

## **On the Genetic Interpretation of Eigenvectors for the Generalized Model of Nonepistatic Selection**

**Vladimir P. Passekov**

Vienna, Preprint ESI 1329 (2003)

May 28, 2003

Supported by the Austrian Federal Ministry of Education, Science and Culture  
Available via <http://www.esi.ac.at>

# On the genetic interpretation of eigenvectors for the generalized model of nonepistatic selection

Vladimir P. Passekov\*

Division of Modeling Problems, Computing Center of Russian Academy of Science  
40 Vavilov Street, Moscow 119991, RUSSIA. E-mail: PASSEKOV@CCAS.RU

## Abstract

A generalized nonepistatic selection model for diploid population with nonoverlapping generations suggested by Karlin and Liberman is considered. In the model, viabilities of genotypes are determined by combinations of neutral/multiplicative interactions between loci. The structure of known expressions for eigenvectors of the linearization matrix at the polymorphic linkage equilibrium fixed point is analyzed.

1. Calculations of marginal characteristics show that the eigenvectors can be interpreted as such deviations from the equilibrium that can be observed only starting from a certain number of loci. That is, marginal one-, two-locus gamete frequencies and so on up to, say,  $k$  loci (the threshold) satisfy multilocus linkage equilibrium relations.

2. In the neutral case, the more loci determine this threshold, the more rapid is the decay of the appropriate disequilibrium. Eigenvalues of linearization matrix are simplified to known expressions under neutrality. Each eigenvalue is associated with the probability that a recombination takes place in the appropriate set of loci. In the continuous-time model of nonepistatic selection, the condition "more

---

\*The author is grateful to the Erwin Schrodinger International Institute for Mathematical Physics in Vienna for support during a stay in winter 2002/2003, where part of this work was done. This study was supported in part by the Russian Foundation for Basic Research, projects no. 03-01-00678 and no. 01-01-00965

recombination” (in the sense of increasing these probabilities for sets of loci by some quantities) enhances the stability of the polymorphic equilibrium (decreasing appropriate eigenvalues by the same quantities).

AMS classification: 92D, 92D10, 92D15

Keywords: population genetics, multilocus models, viability selection, nonepistatic selection, central equilibria, linearization, eigenvectors, disequilibria, rates of decay

## 1 Introduction

It is well known that (nonlinear) neutral models of multilocus population evolution under the influence of recombination process have some properties that resemble the properties of linear systems. One of such an attractive property is the existence of so called Bennet principal components [1, 4, 5]. These components reflect deviations from random gene combinations in gametes, successively consider more and more loci, and decay at constant rates more and more rapidly. In terms of principal components, the neutral model is linearized, and component dynamics resemble the behavior of eigenvectors for linear systems.

Unfortunately, introducing selection pressure into neutral models destroys these attractive features. In Section 2, the model of generalized nonepistatic selection (i. e., viability selection that admits summing multiplicative and neutral combinations of locus interactions) is described. In Section 3, it is shown that in a linear approximation near the polymorphic fixed point with linkage equilibrium structure the model has such a behavior that is qualitatively similar to that of the neutral case, but now disequilibria are represented by the eigenvectors of the linearization matrix. In this sense, the model of nonepistatic selection preserves linear type properties of the neutral model. The above disequilibria can be grouped according to the number of loci (the threshold) at which these disequilibria are detected.

There are a few examples of mathematical models with clear biological interpretations of such mathematical characteristics as the eigenvectors and eigenvalues for the linearization matrix determining the model behavior near

the appropriate equilibrium point. For example, in the model of selfreproducing system  $dN_i/dt = f_i N_i$ ,  $i = 1, 2, \dots$ , the growth rate  $f_i$  of some invader (introduced into a resident population) is an eigenvalue of the appropriate linearization matrix. This eigenvalue determines the fate of the invader and therefore it has the clear biological sense. In the present paper, another example of this kind is provided, and a clear genetical interpretation for eigenvectors (and, partially, for eigenvalues) in the model of nonepistatic selection is suggested.

Further, population genetic model with nonoverlapping generations is considered. In a population, each (*diploid*) individual possesses two sets of *autosomal* (with no relation to sex determination) hereditary information received from paternal and maternal sex cells (*gametes*). Each of sets consists of different discrete units, *genes*, and these two sets form an individual *genotype* coding organism traits (in particular, *viability*, i. e., the probability to survive up to reproduction stage). Differences in surviving at this stage represent an important part of selection process (viability selection).

