

Eigen's Evolutionary Model, Two-Valued Fitness Landscapes, and Isometry Groups Acting on Finite Metric Spaces

Artem Novozhilov

Department of Mathematics
North Dakota State University

Colloquium,
Fargo, North Dakota, USA

September 8, 2015

Collaborators:



Yuri S. Semenov

*Moscow State University of Railway
Engineering, Moscow, Russia*



Alexander S. Bratus

*Lomonosov Moscow State University,
Moscow, Russia*

Mathematical statement of the problem:

Consider the following eigenvalue problem

$$QW\mathbf{p} = \lambda\mathbf{p},$$

for the matrices

$$W = \text{diag}(w_0, \dots, w_{l-1}), \quad w_i \geq 0, \quad l = 2^N,$$

and $Q = (q_{ij})_{2^N \times 2^N}$, such that

$$q_{ij} = q^{N-H_{ij}}(1-q)^{H_{ij}}, \quad i, j = 0, \dots, l-1.$$

Here H_{ij} is the Hamming distance between the binary representations of indices i and j , and $q \in [0, 1]$ is a constant.

Since QW is non-negative and primitive then by the *Perron-Frobenius theorem* there exists a strictly dominant eigenvalue $\bar{w} > |\lambda_j|$ with a positive eigenvector $\mathbf{p} \in \mathbf{R}^l$.

Mathematical statement of the problem:

Consider the following eigenvalue problem

$$QW\mathbf{p} = \lambda\mathbf{p},$$

for the matrices

$$W = \text{diag}(w_0, \dots, w_{l-1}), \quad w_i \geq 0, \quad l = 2^N,$$

and $Q = (q_{ij})_{2^N \times 2^N}$, such that

$$q_{ij} = q^{N-H_{ij}}(1-q)^{H_{ij}}, \quad i, j = 0, \dots, l-1.$$

Here H_{ij} is the Hamming distance between the binary representations of indices i and j , and $q \in [0, 1]$ is a constant.

Since QW is non-negative and primitive then by the *Perron-Frobenius theorem* there exists a strictly dominant eigenvalue $\bar{w} > |\lambda_j|$ with a positive eigenvector $\mathbf{p} \in \mathbf{R}^l$.

Problem: Given W and q determine (or approximate) \bar{w} .

Motivation: Eigen's evolutionary problem

Consider a population of sequences of fixed length N composed of zeros and ones, hence we have $l = 2^N$ different types of sequences. Sequences reproduce and mutate in discrete time. The reproduction of each sequence type indexed by the variable $a = 0, \dots, l - 1$ is determined by its fitness value w_i . The correspondence between the index and the sequence itself is determined through the binary representation

$$a = \alpha_0 + \alpha_1 2 + \dots + \alpha_{N-1} 2^{N-1} = [\alpha_0, \alpha_1, \dots, \alpha_{N-1}], \quad \alpha_k \in \{0, 1\}.$$

During the reproduction the sequences also mutate. We assume that the mutations at different sites are independent and the fidelity (i.e., the probability of the error-free reproduction) per site per replication is given by the same constant $0 \leq q \leq 1$ for each site. Then, invoking simple probabilistic rules, the probability q_{ij} that sequence j will mutate to sequence i is

$$q_{ij} = q^{N-H_{ij}} (1 - q)^{H_{ij}}, \quad i, j = 0, \dots, l - 1,$$

and H_{ij} is the Hamming distance between sequences i and j .

Motivation: Eigen's evolutionary problem

If $\mathbf{p}(t+1)$ denotes the vector of frequencies of different sequences at time $t+1$ then simple bookkeeping leads to the discrete dynamical system

$$\mathbf{p}(t+1) = \frac{\mathbf{QW}\mathbf{p}(t)}{\bar{w}(t)}, \quad \bar{w}(t) = \sum_{i=0}^{l-1} w_i p_i(t).$$

The quantity $\bar{w}(t)$ is called the mean population fitness. It can be shown that there exists a unique globally stable stationary point \mathbf{p} of this system, which is given by the positive eigenvector of the eigenvalue problem

$$\mathbf{QW}\mathbf{p} = \lambda\mathbf{p},$$

corresponding to the strictly dominant eigenvalue $\lambda = \bar{w} = \sum w_i p_i$.

