Fitness Function Distributions over Generalized Search Neighborhoods in the q-ary Hypercube

Andrew M. Sutton   sutton@cs.colostate.edu
Department of Computer Science, Colorado State University, USA

Francisco Chicano   chicano@lcc.uma.es
Department of Lenguajes y Ciencias de la Computación, University of Málaga, Spain

L. Darrell Whitley   whitley@cs.colostate.edu
Department of Computer Science, Colorado State University, USA

Abstract

The frequency distribution of a fitness function over regions of its domain is an important quantity for understanding the behavior of algorithms that employ randomized sampling to search the function. In general, exactly characterizing this distribution is at least as hard as the search problem, since the solutions typically live in the tails of the distribution. However, in some cases it is possible to efficiently retrieve a collection of quantities (called *moments*) that describe the distribution.

In this paper, we consider functions of bounded epistasis that are defined over length-n strings from a finite alphabet of cardinality q. Many problems in combinatorial optimization can be specified as search problems over functions of this type. Employing Fourier analysis of functions over finite groups, we derive an efficient method for computing the exact moments of the frequency distribution of fitness functions over Hamming regions of the q-ary hypercube. We then use this approach to derive equations that describe the expected fitness of the offspring of any point undergoing uniform mutation.

The results we present provide insight into the statistical structure of the fitness function for a number of combinatorial problems. For the graph coloring problem, we apply our results to efficiently compute the average number of constraint violations that lie within a certain number of steps of any coloring. We derive an expression for the mutation rate that maximizes the expected fitness of an offspring at each fitness level. We also apply the results to the slightly more complex frequency assignment problem, a relevant application in the domain of the telecommunications industry. As with the graph coloring problem, we provide formulas for the average value of the fitness function in Hamming regions around a solution and the expectation-optimal mutation rate.

1   Introduction

Many problems in combinatorial optimization, such as graph coloring, register allocation and frequency assignment problems, can be characterized as a search for the

extremum of a real-valued function over the set of length-n strings over a q-ary alphabet. Local search algorithms and mutation-based evolutionary algorithms for such problems can be viewed as processes that perform a biased random walk on the q-ary hypercube graph. Characterizing the frequency distribution of fitness values over regions of this graph is thus an important part of understanding the behavior of such algorithms and the influence of problem structure on this behavior.

In this paper we derive a method for the efficient calculation of exact statistical moments of the frequency distribution of fitness function values over Hamming regions of the q-ary hypercube provided that the function is bounded epistatically by a constant. The retrieval of these moments provides insight into the local structure of the search space and can be used to describe the fitness of states that lie within a certain distance of any given state. Our aim is to provide a framework for studying the statistical structure of local regions of the search space for a broad class of combinatorial optimization problems. This framework is directly relevant to the theory of randomized search heuristics such as evolutionary algorithms, simulated annealing, and variants of randomized local search (hill-climbing). For the graph coloring problem and the frequency assignment problem, we derive specific equations that model the fitness statistics of states sampled by the move and mutation operators of such algorithms.

We present a direct computation of the exact statistics (e.g. mean, variance, and skew) of the distribution of fitness values over regions of the search space for problems over strings. As long as the epistasis of the fitness function is bounded by a constant, our method has polynomial complexity even when the cardinality of the region grows as an exponential in n. Subsequently, we use this result to determine an exact calculation of the *expected fitness* of an offspring parameterized by mutation rate from any given point. Aside from working out the statistics of the fitness of an offspring of a state undergoing mutation in an evolutionary algorithm, it also allows one to compute exactly the rate that results in the maximal expected fitness at any point. We see that the resulting expression illuminates the behavior of mutation on combinatorial problems such as graph coloring and the frequency assignment problem. Currently, our results are mainly of theoretical interest, however we also discuss some different ways they could eventually be used in practice to design more principled algorithms.

This paper generalizes the work of Sutton et al. (2012) who presented a polynomial-time algorithm for computing the moments of pseudo-Boolean functions over regions in the Boolean hypercube, i.e., for $q = 2$. We will generalize this result to complex, epistatically bounded functions over strings from a finite alphabet of q elements where q is $O(1)$ and the corresponding adjacency is the generalized Hamming graph (i.e., the q-ary hypercube). This paper also generalizes a result of Heckendorn (2002) who derived a polynomial-time algorithm for obtaining the summary statistics of the fitness function distribution over the entire search space when $q = 2$.

The remainder of the paper is organized as follows. In the next section, we introduce some technical preliminaries and show that fitness functions over q-ary strings can be written in a Fourier function basis expansion. In Section 3, we relate this Fourier analysis to the Hamming adjacency by proving that the Fourier basis functions are eigenfunctions of generalized Hamming adjacency matrices. This allows us to closely study the statistical relationship between the fitness function and Hamming regions in the q-ary hypercube. In Section 4 we apply these results to obtain an efficient algorithm for computing moments of the fitness distribution over arbitrary Hamming regions. We then turn our attention to uniform mutation in the q-ary hypercube. Specifically, in Section 5 we apply our results to derive equations that express the expected fitness of

an offspring of an arbitrary state undergoing mutation as a function of rate. We then show how the results introduced in this paper can be immediately applied to two specific combinatorial problems, graph coloring and the frequency assignment problem in Sections 6 and 7, respectively. Finally, in Section 8 we conclude the paper.

2   Fourier Analysis of Fitness Functions over Σ_q^n

Let Σ_q denote an alphabet of cardinality q. We are interested in functions over Σ_q^n the set of length-n strings over Σ_q. Combinatorial optimization problems typically have a real-valued fitness function. We will be slightly more general and work with complex functions over Σ_q^n. This makes the analysis somewhat more natural.

Let f be a complex function over Σ_q^n. Without loss of generality, we associate the symbols of the alphabet with an element of $\mathbb{Z}_q$: the cyclic group of integers under addition modulo q. We can then appeal to tools from group theory to express f in its group-theoretic Fourier expansion. Let G be the finite Abelian group

$$G = \underbrace{\mathbb{Z}_q \oplus \mathbb{Z}_q \oplus \cdots \oplus \mathbb{Z}_q}_{n}, \tag{1}$$

where addition is component-wise. This is the n-way direct product[1] of cyclic groups of order q. Since we have associated the symbols of the q-ary alphabet with the elements of the cyclic group $\mathbb{Z}_q$ (which have additional algebraic structure), we can, without loss of generality, express f as a function over G. The set of all complex functions $\{f : G \to \mathbb{C}\}$ is an inner product space with operation

$$\langle f, g \rangle = \sum_{x \in G} f(x)\overline{g(x)},$$

where the overline denotes complex conjugation. This inner product space has a complete orthonormal Fourier basis given by the set of functions $\{\varphi_a : a \in G\}$, where

$$\varphi_a(x) = \prod_{i=1}^{n} \exp\left(\frac{2\pi\sqrt{-1}a[i]x[i]}{q} \right).$$

The notation $x[i]$ denotes the i-th element of the vector $x \in G$. When $q = 2$, this is equivalent to the Walsh basis expansion (Rana et al., 1998; Heckendorn, 1999; Sutton et al., 2012). In that particular case, the functions are real-valued (since the 2's cancel and $\exp(\pi\sqrt{-1}) = -1$). However, when $q > 2$, the Fourier functions become complex.

For simplicity, we will express the above complex exponential function as a primitive q-th *root of unity*. A complex number z is a q-th *root of unity* when z satisfies the equation $z^q = 1$. Furthermore, a q-th root of unity is called *primitive* when there is no such $0 < k < q$ such that $z^k = 1$. In this paper, we will often appeal to the following simple fact entailed by the properties of a geometric series: if z is a q-th root of unity for $q > 1$, then

$$\sum_{\ell=0}^{q-1} z^\ell = \frac{z^q - 1}{z - 1} = 0. \tag{2}$$

---

[1]The direct product of groups is defined as the Cartesian product of the groups under element-wise group operation. In other words, given two groups $(X, \bullet)$ and $(Y, \star)$, an element $(x, y) \times (x', y') \in (X, \bullet) \oplus (Y, \star)$ is given by $(x \bullet x', y \star y')$.

We define
$$\omega_q = \exp\left(\frac{2\pi\sqrt{-1}}{q}\right).$$

Clearly ω_q is a primitive q-th root of unity. The Fourier functions can thus be expressed in terms of this primitive root of unity:

$$\varphi_a(x) = \prod_{i=1}^{n} \omega_q^{a[i]x[i]}.$$

2.1 Some Useful Properties of Fourier Analysis

As we have already seen, any complex function $f : G \to \mathbb{C}$ belongs to the inner product space of functions spanned by the orthogonal Fourier basis; we can write f in its Fourier basis expansion

$$f(x) = \sum_{a \in G} z_a \varphi_a(x). \tag{3}$$

Furthermore, since φ_a also lives in the space of complex functions over G, it is straightforward to show that

$$z_a = \frac{1}{|G|}\langle f, \varphi_a \rangle. \tag{4}$$

One useful property is that, as in the special case of the Walsh coefficients, the zeroth Fourier function is constant. Denote as φ_0 the function corresponding to the zero element $a = (0, 0, \ldots, 0)$. Then

$$\varphi_0(x) = \prod_{i=1}^{n} \omega_q^{x[i]\cdot 0} = \prod_{i=1}^{n} 1 = 1. \tag{5}$$

This implies that z_0 is the average value of f over G:

$$z_0 = \frac{1}{|G|}\langle f, \varphi_0 \rangle = \frac{1}{|G|}\sum_{x \in G} f(x) = \operatorname*{avg}_{x \in G}\{f(x)\}. \tag{6}$$

We also have the following orthogonality property.

$$\sum_{x \in G} \varphi_a(x) = \begin{cases} |G| & \text{if } a = (0, 0, \ldots, 0); \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

This is easy to prove by straightforward manipulation, for example, see Terras (1999, page 169).

In this paper, we will often take advantage of the Fourier spectral properties of fitness functions that have bounded epistasis. We now make this concept rigorous.

Definition. *A function $f : G \to \mathbb{C}$ is* epistatically bounded by k *if it can be expressed as the sum of functions that each depend on at most k positions of their argument.*

Epistatically bounded fitness functions arise from combinatorial problems that are structures built from components of bounded size. For example, the fitness function for many graph-theoretic problems can be expressed as the sum of sub-functions over each edge.

In the lemma that follows, we will make brief use of the following concept of *sequence projection*. We will denote by $|a|$ the number of nonzero elements of $a \in G$.

Definition. *The* projection *of a length-n sequence x onto a length-n sequence b is the length-$|b|$ sequence*

$$\operatorname{proj}_b(x) = (x[i_1], x[i_2], \ldots, x[i_{|b|}])$$

where i_j is the position of the j-th nonzero element of b.

We are now ready to prove the following lemma.