At each *locus* (the location of a gene) there can be one of alternative types (*alleles*) of appropriate gene. Any gamete carries one allele of each locus, and there are two (may be of the same type) alleles of every locus in a diploid genotype. At reproduction stage, individuals produce gametes with *one* set of loci. Due to *recombination* (association in a new way of paternal and maternal genes in produced gametes) the gametes of an individual usually have different types. Under random mating, these gametes form pairs and fuse at fertilization independently, starting offspring development (new generation). The adults are replaced by their newborn offspring (as annual plants), and cycles

newborn offspring  $\rightarrow$  viability selection  $\rightarrow$  reproduction and recombination

repeat. Due to selection, the genetic composition of population (usually described in terms of gamete frequencies) changes, and this is considered as an elementary evolutionary step.

## 2 The generalized model of nonepistatic se-

## lection

In series of papers by Karlin and Liberman [7, 8, 9], [6], a generalized multi-locus multiallele model of nonepistatic selection is introduced and analyzed (see also [2, 3], [11]). A random mating infinite diploid population with non-overlapping generations is considered in discrete time with respect to a set  $L$  of  $l$  autosomal loci that determine viability. The model allows different configurations of multiplicative/neutral interactions between  $l$  loci, and resulting viabilities are combined additively.

For the description of multilocus systems, Kronecker product technique is convenient. Let  $\mathbf{A}_k \equiv (A_{1_k}, A_{2_k}, \dots, A_{n_k})^T$  be the vector of allele symbols for  $k$ th locus. Then a gamete composition (type) can be written as a formal product  $(A_{i_1} A_{i_2} \dots A_{i_l})$ ,  $i_k \in \overline{1, n_k}$ ,  $k = \overline{1, l}$ . The ordered set of such gamete types can be organized as the vector  $\gamma$ :

$$\gamma \equiv \mathbf{A}_1 \otimes \mathbf{A}_2 \dots \otimes \mathbf{A}_l.$$

The same coordinate order is assumed for the vector of gamete type frequencies  $\mathbf{p}$ . Multi-indices  $\mathbf{i} = (i_1, i_2, \dots, i_l)$  of  $\gamma$  ( $\mathbf{p}$ ) coordinates correspond to the allele composition of appropriate gametes  $(A_{i_1} A_{i_2} \dots A_{i_l})$ , where the coordinate  $i_k$  of a multi-index  $\mathbf{i}$  refers to the allele number of  $k$ th locus in the  $i$ th gamete considered. Population state is represented by the vector of multilocus gamete frequencies  $\mathbf{p} \equiv \{p_{\mathbf{i}} \equiv p_{i_1 i_2 \dots i_l}\}$ .

In general, viability selection pressure is determined by a matrix  $\mathbf{V} \equiv \|\mathbf{v}_{\mathbf{ij}}\|$ , where  $v_{\mathbf{ij}}$  represents the viability of the genotype composed of  $\mathbf{i}$ th and  $\mathbf{j}$ th gametes. In coordinatewise notation, population state  $\mathbf{p}$  obeys the following dynamic equation:

$$p'_{\mathbf{i}} = \frac{1}{v(\mathbf{p})} \sum_{U|V} r(U|V) \sum_{\mathbf{j}:\mathbf{j}_U=\mathbf{i}_U} \sum_{\mathbf{k}:\mathbf{k}_V=\mathbf{i}_V} v_{\mathbf{jk}} p_{\mathbf{j}} p_{\mathbf{k}}. \quad (1)$$

Here  $v(\mathbf{p}) \equiv \sum_{\mathbf{i}, \mathbf{j}} v_{\mathbf{ij}} p_{\mathbf{i}} p_{\mathbf{j}}$  is the *mean (population) viability*. We use Lyubich [10] notation for the description of recombination:  $U|V$  denotes the partition of the set  $L$  of the loci in two subsets  $U$  and  $V$  such that recombination does not occur for the loci in the same subset and it takes place otherwise. Coefficient  $r(U|V)$  stands for the probability of such an event in the meiosis, i. e., the probability distribution  $\{r(U|V)\}$  on the partition set is given, which

is called *linkage distribution*. Provided  $U|V$ -recombination, produced gamete consists of alleles for  $U$  loci of one parental gamete and  $V$  loci of the other.

Let there be no sex-structure and the total viability matrix  $\mathbf{V}$  be determined by one-locus viability matrices  $\{\mathbf{V}_k\}$  as described below. For each configuration  $\eta \equiv (\eta_1, \eta_2, \dots, \eta_l)$ ,  $\eta_k = 0$  or  $1$ , a multiplicative/neutral viability component  $\mathbf{V}(\eta)$  is defined as  $\bigotimes_{k=1}^l \mathbf{V}_k^{(\eta_k)}$  where  $\bigotimes$  is the symbol of Kronecker multiplication. Here  $\mathbf{V}_k^{(0)} \equiv \mathbf{e}_k \mathbf{e}_k^T$  (at  $k$ th locus neutrality with no difference between individuals in surviving),  $\mathbf{V}_k^{(1)} = \mathbf{V}_k$ ,  $\mathbf{e}_k$  is the vector of size  $n_k$  with unity coordinates, and  $n_k$  stands for the number of alleles for the  $k$ th locus. Given the coefficients  $\{c(\eta)\}$  that are arbitrary provided the total viability matrix

