Locality vs. randomness – dependence of operator quality on the search state

Karsten Weicker*
Wilhelm–Schickard–Institute
Dept. Computer Architecture
University of Tübingen
Köstlinstr. 6
72074 Tübingen, Germany

Nicole Weicker†
Institute of Computer Science
Dept. Formal Concepts
University of Stuttgart
Breitwiesenstr. 20–22
70565 Stuttgart, Germany

Abstract

By analyzing mutation operators' behavior through their effect on the fitness only, fundamental statements on the interplay of the operator's quality and the state of the search process are enabled for arbitrary optimization problems with arbitrary discrete search spaces. First, it is observed that the defined operator quality has a direct effect on hill climbing search. Moreover, an operator's locality property influences its quality depending on the fitness level. This leads to the major result that a local operator cannot be valued without considering the search state.

1 Introduction

The "No Free Lunch" (NFL) theorem (Wolpert & Macready, 1997) shows that there is no optimal algorithm independently of the considered optimization problem. Therefore, it is necessary to use problem knowledge in the algorithm, e.g. by assessing and selecting operators for the considered problem. Already various measures for judging the suitability of mutation operators have been discussed in literature. This work applies such measures in order to prove that under certain prerequisites the choice of operators depends on the state of the search.

Generally, there are various theoretical approaches to judge operators (cf. Beyer & Rudolph, 1997): global convergence examinations (e.g. Rudolph, 1997) and local performance mea-

*E-mail: weicker@informatik.uni-tuebingen.de
†E-mail: weicker@informatik.uni-stuttgart.de

sures. The latter may be used for valuating the suitability of the operator to the problem. But those results are mostly stated for very simple, abstract model fitness landscapes. Beyer (1994) introduced two local performance measures: *quality gain*, capturing the fitness change from one generation of an evolutionary algorithm to the next, and *progress rate*, measuring the local progression with the "natural" distance to a reference point—according to a distance measure defined on the problem structure only.

Measures like the progress rate are problematic since the "natural" neighboring structure may be uncorrelated with the problem, i.e. small changes in an individual may lead to big fitness changes. This is faced by the inspection of correlation coefficients, which indicate how appropriate an operator is adapted to the "natural" neighboring structure (e.g. Manderick, Weger, & Spiessens, 1991; Hordijk, 1996). This method considers the concrete problem directly. Unfortunately, the "natural" neighboring structure and the structure of the problem are often not interrelated as Jones and Forrest (1995) have shown by examinations of fitness distance correlations. Alternatively, Jones (1995) proposes an operator–given neighborhood relative to an operator as reference basis. This approach was not pursued consequently, since the operator–given neighborhood cannot be computed globally but provides selective insight only.

This work combines a variant of the quality gain with the operator–given neighborhood. In order to do so, an operator defines the operator–given fitness landscape by its possible steps in the search space (Section 2), like it was proposed by Jones (1995). Measures for the assessment of mutation operators are introduced independently of the search strategy (Section 3). In Section 4, the practical relevance of the measures are proven for first–fit hill climbing. As well, those measures are used to prove that at the beginning of the search the random operator is superior to all local operators. In addition, an interdependence of the search state and the factors locality and randomness is shown (Section 5). This is underscored by some experiments in Section 6 before Section 7 concludes.

2 Fundamentals

This section establishes the fundamental conception of operator and problem structure underlying the following sections. The framework of this paper neglects the structure of the problem space but uses the relation between possible solutions induced by the operator and its possible offsprings only. The fitness evaluation for each solution results in a more or less rugged landscape (cf. Jones, 1995). Without loss of generality, only minimization problems on a finite search space are considered where the fitness values are positive. Moreover, an equidistant occurrence of the fitness values is assumed.

A mutation operator is formally defined as a function that maps one point to another point depending on certain random numbers. The operator defines a relation between the points and the possible offsprings.

Definition 1 (Operator) Suppose M is a finite set of points, the regarded *search space.*

An *operator Op on M* is a function $Op : M \times Param_{Op} \rightarrow M$ where $Param_{Op}$ denotes the range of the random numbers used by Op.

For each $m \in M$ the *set of offsprings of m under Op* is defined as the following multiset

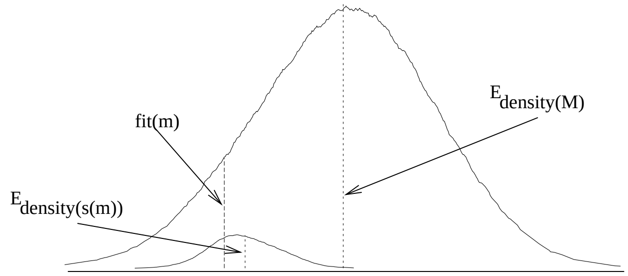

Figure 1: Example to fitness densities

$s_{Op}(m) \subset M \times \mathbb{N}$

$$s_{Op}(m) := \Big\{ (m', h) \in M \times \mathbb{N} \;\Big|\; \exists\, \hat{p} \in Param_{Op} : Op(m, \hat{p}) = m' \text{ and } h = |\{\hat{p} \in Param_{Op} \mid Op(m, \hat{p}) = m'\}| \Big\}.$$

The operator's *offspring fitness range* is defined as

$$I_{Op(m)} := \Big[\min\{fit(\tilde{m}) \mid (\tilde{m}, h) \in s_{Op}(m)\}, \max\{fit(\tilde{m}) \mid (\tilde{m}, h) \in s_{Op}(m)\} \Big].$$

An operator *Op* is called *compact* on $M' \subset M$ iff

$$\forall m \in M' \; \forall \tilde{m} \in M : \Big(\big(fit(\tilde{m}) \in I_{Op(m)}\big) \Rightarrow (\exists (\tilde{m}', h) \in s_{Op}(m) : fit(\tilde{m}') = fit(\tilde{m})) \Big).$$

Notation: In a multiset, the second component denotes the frequency of the element in the first component. Conventional multiset operators are used—especially the union is defined as $(\tilde{m}, h_1) \cup (\tilde{m}, h_2) := (\tilde{m}, h_1 + h_2)$.
In the remainder of the paper, the index *Op* in $s_{Op}(m)$ is omitted where the operator is clear from the context. As well, $s_i(m)$ denotes $s_{Op_i}(m)$.