Lemma 1. *Let $a, b \in G$. Suppose $g_b : G \to \mathbb{C}$ is a function such that, for all $x, y \in G$,*

$$\operatorname{proj}_b(x) = \operatorname{proj}_b(y) \implies g_b(x) = g_b(y).$$

Then we have

$$|a| > |b| \implies \langle g_b, \varphi_a \rangle = 0.$$

Proof. By definition of the inner product,

$$\langle g_b, \varphi_a \rangle = \sum_{x \in G} g_b(x)\overline{\varphi_a}(x)$$

Since $|a| > |b|$ there must be some element u such that $a[u] \neq 0$ and $b[u] = 0$. We can rewrite the inner product as

$$\langle g_b, \varphi_a \rangle = \sum_{j=0}^{q-1} \sum_{x : x[u]=j} g_b(x) \prod_{i=1}^{n} \omega_q^{-a[i]x[i]}$$

$$= \sum_{j=0}^{q-1} \omega_q^{-a[u]j} \sum_{x : x[u]=j} g_b(x) \prod_{i=1}^{u-1} \omega_q^{-a[i]x[i]} \prod_{i=u+1}^{n} \omega_q^{-a[i]x[i]}. \tag{8}$$

Let $x, y \in G$ such that x and y only differ in position u. Since $b[u] = 0$, we have $\operatorname{proj}_b(x) = \operatorname{proj}_b(y)$ and, by supposition, $g_b(x) = g_b(y)$. Hence, for any $j \in \{0, 1, \ldots, q-1\}$,

$$\sum_{x : x[u]=j} g_b(x) \prod_{i=1}^{u-1} \omega_q^{-a[i]x[i]} \prod_{i=u+1}^{n} \omega_q^{-a[i]x[i]} = \sum_{x : x[u]=0} g_b(x) \prod_{i=1}^{u-1} \omega_q^{-a[i]x[i]} \prod_{i=u+1}^{n} \omega_q^{-a[i]x[i]}.$$

We can use this property to rewrite the sum in Equation (8) over $g_b(x)$ so it no longer depends on the index j,

$$\langle g_b, \varphi_a \rangle = \left(\sum_{j=0}^{q-1} \omega_q^{-a[u]j} \right) \left(\sum_{x : x[u]=0} g_b(x) \prod_{i=1}^{u-1} \omega_q^{-a[i]x[i]} \prod_{i=u+1}^{n} \omega_q^{-a[i]x[i]} \right)$$

$$= 0 \qquad\qquad\qquad\qquad \text{by Equation (2).}$$

The final equality comes from the geometric series and the fact that $\omega_q^{-a[u]}$ is a q-th root of unity since $a[u] \neq 0$. □

Since epistatically bounded functions are composed of precisely the type of function in the statement of Lemma 1, we may use its result to prove the following theorem, which imposes a constraint on the location of the nonzero Fourier coefficients of any function over G with bounded epistasis.

Theorem 1. *Suppose f is epistatically bounded by k. Let $a \in G$ and let z_a be the Fourier coefficient corresponding to a. We have*

$$|a| > k \implies z_a = 0.$$

Proof. We have

$$z_a = \frac{1}{|G|} \langle f, \varphi_a \rangle = \frac{1}{|G|} \left\langle \left(\sum_{i=1}^{m} g^{(i)} \right), \varphi_a \right\rangle$$

where $g^{(i)}$ depends on at most k positions. Rewriting,

$$z_a = \frac{1}{|G|} \sum_{i=1}^{m} \langle g^{(i)}, \varphi_a \rangle = 0.$$

The final equality follows from Lemma 1. $\qquad\square$

Epistatically bounded functions thus have sparse representations in the Fourier basis. Indeed, we can immediately bound the number of nonzero coefficients for an epistatically bounded function.

Corollary 1 (to Theorem 1). *If f is epistatically bounded by k, the number of nonzero Fourier coefficients for f is bounded from above by $O(n^k (q-1)^k)$. Moreover, if f can be written as the sum of $m = o(n^k)$ functions that each depend on at most k positions, the count of nonzero Fourier coefficients for f is bounded from above by $O(mq^k)$.*

3 Hamming Regions in the q-ary Hypercube

Randomized local search and many evolutionary algorithms employ elementary moves or mutation operators to locally sample the set of states by performing small perturbations on current states to construct similar neighboring states. For a given current state x, the *neighborhood* of x is defined as $N(x)$ which is the set of states reachable from x by a single move or mutation.

Associated with N is the *adjacency matrix*: a square matrix with each row (and each column) corresponding to a unique state where

$$\boldsymbol{A}_{xy} = \begin{cases} 1 & \text{if } y \in N(x); \\ 0 & \text{otherwise.} \end{cases}$$

In the case where the state set is length-n strings over a q-ary alphabet, $\boldsymbol{A}$ is a $q^n \times q^n$ matrix. Though it is always possible to express N in terms of such a matrix, the analysis can become unwieldy in general because the size of the matrix grows as an exponential in n. However, since we have characterized the state space as a group G, where G is defined in Equation (1), it is possible to use a more compact, group theoretic specification of N. In particular we can define some neighborhoods N using a fixed set of d unique group elements $S = \{s_1, s_2, \ldots, s_d\}$ where $s_i \in G$. In this case, for any $x \in G$,

$$N(x) = \{x + s_1, x + s_2, \ldots, x + s_d\}.$$

The underlying neighborhood graph is isomorphic to the Cayley graph of the subgroup of G generated by S. Indeed, if S generates G, then the underlying neighborhood graph is connected.

The adjacency matrix $\boldsymbol{A}$ is a $|G| \times |G|$ matrix and can be viewed as a linear operator that acts on the function space $\{f : G \to \mathbb{C}\}$. In particular, let g be any complex function over G. The image of g under $\boldsymbol{A}$, denoted $\boldsymbol{A}g$, is also a complex function over G where

$$\boldsymbol{A}g(x) = \sum_{y \in G} \left(\boldsymbol{A}_{xy}\right) g(y) = \sum_{y \in N(x)} g(y). \tag{9}$$

When the neighborhood operator can be described in terms of a fixed set S of group elements, we can make some statements about the Fourier functions and their relationship to the adjacency matrix. We say a function $g : G \to \mathbb{C}$ is an *eigenfunction* of the adjacency matrix $\boldsymbol{A}$ when the following equation holds

$$\boldsymbol{A}g(x) = \lambda g(x) \qquad \text{for all } x \in G,$$

where λ is a scalar value. This simply means that the function obtained by applying the linear operator $\boldsymbol{A}$ to g evaluated at state x is directly proportional to g evaluated at x with proportionality constant λ. We say that a function $f : G \to \mathbb{C}$ is elementary in the context of the adjacency matrix $\boldsymbol{A}$ when we can find a constant value b such that the new function $f - b$ is an eigenfunction of $\boldsymbol{A}$ (Reidys and Stadler, 2002). We are now ready to prove the following.

Lemma 2. *If the neighborhood operator N is constructed by a fixed set S of group elements from G, then the Fourier function φ_a is an eigenfunction of $\boldsymbol{A}$ corresponding to the eigenvalue*

$$\lambda_a = \sum_{s \in S} \varphi_a(s).$$

Proof. Consider the matrix $\boldsymbol{A}$ operating on φ_a. We have

$$\boldsymbol{A}\varphi_a(x) = \sum_{y \in N(x)} \varphi_a(y) \qquad\qquad \text{by (9)}$$

$$= \sum_{s \in S} \varphi_a(x + s) = \sum_{s \in S} \prod_{i=1}^{n} \omega_q^{a[i](x[i]+s[i])}$$

$$= \sum_{s \in S} \prod_{i=1}^{n} \omega_q^{a[i]x[i]} \omega_q^{a[i]s[i]} = \sum_{s \in S} \varphi_a(s)\varphi_a(x)$$

$$= \lambda_a \varphi_a(x),$$

where λ_a is as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Thus for any neighborhood defined by a fixed set S of group operations, the Fourier basis functions are eigenfunctions of the adjacency that arises from that neighborhood.

3.1 The Hamming Adjacency

We define the Hamming distance $d_H : G \times G \to \{0, \ldots, n\}$ as

$$d_H(x, y) = |\{i \in \{1, \ldots, n\} : x[i] \neq y[i]\}|. \tag{10}$$

In the case of strings over a finite alphabet, the most natural mutation operator is single-element changes of strings. The underlying connectivity of this mutation operator is captured by the n-dimensional q-ary hypercube (or Hamming graph) $H(n, q)$.
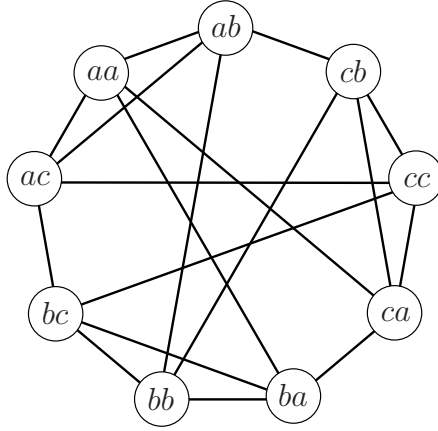
Figure 1: $H(2,3)$: the two-dimensional tertiary hypercube graph (Hamming graph) over $\{a, b, c\}^2$.

In particular, the vertices of $H(n, q)$ correspond to the q^n strings of Σ_q^n and two vertices $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are adjacent if and only if their Hamming distance is one. For a simple example of $H(2, 3)$: the Hamming graph over $\{a, b, c\}^2$, see Figure 1. For the remainder of this paper, we will focus exclusively on this adjacency.

The Hamming adjacency satisfies the antecedent in the claim of Lemma 2 since the single-change neighborhood operator can be constructed by a fixed set S of $(q-1)n$ group elements (in this case, S is also a generating set for G):

$$S = \{x \in G : \exists! i \in \{1, \ldots, n\} : x[i] \neq 0\}.$$

Hence, the claim of Lemma 2 holds for the Hamming adjacency. In particular, letting $\boldsymbol{A}$ expressly refer to the Hamming adjacency matrix, the Fourier function φ_a is an eigenfunction of $\boldsymbol{A}$. To compute the corresponding eigenvalue, note that in the expression

$$\lambda_a = \sum_{s \in S} \varphi_a(s) = \sum_{s \in S} \prod_{i=1}^{n} \omega_q^{a[i]s[i]},$$

the $s[i]$ is nonzero for exactly one i, and each $\ell = 1, 2, \ldots, q-1$ is represented exactly once so we can rewrite this as

$$\lambda_a = \sum_{\ell=1}^{q-1} \sum_{i=1}^{n} \omega_q^{a[i]\ell}. \tag{11}$$

Since ω_q is a primitive q-th root of unity, we have the following property.

$$\sum_{\ell=1}^{q-1} \omega_q^{a[i]\ell} = \sum_{\ell=0}^{q-1} \omega_q^{a[i]\ell} - 1$$

$$= \begin{cases} q - 1 & \text{if } a[i] = 0; \\ -1 & \text{otherwise.} \end{cases} \tag{12}$$

This again follows from the property expressed in Equation (2). If $a[i]$ is nonzero, then $\omega_q^{a[i]}$ is also a q-th root of unity and the summation on the RHS of Equation (12) evaluates to zero. Otherwise, $\omega_q^0 = 1$ and that summation evaluates to q. Putting this together with Equation (11) yields the following expression for the eigenvalue corresponding to $a \in G$. Then

$$\lambda_a = \sum_{i=1}^{n}\sum_{\ell=1}^{q-1} \omega_q^{a[i]\ell} \qquad \text{by (11)}$$
$$= (q-1)(n-|a|) - |a| \qquad \text{by (12)}$$
$$= (q-1)n - q|a|. \qquad (13)$$

3.2 Generalized Adjacencies: Hamming Regions

We define the *Hamming sphere of radius r* around a state $x \in G$ as the set of states $\{y \in G : d_H(x,y) = r\}$, and the *Hamming ball of radius r* around a state $x \in G$ as the set of states $\{y \in G : d_H(x,y) \leq r\}$. In these cases, we refer to x as the *centroid* of the sphere (ball). A set $R \subseteq G$ is called a *Hamming region* if it is a Hamming sphere (ball) for some centroid and radius.

Consider the generalized adjacency $\boldsymbol{A}^{(r)}$ defined as

$$\boldsymbol{A}_{xy}^{(r)} = \begin{cases} 1 & \text{if } d_H(x,y) = r; \\ 0 & \text{otherwise.} \end{cases}$$

This defines a generalized r-change neighborhood, i.e., the neighborhood defined by r-change mutations. In particular, the generalized r-change neighborhood around a state x is the Hamming sphere of radius r around x.

