

# Linear Algebra of the Quasispecies Model

Artem Novozhilov

Department of Mathematics  
North Dakota State University

SIAM Conference on the Life Sciences,  
Charlotte, North Caroline, USA

August 6, 2014

## Collaborators:



Yuri S. Semenov  
Moscow State University of Railway  
Engineering, Moscow, Russia

Alexander S. Bratus  
Lomonosov Moscow State University  
Moscow, Russia



# Quasispecies theory:



Manfred Eigen, born 1927

- ▶ M. Eigen, *Naturwissenschaften*, 58(10), 1971:465–523
- ▶ M. Eigen, P. Schuster, *The Hypercycle*, Springer, 1979
- ▶ M. Eigen, J. McCaskill, P. Schuster, *J Phys Chem*, 92(24), 1988:6881-6891

## DIE NATURWISSENSCHAFTEN

58. Jahrgang, 1971

Heft 10 Oktober

### Selforganization of Matter and the Evolution of Biological Macromolecules

MANFRED EIGEN\*

Max-Planck-Institut für Biophysikalische Chemie,  
Karl-Friedrich-Bonhoeffer-Institut, Göttingen-Nikolausberg

<i>I. Introduction</i> . . . . .	465	<i>V. Selforganization via Cyclic Catalysis: Proteins</i> . . . . .	498
I.1. Cause and Effect . . . . .	465	V.1. Recognition and Catalysis by Enzymes . . . . .	498
I.2. Prerequisite of Selforganization . . . . .	467	V.2. Selforganizing Enzyme Cycles (Theory) . . . . .	499
I.2.1. Evolution Must Start from Random Events . . . . .	467	V.2.1. Catalytic Networks . . . . .	499
I.2.2. Instruction Requires Information . . . . .	467	V.2.2. The Selfreproducing Loop and Its Variants . . . . .	499
I.2.3. Information Originates or Gains Value by Selection . . . . .	469	V.2.3. Competition between Different Cycles: Selection . . . . .	501
I.2.4. Selection Occurs with Special Substances under Special Conditions . . . . .	470	V.3. Can Proteins Reproduce Themselves? . . . . .	501
<i>II. Phenomenological Theory of Selection</i> . . . . .	473	<i>VI. Selfordering by Encoded Catalytic Function</i> . . . . .	503
II.1. The Concept "Information" . . . . .	473	VI.1. The Requirement of Cooperation between Nucleic Acids and Proteins . . . . .	503
II.2. Phenomenological Equations . . . . .	474	VI.2. A Selfreproducing Hyper-Cycle . . . . .	503
II.3. Selection Strains . . . . .	476	VI.2.1. The Model . . . . .	503
II.4. Selection Equilibrium . . . . .	479	VI.2.2. Theoretical Treatment . . . . .	505
II.5. Quality Factor and Error Distribution . . . . .	480	VI.3. On the Origin of the Code . . . . .	505
II.6. Kinetics of Selection . . . . .	481	<i>VII. Evolution Experiments</i> . . . . .	511
<i>III. Stochastic Approach to Selection</i> . . . . .	484	VII.1. The $Q\beta$ -Replicase System . . . . .	511
III.1. Limitations of a Deterministic Theory of Selection . . . . .	484	VII.2. Darwinian Evolution in the Test Tube . . . . .	512
III.2. Fluctuations around Equilibrium States . . . . .	484	VII.3. Quantitative Selection Studies . . . . .	513
III.3. Fluctuations in the Steady State . . . . .	485	VII.4. "Minus One" Experiments . . . . .	514
III.4. Stochastic Models as Markov Chains . . . . .	487	<i>VIII. Conclusion</i> . . . . .	515
III.5. Quantitative Discussion of Three Prototypes of Selection . . . . .	487	VIII.1. Limits of Theory . . . . .	515
<i>IV. Selforganization Based on Complementary Recognition: Nucleic Acids</i> . . . . .	490	VIII.2. The Concept "Value" . . . . .	515
IV.1. True "Selfinstruction" . . . . .	490	VIII.3. "Dispersion" and the "Origin of Information" . . . . .	516
IV.2. Complementary Instruction and Selection (Theory) . . . . .	492	VIII.4. The Principles of Selection and Evolution . . . . .	517
IV.3. Complementary Base Recognition (Experimental Data) . . . . .	494	VIII.5. "Indeterminate", but "Inevitable" . . . . .	518
IV.3.1. Stable Pair Formation . . . . .	494	VIII.6. Can the Phenomenon of Life be Explained by Our Present Concepts of Physics? . . . . .	520
IV.3.2. Cooperative Interactions in Oligo- and Polynucleotides . . . . .	495	<i>IX. Deutsche Zusammenfassung</i> . . . . .	520
IV.3.3. Conclusions about Recognition . . . . .	496	Acknowledgements . . . . .	522
		Literature . . . . .	522