$$\begin{aligned} \mathbf{V} &\equiv \sum_{\eta} c(\eta) \mathbf{V}(\eta) = \sum_{\eta} c(\eta) \mathbf{V}_1^{(\eta_1)} \otimes \dots \otimes \mathbf{V}_l^{(\eta_l)} \\ &= \sum_{\eta} c(\eta) \bigotimes_{k=1}^l \mathbf{V}_k^{(\eta_k)}, \quad \mathbf{V}_k^{(0)} \equiv \mathbf{e}_k \mathbf{e}_k^T, \quad \mathbf{V}_k^{(1)} = \mathbf{V}_k \end{aligned} \quad (2)$$

has nonnegative entries, we have *the generalized nonepistatic scheme of viability selection*.

Karlin and Liberman studied properties of *central equilibria* (multilocus polymorphic equilibria with random combinations of different locus alleles in gametes) admitting representation

$$\mathbf{p}^* = \mathbf{p}^{1*} \otimes \mathbf{p}^{2*} \otimes \dots \otimes \mathbf{p}^{l*} \equiv \bigotimes_{k=1}^l \mathbf{p}^{k*}. \quad (3)$$

Here each of  $\mathbf{p}^{k*}$  satisfies the condition  $\mathbf{D}(\mathbf{p}^{k*}) \mathbf{V}_k \mathbf{p}^{k*} = v_k^* \mathbf{p}^{k*}$ , i. e.,  $\mathbf{p}^{k*}$  is the polymorphic equilibrium for one-locus selection model with the viability matrix  $\mathbf{V}_k$ ,  $v_k^*$  is the mean viability at  $\mathbf{p}^{k*}$ , and  $\mathbf{D}(\cdot)$  is the diagonal matrix with the coordinates of appropriate vector on the main diagonal.

The linearization matrix  $\mathbf{B}$  for (1) with account of (2) at the equilibrium  $\mathbf{p}^*$  with polymorphic one-locus components  $\mathbf{p}^{k*}$  can be written in the form

$$\begin{aligned} \mathbf{B} &= \frac{2}{v^*} \sum_{\eta} c(\eta) v^*(\eta) \\ &\times \sum_{U|V} r(U|V) \bigotimes_{k=1}^l (\|\mathbf{D}(\mathbf{p}^{k*}) \mathbf{V}_k / v_k^*\|^{1-\eta_k} \|\mathbf{p}^k \mathbf{e}_k^T\|^{\eta_k})^{\delta_{kU}}, \end{aligned}$$

where  $v^*$  is the mean viability at  $\mathbf{p}^*$ ,  $v^*(\eta) \equiv \prod_k (v_k^*)^{\eta_k}$ , and  $\delta_{kU} \equiv \begin{cases} 1, & k \in U \\ 0, & k \notin U \end{cases}$  is an indicator function.

Eigenvectors  $\{\mathbf{u}_i\}$  of matrix  $\mathbf{B}$  are known to be [7]–[9]

$$\mathbf{u}_i \equiv \mathbf{u}_{i_1 i_2 \dots i_l} = \bigotimes_{k=1}^l \mathbf{u}_{i_k}, \quad (4)$$

$$\mathbf{u}_{i_k} : \begin{cases} \mathbf{u}_{i_k} = \mathbf{p}^{k*} & \text{for } i_k = 1 \\ (\mathbf{e}, \mathbf{u}_{i_k}) = 0 & \text{for } i_k = \overline{2, n_k}, \end{cases} \quad \mathbf{e}_k \equiv (1, \dots, 1), \quad (5)$$

where, as before,  $n_k$  is the number of alleles at the  $k$ th locus, and vector  $\mathbf{e}_k$  has the size  $n_k$ .

The aim of the present paper is to give an interpretation to these eigenvectors in terms of disequilibria (deviations from random combinations of different alleles on levels two, three and so on loci).

### 3 The genetic interpretation of eigenvectors

#### 3.1 Eigenvectors of linearization matrix as specific deviations from linkage equilibrium

Let subscript  $U$  below refer to the subset of indices  $i_1, i_2, \dots, i_l$  corresponding to the set  $U$  of the loci considered. The following is true.