Theorem 2. *Let $\boldsymbol{A}^{(r)}$ denote the radius-r adjacency in the q-ary hypercube. The Fourier function φ_a is an eigenfunction of $\boldsymbol{A}^{(r)}$ corresponding to eigenvalue*

$$\lambda_a^{(r)} = \sum_{j=0}^{r} \binom{|a|}{j}\binom{n-|a|}{r-j}(q-1)^{r-j}(-1)^j.$$

Proof. Let $x,y \in G$ such that $d_H(x,y) = r$. Then there is a group element s where $|s| = r$ and $y = x + s$. Indeed, the fixed set of group elements that define this adjacency is the set $\{s \in G : |s| = r\}$. Hence, by Lemma 2, φ_a is an eigenfunction of $\boldsymbol{A}^{(r)}$ corresponding to eigenvalue

$$\lambda_a^{(r)} = \sum_{s:|s|=r} \varphi_a(s) = \sum_{s:|s|=r}\prod_{i=1}^{n}\omega_q^{a[i]s[i]}.$$

We partition this sum by the indices of the nonzero elements of s. Denote as $[n] = \{1,2,\ldots,n\}$. Since, if $s[i] = 0$, then $\omega_q^{a[i]s[i]} = 1$ we can write

$$\sum_{s:|s|=r}\prod_{i=1}^{n}\omega_q^{a[i]s[i]} = \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq[n]}\left(\sum_{\ell=1}^{q-1}\omega_q^{a[i_1]\ell}\right)\left(\sum_{\ell=1}^{q-1}\omega_q^{a[i_2]\ell}\right)\cdots\left(\sum_{\ell=1}^{q-1}\omega_q^{a[i_r]\ell}\right).$$

For a given $j \in \{0,1,\ldots,r\}$, there are exactly $\binom{|a|}{j}$ ways to choose j indices i_k where $a[i_k] \neq 0$ and $\binom{n-|a|}{r-j}$ ways to choose the remaining $r-j$ indices i'_k where $a[i'_k] = 0$. Summing over j and appealing to Equation (12) yields the claimed result. $\square$

Theorem 2 covers the special case of Equation (13) when $r = 1$. The set of eigenvalues $\lambda_a^{(r)}$ corresponds to the *generalized* Krawtchouk polynomials. These polynomials have applications in the theory of error correcting codes (MacWilliams and Sloane, 1977) in which generalized Hamming schemes are important.

From the theory of Krawtchouk polynomials (see e.g., Chihara and Stanton (1990)), we obtain the following generating function for $\lambda_a^{(r)}$, which will be useful in subsequent sections.

$$\sum_{r=0}^{n} \lambda_a^{(r)} z^r = (1 + (q-1)z)^{n-|a|}(1-z)^{|a|}. \tag{14}$$

Let $\boldsymbol{B}^{(r)}$ denote the matrix in which element $\boldsymbol{B}_{xy}^{(r)}$ is 1 if and only if the Hamming distance between x and y is less than or equal to r and 0 otherwise. We will call $\boldsymbol{B}^{(r)}$ the radius-r *ball matrix* because it encapsulates Hamming ball membership. Since the adjacencies are disjoint, this matrix can be written as a sum of adjacency matrices as follows.

$$\boldsymbol{B}^{(r)} = \sum_{i=0}^{r} \boldsymbol{A}^{(i)}. \tag{15}$$

In Theorem 2 we proved that the Fourier function φ_a is an eigenfunction of the generalized Hamming adjacency matrix $\boldsymbol{A}^{(r)}$ with eigenvalue $\lambda_a^{(r)}$. As a consequence, φ_a is also an eigenfunction of the ball matrix $\boldsymbol{B}^{(r)}$ with eigenvalue

$$\gamma_a^{(r)} = \sum_{i=0}^{r} \lambda_a^{(i)}. \tag{16}$$

The eigenvalues $\lambda_a^{(r)}$ and $\gamma_a^{(r)}$ depend only on the number of nonzero elements of a, that is $|a|$. In subsequent sections, it will be convenient to write the eigenvalues only in terms of their *order* $p \in \{0, 1, \ldots, n\}$. When doing so, we will use square brackets in the subindex:

$$\lambda_{[p]}^{(r)} = \sum_{j=0}^{r} \binom{p}{j}\binom{n-p}{r-j}(q-1)^{r-j}(-1)^j, \tag{17}$$

$$\gamma_{[p]}^{(r)} = \sum_{i=0}^{r} \lambda_{[p]}^{(i)}. \tag{18}$$

4 Moments of Epistatically Bounded Fitness Functions

From an analytical perspective, the distribution of fitness function codomain values over a Hamming region in the q-ary hypercube provides insight into how an evolutionary algorithm might behave in that region of the search space. Suppose $R \subseteq G$ is a Hamming region. Consider the random process of drawing a state uniformly at random from R. The resulting state can be characterized as a random variable over R, which we denote as X_R. The fitness of this uniformly drawn state is measured by the random variable $f(X_R)$.

Because the states are sampled uniformly at random across R, the moments of the frequency distribution of f over R are precisely captured by the moments of the distribution of the random variable $f(X_R)$. For instance, the average fitness in R is simply

$$\mathrm{E}(f(X_R))$$

and the variance of fitness values in R is

$$\mathrm{E}\big(f(X_R)^2\big) - \mathrm{E}(f(X_R))^2 \,.$$

The (complementary) cumulative distribution function of $f(X_R)$ provides an exact count of states that lie above a specific value. For instance, the expression

$$|R| \Pr\{f(X_R) > f(x)\} \tag{19}$$

exactly counts the number of states in the region that have strictly greater fitness than a state x. Note that since R is a finite set, the distribution of $f(X_R)$ is necessarily discrete.

An exact computation of this distribution is not tractable in general, indeed the special case of $q = 2$ is already #P-hard (Sutton et al., 2011a, Theorem 1). Clearly, we would expect it to be as hard as the corresponding search problem when R is a Hamming ball of radius $r = \Omega(n)$. Thus, instead of a complete characterization, we are interested in efficiently computing exact quantities that describe the shape, location, and spread of the distribution of $f(X_R)$. Such quantities are called the *moments* of $f(X_R)$, which are equivalent to the moments of the frequency distribution of f over R. We now apply the foregoing results to describe how one can efficiently compute the first such moment, a quantity that captures the *expectation* of $f(X_R)$. This corresponds to the average codomain value of f over the region R.

For any random variable X over G and any complex function $g : G \to \mathbb{C}$, the expectation of the image of X under g is given by

$$\mathrm{E}(g(X)) = \sum_{x \in G} g(x) \Pr\{X = x\}. \tag{20}$$

This allows us to derive a compact expression for the average value in the image of R under Fourier functions. In particular we have the following lemma. Again, for a region R we denote as X_R as a random variable which samples states in R uniformly at random.

Lemma 3. *Let $S = \{y : d_H(x,y) = r\}$ be the radius r sphere centered on an arbitrary state x. For any Fourier function φ_a, we have the following identity*

$$\mathrm{E}(\varphi_a(X_S)) = |S|^{-1} \lambda_a^{(r)} \varphi_a(x),$$

where $\lambda_a^{(r)}$ is given by Theorem 2.

Proof. By Equation (20),

$$\mathrm{E}(\varphi_a(X_S)) = \sum_{y \in G} \varphi_a(y) \Pr\{X_S = y\}.$$

Since any state y is drawn uniformly at random,

$$\Pr\{X_S = y\} = \begin{cases} |S|^{-1} & \text{if } y \in S; \\ 0 & \text{otherwise,} \end{cases}$$

yielding

$$\mathrm{E}(\varphi_a(X_S)) = |S|^{-1} \sum_{y : d_H(x,y) = r} \varphi_a(y) = |S|^{-1} \boldsymbol{A}^{(r)} \varphi_a(x).$$

The latter equality holds by Equation (9). The claim then follows immediately from Theorem 2. $\qquad \square$

We can also prove the analogous result for an arbitrary Hamming ball.

Lemma 4. *Let $B = \{y : d_H(x,y) \leq r\}$ be the radius r ball centered on an arbitrary state x. For any Fourier function φ_a, we have the following identity*

$$\mathrm{E}(\varphi_a(X_B)) = |B|^{-1}\gamma_a^{(r)}\varphi_a(x).$$

Proof. This follows directly from Lemma 3 and Equations (15) and (16). □

4.1 The Average Fitness in a Hamming Region

Let R be a Hamming region (sphere or ball) of radius r for some $r \in \{0, 1, \ldots, n\}$ and some $x \in G$. The average of f over R is equivalent to the expectation of the random variable $f(X_R)$ which measures the fitness of a state X_R drawn uniformly at random from R.

For asymptotically large Hamming regions, i.e., $|R| = \Theta(q^n)$, the average fitness in R is intractable to compute directly. However, when f is epistatically bounded, we can apply the analysis presented in this paper to derive a compact expression to efficiently compute the exact average. Let $B = \{y : d_H(x,y) \leq r\}$ be a ball of radius r around an arbitrary state x (the case is analogous for Hamming spheres).

We begin by rewriting f in its Fourier basis expansion, and employing linearity of expectation

$$\mathrm{E}(f(X_B)) = \sum_{a \in G} z_a \mathrm{E}(\varphi_a(X_B))$$

and by Lemma 4,

$$\mathrm{E}(f(X_B)) = |B|^{-1} \sum_{a \in G} \gamma_a^{(r)} z_a \varphi_a(x). \tag{21}$$

Because f is supposed to have a sparse representation in the Fourier basis, the resulting sum is over only polynomially many terms, even when $|B| = \Theta(q^n)$.

This equation provides one with a polynomial-time algorithm for computing the first moment of the fitness function over any Hamming ball in the q-ary cube. Let $Z = \{z_0, \ldots\}$ denote the set of nonzero Fourier coefficients for f. The average value of f over B can be computed by the following algorithm.

FIRSTMOMENT(x, r, Z)

$\quad s \leftarrow 0$
$\quad$**for** each $z_a \in Z$ **do**
$\quad\quad s \leftarrow s + \gamma_a^{(r)} z_a \varphi_a(x)$
$\quad$**return** $|B|^{-1} \times s$

The sum accumulated in s corresponds to the sum in (21) and contains exactly $|Z|$ terms. Since we have supposed f to be epistatically bounded by k, Corollary 1 holds and provides a bound on the cardinality of Z. Since the eigenvalues can be precomputed, the time complexity of FIRSTMOMENT is dominated by the loop through the Fourier coefficients and is hence bounded above by $O(n^k(q-1)^k)$. In the case that f can be written as a sum over $m = o(n^k)$ functions that each depend on at most k positions, the bound can be improved to $O(mq^k)$.

In the case of a sphere $S = \{y : d_H(x,y) = r\}$ of radius r, the same algorithm can be used changing the factor $\gamma_a^{(r)}$ by $\lambda_a^{(r)}$ and $|B|$ by $|S|$.

