

Locality Bounds for Nonredundant Binary-Integer Representations

Hrishee Shastri
Reed College, Portland, OR, USA
shastrih@reed.edu

Eitan Frachtenberg
Reed College, Portland, OR, USA
eitan@reed.edu

Abstract—Genetic and Evolutionary Algorithms (GEAs) rely on operators such as mutation and recombination to introduce variation to the genotypes. Because of their crucial role and effect on GEA performance, several studies have attempted to model and quantify the variation induced by different operators on various genotypic representations and GEAs. One metric of particular interest is the locality of genetic operators and representations, or how sensitive the phenotype is to small changes in genotype. Consequently, there is a considerable body of empirical work on the effects that different representations have on locality, with an emphasis on several popular representations, such as Gray encoding, and popular variation operators, such as single-bit mutation and single-point crossover.

Here, we compute and prove tight upper and lower bounds on locality. We first precisely define our locality metrics for the single-point mutation and generic crossover operators by reformulating Rothlauf’s seminal definitions specific to the binary-to-integer domain of representations. We then prove lower and upper bounds for single-point mutation locality by reducing the problem to mappings on hypercubes, and present constructive algorithms to generate representations of both optimal and pessimal locality. We also compute asymptotic bounds for generalized locality under any crossover operator.

Our primary result is that the single-point locality of standard binary encoding is provably as good as Binary-Reflected Gray encoding, while other Gray encodings, which we construct, have worse locality. Another important result is that the generalized locality of any nonredundant binary-to-integer representation quickly converges to the same value, meaning that this metric cannot discriminate among representations and may therefore lose its usefulness for binary-integer representations.

I. INTRODUCTION

Genetic and Evolutionary Algorithms (GEAs) solve optimization and search problems by codifying a population of possible solutions, evaluating them for fitness, and iteratively modifying them in an attempt to improve their fitness. The digital manifestation of the solutions are called genotypes, and their interpretations into the specific problem domain are called phenotypes. The function that maps from genotypes to phenotypes is simply called the *representation*, and it can have a

significant impact on the success and speed of the GEA to approximate optimal solutions. Consequently, many empirical and theoretical studies investigated the effects of representation on GEA performance under different variation operators (Sec. II).

Variation of genotypes is a key component of evolution, both biological [1] and computational [2]. Perhaps the most common operator for variation in GEAs is mutation, which may be combined with another operator, recombination [3]. Typical implementations of the mutation operator embody a localized change to an individual genotype. Contrast this with recombination, which requires two or more individuals and often involves nonlocal changes to the genotypes.

One popular example is point mutation, a simple mutation operator which randomly changes one allele at a time. Even within point mutation there are variations and parameters that control the magnitude of the genotypic change. Thus, in a search GEA, mutation can be used both for exploration—sampling many disparate parts of the search space—and for exploitation—thoroughly searching in a localized subspace [4].

An important property of mutation that can increase the predictability and interpretability of the GEA is to have “good locality,” which we informally define here as the property that small variations in the genotype lead to small variations in the phenotype [5]. Good locality implies better control of the GEA, because tuning mutation for a certain magnitude of changes in the genotype—the inputs to the search—leads to an expectation of the magnitude of changes in the phenotype—the outcome of the search. Note that the mutation operator is intricately tied to the representation. It is the combination of the mutation operator and representation that determines the magnitude of phenotypic change. Often, the discussion of locality assumes a fixed mutation operator, such as uniform bit flips or Gaussian differences, and focuses on the representation, such as standard binary or Gray.

In his seminal book, Rothlauf presented a theoretical framework for comparing representations based on three aspects: redundancy, scaling, and locality [6]. Rothlauf proposed two metrics to quantify the locality of repre-

sentations, one specifically for point mutation (simply called “locality” in the book) and one for any variation operator (called “distance distortion”).

This paper focuses on these metrics when applied to a widely used subclass of representations, namely nonredundant translations from bitstring genotypes to nonnegative integer phenotypes. Our main contributions are proofs of several properties and limitations of these two metrics. For point mutation, we show not only that standard binary (SB) and Binary-Reflected Gray (BRG) representations have identical locality for any bitstring length, but also that it is optimal. For generalized locality, we show that this metric—both Rothlauf’s and our derived form—is not actually useful, because its value does not meaningfully depend on the representation.

In the next section, we delve deeper into Rothlauf’s and others’ definitions of locality and how they have been used in the past. Then, in Sec. III, we derive new lower and upper bounds on point locality, as well as asymptotic limits for general locality. The main result of our bounds is that Gray encoding offers no locality advantage over standard binary encoding, even though some studies that found Gray encoding empirically superior, for reasons other than locality. Finally, we discuss possible ramifications for this finding and propose ideas for future research in Sec. IV.

II. BACKGROUND

Because of the strong impact of representations on GEA performance, there have been several attempts to formalize and measure the effects of representation locality [7]–[12]. These measures have been applied to a variety of phenotype classes, such as floating-point numbers [5], computer programs [7], [13], permutations [14], and trees [15], [16]. Many of these approaches focus on measuring the effect of genotypic changes on fitness distances [9]. A more general approach is to limit the scope to the effect on phenotypic distances, because it provides a locality measure of a representation that is independent of fitness function.

The foundational locality definitions for our study and several others [17], [18], come from Rothlauf’s treatise of the theory of GEA representations [6]. Rothlauf defines the locality d_m as:

$$d_m \stackrel{\text{def}}{=} \sum_{d_{x,y}^g = d_{min}^g} |d_{x,y}^p - d_{min}^p| \quad (\text{ibid. Eq. 3.23}),$$

where for every two distinct genotypes x, y , $d_{x,y}^g$ is the genotypic distance between x and y , and $d_{x,y}^p$ is the phenotypic distance between their respective phenotypes, based on our metrics of choice for genotypic and phenotypic spaces. Similarly, d_{min}^g and d_{min}^p represent the minimum possible distance between genotypes or

phenotypes, respectively.¹ For example, for nonredundant representations of integers as bitstrings (the focus of this paper), genotypic distances are measured by Hamming distance and phenotypic distances use the usual Euclidean metric in \mathbb{N} .