For simplicity, a finite range of the random numbers with uniform distribution is assumed. Therefore, counting arguments are used in the proofs.

Note that the operator–given fitness landscape differs from the frequently examined usual landscape which depends on the problem's representation only.

Since the operator–defined landscape is expensive to compute, the fitness density is introduced, which may be approximated easily (e.g. Rose, Ebeling, & Asselmeyer, 1996).

Definition 2 (Fitness density) Suppose M is a search space and $fit : M \to \mathbb{R}^+$ a fitness function that assigns a nonnegative real value to each point in M. Then, the *fitness density* $density(M') : fit(M') \to [0, 1]$ for $M' \subseteq M$ is defined by

$$density(M')(r) := \frac{|\{m \in M' \mid fit(m) = r\}|}{|M'|}$$

Notation: In the remainder of the paper, the mean $E_{density(M')}[X]$ is denoted by $E_{density(M')}$ and the variance $V_{density(M')}[X]$ is denoted by $V_{density(M')}$.

Note that the offspring distribution is regarded relative to the fitness instead of the problem structure. This is sensible since the observed behavior (fitness) is the only criterion for the selection and survival of individuals—not some relation on the underlying genotype (cf. Fogel, 1994, p. 3). Especially the variance differs completely from the variance that is used as strategy parameter in evolutionary strategies.

Such a fitness density may be computed for both the complete problem space and the offsprings of a single point or a set of points. Therewith, quality measures for operators, as introduced in the next section, may be computed by estimated approximation.

Figure 1 shows how the fitness density may be applied to the complete search space as well as an operator's offspring range. Point $m \in M$ is a representative of fitness level $\mathit{fit}(m)$. During a search, the current point's fitness level may be viewed as the state of the search process. Note that this example's intention is clarification only—therefore, the densities are not properly scaled.

3 Measures for operators

This section introduces two measures for the comparison of operators. They make it possible to rate operators relative to the fitness level. This is achieved by choosing the density of the operator's offsprings as reference quantity. The measures are analyzed with respect to the fitness level in Section 5.

For the time being, all criteria are defined for single points in the search space. This may be extended to fitness ranges by averaging the quality of all points with a particular fitness.

Definition 3 (Measures for operators) Suppose Op is an operator on M and $\mathit{better}(m, N) := \{(m', h) \in N \mid \mathit{fit}(m') < \mathit{fit}(m)\} \subset M \times \mathbb{N}$ denotes the multiset of points with better fitness. In this definition b abbreviates $\mathit{better}(m, s(m))$.

1. The *improvement probability of Op with respect to point* $m \in M$ is defined by

$$\mathit{Improve}_{Op}(m) \quad := \quad \frac{|b|}{|s(m)|}.$$

2. The *relative expected improvement of Op with respect to point* $m \in M$ is defined by

$$\begin{aligned} \mathit{Relprogress}_{Op}(m) \quad &:= \quad E_{density(s(m))}[\max\{\mathit{fit}(m) - X, 0\}] \\ &= \quad \mathit{Improve}_{Op}(m) \frac{1}{|b|} \sum_{(m',h') \in b} (\mathit{fit}(m) - \mathit{fit}(m'))h' \\ &= \quad \frac{1}{|s(m)|} \sum_{(m',h') \in b} (\mathit{fit}(m) - \mathit{fit}(m'))h'. \end{aligned}$$

The improvement probability captures the probability with which the application of an operator Op yields an improvement from a single point m. Hence for first–fit hill climbing, the expected number of evaluations until a better successor is found may be computed directly (see remark after Definition 4). Apparently, the improvement probability is a suitable measure to estimate how often the operator needs to be applied in order to get an improvement. Nevertheless, suppose all points in the search space have distinct fitness values. Then, similarly to long path problems (e.g. Horn, Goldberg, & Deb, 1994), an operator may be constructed arranging the search space in such a way that each point has a single successor with the next occurring better fitness value. Then, the improvement probability equals 1.0 for all points except the optimum and is obviously unsuitable as a sole quality measure.

The relative expected improvement measures the expected fitness change if an improvement takes place qualified by the improvement probability. This measure cannot be mislead by long path problems.

Those or similar measures have already been introduced previously. Beyer (1994) computed both quality measures for mutation in $(1 \overset{+}{,} \lambda)$–evolution strategies and Fogel and Ghozeil (1996) determined the relative expected improvement in empirical examinations for evolution strategies as well as discrete problems. In our work, the relative expected improvement is considered depending on the fitness level. Moreover, it is examined how the quality is influenced by operator properties. This enables theoretical results for arbitrary discrete problems.

4 Relative expected improvement and search

This section shows that the relative expected improvement of an operator has considerable relevance for the expected progress of a search. The search process is considered as an iterative application of an operator where a successor is selected from the operator's offsprings.

In this paper, only first–fit hill climbing is regarded in which the operator is applied to the current point until a point with better fitness is found. Then, this new point (the first one fitting in the better range) becomes the current point. The search strategy in first–fit hill climbing is equivalent to $(1+1)$–strategies in evolution strategy terminology. For simplicity, no termination criteria is considered here.

Definition 4 (First–fit hill climbing) Suppose M is a search space, Op an operator on M, and $m \in M$. Then, the *set of reachable points after n evaluations from starting point m using first–fit hill climbing* is defined as

$$\begin{aligned} \mathit{reach}_1(m, Op) &:= b(m) \cup \bigcup_{(m',h') \in w(m)} \{(m, h')\} \\ \mathit{reach}_{n+1}(m, Op) &:= \bigcup_{(m',h') \in \mathit{reach}_n(m,Op)} b(m') \cup \bigcup_{(m'',h'') \in w(m')} \{(m', h'')\} \end{aligned}$$

where $b(\tilde{m})$ denotes $\mathit{better}(\tilde{m}, s(\tilde{m}))$ and $w(\tilde{m})$ denotes $s(\tilde{m}) \setminus b(\tilde{m})$.