4.2 Higher Moments

We now generalize the above results to calculate higher moments of the frequency distribution of f over Hamming regions. The first moment, for which we gave an efficient computation in Section 4.1, provides one with a measurement of the *location* of the frequency distribution of f over a Hamming region. The c-th moment of the frequency distribution of f over a Hamming region R is defined as the expectation of the c-th power of f:

$$\mathrm{E}(f^c(X_R)).$$

Higher moments are called the *parameters* of a distribution and yield further insight into its *spread* and *shape* (cf. Rényi, 1970). Often, one is interested in the *central* moments of a distribution which center the measurement about the mean. The c-th central moment μ_c can be obtained from the above expression by the transformation

$$\mu_c = \sum_{i=0}^{c} \binom{c}{i} (-1)^{c-i} \mathrm{E}\big(f^i(X_R)\big) \, \mathrm{E}(f(X_R))^{c-i}.$$

For example, the variance, which describes how far the fitness values deviate from their expectation in R, is $\mu_2 = \mathrm{E}\big(f^2(X_R)\big) - \mathrm{E}(f(X_R))^2$. When the c-th central moment is normalized by the c-th power of the standard deviation (i.e., the positive square root of the variance), the result is c-th *standardized moment* of the distribution. This quantity is a standard measurement of the shape of the distribution such as skewness ($c = 3$) and kurtosis ($c = 4$).

It is easy to show that if f has a sparse representation in the Fourier basis and $c = O(1)$, so must f^c.

Lemma 5. *For any $a, b, x \in G$,*

$$\varphi_a(x)\varphi_b(x) = \varphi_{a+b}(x)$$

Proof. The function product is

$$\varphi_a(x)\varphi_b(x) = \prod_{i=1}^{n} \omega_q^{a[i]x[i]} \prod_{i=1}^{n} \omega_q^{b[i]x[i]} = \prod_{i=1}^{n} \omega_q^{(a[i]+b[i])x[i]} = \varphi_{a+b}(x).$$

$\square$

It then becomes straightforward to derive the Fourier series for f^c.

$$
\begin{aligned}
f^c(x) &= \left(\sum_{a \in G} z_a \varphi_a(x) \right)^c \\
&= \sum_{a_1 \in G} \sum_{a_2 \in G} \cdots \sum_{a_c \in G} z_{a_1} z_{a_2} \ldots z_{a_c} \varphi_{a_1 + a_2 + \cdots + a_c}(x).
\end{aligned}
$$

The coefficient corresponding to $\varphi_{a_1 + \cdots + a_c}$ is nonzero if and only if $z_{a_i} \neq 0$ for $i = 1, 2, \ldots c$. But z_{a_i} is a Fourier coefficient for f. Since there are at most $|Z|$ nonzero coefficients for f, there are at most $|Z|^c$ nonzero coefficients for f^c. The c-th moment of f over B is simply the first moment of f^c over B. The c-th moment can be computed by the following algorithm.

MOMENT(x, r, Z, c)
  $Z_c \leftarrow \emptyset$
  for each $(z_{a_1}, \ldots, z_{a_c}) \in Z \times Z \times \cdots \times Z$ **do**
    $z_{a_1 + \cdots + a_c} \leftarrow z_{a_1} z_{a_2} \cdots z_{a_c}$
    $Z_c \leftarrow Z_c \cup \{z_{a_1 + \cdots + a_c}\}$
  return FIRSTMOMENT(x, r, Z_c)

The time complexity of MOMENT is dominated by the call to FIRSTMOMENT for function f^c. Since f is epistatically bounded by k, it follows that f^c is epistatically bounded by $c \times k$ and the runtime is bounded by $O\left(n^{ck}(q-1)^{ck}\right)$. Again, if f can be expressed as a linear combination of $m = o(n^k)$ terms that depend on at most k positions, the bound is improved to $O\left(m^c q^{ck}\right)$. Since we have assumed c and k to be constants, the complexity is polynomial in n. However, we remark here that it is of course practically constrained to modest values of c.

5 Expected Fitness After a Mutation

The mutation operator has capital importance in evolutionary algorithms. Let us consider the following mutation operator in the q-ary hypercube. For each element $x[i]$ in the solution vector, change its value with probability ρ. Here, ρ is typically called the *mutation rate*. The change can be made to any of the $q - 1$ possible values that are different from $x[i]$ with the same probability $1/(q-1)$. The *mutation rate* parameter is a quantity that describes how much an individual is perturbed to generate an offspring during search. One is often interested in deriving the mutation rate that maximizes the probability of generating a successful offspring, i.e., one with strictly improving fitness. Such a quantity has an immediate use in runtime analysis since it describes the success probability of a geometric random variable that measures the time until an improving move is taken.

  The optimal mutation rate for general functions is currently not well-understood. For very simple linear ONEMAX functions over binary strings, the mutation rate was investigated by Bäck (1992). In practice, the commonly recommended mutation rate for functions over binary strings of length n is the so-called "standard mutation rate" of $1/n$. In the case of linear functions over binary strings, Witt (2012) has proved that the optimal mutation rate is indeed $1/n$, and even derived a bound of $en \ln n + O(n)$ on the corresponding runtime of a simple evolutionary algorithm at this rate, which is tight up to lower-order terms.

  The case of nonlinear functions is less clear. Jansen and Wegener (2000) introduce a function over binary strings on which the standard rate of $1/n$ is provably suboptimal because it takes an exponential number of steps to find a global optimum with high probability. Similarly, Böttcher et al. (2010) proved the standard $1/n$ mutation rate is suboptimal for the classical LEADINGONES function. They also introduce an *adaptive* mutation rate that depends on the current fitness and improves the optimization time by 12% over a nonadaptive rate. In the current paper, we find that the expectation-best mutation rate is a function of current solution.

  Doerr et al. (2012) consider the class of strictly monotonic functions over binary strings and show that, for mutation rates of c/n, the choice of constant c has a strong effect on the optimization time. They present a particular monotonic function for which changing the mutation rate of the (1+1)-EA by a constant factor leads to an exponential performance gap. These results stress the importance of an understanding of the relationship between the mutation rate and the function being searched, and that the issue

of how best to set the mutation rate for general optimization is far from resolved.

For epistatically bounded nonlinear functions over binary strings, Chicano and Alba (2011) showed that the expected fitness of the offspring is a polynomial in the mutation rate. Sutton et al. (2011b) also investigated some properties of this polynomial. In this section we extend these results to the q-ary hypercube.

Let $M_\rho(x)$ be a random variable that takes the value in G after applying the mutation operator (with mutation rate ρ) to x. Then, $\Pr\{M_\rho(x) = y\}$ is the probability of obtaining y after the mutation operator is applied to x. We are interested in computing the expected value of a function f over G after a mutation, that is, $\mathrm{E}(f(M_\rho(x)))$. The following theorem provides a first result.

Theorem 3. *For φ_a with $a \in G$ we have*

$$\mathrm{E}(\varphi_a(M_\rho(x))) = \left(1 - \frac{q}{q-1}\rho\right)^{|a|} \varphi_a(x). \tag{22}$$

Proof. The probability of jumping from x to y through a mutation is

$$\Pr\{M_\rho(x) = y\} = \left(\frac{\rho}{q-1}\right)^{d_H(x,y)} (1-\rho)^{n-d_H(x,y)},$$

where $d_H(x, y)$ is the Hamming distance between solutions x and y as defined in Equation (10). We can write the expectation $\mathrm{E}(\varphi_a(M_\rho(x)))$ as

$$\begin{aligned}
\mathrm{E}(\varphi_a(M_\rho(x))) &= \sum_{y \in G} \Pr\{M_\rho(x) = y\}\varphi_a(y) \\
&= \sum_{r=0}^{n} \sum_{\substack{y \in G \\ d_H(x,y)=r}} \left(\frac{\rho}{q-1}\right)^{d_H(x,y)} (1-\rho)^{n-d_H(x,y)}\varphi_a(y) \\
&= \sum_{r=0}^{n} \left(\frac{\rho}{q-1}\right)^{r} (1-\rho)^{n-r} \sum_{\substack{y \in G \\ d_H(x,y)=r}} \varphi_a(y) \\
&= \sum_{r=0}^{n} \left(\frac{\rho}{q-1}\right)^{r} (1-\rho)^{n-r} \lambda_a^{(r)} \varphi_a(x).
\end{aligned} \tag{23}$$

Now, we can use the generating function (14) to simplify the result:

$$\begin{aligned}
\sum_{r=0}^{n} \left(\frac{\rho}{q-1}\right)^{r} (1-\rho)^{n-r}\lambda_a^{(r)} &= (1-\rho)^n \sum_{r=0}^{n} \left(\frac{\rho}{(q-1)(1-\rho)}\right)^{r} \lambda_a^{(r)} \\
&= (1-\rho)^n \left(1 + (q-1)\frac{\rho}{(q-1)(1-\rho)}\right)^{n-|a|} \left(1 - \frac{\rho}{(q-1)(1-\rho)}\right)^{|a|} \\
&= (1-\rho)^n \left(1 + \frac{\rho}{1-\rho}\right)^{n-|a|} \left(1 - \frac{\rho}{(q-1)(1-\rho)}\right)^{|a|} \\
&= (1-\rho)^n \frac{1}{(1-\rho)^{n-|a|}} \left(1 - \frac{\rho}{(q-1)(1-\rho)}\right)^{|a|} = \left((1-\rho) - \frac{\rho}{q-1}\right)^{|a|} \\
&= \left(1 - \left(1 + \frac{1}{q-1}\right)\rho\right)^{|a|} = \left(1 - \frac{q}{q-1}\rho\right)^{|a|},
\end{aligned} \tag{24}$$

and we get (22). $\qquad\square$

Corollary 2. *Given an arbitrary function $f : G \to \mathbb{C}$ with Fourier expansion*

$$f = \sum_{a \in G} z_a \varphi_a, \tag{25}$$

then the expected fitness after the mutation of a solution x with probability ρ is given by

$$E(f(M_\rho(x))) = \sum_{a \in G} z_a \left(1 - \frac{q}{q-1}\rho \right)^{|a|} \varphi_a(x). \tag{26}$$

Proof. Using Theorem 3 we can write

$$E(f(M_\rho(x))) = E\left(\sum_{a \in G} z_a \varphi_a(M_\rho(x)) \right) = \sum_{a \in G} z_a E(\varphi_a(M_\rho(x)))$$

$$= \sum_{a \in G} z_a \left(1 - \frac{q}{q-1}\rho \right)^{|a|} \varphi_a(x),$$

which proves the statement. $\square$

Notice that these results are reduced to those of Chicano and Alba (2011) when $q = 2$. In this case we have for a general function f

$$E(f(M_\rho(x))) = \sum_{a \in G} z_a \left(1 - 2\rho \right)^{|a|} \varphi_a(x). \tag{27}$$

5.1 Discussion

The above derivations allow for determining the mutation rate ρ that maximizes the expected fitness of the offspring of an individual. Unfortunately, such a value is not immediately useful for the following reasons. First, if the fitness function together with mutation operation yields an elementary landscape, then the expected fitness after the mutation of a solution is simplified due to the fact that only the Fourier coefficients z_a with $a = 0$ and $|a| = p$ for some constant p are nonzero. That is, Equation (26) is simplified to

$$E(f(M_\rho(x))) = z_0 + \left(1 - \frac{q}{q-1}\rho \right)^{p} (f(x) - z_0).$$

In this case, the dependence of the expected value on the mutation rate ρ is a simple polynomial with root $\rho = 1 - 1/q$ and the expectation-optimal expected values can be found at the extremes of the interval ($\rho = 0$ or $\rho = 1$) or at the root of the polynomial $\rho = 1 - 1/q$. None of these optimal values provide new information from the problem at hand. In other words, the optimal mutation rates do not depend on the problem; they are general. We will see later an example of this in Section 6. If the fitness function is the sum of several elementary landscapes, then the expectation-optimal mutation rate obtained using Equation (26) becomes problem-dependent and the equation is potentially more useful.

