$ otherwise, where $|x|_1 := \sum_{i=1}^n x_i$. JUMP_m has a local optimum separated from the global optimum by a gap of size m . The EA must either jump the gap, or make a random walk in the gap. We consider tournament selection, but the theorem can easily be generalised to other selection mechanisms.

Theorem 5. *For any $m < n(1 - \varepsilon)/5, 0 < \varepsilon < 1$, with $m = \omega(\ln n)$, the probability that a non-elitist EA using k -tournament selection, $k \geq 1$, population size $\lambda = \text{poly}(n)$, and bitwise mutation rate $\chi/n \leq 1/m$, optimises JUMP_m within e^{cm} generations is $e^{-\Omega(m)}$, for some constant $c > 0$.*

Proof. We apply Theorem 4, with the parameters $a(n) = 0$, $b(n) = m$, $\alpha_0 = 1$, and optimum $x^* = 1^n$. Individuals with no more than m 0-bits, which we call *gap-individuals*, never win tournaments containing individuals with less than m 0-bits. Optimistically assuming that all individuals are gap-individuals, the individuals are selected uniformly at random. Hence, the first condition holds. We have $\psi = \delta$, so the second condition holds for any $\delta, 0 < \delta < 1$. By assumption, $n/5 > m$, and choosing δ sufficiently small also gives $(n/2)(1 - \sqrt{\psi(2 - \psi)}) > m$, so the third condition also holds. \square

The function $\text{NEEDLE}(x) := \prod_{i=1}^n x_i$ can be analysed analogously to JUMP_m .

Theorem 6. *The probability that a non-elitist EA using k -tournament selection, $k \geq 1$, population size $\lambda = \text{poly}(n)$, and bitwise mutation, optimises NEEDLE within e^{cn} generations is $e^{-\Omega(n)}$, for some constant $c > 0$.*

We now consider scenarios where the selective pressure is too low.

Corollary 1. *The probability that a non-elitist EA with population size $\lambda = \text{poly}(n)$, bitwise mutation rate χ/n , and maximal reproductive rate bounded by $\alpha_0 < e^x - \delta$, for a constant $\delta > 0$, optimises any function with a polynomial number of optima within e^{cn} generations is $e^{-\Omega(n)}$, for some constant $c > 0$.*

Proof. By Theorem 4, choosing $a(n) = 0$ and $b(n) = c'n$, $c' > 0$ sufficiently small, a given optimum is found in e^{cn} generations with probability $e^{-\Omega(n)}$. By union bound, the probability that any of $r = \text{poly}(n)$ optima is found within e^{cn} generations is $re^{-\Omega(n)} = e^{-\Omega(n)}$. \square

A larger reproductive rate than e^χ is therefore necessary for a non-elitist EA to be effective. Table 1 summarises parameter settings in some common selection mechanisms that render non-selective EAs ineffective. The parameter $\eta \in [1, 2]$ in linear ranking selection is directly related to the reproductive rate by $\alpha_0 = \eta$. In k -tournament selection, and in (μ, λ) -selection, the reproductive rate is bounded by $\alpha_0 \leq k$, respectively $\alpha_0 < \lambda/\mu$. The reproductive rate in cellular EAs [1] is bounded by $\alpha_0 \leq \Delta(G)$, i.e. the degree of the neighbourhood graph G . E.g., non-elitist cEAs with $\chi = 1$ on the ring graph are ineffective.

4 Conclusion

A new drift theorem for analysis of non-elitist populations has been introduced. The conditions of the theorem decouple the effects of the selection mechanism and the variation operator — they are w.r.t. a random walk of a single individual, and not the population as a whole — thus simplifying the analysis of EAs greatly. The proof of the theorem combines results about multi-type branching processes with drift analysis. A special case of the theorem for Pseudo-boolean functions is derived, and applied on several selection mechanisms and example functions.

Acknowledgements This work was supported by Deutsche Forschungsgemeinschaft (DFG) grant no. WI 3552/1-1, and EPSRC grant no. EP/D052785/1.

References

1. E. Cantú-Paz. A survey of parallel genetic algorithms. *Calculateurs Paralleles, Reseaux et Systems Repartis*, 10(2):141–171, 1998.
2. O. Giel and I. Wegener. Evolutionary algorithms and the maximum matching problem. In *Proc. of the 20th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2003)*, pages 415–426, 2003.
3. P. Haccou, P. Jagers, and V. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press, 2005.
4. B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14(3):502–525, 1982.
5. J. He and X. Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.
6. L. Y. Kolotilina. Bounds and inequalities for the Perron root of a nonnegative matrix. *Journal of Mathematical Sciences*, 121(4):2481–2507, November 2004.
7. P. K. Lehre and X. Yao. On the impact of the mutation-selection balance on the runtime of evolutionary algorithms. In *Proc. of FOGA 2009*, pages 47–58, New York, NY, USA, 2009. ACM.
8. H. Minc. On the maximal eigenvector of a positive matrix. *SIAM Journal on Numerical Analysis*, 7(3):424–427, 1970.
9. F. Neumann, P. S. Oliveto, and C. Witt. Theoretical analysis of fitness-proportional selection: landscapes and efficiency. In *Proc. of GECCO 2009*, pages 835–842, New York, NY, USA, 2009. ACM.
10. G. H. Sasaki and B. Hajek. The time complexity of maximum matching by simulated annealing. *Journal of the ACM*, 35(2):387–403, 1988.
11. E. Seneta. *Non-Negative Matrices*. George Allen & Unwin Ltd., London, 1973.