

Highlights

Prolonged Transient Persistence of Polygenic Polymorphism under Stabilizing Selection Alone

Xiang Shi, Jiayi Kang, Xin Zhao, Nan Sun, Stephen S.-T. Yau, Qi Wu

- The system admits a unique algebraic fully polymorphic equilibrium candidate, which is feasible if and only if all its components lie in $(0, 1)$.
- The equilibrium exhibits strong time-scale separation between rapid mean restoration and slow locus-differentiating instability.
- Nonlinear simulations show that escape from the polymorphic neighbourhood is substantially delayed, yielding prolonged deterministic finite-time persistence of polygenic polymorphism.

Prolonged Transient Persistence of Polygenic Polymorphism under Stabilizing Selection Alone

Xiang Shi^{b,1}, Jiayi Kang^{c,1}, Xin Zhao^a, Nan Sun^c, Stephen S.-T. Yau^{b,*} and Qi Wu^{a,**}

^aBeijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing, 101408, China

^bDepartment of Mathematical Science, Tsinghua University, Beijing, 100084, China

^cHetao Institute of Mathematics and Interdisciplinary Sciences (HIMIS), Shenzhen, 518000, China

ARTICLE INFO

Keywords:

population genetics,
quantitative trait,
genetic diversity,
dynamic analysis,
complex trait

ABSTRACT

In evolutionary theories of quantitative traits, a mutation–selection balance is essential to preserve polymorphism, where stabilizing selection is widely considered to eliminate genetic variation at underlying loci. To isolate the role of selection, we re-examine the effect of stabilizing selection on genetic polymorphism in a mathematically rigorous manner. Following the classical stabilizing-selection model, we explicitly decompose the population mean μ across all n contributing loci, transforming the model within a n -dimensional dynamical system. The fully polymorphic stationarity equations have a unique algebraic candidate, which is a biologically feasible interior equilibrium if and only if all its components lie in $(0, 1)$. Linearization at the equilibrium point reveals a saddle-point structure: under equal allelic effects, exactly one eigenvalue is negative, and $n - 1$ eigenvalues are positive. Under equal allelic effects, the linear stable and unstable eigenspaces are $E^s = \text{span}\{\mathbf{1}\}$ and $E^u = \ker(\mathbf{1}^\top)$, respectively; the corresponding local nonlinear invariant manifolds are tangent to these spaces at $\mathbf{p}^*$. The stable mode restores the population mean toward the equilibrium value μ^* , which need not equal the optimum θ , whereas the unstable modes are mean-neutral to first order. The relative rate ratio $|\lambda_s|/\lambda_u = 2n - 1$ shows that collective relaxation is faster than locus differentiation, while the absolute local residence time is controlled by λ_u and the initial unstable displacement. A nonzero unstable component causes local departure from the fully polymorphic saddle, but the local analysis does not classify the global limiting state. In the full nonlinear system, the non-trivial equilibrium is not asymptotically stable. Simulations show that escape from the polymorphic neighborhood is substantially delayed, producing long finite-time persistence of polymorphism. Consequently, the saddle geometry can generate potentially long deterministic finite-time persistence of polygenic polymorphism, but it does not provide asymptotically stable or finite-population maintenance in the absence of mutation.

1. Introduction

Maintaining genetic diversity at loci underlying quantitative traits is a central problem in evolutionary quantitative genetics. The classical view that stabilizing selection alone is insufficient to maintain genetic polymorphism dates back to Wright[1] and was first systematically investigated by Robertson[2]. Most early analyses focused on single-locus systems, where stabilizing selection was generally expected to drive alleles toward fixation or loss rather than to maintain stable polymorphism. Polygenic formulations were studied by Latter³ and Bulmer⁴, and Barton later developed a general multilocus framework for analysing the stability of polymorphic equilibria⁵. In this framework, the Hessian matrix of the logarithm of mean fitness was used to characterize how the system responds to perturbations around an equilibrium. For n biallelic loci contributing to a single quantitative trait, Barton's analysis showed that, under stabilizing selection alone, a stable polymorphic equilibrium can be maintained at no more than one locus[5]. This result provided a rigorous theoretical basis for the classical view that mutation is essential for preserving heritable genetic variation under stabilizing selection[6].

Consequently, much subsequent work has examined the maintenance of allelic polymorphism within the framework of mutation–selection balance. Previous studies have analysed how polymorphism can be maintained in models incorporating dominance effects[7], epistasis and genetic linkage[8, 9], and heterogeneity in allelic effects across loci[10, 11]. Later work further used eigenvalue-based analyses to reveal threshold behaviour in the maintenance of polymorphism under mutation–selection balance[11]. Together, these studies established mutation–selection balance as a central theoretical framework for explaining the stable maintenance of quantitative genetic variation.

Stabilizing selection in the absence of mutation has also been investigated in several contexts. Some studies examined the stability of interior equilibria in stabilizing-selection models with a small number of loci[8], whereas others considered the maintenance of polymorphism under strong selection[12]. Waxman[13] systematically studied the evolutionary dynamics of quantitative traits under stabilizing selection without mutation in an infinite population, deriving analytical solutions for initially Gaussian trait distributions. Nevertheless, these studies did not fully characterize the local dynamical structure of the interior equilibrium in a general multilocus allele-frequency system. In particular,

*Corresponding author.

**Corresponding author.

¹These authors contributed equally to this work.

how the stable and unstable directions around such an equilibrium shape finite-time polymorphism in the complete absence of mutation remains insufficiently understood.

In this study, we revisit stabilizing selection without mutation by analysing the Jacobian matrix and its eigenvalue spectrum in a coupled multilocus dynamical system. Following the classical stabilizing-selection model, we explicitly decompose the population mean μ across all n contributing loci, thereby formulating the dynamics as an n -dimensional system of allele-frequency equations. We prove that the fully polymorphic stationarity equations admit a unique algebraic candidate $\mathbf{p}^*$, which is an interior equilibrium precisely under the stated feasibility condition. In the equal-effects case, linearization around this equilibrium reveals a saddle-point structure, with one negative eigenvalue of relatively large magnitude and $n - 1$ positive eigenvalues of relatively small magnitude. The stable direction corresponds to collective shifts of allele frequencies that restore the population mean toward its equilibrium value μ^* , whereas the unstable directions correspond to redistributions of allele frequencies across loci that leave the population mean unchanged to first order. This spectral structure produces a strong separation of time scales: mean-restoring dynamics are rapid, while locus-differentiating instability unfolds much more slowly. Thus, although the interior equilibrium is not asymptotically stable, trajectories of the full nonlinear system can remain close to it for extended transient periods, with the characteristic unstable timescale given by

$$\tau_u = \frac{1}{sa^2p^*(1-p^*)}.$$

This provides a quantitative mechanism for deterministic finite-time persistence near the saddle; it is not a claim of stable or finite-population maintenance without mutation.