The *expected fitness change by first–fit hill climbing after n evaluations using operator Op relative to m* is

$$\mathit{Searchprogress}_1^{Op}(m) := \mathit{Relprogress}_{Op}(m)$$

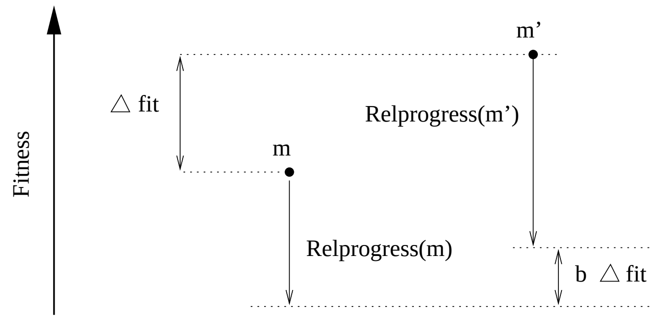

Figure 2: Condition for Relprogress in Proposition 1

$$Searchprogress^{Op}_{n+1}(m) \quad := \quad fit(m) - \frac{1}{|reach_n|} \sum_{(m',h') \in reach_n} \big(fit(m') - Relprogress_{Op}(m')\big) \; h'$$

where $reach_n$ abbreviates the multiset $reach_n(m, Op) \subset M \times \mathbb{N}$.

Note that the sets of reachable points are multisets. Therefore, a single point of M may occur several times in the reachable set.

Remark For first–fit hill climbing, the mean number of necessary fitness evaluations until an improvement takes place is $\left\lceil p + \frac{1}{p} - 1 \right\rceil$ where $p := Improve_{Op}(m)$, supposed $p > 0$, i.e. an improvement takes place at all.

The following proposition states the interrelation between the quality of different operators and the expected result of their searches.

Proposition 1 *Suppose $m \in M$ is arbitrary but fixed, Op_1, Op_2 are operators, and the reachable range of points within n steps is*

$$N \quad := \quad \big\{ m' \;\big|\; \exists\ 1 \leq i < n, j \in \{1,2\}, h > 0 : \; (m', h) \in reach_i(m, Op_j) \big\}.$$

Assume moreover that, $\exists\ a, b \in \mathbb{R},\ 0 \leq b < 1\ \forall m' \in N$:

$$Relprogress_{Op_1}(m') = a + b \; fit(m').$$

*Then, from $Relprogress_{Op_1}(m') \geq Relprogress_{Op_2}(m')$ for all $m' \in N$
it follows that $Searchprogress^{Op_1}_{n}(m) \;\geq Searchprogress^{Op_2}_{n}(m)$.*

In order to show this proposition, the linearity precondition that $Relprogress_{Op_1}$ depends directly on the fitness is necessary. This condition is a hard requirement in practice but it should be fulfilled at least in sections of the fitness range.

Figure 2 demonstrates the linearity condition which means that there is a linear dependency between the difference of *Relprogress* applied to two points and the difference of the point's fitness values. This guarantees that the worse individual cannot overtake the better individual. As long as this is the case, *Relprogress* is a sufficient criterion for the convergence speed of the search process.

Proof (Proposition 1): By induction on i

$$A_i \quad \equiv \quad E_{density(reach_i(m,Op_2))} \;\geq\; E_{density(reach_i(m,Op_1))}$$

is shown from which the proposition follows directly. A_1 holds per definition because of the superiority regarding Relprogress. Now, suppose A_{i-1} is true. Statement A_i is equivalent to

$$\begin{aligned}
&\frac{1}{|reach_{i-1}(m,Op_2)|} \sum_{(m',h')\in reach_{i-1}(m,Op_2)} Relprogress_{Op_2}(m')\ h' \\
&\quad - \frac{1}{|reach_{i-1}(m,Op_1)|} \sum_{(m',h')\in reach_{i-1}(m,Op_1)} Relprogress_{Op_1}(m')\ h' \\
&\qquad \leq E_{density(reach_{i-1}(m,Op_2))} - E_{density(reach_{i-1}(m,Op_1))}
\end{aligned}$$

which is shown now. First, we assume that $|reach_{i-1}(m,Op_1)| = |reach_{i-1}(m,Op_2)|$. Then, it is possible to arrange all elements of both reachable sets in pairs $(m_k^{r_1}, m_k^{r_2})$ where $k \in I := \left[1, |reach_{i-1}(m,Op_1)|\right]$ and $m_k^{r_j} \in reach_{i-1}(m,Op_j)$. In the first following transformation, the linearity condition of *Relprogress* is used and in the second step $Relprogress_{Op_1}(m') \geq Relprogress_{Op_2}(m')$ is applied.

$$\begin{aligned}
&\sum_{k\in I} Relprogress_{Op_2}(m_k^{r_2}) - \sum_{k\in I} Relprogress_{Op_1}(m_k^{r_1}) \\
&\quad = \sum_{k\in I} Relprogress_{Op_2}(m_k^{r_2}) - \sum_{k\in I} Relprogress_{Op_1}(m_k^{r_2} + b(fit(m_k^{r_1}) - fit(m_k^{r_2})) \\
&\quad \leq \sum_{k\in I} b(fit(m_k^{r_2}) - fit(m_k^{r_1})) \\
&\quad \leq \sum_{(m',h')\in reach_{i-1}(m,Op_2)} fit(m')\ h' \;-\; \sum_{(m',h')\in reach_{i-1}(m,Op_1)} fit(m')\ h'
\end{aligned}$$

The last step uses $0 \leq b < 1$ and with the induction hypothesis A_i follows. For the case $|reach_{i-1}(m,Op_1)| \neq |reach_{i-1}(m,Op_2)|$, a padding technique with a hypothetical element is used. □

5 Operators, locality, and the state of the search process

Inspired by the NFL (Wolpert & Macready, 1997), which questions the "universal" algorithm, and the self–adaption of evolution strategies, the focus of this work is the examination of operators depending on the fitness level resp. the state of the search. The general idea is to use different operators at different moments during search since for many operators self–adaption is not intuitively feasible (e.g. for permutation operators). This section examines the random operator compared to any local operator (Proposition 2) and, subsequently, how the variance of the operator's offspring density influences the behavior of local operators depending on the fitness level (Proposition 3). Those results are generalized in Claim 1.

