Detailed analysis of an Eigen quasispecies model in a periodically moving sharp-peak landscape

Armando G. M. Neves
aneves@mat.ufmg.br

Departamento de Matemática
Universidade Federal de Minas Gerais
BRASIL
Supported by PRPG/UFMG

6o BIOMAT, La Falda, August 4 to 7, 2014

The problem

Under which conditions on the mutation rates can a virus survive if its environment (immune system persecution) changes periodically?

The Eigen quasispecies model

- Introduced by Eigen in the 1970's to study the origin of life.
- Later used to study virus replication, taking into account the possibility of replication errors.
- A virus genome is $\sigma = (s_1, s_2, \ldots, s_\ell)$ with $s_i \in \{0, 1\}$.
- ℓ is large, in the range 10^3 to 10^5 for viruses.
- Genome space is $\Lambda = \{0, 1\}^\ell$.
- Phase transitions, methods from Statistical Mechanics and Quantum Field Theory. Interest of physicists.

The Eigen model in general

- If $p_\sigma(t)$ be the virus population with genome σ in generation t, then

$$p_\sigma(t+1) = \sum_{\sigma' \in \Lambda} W_{\sigma,\sigma'} f(\sigma', t) p_{\sigma'}(t) \ . \tag{1}$$

The Eigen model in general

- If $p_\sigma(t)$ be the virus population with genome σ in generation t, then

$$p_\sigma(t+1) = \sum_{\sigma' \in \Lambda} W_{\sigma,\sigma'} f(\sigma', t) p_{\sigma'}(t) \ . \tag{1}$$

- $f(\sigma, t)$ is the *fitness* at time t of individuals with genome σ, i.e., the number of its descendants in generation $t+1$.

The Eigen model in general

- If $p_\sigma(t)$ be the virus population with genome σ in generation t, then

$$p_\sigma(t+1) = \sum_{\sigma' \in \Lambda} W_{\sigma,\sigma'} f(\sigma', t) p_{\sigma'}(t) \ . \tag{1}$$

- $f(\sigma, t)$ is the *fitness* at time t of individuals with genome σ, i.e., the number of its descendants in generation $t+1$.
- $W_{\sigma,\sigma'}$ is the probability that an individual with genome σ' has a descendant with genome σ.

The Eigen model in general

- If $p_\sigma(t)$ be the virus population with genome σ in generation t, then

$$p_\sigma(t+1) = \sum_{\sigma' \in \Lambda} W_{\sigma,\sigma'} f(\sigma', t) p_{\sigma'}(t) \ . \tag{1}$$

- $f(\sigma, t)$ is the *fitness* at time t of individuals with genome σ, i.e., the number of its descendants in generation $t+1$.
- $W_{\sigma,\sigma'}$ is the probability that an individual with genome σ' has a descendant with genome σ.
- We shall use Hamming distance $d(\sigma, \sigma')$ to measure distance between genomes:

The Eigen model in general

- If $p_\sigma(t)$ be the virus population with genome σ in generation t, then

$$p_\sigma(t+1) = \sum_{\sigma' \in \Lambda} W_{\sigma,\sigma'} f(\sigma', t) p_{\sigma'}(t) \ . \tag{1}$$

- $f(\sigma, t)$ is the *fitness* at time t of individuals with genome σ, i.e., the number of its descendants in generation $t+1$.
- $W_{\sigma,\sigma'}$ is the probability that an individual with genome σ' has a descendant with genome σ.
- We shall use Hamming distance $d(\sigma, \sigma')$ to measure distance between genomes:

$$d(\sigma, \sigma') \ = \ \sum_{i=1}^{\ell} |s_i - s'_i| \ .$$

Mutation matrix

- Let μ be the *per site* mutation probability.
- Naturally, $W_{\sigma\sigma'} = \mu^d (1-\mu)^{\ell-d}$, where d is the Hamming distance between σ and σ'.
- As μ is very small, of order 10^{-7} or less, a useful simplification is taking

$$W_{\sigma\sigma'} = \begin{cases} 1-\beta, & \text{if} \quad d(\sigma,\sigma') = 0 \\ \mu, & \text{if} \quad d(\sigma,\sigma') = 1 \\ 0, & \text{if} \quad d(\sigma,\sigma') > 1 \end{cases} , \tag{2}$$

where $\beta \equiv \mu\ell$ is the genome mutation probability.

The sharp-peak fitness landscape

- A simple and popular choice for the fitness is the *sharp-peak landscape* (SPL):

$$f(\sigma, t) = \begin{cases} 1 + k, & \text{if} \quad \sigma = \sigma_0(t) \\ 1, & \text{if} \quad \sigma \neq \sigma_0(t) \end{cases} . \tag{3}$$

The sharp-peak fitness landscape

- A simple and popular choice for the fitness is the *sharp-peak landscape* (SPL):

$$f(\sigma, t) = \begin{cases} 1 + k, & \text{if} \quad \sigma = \sigma_0(t) \\ 1, & \text{if} \quad \sigma \neq \sigma_0(t) \end{cases} . \tag{3}$$

- The fittest genome $\sigma_0(t)$ at time t is called the *wild type* or *master sequence*.
- Parameter $k > 0$ is called the selective advantage of the master sequence above all other genomes.

