Macroscopic features of the NK Landscape and Why Ruggedness Isn't Really the Problem for Search.

Jon Atwell

December 8, 2020

Search on Fitness Landscapes

Organizations face complicated decision-making environments. Often the challenges are linked to competition, but another fundamental obstacle is the fact that changing one aspect of the firm's organization or operation can have unpredictable effects on the performance of the whole. This unpredictability comes from the diverse interdependencies in the components, whether of the organization itself or its final products. Few, if any, components are siloed operationally from the others and altering the state of one alters the performance of others. If these interdependencies were known in advance, the effects would be predicable and decisions could assess the effects of all changes prior to making a decision. But given a plethora of interdependencies and the limited ability of managers to gather information about them, such interdependencies necessitate that organizations learn through search and experimentation.

The need to learn by doing raises the question of how to best search for good, or even optimal, solutions. Levinthal **Levinthal1997** started to formalize this question by importing Kauffman's NK model of genetic interdependencies (**Kauffman1989**) to management research. The NK model and the question of how to best search has loomed large in the organizational learning literature since (see **Baumann2019** for a review). The NK model has in fact been so fruitful that some have labeled a classic model **Borgonovo2020** This paper takes the position that understanding how to best search the relevant decision-making environment is the fundamental

1

question of organizational learning, but that the study of search behaviors can never be decoupled from the study of landscape over which search is taking place. The features of the landscape determine how successful various search behaviors will be. This claim is the central impetus.

Modifications to the NK and alternative landscapes have been proposed and results clearly indicate that best search behavior is indeed highly dependent upon the features of the landscape **Rivkin2007**; **Winter2007**; **Callander2011**; **Rahmandad2019** Some have gone so far as to suggest landscape design as goal of organizations **Levinthal1999**

The theoretical problem of organizational learning is thus both a question of what type of environment the firm is facing and how to best search it. If the firm knows what type of environment it is facing, the question of how to search is much more tractable, but from a firm's perspective the former is generally not knowable and it must learn about both the features of the environment and how to search it ~~by doing~~.

But the learning strategy firms deploy can be influenced by the theoretical work on the topic and this paper argues that a key insight of the NK model is misleading. That putative insight is that ruggedness—the existence of many local optima—is an impediment to successful search or imitation because local optima confuse boundedly rational actors. The assumption **(confusing statement, what assumption ??)** is that any particular local optimum looks better than all alternatives if the actor cannot construct knowledge about broader patterns in the landscape. That form of bounded rationality, however, might not meaningfully describe human actors. Certainly, human rationality is bounded by a variety of cognitive limitations, but if human cognition is good at one thing, it is extracting patterns and trends. This coarsening of diverse information into abstract models is what allows us to problem solve better than any other species and even natural selection itself in some domains. So while model-based search is boundedly rational insofar as models can be constructed based on samples and not complete information, thereby necessarily accommodating errors, humans might not be stymied by local features like ruggedness.

In models of search on the canonical NK landscape, actors generally do not perform model-based search. Instead they are imbued with something that might be called Markovian rationality; they forget prior states and base next moves solely on information accessible from the current state **- a myopic search strategy**. This ensures that progress is slow and often needs rules about when to throw any sense of rationality into the wind and roll

2

the dice to start over from a random point **(Maybe frame this in terms of exploration vs. exploitation)**. Similar behavior has been observed in human subject experiments **Billinger2014**; **Vuculescu2017** and this suggests that humans are not model-based learners, are stymied by ruggedness and are in fact Markovianly rational.

This paper argues no such inferences can be made based on the NK model because the primary tunable parameter of the model tunes both micro **(e.g. local changes in fitness)** and macroscopic features of the landscape **(such as the dispersion of local optima and gradient of the landscape from the global minimum to maximum)**. We demonstrate that the macroscopic changes are antithetical to model-based learning and are in fact the cause of suboptimal search outcomes, not ruggedness. The upshot of this analysis is that ruggedness is likely not true obstacle in search and can be in some settings a source of positive learning because it encourages broader sampling for the purpose of constructing models.