## Mathematical side of the story:

The variables of the dynamical system are the concentrations of individual polynucleotide sequences:  $[I_i] = c_i(t)$ . We are interested, essentially, in the relative concentrations of the different species

$$x_i(t) = c_i(t) / \sum_{i=1}^n c_i(t); \quad i = 1, 2, \dots, n \quad (12)$$

The resulting kinetic equations, around which quasi-species theory centers, are then

$$dx_i(t)/dt \equiv \dot{x}_i(t) = (W_{ii} - \bar{E}(t))x_i(t) + \sum_{k \neq i} W_{ik}x_k(t); \\ i, k = 1, 2, \dots, n \quad (13)$$

The mean excess production

$$\bar{E}(t) = \sum_{i=1}^n x_i(t)E_i \quad (14)$$

of the population may be physically compensated by a dilution

**Ref:** M. Eigen, J. McCaskill, P. Schuster, J Phys Chem, 92(24), 1988: 6881-6891

## Model statement:

Consider a population of sequences of fixed length  $N$  composed of two-letter alphabet, say,  $\{0, 1\}$ , therefore  $2^N$  different sequences.

The population is subject to two evolutionary forces. First evolutionary force is *selection*, which is included in the system through the *Malthusian fitness*, defined here for simplicity as

$$m(\text{particular sequence } \sigma) = m(H_\sigma),$$

where  $H_\sigma$  is the Hamming norm of this sequence, i.e., number of 1s in sequence  $\sigma$ . In this way we do not distinguish between sequences with the same number of 1s and hence reduce the dimensionality of the problem from  $2^N \times 2^N$  to  $(N + 1) \times (N + 1)$ . Hence, we consider at this point only *permutation invariant* fitness landscapes

$$\mathbf{M} = \text{diag}(m_0, \dots, m_N) \quad \text{or} \quad \mathbf{m} = (m_0, \dots, m_N)^\top.$$

## Model statement:

The second evolutionary force is *mutation*.

In particular, assuming  $N + 1$  classes of sequences, we have that the mutations  $\mu_{ij}$  (i.e., the mutation rate from class  $j$  to class  $i$ ) can be described by the matrix

$$\mathcal{M} = (\mu_{ij}) = \mu \mathbf{Q} = \mu \begin{bmatrix} -N & 1 & 0 & 0 & \dots & \dots & 0 \\ N & -N & 2 & 0 & \dots & \dots & 0 \\ 0 & N-1 & -N & 3 & \dots & \dots & 0 \\ 0 & 0 & N-2 & -N & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 2 & -N & N \\ 0 & 0 & \dots & \dots & 0 & 1 & -N \end{bmatrix},$$

where  $\mu$  is the mutation rate per site per sequence per replication event.

## Model statement:

Let  $\mathbf{p}(t)$  denote the vector of frequencies of different classes of sequences, then, assuming uncoupled reproduction and mutation events, we arrive at

$$\dot{\mathbf{p}}(t) = (\mathbf{M} + \mu\mathbf{Q})\mathbf{p}(t) - \bar{m}(t)\mathbf{p}(t),$$

where

$$\bar{m}(t) = \mathbf{m} \cdot \mathbf{p}(t) = \sum_{i=0}^N m_i p_i(t)$$

is the mean population fitness.

This model is often called a paramuse of Crow–Kimura quasispecies model with permutation invariant fitness landscape.