2. Model Formulation

We first formulate the allele-frequency dynamics of a multilocus quantitative trait under stabilizing selection in the absence of mutation. This formulation makes explicit the dependence of the population mean on all contributing loci, thereby converting the classical single-locus expression into a coupled n -dimensional dynamical system. We then characterize the fully polymorphic equilibrium candidate and its feasibility, before analysing its local stability and the resulting separation of time scales.

2.1. Interior equilibrium: existence, uniqueness, and feasibility

We consider a quantitative trait determined by n bi-allelic loci, with alleles B_i and b_i at locus i ($i = 1, \dots, n$). The population frequency of allele B_i is denoted by p_i . Let Z denote the random phenotypic value of an individual and let z denote a realized phenotypic value. We assume $Z \sim \mathcal{N}(\mu, \sigma^2)$. Under random mating, Hardy-Weinberg equilibrium, additive gene action with no dominance or epistasis, and linkage equilibrium, the mean and variance

are expressed as

$$\mu = m + 2 \sum_{i=1}^n a_i p_i, \quad \sigma^2 = 2 \sum_{i=1}^n a_i^2 p_i (1 - p_i) + \sigma_E^2, \quad (1)$$

where $a_i > 0$ is the additive effect of substituting b_i with B_i , m is a baseline phenotypic value (the minimum possible mean when all $p_i = 0$), and σ_E^2 is the environmental variance. Here linkage equilibrium is used as a dynamical closure, or equivalently as a leading-order quasi-linkage-equilibrium approximation when recombination is fast relative to selection; it is not merely an assumption of zero initial linkage disequilibrium.

Stabilising selection acts on the trait through a Gaussian fitness function

$$W(z) = \exp\left(-\frac{(z - \theta)^2}{2w^2}\right), \quad (2)$$

where θ is the optimal phenotypic value and w^2 measures the width of the fitness peak. For $Z \sim \mathcal{N}(\mu, \sigma^2)$, Gaussian integration gives an exact expression for $\bar{W}$. Under the leading-order weak-selection approximation derived in Appendix A, we retain

$$\ell_0(\mathbf{p}) := -\frac{s}{2} [\sigma^2(\mathbf{p}) + \{\mu(\mathbf{p}) - \theta\}^2].$$

A sufficient uniform weak-selection regime along the trajectory is $s\sigma^2(\mathbf{p}) \ll 1$ and $s\{\mu(\mathbf{p}) - \theta\}^2 \ll 1$. At this order, Wright's formula for the one-generation allele-frequency increment is

$$\Delta p_i = \frac{p_i(1 - p_i)}{2} \frac{\partial \ell_0}{\partial p_i} + O(s^2), \quad (3)$$

and substituting the expressions for μ and σ^2 , one obtains the single-locus dynamics

$$\frac{dp_i}{dt} = \frac{1}{2} sa_i p_i (1 - p_i) [a_i (2p_i - 1) + 2(\theta - \mu)], \quad (4)$$

Here t is measured in generations and Eq. (4) is the leading-order continuous-time weak-selection approximation obtained by replacing the one-generation increment by its continuous interpolation; terms of order s^2 and higher are omitted.

To capture the coupling among loci, we treat μ as an explicit function of all allele frequencies. Using (1), the mean can be decomposed as

$$\mu = 2a_i p_i + m + 2 \sum_{j \neq i} a_j p_j. \quad (5)$$

Substituting this decomposition into (4) yields the dynamics of p_i that incorporate the feedback from the focal locus to the population mean:

$$\frac{dp_i}{dt} = \frac{1}{2} sa_i p_i (1 - p_i) \left[-a_i (2p_i + 1) - 4 \sum_{j \neq i} a_j p_j + 2(\theta - m) \right].$$

(6)

For all $i = 1, \dots, n$, these equations form an n -dimensional coupled dynamical system, which we analyze in the following sections.

Remark 1 (Gradient representation). *The coupled system has the generalized gradient form*

$$\dot{\mathbf{p}} = G(\mathbf{p}) \nabla_{\mathbf{p}} \ell_0(\mathbf{p}), \quad G(\mathbf{p}) = \frac{1}{2} \text{diag}\{p_i(1-p_i)\}_{i=1}^n.$$

On $(0, 1)^n$, introduce the angular coordinates

$$p_i = \sin^2\left(\frac{y_i}{2}\right), \quad 0 < y_i < \pi, \quad \Phi(\mathbf{y}) = \frac{1}{2} \ell_0(\mathbf{p}(\mathbf{y})).$$

Since $dp_i/dy_i = \sqrt{p_i(1-p_i)}$, the transformed system is

$$\dot{\mathbf{y}} = \nabla_{\mathbf{y}} \Phi(\mathbf{y}),$$

where

$$\Phi(\mathbf{y}) = -\frac{s}{4} \left[\sigma_E^2 + \frac{1}{2} \sum_{i=1}^n a_i^2 \sin^2 y_i + \left\{ m + \sum_{i=1}^n a_i(1 - \cos y_i) - \theta \right\}^2 \right].$$

Consequently,

$$\frac{d\Phi}{dt} = \|\nabla_{\mathbf{y}} \Phi\|^2 \geq 0.$$

For $n = 2$ and $a_1 = a_2 = a$, let $d = sa^2 p^*(1-p^*)$ and $y_1^* = y_2^* = 2 \arcsin \sqrt{p^*}$. Then

$$\nabla_{\mathbf{y}}^2 \Phi(\mathbf{y}^*) = d \begin{pmatrix} -1 & -2 \\ -2 & -1 \end{pmatrix},$$

whose eigenvalues are $-3d$ and d , with eigenvectors proportional to $(1, 1)^T$ and $(1, -1)^T$, respectively. Thus the stationary point has the classical saddle geometry.