Motivation: Eigen's evolutionary problem

If $\mathbf{p}(t+1)$ denotes the vector of frequencies of different sequences at time $t+1$ then simple bookkeeping leads to the discrete dynamical system

$$\mathbf{p}(t+1) = \frac{QW\mathbf{p}(t)}{\bar{w}(t)}, \quad \bar{w}(t) = \sum_{i=0}^{l-1} w_i p_i(t).$$

The quantity $\bar{w}(t)$ is called the mean population fitness. It can be shown that there exists a unique globally stable stationary point \mathbf{p} of this system, which is given by the positive eigenvector of the eigenvalue problem

$$QW\mathbf{p} = \lambda\mathbf{p},$$

corresponding to the strictly dominant eigenvalue $\lambda = \bar{w} = \sum w_i p_i$.

Problem: Given the *fitness landscape* W and *fidelity* q determine (or approximate) the mean equilibrium population fitness \bar{w} .

Motivation: Ising model

Consider a rectangular lattice, with each vertex (particle) can be in two states (spins). The interaction between the particles tend to align the spins whereas the thermal movement has a randomizing effect. At a critical temperature the latter becomes so strong that a phase transition occurs from ordered into the disordered phase. If the interactions do not span more than two neighboring rows the transfer matrix method solves the problem of finding the partition function at least formally.

Motivation: Ising model

Consider a rectangular lattice, with each vertex (particle) can be in two states (spins). The interaction between the particles tend to align the spins whereas the thermal movement has a randomizing effect. At a critical temperature the latter becomes so strong that a phase transition occurs from ordered into the disordered phase. If the interactions do not span more than two neighboring rows the transfer matrix method solves the problem of finding the partition function at least formally.

Problem: Given the energy function $E(c)$, which depends on the details of interaction between the particles and inverse temperature β determine the ground state energy \bar{w} , which can be found as the leading eigenvalue of the transfer matrix.

A personal remark:

While the problem stated above is quite natural, its general analytical solution is probably beyond our reach. Note that even numerically one has to deal with problems of dimension $2^N \times 2^N$, which is unrealistic even for moderate biologically relevant values of N (say, of order 100). Therefore, some specific simplifications should be made to achieve even partial progress.

A personal remark:

While the problem stated above is quite natural, its general analytical solution is probably beyond our reach. Note that even numerically one has to deal with problems of dimension $2^N \times 2^N$, which is unrealistic even for moderate biologically relevant values of N (say, of order 100). Therefore, some specific simplifications should be made to achieve even partial progress.

The only known case with an exact explicit analytical solution is given for the multiplicative fitness landscape, when

$$w_i = \prod_{k: \alpha_k=1} r_{k+1}, \quad i = 1, \dots, l-1, \quad w_0 = 1,$$

where $0 < r_j < 1$, $i = 1, \dots, N$ is a multiplicative contribution of the j -th site.

Permutation invariant fitness landscapes:

Another way to be able to achieve some progress is to reduce the dimensionality of the problem from $2^N \times 2^N$ to $(N + 1) \times (N + 1)$ by considering the permutation invariant fitness landscapes

$$w_i = w_{H_i}, \quad i = 0, \dots, 2^N - 1,$$

and $H_i := H_{0i}$ is the Hamming norm of sequence i , i.e., the number of ones in this sequence. Therefore we find that instead of following 2^N types of sequences we need to calculate only $N + 1$ classes of sequences (e.q., class 4 includes all possible sequences with 4 ones, the total number is $\binom{N}{4}$). Now

$$q_{kl} = \sum_{i=l+k-N}^{\min\{k,l\}} \binom{k}{i} \binom{N-k}{l-i} (1-q)^{k+l-2i} q^{N-(k+l-2i)}, \quad k, l = 0, \dots, N,$$

are the mutation probabilities from class l to class k .

