NM Landscapes: Beyond NK

Narine Manukyan
University of Vermont (UVM)
Burlington, VT, USA
Narine.Manukyan@uvm.edu

Margaret J. Eppstein
University of Vermont (UVM)
Burlington, VT, USA
Maggie.Eppstein@uvm.edu

Jeffrey S. Buzas
University of Vermont (UVM)
Burlington, VT, USA
Jeffrey.Buzas@uvm.edu

ABSTRACT

For the past 25 years, *NK* landscapes have been the classic benchmarks for modeling combinatorial fitness landscapes with epistatic interactions between up to $K + 1$ of N binary features. However, the ruggedness of *NK* landscapes grows in large discrete jumps as K increases, and computing the global optimum of unrestricted NK landscapes is an NP-complete problem. Walsh polynomials are a superset of NK landscapes that solve some of the problems. In this paper, we propose a new class of benchmarks called *NM* landscapes, where M refers to the Maximum order of epistatic interactions between N features.*NM* landscapes are much more smoothly tunable in ruggedness than *NK* landscapes and the location and value of the global optima are trivially known. For a subset of *NM* landscapes the location and magnitude of global minima are also easily computed, enabling proper normalization of fitnesses. NM landscapes are simpler than Walsh polynomials and can be used with alphabets of any arity, from binary to real-valued. We discuss several advantages of *NM* landscapes over *NK* landscapes and Walsh polynomials as benchmark problems for evaluating search strategies.

1. INTRODUCTION

Simulated landscapes are widely used for evaluating search strategies, where the goal is to find the landscape location with maximum fitness value [2]. Without loss of generality and for notational simplicity, we assume function maximization, rather than minimization, throughout this paper. *NK* Landscapes [3] have been the classic benchmarks for generating landscapes with epistatic interactions. They are described by two parameters: N specifies the number of binary features or loci and K specifies the degree of epistatic interactions among the features, where the maximum order of interactions is $K + 1$ [2]. *NK* landscapes have been used in many applications and widely studied in theory, as they can generate landscapes with various ruggedness using different values of K.

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM ACM 978-1-4503-2881-4/14/07.
http://dx.doi.org/10.1145/2598394.2598403.

However, NK landscapes have several limitations as benchmark problems. Buzas and Dinitz [1] recently showed that the expected number of local peaks in *NK* landscapes rises in large discrete jumps as K is increased, but actually decreases as a function of the number of interaction terms in an equivalent parametric interaction model. Additionally, the problem of finding the location and value of the global optimum of unrestricted *NK* landscapes with $K > 1$ is NP-complete, although for restricted classes one can use dynamic programming or approximation algorithms [5].

Walsh polynomials are a superset of NK landscapes that overcome some of the limitations of NK landscapes (for example they allow more control over which interactions are present and one could construct Walsh polynomials that have known global maximum [4]). However, the way Walsh polynomials are defined limits there use to binary features. In this paper, we introduce a different, more flexible subset of general interaction models that we dub *NM* landscapes. *NM* landscapes transparently incorporate any number of epistatic feature interactions. Moreover, they also (a) have known value and location of the maximum fitness, (b) work with alphabets of any arity, including both discrete and real-valued alphabets, and (c) when coefficients are chosen properly, have smoothly tunable ruggedness. A subset of these landscapes also has known value and location of the minimum fitness.

2. NM LANDSCAPES

Parametric interaction models have been long used in statistics to study effects of multiple features on an outcome.

They are easy to define and the interactions are transparent and easy to interpret (unlike in *NK* landscapes or Walsh polynomials). A fitness landscape F can be defined for N binary features using an interaction model of the form:

$$F(\mathbf{x}) = \sum_{k=1}^{m} \beta_{U_k} \prod_{i \in U_k} x_i \quad (1)$$

where $x_i \in \{-1, 1\}$, $\mathbf{x} = [x_1, x_2, ...x_N]$, m is the number of terms, and each of m coefficients β_{U_k} can take any value in $\mathbb{R}$. For $k = 1 \ldots m$, we define $U_k \subseteq \{1, 2, \ldots, N\}$. We adopt the convention that when $U_k = \emptyset$, $\prod_{j \in U_k} x_j \equiv 1$.

In [1] the authors show that for every *NK* landscape with a given K, one can create an equivalent parametric interaction model, where the maximum order of interactions is $K + 1$. They show that the *NK* algorithm dictates that the interaction model contain all main effects and sub-interactions contained in higher order interactions. Thus, *NK* landscapes are a very restricted subset of general interaction models. While both NK landscapes and binary interaction models can be represented as Walsh polynomials, interaction models represent a larger set of functions as they naturally extend to real-valued features. However, in general the location and the value of the maximum fitness is unknown (for both general interaction models and Walsh polynomials).