An equilibrium point $\mathbf{p}^* = (p_1^*, \dots, p_n^*)^T$ satisfies $dp_i/dt = 0$ for all i . From Equation (6), this condition implies that, at each locus i , one of the following must hold:

$$\begin{cases} p_i = 0, \\ p_i = 1, \\ -2a_i p_i - 4 \sum_{j \neq i} a_j p_j + 2(\theta - m) - a_i = 0. \end{cases} \quad (7)$$

The first two cases correspond to boundary equilibria, where allele B_i is either lost or fixed. If the third condition is imposed at every locus, the resulting linear system defines a unique algebraic candidate for a fully polymorphic equilibrium. It is an interior equilibrium only when $\mathbf{p}^* \in (0, 1)^n$. We focus on this interior solution because it is the relevant equilibrium for analysing whether stabilizing selection alone

can support multilocus polymorphism, at least over finite transient time scales.

Define $\mathbf{p} = (p_1, \dots, p_n)^T \in \mathbb{R}^n$. The n non-trivial equilibrium conditions in Equation (7) can be written in matrix form as

$$\mathbf{A}\mathbf{p} = \mathbf{h}, \quad (8)$$

where the coefficient matrix $\mathbf{A}$ and vector $\mathbf{h}$ are defined by

$$\mathbf{A}_{ij} = \begin{cases} -2a_i, & i = j, \\ -4a_j, & i \neq j, \end{cases} \quad \mathbf{h} = (a_1 - 2(\theta - m), \dots, a_n - 2(\theta - m))^T.$$

Theorem 1. *Suppose that $a_i > 0$ for all $i = 1, 2, \dots, n$, and $n \geq 2$. Then the determinant of $\mathbf{A}$ is*

$$\det(\mathbf{A}) = 2^n(1 - 2n) \prod_{i=1}^n a_i \neq 0.$$

Consequently, $\mathbf{A}$ is invertible, and the linear system $\mathbf{A}\mathbf{p} = \mathbf{h}$ has a unique solution $\mathbf{p}^* \in \mathbb{R}^n$.

Proof. See Appendix A. $\square$

Theorem 1 guarantees the existence and uniqueness of an algebraic solution to the interior equilibrium equations. However, biological feasibility further requires that every component of this solution lies strictly between zero and one, namely $p_i^* \in (0, 1)$ for all i . This condition ensures that all loci remain polymorphic at the interior equilibrium. We therefore derive the closed-form expression of the solution and characterize the parameter regime in which it is biologically feasible.

Theorem 2. *Under the assumptions of Theorem 1, the unique solution $\mathbf{p}^*$ to (8) has the explicit form*

$$p_i^* = \frac{1}{2} + \frac{\theta - m - \sum_{j=1}^n a_j}{a_i(2n - 1)}, \quad i = 1, \dots, n.$$

Moreover, the condition $p_i^* \in (0, 1)$ for all i holds if and only if

$$\max_i \left(\sum_{j=1}^n a_j - \frac{2n-1}{2} a_i \right) < \theta - m < \min_i \left(\sum_{j=1}^n a_j + \frac{2n-1}{2} a_i \right).$$

Equivalently, if $a_{\min} := \min_i a_i$, feasibility is characterized by

$$\left| \theta - m - \sum_{j=1}^n a_j \right| < \frac{2n-1}{2} a_{\min}.$$

Proof. See Appendix A. $\square$

These inequalities define the parameter region in which the unique algebraic solution is a biologically feasible interior equilibrium. Within this region, all loci are polymorphic at equilibrium. Outside this region, the linear system still has a unique algebraic solution, but at least one component lies

outside the admissible interval $(0, 1)$, indicating that a fully polymorphic interior equilibrium is not biologically feasible.

The corresponding equilibrium population mean is

$$\mu^* = m + 2 \sum_{i=1}^n a_i p_i^* = \theta + \frac{\theta - m - \sum_{j=1}^n a_j}{2n - 1}.$$

Hence $\mu^* = \theta$ if and only if $\theta - m = \sum_j a_j$. Moreover, feasibility implies

$$|\mu^* - \theta| < \frac{a_{\min}}{2}.$$

Corollary 1. *In the special case where $a_i = a > 0$ for all i , the equilibrium frequencies are identical across loci and reduce to*

$$p_i^* = p^* = \frac{1}{2} + \frac{\theta - m - na}{a(2n - 1)} = -\frac{1}{4n - 2} + \frac{\theta - m}{a(2n - 1)}, \quad \forall i.$$

The feasibility condition simplifies to

$$\frac{a}{2} < \theta - m < \frac{a(4n - 1)}{2},$$

or equivalently,

$$m + \frac{a}{2} < \theta < m + \frac{a(4n - 1)}{2}.$$

In this symmetric case, the full attainable range of the population mean μ , defined in Eq. (5), is $[m, m + 2na]$, whereas the feasibility interval for the optimum is

$$m + \frac{a}{2} < \theta < m + 2na - \frac{a}{2}.$$

Thus, the selective optimum must lie within the attainable phenotypic range, excluding narrow boundary regions of width $a/2$. Since the allelic effect a is typically small relative to the trait scale, this feasible interval approximates the full phenotypic range for large n . Therefore, in the equal-effect case, a fully polymorphic equilibrium exists if and only if $m + a/2 < \theta < m + 2na - a/2$, rather than for every optimum in the full attainable interval $(m, m + 2na)$.

2.2. Linear stability analysis: saddle-point spectrum

We next examine the local stability of the interior equilibrium $\mathbf{p}^*$ whose existence, uniqueness and feasibility were established above. Linearizing the coupled dynamical system (6) around $\mathbf{p}^*$ yields a Jacobian matrix whose eigenvalue spectrum determines the local nature of the equilibrium. Throughout this section, we assume $a_i > 0$ for all i and $n \geq 2$.