In order to formulate those statements, the following definition introduces two properties of local operators. The first property (*locality on average*) makes a statement only on the mean fitness change of the operator for all possible offsprings where the second property (*locally random*) assumes a certain fitness distribution for the offsprings of an operator.

Remark: Note that for all compact operators on M' and a $m \in M'$ there exists a real valued, continual density $\varphi_{s_{Op}(m)} : I_{Op(m)} \to [0,1]$ with

1. $\varphi_{s_{Op}(m)}(\mathit{fit}(m')) = \mathit{density}(s_{Op}(m))(\mathit{fit}(m'))$ for all $m' \in s_{Op}(m)$
2. $E_{\varphi_{s_{Op}(m)}} = E_{\mathit{density}(s_{Op}(m))(\mathit{fit}(m))}$ and $V_{\varphi_{s_{Op}(m)}} = V_{\mathit{density}(s_{Op}(m))(\mathit{fit}(m))}$.

Since equidistant fitness values are assumed it is always possible to find such a density within a small error range which is neglected in the remainder.

Definition 5 (locality) Suppose M is a search space, $M' \subseteq M$, and Op is a compact operator on M'. Then Op is called

1. *δ–local on average for* M' with $0 < \delta \in \mathbb{R}$ iff it holds for all $m \in M'$ that

$$\left|\mathit{fit}(m) - E_{\mathit{density}(s(m))}\right| < \delta.$$

2. *locally random with density* Φ *for* M' with $V_\Phi = 1$ iff

$$\begin{aligned} &(1) \quad \forall m \in M' \; \forall r \in I_{Op(m)} : \\ &\qquad \varphi_{s_{Op}(m)}(r) = \frac{1}{\varepsilon_{s(m)}} \Phi(E_\Phi - \frac{1}{\varepsilon_{s(m)}}(E_{\varphi_{s_{Op}(m)}} - \mathit{fit}(m'))), \text{ and} \\ &(2) \quad E_{\varphi_{s_{Op}(m)}} - \mathit{fit}(m) = \frac{\varepsilon^2_{s(m)}}{\varepsilon^2_M}(E_{\mathit{density}(M)} - \mathit{fit}(m)) \end{aligned}$$

where $\varphi_{s_{Op}(m)}$ is defined by the previous remark and $\varepsilon_{M''} := \sqrt{V_{\mathit{density}(M'')}}$ for $M'' \subseteq M$.

The first definition is an analogy to the examinations of correlation coefficients by Manderick et al. (1991) where the mean offspring fitness was inspected relative to the parental fitness. An operator is considered more local if δ is smaller. Since only the mean fitness is regarded, the definition of δ–local on average covers only one aspect of locality. Nevertheless, this suffices to show Proposition 2.

The second part of the definition regards the density of the possible offsprings' fitness values and demands that these densities of a set of points may be transformed to a common density. In addition, the offspring density is shifted towards the mean of the search space. With a bell–shaped $\mathit{density}(M)$ in mind, such a displacement is sensible since the choice range is bigger towards $E_{\mathit{density}(M)}$ than towards the border of $\mathit{density}(M)$. Such a bell–shaped distribution as well as a displacement have been corroborated by empirical examinations of various problems. Note that the displacement's size is chosen rather arbitrarily. Similar results to the following propositions may be proven using various different displacements.

Since those locality properties are defined in terms of the fitness density, those definitions differ from the usually considered locality which is defined relative to the underlying genotype. The proposed locality depends on the operator's behavior, i.e. its effect regarding the fitness only.

Each operator which is locally random with Φ is δ–local with $\delta = \frac{\varepsilon^2_{s(m)}}{\varepsilon^2_M}(E_{density(M)} - fit(m))$. The bigger the variance of the offspring density is (i.e. the operator is more random), the more the offspring density is shifted towards the mean of the search space. As a consequence the operator is δ–local with a bigger δ and, therefore, less local as defined by the first part of the definition.

As a first consequence from the locality definition, the following proposition states what is well–known from experiments: at the beginning of a search the random operator is superior to any local operator with respect to the relative expected improvement.

Proposition 2 *Suppose locOp is a δ–local operator for*

$$\begin{aligned} M' \;=\; \{m \in M \;\mid\; & better(m, M) \geq \frac{1}{2}|M| \text{ and} \\ & fit(m) \geq E_{density(M)} + 2\delta + 2q\delta\} \end{aligned}$$

where $1 < q \in \mathbb{R}$ is defined by

$$\max\left\{fit(m') - fit(m) \;\middle|\; m' \in s_{locOp}(m) \text{ and } fit(m) \geq E_{density(M)}\right\} \;\leq\; q\delta.$$

Then, $Relprogress_{locOp}(m) \;\leq\; Relprogress_{randomOp}(m)$ for all $m \in M'$.

Proof: Using Definition 5 and prerequisite $fit(m') \leq fit(m) + q\delta$, $m' \in s_{locOp}(m)$, the following equation results with $s := s_{locOp}(m)$ and $better := better(m, s)$.

$$\begin{aligned} Relprogress_{locOp}(m) \;&=\; \frac{1}{|s|} \sum_{(m',h) \in better} (fit(m) - fit(m'))\; h \\ &\leq\; \delta \left(1 + \frac{|s| - |better|}{|s|} q\right) \;\leq \delta(1+q) \end{aligned}$$

As well, using the conditions of M' the following equation holds for $m \in M'$.

$$\begin{aligned} Relprogress_{randomOp}(m) \;&\geq\; \frac{|better(m, M)|}{|M|}(fit(m) - E_{density(M)}) \\ &\geq\; \delta(1+q) \end{aligned}$$

And the proposition follows. □

This result shows that, at the beginning, random search is useful within a certain range of the worse part of the search space. Moreover, it implies that it is sensible to initialize several times (e.g. by using populations or selecting a starting point from a pool of randomly initialized individuals) in order to avoid wasting expensive fitness evaluations by local search.