Second, the expected value of the fitness function can have little to do with the real probability of finding an improving solution. As an illustration, if $\rho = 0$ the expected value is exactly $f(x)$ because there is no mutation. If the landscape is elementary and $f(x) > z_0$, this is the maximum possible expected value. However, it is clear that the probability of finding an improving solution is zero if $\rho = 0$.

Third, mutation is a local operation. The mutation rate selected is just used in a single application of the mutation operator during one iteration of an evolutionary algorithm. Thus, this expectation-optimal mutation rate can ensure some kind of optimality in only one step of the algorithm. In this sense, the use of the expectation-optimal mutation rate is a greedy approach. It is well-known that for some optimization problems a greedy approach can sometimes yield poor results.

The second limitation above could be addressed by using the probability that mutating a state x produces an offspring with a strictly improving fitness value. Similar to the complementary cumulative distribution function of $f(X_R)$ in Equation (19), this probability, expressed as a function of x and ρ, is

$$\Pr\left\{f(M_\rho(x)) > f(x)\right\}. \tag{28}$$

Given a sequence of moments, finding the corresponding probability distribution is known as the *moment problem*. Unfortunately, the number and degree of moments necessary to determine the cumulative probability in Equation (28) may render its exact solution intractable.

However, Sutton et al. (2011a) show how a fitness distribution can be approximated using a small number of moments. The distribution is obtained by multiplying the vector of moments by the inverse and transposed Vandermonde matrix constructed from the set of codomain values of the fitness function.

The result is a vector of polynomials that each approximate the probability of the offspring attaining a particular fitness value. Thus, the probability that an offspring has higher fitness than a state x is given by the sum over the set of polynomials that correspond to fitness values strictly greater than $f(x)$. Since the sum of polynomials is again a polynomial, this approximate probability can be expressed a polynomial in ρ. One can then solve for the value of ρ that gives the maximal value in the interval $[0, 1]$.

In practice, this numerically-derived value could be used by algorithm designers to determine good parameter settings for the mutation rate. Moreover, this idea could potentially be applied to a more theoretical result. Specifically, if the above derivations could be relaxed in such a way that allow for good bounds on the improvement probability, it would be possible to establish runtime results for simple genetic algorithms (without crossover) parameterized by ρ. We defer these ideas to future work.

6   Results on Graph Coloring

The graph coloring problem is a classical combinatorial optimization problem in which one is interested in finding a conflict-minimal q-coloring of a graph. The fitness of a coloring, given by the number of conflicts it induces, is epistatically bounded since it can be expressed as a function that depends only on second-order interactions. In this section, we apply to graph coloring the results presented in the foregoing sections.

Let $\Gamma = (V, E)$ be a graph with $|V| = n$. A q-*coloring* of Γ is an assignment $\mathcal{A} : V \to \{0, 1, \ldots, q-1\}$. The set of all q-colorings of Γ is isomorphic to $\{0, 1, \ldots, q-1\}^n$. The objective of the graph coloring problem is to find a coloring x that minimizes the count of edges incident to vertices that have the same color in x.

The fitness function can be characterized as

$$f(x) = \sum_{(u,v) \in E} \delta_{(u,v)}(x), \tag{29}$$

where

$$\delta_{(u,v)}(x) = \begin{cases} 1 & \text{if } x[u] = x[v]; \\ 0 & \text{otherwise.} \end{cases}$$

The following lemma allows us to characterize the sum of a series of Fourier functions over all elements of G in which two elements are fixed. This will be useful for the calculation of the Fourier coefficients of the graph coloring problem.

Lemma 6. *Let $a \in G$ be arbitrary. Let G' denote the direct product of $n-2$ cyclic groups of order q, that is*

$$G' = \underbrace{\mathbb{Z}_q \oplus \mathbb{Z}_q \oplus \cdots \oplus \mathbb{Z}_q}_{n-2}.$$

Let $a' \in G'$ be the element obtained from a by removing the components in positions u and v. Then

$$\sum_{\substack{x \in G: \\ x[u]=p, x[v]=r}} \varphi_a(x) = \omega_q^{a[u]p + a[v]r} \sum_{x' \in G'} \varphi_{a'}(x').$$

Proof. Without loss of generality, suppose $u < v$. We thus have the following.

$$\sum_{\substack{x \in G: \\ x[u]=p, x[v]=r}} \varphi_a(x) = \sum_{\substack{x \in G: \\ x[u]=p, x[v]=r}} \prod_{i=1}^{n} \omega_q^{a[i]x[i]}$$

$$= \sum_{\substack{x \in G: \\ x[u]=p, x[v]=r}} \left(\prod_{i=1}^{u-1} \omega_q^{a[i]x[i]} \right) \omega_q^{a[u]p} \left(\prod_{i=u+1}^{v-1} \omega_q^{a[i]x[i]} \right) \omega_q^{a[v]r} \left(\prod_{i=v+1}^{n} \omega_q^{a[i]x[i]} \right)$$

$$= \omega_q^{a[u]p} \omega_q^{a[v]r} \sum_{x' \in G'} \varphi_{a'}(x'),$$

which yields the claimed result. □

We are now ready to prove the following.

Theorem 4. *Let f be the fitness function for the q-coloring problem on a graph $\Gamma = (V, E)$ where $|V| = n$. We say that a group element $a \in G$ is a* balanced mask *of the edge $(u, v) \in E$ if and only if the following conditions hold:*

1. *$|a| = 2$,*

2. *$a[u] \neq 0$ and $a[v] \neq 0$ are the two nonzero entries of a*

3. *$a[u] + a[v] \equiv 0 \pmod{q}$.*

Then the Fourier coefficients of f are

$$z_a = \begin{cases} \frac{|E|}{q} & \text{if } |a| = 0; \\ \frac{1}{q} & \text{if } a \text{ is a balanced mask of some } (u, v) \in E; \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Substituting the definition of f from Equation (29) into the expression for the Fourier coefficient in (4) (and using the fact that $|G| = q^n$) we can write

$$z_a = \frac{1}{q^n} \sum_{x \in G} \sum_{(u,v) \in E} \delta_{(u,v)}(x) \overline{\varphi_a(x)}.$$

Because $\delta_{(u,v)}(x)$ is nonzero only for $x[u] = x[v] = \ell$ for each $\ell \in \{0, 1, \dots, q-1\}$ we can write the Fourier coefficient corresponding to $a \in G$ as

$$z_a = \frac{1}{q^n} \sum_{\ell=0}^{q-1} \sum_{(u,v) \in E} \sum_{\substack{x \in G: \\ x[u]=x[v]=\ell}} \overline{\varphi_a(x)}.$$

Letting G' be the direct product of $n-2$ cyclic groups of order q and $a' \in G'$ be the element obtained by removing the components in positions u and v from a, we can appeal to Lemma 6 to yield the following.

$$z_a = \frac{1}{q^n} \sum_{\ell=0}^{q-1} \sum_{(u,v) \in E} \overline{\omega_q^{(a[u]+a[v])\ell}} \sum_{x' \in G'} \overline{\varphi_{a'}(x')}. \tag{30}$$

We now make a case distinction for a.

Case 1: $|a| = 0$, that is, $a = (0, 0, \dots, 0)$. In this case, $|a'| = 0$ in (30) so by (7) we have

$$z_a = \frac{1}{q^n} \sum_{\ell=0}^{q-1} \sum_{(u,v) \in E} \overline{\omega_q^{(a[u]+a[v])\ell}} |G'|$$

$$= \frac{1}{q^n} \sum_{(u,v) \in E} \sum_{\ell=0}^{q-1} \overline{\omega_q^{(a[u]+a[v])\ell}} q^{n-2}.$$

Furthermore, since for any u and v, $a[u] + a[v] = 0$, we have $\overline{\omega_q^{(a[u]+a[v])\ell}} = 1$ yielding

$$= \frac{1}{q^n} \sum_{(u,v) \in E} q^{n-1} = \frac{|E|}{q}.$$

Case 2: $|a| = 1$. Suppose that $a[w] \neq 0$ is the single nonzero component of a. Considering the interior summations of Equation (30), for each $(u,v) \in E$, either $u = w$, $v = w$, or $u \neq w \neq v$. If $u = w$ or $v = w$, then the term

$$\sum_{\ell=0}^{q-1} \overline{\omega_q^{(a[u]+a[v])\ell}} = 0$$

since in this case $0 < a[u] + a[v] < q$ making $\omega_q^{a[u]+a[v]}$ a q-th root of unity and the sum vanishes by Equation (2). On the other hand, if $u \neq w \neq v$, then $|a'| > 0$ and the term

$$\sum_{x' \in G'} \overline{\varphi_{a'}(x)} = 0$$

by Equation (7). Since these are the only possibilities for $|a| = 1$, it must be the case that $z_a = 0$ when $|a| = 1$.

Case 3: $|a| = 2$. Let $a[w] \neq 0$ and $a[t] \neq 0$ be the two only nonzero components of a. If $(w, t) \notin E$, then clearly (30) evaluates to zero. On the other hand, if $(w, t) \in E$, then we have

$$z_a = \frac{1}{q^n} \sum_{\ell=0}^{q-1} \overline{\omega_q^{(a[w]+a[t])\ell}} \sum_{x' \in G'} \overline{\varphi_{a'}(x')}$$

and since $|a'| = 0$, by Equation (7) we have

$$= \frac{1}{q^n} \sum_{\ell=0}^{q-1} \overline{\omega_q^{(a[w]+a[t])\ell}} q^{n-2}$$

$$= \frac{q^{n-2}}{q^n} \sum_{\ell=0}^{q-1} \overline{\omega_q^{(a[w]+a[t])\ell}} = \begin{cases} \frac{1}{q} & \text{if } a[w] + a[t] \equiv 0 \pmod{q}; \\ 0 & \text{otherwise.} \end{cases}$$

The final step follows from the fact that if $a[w] + a[t] \not\equiv 0 \pmod{q}$, then $\omega_q^{a[w]+a[t]}$ is a q-th root of unity and $\sum_{\ell=0}^{q-1} \omega_q^{a[w]+a[t]} = 0$. Otherwise, $\omega_q^{a[w]+a[t]} = 1$ and the sum evaluates to q. Hence, if $|a| = 2$, then $z_a = \frac{1}{q}$ if and only if the nonzero components of a correspond to an edge in E *and* their sum is equivalent to zero modulo q, otherwise $z_a = 0$. Put another way, if $|a| = 2$, then z_a is nonzero (and equal to $\frac{1}{q}$) if and only if a is a balanced mask of an edge $(w, t) \in E$.

Case 4: $|a| > 2$. In this case, $|a'| > 0$ so again by Equation (7), $\sum_{x' \in G'} \overline{\varphi_{a'}(x)} = 0$, and Equation (30) evaluates to zero. Hence for the case of $|a| > 2$, $z_a = 0$.

These cases cover all $a \in G$ and yield the claimed result. $\qquad\qquad\square$

Thus, the graph coloring fitness function that counts constraint violations in a q-coloring corresponding to $x \in G$ has a sparse representation in the Fourier basis: its Fourier coefficients are nonzero for $a \in G$ which have exactly two nonzero entries corresponding to balanced masks of edges in Γ.

6.1 The Average Number of Constraint Violations Within r Steps

We now apply the above results to compute the average number of constraint violations that occur within a sphere or ball of radius r centered on an arbitrary coloring. In doing so, we generalize a well-known result of Grover (1992) that relates the average number of constraint violations in the single-change neighborhood with the number of violations in the current coloring.