Mostly from interdependencies between subunits, but scholars of organizations have yet to deeply consider the *manifestations* of that complexity. (For a recent exception, see **Rahmandad2019** in MGMT studies. See **Newman1998** in biology.) In particular, while the NK model of rugged landscapes **Kauffman1989**; **Levinthal2007**; **Rivkin2000** has been generative for theorizing about organizational decision-making and design (one of the two most fundamental models in the behavioral theory of the firm, along with the garbage can model), remarkably little has been said within the management and organizational sciences literature about the mathematical characteristics of the model and how those relate to search behavior.

The feature of "ruggedness" for which the NK model is known has been well documented, but ruggedness is a micro-level feature of the landscape. As a model encoding interdependencies, the NK landscape naturally tunes macro-level and other micro-level features in the course of tuning ruggedness. Thus, effects on outcomes variables related to search cannot be directly attributed to ruggedness because the other features are not controlled for. On first blush, this might seem beside the point. After all Levinthal's goal in importing the model from biology was to make the point that interdependencies make search and decision making hard. While that is true, Kauffman's decision maker in the model was nature herself (in spite of the model not introducing any of the selective dynamics found in population genetics.) Natural selection, while an incredible search tool, does not learn in a model-based way has humans do and is easily stymied by local optima.

In that sense, local ruggedness the primary feature Kauffman needed to make his point and macro-level features did not really matter.

Features of the NK model

Kauffman's NK model was proposed to capture interactions between alleles in DNA. In this setting, there are epistatic interactions between alleles, meaning the protein defined by the whole segment depends on the (nucleo)-base present at all alleles in the segment. The model skips modeling the protein and instead goes straight to defining the fitness of the protein. To do this, it looks at a segment of length N and then defines a substring length parameter K. The epistatic interactions for a given allele are then restricted to K other alleles, and each of the 2^{K+1} permutations is assigned a fitness drawn uniform at random.

The choice to assign fitness this way was justified by noting that we have little-to-no knowledge about the interaction effects. Which interactions are present is arbitrary for the same reason. Valuable work has gone into modeling the interaction structure (c.f. Rivkin and Siggelkow), but to our knowledge studying the effects of different distributions over the fitness value of the substrings has not been done. [Someone has to have done some of this...we should look into it.]

A straightforward implication of this formulation is that changing a single bit within the substring has the same expected effect on substring fitness as changing all of them, suggesting the model places significantly more weight on higher order interaction effects than the main effects. Indeed, Buzas and Dinitz link the NK model to linear regression models with higher order interaction effects and show that for low K models the main effects are dominant. For higher K, when the interactions actually come into play, the higher order interactions have the same expected magnitude as the main effects.

Finding evidence of such effects in an empirical setting would be quite shocking, but our goal is not to critique the model specification. Instead we focus the chief emergent feature of the landscape, *ruggedness*, or the count of local optima. It is this feature that figures large in management literature because of the putative inability of boundedly rational agents to conduct intelligent search when local optima abound. While that may well be the case, the tunable parameter of the NK model alters more than ruggedness.

As noted above it increases the importance of higher order interaction effects and decreases the importance of main effects, which has consequences

for the macroscopic structure. In particular, the NK landscape has a "central massif" for low K in which local optima cluster in the search space around the global optimum, but as K increases this local optima become dispersed around the space in small clusters. Leaving those smaller clusters through local search is unlikely and long distance search is likely to results in being in very similar clusters, not the one with the global optimum. All of this is coming from Herrman, Ochoa and Rothlauf, 2016

These results that the primary source of search failure or frustration is the restructuring of the local optima networks, not the presence of many local optima. We propose here a model that allows us to disentangle the effects of local optima count and their correlations within space. The Dirichlet distribution plays a crucial element in the setup so we next introduce it.

What is the Dirichlet distribution?

The Dirichlet family of distributions is the multivariate Beta distribution, which is defined by K, the number of dimensions or categories and $\boldsymbol{\alpha}$, a K-length vector of a single positive real $\{\alpha_1, alpha_2, ..., \alpha_K\}$. Each α_i in this $\boldsymbol{\alpha}$ vector of *concentration* parameters determines the mean for the values along the corresponding dimension of draws from the distribution. The mean is $\frac{\alpha_i}{\sum_i^K \alpha_i}$. Accordingly, when $a_i = a_j$ for all i, j the expected value of a draw from the dirichlet is $1/K$ for all dimensions.