This definition “describes how well neighboring genotypes correspond to neighboring phenotypes” (ibid. p. 77), which is valuable for measuring small genotypic changes, typically resulting from a single mutation. Extending this notion to include large genotypic changes, e.g., from a recombination operator, Rothlauf defines *distance distortion* d_c as:

$$d_c \stackrel{\text{def}}{=} \frac{2}{n_p(n_p - 1)} \sum_{i=1}^{n_p} \sum_{j=i+1}^{n_p} |d_{x_i, x_j}^p - d_{x_i, x_j}^g| \quad (\text{Eq. 3.24}),$$

where n_p is the size of the search space, and $d_{x_i, x_j}^p, d_{x_i, x_j}^g$ are the phenotypic and genotypic distance, respectively, between two individuals x_i, x_j . The term $\frac{2}{n_p(n_p - 1)}$ is equal to $\frac{1}{\binom{n_p}{2}}$, the proportion of each distinct pair of individuals. This definition “describes how well the phenotypic distance structure is preserved when mapping Φ_p on Φ_g ”, where Φ_p is the phenotypic search space and Φ_g is the genotypic search space (ibid. p. 84).

Similar in spirit, Gottlieb and Raidl [9] also defined a pair of locality metrics for mutation and crossover operators called *mutation innovation* and *crossover innovation*. They additionally defined *crossover loss*, which measures the number of phenotypic properties that are lost by crossover. These metrics are probabilistic and empirical in nature, so they are harder to reason about analytically. But they have been demonstrated in practice to predict relative GEA performance on the multidimensional knapsack problem [5], [19].

In a different study, Chiam et al. defined the concept of *preservation*, which “measure the similarities between the genotype and phenotype search space” [17]. Their study uses Hamming distance between genotypes and L2 norms between phenotypes to define analogous metrics to Rothlauf’s (called *proximity preservation* and *remoteness preservation*), but looking in both directions of the genotype-phenotype mapping. The authors demonstrated empirically (as we prove in the next section), that SB and BRG encodings have the same genotype-to-phenotype locality, but not phenotype-to-genotype locality. They also predicted, based on this similarity, that crossover-based GEAs would perform about the same with both binary and Gray encodings, which contradicts some past empirical results [20].

A different approach to approximating locality was given by Mathias and Whitley [21], and indepen-

¹This definition has changed from the book’s first edition to better match reader intuition and for simpler computation.

dently, by Weicker [22], using the various norms of the phenotype-genotype distance matrix for different representations. The studies did not draw performance predictions from these metrics.

In this study we build on Rothlauf’s definitions because they are both independent from the fitness function, and explicitly computable in our domain. It is worth repeating that our definitions of locality are simply reformulations of Rothlauf’s definitions for the domain of nonredundant binary-integer representations, which we will refer to “representations” for brevity. In the next section we proceed to analyze the theoretical limits of these metrics.

III. RESULTS

A. Definitions

We define a representation $r : \{0, 1\}^\ell \rightarrow [0, 2^\ell)$ as a bijection between the set of ℓ -bit bitstrings $\{0, 1\}^\ell$ and the discrete integer interval $[0, 2^\ell)$. This ensures that the representation is not redundant – i.e., every integer in the interval $[0, 2^\ell)$ is represented by exactly one ℓ -bit bitstring, and the number of search-space points n_p is exactly 2^ℓ . A representation r can therefore be equivalently described as a permutation $\pi : [0, 2^\ell) \rightarrow [0, 2^\ell)$, where $\pi(i) = j$ if and only if the SB representation of i maps to j under r . Consequently, we can write π as a 2^ℓ -tuple, where the i^{th} coordinate (starting at 0) is $\pi(i)$. We also use the notation \hat{s}_i to denote the binary string produced from flipping the i^{th} coordinate of the binary string $s \in \{0, 1\}^\ell$. Formally, $\hat{s}_i = s \oplus 2^i$, where \oplus denotes exclusive-or.

We can now develop an equivalent definition to Rothlauf’s locality that is specific to our domain, using units that we find more intuitive. We define the *point locality* p_r for a nonredundant bitstring-to-integer representation r as the mean change in phenotypic value for a uniformly random single-bit flip in the genotype. More formally:

Definition 1. The point locality p_r for a representation r is $p_r \stackrel{\text{def}}{=} \frac{1}{\ell} \sum_{s \in \{0, 1\}^\ell} |r(\hat{s}_i) - r(s)|$ $\forall i \in [0, \ell)$. Explicitly,

$$p_r \stackrel{\text{def}}{=} \frac{\sum_s \sum_i |r(\hat{s}_i) - r(s)|}{2^\ell \cdot \ell}. \quad (1)$$

Note that for a fixed value of ℓ , our definition of p_r is just a linear transformation of Rothlauf’s d_m . In our domain, d_m simply sums the phenotypic distances minus one between all distinct genotypic neighbors, while p_r computes the average phenotypic distance between all ordered pairs of genotypic neighbors. Note that in Rothlauf’s d_m , d_{min}^p occurs in the summation $\ell \cdot 2^\ell$ times. Coupled with the fact that $d_{min}^p = 1$ in our domain, we have the relationship:

$$p_r = \frac{2d_m + \ell 2^\ell}{\ell 2^\ell} = \frac{d_m}{\ell 2^{\ell-1}} + 1.$$