2.2.1. Linearization and Jacobian structure

For convenience, write the dynamics (6) as $\dot{p}_i = f_i(\mathbf{p})$, where

$$f_i(\mathbf{p}) = \frac{1}{2} s a_i p_i (1 - p_i) \left[-a_i (2p_i + 1) - 4 \sum_{j \neq i} a_j p_j + 2(\theta - m) \right].$$

At the interior equilibrium $\mathbf{p}^*$, the bracketed term vanishes by the interior equilibrium condition. When differentiating f_i , the derivative of the prefactor $p_i(1 - p_i)$ therefore contributes a term proportional to this bracketed term, and this contribution vanishes at $\mathbf{p}^*$. Consequently, the Jacobian is determined by the derivative of the bracketed term, multiplied by the equilibrium value of the prefactor.

Define the positive scalars

$$q_i := s a_i p_i^* (1 - p_i^*) > 0, d_i := s a_i^2 p_i^* (1 - p_i^*) > 0, i = 1, \dots, n.$$

Theorem 3 (Jacobian matrix at the interior equilibrium). *At the equilibrium $\mathbf{p}^*$, the Jacobian matrix J has entries*

$$J_{ij} = \begin{cases} -d_i, & i = j, \\ -2q_i a_j, & i \neq j. \end{cases} \quad (9)$$

Equivalently, J admits the factorisation

$$J = -D_q M, \quad (10)$$

where $D_q = \text{diag}(q_1, \dots, q_n)$ and M is the matrix with entries $M_{ij} = (2 - \delta_{ij})a_j$ (i.e., $M_{ii} = a_i$, $M_{ij} = 2a_j$ for $i \neq j$).

Proof. Let

$$g_i(\mathbf{p}) = -a_i(2p_i + 1) - 4 \sum_{j \neq i} a_j p_j + 2(\theta - m).$$

Then

$$f_i(\mathbf{p}) = \frac{s a_i}{2} p_i (1 - p_i) g_i(\mathbf{p}).$$

By the product rule,

$$\frac{\partial f_i}{\partial p_j} = \frac{s a_i}{2} \left[(1 - 2p_i) g_i(\mathbf{p}) \delta_{ij} + p_i (1 - p_i) \frac{\partial g_i}{\partial p_j} \right],$$

At the interior equilibrium, $g_i(\mathbf{p}^*) = 0$, so the derivatives are

$$\frac{\partial g_i}{\partial p_i} = -2a_i, \quad \frac{\partial g_i}{\partial p_j} = -4a_j \quad (j \neq i).$$

Substituting these derivatives gives

$$\frac{\partial f_i}{\partial p_i} \Big|_{\mathbf{p}^*} = \frac{s a_i}{2} p_i^* (1 - p_i^*) (-2a_i) = -s a_i^2 p_i^* (1 - p_i^*) = -d_i,$$

and for $j \neq i$,

$$\frac{\partial f_i}{\partial p_j} \Big|_{\mathbf{p}^*} = \frac{s a_i}{2} p_i^* (1 - p_i^*) (-4a_j) = -2s a_i a_j p_i^* (1 - p_i^*) = -2q_i a_j.$$

This proves Equation (9). The factorised form follows from $M = 2\mathbf{1}\mathbf{a}^T - \text{diag}(\mathbf{a})$, where $\mathbf{1} = (1, \dots, 1)^T$ and $\mathbf{a} = (a_1, \dots, a_n)^T$. Equivalently, $M_{ij} = (2 - \delta_{ij})a_j$, and hence $J = -D_q M$. $\square$

The structure of this Jacobian is reminiscent of the stability matrix derived by Barton for mutation–selection balance⁵. Here, however, the same type of spectral analysis is applied to the pure-selection limit, allowing us to isolate the local dynamical consequences of stabilizing selection in the complete absence of mutation.

2.2.2. Saddle-point spectrum and eigenspaces

We next characterize the eigenvalue spectrum of the Jacobian matrix J . The key observation is that J can be written as a diagonal matrix plus a rank-one perturbation. This structure leads to a secular equation whose roots determine the eigenvalues. For clarity, we first consider the case in which the quantities d_i are pairwise distinct. Strict interlacing is asserted only in the pairwise-distinct case. The general sign structure, including repeated values, is proved separately by symmetrisation and Sylvester's law of inertia. Let

$$D = \text{diag}(d_1, \dots, d_n), \mathbf{q} = (q_1, \dots, q_n)^\top, \mathbf{a} = (a_1, \dots, a_n)^\top.$$

Since $d_i = q_i a_i$, the Jacobian can be written as

$$J = D - 2\mathbf{q}\mathbf{a}^\top.$$

Thus J is a diagonal matrix perturbed by the rank-one coupling term $-2\mathbf{q}\mathbf{a}^\top$.

Theorem 4 (Secular equation). *Suppose that $d_i \neq d_k$ for all $i \neq k$. Then none of the values d_i is an eigenvalue of J , and λ is an eigenvalue of J if and only if*

$$F(\lambda) := \sum_{i=1}^n \frac{d_i}{d_i - \lambda} = \frac{1}{2}. \quad (11)$$

Consequently, all eigenvalues of J are real.

Proof. Write

$$J - \lambda I = D - \lambda I - 2\mathbf{q}\mathbf{a}^\top.$$

For $\lambda \notin \{d_1, \dots, d_n\}$, the matrix $D - \lambda I$ is invertible. Applying the matrix determinant lemma gives

$$\det(J - \lambda I) = \det(D - \lambda I) [1 - 2\mathbf{a}^\top (D - \lambda I)^{-1} \mathbf{q}].$$

Since $d_i = a_i q_i$, the equation becomes

$$\det(J - \lambda I) = \prod_{i=1}^n (d_i - \lambda) \left[1 - 2 \sum_{i=1}^n \frac{d_i}{d_i - \lambda} \right].$$

Thus, for $\lambda \notin \{d_1, \dots, d_n\}$, the characteristic equation $\det(J - \lambda I) = 0$ is equivalent to

$$\sum_{i=1}^n \frac{d_i}{d_i - \lambda} = \frac{1}{2}.$$

It remains to note that, when the d_i are distinct, none of the points $\lambda = d_k$ is an eigenvalue. Indeed, although the expression above has a pole at $\lambda = d_k$, the characteristic polynomial has the finite limiting value

$$\lim_{\lambda \rightarrow d_k} \prod_{i=1}^n (d_i - \lambda) \left[1 - 2 \sum_{i=1}^n \frac{d_i}{d_i - \lambda} \right] = -2d_k \prod_{i \neq k} (d_i - d_k),$$

which is non-zero because $d_k > 0$ and $d_i \neq d_k$ for all $i \neq k$. Therefore, the eigenvalues of J are exactly the roots of the secular equation.