When the α's are not the same, $E[X_i] = \frac{\alpha_i}{\alpha_0}$ where $\alpha_0 = \sum_i^K \alpha_i$. Accordingly dimensions corresponding higher α's have greater expected values.

The variance is the interesting moment of the family. It is defined as follows: $Var[X] = \frac{\tilde{\alpha}_i(1-\tilde{\alpha}_i)}{1+\alpha_0}$ where $\alpha_0 = \sum \alpha_i$ and $\tilde{\alpha}_i = \frac{\alpha_i}{\alpha_0}$

Because the denominator is $\alpha_0 + 1$, $Var[X_i]$ is inversely related the sum of the concentration parameters. So the variances for $\alpha = 1$ (i.e. the symmetrical case) are roughly 100 times greater than $\alpha = 100$.

Defining the Dirichlet Dot Product Landscape

We define a fitness landscape using two entities; a Euclidean Vector Space $\vec{E}_N$ over $\mathbb{R}^N$ and a Dirichlet K of dimension N parameterized by $\boldsymbol{\alpha}$. N is the same positive integer in both entities and corresponds to the dimensionality

of search problem, as in the NK model.

Let location $\langle \ell \rangle \in \vec{E_N}$.

Let $X = (X_1, X_2, ..., X_N)$ be a draw from the Dirichlet K with $\boldsymbol{\alpha}$.

The fitness of location $\langle \ell \rangle$ is then simply
$F(\langle \ell \rangle) = \langle \ell \rangle^T \bullet X$ where $\bullet$ denotes the dot product (i.e. $\ell_1 \cdot X_1 + \ell_2 \cdot X_2 + ... + \ell_N \cdot X_N$)

Note that because the vector space is affine, the dimensions can be re-oriented so that (1,1,...1) and (0,0,...,0) are no longer the maximum and minimum; an arbitrary mask can be applied to the locations without effects that properties of the landscape.

The family of landscapes encompassed by this model is expansive because, as defined, $\vec{E}$ covers the real number line. In discussing the model, we will consider $\langle 0, 1 \rangle^N$ (the continuous unit vector space) because of the ease of interpretability. In the analysis, we will restrict ourselves to the hypercube $\{1, 0\}^N$, as is standard in the NK literature. Other interesting domains include the following:

- $\{-1, 1\}^N$, allowing for negative contributions

- $\{-1, 0, 1\}^N$, allowing null and negative contributions

- $[-1, 1]$, continuous case but with negative contributions

- $[0, M_i]$, with $M_i \neq M_j$ for some i, j (i.e. allowing asymmetrical domains)

- $\{0, 1/M, 2/M...1\}^N$, a discretized domain.

- Any combinations of the above, to emphasize the heterogeneity of component contributions.

Implications

Coupling in this model is quite different than in the NK model. Rather than having precise functional interdependencies (with imprecise effects), we assume that there are interdependencies across all components and instead control the strength of the interdependencies though the sum of the α_i values. Mathematically, the components are interdependent because the fitness contribution of a given component is assigned through the Dirichlet draw and, in virtue of all probabilities needing to sum to one, an increase in one component comes at the expense of one or more other components.

Starting with the case when $\alpha = 1$ (for all dimensions, the symmetrical and flat dirichlet) the variance of the value for each dimension is $.00\overline{81}$. The expected value of the difference between the value of two draws from the dirichlet is approximately $.09$ [I simulated this...is there a simple reason it should be is?]. In expectation half of the changes are positive, half negative.

The Dirichlet is not the sole source of the structure of interdependencies, however. The contribution to fitness of a given attribute/dimension depends on the states of all dimensions. To see this clearly, let $F(\ell_j)$ be the fitness of a location ℓ_j and let X_{i,ℓ_j} be the value of the Dirichlet draw X_{i,ℓ_j} for the ith dimension and for location ℓ_j.

Now consider the location $\ell_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. The fitness of the location, F_{ℓ_1}, is the value of the draw for only the first dimension, X_{1,ℓ_1} because all other dimensions are multiplied by zero before being added. Next consider $\ell_2 = (1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$. The second dimension of the draw, X_{2,ℓ_2} is added to the overall fitness for a value of $F(\ell_2)$. But crucially, $F(\ell_2) \neq F(\ell_1) + X_{2,\ell_2}$ because X_{ℓ_2} is a different draw from the Dirichlet than X_{ℓ_1} and the contribution of first dimension is changed (i.e. $X_{1,\ell_1} \neq X_{1,\ell_2}$).