**Ref:** Baake and Gabriel, Annual Reviews of Computational Physics VII, 1999: 203–264

**Ref:** Crow and Kimura, An introduction to population genetics theory, 1970

## Elementary results:

The asymptotic behavior of the quasispecies model is determined by the equilibrium  $\mathbf{p} = \lim_{t \rightarrow \infty} \mathbf{p}(t)$ , which solves the eigenvalue problem

$$(\mathbf{M} + \mu\mathbf{Q})\mathbf{p} = \bar{m}\mathbf{p},$$

where

$$\bar{m} = \mathbf{m} \cdot \mathbf{p} = \sum_{i=0}^N m_i p_i.$$

By Perron–Frobenius theorem it follows that there is a unique positive solution  $\mathbf{p} > 0$ , which is the right eigenvector of  $\mathbf{M} + \mu\mathbf{Q}$  corresponding to the simple real dominant eigenvalue  $\lambda = \bar{m}$ .

This vector  $\mathbf{p}$  was called by Eigen the *quasispecies*. It is globally stable for the quasispecies system. We are mostly interested in properties of  $\bar{m}$  and  $\mathbf{p}$  depending on the fitness landscape  $\mathbf{M}$  and mutation rate  $\mu$ , therefore, we use the notation  $\bar{m} = \bar{m}(\mu)$  and  $\mathbf{p} = \mathbf{p}(\mu)$  for the mean fitness and equilibrium distribution.



## Known results:

- ▶ Thompson and McBride, Math Biosciences, 21: 127–142, 1974  
The quasispecies model is essentially linear.
- ▶ Rumschitzki, J of Math Biol, 24: 667–680, 1987  
For zero epistasis fitness landscape the spectral properties of the Eigen evolutionary matrices can be inferred with the representation of  $\mathcal{M}$  and  $M$  using tensor products.
- ▶ Swetina and Schuster, Bioph Chem, 16: 329–345, 1982  
Numerical analysis of single peaked fitness landscape yields the *error threshold*.

# Known results: The error threshold

Ref: Swetina and Schuster, Bioph Chem, 16: 329–345, 1982

Consider the single peaked fitness landscape

$$M = \text{diag}(m_0, 0, \dots, 0), \quad m_0 > 0.$$

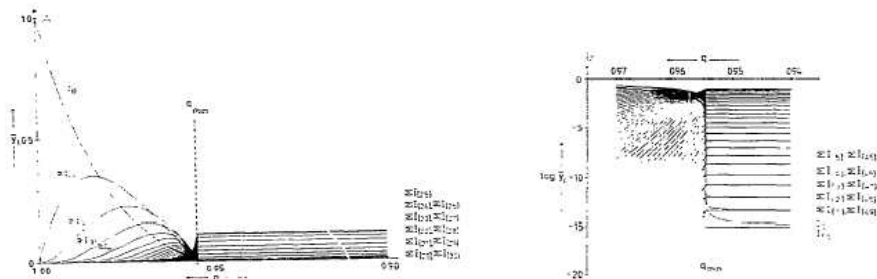
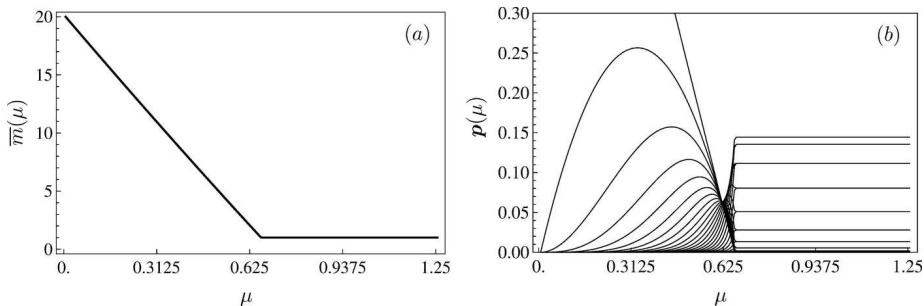


Fig. 10. Distribution of mutant classes as a function of the single-digit accuracy  $q$  for  $\nu = 50$ . Note the sharpness of the transition from direct to stochastic replication around  $q_{\min}$ . This is seen best on the logarithmic plot. In the domain of stochastic replication individual concentrations become exceedingly small:  $\xi_i = 8.9 \times 10^{-16}$ ,  $i = 0 \dots 2^{30} - 1$ . For basic definitions and numerical values see fig. 7.