Finally, the function

$$F(\lambda) = \sum_{i=1}^n \frac{d_i}{d_i - \lambda}$$

is strictly increasing on every interval of its domain, because

$$F'(\lambda) = \sum_{i=1}^n \frac{d_i}{(d_i - \lambda)^2} > 0.$$

As shown below, this monotonicity yields exactly one real root in each relevant interval. Hence all eigenvalues of J are real. $\square$

Theorem 5 (Spectral sign structure). *Let $0 < d_1 < d_2 < \dots < d_n$. Then J has exactly one strictly negative eigenvalue $\lambda_1 \in (-\infty, 0)$ and exactly $n-1$ strictly positive eigenvalues $\lambda_{k+1} \in (d_k, d_{k+1})$ for $k = 1, \dots, n-1$.*

Proof. Consider $F(\lambda) = \sum_{i=1}^n d_i/(d_i - \lambda)$. Since $F'(\lambda) = \sum_{i=1}^n d_i/(d_i - \lambda)^2 > 0$, F is strictly increasing on each interval of its domain. On the interval $(-\infty, d_1)$, we have

$$\lim_{\lambda \rightarrow -\infty} F(\lambda) = 0, \quad F(0) = n.$$

Since $n \geq 2$, the equation $F(\lambda) = 1/2$ has exactly one solution in $(-\infty, 0)$. This gives one strictly negative eigenvalue. Next, on each interval (d_k, d_{k+1}) for $k = 1, \dots, n-1$, the function F satisfies

$$\lim_{\lambda \rightarrow d_k^+} F(\lambda) = -\infty, \quad \lim_{\lambda \rightarrow d_{k+1}^-} F(\lambda) = +\infty.$$

So the equation $F(\lambda) = 1/2$ has exactly one solution in (d_k, d_{k+1}) . These roots give $n-1$ strictly positive eigenvalues. $\square$

Corollary 2 (General sign structure). *For arbitrary $d_i > 0$, including repeated values, J is diagonalizable over $\mathbb{R}$ and has exactly one strictly negative eigenvalue and exactly $n-1$ strictly positive eigenvalues.*

Proof. Define

$$P = \text{diag}\left(\sqrt{\frac{q_1}{a_1}}, \dots, \sqrt{\frac{q_n}{a_n}}\right), \quad \mathbf{v} = (\sqrt{d_1}, \dots, \sqrt{d_n})^\top.$$

Since $d_i = a_i q_i$, direct calculation gives

$$P^{-1}JP = D - 2\mathbf{v}\mathbf{v}^\top = D^{1/2}(I - 2\mathbf{1}\mathbf{1}^\top)D^{1/2}.$$

Thus J is similar to a real symmetric matrix. This symmetric matrix is congruent to $I - 2\mathbf{1}\mathbf{1}^\top$, which has eigenvalue $1 - 2n < 0$ along $\mathbf{1}$ and eigenvalue $1 > 0$ with multiplicity $n-1$ on $\ker(\mathbf{1}^\top)$. Sylvester's law of inertia proves the claim. $\square$

Corollary 3 (Saddle-point structure). *Every feasible fully polymorphic equilibrium $\mathbf{p}^*$ is a hyperbolic saddle point. The linearized system has a one-dimensional stable eigenspace associated with the unique negative eigenvalue, and an $(n-1)$ -dimensional unstable eigenspace associated with the positive eigenvalues. Equivalently, in the full nonlinear system, the local stable and unstable manifolds of $\mathbf{p}^*$ have dimensions 1 and $n-1$, respectively, and are tangent to these eigenspaces at $\mathbf{p}^*$.*

Corollary 4 (Equal allelic effects). *Suppose that all allelic effects are equal, $a_i = a > 0$ for all i . Then $p_i^* = p^*$ for all i , and hence $d_i = d = sa^2 p^*(1-p^*)$ for all i . In this case, the eigenvalues of J are*

$$\lambda_s = -(2n-1)d, \quad \lambda_u = d \quad (\text{with multiplicity } n-1).$$

The stable eigenspace is

$$E^s = \text{span}\{\mathbf{1}\}, \quad \mathbf{1} = (1, 1, \dots, 1)^T,$$

and the unstable eigenspace is

$$E^u = \ker(\mathbf{1}^T) = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0 \right\}.$$

Proof. When $a_i = a$ for all i , the Jacobian reduces to

$$J = dI - 2d \mathbf{1}\mathbf{1}^T.$$

For the collective direction $\mathbf{1}$,

$$J\mathbf{1} = d\mathbf{1} - 2d \mathbf{1}\mathbf{1}^T \mathbf{1} = d\mathbf{1} - 2dn\mathbf{1} = -(2n-1)d\mathbf{1}.$$

Thus $\mathbf{1}$ is an eigenvector with eigenvalue $\lambda_s = -(2n-1)d$.

For any vector $\mathbf{x}$ satisfying $\mathbf{1}^T \mathbf{x} = 0$, we have

$$J\mathbf{x} = d\mathbf{x} - 2d \mathbf{1}\mathbf{1}^T \mathbf{x} = d\mathbf{x}.$$

Therefore every vector in $\ker(\mathbf{1}^T)$ is an eigenvector with eigenvalue d . Since this subspace has dimension $n-1$, the eigenvalue d has multiplicity $n-1$. $\square$

The equal-effects case reveals the dynamical meaning of the saddle-point spectrum. The stable direction $\text{span}\{\mathbf{1}\}$ corresponds to collective shifts of all allele frequencies in the same direction. Such perturbations change the population mean μ and are rapidly restored by stabilizing selection toward the equilibrium mean μ^* , which equals θ only when $\theta - m = na$. By contrast, the unstable subspace $\ker(\mathbf{1}^T)$ consists of redistributions of allele frequencies across loci whose total change is zero. Because $a_i = a$, these redistributions leave the population mean unchanged to first order, but they differentiate loci from one another and cause departure from a sufficiently small neighbourhood of the fully polymorphic saddle when the unstable component is nonzero. This local statement does not classify the global limiting state.