The expected value of the addition of the contribution of the second dimension is greater than the expected change in contribution of the first dimension. Accordingly, the main effect of changing the state of the second dimension to 1 is positive. However, because the variance in the Dirichlet draw, there is a difference between total fitness of the location $F(\ell_2)$ and the fitness of $\ell_1 + X_{2,\ell_2}$. This difference, $X_{1,\ell_1} - X_{1,\ell_2}$, is the magnitude of the interaction between dimensions/components 1 and 2. It can be either positive or negative, suggesting positive or negative complementarity between them.

Expanding this reasoning to all dimensions, we can say that $X_{1,\ell_1} - X_{1,\ell_j}$ is the net complementarity of dimension 1 with all other dimensions. One can reconstruct the individual complementaries by traversing the shortest

path from the location with a single 1 on the focal dimension, call it ℓ_H (for home), to arbitrary location ℓ_j. The differences between fitnesses at ℓ_H and the locations within a single step are the first order interaction terms. The difference between the locations one step away and those two steps away, minus the first order interaction term is the value of the second order interaction and so on.

In order for a landscape to be complex, there need to be negative complementaries and the Dirichlet ensures these are present in expectation for lower values of alpha. If there weren't consistently negative complementaries, the search problem is straightforward hillclimbing. Such negative complementaries are present in the NK model, increasingly so as K increases. However, they are substantial in magnitude for high K and have the same expected magnitude as the main effects. This yields ruggedness, but also wide swings in fitness and a disintegration of macroscopic structure.

The Dirichlet dot product model permits control of the expected value of the main effect relative to the higher order effects, which allows one to create landscape ensembles with both ruggedness and structure to the search problem. For example, when $N = 10$ and $\alpha = 2$, the distance from the top quartile of fitness scores to the global maximum and the autocorrelation structure is closely comparable to NK models with K=0. However, whereas the latter has a single global and local optimum, the former has s a count of local optima on the order of 50. It is both rugged and searchable.

The Dirichlet dot product also has parameters for which is in comparable to high K NK models. For example, when $N = 10$ and $\boldsymbol{\alpha} = 100$, it has a single global and local optimum and comparable autocorrelation structure.

It also permits exploring the consequences of asymmetrical α's. This is a natural choice, as so components are likely to impact fitness more than others. The fact that the variance decreases as the weight increases also is a natural correspondence in so far as functionally central choices are less likely to be affected by peripheral choices.

Notes about fitness distributions

Why the low concentration parameters lead to a bi-modal distribution of fitnesses:

For concentration parameters ¡1, the density of the PDF is highest in the corners of the simplex. The lower the parameter, the more extreme the shift. So while the expected value of a draw for single dimension is still 1/N and

8

the sampling process is ergodic, 1 out of every N draws the value for the dimension is close to 1. The other N-1 draws are close to zero.

So let's make the concentration parameters extreme, say alpha=.0001 so that a draw is more or less the selection of a single dimension as certain. When thinking about the dot product then, we should note that for a given bit/characteristics in the location vector, in half of the landscape that bit is equal to 1 and for the other half it is 0. This suggests that the dimension that gets a 1 from the dirichlet draw will have a corresponding location bit equal to one with probably .5. Then in taking the dot product of the dirichlet draw and the location, the fitness will be equal to 1. The other times it is 0. Hence the bimodal PDF of the fitness values.

When alpha is not that extreme, say .05, there are some locational bits that have a positive dirichlet weighting and their inclusion in the dot product yields fitnesses that start to fill in the range between 0 and 1.

I think we can extend this style of reasoning to the other extreme alpha values. When alpha is high, say 100, the expected value of any given dimension is again. 1/N and the variance is $N/100 * (1/100 * N + 1)$. So for $N = 10$, $var(a_i) = .00008$. Accordingly, the primary driver on the fairly normally distributed fitness values is the location vector. The weighting for each is the same so the dot product depends on the locations only. Because the hamming distances are distributed normally, the fitness are too.