**Result I.** *Let eigenvector (4) have exactly  $k \geq 1$  non-unity subscripts (corresponding to a subset  $K$  of the loci considered).*

*Then  $\mathbf{u}_i$  can be interpreted as such a genetically admissible deviation  $\Delta$  (that is,  $\mathbf{p} = \mathbf{p}^* + \Delta$  satisfies:  $\sum_i p_i = 1$ ,  $p_i \geq 0$ ) from the equilibrium state  $\mathbf{p}^*$  that keeps the linkage equilibrium structure of marginal gamete frequencies up to the level of  $k$  loci (except for  $\mathbf{p}_K$ ). Deviations from random combinations on higher levels are determined by the existence of  $K$ -locus deviation. If the enumeration of loci is such that*

$$K = \{1, 2, \dots, k\}, \quad i_j \neq 1 \text{ if } j \in K,$$

then perturbed marginal frequencies of  $U$ -locus gametes have the following form:

$$\mathbf{p}_U = \begin{cases} \mathbf{p}_U^*, & U \subset K \text{ or } U \cap K \neq K \\ \mathbf{p}_K^* + \bigotimes_{j=1}^k \mathbf{u}_{i_j} \equiv \mathbf{p}_K^* + \Delta_K, & U = K \\ (\mathbf{p}_K^* + \Delta_K) \otimes (\bigotimes_{j \in U \setminus K} \mathbf{p}^{j*}), & U \supset K \end{cases}, \quad (6)$$

where  $\mathbf{p}_U$  ( $\Delta_U$ ) is the  $U$ -locus gamete frequency (deviation).

**Proof.** Let us show that eigenvectors  $\{\mathbf{u}_i\}$  can be interpreted as genetically admissible deviations  $\{\Delta_i \equiv \mathbf{u}_i\}$  from  $\mathbf{p}^*$  if their multideces have at least one non-unity coordinate. Since we consider a linear dynamics,  $\{\Delta_i\}$  are determined up to multiplication factor, i. e., vector-deviation is admissible if the sum of its coordinates equals zero. By successive applications of the property of Kronecker products (where  $\mathbf{A}_k \mathbf{B}_k$  are defined)

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) = (\mathbf{A}_1 \mathbf{B}_1) \otimes (\mathbf{A}_2 \mathbf{B}_2),$$

we have

$$(\mathbf{e}, \Delta_i) = \left( \bigotimes_{k=1}^l \mathbf{e}_k \right)^T \left( \bigotimes_{k=1}^l \mathbf{u}_{i_k} \right) = \bigotimes_{k=1}^l (\mathbf{e}_k, \mathbf{u}_{i_k}) = 0.$$

Here the last equality holds due to (5):  $(\mathbf{e}_k, \mathbf{u}_{i_k}) = 0$  for  $i_k \neq 1$ .

Note that the vector of marginal  $U$ -locus deviations  $\Delta_U$  linearly depends on  $\Delta$  (with the same matrix of transformation  $\mathbf{B}_U$  as that of the dependence for marginal  $U$ -locus gamete frequencies  $\mathbf{p}_U$  on population state  $\mathbf{p}$ ):

$$\Delta_U = \mathbf{B}_U \Delta \equiv \left[ \bigotimes_{k=1}^l (\mathbf{e}_k^T)^{1-\delta_{kU}} \right] \Delta, \quad (\mathbf{e}_k^T)^0 \equiv \mathbf{I}_k, \quad (\mathbf{e}_k^T)^1 \equiv \mathbf{e}_k^T.$$

Here  $\mathbf{I}_k$  is the identity matrix of appropriate size. The correctness of this transformation is especially easy to see if we relate  $\mathbf{p}_U$  with  $\mathbf{p} = \bigotimes_{k=1}^l \mathbf{p}^k$ .

Analogously,

$$\mathbf{e}_U \equiv \bigotimes_{k \in U} \mathbf{e}_k = \bigotimes_{k=1}^l (\mathbf{p}_k^T)^{1-\delta_{kU}} \mathbf{e}_k, \quad (\mathbf{p}_k^T)^0 \equiv \mathbf{I}_k, \quad (\mathbf{p}_k^T)^1 \equiv \mathbf{p}_k^T.$$



Obviously, since  $(\mathbf{p}, \mathbf{e}) \equiv (\mathbf{p}, \bigotimes_{k=1}^l \mathbf{e}_k) = 1$  and  $(\Delta, \mathbf{e}) = 0$  for any admissible  $\mathbf{p}$ ,  $\Delta$ , respectively, appropriate marginal quantities  $\mathbf{p}_U$  and  $\Delta_U$  satisfy similar relations:  $(\mathbf{p}_U, \mathbf{e}_U) = 1$  and  $(\Delta_U, \mathbf{e}_U) = 0$ . For example,

$$\begin{aligned} (\mathbf{p}_U, \mathbf{e}_U) &= \mathbf{p}^T \mathbf{B}_U^T \mathbf{e}_U \equiv \mathbf{p}^T \left[ \bigotimes_{k=1}^l (\mathbf{e}_k^T)^{1-\delta_{kU}} \right] \left[ \bigotimes_{k=1}^l (\mathbf{p}_k^T)^{1-\delta_{kU}} \mathbf{e}_k \right] \\ &= \mathbf{p}^T \bigotimes_{k=1}^l \left[ (\mathbf{e}_k^T)^{1-\delta_{kU}} (\mathbf{p}_k^T)^{1-\delta_{kU}} \mathbf{e}_k \right] = \mathbf{p}^T \bigotimes_{k=1}^l \mathbf{e}_k = 1. \end{aligned}$$