The resulting separation of time scales is

$$\frac{|\lambda_s|}{\lambda_u} = 2n-1.$$

Equivalently,

$$\tau_s = \frac{1}{(2n-1)d}, \quad \tau_u = \frac{1}{d}, \quad \frac{\tau_u}{\tau_s} = 2n-1.$$

Thus mean-restoring dynamics are faster than locus-differentiating instability by a factor of $2n-1$. This is a relative rate comparison: at fixed s , a , and p^* , the absolute unstable time $\tau_u = 1/[sa^2 p^*(1-p^*)]$ is independent of n . The local residence time also depends logarithmically on the initial unstable displacement. If the ODE is interpreted as a first-order approximation to the discrete-generation recursion, the corresponding multipliers are $\rho_s = 1 - (2n-1)d + O(s^2)$ and $\rho_u = 1 + d + O(s^2)$; thus the continuous rate interpretation requires $(2n-1)d \ll 1$ and $d \ll 1$.

For an interior trajectory,

$$\frac{d}{dt} \log \left(\frac{p_i}{1-p_i} \right) = \frac{sa_i}{2} [a_i(2p_i-1) + 2\{\theta - \mu(\mathbf{p})\}].$$

The right-hand side is bounded on $[0, 1]^n$, so a trajectory starting in $(0, 1)^n$ cannot reach $p_i = 0$ or $p_i = 1$ in finite continuous time. Accordingly, fixation or loss in this ODE denotes an asymptotic boundary limit or a threshold-based numerical classification, not a finite-time consequence of the local saddle theorem.

3. Numerical experiments and results

The analytical results above establish a one-dimensional stable eigenspace, an $(n-1)$ -dimensional unstable eigenspace, and the spectral separation $|\lambda_s|/\lambda_u = 2n-1$ at the interior equilibrium. We next examine how this local geometry is expressed in the full nonlinear dynamics. The first experiment compares trajectories initiated along the stable and unstable eigenspaces and under a mixed perturbation. The second tests the predicted dependence of the transient lifetime on the initial amplitude of an unstable perturbation. These simulations evaluate consequences of the analytical saddle structure; they are not used to infer that structure.

3.1. Transient dynamics under stable, unstable, and mixed perturbations

We used

$$n = 100, \quad a = 0.01, \quad m = 0.30, \quad s = 0.05,$$

and defined the population mean by

$$\mu(\mathbf{p}) = m + 2 \sum_{i=1}^n a_i p_i, \quad a_i = a.$$

The phenotypic optimum was $\theta = 1.29$. Solving the interior fixed-point condition gives

$$p_i^* = p^* = 0.495, \quad i = 1, \dots, n.$$

Linearization gives the characteristic time scales

$$\tau_s = 4.02 \times 10^3 \text{ generations}, \quad \tau_u = 8.00 \times 10^5 \text{ generations}.$$

For equal allelic effects, their ratio follows analytically from the eigenvalue spectrum:

$$\frac{\tau_u}{\tau_s} = \frac{|\lambda_s|}{\lambda_u} = 2n - 1 = 199.$$

Consequently, mean-restoring relaxation is predicted to be about two orders of magnitude faster than locus-redistributing growth.

To construct reproducible controlled perturbations, we generated a random vector and projected it onto E^u . The resulting zero-sum vector was scaled so that

$$\mathbf{1}^\top \boldsymbol{\eta}_u = 0, \quad \max_i |\eta_{u,i}| = 1.$$

With this normalization, ε_u represents the maximum allele-frequency displacement in the unstable direction. We set $\varepsilon_u = 0.008$ and chose the stable amplitude so that the stable and unstable modal norms were initially comparable:

$$\varepsilon_s = \sqrt{n}(0.5 - p^*) = 0.050.$$

For the fixed projected direction used here, $\|\varepsilon_u \boldsymbol{\eta}_u\| = 0.047$, so the stable component is initially slightly larger. The three initial conditions were

$$\begin{aligned} \mathbf{p}_s(0) &= \mathbf{p}^* + \varepsilon_s \mathbf{e}_s, \\ \mathbf{p}_u(0) &= \mathbf{p}^* + \varepsilon_u \boldsymbol{\eta}_u, \\ \mathbf{p}_m(0) &= \mathbf{p}^* + \varepsilon_s \mathbf{e}_s + \varepsilon_u \boldsymbol{\eta}_u. \end{aligned}$$

The initial allele frequencies of unstable and mixed situations ranged approximately from 0.49 to 0.51. All initial conditions lay within the admissible state space, and no boundary clipping or post hoc adjustment was required.

In the equal-effect system, $E^s = \text{span}\{\mathbf{1}\}$ is the tangent direction of the diagonal set

$$\mathcal{D} = \{\mathbf{p} \in (0, 1)^n : p_1 = \dots = p_n\}.$$

Because $\varepsilon_s = \sqrt{n}(0.5 - p^*)$, the stable-only initial condition satisfies $\mathbf{p}_s(0) = 0.5 \mathbf{1}$ and therefore lies exactly on $\mathcal{D}$. The diagonal set is invariant under the full nonlinear dynamics and coincides locally with the stable manifold of $\mathbf{p}^*$. Thus, the stable-only experiment represents a finely tuned trajectory on the stable manifold rather than a generic perturbation of the polymorphic equilibrium. By contrast, the unstable-only and mixed conditions contain locus-redistributing components transverse to this manifold. In these two conditions, the unstable component has zero sum and therefore does not contribute to the initial change in the population mean. Although the unstable-only displacement initially lies entirely in E^u , this subspace is generally not invariant under the full nonlinear dynamics; nonlinear coupling may therefore generate a stable component at later times.