Example:

Consider the simplest possible permutation invariant fitness landscape

$$\mathbf{W} = \text{diag}(w + s, w, \dots, w), \quad w > 0, s > 0.$$

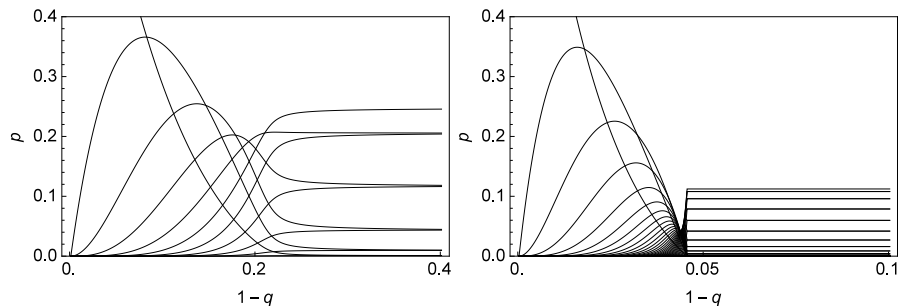
which is called the single (or sharply) peaked landscape. Using the trick from the previous slide we can numerically solve the problem and find

Example:

Consider the simplest possible permutation invariant fitness landscape

$$\mathbf{W} = \text{diag}(w + s, w, \dots, w), \quad w > 0, s > 0.$$

which is called the single (or sharply) peaked landscape. Using the trick from the previous slide we can numerically solve the problem and find



Here on the left $N = 10$ and on the right $N = 50$.

Maximum principle:

For the permutation invariant fitness landscapes also a very powerful approximations for \bar{w} can be found:

Theorem: Let the fitness values of a permutation invariant fitness landscape have a continuous approximation

$$w_i = W\left(\frac{i}{N}\right) + \mathcal{O}\left(\frac{1}{N}\right),$$

and

$$q = \frac{\mu}{N},$$

where μ is independent of N . Then

$$\bar{w} \approx \sup_{x \in [0,1]} \left\{ W(x) \exp\left(-\left(\sqrt{\mu(1-x)} - \sqrt{\mu x}\right)^2\right) \right\}.$$

Ref: Baake, Georgii: J Math Bio, 2007, 54:257-303

Open questions:

- ▶ What if there is no continuous approximation for w_i ?
- ▶ Can we also find the stationary distribution \mathbf{p} ?
- ▶ Can we work with permutation non-invariant fitness landscapes?

Open questions:

- ▶ What if there is no continuous approximation for w_i ?
- ▶ Can we also find the stationary distribution \mathbf{p} ?
- ▶ Can we work with permutation non-invariant fitness landscapes?

Two valued fitness landscapes:

Consider again the eigenvalue problem

$$QWp = \lambda p,$$

where

$$q_{ij} = q^{N-H_{ij}}(1-q)^{H_{ij}}, \quad i, j = 0, \dots, 2^N - 1,$$

and $W = \text{diag}(w_0, \dots, w_{l-1})$, $l = 2^N$. For the following I will assume that

$$w_i = \begin{cases} w + s, & i \in A, \\ w, & i \notin A, \end{cases}$$

where $A \subseteq \{0, 1, \dots, l-1\}$ and $w \geq 0$, $s > 0$. That is

$$W = wI + sE_A = wI + s \sum_{a \in A} E_a,$$

where E_a is the matrix with element $e_{aa} = 1$ and all other entries being zero.