We also develop an equivalent definition to Rothlauf’s distance distortion, called *general locality*, that is tailored to our domain. We define the general locality g_r for a representation r as the mean difference between phenotypic and genotypic distances between each unique pair of individuals. More formally:

Definition 2.

$$g_r \stackrel{\text{def}}{=} \frac{1}{\binom{2^\ell}{2}} \sum_{(s_1, s_2) \in S_\ell} |d^p(s_1, s_2) - d^g(s_1, s_2)|, \quad (2)$$

where S_ℓ is the set of all unordered pairs of $\{0, 1\}^\ell$ (so $|S_\ell| = \binom{2^\ell}{2}$), $d^p(s_1, s_2)$ is the phenotypic distance $|r(s_1) - r(s_2)|$, and $d^g(s_1, s_2)$ is the genotypic (Hamming) distance between s_1 and s_2 . Note that our definition of general locality mirrors Chiam’s remoteness preservation and is also equivalent to Rothlauf’s d_c metric, since $\sum_{i=0}^{n_p} \sum_{j=i+1}^{n_p}$ is identical to $\sum_{(s_1, s_2) \in S_\ell}$.

It is worth repeating previous observations that high values of d_m , d_c , p_r , or g_r actually denote low locality while low values denote high locality [7]. To avoid confusion, we will refer to low metric values as strong locality and high metric values as weak locality.

B. Single-Bit Mutation

Our analysis of single-bit mutation proves lower and upper bounds on p_r , computes p_r for SB and Gray representations, and verifies the existence of representations with maximum p_r .

Theorem 1 (Lower bound). $p_r \geq \frac{2^\ell - 1}{\ell}$.

Proof. We reduce the problem of minimizing locality to another problem, that of enumerating nodes on a hypercube while minimizing neighbor distances, for which lower and upper bounds have already been established by Harper [23]. We use the term $\Delta_{ij} \stackrel{\text{def}}{=} |i - j|$ to denote the absolute difference between the numbers assigned to two adjacent vertices i and j on the unit ℓ -cube. The unit ℓ -cube consists of all elements in $\{0, 1\}^\ell$. Two vertices in the ℓ -cube are adjacent if they differ by only one bit (i.e., they have Hamming distance 1). Note that assigning a number n to a vertex i can be thought of as a representation r mapping i to n , or $r(i) = n$. Therefore, we have $\Delta_{\hat{s}_i s} = |r(\hat{s}_i) - r(s)|$ for adjacent vertices \hat{s}_i and s , since a 1-bit difference is equivalent to a single bit-flip mutation. Furthermore, Harper defines the sum $\sum \Delta_{\hat{s}_i s}$ to be the sum of the absolute difference between two adjacent vertices \hat{s}_i and s that runs over all possible pairs of neighboring vertices in the ℓ -cube. Note that

$$2 \sum \Delta_{\hat{s}_i s} = \sum_s \sum_i |r(\hat{s}_i) - r(s)|,$$

since the RHS computes $|r(\hat{s}_i) - r(s)|$ twice for every ordered pair.

Harper proved that $\sum \Delta_{\hat{s}_i s} \geq 2^{\ell-1}(2^\ell - 1)$. Therefore,

$$\begin{aligned} p_r &= \frac{\sum_s \sum_i |r(\hat{s}_i) - r(s)|}{2^\ell \cdot \ell} \\ &= \frac{2 \sum \Delta_{\hat{s}_i s}}{2^\ell \cdot \ell} \\ &\geq \frac{2(2^{\ell-1}(2^\ell - 1))}{2^\ell \cdot \ell} \quad (\textit{ibid.}) \\ &= \frac{2^\ell - 1}{\ell}. \end{aligned}$$

□

Corollary 2. Standard binary encoding is optimal, meaning that it has the strongest point locality $\frac{2^\ell - 1}{\ell}$.

Proof. SB encoding is also a representation—call it SB . We consider p_{SB} :

$$p_{SB} = \frac{\sum_s \sum_i |SB(\hat{s}_i) - SB(s)|}{2^\ell \cdot \ell}.$$

The inner sum $\sum_i |SB(\hat{s}_i) - SB(s)|$ computes the sum of all differences obtained from flipping the i^{th} bit of a given SB string s . Flipping the i^{th} bit elicits an absolute phenotypic difference of 2^i for any s and i , reducing the inner sum to:

$$\sum_i |SB(\hat{s}_i) - SB(s)| = \sum_{i=0}^{\ell-1} 2^i = 2^\ell - 1,$$

since it is a geometric series with common ratio 2. Now since there are 2^ℓ elements in $\{0, 1\}^\ell$, the outer sum reduces to $\sum_s (2^\ell - 1) = (2^\ell)(2^\ell - 1)$. Combining these lets us compute p_{SB} :

$$\begin{aligned} p_{SB} &= \frac{\sum_s \sum_i |SB(\hat{s}_i) - SB(s)|}{2^\ell \cdot \ell} \\ &= \frac{2^\ell(2^\ell - 1)}{2^\ell \cdot \ell} \\ &= \frac{2^\ell - 1}{\ell}, \end{aligned}$$

which is the lower bound given by Theorem 1. Thus, SB has optimal point locality. □

Corollary 3. Binary Reflected Gray (BRG) encoding is also optimal.

Proof. Let BRG denote the representation for Binary Reflected Gray and

$$p_{BRG} = \frac{\sum_s \sum_i |BRG(\hat{s}_i) - BRG(s)|}{2^\ell \cdot \ell}.$$

We start by proving the following two lemmas:

Lemma 4. $\sum_{s \in \{0,1\}^\ell} |BRG(\hat{s}_{\ell-1}) - BRG(s)| = 2^{2\ell-1}$.

In other words, the sum of the differences obtained by flipping the leftmost bit over all ℓ -bit bitstrings in BRG encoding is $2^{2\ell-1}$.