## Known results: The error threshold

Consider the single peaked fitness landscape

$$\mathbf{M} = \text{diag}(m_0, m_1, \dots, m_1), \quad m_0 > m_1.$$



**Figure :** Error threshold in the quasispecies model with the single peaked fitness landscape ( $\mathbf{m} = (m_0, m_1, \dots, m_1)$ ,  $m_0 > m_1$ ). The parameters are  $N = 30$ ,  $m_0 = 20$ ,  $m_1 = 1$ . (a) The mean population fitness  $\bar{m}(\mu)$  versus the mutation rate; (b) the stationary quasispecies distribution versus the mutation rate

## Known results: Statistical Physics

- ▶ Leuthäusser, I. (1986). An exact correspondence between Eigen's evolution model and a two-dimensional Ising system. *The Journal of Chemical Physics*, 84(3), 1884-1885.
- ▶ Tarazona, P. (1992). Error thresholds for molecular quasispecies as phase transitions: From simple landscapes to spin-glass models. *Physical Review A*, 45(8), 6038.
- ▶ Baake, E., Baake, M., & Wagner, H. (1997). Ising quantum chain is equivalent to a model of biological evolution. *Physical Review Letters*, 78(3), 559-562.
- ▶ Galluccio, S. (1997). Exact solution of the quasispecies model in a sharply peaked fitness landscape. *Physical Review E*, 56(4), 4526.
- ▶ Baake, E., & Wagner, H. (2001). Mutation-selection models solved exactly with methods of statistical mechanics. *Genetical Research*, 78(01), 93-117.
- ▶ Saakian, D. B., & Hu, C. K. (2006). Exact solution of the Eigen model with general fitness functions and degradation rates. *Proceedings of the National Academy of Sciences of the USA*, 103(13), 4935-4939.

## Known results: Maximum principle

- ▶ Hermisson, J., Redner, O., Wagner, H., & Baake, E. (2002). Mutation-selection balance: Ancestry, load, and maximum principle. *Theoretical Population Biology*, 62(1), 9-46.
- ▶ Baake, E., & Georgii, H. O. (2007). Mutation, selection, and ancestry in branching models: a variational approach. *Journal of Mathematical Biology*, 54(2), 257-303.

Assume that  $m_i = Nr_i = Nr(x_i)$ ,  $x_i = \frac{i}{N} \in [0, 1]$  and define  $g(x) = \mu(1 - 2\sqrt{x(1-x)})$ . Then the mean fitness  $\bar{m}(\mu) = N\bar{r}$  is given by

$$\bar{r} \approx \bar{r}_\infty = \sup_{x \in [0,1]} (r(x) - g(x)).$$

$r(x)$  may have only finite number of discontinuities and be either left or right continuous at every point.

## Main idea:

We consider the eigenvalue problem

$$(\mathbf{M} + \mu \mathbf{Q})\mathbf{p} = \bar{m} \mathbf{p}, \quad \bar{m} = \mathbf{m} \cdot \mathbf{p},$$

where  $\mathbf{p} = \mathbf{p}(\mu)$ ,  $\bar{m} = \bar{m}(\mu)$  with a fixed fitness landscape  $\mathbf{m}$ .

We claim that this problem simplifies in the coordinates of the basis composed of the eigenvectors of the matrix  $\mathbf{Q} = \mathbf{Q}_N$ . Recall that

$$\mathbf{M} = \text{diag}(m_0, \dots, m_N), \quad \mathbf{Q}_N = \begin{bmatrix} -N & 1 & 0 & \dots & \dots & 0 \\ N & -N & 2 & \dots & \dots & 0 \\ 0 & N-1 & -N & \dots & \dots & 0 \\ 0 & 0 & N-2 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 & -N \end{bmatrix}$$