Let $x \in G$ correspond to a q-coloring of Γ. Let $B = \{y : d_H(x, y) \leq r\}$ denote the Hamming ball of radius r around x. Substituting into Equation (21) results in an expression for this average value of f in B.

$$\mathrm{E}(f(X_B)) = |B|^{-1} \sum_{a \in G} \gamma_a^{(r)} z_a \varphi_a(x) \qquad\qquad \text{by (21)}$$

Since $|a| = 0$ or $|a| = 2$, we employ the notation defined in (18) to write

$$
\begin{aligned}
&= |B|^{-1} \left(\gamma_{[0]}^{(r)} z_0 + \gamma_{[2]}^{(r)} \sum_{\substack{a \in G: \\ |a| = 2}} z_a \varphi_a(x) \right) \\
&= |B|^{-1} \left(\gamma_{[0]}^{(r)} z_0 + \gamma_{[2]}^{(r)} \left(f(x) - z_0 \right) \right) \\
&= |B|^{-1} \left(\frac{\gamma_{[0]}^{(r)} |E|}{q} + \gamma_{[2]}^{(r)} \left(f(x) - \frac{|E|}{q} \right) \right),
\end{aligned}
\tag{31}
$$

where $\gamma_{[2]}^{(r)}$ and $\gamma_{[0]}^{(r)}$ are given by Equation (18).

The same can be done for the spheres. Let $S = \{y : d_H(x, y) = r\}$ denote the Hamming sphere of radius r centered on x, a q-coloring of the graph. Then we have

$$
\mathrm{E}(f(X_S)) = |S|^{-1} \left(\lambda_{[0]}^{(r)} z_0 + \lambda_{[2]}^{(r)} \sum_{\substack{a \in G: \\ |a| = 2}} z_a \varphi_a(x) \right) = |S|^{-1} \left(\lambda_{[0]}^{(r)} z_0 + \lambda_{[2]}^{(r)} \left(f(x) - z_0 \right) \right)
$$

$$
= |S|^{-1} \left(\frac{\lambda_{[0]}^{(r)} |E|}{q} + \lambda_{[2]}^{(r)} \left(f(x) - \frac{|E|}{q} \right) \right),
\tag{32}
$$

where $\lambda_{[2]}^{(r)}$ and $\lambda_{[0]}^{(r)}$ are given by Equation (17).

Note that in the simple case, when $r = 1$, we have $|S| = (q-1)n$ and the above equation simplifies to

$$
\mathrm{E}(f(X_S)) = \frac{1}{(q-1)n} \left(\lambda_{[0]} z_0 + \lambda_{[2]} \left(f(x) - z_0 \right) \right).
$$

Substituting the proper values for $\lambda_{[p]}$ we have

$$
\begin{aligned}
\mathrm{E}(f(X_S)) &= \frac{1}{(q-1)n} \left((q-1)n z_0 + \left((q-1)n - 2q \right) \left(f(x) - z_0 \right) \right) \\
&= z_0 + \left(1 - \frac{2q}{(q-1)n} \right) \left(f(x) - z_0 \right) \\
&= f(x) - \frac{2q}{(q-1)n} \left(f(x) - z_0 \right).
\end{aligned}
$$

From Theorem 4 we get $z_0 = \frac{|E|}{q}$ which finally yields

$$
= f(x) - \frac{2q}{(q-1)n} \left(f(x) - \frac{|E|}{q} \right).
$$

This exactly corresponds to the result of Grover (1992) who derived the same expression for the average number of constraint violations in the immediate Hamming neighbors of a coloring. Equation (31) provides a similar compact expression for the average number of constraint violations within a Hamming region. We also remark here that the higher moments of the frequency distribution of the count of constraint violations could be easily computed using the method presented in Section 4.2.

6.2 Expectation-optimal Mutation Rate

In Section 5 we provided closed-form formulas for the expected value after mutation of an arbitrary fitness function f. In this section we particularize the expression to the case of the graph coloring problem and we discuss the result from a practical point of view. We start with the main result of this section.

Theorem 5. *In the graph coloring problem, the expected value of the fitness function after the mutation of a solution x with probability ρ is given by:*

$$\mathrm{E}(f(M_\rho(x))) = \frac{|E|}{q} + \left(1 - \frac{q}{q-1}\rho\right)^2 \left(f(x) - \frac{|E|}{q}\right) \tag{33}$$

Proof. According to Theorem 4 the fitness function of the graph coloring problem can be written as:

$$f(x) = z_0 + \sum_{\substack{a \in G: \\ |a|=2}} z_a \varphi_a(x) \tag{34}$$

where $z_0 = |E|/q$ and $z_a = 1/q$. Then, using Equation (26) we can write the expected value after mutation as:

$$\mathrm{E}(f(M_\rho(x))) = \sum_{a \in G} \left(1 - \frac{q}{q-1}\rho\right)^{|a|} z_a \varphi_a(x) = z_0 + \sum_{\substack{a \in G: \\ |a|=2}} \left(1 - \frac{q}{q-1}\rho\right)^{|a|} z_a \varphi_a(x)$$

$$= z_0 + \left(1 - \frac{q}{q-1}\rho\right)^2 \sum_{\substack{a \in G: \\ |a|=2}} z_a \varphi_a(x) = z_0 + \left(1 - \frac{q}{q-1}\rho\right)^2 (f(x) - z_0)$$

$$= \frac{|E|}{q} + \left(1 - \frac{q}{q-1}\rho\right)^2 \left(f(x) - \frac{|E|}{q}\right) \tag{35}$$

which proves the statement. $\square$

We are interested in studying the expected fitness under mutation with respect to the mutation rate ρ. Equation (33) is a quadratic equation in ρ. Furthermore, the vertex of the parabola is in $\rho = 1 - 1/q$. In Figure 2 we plot $\mathrm{E}(f(M_\rho(x)))$ against ρ.

Our first observation is that the expected value of the fitness function is $f(x)$ when $\rho = 0$, since in this case we do not change the individual. Then, the expected value approaches the average value $|E|/q$ in a quadratic way. The expected value is exactly the average value when $\rho = 1 - 1/q$ (the vertex of the parabola). This result can also be explained from a different point of view. If $\rho = 1 - 1/q$, the probability of keeping the current value of a solution component $x[i]$ is $1/q$ and the probability of changing this value to a different one is $(q-1)/q$. There are $q - 1$ values to which the current one can be changed and the probability of selecting one of these values is $1/(q-1)$. Then, the probability of changing the value of $x[i]$ to a given new value is

$$\frac{q-1}{q} \cdot \frac{1}{q-1} = \frac{1}{q},$$

exactly the same as keeping the current value of $x[i]$. In summary, for this particular mutation rate, the mutation operator has the effect of replacing x with a state chosen uniformly at random from the search space. The expected value is thus the average value of the fitness function over the entire search space.
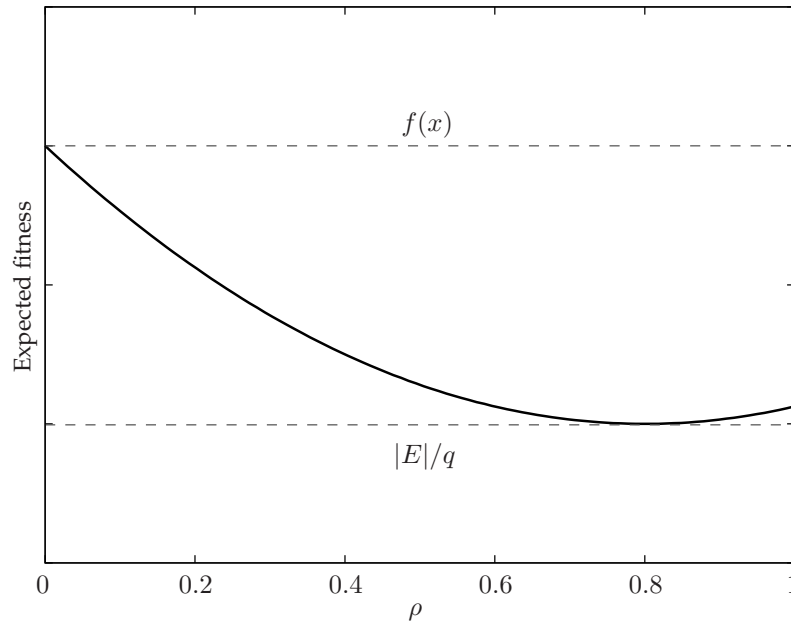
Figure 2: Expected fitness of an offspring of x, i.e., $\mathrm{E}(f(M_\rho(x)))$, plotted against mutation rate ρ for the graph coloring problem. The fitness of the current solution x is denoted by $f(x)$ whereas $|E|/q$ is the average fitness over all solutions.

The plot in Figure 2 considers the case in which $f(x) > |E|/q$. If $f(x) < |E|/q$ then the plot is inverted along the X axis and the parabola has a maximum at $\rho = 1 - 1/q$. We can summarize the previous considerations in the following rules:

- If $f(x) > |E|/q$ the maximum expected value of the fitness function after mutation is reached when $\rho = 0$, that is, when no mutation is applied.

- If $f(x) < |E|/q$ the maximum expected value of the fitness function after mutation is reached when $\rho = 1 - 1/q$, which is equivalent to picking a random solution from the search space.

- If $f(x) = |E|/q$ the expected mutation is always $|E|/q$ independently of the value of ρ.

This analysis of the expected value of the fitness function for the graph coloring problem provides some information of the mutation process but does not give interesting values for the mutation rate. The reason is that the graph coloring problem is what we call an elementary landscape, and the expected value of an elementary landscape after a mutation operation has always the maximum and minimum values in the extremes ($\rho = 0$ and $\rho = 1$) and/or a trivial value inside the interval ($\rho = 1 - 1/q$ in our case). More interesting values can be obtained when the function is not an elementary landscape (e.g., see Chicano and Alba (2011)). This happens in the case of the frequency assignment problem, which we now explore.

7   Results on Frequency Assignment

The frequency assignment problem is the last step in the layout of a radio network, like a 2G (second generation) cellular mobile network. Frequency assignment entails the assignment of a channel (a frequency band) to every transceiver (TRX) in the radio network. The optimization problem arises because the usable radio spectrum is generally very scarce and, consequently, channels have to be reused by many TRXs in the network (Eisenblätter, 2001).

The multiple use of the same channel may cause interferences that might reduce the quality of service down to unsatisfactory levels. Indeed, significant interference may occur if the same or adjacent-channels are used in neighboring overlapping cells. Two matrices are often used to measure the interferences. In the so-called *interference matrix*, denoted by M, each element m_{ij} indicates the degradation of the network quality if TRXs i and j operate on the same channel. This is called *co-channel interference*. In addition to co-channel interference there also may exist a so-called *adjacent-channel interference*, which occurs when two TRXs operate on adjacent channels (i.e., one TRX operates on channel p and the other on channel $p+1$ or $p-1$). Co-channel and adjacent-channel interferences are the most important ones in the design of a radio network. But we could also be interested in considering interferences due to overlapping of channels with a larger separation. This is in accordance with real-world applications, since the amount of interference between two channels depends on the separation of the channels (Aardal et al., 2003).

Thus, in our generalized form of the frequency assignment problem, we consider both co-channel interference and adjacent channel interference as well as interferences due to frequencies with a larger separation.

We can assign a cost to each possible interference that can occur in a channel assignment. Then, the objective is to minimize the cost due to interferences in a radio network. An additional generalization of this problem also considers the possibility of additional costs due to the mere fact that a given channel is used by a given TRX, e.g., a fee could be charged to a telecommunication company for using a channel in a given location. In general, the set of channels that can be assigned to each TRX might be different. We assume that the valid channels of each TRX are sorted and we use an integer number to represent their position in the sorted set. We also assume, without loss of generality, that the number of valid channels is the same, q, in all the TRXs. We denote with n the number of TRXs in the radio network.