The error catastrophe

- In the static SPL, if β is too large, or k too small, the virus population will not be concentrated within genomes close to the master sequence, being spread throughout genome space.
- In the static SPL, this *error catastrophe* will occur if $\beta > \beta_u^{\text{static}}$, where

$$\beta_u^{\text{static}} = \frac{k}{1+k} \,. \tag{4}$$

- The error catastrophe is a transition between a *localized phase* in Λ, the quasispecies, and a delocalized phase in Λ, in which the virus population is not able to maintain genetic identity.

The Nilsson-Snoad model

- Nilsson and Snoad proposed in Phys. Rev. Lett. **84** (2000) a time-dependent version of the SPL in which at every τ generations the master sequece hops to a random nearest neighbor in Λ.

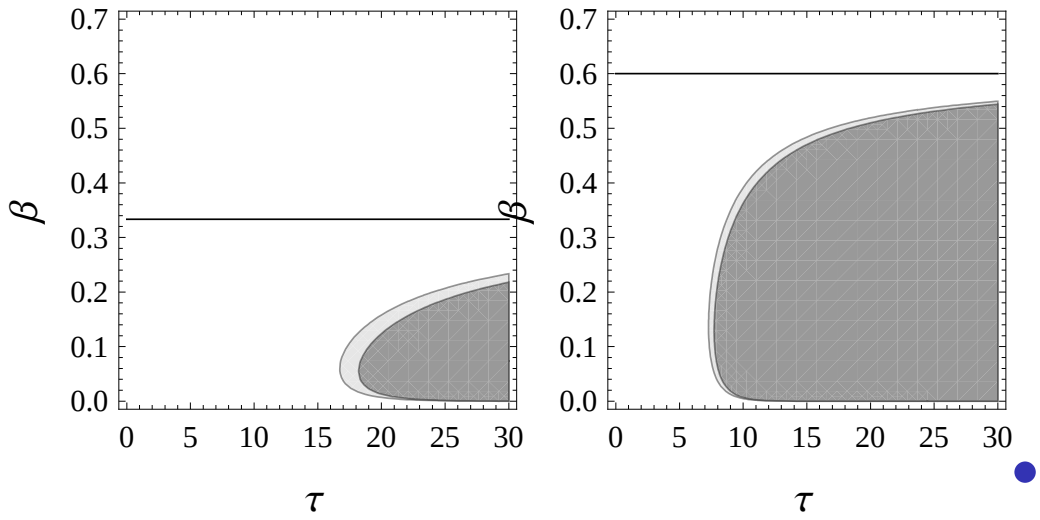
The Nilsson-Snoad model

- Nilsson and Snoad proposed in Phys. Rev. Lett. **84** (2000) a time-dependent version of the SPL in which at every τ generations the master sequece hops to a random nearest neighbor in Λ.
- The idea is to model a viral population forced to periodically change its master sequence due to persecution by an immune system.

The Nilsson-Snoad model

- Nilsson and Snoad proposed in Phys. Rev. Lett. **84** (2000) a time-dependent version of the SPL in which at every τ generations the master sequece hops to a random nearest neighbor in Λ.
- The idea is to model a viral population forced to periodically change its master sequence due to persecution by an immune system.
- Nilsson and Snoad treated the model using several questionable approximations. They found out not only the well-known error catastrophe characterized by an upper threshold β_u^{NS}, but also an adaptability catastrophe characterized by a lower threshold β_l^{NS}.
- A quasispecies will exist if $\beta_l^{\mathrm{NS}} < \beta < \beta_u^{\mathrm{NS}}$.

Results



$k = 0.5$ $k = 1.5$

- In Phys. Rev. E **82**(3):031915 (2010), we have shown that the conclusions by Nilsson and Snoad about the existence of upper and lower thresholds were correct.
- But their approximation scheme was not so much accurate, particularly for small values of the selective advantage k.

Some ideas about our techinques

- Nilsson and Snoad divide the virus population into 3 classes: viruses in the present master sequence, viruses in the next master sequence and all others.
- Existence of a quasispecies turns out to be the calcuation of the dominant eigenvalue of a 3×3 matrix.
- We divide instead the population into $M + 1$ classes: each of the M genomes which are going to be master sequences at some time plus one class for all other genomes.
- M should be of order 2^{ℓ}, but smaller values produce almost the same results.
- We seek the dominant eigenvalue of the non-negative matrix $A = S^{-1} E_1^{\tau}$, where E_1 gives the evolution for one generation while the master sequence remains unchanged and S represents the shift of the master sequence after τ generations.