Proof. Consider the recursive nature of BRG codes [24]. Let L_ℓ be the ordered list of ℓ -bit BRG codes where $L_\ell[i]$ is the bitstring that maps to i . Note that 2^ℓ is the length of L_ℓ and $[\]$ denotes list indexing. The left half of L_ℓ contains $L_{\ell-1}$ prefixed with 0 and the right half of L_ℓ contains $L_{\ell-1}$ in reverse order prefixed with 1. Flipping the leftmost bit of $L_\ell[i]$ will yield $L_\ell[2^\ell - 1 - i]$. Thus,

$$\begin{aligned} &\sum_{s \in \{0,1\}^\ell} |BRG(\hat{s}_{\ell-1}) - BRG(s)| \\ &= \sum_{i=0}^{2^\ell-1} |BRG(L_\ell[2^\ell - 1 - i]) - BRG(L_\ell[i])| \\ &= \sum_{i=0}^{2^\ell-1} |2^\ell - 1 - i - i| \\ &= \sum_{i=0}^{2^\ell-1} |2^\ell - (2i + 1)|. \end{aligned}$$

Note that for a given $i \in [0, 2^\ell - 1]$,

$$|2^\ell - (2i + 1)| = \begin{cases} 2^\ell - (2i + 1) & i < 2^{\ell-1} \\ (2i + 1) - 2^\ell & i \geq 2^{\ell-1}, \end{cases}$$

which lets us split the sum to

$$\begin{aligned} &= \sum_{i=0}^{2^{\ell-1}-1} (2^\ell - (2i + 1)) + \sum_{i=2^{\ell-1}}^{2^\ell-1} ((2i + 1) - 2^\ell) \\ &= \sum_{i=0}^{2^{\ell-1}-1} 2^\ell - \sum_{i=0}^{2^{\ell-1}-1} (2i + 1) + \sum_{i=2^{\ell-1}}^{2^\ell-1} (2i + 1) - \sum_{i=2^{\ell-1}}^{2^\ell-1} 2^\ell \\ &= (2^{\ell-1})(2^\ell) - \sum_{i=0}^{2^{\ell-1}-1} (2i + 1) + \sum_{i=2^{\ell-1}}^{2^\ell-1} (2i + 1) - (2^{\ell-1})(2^\ell) \\ &= \sum_{i=2^{\ell-1}}^{2^\ell-1} (2i + 1) - \sum_{i=0}^{2^{\ell-1}-1} (2i + 1). \end{aligned}$$

Splitting the left sum and using the facts that the sum of the first n odd numbers is n^2 ,

$$\begin{aligned} &= \left(\sum_{i=0}^{2^{\ell-1}-1} (2i + 1) - \sum_{i=0}^{2^{\ell-1}-1} (2i + 1) \right) - (2^{\ell-1})^2 \\ &= (2^\ell)^2 - (2^{\ell-1})^2 - (2^{\ell-1})^2 \\ &= 2^{2\ell} - 2 \cdot 2^{2\ell-2} \\ &= 2^{2\ell} - 2^{2\ell-1} \\ &= 2^{2\ell-1}. \end{aligned}$$

Thus $\sum_{s \in \{0,1\}^\ell} |BRG(\hat{s}_{\ell-1}) - BRG(s)| = 2^{2\ell-1}$. □

Lemma 5. $\sum_s \sum_i |BRG(\hat{s}_i) - BRG(s)| = 2^{2^\ell} - 2^\ell$.

Proof. We proceed with induction on ℓ . For the base case ($\ell = 1$), the set $\{0, 1\}^1$ contains two BRG codes, $\{0, 1\}$, which corresponds to the integers 0 and 1, respectively. Thus $\sum_{s \in \{0, 1\}^1} \sum_{i=0}^{1-1} |BRG(\hat{s}_i) - BRG(s)| = 1 + 1 =$

$$2 = 2^{2^1} - 2^1.$$

For the inductive hypothesis (I.H.), assume for some $\ell \in \mathbb{N}$ that $\sum_s \sum_i |BRG(\hat{s}_i) - BRG(s)| = 2^{2^\ell} - 2^\ell$. We must now show that $\sum_{s \in \{0, 1\}^{\ell+1}} \sum_{i=0}^{\ell} |BRG(\hat{s}_i) - BRG(s)| = 2^{2^{\ell+1}} - 2^{\ell+1}$. Note that in the inductive step, we are working with strings of length $\ell + 1$.

$$\begin{aligned} \sum_{s \in \{0, 1\}^{\ell+1}} \sum_{i=0}^{\ell} |BRG(\hat{s}_i) - BRG(s)| &= \sum_{s \in \{0, 1\}^{\ell+1}} \left(|BRG(\hat{s}_\ell) - BRG(s)| + \sum_{i=0}^{\ell-1} |BRG(\hat{s}_i) - BRG(s)| \right) \\ &= \sum_{s \in \{0, 1\}^{\ell+1}} \sum_{i=0}^{\ell-1} |BRG(\hat{s}_i) - BRG(s)| + \sum_{s \in \{0, 1\}^{\ell+1}} |BRG(\hat{s}_\ell) - BRG(s)|. \end{aligned}$$

By I.H. and the fact that there are two copies of the ℓ -bit BRG code in the $(\ell + 1)$ -bit BRG code,

$$= 2 \cdot (2^{2^\ell} - 2^\ell) + \sum_{s \in \{0, 1\}^{\ell+1}} |BRG(\hat{s}_\ell) - BRG(s)|.$$