## Proposition: For the matrix $Q = Q_N$ :

1. The eigenvalues of  $Q_N$  are simple (all have algebraic multiplicities one) and given by

$$q_k = -2k, \quad k = 0, \dots, N.$$

2. Let  $\mathbf{v}_k^\top = (c_{0k}, \dots, c_{Nk})$  be the right eigenvector of  $Q_N$  corresponding to  $q_k$  and normalized such that  $c_{0k} = 1$ ,  $\mathbf{C} = \mathbf{C}_N = (c_{ik})_{(N+1) \times (N+1)}$  be the matrix composed of  $\mathbf{v}_k$  ( $\mathbf{v}_k$  is the  $k$ -th column of  $\mathbf{C}_N$ ). Then the generating function for the elements of the  $k$ -th column has the form

$$P_k(t) = \sum_{i=0}^N c_{ik} t^i = (1-t)^k (1+t)^{N-k}, \quad k = 0, \dots, N.$$

3.  $\mathbf{C}^2 = 2^N \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix, or, equivalently,

$$\mathbf{C}^{-1} = 2^{-N} \mathbf{C}.$$

4. 1-norm of  $\mathbf{C}$  is

$$\|\mathbf{C}\|_1 = \max_{0 \leq k \leq N} \sum_{i=0}^N |c_{ik}| = 2^N.$$

## The key idea of the proof:

Let

$$\mathbf{Q}_N = \begin{bmatrix} -N & 1 & 0 & \dots & \dots & 0 \\ N & -N & 2 & \dots & \dots & 0 \\ 0 & N-1 & -N & \dots & \dots & 0 \\ 0 & 0 & N-2 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 & -N \end{bmatrix}.$$

Consider the linear operator

$$\mathcal{Q}_N: P(t) \longrightarrow (1 - t^2)P'(t) - N(1 - t)P(t),$$

acting on the  $(N + 1)$ -dimensional vector space  $\mathbf{R}_N[t]$  of the polynomials of degree less or equal  $N$ . Direct calculations show that matrix  $\mathbf{Q}_N$  is the matrix of  $\mathcal{Q}_N$  in the standard basis  $\{1, t, \dots, t^N\}$  of  $\mathbf{R}_N[t]$ .



## Behavior for $\mu \rightarrow \infty$ :

Let  $\mathbf{x} = \mathbf{C}^{-1}\mathbf{p}$ , where  $\mathbf{C} = (c_{ij})$  is the matrix composed of the eigenvectors of  $\mathbf{Q}$ ,

$$c_{ik} = \sum_{j=0}^N (-1)^j \binom{k}{j} \binom{N-k}{i-j},$$

and  $\hat{\mathbf{x}} = 2^{-N}(1, 0, \dots, 0)^\top$ . Then

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{1}{\mu} \|\mathbf{M}\|_1,$$

and

$$\|\mathbf{x}'(\mu)\|_1 \leq \frac{2N}{2\mu - (2^{N+1} + 1)\|\mathbf{M}\|_1}.$$

Therefore,

$$\mathbf{p}(\mu) \rightarrow \hat{\mathbf{p}} = 2^{-N} \left( \binom{N}{0}, \dots, \binom{N}{N} \right).$$

## Parametric solution:

Consider fitness landscape such that  $m_j > 0$  for some  $j$ , and all the rest  $m_i = 0$  for  $i \neq j$ . Then a parametric solution to the eigenvalue problem

$$(M + \mu Q)\mathbf{p} = \bar{m}\mathbf{p}$$

can be found in the form

$$\begin{aligned}x_k(s) &= 2^{-N} \frac{c_{kj}}{1 + ks}, \quad k = 0, \dots, N, \\p_i(s) &= 2^{-N} \sum_{k=0}^N \frac{c_{ik}c_{kj}}{1 + ks}, \quad i = 0, \dots, N, \\\bar{m}(s) &= m_j p_j(s), \quad \mu = \frac{s}{2} \bar{m}(s).\end{aligned}$$

Here  $\mathbf{x} = \mathbf{C}^{-1}\mathbf{p}$ .