In order to take into account all the previous considerations and keep a compact formulation of the problem we define an array of weights $w \in \mathbb{R}^{n \times n \times q \times q}$ in which we denote each element with $w_{u,v}^{p,r}$, where $u, v \in \{1, \ldots, n\}$ and $p, r \in \{0, \ldots, q-1\}$. We can interpret the element $w_{u,v}^{p,r}$ as the cost of having channel p in TRX u and channel r in TRX v. One solution for this problem is a map from the TRXs set, denoted with V, to the set of possible channels $F = \{0, 1, \ldots, q-1\} = \mathbb{Z}_q$. Thus, the solution space is $G = F^V = \mathbb{Z}_q^n$. Using the array of weights we can define the cost function as:

$$f(x) = \sum_{u,v=1}^{n} w_{u,v}^{x[u],x[v]}. \tag{36}$$

The cost element $w_{v,u}^{r,p}$ has the same meaning as $w_{u,v}^{p,r}$, so we can set one of them to zero. However, for the sake of clarity and without loss of generality, we will take the convention that $w_{u,v}^{p,r} = w_{v,u}^{r,p}$ for all u, v, p, r. Then, the cost element $w_{u,v}^{p,r}$ with $u \neq v$ must be interpreted as half the cost of having channel p in TRX u and channel r in TRX

v. If $u = v$ the element $w_{u,u}^{p,p}$ is the additional cost of having channel p in TRX u. The values $w_{u,u}^{p,r}$ with $p \neq r$ are zero.

Let us rewrite the cost function (36) in a more convenient way, as a linear combination of functions:

$$f(x) = \sum_{u,v=1}^{n} \sum_{p,r=0}^{q-1} w_{u,v}^{p,r} \delta_{u,v}^{p,r}(x), \tag{37}$$

where

$$\delta_{u,v}^{p,r}(x) = \begin{cases} 1 & \text{if } x[u] = p \wedge x[v] = r; \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

Now we can focus on the functions $\delta_{u,v}^{p,r}$. The following lemma provides a first result.

Lemma 7. *Given $u, v \in \{1, 2, \ldots, n\}$ and $p, r \in \mathbb{Z}_q$ the Fourier coefficients of the function $\delta_{u,v}^{p,r}$ defined in (38) are*

$$z_a = \begin{cases} \dfrac{\omega_q^{-a[u]p - a[v]r}}{q^2} & \text{if } a[t] = 0 \text{ for all } t \in \{1, 2, \ldots, n\} \text{ such that } t \neq u \text{ and } t \neq v. \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

Proof. Substituting the definition of $\delta_{u,v}^{p,r}$ into the expression for the Fourier coefficients (4) we can write

$$z_a = \frac{1}{q^n} \sum_{x \in G} \delta_{u,v}^{p,r}(x) \overline{\varphi_a(x)}.$$

Since $\delta_{u,v}^{p,r}(x)$ is one when $x[u] = p$ and $x[v] = r$ and zero otherwise, we can write

$$z_a = \frac{1}{q^n} \sum_{\substack{x \in G: \\ x[u]=p, x[v]=r}} \overline{\varphi_a(x)};$$

and using Lemma 6 we have

$$z_a = \frac{\overline{\omega_q^{a[u]p + a[v]r}}}{q^n} \sum_{x' \in G'} \overline{\varphi_{a'}(x')}.$$

where G' is the direct product of $n-2$ cyclic groups of order q and $a' \in G'$ is the element obtained by removing the components in positions u and v from a. Equation (7) states that the sum in the previous expression will be $|G'| = q^{n-2}$ only if $a' = 0$, that is, if $a[t] = 0$ for all $t \in \{1, 2, \ldots, n\}$ such that $t \neq u$ and $t \neq v$. The sum will be zero if any of the $a[t] \neq 0$ for t different from u and v. Replacing the sum in the previous expression by zero or q^{n-2} and taking into account that $\overline{\omega_q^b} = \omega_q^{-b}$ we find (39). $\square$

Using this lemma we can prove the following theorem.

Theorem 6. *The Fourier coefficients of the fitness function of the frequency assignment problem*

A. M. Sutton, F. Chicano, and L. D. Whitley

f(x) are

$$
z_a = \begin{cases}
\dfrac{1}{q^2} \displaystyle\sum_{u,v=1}^{n} \sum_{p,r=0}^{q-1} w_{u,v}^{p,r} & \text{if } |a| = 0; \\[3ex]
\dfrac{2}{q^2} \displaystyle\sum_{v=1}^{n} \sum_{p,r=0}^{q-1} w_{t,v}^{p,r} \omega_q^{-a[t]p} - \sum_{p=0}^{q-1} w_{t,t}^{p,p} \omega_q^{-2a[t]p} & \text{if } |a| = 1 \text{ and } a[t] \neq 0; \\[3ex]
\dfrac{2}{q^2} \displaystyle\sum_{p,r=0}^{q-1} w_{u,v}^{p,r} \omega_q^{-a[u]p - a[v]r} & \text{if } |a| = 2,\, a[u] \neq 0 \text{ and } a[v] \neq 0; \\[3ex]
0 & \text{if } |a| > 2.
\end{cases}
\tag{40}
$$

Proof. We can find the expression for the Fourier coefficients z_a of $f(x)$ by computing the weighted sum of Equation (37) using the Fourier coefficients of the δ functions instead of the functions themselves:

$$
z_a = \sum_{u,v=1}^{n} \sum_{p,r=0}^{q-1} w_{u,v}^{p,r} z_a^{u,v,p,r},
\tag{41}
$$

where we include the parameters u, v, p, r in the Fourier coefficients of the δ functions. We now distinguish four cases depending on the number of nonzero elements of a.

- Case 1: $|a| = 0$. According to Lemma 7, $z_a^{u,v,p,r} = 1/q^2$ and Equation (41) can be simplified to:

$$
z_a = \frac{1}{q^2} \sum_{u,v=1}^{n} \sum_{p,r=0}^{q-1} w_{u,v}^{p,r}.
$$

- Case 2: $|a| = 1$. Let us denote with $a[t]$ the nonzero entry of a. The coefficient $z_a^{u,v,p,r}$ is nonzero if $t = u$ or $t = v$. Then, we can simplify Equation (41) to

$$
\begin{aligned}
z_a &= \frac{1}{q^2} \sum_{v=1}^{n} \sum_{p,r=0}^{q-1} \left(w_{t,v}^{p,r} \omega_q^{-a[t]p} + w_{v,t}^{p,r} \omega_q^{-a[t]r} \right) - \sum_{p=0}^{q-1} w_{t,t}^{p,p} \omega_q^{-2a[t]p} \\
&= \frac{1}{q^2} \sum_{v=1}^{n} \sum_{p,r=0}^{q-1} \left(w_{t,v}^{p,r} \omega_q^{-a[t]p} + w_{t,v}^{r,p} \omega_q^{-a[t]r} \right) - \sum_{p=0}^{q-1} w_{t,t}^{p,p} \omega_q^{-2a[t]p} \\
&= \frac{2}{q^2} \sum_{v=1}^{n} \sum_{p,r=0}^{q-1} w_{t,v}^{p,r} \omega_q^{-a[t]p} - \sum_{p=0}^{q-1} w_{t,t}^{p,p} \omega_q^{-2a[t]p}.
\end{aligned}
\tag{42}
$$

- Case 3: $|a| = 2$. Let us denote with $a[u]$ and $a[v]$ the two only entries in a that are nonzero. Then, only the coefficients $z_a^{u,v,p,r}$ with any arbitrary value for p and r are

nonzero. Equation (41) can be simplified to:

$$
\begin{aligned}
z_a &= \frac{1}{q^2} \sum_{p,r=0}^{q-1} \left(w_{u,v}^{p,r} \omega_q^{-a[u]p-a[v]r} + w_{v,u}^{p,r} \omega_q^{-a[v]p-a[u]r} \right) \\
&= \frac{1}{q^2} \sum_{p,r=0}^{q-1} \left(w_{u,v}^{p,r} \omega_q^{-a[u]p-a[v]r} + w_{u,v}^{r,p} \omega_q^{-a[u]r-a[v]p} \right) \\
&= \frac{2}{q^2} \sum_{p,r=0}^{q-1} w_{u,v}^{p,r} \omega_q^{-a[u]p-a[v]r}.
\end{aligned}
\tag{43}
$$

- Case 4: $|a| > 2$. In this case $z_a^{u,v,p,r} = 0$ and, as a consequence, $z_a = 0$.

These cases cover all the possible values for $a \in G$ and proves the statement. $\qquad\square$

As in the case of the graph coloring problem, we can observe that the frequency assignment problem has also bounded epistasis and thus the fitness function has a sparse representation in the Fourier basis. In this case, the Fourier coefficients are nonzero only for the elements $a \in G$ with nonzero values in two or less entries.

In the following it will be convenient to group together the terms in the Fourier expansion related to elements $a \in G$ having the same number of nonzero entries, the so-called elementary components of the function (Chicano et al., 2011). We identify here three components, related to the terms in which $|a| = 0$, $|a| = 1$ and $|a| = 2$. The definition of the elementary components are:

$$
f_{[0]} = z_0,
\tag{44}
$$

$$
f_{[1]}(x) = \sum_{\substack{a \in G \\ |a|=1}} z_a \varphi_a(x),
\tag{45}
$$

$$
f_{[2]}(x) = \sum_{\substack{a \in G \\ |a|=2}} z_a \varphi_a(x),
\tag{46}
$$

where we omit the variable in $f_{[0]}$ to highlight that it is a constant (function). The fitness function $f(x)$ is just the sum of these three terms: $f(x) = f_{[0]} + f_{[1]}(x) + f_{[2]}(x)$.

7.1 The Average Fitness in Hamming Regions

We can use the Fourier expansion of the previous section to compute the average value of the fitness function in an arbitrary Hamming region (ball or sphere) of size r. Let $x \in G$ corresponds to a map of TRXs to channels. Let $B = \{y : d_H(x,y) \le r\}$ denote the Hamming ball of radius r around x. Substituting into Equation (21) results in an expression for this average value of f in B.

$$
\begin{aligned}
\mathrm{E}(f(X_B)) &= |B|^{-1} \left(\gamma_{[0]}^{(r)} z_0 + \gamma_{[1]}^{(r)} \sum_{\substack{a \in G: \\ |a|=1}} z_a \varphi_a(x) + \gamma_{[2]}^{(r)} \sum_{\substack{a \in G: \\ |a|=2}} z_a \varphi_a(x) \right) \\
&= |B|^{-1} \left(\gamma_{[0]}^{(r)} f_{[0]} + \gamma_{[1]}^{(r)} f_{[1]}(x) + \gamma_{[2]}^{(r)} f_{[2]}(x) \right),
\end{aligned}
\tag{47}
$$

where $\gamma_{[0]}^{(r)}$, $\gamma_{[1]}^{(r)}$ and $\gamma_{[2]}^{(r)}$ are given by Equation (18).

The same can be done for the spheres. Let $S = \{y : d_H(x, y) = r\}$ denote the Hamming sphere of radius r centered on x, a map from TRXs to channels. Then we have

$$
\begin{aligned}
\mathrm{E}(f(X_S)) &= |S|^{-1} \left(\lambda_{[0]}^{(r)} z_0 + \lambda_{[1]}^{(r)} \sum_{\substack{a \in G: \\ |a|=1}} z_a \varphi_a(x) + \lambda_{[2]}^{(r)} \sum_{\substack{a \in G: \\ |a|=2}} z_a \varphi_a(x) \right) \\
&= |S|^{-1} \left(\lambda_{[0]}^{(r)} f_{[0]} + \lambda_{[1]}^{(r)} f_{[1]}(x) + \lambda_{[2]}^{(r)} f_{[2]}(x) \right).
\end{aligned}
\tag{48}
$$

where $\lambda_{[0]}^{(r)}$, $\lambda_{[1]}^{(r)}$ and $\lambda_{[2]}^{(r)}$ are given by Equation (17).