The following proposition examines for two operators which are locally random with different extents of randomness under which circumstances one operator is superior.

Proposition 3 *Suppose M is a search space, Op_i $(i = 1,2)$ are locally random operators with density Φ for $M' := \{m \in M \mid fit(m) < E_{density(M)}\}$ where Φ is strictly monotonous increasing for $(-\infty, E_\Phi)$ and*

$$V_{density(s_1(m))} \quad \leq \quad V_{density(s_2(m))}$$

for all $m \in M'$. Let $\varepsilon_i < \varepsilon_M$ $(i = 1,2)$ and $\substack{\min \\ \max}\ \varepsilon_i :=\substack{\min \\ \max} \left\{\sqrt{V_{density(s_i(m))}} \mid m \in M'\right\}$ $(i = 1,2)$. Then,

1. *for all $m \in M'$ with*

$$fit(m) \in \left[0, \left(1 - \frac{\varepsilon_M^2}{\min \varepsilon_2 \ \min \varepsilon_1 + \varepsilon_M^2}\right) E_{density(M)}\right)$$

   it holds that $Relprogress_{Op_1}(m) \geq Relprogress_{Op_2}(m)$.

2. *for all $m \in M'$ with*

$$fit(m) \in \left(E_{density(M)} - \frac{\varepsilon_M^2}{3 \max \varepsilon_2 \ \max \varepsilon_1}, E_{density(M)}\right)$$

   and $fit(m) - \min\left(I_{Op_1} \cup I_{Op_2}\right) > 3\varepsilon_2$
   it holds that $Relprogress_{Op_2}(m) \geq Relprogress_{Op_1}(m)$.

Remark Op_2 is more random (less local) than Op_1.

The essential idea of the proposition is to require a common density into which the offspring densities may be transformed. Thus, they become comparable. Admittedly, such a prerequisite is of theoretical nature and can scarcely be fulfilled in practice. Nevertheless, the proposition shows a basic trend.

Proof (Proposition 3): Using Definition 3 and 2 it holds for $m \in M'$:

$$\begin{aligned} & Relprogress_{Op_1}(m) - Relprogress_{Op_2}(m) \\ = & \sum_{r \in \tilde{I}} (fit(m) - r)(density(s_1(m))(r) - density(s_2(m))(r)) \end{aligned} \tag{1}$$

with $\tilde{I} := [\min I_{Op_1(m)} \cup I_{Op_2(m)}, fit(m)]$ and $I_{Op_i(m)} \subset \mathbb{R}$ $(i = 1,2)$ defined in Definition 1. Since $Op_i(m)$ are equidistant and compact there exists a $n \in \mathbb{N}$ $(n \geq 1)$ such that $\{r_1, \ldots, r_n\}$ are all occuring fitness values in $\tilde{I}$ with $fit(m) - r_j = j$—without loss of generality a fitness distance 1 is assumed.

Note that by Definition 5 in each summation term of (1)

$$\begin{aligned} & density(s_1(m))(r) - density(s_2(m))(r) \\ = & \frac{(\varepsilon_2 - \varepsilon_1)}{\varepsilon_1 \varepsilon_2} \Phi(transform_1^{(m)}(r)) \\ & \qquad + \frac{\varepsilon_1}{\varepsilon_1 \varepsilon_2} (\Phi(transform_1^{(m)}(r)) - \Phi(transform_2^{(m)}(r))) \end{aligned} \tag{2}$$

Furthermore, let $transform_i^{(m)}(r) = E_\Phi - \frac{1}{\varepsilon_i}(E_{density(s_i(m))} - r)$ ($i = 1, 2$ and $\varepsilon_i := \varepsilon_{s_i(m)}$). Then

$$\begin{aligned} & transform_1^{(m)}(r) - transform_2^{(m)}(r) \\ & \quad = \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_M^2}(E_{density(M)} - fit(m)) - \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 \varepsilon_1}(fit(m) - r). \end{aligned}$$

To prove the first part of the proposition it will be shown

$$\forall r_j \in \{r_1, \ldots, r_n\} \quad : \quad density(s_1(m))(r_j) - density(s_2(m))(r_j) \geq 0.$$

In order to show $density(s_1(m))(r) - density(s_2(m))(r) \geq 0$, it is sufficient to show $transform_1^{(m)}(r) - transform_2^{(m)}(r) \geq 0$ since (2) and Φ being strictly monotonous increasing for $(0, E_\Phi)$.

For $fit(m) \leq \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 \varepsilon_2 + \varepsilon_M^2} E_{density(M)}$ it holds that

$$\begin{aligned} & transform_1^{(m)}(r) - transform_2^{(m)}(r) \\ = \; & \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_M^2} E_{density(M)} - \frac{(\varepsilon_2 - \varepsilon_1)(\varepsilon_1 \varepsilon_2 + \varepsilon_M^2)}{\varepsilon_1 \varepsilon_2 \varepsilon_M^2} fit(m) + \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 \varepsilon_1} r \\ \geq \; & \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_M^2} E_{density(M)} - \frac{(\varepsilon_2 - \varepsilon_1)(\varepsilon_1 \varepsilon_2 + \varepsilon_M^2)}{\varepsilon_1 \varepsilon_2 \varepsilon_M^2} \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 \varepsilon_2 + \varepsilon_M^2} E_{density(M)} + \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 \varepsilon_1} r \\ \geq \; & 0 \end{aligned}$$

since all fitness values are positive.