By Lemma 4,

$$\begin{aligned} &= 2 \cdot (2^{2^\ell} - 2^\ell) + 2^{2^{\ell+1}-1} \\ &= 2^{2^\ell+1} - 2^{\ell+1} + 2^{2^\ell+1} \\ &= 2 \cdot 2^{2^\ell+1} - 2^{\ell+1} \\ &= 2^{2^{\ell+1}} - 2^{\ell+1}. \end{aligned}$$

□

Now we can prove Corollary 3. Considering p_{BRG} :

$$\begin{aligned} p_{BRG} &= \frac{\sum_s \sum_i |BRG(\hat{s}_i) - BRG(s)|}{2^\ell \cdot \ell} \\ &= \frac{2^{2^\ell} - 2^\ell}{2^\ell \cdot \ell} \quad \text{Lemma 5} \\ &= \frac{2^\ell(2^\ell - 1)}{2^\ell \cdot \ell} \\ &= \frac{2^\ell - 1}{\ell}, \end{aligned}$$

which is the lower bound given by Theorem 1. □

Note that this equivalence in locality has already been demonstrated empirically for small values of ℓ [17], but this proof holds for all values of ℓ .

Claim 6. There exists a Gray encoding g with suboptimal point locality $p_g > \frac{2^\ell - 1}{\ell}$ for any $\ell \geq 3$.

Proof. Recall the ℓ -cube, $\{0, 1\}^\ell$, that contains all ℓ -bit bitstrings. A bitstring $b \in \{0, 1\}^\ell$ has ℓ neighbors that are all Hamming distance one away. We can therefore reduce the problem of constructing a Gray code to constructing a Hamiltonian path on the ℓ -cube. Recall in Theorem 1

that we mapped the problem of minimizing locality to that of minimizing $\sum \Delta_{\hat{s}_i s}$. In the same paper, Harper [23] formulates an algorithm that provably generates all representations that minimize $\sum \Delta_{\hat{s}_i s}$, as follows:

1. Assign 0 to a random vertex.
2. For i from 1 to $2^\ell - 1$, assign i to the vertex with the highest number of already labeled neighbors. If there are multiple vertices with this property, choose one at random.

Our goal is to construct a Hamiltonian path that violates this algorithm, which in turn will have suboptimal point locality, because Harper's algorithm generates all representations with optimal point locality.

Our modified algorithm starts by assigning 0 to vertex $0^{\ell-3}000$, where the notation $0^{\ell-3}$ denotes an $\ell-3$ length bitstring of all 0s. We then assign 1 to $0^{\ell-3}001$ and assign 2 to $0^{\ell-3}011$. The above algorithm would force us to assign 3 to $0^{\ell-3}010$ if we wanted to produce an optimal Gray code. Instead, we assign 3 to $0^{\ell-3}111$, which violates the algorithm. Assuming $\ell \geq 3$, we can complete the remainder of the Hamiltonian path from $0^{\ell-3}111$ to $0^{\ell-3}000$ with the remaining edges [25], thus constructing a Gray code g with $p_g > \frac{2^\ell - 1}{\ell}$. □

Theorem 7 (Upper bound). $p_r \leq 2^{\ell-1}$.

Proof. We use another result from Harper [23], where he proved that $\sum \Delta_{\hat{s}_i s} \leq \ell 2^{2^{\ell-1}}$. We have

$$\begin{aligned} p_r &= \frac{\sum_s \sum_i |r(\hat{s}_i) - r(s)|}{2^\ell \cdot \ell} \\ &= \frac{2 \sum \Delta_{\hat{s}_i s}}{2^\ell \cdot \ell} \\ &\leq \frac{2(\ell 2^{2^{\ell-1}})}{2^\ell \cdot \ell} \quad (\textit{ibid.}) \\ &= \frac{2^{2^{\ell-1}}}{2^{\ell-1}} \\ &= 2^{\ell-1} \end{aligned}$$

$$= 2^{\ell-1}.$$

Thus $p_r \leq 2^{\ell-1}$. \square

Claim 8. There exists a representation r with upper bound point locality $p_r = 2^{\ell-1}$.

Proof. In this proof, we reduce the problem of constructing a representation r with upper bound locality $p_r = 2^{\ell-1}$ to that of assigning integers in $[0, 2^\ell)$ to vertices in the ℓ -cube such that $\sum \Delta_{\hat{s}_i s}$ is maximized. Harper [23] also demonstrates an algorithm assigning numbers to vertices that provably maximizes $\sum \Delta_{\hat{s}_i s}$, as follows:

1. Assign 0 to a random vertex. Let n be the number of '1' bits in that vertex.
2. Assign the integers $\{1, \dots, 2^{\ell-1} - 1\}$ randomly to vertices (bitstrings) whose number of constituent 1s have the same parity as n .
3. Assign the remaining numbers $\{2^{\ell-1}, \dots, 2^\ell - 1\}$ to the leftover vertices at random. These are the vertices whose number of constituent 1s have the opposite parity as n .

Maximizing $\sum \Delta_{\hat{s}_i s}$ is equivalent to maximizing p_r , so such a representation exists. \square

Having explored the properties of point locality for single-bit mutations, we turn our attention to general locality and distance distortion for any variation operator.

C. General Locality

We begin our analysis of general locality by proving a lower bound on its value, and continue by proving the asymptotic equivalence of all nonredundant binary integer representations under this metric.

Theorem 9. $g_r \geq \frac{1}{\binom{2^\ell}{2}} \left(\frac{1}{6} (2^\ell - 1)(2^\ell)(2^\ell + 1) - \ell 2^{2(\ell-1)} \right)$.