We use this type calculations to analyze eigenvectors  $\{\mathbf{u}_i\}$ , considering them as deviations  $\Delta$  from  $\mathbf{p}^*$ , and induced marginal deviations  $\Delta_U$ . Thus,

$$\begin{aligned} \Delta_U &= \mathbf{B}_U \mathbf{u}_i \equiv \left[ \bigotimes_{k=1}^l (\mathbf{e}_k^T)^{1-\delta_{kU}} \right] \left( \bigotimes_{k=1}^l \mathbf{u}_{ik} \right) = \bigotimes_{k=1}^l (\mathbf{e}_k^T)^{1-\delta_{kU}} \mathbf{u}_{ik} \\ &= \left[ \bigotimes_{k \in K} (\mathbf{e}_k^T)^{1-\delta_{kU}} \mathbf{u}_{ik} \right] \left[ \bigotimes_{k \notin K} (\mathbf{e}_k^T)^{1-\delta_{kU}} \mathbf{p}^{k*} \right]. \end{aligned}$$

Here the second factor in the last expression for  $\Delta_U$  never equals zero. The first one turns out to be zero if there exists at least one  $j$  such that  $j \in K$  and  $j \notin U$ . In this case, the appropriate cofactor  $(\mathbf{e}_k^T)^{1-\delta_{kU}} \mathbf{u}_{ik} = (\mathbf{e}_k, \mathbf{u}_{ik}) = 0$  by (5). Therefore,

$$\Delta_U = \mathbf{0} \quad \text{if } U \subset K \text{ or } U \cap K \neq K \text{ for disjoint or intersecting sets.}$$

Hence if the multi-index of vector  $\mathbf{u}_i$  has exactly one (say, the first in the order) non-unity coordinate ( $i_1 \neq 1$ ,  $K = \{1\}$ ), then

$$\begin{aligned} \Delta_{\{1\}} &= \mathbf{B}_1 \mathbf{u}_i \equiv \left( \mathbf{I}_1 \otimes (\bigotimes_{k>1} \mathbf{e}_k^T) \right) \left( \mathbf{u}_{i1} \otimes (\bigotimes_{k>1} \mathbf{p}^{k*}) \right) \\ &= \mathbf{I}_1 \mathbf{u}_{i1} \prod_{k>1} (\mathbf{e}_k, \mathbf{p}^{k*}) = \mathbf{u}_{i1}, \end{aligned}$$

$$\Delta_{\{k\}} = \mathbf{B}_k \mathbf{u}_i = (\mathbf{e}_1, \mathbf{u}_{i1}) \prod_{j \neq 1, k} (\mathbf{e}_j, \mathbf{p}^{j*}) \mathbf{p}^{k*} = \mathbf{0}, \quad k \neq 1$$

since  $\mathbf{u}_{i1}$  and  $\mathbf{e}_1$  are orthogonal if  $i_1 \neq 1$  by (5). Thus, eigenvector  $\mathbf{u}_i$  with exactly one non-unity coordinate of multi-index has the sense of such a genetically admissible deviation that induces zero increments for all locus allele frequencies, excluding the locus corresponding to the non-unity coordinate. Deviations of allele frequencies for this locus are given by the appropriate factor,  $\mathbf{u}_{i1}$ , in the Kronecker product representation of the eigenvector considered.

If the multi-index of eigenvector  $\mathbf{u}_i$  has exactly two non-unity coordinates, then induced increments of all one- and two-locus gametes equal zero (excluding the case when both locus numbers correspond to non-unity coordinates of gamete multi-index). For definiteness, consider the case when the

first two multi-index coordinates of  $\mathbf{u}_i$  are different from unity ( $K = \{1, 2\}$ ). Then

$$\begin{aligned}\Delta_{\{1\}} &= \mathbf{B}_1 \mathbf{u}_i \equiv \left( \mathbf{I}_1 \otimes \mathbf{e}_2^T \otimes \left( \bigotimes_{k>2}^l \mathbf{e}_k^T \right) \right) \left( (\mathbf{u}_{i_1} \otimes \mathbf{u}_{i_2} \otimes \left( \bigotimes_{k>2}^l \mathbf{p}^{k*} \right)) \right) \\ &= \mathbf{u}_{i_1}(\mathbf{e}_2, \mathbf{u}_{i_2}) = \mathbf{0}, \\ \Delta_{\{j,k\}} &= \mathbf{B}_{\{j,k\}} \mathbf{u}_i = \delta_{\{1,2\}\{j,k\}} \mathbf{u}_{i_j} \otimes \mathbf{u}_{i_k}\end{aligned}$$

where (5) is taken into account as well as the equality to zero of the indicator function  $\delta$  if sets  $\{1, 2\}$  and  $\{i, j\}$  are different.