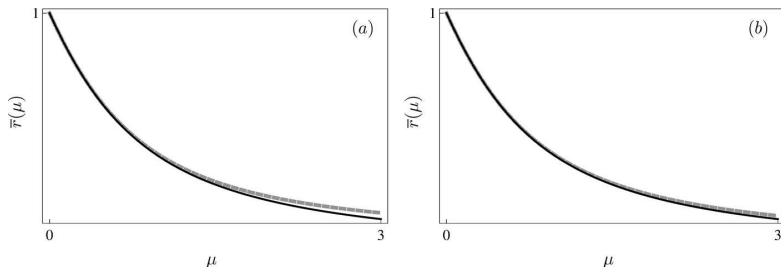
## Example 1:

Let  $N = 2A$  and

$$\mathbf{m} = (0, \dots, 0, N, 0, \dots, 0),$$

where  $N$  is exactly on the  $A$ -th position. Consider a scaled fitness landscape  $N\mathbf{r} = \mathbf{m}$  and  $\bar{r}(\mu) = \bar{m}(\mu)/N$ . Then, using the properties of the mutation matrix  $\mathbf{Q}$  and the parametric formulas from the previous slide, it can be proved that

$$\bar{r} \approx \bar{r}_\infty = \lim_{N \rightarrow \infty} \bar{r}(\mu) = \sqrt{\mu^2 + 1} - \mu.$$



**Figure :** Comparison of the limit distribution (black) with the numerical computations (gray, dashed). (a) The sequence length is  $N = 100$ , (b)  $N = 200$

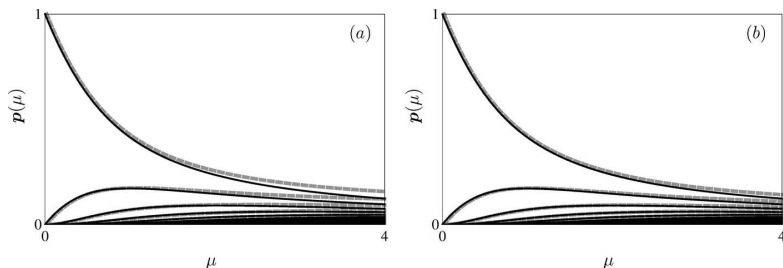
## Example 1:

Let  $N = 2A$  and

$$\mathbf{m} = (0, \dots, 0, N, 0, \dots, 0),$$

where  $N$  is exactly on the  $A$ -th position. Consider a scaled fitness landscape  $N\mathbf{r} = \mathbf{m}$  and  $\bar{r}(\mu) = \bar{m}(\mu)/N$ . Denoting  $\bar{r}_\infty = \sqrt{\mu^2 + 1} - \mu$  we can prove that

$$\lim_{N \rightarrow \infty} p_{A \pm k}(\mu) = \bar{r}_\infty \left( \frac{1 - \bar{r}_\infty}{1 + \bar{r}_\infty} \right)^k, \quad k = 0, \dots, A.$$



**Figure :** Comparison of the limit distribution (black) with the numerical computations (gray, dashed) for  $N = 100$  (a) and  $N = 200$  (b)

## Example 2: Single peaked landscape

Consider the single peaked landscape, defined as

$$\mathbf{m} = (N, 0, \dots, 0) = N\mathbf{r} = N(1, 0, \dots, 0).$$

Consider a scaled fitness landscape  $N\mathbf{r} = \mathbf{m}$  and  $\bar{r}(\mu) = \bar{m}(\mu)/N$ . Then, using the properties of the mutation matrix  $\mathbf{Q}$  and the parametric formulas it can be proved that

$$\begin{aligned}\bar{r}_\infty &= \lim_{N \rightarrow \infty} p_0 = 1 - \mu, \\ \lim_{N \rightarrow \infty} p_i &= (1 - \mu)\mu^i, \quad i \geq 1.\end{aligned}$$

if  $\mu < 1$  and  $p_j = 0$  for all  $j$  if  $\mu \geq 1$ .