In order to prove the second part of the proposition, it is used that there exists an $\alpha \geq \frac{1}{2}$ such that $\Phi(transform_1^{(m)}(r)) - \Phi(transform_2^{(m)}(r)) \leq \alpha(transform_1^{(m)}(r) - transform_1^{(m)}(r))$ holds. Using this condition, (2), and the inequality $\Phi(transform_1^{(m)}(r_j)) \leq 1$ in (1) one gets

$$\begin{aligned} & Relprogress_{Op_1}(m) - Relprogress_{Op_2}(m) \\ \leq \; & \sum_{j=1}^{n} j \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1 \varepsilon_2} + \frac{\alpha}{\varepsilon_2} \left(transform_1^{(m)}(r_j) - transform_2^{(m)}(r_j) \right) \right) \\ = \; & \sum_{j=1}^{n} j \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1 \varepsilon_2} + \frac{\alpha}{\varepsilon_2} \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_M^2}(E_{density(M)} - fit(m)) - \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1 \varepsilon_2} j \right) \right) \\ = \; & \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1 \varepsilon_2} \frac{n(n+1)}{2} \left(1 + \frac{\varepsilon_1 \alpha}{\varepsilon_M^2}(E_{density(M)} - fit(m)) - \frac{\alpha}{\varepsilon_2} \frac{2n+1}{3} \right) \end{aligned}$$

For $\alpha \geq \frac{1}{2}$, $fit(m) \geq E_{density(M)} - \frac{\varepsilon_M^2}{3 \varepsilon_1 \varepsilon_2}$, and $n = fit(m) - \min\left(I_{Op_1} \cup I_{Op_2}\right) > 3\varepsilon_2$ one gets

$$\begin{aligned} & Relprogress_{Op_1}(m) - Relprogress_{Op_2}(m) \\ \leq \; & \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1 \varepsilon_2} \frac{n(n+1)}{2} \left(1 - \frac{\alpha}{\varepsilon_2} \frac{2n+1}{3} + \frac{\varepsilon_1 \alpha}{\varepsilon_M^2} \left(E_{density(M)} - \left(E_{density(M)} - \frac{\varepsilon_M^2}{3 \varepsilon_1 \varepsilon_2} \right) \right) \right) \\ \leq \; & \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1 \varepsilon_2} \frac{n(n+1)}{2} \left(1 - \frac{n}{3 \varepsilon_2} \right) \leq 0 \end{aligned}$$

□

This result shows that a more random operator is superior to a local operator in the range around $E_{density(M)}$ since the bigger variance causes bigger improvements. The more local an operator is, the bigger is the range in which the random operator is superior. Close to the optimum, this effect is neutralized by the bigger displacement of the offspring range towards $E_{density(M)}$.

As a consequence, a local operator is only sensibly applicable towards the optimum. Evolutionary algorithms achieve this random influence by additional operators, like e.g. recombination in evolution strategies.

The proposition implies that there is no global measure to assess an operator concerning cost and convergence speed since locality and randomness dominate in different fitness ranges. Moreover, for all local operators within the limits of the proposition there is a fitness range where they are superior as well as a fitness range where they are inferior. Clearly, this is not valid for self–adapting operators since they do not fit in the proposed framework.

Since the authors believe that those statements are true for a wider class of problems and operators, the following claim generalizes the propositions' results without the restricting prerequisites.

Claim 1 *For non–trivial local operators, i.e. for each non–extreme point there are better and worse offsprings, the following statements are proposed with respect to the relative expected improvement as quality measure.*

1. *The quality of a locally random operator cannot be viewed independently of the search state.*
2. *The more local a locally random operator, the better it is in the range close to the optimum.*
3. *In the area close to $E_{density(M)}$ a less local operator is superior.*
4. *For a search with first–fit hill climbing, it is advantageous to switch operators during search.*

This partially proven claim may give deep insight in the mechanism and effectiveness of operators. As a consequence, the state of the search process, i.e. the fitness level, is crucial for the judgment of operators. This does not only lead to self–adapting operators but also a change of operators during the search process might be sensible. To some extent, such a change is already realized in some hybrid methods where extensive operators alternate with local operators with genotypic locality (cf. Hart, 1994).

Summarizing, the random operator is unrivaled for nearly the complete worse half of the search space. By the examination of operators where the randomness is limited by locality, a diminishing positive influence of randomness is shown, the closer search gets to the optimum. The relevance of locality increases inversely proportional. Thus, the results of this section give deep insight in the character of operators, their behavior, and the influence of this quality on searches.

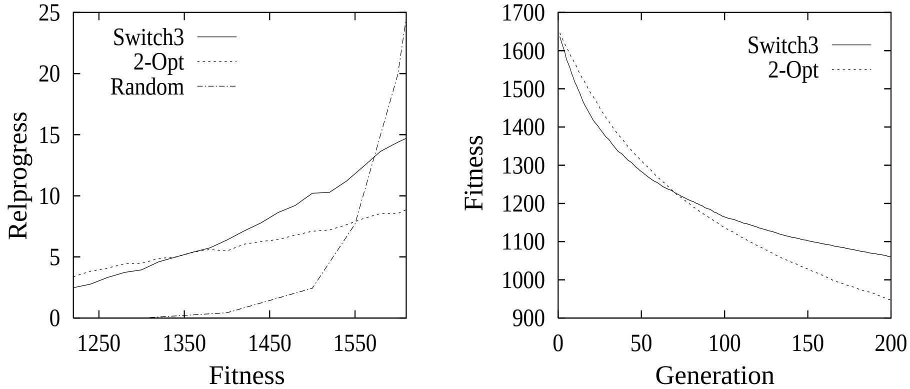

Figure 3: Comparison of relative expected improvement and search with first–fit hill climbing for TSP benchmark

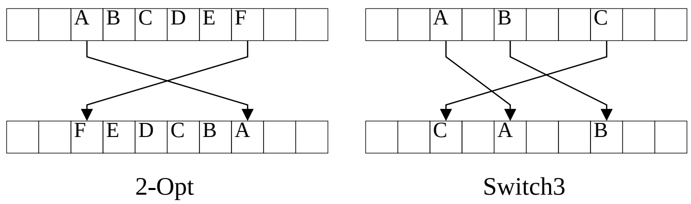

Figure 4: Schematic description of used permutation operators in the TSP example

6 Experiments

The previous results are valid under certain prerequisites which are necessary to carry out the theoretical proofs. Since those prerequisites in their strict form can neither be verified nor computed for practical problems, the statements of Claim 1 are validated empirically for three exemplary toy optimization problems: two permutation problems (traveling salesperson problem and a complexity measure) and a graph coloring problem. This is intended to emphasize the practical relevance of this work's results. For all examples, the quality measures are approximated statistically and compared to empirical results of several hill climbing runs.