Some technicalities:

After some (not elementary) manipulations, it can be shown that

$$p_b = \frac{s}{l} \sum_{a \in A} \sum_{k=0}^{l-1} \frac{(2q-1)^{H_k} t_{bk} t_{ak}}{\bar{w} - w(2q-1)^{H_k}} p_a,$$

and since only the components p_a , $a \in A$ are involved in the right hand side, I can rewrite the last problem in the matrix form

$$\mathbf{p}_A = \mathbf{M} \mathbf{p}_A, \quad \mathbf{M} = (m_{ab})_{|A| \times |A|}.$$

Here t_{ij} are the elements of matrix \mathbf{T} , which I know how to calculate explicitly and which puts \mathbf{Q} into diagonal form.

After yet another (not elementary) manipulation, I can represent the elements of the matrix \mathbf{M} as

$$m_{ab} = \frac{s}{l\bar{w}} \sum_{c=0}^{\infty} \left(\frac{w}{\bar{w}}\right)^c (1 - (2q-1)^{c+1})^{H_{ab}} (1 + (2q-1)^{c+1})^{N-H_{ab}}$$

Some technicalities:

From $\mathbf{p}_A = \mathbf{M}\mathbf{p}_A$ I have that

$$\sum_{b \in A} p_b = \sum_{b \in A} \sum_{a \in A} m_{ba} p_a = \sum_{a \in A} p_a \sum_{b \in A} m_{ba}.$$

Some technicalities:

From $\mathbf{p}_A = \mathbf{M}\mathbf{p}_A$ I have that

$$\sum_{b \in A} p_b = \sum_{b \in A} \sum_{a \in A} m_{ba} p_a = \sum_{a \in A} p_a \sum_{b \in A} m_{ba}.$$

Key assumption: The sum $\sum_{b \in A} m_{ab}$ does not depend on $a \in A$.

Some technicalities:

From $\mathbf{p}_A = \mathbf{M}\mathbf{p}_A$ I have that

$$\sum_{b \in A} p_b = \sum_{b \in A} \sum_{a \in A} m_{ba} p_a = \sum_{a \in A} p_a \sum_{b \in A} m_{ba}.$$

Key assumption: The sum $\sum_{b \in A} m_{ab}$ does not depend on $a \in A$.

This implies that $\sum_{b \in A} m_{ba} = 1$ and, from the calculations in the previous slide

$$\frac{l\bar{w}}{s} = \sum_{c=0}^{\infty} \left(\frac{w}{\bar{w}}\right)^c \sum_{b \in A} (1 - (2q - 1)^{c+1})^{H_{ab}} (1 + (2q - 1)^{c+1})^{N - H_{ab}},$$

if the inner sum does not depend on $a \in A$.

Some technicalities:

From $\mathbf{p}_A = \mathbf{M}\mathbf{p}_A$ I have that

$$\sum_{b \in A} p_b = \sum_{b \in A} \sum_{a \in A} m_{ba} p_a = \sum_{a \in A} p_a \sum_{b \in A} m_{ba}.$$

Key assumption: The sum $\sum_{b \in A} m_{ab}$ does not depend on $a \in A$.

This implies that $\sum_{b \in A} m_{ba} = 1$ and, from the calculations in the previous slide

$$\frac{l\bar{w}}{s} = \sum_{c=0}^{\infty} \left(\frac{w}{\bar{w}}\right)^c \sum_{b \in A} (1 - (2q - 1)^{c+1})^{H_{ab}} (1 + (2q - 1)^{c+1})^{N - H_{ab}},$$

if the inner sum does not depend on $a \in A$.

Question: When does our *key assumption* hold?

Main proposition:

Consider 1-skeleton of the N -dimensional cube $[0, 1]^N$ with the set of vertices V . The vertices a and b are connected by the (unique) edge e_{ab} if $H_{ab} = 1$. The Hamming distance between vertices u and v is the length of a shortest path connecting these vertices, that is, the number of edges in this path. The set V , due to the binary representation

$$a = \alpha_0 + \alpha_1 2 + \cdots + \alpha_{N-1} 2^{N-1} = [\alpha_0, \alpha_1, \dots, \alpha_{N-1}], \quad \alpha_k \in \{0, 1\},$$

can be identified with the set of indices $X = X_N = \{0, 1, \dots, 2^N - 1\}$ with the Hamming distance.