For each trajectory, we decomposed the displacement from the equilibrium as

$$\mathbf{p}(t) - \mathbf{p}^* = c_s(t) \mathbf{e}_s + \mathbf{u}(t), \quad \mathbf{u}(t) \in E^u.$$

We then plotted the normalized modal amplitudes

$$\frac{|c_s(t)|}{\|\delta \mathbf{p}(0)\|} \quad \text{and} \quad \frac{\|\mathbf{u}(t)\|}{\|\delta \mathbf{p}(0)\|}, \quad \delta \mathbf{p}(0) = \mathbf{p}(0) - \mathbf{p}^*.$$

This common reference avoids division by a vanishing initial component and preserves the relative magnitudes of the stable and unstable projections.

To quantify finite-time persistence and the gradual loss of within-locus diversity, we also recorded the mean heterozygosity

$$H(t) = \frac{1}{n} \sum_{i=1}^n 2p_i(t)(1 - p_i(t)),$$

the number of polymorphic loci

$$N_{\text{poly}}(t) = \#\{i : \varepsilon_b < p_i(t) < 1 - \varepsilon_b\}, \quad \varepsilon_b = 10^{-3},$$

and the spread of allele frequencies across loci, measured by

$$\min_i p_i(t) \quad \text{and} \quad \max_i p_i(t).$$

The modal amplitudes directly test the spectral prediction, whereas heterozygosity and the frequency extrema describe its continuous genetic consequences. By contrast, N_{poly} is a threshold-based indicator and is therefore expected to respond only after allele frequencies have moved sufficiently close to the boundaries.

The full nonlinear ODE system was integrated numerically using the adaptive Dormand–Prince RK45 method with relative and absolute tolerances of 10^{-10} and 10^{-13} , respectively. Each trajectory was integrated to 4.0×10^6 generations. Modal amplitudes below 10^{-11} were treated as numerically unresolved and were omitted from the logarithmic plots. This heuristic threshold affected only the visualisation; the integrated trajectories were neither altered nor smoothed.

Repeating the simulations with the higher-order DOP853 method and tighter tolerances produced indistinguishable trajectories, with the maximum allele-frequency difference below 10^{-13} , confirming that the reported dynamics are insensitive to the numerical solver settings.

The horizontal axes are measured in generations. The mode-decomposition panels cover the first 1.0×10^6 generations, whereas the heterozygosity and allele-frequency panels cover the first 4.0×10^6 generations. The unstable-only and mixed mode panels include early-time insets covering the first 1000 generations.

The full nonlinear trajectories reproduce the qualitative time-scale separation predicted by the local linearization (Fig. 1a–1c). Under the stable-only perturbation, the stable component decays rapidly, while no unstable component is resolved. In the equal-effect case, the diagonal $p_i = r$ is invariant and the dynamics restricted to it satisfy

$$\dot{r} = -sa^2(2n - 1)r(1 - r)(r - p^*).$$

Consequently, the stable-only trajectory approaches the interior equilibrium monotonically, without generating an unstable component, and polymorphism is maintained along this