Note that in the simplest case, when $r = 1$, we have $|S| = (q-1)n$ and the above equation simplifies to

$$
\mathrm{E}(f(X_S)) = \frac{1}{(q-1)n} \left(\lambda_{[0]} f_{[0]} + \lambda_{[1]} f_{[1]}(x) + \lambda_{[2]} f_{[2]}(x) \right).
$$

Substituting the proper values for $\lambda_{[p]}$ we have

$$
\begin{aligned}
\mathrm{E}(f(X_S)) &= \frac{1}{(q-1)n} \left((q-1)n f_{[0]} + ((q-1)n - q) f_{[1]}(x) + ((q-1)n - 2q) f_{[2]}(x) \right) \\
&= \frac{1}{(q-1)n} \left((q-1)n (f_{[0]} + f_{[1]}(x) + f_{[2]}(x)) - q f_{[1]}(x) - 2q f_{[2]}(x) \right) \\
&= \frac{1}{(q-1)n} \left((q-1)n f(x) - q(f_{[1]}(x) + 2 f_{[2]}(x)) \right) \\
&= f(x) - \frac{q}{(q-1)n} (f_{[1]}(x) + 2 f_{[2]}(x)).
\end{aligned}
$$

As in the case of the graph coloring problem, the higher moments of the frequency distribution of the values of the fitness function could be easily computed using the method presented in Section 4.2.

7.2 Expectation-optimal Mutation Rate

Let us now particularize the results of Section 5 to the case of the frequency assignment problem. The following theorem presents the main result.

Theorem 7. *In the frequency assignment problem, the expected value of the fitness function after the mutation of a solution x with probability ρ is given by:*

$$
\mathrm{E}(f(M_\rho(x))) = f_{[0]} + \left(1 - \frac{q}{q-1} \rho \right) f_{[1]}(x) + \left(1 - \frac{q}{q-1} \rho \right)^2 f_{[2]}(x).
\tag{49}
$$

Proof. According to Theorem 4 the fitness function of the graph coloring problem can be written as:

$$
f(x) = \sum_{\substack{a \in G: \\ |a| \leq 2}} z_a \varphi_a(x).
\tag{50}
$$

Then, using Equation (26) we can write the expected value after mutation as:

$$\mathrm{E}(f(M_\rho(x))) = \sum_{a \in G} \left(1 - \frac{q}{q-1}\rho\right)^{|a|} z_a \varphi_a(x)$$

$$= z_0 + \left(1 - \frac{q}{q-1}\rho\right) \sum_{\substack{a \in G: \\ |a|=1}} z_a \varphi_a(x) + \left(1 - \frac{q}{q-1}\rho\right)^2 \sum_{\substack{a \in G: \\ |a|=2}} z_a \varphi_a(x),$$

and using the definition of $f_{[0]}$, $f_{[1]}(x)$ and $f_{[2]}(x)$ given in Eqs. (44), (45) and (46) we have

$$\mathrm{E}(f(M_\rho(x))) = f_{[0]} + \left(1 - \frac{q}{q-1}\rho\right) f_{[1]}(x) + \left(1 - \frac{q}{q-1}\rho\right)^2 f_{[2]}(x), \tag{51}$$

which proves the statement. □

According to the previous theorem, the expected value of the fitness function after a mutation is a quadratic equation with respect to the mutation rate ρ. However, the analysis of this quadratic equation provides more information than in the case of the graph coloring problem. The reason is that the frequency assignment is not an elementary landscape and the behavior of the elementary components $f_{[1]}(x)$ and $f_{[2]}(x)$ is different, in general, from the behavior of $f(x)$.

We are interested in finding the expectation-optimal mutation rate for a given solution x. Let us consider two cases:

- Case $f_{[2]}(x) = 0$. In this case Equation (49) is a linear equation in ρ and the expectation-optimal mutation rate must be in one of the extremes of the interval: $\rho = 0$ or $\rho = 1$. In particular, since we are minimizing, the expectation-optimal mutation rate will be $\rho = 1$ if $f_{[1]}(x) > 0$ and $\rho = 0$ if $f_{[1]}(x) < 0$. In the case $f_{[1]}(x) = 0$ the mutation rate is irrelevant and the expected value of the fitness function is $f_{[0]}$.

- Case $f_{[2]}(x) \neq 0$. By deriving Equation (49) with respect to ρ we can find the vertex of the parabola, which is:

$$\rho^* = \frac{q-1}{q}\left(1 + \frac{f_{[1]}(x)}{2f_{[2]}(x)}\right) \tag{52}$$

If this vertex is in the interval $[0, 1]$, then it could be the expectation-optimal mutation rate. The expected value of the fitness function after a mutation when this mutation rate is used is

$$\mathrm{E}(f(M_{\rho^*}(x))) = f_{[0]} - \frac{f_{[1]}^2(x)}{4f_{[2]}(x)} \tag{53}$$

If this expected value is lower than the expected values in the extremes, $\mathrm{E}(f(M_0(x)))$ and $\mathrm{E}(f(M_1(x)))$, then ρ^* is the expectation-optimal mutation rate. Since, $f_{[0]}$ is the expected value of the fitness function when a random solution is selected from the search space, Equation (53) reveals that ρ^* is better than a simple random selection if $f_{[2]}(x) > 0$. If $f_{[2]}(x) < 0$ a random selection of the solution provides a lower expected value for the fitness function. Finally, if ρ^* is not in the interval $[0, 1]$, then the expectation-optimal mutation rate must be in the extremes: $\rho = 0$ or $\rho = 1$.

8 Conclusions

In this work we have provided closed-form formulas and algorithms for computing the exact moments of the frequency distribution for functions with bounded epistasis defined over length-n strings of a finite alphabet. The results presented in this paper are relevant in the context of combinatorial optimization problems in which candidate solutions are represented by q-ary strings.

In addition to the moments, we also derived an exact expression for the expected value of the fitness function after the application of a mutation operator. Using this expression it is possible to compute the mutation rate for which this expected value is optimum. These derivations are mainly of theoretical interest, and we present them as a step toward a more rigorous understanding of the statistics of combinatorial search spaces. However, we also have briefly discussed potential practical applications of this work. In particular, we presented a connection to previous work on approximating fitness distributions using a linear program that could in principle allow one to estimate the mutation rate that maximizes the probability of producing an offspring that has improving fitness. This result could be used for finding good parameter values during search, but it might also be useful for finding lower bounds on the probability of improvement, and therefore an upper bound for the runtime of a simple evolutionary algorithm. We leave this as an open research question.

Finally, we applied the derivations to the graph coloring problem, a well-known NP-hard problem, and the frequency assignment problem, another NP-hard problem with relevance in the telecommunication industry. For both problems, we presented exact closed-form formulas for the average value of the fitness function in Hamming regions around any state. We also derived, in both cases, the exact formula for the expected fitness of an offspring produced by uniform mutation at a specific rate and discussed the results obtained.

Acknowledgments

References

Aardal, K. I., van Hoesen, S. P. M., Koster, A. M. C. A., Mannino, C., and Sassano, A. (2003). Models and solution techniques for frequency assignment problems. *4OR*, 1(4):261 – 317.

Bäck, T. (1992). The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm. In Männer, R. and Manderick, B., editors, *Parallel Problem Solving from Nature 2*, pages 85–94, Amsterdam. Elsevier.

Böttcher, S., Doerr, B., and Neumann, F. (2010). Optimal fixed and adaptive mutation rates for the leadingones problem. In Schaefer, R., Cotta, C., Kolodziej, J., and Rudolph, G., editors, *PPSN (1)*, volume 6238 of *Lecture Notes in Computer Science*, pages 1–10. Springer.

Chicano, F. and Alba, E. (2011). Exact computation of the expectation curves of the bit-flip mutation using landscapes theory. In Krasnogor, N. and Lanzi, P. L., editors, *Proceedings of the Thirteenth Annual Conference on Genetic and Evolutionary Computation*, pages 2027–2034, Dublin, Ireland. ACM.

Chicano, F., Whitley, L. D., and Alba, E. (2011). A methodology to find the elementary landscape decomposition of combinatorial optimization problems. *Evolutionary Computation*, 19(4):597–637.

Chihara, L. and Stanton, D. (1990). Zeros of generalized Krawtchouk polynomials. *Journal of Approximation Theory*, 60:43–57.

Doerr, B., Jansen, T., Sudholt, D., Winzen, C., and Zarges, C. (2012). Mutation rate matters even when optimizing monotonic functions. *Evolutionary Computation*, (to appear), doi: 10.1162/EVCO_a_00055.

Eisenblätter, A. (2001). *Frequency Assignment in GSM Networks: Models, Heuristics, and Lower Bounds*. PhD thesis, Technische Universität Berlin.

Grover, L. K. (1992). Local search and the local structure of NP-complete problems. *Operations Research Letters*, 12:235–243.

Heckendorn, R. B. (1999). *Walsh Analysis, Epistasis, and Optimization Problem Difficulty for Evolutionary Algorithms*. PhD thesis, Colorado State University, Fort Collins, CO.

Heckendorn, R. B. (2002). Embedded landscapes. *Evolutionary Computation*, 10(4):345–369.

Jansen, T. and Wegener, I. (2000). On the choice of the mutation probability for the (1+1) EA. In Schoenauer, M., Deb, K., Rudolph, G., Yao, X., Lutton, E., Merelo, J., and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature VI*, volume 1917 of *Lecture Notes in Computer Science*, pages 89–98. Springer Berlin / Heidelberg.

MacWilliams, F. J. and Sloane, N. J. A. (1977). *The Theory of Error Correcting Codes*. North Holland Publishing Company.

Rana, S., Heckendorn, R. B., and Whitley, L. D. (1998). A tractable Walsh analysis of SAT and its implications for genetic algorithms. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, pages 392–397.

Reidys, C. M. and Stadler, P. F. (2002). Combinatorial landscapes. *SIAM Review*, 44(1):3–54.

Rényi, A. (1970). *Foundations of Probability*. Holden-Day.

Sutton, A. M., Whitley, L. D., and Howe, A. E. (2011a). Approximating the distribution of fitness over Hamming regions. In *Proceedings of Foundations of Genetic Algorithms XI*, pages 93–104.

Sutton, A. M., Whitley, L. D., and Howe, A. E. (2011b). Mutation rates of the (1+1)-EA on pseudo-boolean functions of bounded epistasis. In Krasnogor, N. and Lanzi, P. L., editors, *Proceedings of the Thirteenth Annual Conference on Genetic and Evolutionary Computation*, pages 973–980, Dublin, Ireland. ACM.

Sutton, A. M., Whitley, L. D., and Howe, A. E. (2012). Computing the moments of k-bounded pseudo-Boolean functions over Hamming spheres of arbitrary radius in polynomial time. *Theoretical Computer Science*, 425:58–74.

Terras, A. (1999). *Fourier Analysis on Finite Groups and Applications, Cambridge U. Press, Cambridge*. Cambridge University Press.

Witt, C. (2012). Optimizing linear functions with randomized search heuristics - the robustness of mutation. In Dürr, C. and Wilke, T., editors, *STACS*, volume 14 of *LIPIcs*, pages 420–431. Schloß Dagstuhl - Leibniz-Zentrum für Informatik.