We introduce a subset of interaction models called NM landscapes, where N is the number of features and all interactions in the model are of order $\leq M$. NM models comprise the set of all interaction models described by Eq. (1), with the added constraint that all coefficients β_{U_k} are non-negative. In this work, each β_{U_k} is randomly created as follows:

$$\beta_{U_k} = e^{-abs(\mathbb{N}(0,\sigma))} \quad (2)$$

where $\mathbb{N}(0,\sigma)$ is a random number drawn from a Gaussian distribution with 0 mean and standard deviation of σ. As the values of σ increase so does the standard deviation of the number of local peaks in the landscapes. We use $\sigma = N$ in our experiments. It is possible that other non-negative coefficient distributions may also yield landscapes with desirable characteristics. However, while any non-negative distribution will yield known location and value of the global maximum, not all non-negative distributions lead to landscapes with smoothly tunable number of local peaks.

Based on the above definition a maximum fitness $F_{optimal}$ of an NM landscape with a binary alphabet is achieved when all features are set to 1 ($x_i = 1$, $\forall i$) and the value of $F_{optimal}$ of an NM landscape with a binary alphabet is:

$$F_{optimal} = \sum_{i=1}^{N} \beta_{U_k} \quad (3)$$

Furthermore, NM landscapes can be defined on discrete alphabets of any arity or on real-valued alphabets, and the value and location of a global maximum is independent of the discretization of the alphabet. For any alphabet in the range $[-a, b]$, where $a \leq b$, the optimal fitness is achieved when all the features are equal to b ($x_i = b$, $\forall i$) and the optimal fitness is:

$$F(\mathbf{x}) = \sum_{\forall \beta_{U_k}} \beta_{U_k} b^{|U_k|} \quad (4)$$

There are many ways that the complexity of an NM landscape can be varied. For example, increasing the maximum order of interactions (M) will increase the ruggedness. In addition, in NM landscapes one could easily draw the random coefficients β_{U_k} for different orders of interactions from distributions with different maximum values (e.g., by varying σ in Eq. (2)). Our experiments show that the number of local peaks L in NM landscapes increases relatively smoothly (Fig. 1), and the lag 1 autocorrelation decreases relatively smoothly, as we increase the number of terms (m) in the model for a given maximum order of interaction M and as we increase the maximum order of interactions M (i.e., as we cross a vertical line on Fig. 1). In NM models one can easily control which interactions are present or absent. For example one can only include odd order interactions, which will result in landscapes, where the global minimum is equal to the global maximum with opposite sign.

3. CONCLUSIONS

In conclusion, NM landscapes are parametric interaction models that have non-negative coefficients. They naturally incorporate epistasis in a simple and transparent model. NM landscapes are well-defined for alphabets of any arity (from binary to real-valued), as long as the minimum value

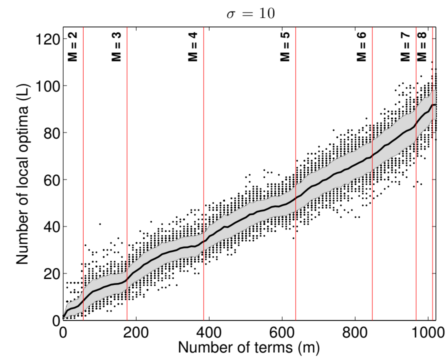

Figure 1: Number of local peaks in landscape as we increase the number of terms (x-axis) and maximum order of interactions M (labels near top), for NM landscapes with $N = 10$ and $\sigma = 10$. The gray area shows the standard deviation and black line shows the mean for 100 random NM landscapes.

in the alphabet is negative with absolute value $\leq$ the positive maximum value. This combination of constraints ensures that a global maximum is located at the point where all decision variables have their maximum value, with the optimal value equal to the sum of the model coefficients. By further restricting which combinations of interactions are present, a subset of NM models can also be constructed with known location and value of the global minimum. One can easily control or analyze which terms and interactions are present in an NM landscape. By using an appropriate non-negative distribution for the coefficients, the resulting NM landscape models have smoothly tunable ruggedness. The coefficient distribution can be tuned in a variety of other ways that control the distribution of fitnesses in the landscape, depending on what is most appropriate for a particular application. NM landscapes are thus a simple but powerful class of models that offer many benefits over NK landscapes and Walsh polynomials.

4. ACKNOWLEDGMENTS

This work was funded in part by the NIH Eunice Kennedy Shriver National Institute of Child Health & Human Development award #1R21HD068296.

5. REFERENCES

[1] J. Buzas and J. Dinitz. An analysis of NK landscapes: Interaction structure, statistical properties and expected number of local optima. *IEEE Transactions on Evolutionary Computation*, in press, DOI10.1109/TEVC.2013.2286352, 2014.

[2] S. Kauffman. *The origins of order: Self organization and selection in evolution.* Oxford University Press, 1993.

[3] S. A. Kauffman and E. D. Weinberger. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of theoretical biology*, 141(2):211–245, 1989.

[4] R. Tanese. *Distributed genetic algorithms for Function Optimization.* PhD thesis, The University of Michigan, Ann Arbor, MI, 1989.

[5] A. H. Wright, R. K. Thompson, and J. Zhang. The computational complexity of NK fitness functions. *IEEE Transactions on Evolutionary Computation*, 4(4):373–379, 2000.