Main proposition:

Consider 1-skeleton of the N -dimensional cube $[0, 1]^N$ with the set of vertices V . The vertices a and b are connected by the (unique) edge e_{ab} if $H_{ab} = 1$. The Hamming distance between vertices u and v is the length of a shortest path connecting these vertices, that is, the number of edges in this path. The set V , due to the binary representation

$$a = \alpha_0 + \alpha_1 2 + \cdots + \alpha_{N-1} 2^{N-1} = [\alpha_0, \alpha_1, \dots, \alpha_{N-1}], \quad \alpha_k \in \{0, 1\},$$

can be identified with the set of indices $X = X_N = \{0, 1, \dots, 2^N - 1\}$ with the Hamming distance.

Proposition. Let G be a group that acts on the metric space $X = \{0, 1, \dots, l - 1\}$ by isometries (i.e., $G \leq \text{Iso}(X)$) and let A be a G -orbit. Then the equality

$$\frac{l\bar{w}}{s} = \sum_{c=0}^{\infty} \left(\frac{w}{\bar{w}}\right)^c \sum_{b \in A} (1 - (2q - 1)^{c+1})^{H_{ab}} (1 + (2q - 1)^{c+1})^{N - H_{ab}},$$

holds.

Proof Since G acts transitively on A and preserves the Hamming distance H_{ab} , the inner sum in the equality does not depend on $a \in A$.

Main result:

Corollary. Let G be a group that acts on the metric space $X = \{0, 1, \dots, l-1\}$ by isometries (i.e., $G \leq \text{Iso}(X)$) and let A be a G -orbit. Then the leading eigenvalue of the eigenvalue problem

$$QWp = \bar{w}p,$$

where

$$w_i = \begin{cases} w + s, & i \in A, \\ w, & i \notin A, \end{cases}$$

where $A \subseteq \{0, 1, \dots, l-1\}$ and $w \geq 0$, $s > 0$, can be found as a root of an algebraic equation of degree at most $N + 1$.

Let

$$F_A(z) := \frac{1}{2^N} \sum_{b \in A} (1-z)^{H_{ab}} (1+z)^{N-H_{ab}} = \sum_{d=0}^N h_d z^d.$$

Then this equation has the form

$$\sum_{d=0}^N \frac{h_d (2q-1)^d}{\bar{w} - w(2q-1)^d} = \frac{1}{s}.$$

Examples:

Example 1. Let $G = S_N$, the symmetric group, which acts on X by isometries and is a proper subgroup of $\text{Iso}(X_N)$. The S_N orbits are the subsets of

$$A_p = \{a \in X \mid H_a = p\}, \quad p = 0, 1, \dots, N.$$

Then the polynomial

$$F_{A_p}(z) = \frac{1}{2^N} \sum_{k=0}^p \binom{p}{k} \binom{N-p}{k} (1-z)^{2k} (1+z)^{N-2k} = \sum_{d=0}^N h_d z^d$$

defines the algebraic equation

$$\sum_{d=0}^N \frac{h_d (2q-1)^d}{\bar{w} - w(2q-1)^d} = \frac{1}{s}$$

for the dominant eigenvalue.

Examples:

Example 1.1 Let $p = 0$ from the previous (or, equivalently, consider the trivial group $G = \{1\}$). Then the algebraic equation for the leading eigenvalue takes the form

$$\frac{1}{2^N} \sum_{d=0}^N \binom{N}{d} \frac{(2q-1)^d}{\bar{w} - w(2q-1)^d} = \frac{1}{s}.$$

Examples:

Example 1.1 Let $p = 0$ from the previous (or, equivalently, consider the trivial group $G = \{1\}$). Then the algebraic equation for the leading eigenvalue takes the form

$$\frac{1}{2^N} \sum_{d=0}^N \binom{N}{d} \frac{(2q-1)^d}{\bar{w} - w(2q-1)^d} = \frac{1}{s}.$$