Some ideas about our techinques 2

- By the Perron-Frobenius theory for non-negative matrices, the dominant eigenvalue λ_{PF} is given by the maximum over non-negative vectors of the Collatz-Wielandt function

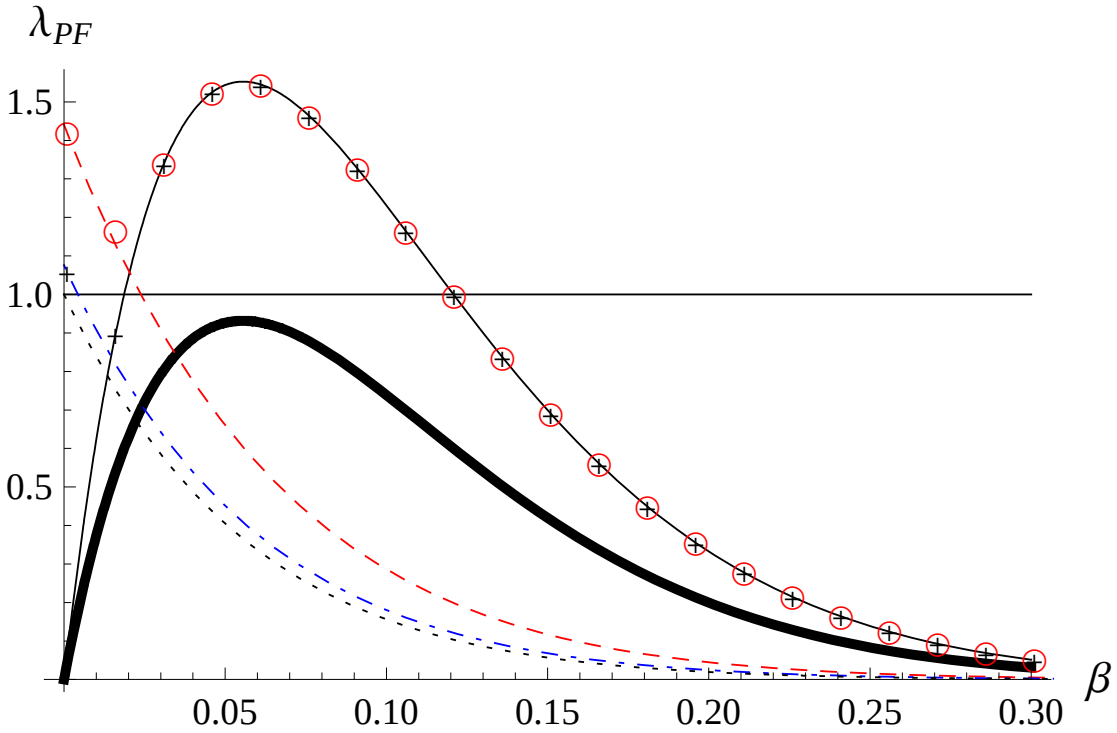
$$f_A(v) = \min_{v_i \neq 0} \frac{(Av)_i}{v_i} .$$

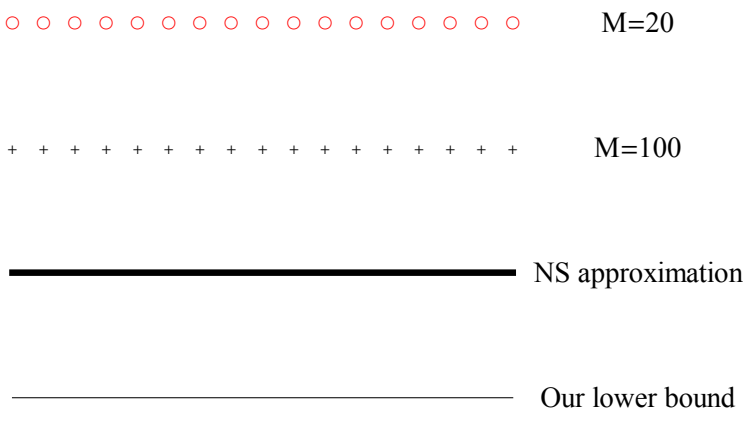
- The vector v which maximizes the above function is an eigenvector corresponding to λ_{PF}.
- For any vector v, $f_A(v)$ is a lower bound to λ_{PF}. If v is a good approximant to the dominant eigenvector, $f_A(v)$ will be a large lower bound approximating λ_{PF}.
- If e_k is the k-th vector in the canonical basis for $\mathbb{R}^M$, a good guess for the dominant eigenvector is $v(\delta) = \delta e_1 + (1 - \delta) e_M$.
- It is straightforward to find the value of $\delta_{\max} \in [0, 1]$ maximizing $f_A(v(\delta))$.

Some ideas about our techinques 3

Surprisingly, $f_A(v(\delta_{\max}))$ is not only a lower bound, but a very good approximation for λ_{PF}.

$$\lambda_{PF} \approx \frac{(1+k)^{\tau}(2+k)}{k\,\ell}\,\beta\,(1-\beta)^{\tau-1}\ .$$





$k = 0.5,\ \tau = 18,\ \ell = 100$