Proof. $g_r \stackrel{\text{def}}{=} \frac{1}{\binom{2^\ell}{2}} \sum_{(s_1, s_2) \in S_\ell} |d^p(s_1, s_2) - d^g(s_1, s_2)|$ (by definition). Using the triangle inequality, we get

$$\begin{aligned} g_r &\geq \frac{1}{\binom{2^\ell}{2}} \sum_{(s_1, s_2) \in S_\ell} (d^p(s_1, s_2) - d^g(s_1, s_2)) \\ &= \frac{1}{\binom{2^\ell}{2}} \left(\sum_{(s_1, s_2) \in S_\ell} d^p(s_1, s_2) - \sum_{(s_1, s_2) \in S_\ell} d^g(s_1, s_2) \right) \\ &= \frac{1}{\binom{2^\ell}{2}} (P - G), \end{aligned}$$

where we let $P = \sum_{(s_1, s_2) \in S_\ell} d^p(s_1, s_2)$ and $G = \sum_{(s_1, s_2) \in S_\ell} d^g(s_1, s_2)$. Since P only deals with phenotypes in \mathbb{N} , it is equivalent to

$$P = \sum_{i=0}^{2^\ell-1} \sum_{j=i+1}^{2^\ell-1} (j - i).$$

Let the outer sum fix i . The inner sum computes the sum of numbers from 1 to $2^\ell - 1 - i$. This reduces P to

$$= \sum_{i=1}^n i + \sum_{i=1}^{n-1} i + \dots + \sum_{i=1}^1 i$$

where we let $n = 2^\ell - 1$ for simplicity. Using the facts that the sum of the first m natural numbers is $\frac{1}{2}m(m+1)$ and the sum of the first m squares is $\frac{1}{6}m(m+1)(2m+1)$, we have

$$\begin{aligned} &= \frac{1}{2} \sum_{i=1}^n i(i+1) \\ &= \frac{1}{2} \left(\sum_{i=1}^n i^2 + \sum_{i=1}^n i \right) \\ &= \frac{1}{2} \left(\frac{1}{6} n(n+1)(2n+1) + \frac{1}{2} n(n+1) \right) \\ &= \frac{n(n+1)}{4} \left(\frac{2n+1}{3} + 1 \right) \\ &= \frac{n(n+1)}{4} \left(\frac{2n+4}{3} \right) \\ &= \frac{1}{6} n(n+1)(n+2). \end{aligned}$$

Substituting $n = 2^\ell - 1$ back into the equation yields:

$$= \frac{1}{6} (2^\ell - 1)(2^\ell)(2^\ell + 1).$$

Since G is the sum of the Hamming distances between all unique pairs of bitstrings, it is equivalent to

$$G = \frac{1}{2} 2^\ell \sum_{i=1}^{\ell} i \binom{\ell}{i},$$

because for each of the 2^ℓ bitstrings, a bitstring has $\binom{\ell}{i}$ other bitstrings with Hamming distance i (choose i of the ℓ bits to be flipped). We divide by two because we count each pair twice. We can simplify G , obtaining

$$\begin{aligned} &= 2^{\ell-1} \sum_{i=1}^{\ell} i \cdot \frac{\ell}{i} \binom{\ell-1}{i-1} \\ &= \ell 2^{\ell-1} \sum_{i=1}^{\ell} \binom{\ell-1}{i-1} \\ &= \ell 2^{\ell-1} 2^{\ell-1} \\ &= \ell 2^{2(\ell-1)}. \end{aligned}$$

Substituting P and G back into g_r yields

$$\begin{aligned} g_r &\geq \frac{1}{\binom{2^\ell}{2}} (P - G) \\ &= \frac{1}{\binom{2^\ell}{2}} \left(\frac{1}{6} (2^\ell - 1)(2^\ell)(2^\ell + 1) - \ell 2^{2(\ell-1)} \right). \end{aligned}$$

\square

Theorem 10. $g_r \sim \frac{1}{\binom{2^\ell}{2}} (\frac{1}{6}(2^\ell - 1)(2^\ell)(2^\ell + 1) - \ell 2^{2(\ell-1)})$ for any representation r on ℓ bits. That is, as ℓ grows, the value of g_r is independent of the actual representation.

Proof. The key intuition behind this proof is that for nonredundant binary-integer representations, as ℓ grows, the phenotypic distances grow at an asymptotically greater rate than the genotypic distances. We can separate the phenotypic distances from the genotypic distances by partitioning S_ℓ into two sets $S_\ell = S_\ell^p \sqcup S_\ell^g$, where \sqcup denotes disjoint union:

$$S_\ell^p = \{(s_1, s_2) \in S_\ell | d^p(s_1, s_2) > d^g(s_1, s_2)\},$$

$$S_\ell^g = \{(s_1, s_2) \in S_\ell | d^p(s_1, s_2) \leq d^g(s_1, s_2)\}.$$

In other words, S_ℓ^p contains all the pairs in S_ℓ where the two bitstrings have greater phenotypic (Euclidean) distance than genotypic (Hamming) distance, and S_ℓ^g contains all pairs where the two bitstrings have greater or equal genotypic distance than phenotypic distance. We can rewrite g_r as (letting $C = \frac{1}{\binom{2^\ell}{2}}$)

$$g_r = \frac{1}{\binom{2^\ell}{2}} \sum_{(s_1, s_2) \in S_\ell} |d^p(s_1, s_2) - d^g(s_1, s_2)|$$