Example 1.2 Let $p = 1$, i.e., $A_1 = \{1, 2, 4, 8, \dots, 2^{N-1}\}$. Then the algebraic equation for the leading eigenvalue takes the form

$$\frac{1}{2^N} \sum_{d=0}^N \frac{(N-2d)^2}{N} \binom{N}{d} \frac{(2q-1)^d}{\bar{w} - w(2q-1)^d} = \frac{1}{s}.$$

Examples:

Example 2. According to Cayley's theorem each finite group G can be embedded into symmetric group S_n , $n = |G|$. Moreover, since there are standard embeddings $S_n \rightarrow S_{n+1} \rightarrow S_{n+2} \rightarrow \dots$ there is no problem, in principle, to construct the action of any finite group G on the set X_N for $N \geq n$. This gives us a virtually unlimited list of the two-valued fitness landscapes, which are not permutation invariant.

Examples:

Example 2. According to Cayley's theorem each finite group G can be embedded into symmetric group S_n , $n = |G|$. Moreover, since there are standard embeddings $S_n \rightarrow S_{n+1} \rightarrow S_{n+2} \rightarrow \dots$ there is no problem, in principle, to construct the action of any finite group G on the set X_N for $N \geq n$. This gives us a virtually unlimited list of the two-valued fitness landscapes, which are not permutation invariant.

Let

$$G = Q_8 = \{\pm 1, \pm i, \pm j, \pm k \mid i^2 = j^2 = k^2 = -1, ij = k, jk = i, ki = j\}$$

(-1 commutes with each element of Q_8) be the classical quaternion group of order 8. The embedding $Q_8 \rightarrow S_8$ is chosen so that $i \rightarrow (0213)(4657)$, $j \rightarrow (0415)(2736)$.

Direct calculations yield the polynomial

$$F_{A,N}(z) = \frac{1}{2^N} ((1+z)^N + 3(1-z)^2(1+z)^{N-2} + 4(1-z)^6(1+z)^{N-6})$$

for the G -orbit $A = \{7, 11, 13, 14, 112, 176, 208, 224\}$.

Examples:

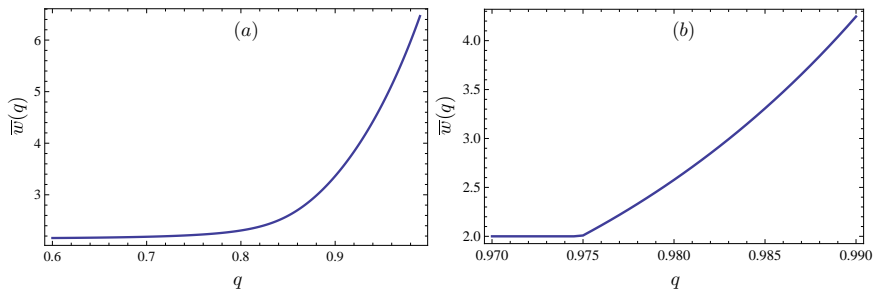


Figure : The leading eigenvalue \bar{w} depending on the fidelity q for the two-valued fitness landscape with $w = 2$, $s = 5$ and the set A as in Example 2. (a) $N = 8$ (this case was also checked numerically, using the full matrix \mathbf{QW}), (b) $N = 50$.

General construction for the Eigen evolutionary problem:

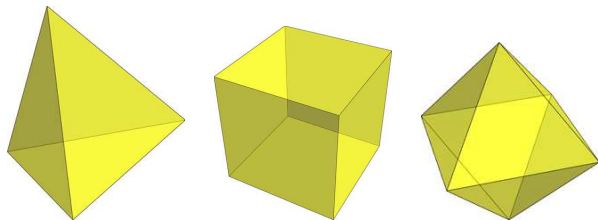


Figure : Examples of regular polytopes in dimension 3: tetrahedron (regular simplex), cube, octahedron

Question: Why to stick with the geometry of the hypercube?