In general case, the multi-index of eigenvector  $\mathbf{u}_i$  has exactly  $k \leq l$  (for definiteness, the first in the order) non-unity coordinates. and  $K = \{1, 2, \dots, k\}$  is the set of loci corresponding to these coordinates. Then

$$\Delta_U = \begin{cases} \mathbf{0}, & U \subset K \text{ or } U \cap K \neq K \\ \bigotimes_{j=1}^k \mathbf{u}_{i_j} = \Delta_K, & U = K \\ \Delta_K \bigotimes_{j \in U \setminus K} \mathbf{p}^{j*}, & U \supset K \end{cases} \quad (7)$$

In this case,  $\mathbf{u}_i$  can be interpreted as some genetically admissible deviation that induces zero increments of  $U$ -locus gamete frequencies and keeps the random combination of alleles on the level of  $U$  loci if, for example,  $U \subset K$  (but the enlargement of  $U$  above  $K$  violates linkage equilibrium structure). The picture is especially clear if we consider the frequencies of  $U$ -locus gametes  $\mathbf{p}_U$  induced by this deviation  $\Delta_i \equiv \mathbf{u}_i$ . The frequencies, obviously, satisfy formula (6) and this completes the proof.

**Remark.** Note that deviations  $\{\mathbf{u}_i\}$  from  $\mathbf{p}^*$  can be of two kinds: deviations that perturb only allele frequencies of one locus and deviations keeping all allele frequencies, but violating random combinations of different locus alleles in gametes.

Indeed, the case  $|K| = 1$  ( $|K|$  stands for the number of elements in the set  $K$ ) is the only case, where allele frequencies are perturbed. By (7),  $\Delta_U$  can have non-zero value only if  $U \supseteq K$ , whence a *necessary condition* for  $\Delta_U \neq \mathbf{0}$  is  $|U| \geq |K|$ . If  $U = \{j\}$  (vector  $\Delta_U$  gives perturbations of allele frequencies for  $j$ th locus), then  $\Delta_U = \mathbf{0}$  for  $|K| > 1$  since  $|U| = 1 < |K|$ . Therefore, only eigenvectors  $\{\mathbf{u}_i\}$  such that index  $i$  has exactly one non-unity coordinate ( $|K| = 1$ ) induce non-zero perturbations of allele frequencies (for one locus). These eigenvectors do not violate random combinations of alleles in  $U$ -locus gametes ( $|U| > 1$ ), i. e. induce zero perturbations of the gamete frequencies.

On the other hand, eigenvectors  $\{\mathbf{u}_i\}$  such that index  $i$  has more than one non-unity coordinates can violate only random combinations of alleles in gametes, and  $\mathbf{p}_U \neq \mathbf{p}_U^*$  for sufficiently large sets of loci  $U$  ( $U \supseteq K$ ). Such eigenvectors cannot induce perturbations of allele frequencies as it is shown above. In other words, eigenvectors (4) with exactly  $k > 1$  multi-index non-unity coordinates (corresponding to a set  $K$  of loci) represent such deviations from linkage equilibrium state  $\mathbf{p}^*$  that break random combinations of alleles, starting from the level of  $K$ -locus gametes. Deviations on the level above  $k$  loci are determined by the existence of  $K$ -locus deviation. Marginal gamete frequencies up to the level of  $k$  loci (except for  $p_K$ ) satisfy multilocus linkage equilibrium relations.

### 3.2 Eigenvectors and the decay of disequilibria in the neutral case

Under nonepistatic selection, eigenvalues have a rather complicated form [7, 8, 9]:

$$\frac{2}{v^*} \sum_{\eta} c(\eta) v^*(\eta) \sum_{U|V} r(U|V) \prod_{k=1}^l [\lambda_{i_k}^{\eta_k} \mu_{i_k}^{1-\eta_k}]^{\delta_{kU}}. \quad (8)$$

Here  $\{\lambda_{i_k}\}$ ,  $\lambda_{1_k} = 1$  and  $\{\mu_{i_k}\}$ ,  $\mu_{1_k} = 1$ ,  $\mu_{i_k} = 0$ ,  $i_k = \overline{2, n_k}$  are eigenvalues of matrices  $\mathbf{D}(\mathbf{p}^{k*}) \mathbf{V}_k / v_k^*$  and  $\mathbf{D}(\mathbf{p}^{k*}) \mathbf{e}_k \mathbf{e}_k^T = \|\mathbf{p}^{k*} \mathbf{e}_k\|^2$ ,  $k = \overline{1, l}$ , respectively, and we preclude the case  $i_1 = i_2 = \dots = i_l = 1$ . At neutrality, the eigenvalues take very simple form.