6.1 Traveling salesperson problem (TSP)

The symmetric TSP benchmark `eil51.tsp` from TSPLIB with the following operators illustrated in Figure 4 is used for this section. Operator *2–Opt* (cf. Lin & Kernighan, 1973) reverses a substring of the permutation and operator *Switch3* exchanges three numbers of the permutation. Additionally, operator *Random* is regarded which chooses a completely new random permutation.

The left part of Figure 3 shows the approximated quality of the operators in dependence of the fitness level, i.e. the current fitness value. It shows clearly the superior quality of

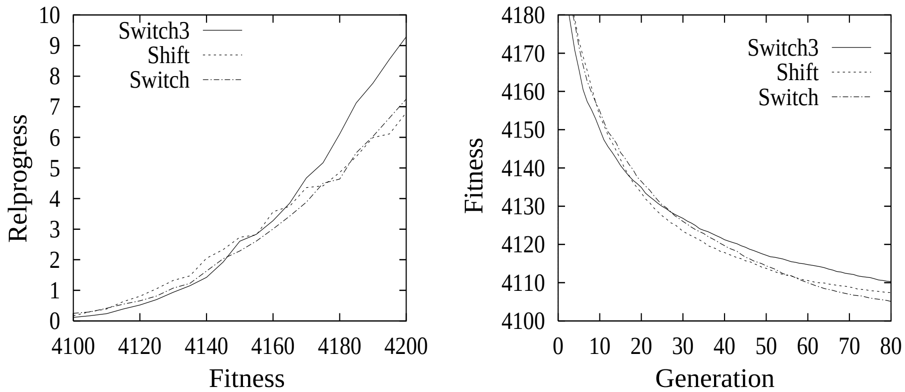

Figure 5: Comparison of relative expected improvement and search with first–fit hill climbing for transformed C20

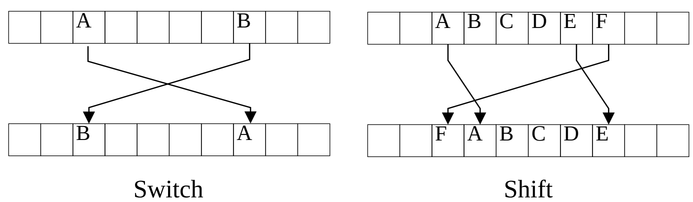

Figure 6: Schematic description of used permutation operators in the C20 example

the random operator in the worse fitness range (cf. Proposition 2). In the better range a crossing of the operator quality of 2–Opt and Switch3 around fitness 1350 is recognizable. Since examinations have shown that 2–Opt has a smaller variance than Switch3, the figure reflects the predicted behavior according to Proposition 3. This change in the operator quality is also obvious in the observed behavior of search with first–fit hill climbing as it is displayed in the right part of Figure 3. From the predicted fitness level 2–Opt improves compared to Switch3. The curves in the right diagram intersect with a slight delay around fitness 1220.

6.2 Complexity measure “C” on permutations

The second optimization problem is a complexity measure for permutations introduced by Claus (1991). This maximization problem is transformed into a minimization problem in order to fit in the suggested framework.

The used operators are Switch3, the *Switch* operator which exchanges two numbers in the permutation, and the *Shift* operator which shifts a substring of the permutation to the right and fills the first position with the last number of the substring (cf. Figure 6).

In Figure 5, a crossing of Relprogress of Shift and Switch3 is obvious around fitness 4155. This is plausible since Switch3 has a bigger variance. Also, a second intersection of Shift and Switch around fitness 4114 is recognizable. Both changes of quality are reflected in

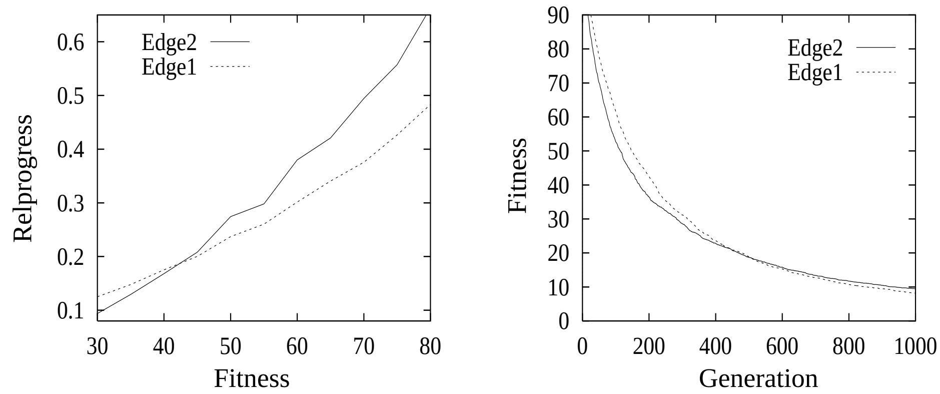

Figure 7: Comparison of relative expected improvement and search with first–fit hill climbing for graph coloring problem

the curve on the right side around fitness 4139 resp. 4110. This example shows that three operators are superior to the other operators in three different fitness ranges.

6.3 Graph coloring

For the benchmark graph coloring problem `myciel6.col` with 95 vertices, 755 edges, and 7 colors a intuitive representation is used where each vertex a number from 1 to 7 is assigned, representing a color. The fitness is the number of edges with violated color constraints.

The used operators are *Edge1* and *Edge2*. Both operators choose a random edge in the graph. Then, Edge1 assigns a random color to one vertex of the edge—the color must be distinct from the other vertex' color. Edge2 assigns two random but distinct colors to both vertices.

In the left part of Figure 7, an intersection of Relprogress curves of Edge1 and Edge2 is obvious around fitness 43. Contrary to both previous examples, this result was gained using samples from several search runs instead of random samples. Due to the problem size and the location of the intersection close to the optimum, no intersection could be observed using random samples from the complete search space only. Since Edge2 has a bigger variance than Edge1, the crossing is in accordance with the theoretical results. The right part of Figure 7 shows the effect of such a crossing around fitness 20 in empirically collected data of hill climbing runs.