General construction for the Eigen evolutionary problem:

Consider a quadruple $(X, d, \Gamma, \mathbf{w})$ where (X, d) is a finite metric space with integer distances between points of diameter N and cardinality $l = |X|$, a group $\Gamma \leq \text{Iso}(X)$ is a fixed group that acts transitively on X and a *fitness landscape* $\mathbf{w} = (w_x)^\top$ is a vector-column of non-negative real numbers called *fitnesses* indexed by $x \in X$. The quadruple $(X, d, \Gamma, \mathbf{w})$ will be called *homogeneous* Γ -landscape. In other words, the sequences of the population are encoded by $x \in X$.

Consider also the diagonal matrix $\mathbf{W} = \text{diag}(w_x)$ of order l called the *fitness matrix*, the symmetric distance matrix $\mathbf{D} = (d(x, y))_{l \times l}$ with integer entries of the same order and the symmetric matrix $\mathbf{Q} = ((1 - q)^{d(x, y)} q^{N - d(x, y)})_{l \times l}$ for $q \in [0, 1]$. Finally, we introduce the *distance polynomial*

$$P_X(q) = \sum_{x \in X} (1 - q)^{d(x, x_0)} q^{N - d(x, x_0)}, \quad x_0 \in X.$$

Since Γ acts transitively on X this polynomial is independent on the choice of $x_0 \in X$ and is the sum of entries in each row (column) of \mathbf{Q} .

General construction for the Eigen evolutionary problem:

Definition: The problem to find the leading eigenvalue $\bar{w} = \bar{w}(q)$ of the matrix $\frac{1}{P_X(q)}\mathbf{QW}$ and the eigenvector $\mathbf{p} = \mathbf{p}(q)$ satisfying

$$\mathbf{QW}\mathbf{p} = P_X(q)\bar{w}\mathbf{p}, \quad p_x = p_x(q) \geq 0, \quad \sum_{x \in X} p_x(q) = 1$$

will be called *the generalized algebraic Eigen's quasispecies problem*.

This problem turns into the classical Eigen's evolutionary problem for the N -dimensional binary cube $X = \{0, 1\}^N$ with the Hamming metric and $\Gamma = \text{Iso}(X)$. In the classical case $P_X(q) \equiv 1$

Simplicial fitness landscape:

Let $X = \{0, 1, \dots, n\}$ and $d(i, j) = 1$ if $i \neq j$, $d(i, i) = 0$. Hence, X is a metric space with the trivial metric, $N = \text{diam}(X) = 1$ and $l = |X| = n + 1$.

Let $A \subset \{0, 1, \dots, n\}$. Consider the landscape

$$w_k = \begin{cases} w + s, & k \in A, \\ w, & k \notin A. \end{cases}$$

Working along the same lines we find that the leading eigenvalue can be found as a root of an algebraic equation

$$\frac{|A|}{(n+1)(\bar{u}-u)} + \left(1 - \frac{|A|}{n+1}\right) \frac{2q-1}{(q+n(1-q))\bar{u} - (2q-1)u} = 1, \quad u = \frac{w}{s}, \quad \bar{u} = \frac{\bar{w}}{s}$$

of degree 2. (A conjecture: for the hyperoctahedral landscapes we will get a cubic equation)

Acknowledgements:

Startup grant from Department of Mathematics, NDSU



ND EPSCoR and NSF grant # EPS-0814442



Thank you for your attention!

e-mail: artem.novozhilov@ndsu.edu

site: <https://www.ndsu.edu/pubweb/novozhil/>

References:

- ▶ Semenov, Yuri S., and Artem S. Novozhilov. "On Eigen's quasispecies model, two-valued fitness landscapes, and isometry groups acting on finite metric spaces." arXiv preprint arXiv:1503.03343 (2015).