**Result II** [12]. *Let the eigenvector  $\mathbf{u}_i$  in (4) (with multi-index coordinate  $i_k \neq 1$  if  $k$  belongs to the set  $K$  of loci) is considered as a deviation from multilocus random proportions (3), i. e., as the deviation that breaks random combinations of alleles, starting from the level of  $K$ -locus gametes.*

*Then in the neutral case with no selection, the rate of decay of this deviation is equal to  $1 - r(K)$ , where  $r(K)$  is the probability that recombination takes place between the loci in the set  $K$ .*

**Proof.** Under neutrality conditions, linearization matrix  $\mathbf{B}$  turns out to be

$$2 \sum_{U|V} r(U|V) \bigotimes_{m=1}^l \|\mathbf{p}^m \mathbf{e}_m^T\|^{\delta_{mU}}$$

with the same eigenvectors (4) as before, but instead of (8) associated eigenvalues become

$$2 \sum_{U|V} r(U|V) \prod_{m=1}^l \mu_{i_m}^{\delta_{mU}} = \sum_{U|V} r(U|V) \left( \prod_{m=1}^l \mu_{i_m}^{\delta_{mU}} + \prod_{m=1}^l \mu_{i_m}^{\delta_{mV}} \right). \quad (9)$$

For any partition, a particular summand in parentheses is positive if all zero-valued  $\mu_{i_m}$  are raised to the power 0. For a fixed partition  $U|V$ , it is possible only when all of  $m \in K$  belong to the same set ( $V$  or  $U$ ), that is, the loci in  $K$  have not recombined. In this case (let, for definiteness,  $K \subseteq V$ ),  $\prod_{m=1}^l \mu_{i_m}^{\delta_{mU}} = 1$  and  $\prod_{m=1}^l \mu_{i_m}^{\delta_{mV}} = 0$ . So, in (9) we have the sum of probabilities for recombination events that leave the set  $K$  of loci unbroken. This sum, obviously, equals  $1 - r(K)$ , where  $r(K)$  is the total recombination probability for the loci in  $K$ . The proof is completed.

Since from  $K \supset K'$  it follows that  $r(K) \geq r(K') > 0$  under non-rigid linkage distribution (that allows to recombine any locus pair), we have obvious dynamics properties for the multilocus neutral model. The most quick rate convergence to zero is for deviations (expressed in the form (7)) from linkage equilibrium relations that can be observed only on the level of  $L$ -locus gametes. Then deviations stand that admit linkage equilibrium relations for marginal gamete frequencies up to the level  $l - 1$  loci (with one exception) and so on with the most slow convergence for deviation from random allele combinations for some locus pair.

## 4 Discussion

The results received are in agreement with the previously known. For example, in the neutral case see exact results [1] and [10] on the behavior (depending on the number of loci considered) of linkage disequilibria in time.

In the case of nonepistatic selection on the set of diallele loci, the linearized dynamics of disequilibria is determined by a triangular linearization matrix [2] whence the similarity of disequilibrium behavior to that of neutral case can be deduced.

Our considerations concerns the analogous properties of multiallele multilocus genetic systems with the most clear and simple behavior in a linear

approximation because of diagonalized dynamics. The simplification is due to choosing disequilibria as eigenvectors of linearization matrix. In addition, disequilibria (eigenvectors) are easily classified with respect to the number of loci determining deviations from linkage equilibrium relations.

The given above eigenvector interpretation still holds for some other models of selection keeping the same eigenvector form for appropriate linearization matrices. For example, it is the case for some models of selection in bisexual populations [8].

The model of nonepistatic viability selection with continuous time in population without age-structure

$$dp_{\mathbf{i}}/dt = p_{\mathbf{i}}(v_{\mathbf{i}} - v) + b \sum_{U|V} r(U|V) \left( \sum_{\mathbf{j}: \mathbf{j}_U = \mathbf{i}_U} \sum_{\mathbf{k}: \mathbf{k}_V = \mathbf{i}_V} p_{\mathbf{j}} p_{\mathbf{k}} - p_{\mathbf{i}} \right). \quad (10)$$

gives another example. Generally speaking, here the *general* case of viabilities is allowed,  $v_{\mathbf{i}} \equiv \sum_{\mathbf{j}} v_{\mathbf{ij}} p_{\mathbf{i}} p_{\mathbf{j}}$ ,  $v \equiv \sum_{\mathbf{i}, \mathbf{j}} v_{\mathbf{ij}} p_{\mathbf{i}} p_{\mathbf{j}}$ ,  $b$  is the fertility rate, processes of selection, recombination, and reproduction proceed simultaneously. This model is known [13] to be the limiting case for the discrete-time model (1) under the assumption of weak selection and weak recombination (or for the continuous model given in terms of genotype frequencies if the selection pressure is much weaker than the fertility one).