$$= C(P(\ell) + G(\ell)),$$

where we let

$$P(\ell) = \sum_{(s_1, s_2) \in S_\ell^p} (d^p(s_1, s_2) - d^g(s_1, s_2)),$$

$$G(\ell) = \sum_{(s_1, s_2) \in S_\ell^g} (d^g(s_1, s_2) - d^p(s_1, s_2)).$$

Note that $|S_\ell^g| \leq \frac{1}{2}(2\ell - 1)2^\ell$ since each of the 2^ℓ bitstrings can have at most $2\ell - 1$ bitstrings for which their genotypic distance is greater or equal to their phenotypic distance. This is because for an integer $i \geq \ell$, the integers between $i - \ell$ and $i + \ell$ can be represented by bitstrings with genotypic distances greater than phenotypic distances $j \leq \ell$, where we subtract 1 due to the fact that each bitstring only has one other bitstring with Hamming distance ℓ . We then divide by two since we count each pair twice. Thus $|S_\ell^g| = \mathcal{O}(\ell 2^\ell)$. We can now say $|S_\ell^p| = |S_\ell| - |S_\ell^g| \geq \binom{2^\ell}{2} - \frac{1}{2}(2\ell - 1)2^\ell = \frac{2^\ell(2^\ell - 2\ell)}{2}$, so $|S_\ell^p| = \Omega(2^{2\ell})$. Consider

$$\lim_{\ell \rightarrow \infty} \frac{|S_\ell|}{|S_\ell^p|} = \lim_{\ell \rightarrow \infty} \frac{|S_\ell^p| + |S_\ell^g|}{|S_\ell^p|}$$

$$= \lim_{\ell \rightarrow \infty} \left(1 + \frac{|S_\ell^g|}{|S_\ell^p|} \right)$$

$$= 1 + 0 = 1,$$

since $|S_\ell^p| = \Omega(2^{2\ell})$ grows faster than $|S_\ell^g| = \mathcal{O}(\ell 2^\ell)$. Thus $|S_\ell^p|$ dominates $|S_\ell^g|$, and $|S_\ell| \sim |S_\ell^p|$.

We can now perform a similar analysis for $P(\ell)$ and $P(\ell) + G(\ell)$. Note that $G(\ell) \leq |S_\ell^g|(\ell - 1) \leq \frac{1}{2}\ell^2 2^\ell - \frac{1}{2}\ell 2^\ell$ since any pair in S_ℓ^g can have a maximum $d^g - d^p$ of $\ell - 1$. Thus $G(\ell) = \mathcal{O}(\ell^2 2^\ell)$. Also note that $P(\ell) \geq |S_\ell^p| \geq \frac{2^\ell(2^\ell - \ell - 1)}{2}$ since any pair in S_ℓ^p can have a minimum $d^p - d^g$ of 1. Thus $P(\ell) = \Omega(2^{2\ell})$. Consider

$$\lim_{\ell \rightarrow \infty} \frac{P(\ell) + G(\ell)}{P(\ell)} = \lim_{\ell \rightarrow \infty} 1 + \frac{G(\ell)}{P(\ell)}$$

$$= 1 + 0 = 1,$$

since $P(\ell) = \Omega(2^{2\ell})$ grows faster than $G(\ell) = \mathcal{O}(\ell^2 2^\ell)$, and so $P(\ell) + G(\ell) \sim P(\ell)$. Now we can make a statement about g_r . We have

$$\frac{1}{C} g_r = P(\ell) + G(\ell)$$

$$\lim_{\ell \rightarrow \infty} \frac{g_r}{CP(\ell)} = \lim_{\ell \rightarrow \infty} \frac{P(\ell) + G(\ell)}{P(\ell)}$$

$$\lim_{\ell \rightarrow \infty} \frac{g_r}{CP(\ell)} = 1.$$

Thus $\frac{g_r}{C} \sim P(\ell)$. Since $|S_\ell^p| \sim |S_\ell|$ and $CP(\ell) \sim g_r$, we replace S_ℓ^p with S_ℓ in the $P(\ell)$ summation to obtain (recall $C = \frac{1}{\binom{2^\ell}{2}}$)

$$g_r \sim \frac{1}{\binom{2^\ell}{2}} \sum_{(s_1, s_2) \in S^\ell} (d^p(s_1, s_2) - d^g(s_1, s_2))$$

$$= \frac{1}{\binom{2^\ell}{2}} \left(\sum_{(s_1, s_2) \in S^\ell} d^p(s_1, s_2) - \sum_{(s_1, s_2) \in S^\ell} d^g(s_1, s_2) \right)$$

$$= \frac{1}{\binom{2^\ell}{2}} \left(\frac{1}{6}(2^\ell - 1)(2^\ell)(2^\ell + 1) - \ell 2^{2(\ell-1)} \right),$$

which was found in the proof of Theorem 9. Thus g_r and $\frac{1}{\binom{2^\ell}{2}} (\frac{1}{6}(2^\ell - 1)(2^\ell)(2^\ell + 1) - \ell 2^{2(\ell-1)})$ are asymptotically equal for any representation r . \square

IV. CONCLUSION AND FUTURE WORK

Various properties of GEA representations have been studied to help explain GEA behaviors, such as the number of induced local optima or the existence of Hamming cliffs. Among these properties, locality is particularly interesting to analyze, because it is independent of the fitness landscape and can be computed precisely, which is not always the case for other properties.

We make the distinction between point locality, which estimates the sensitivity of phenotypic values to single-bit mutations in the genotype, and generalized locality, which estimates the same sensitivity under more disruptive genotypic operators, such as crossover. By formally deriving tight bounds on point locality, we can compare directly the effects of different binary representation on

locality. For example, we can prove that no representation has “better” (lower) locality than standard binary encoding or binary-reflected Gray encoding, but that some other Gray encodings have worse locality.

Moreover, we can easily construct representations of both “optimal” and “pessimal” representations. The former offers more control over the GEA simulation, because small changes in the genotype lead to small changes in the phenotype, a desirable trait for the exploitation phase of the GEA. On the other hand, a representation with poor locality is closer to random search, which may be more desirable for the exploration phase of the GEA, or conversely, for fixed-budget GEAs that need to cover a large subset of the search space with as few computational resources as possible.