7 Conclusion and future work

Where the "No free lunch" theorem has stated that there is no algorithm that performs better on all problems than any other algorithm, this work shows in addition that for simple hill climbing the operator must be evaluated according to the state of the search. This is achieved by averaging and valuing operators locally with respect to the operator–defined neighborhood. With this approach, global convergence statements are not possible—and not sensible in our view. Rather, it puts a new complexion on the interplay of problem and operator depending on fitness levels. Provided a certain locality of an operator, it is shown that randomness and locality are factors influencing the operator's quality depending on the fitness level. Close to the optimum the locality takes effect which turns out to be unfavorable in the worse fitness range. The influence of randomness is contrary. This enables new insights into the relevance of randomness for the quality of operators. The necessity of decreasing random influence towards the end of the search was already observed frequently (e.g. Davis, 1989). Yet, the presented work analyzes the interplay of the different factors. Moreover, the results show that it is not sufficient to assess an operator at the beginning of the search process using statistical measures for a random walk since this enables only statements for the fitness level around the center of search space's fitness distribution and not the area close to the optimum which is of particular interest.

Besides the adaptation of mutation probabilities, our examples imply that an interplay of different random and local operators might be sensible. Clearly, in the examples used in this paper, it is insignificant if the performance of a search using one operator is slightly worse than a search using a different operator. But in real world problems with huge search spaces and expensive fitness evaluations, the results of this work have more significance.

The presented framework establishes a basis for the evaluation and comparison of operators which is applicable for arbitrary finite problems. Besides theoretical statements the framework is of practical relevance. A substantial advantage of the proposed measure of quality is the fact that it is empirically ascertainable and actually relevant to the search process. In order to prove the theoretical statements strict prerequisites are required. Future work will analyze if and with which consequences the preconditions may be relaxed. Among others, it is only shown that for first–fit hill climbing the operator's quality influences the costs of a search. Nevertheless, this work provides the framework for further examinations of other search strategies as for instance threshold algorithms. Besides, future research will examine local optima, which is considered only implicitly in the relative expected improvement, and especially their influence on the operator quality, with the proposed methods.

Acknowledgements The authors thank Volker Claus, the anonymous reviewers, and all FOGA participants (especially Michael Vose) for their valuable input and helpful comments.

References

Beyer, H.-G. (1994). Towards a theory of 'evolution strategies': progress rates and quality gain for $(1 \overset{+}{,} \lambda)$–strategies on (nearly) arbitrary fitness functions. In Y. Davidor, H.-P. Schwefel, & R. Maenner (Eds.), *Parallel Problem Solving from Nature—PPSN III (Int. Conf. on Evolutionary Computation, Jerusalem, 1994) (Lecture Notes in Computer Science 866)* (pp. 58–67). Berlin: Springer.

Beyer, H.-G., & Rudolph, G. (1997). Local performance measures. In T. Bäck (Ed.), *Handbook of Evolutionary Computation.* Bristol, Oxford: IOP Puplishing Ltd and Oxford University Press.

Claus, V. (1991). Complexity measures on permutations. In Buchmann, Ganzinger, & Paul (Eds.), *Informatik: Festschrift zum 60. Geburtstag von Günter Hotz.* Leipzig: Teubner Texte Informatik.

Davis, L. (1989). Adapting operator probabilities in genetic algorithms. In J. D. Schaffer (Ed.), *Proc. 3rd Int. Conf. on Genetic Algorithms (Fairfax, VA, 1989)* (pp. 61–69). San Mateo, CA: Morgan Kaufmann.

Fogel, D. B. (1994). An introduction to simulated evolutionary optimization. *IEEE Trans. on Neural Networks*, *5*(1), 3–14.

Fogel, D. B., & Ghozeil, A. (1996). Using fitness distributions to design more efficient evolutionary computations. In *Proc. of 1996 IEEE Conf. on Evolutionary Computation* (pp. 11–19). New York: IEEE Press. (Keynote lecture)

Hart, W. E. (1994). *Adaptive global optimization with local search.* Unpublished doctoral dissertation, University of California, San Diego.

Hordijk, W. (1996). A measure of landscapes. *Evolutionary Computation*, *4*(4).

Horn, J., Goldberg, D. E., & Deb, K. (1994). Long path problems. In Y. Davidor, H.-P. Schwefel, & R. Maenner (Eds.), *Parallel Problem Solving from Nature—PPSN III (Int. Conf. on Evolutionary Computation, Jerusalem, 1994) (Lecture Notes in Computer Science 866).* Berlin: Springer.

Jones, T. (1995). *Evolutionary algorithms, fitness landscapes and search.* Unpublished doctoral dissertation, University of New Mexico, Albuquerque, New Mexico.

Jones, T., & Forrest, S. (1995). Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In L. Eshelman (Ed.), *Proc. of the 6th Int. Conf. on Genetic Algorithms* (pp. 184–192). San Francisco, CA: Morgan Kaufmann.

Lin, S., & Kernighan, B. (1973). An efficient heuristic procedure for the traveling salesman problem. *Operations Res.*, *21*, 498–516.

Manderick, B., Weger, M. de, & Spiessens, P. (1991). The genetic algorithm and the structure of the fitness landscape. In R. Belew & L. Booker (Eds.), *Proc. of the 4th Int. Conf. on Genetic Algorithms (San Diego, CA, 1991)* (pp. 143–150). San Mateo, CA: Morgan Kauffmann.

Rose, H., Ebeling, W., & Asselmeyer, T. (1996). The density of states – a measure of the difficulty of optimization problems. In H. Voigt, W. Ebeling, I. Rechenberg, & H.-P. Schwefel (Eds.), *Parallel Problem Solving from Nature – PPSN IV (Int. Conf. on Evolutionary Computation, Berlin, 1996) (Lecture Notes in Computer Science 1141).* Berlin: Springer.

Rudolph, G. (1997). *Convergence properties of evolutionary algorithms.* Hamburg: Verlag Dr. Kovac.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Trans. On Evolutionary Computation*, *1*(1), 67–82.