Due to additive pressure of selection and recombination, linearization matrix for (10) equals the sum of linearization matrices for pure selection submodel (it is formally one-locus) and pure recombination one (formally neutral multilocus). One-locus linearization matrix at the polymorphic equilibrium for viability selection model is known to be  $\mathbf{D}(\mathbf{p}^*)\mathbf{V}$  with right eigenvectors that are orthogonal to  $\mathbf{e} \equiv \bigotimes_{k=1}^l \mathbf{e}_k$ . In the case of *nonepistatic* selection, these eigenvectors are given by (4) and they are eigenvectors of the linearization matrix for recombination submodel. If we denote associated eigenvalues of selection submodel by  $\{\lambda_{\mathbf{i}} \equiv \lambda_{i_1 i_2 \dots i_l}\}$ , the eigenvalues for the total linearization matrix of continuous-time model (10) take the form

$$\lambda_{i_1 i_2 \dots i_l} - br(K), \quad K \equiv \{k : i_k \neq 1\}$$

due to Result II for the neutral model of recombination.

The eigenvalues  $\{\lambda_{i_1 i_2 \dots i_l}\}$  of "selection" linearization matrix at the polymorphic linkage equilibrium fixed point (3) are expressed via the eigenvalues

$\{\lambda_{i_k}\}$  of one-locus components  $\{\mathbf{D}(\mathbf{p}^{k*})\mathbf{V}_k\}$ . As a result, we derive the eigenvalues for *nonepistatic* selection model with continuous time in the form [13]

$$\sum_{\eta} c(\eta) \prod_{k=1}^l (\lambda_{i_k}^{\eta_k} \mu_{i_k}^{1-\eta_k}) - br(K), \quad K \equiv \{k : i_k \neq 1\}.$$

It is easy to see that increasing recombination (in the sense of increasing total recombination parameters  $r(K)$ ) enhances the stability of the linkage equilibrium fixed point because of decreasing the appropriate eigenvalues for the total linearization matrix by the same quantities.

## References

- [1] Bennet, J. H. 1954 On the theory of random mating. *Ann. Eugen.*, **18**, no. 2, 311—317
- [2] Christiansen, F. B. 1990. The generalized multiplicative model for viability selection at multiple loci. *J. Math. Biol.*, 1990, **20**, no. 2, 99—129.
- [3] Christiansen, F. B. 1999. *The Population Genetics of Multiple Loci*. Chichester: Wiley, 365 p.
- [4] Dawson, K. J. 2000 The decay of linkage disequilibrium under random union of gametes: how to calculate Bennett’s principal components. *Theor. Popul. Biol.*, **58**, no. 1, 1—20
- [5] Dawson, K. J. 2002 The evolution of population under recombination: how to linearize the dynamics. *Linear Algebra Appl.* **348**, 115—137
- [6] Karlin, S., Avni, H. 1981 Analysis of central equilibria in multilocus systems: a generalized symmetric viability regime. *Theor. Popul. Biol.*, **20**, no. 2, 241—280.
- [7] Karlin, S., Liberman, U. 1979 Central equilibria in multilocus systems. I. Generalized nonepistatic selection regimes. *Genetics*, **91**, no. 4, 777—798.
- [8] Karlin, S., Liberman, U. 1979 Central equilibria in multilocus systems. II. Bisexual generalized nonepistatic selection models. *Genetics*, **91**, no. 4, 700—816.

- [9] Karlin, S, Liberman, U. 1979 Representation of nonepistatic selection models and analysis of multilocus Hardy-Weinberg equilibrium configurations. *J. Math. Biol.*, **7**, no. 4, 353—374.
- [10] Lyubich, Yu.I. 1992. *Mathematical Structures in Population Genetics*. Berlin-Heidelberg-New York: Springer (Biomathematics, **22**), 373 p.
- [11] Lyubich, Yu., Kirzhner, V. Finiteness of equilibria set for nonepistatic selection under multilocus Mendel dominance. *Appl. Math. Let.* (in press)
- [12] Passekov, V.P. 1994. On the perturbation analysis of neutral multi-locus model of population genetics. In: *Operation Research*. Moscow: Computing Center of Russian Academy of Sciences, 51–86 (in Russian).
- [13] Svirezhev, Y.M., Passekov, V.P. 1990. *Fundamentals of Mathematical Evolutionary Genetics*. Dordrecht et al.: Kluwer Academic Publishers, 395 p.