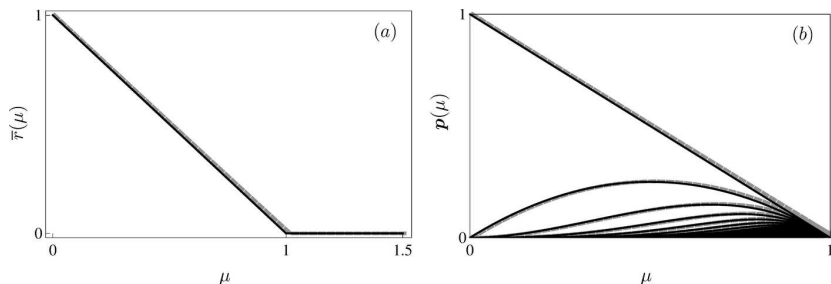
**Ref:** Saakian, D., and C.-K. Hu. Physical Review E 69.4 (2004)

**Ref:** Bratus, A., Novozhilov, A., and Y. Semenov, Mathematical Biosciences, (2014), submitted

## Example 2: Single peaked landscape

Comparison with numerical computations for  $N = 100$ .

$$\bar{r}_\infty = \lim_{N \rightarrow \infty} p_0 = 1 - \mu, \quad \lim_{N \rightarrow \infty} p_i = (1 - \mu)\mu^i, \quad i \geq 1$$



**Figure :** Comparison of numerical solutions of the quasispecies model (gray, dashed) with the single peaked fitness landscape for  $N = 100$  with the geometric limit distribution (black). (a) The expressions for the leading eigenvalue; (b) the quasispecies distribution

# Eigenvalue problem for probability generating functions (pgf):

Recall that we solve the eigenvalue problem

$$(M + \mu Q)\mathbf{p} = \bar{m}\mathbf{p},$$

and matrix  $Q$  is the representation of the linear differential operator

$$Q_N: P(t) \longrightarrow (1 - t^2)P'(t) - N(1 - t)P(t),$$

acting on the space of polynomials, in the standard basis  $\{1, t, \dots, t^N\}$ . Let  $P(t)$  be the pgf of  $\mathbf{p}$ , i.e.,  $P(t) = \sum_{k=0}^N p_k t^k$ . Then the original problem can be replaced with the eigenvalue problem

$$N\mathbf{r} \circ P(t) + Q_N P(t) = N\bar{r}P(t),$$

where  $\mathbf{r} = \mathbf{m}/N$  and  $\mathbf{r} \circ P(t) = \sum_{k=0}^N r_k p_k t^k$ .

## Eigenvalue problem for pgf:

Hence we have the problem to find a pgf  $P(t)$  satisfying

$$N\mathbf{r} \circ P(t) + \mu(1 - t^2)P'(t) - \mu N(1 - t)P(t) = N\bar{r}P(t),$$

where

$$\bar{r} = \bar{r}(\mu) = \mathbf{r} \circ P(1).$$

This problem, due to Perron–Frobenius theorem, has a unique solution  $P(t)$ . After dividing by  $N$  and assuming that

$$\frac{\mu}{N}(1 - t^2)P'(t) \rightarrow 0, \quad N \rightarrow \infty,$$

we obtain the limit problem

$$-\mu(1 - t)P(t) + \mathbf{r}_\infty \circ P(t) = \bar{r}_\infty P(t).$$



## Example: The single peaked landscape:

Start with

$$-\mu(1-t)P(t) + \mathbf{r}_\infty \circ P(t) = \bar{r}_\infty P(t).$$

Let

$$\mathbf{r}_\infty = (1, 0, \dots, 0, \dots).$$

Then

$$\mathbf{r}_\infty \circ P(t) = P(0).$$

Therefore,

$$-\mu(1-t)P(t) + P(0) = \bar{r}_\infty P(t).$$

Plugging  $t = 0$ , we find  $\bar{r}_\infty = 1 - \mu$ . Which implies that

$$P(t) = \frac{1 - \mu}{1 - \mu t},$$

is the pgf of the geometric distribution.

# 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](mailto:artem.novozhilov@ndsu.edu)

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

## References:

- ▶ Bratus, A. S., Novozhilov, A. S., & Semenov, Y. S. (2014). Linear algebra of the permutation invariant Crow–Kimura model of prebiotic evolution. Submitted to Math Biosciences.
- ▶ Semenov, Y. S., Bratus, A. S., & Novozhilov, A. S. (2014). On the behavior of the leading eigenvalue of the Eigen evolutionary matrices. arXiv:1405.6411
- ▶ Semenov, Y. S., & Novozhilov, A. S. (2014) Exact solutions for the selection-mutation equilibrium in the Crow-Kimura evolutionary model. arXiv.