Both Rothlauf’s locality metric and our own equivalent point locality focus on the single-bit mutation operator. Future work could explore other locality metrics or look at different operators, and perhaps combine them with a crossover operator to expand on distance distortion. In addition, we proved that generalized locality is not a useful metric for nonredundant binary-integer representations, because it quickly converges to the same value, regardless of representation. Future work could compute the bounds for crossover locality metrics that are better suited to describe this representation domain. Another path for future research might be to use other statistics as estimators of locality that grow more slowly with the bitstring length, such as the minimum phenotypic distance, maximum, or standard deviation. Finally, we also plan to analyze the expected locality of random representations to better understand the effects on the mutation operator of changing representations dynamically in adaptive GEAs.

REFERENCES

- [1] T. Mitchell-Olds, J. H. Willis, and D. B. Goldstein, “Which evolutionary processes influence natural genetic variation for phenotypic traits?” *Nature Reviews Genetics*, vol. 8, no. 11, p. 845, 2007.
- [2] D. E. Goldberg, K. Deb, and J. H. Clark, “Genetic algorithms, noise, and the sizing of populations,” *Complex Systems*, vol. 6, pp. 333–333, 1992.
- [3] T. Bäck, D. B. Fogel, and Z. Michalewicz, *Evolutionary computation 1: Basic algorithms and operators*. CRC press, 2018.
- [4] M. Črepinšek, S.-H. Liu, and M. Mernik, “Exploration and exploitation in evolutionary algorithms: A survey,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, p. 35, 2013.
- [5] F. B. Pereira, J. M. Marques, T. Leitão, and J. Tavares, “Analysis of locality in hybrid evolutionary cluster optimization,” in *Proceedings of the IEEE International Conference on Evolutionary Computation*. IEEE, 2006, pp. 2285–2292.
- [6] E. Galván-López, J. McDermott, M. O’Neill, and A. Brabazon, “Defining locality as a problem difficulty measure in genetic programming,” *Genetic Programming and Evolvable Machines*, vol. 12, no. 4, pp. 365–401, 2011.
- [7] F. Rothlauf, *Representations for Genetic and Evolutionary Algorithms*, 2nd ed. Heidelberg, Germany: Springer, 2006.
- [8] J. Gottlieb, B. A. Julstrom, G. R. Raidl, and F. Rothlauf, “Prüfer numbers: A poor representation of spanning trees for evolutionary search,” in *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO’01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 343–350. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2955239.2955297>
- [9] J. Gottlieb and G. R. Raidl, “Characterizing locality in decoder-based eas for the multidimensional knapsack problem,” in *European Conference on Artificial Evolution*. Springer, 1999, pp. 38–52.
- [10] T. Jones and S. Forrest, “Fitness distance correlation as a measure of problem difficulty,” in *Proceedings of the 6th International Conference on Genetic Algorithms*, 1995, pp. 184–192.
- [11] B. Manderick, “The genetic algorithm and the structure of fitness landscape,” in *Proceedings of the 4th International Conference on Genetic Algorithms*, 1991, pp. 143–150.
- [12] S. Ronald, “Robust encodings in genetic algorithms: A survey of encoding issues,” in *Proceedings of IEEE international conference on Evolutionary Computation*. IEEE, 1997, pp. 43–48.
- [13] F. Rothlauf and M. Oetzel, “On the locality of grammatical evolution,” in *European Conference on Genetic Programming*. Springer, 2006, pp. 320–330.
- [14] E. Galván-López and M. O’Neill, “On the effects of locality in a permutation problem: The sudoku puzzle,” in *Proceedings of the IEEE Symposium on Computational Intelligence and Games*. IEEE, 2009, pp. 80–87.
- [15] N. X. Hoai, R. I. McKay, and D. Essam, “Representation and structural difficulty in genetic programming,” *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 2, pp. 157–166, 2006.
- [16] F. Rothlauf and D. Goldberg, “Tree network design with genetic algorithms—an investigation in the locality of the pruefernumber encoding,” in *Late Breaking Papers at the Genetic and Evolutionary Computation Conference*, 1999, pp. 238–244.
- [17] S. Chiam, C. Goh, and K. Tan, “Issues of binary representation in evolutionary algorithms,” in *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*. IEEE, 2006, pp. 1–8.
- [18] E. Galván-López, J. McDermott, M. O’Neill, and A. Brabazon, “Towards an understanding of locality in genetic programming,” in *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*. ACM, 2010, pp. 901–908.
- [19] G. R. Raidl and J. Gottlieb, “Empirical analysis of locality, heritability and heuristic bias in evolutionary algorithms: A case study for the multidimensional knapsack problem,” *Evolutionary Computation*, vol. 13, no. 4, pp. 441–475, 2005.
- [20] R. Hinterding, H. Gielewski, and T. C. Peachey, “The nature of mutation in genetic algorithms,” in *Proceedings of the 6th International Conference on Genetic Algorithms*, 1995, pp. 65–72.
- [21] K. E. Mathias and L. D. Whitley, “Transforming the search space with gray coding,” in *Proceedings of the First IEEE Conference on Evolutionary Computation*. IEEE, 1994, pp. 513–518.
- [22] K. Weicker, “A binary encoding supporting both mutation and recombination,” in *Proceedings of the International Conference on Parallel Problem Solving from Nature*. Springer, 2010, pp. 134–143.
- [23] L. H. Harper, “Optimal assignments of numbers to vertices,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 12, no. 1, pp. 131–135, 1964.
- [24] J. Rowe, D. Whitley, L. Barbulescu, and J.-P. Watson, “Properties of Gray and binary representations,” *Evolutionary computation*, vol. 12, no. 1, pp. 47–76, 2004.
- [25] T. Dvořák and P. Gregor, “Hamiltonian paths with prescribed edges in hypercubes,” *Discrete Mathematics*, vol. 307, no. 16, pp. 1982 – 1998, 2007.