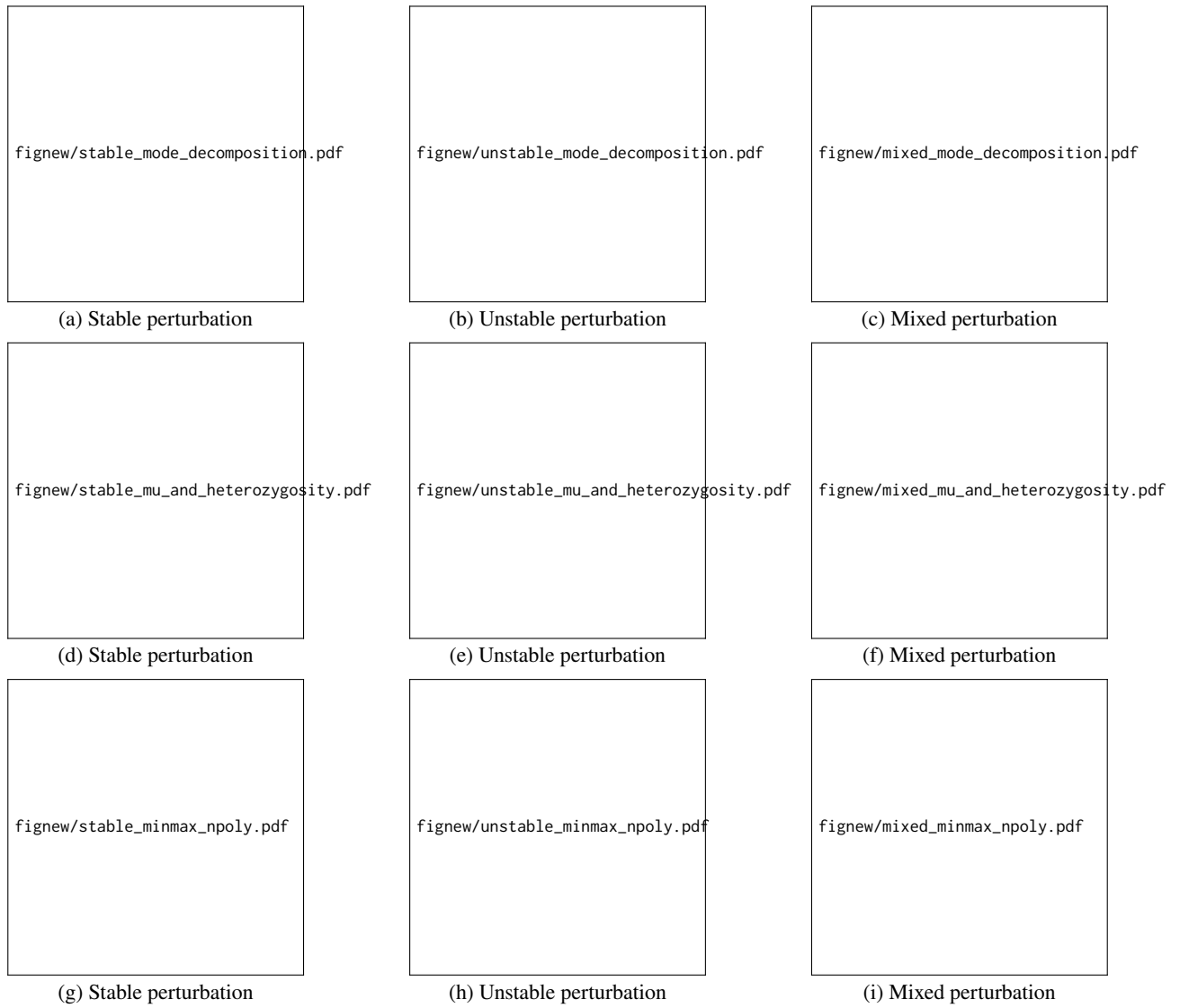


Figure 1: Transient polymorphism in the full nonlinear dynamics under stable, unstable, and mixed perturbations. The first row shows the normalised stable and unstable modal amplitudes, the second row shows normalised heterozygosity, and the third row shows the range of allele frequencies and the proportion of polymorphic loci. The horizontal axes are measured in generations, and the vertical dashed line marks the unstable time scale τ_u . The dotted horizontal line denotes the numerical reporting threshold. The early-time inset in the unstable-only panel resolves the nonlinear generation of a small stable projection from an initially pure unstable displacement. The inset in the mixed panel shows the crossover from initial stable-mode dominance to unstable-mode dominance within the first 1000 generations. The rapid relaxation of the stable component and the much slower growth of the unstable component demonstrate the predicted separation between mean restoration and locus-level differentiation.

precisely specified stable manifold. This maintenance is conditional rather than robust: any nonzero locus-redistributing component moves the trajectory away from the stable manifold and is subsequently amplified on the much longer unstable time scale. Under the unstable-only perturbation, the unstable component grows gradually on the much longer unstable time scale, accompanied by a small stable projection generated through nonlinear coupling. The early-time inset makes this generation from an initially vanishing stable projection visible. In the mixed condition, the initial stable and unstable norms are 0.050 and 0.047, respectively. The stable component therefore dominates initially but decays

rapidly, and the two modal amplitudes cross after approximately 235 generations. The mixed trajectory subsequently becomes dominated by the growing unstable component, which closely tracks the corresponding component of the unstable-only trajectory.

The heterozygosity trajectories provide a complementary measure of locus-level redistribution (Fig. 1d–1f). In the stable-only case, the allele frequencies return toward the interior equilibrium and the normalised heterozygosity remains close to one. By contrast, the unstable-only and mixed trajectories exhibit a prolonged initial plateau followed by

a decline. The decline slows as loci approach the boundary regions, where the factor $p_i(1 - p_i)$ suppresses further deterministic change. After the initial fast transient, these two trajectories become nearly indistinguishable, indicating that their subsequent behaviour is governed primarily by the locus-redistributing component.

The allele-frequency extrema provide consistent evidence of locus-level redistribution (Fig. 1g–1i). They remain close to p^* in the stable-only case but move progressively toward opposite boundaries in the unstable-only and mixed cases. Despite the accompanying decline in heterozygosity, all loci remain within the operational polymorphic range $10^{-3} < p_i < 1 - 10^{-3}$, so N_{poly}/n remains equal to one. Thus, the continuous measures reveal substantial erosion of within-locus diversity before the threshold-based polymorphism count changes. The displayed trajectories indicate movement toward lower-dimensional boundary regions; exact fixation or loss is approached asymptotically in the deterministic ODEs and is not observed within the reported interval.

Taken together, these results distinguish conditional maintenance on the stable manifold from robust maintenance under general perturbations. Stabilising selection can preserve polymorphism for an initial condition lying exactly on the invariant diagonal stable manifold, where all allele frequencies move collectively toward the interior equilibrium. However, this stable manifold is only one-dimensional within the n -dimensional allele-frequency space. A generic perturbation contains a component corresponding to frequency redistribution among loci, and this component eventually drives the trajectory away from the polymorphic equilibrium. Stabilising selection therefore produces prolonged transient polymorphism for general perturbations, rather than an asymptotically stable polymorphic balance. The local analysis establishes departure from the interior equilibrium but does not by itself determine the ultimate limiting boundary state. This transient mechanism complements, but does not replace, classical mutation–selection balance.

3.2. Escape-time scaling under unstable perturbations

To quantify the duration of transient polymorphism, we examined how the escape time depends on the magnitude of an initial perturbation along the unstable eigenspace. Trajectories were initialised as

$$\mathbf{p}(0) = \mathbf{p}^* + \varepsilon \boldsymbol{\eta}_u, \quad \boldsymbol{\eta}_u \in E^u, \quad \|\boldsymbol{\eta}_u\| = 1,$$

where ε denotes the perturbation magnitude. The values of ε were logarithmically spaced from 10^{-5} to $10^{-1.5}$. All resulting initial conditions remained within the biologically admissible state space $(0, 1)^n$, so no amplitude rescaling was required.

Near the interior equilibrium, the linearized dynamics predict

$$\|\mathbf{u}(t)\| \approx \|\mathbf{u}(0)\| e^{\lambda_u t}.$$

If escape from a local neighbourhood is defined by

$$\|\mathbf{u}(t_{\text{esc}})\| = \delta_{\text{esc}},$$

then, to leading order as $\varepsilon \rightarrow 0$,

$$t_{\text{esc}} \approx \frac{1}{\lambda_u} \log \frac{\delta_{\text{esc}}}{\|\mathbf{u}(0)\|} = \tau_u \log \frac{\delta_{\text{esc}}}{\|\mathbf{u}(0)\|}.$$

Because $\|\mathbf{u}(0)\| = \varepsilon$, this gives

$$\frac{t_{\text{esc}}}{\tau_u} \approx \log \delta_{\text{esc}} + \log(1/\varepsilon).$$

For both escape criteria described below, we plotted the normalised escape time t_{esc}/τ_u against $\log(1/\varepsilon)$. The slope-one prediction applies directly to the unstable-amplitude criterion and provides a reference for the heterozygosity-based criterion.

We tested this prediction using two complementary definitions of escape. For the unstable-amplitude criterion, escape was recorded when

$$\|\mathbf{u}(t_{\text{esc}})\| = \delta_{\text{esc}}, \quad \delta_{\text{esc}} = 0.08.$$

This operational threshold exceeds all initial perturbation amplitudes while remaining well within the interior of the state space.

For the heterozygosity criterion, escape was defined as the first time at which

$$\frac{H(t_{\text{esc}})}{H(0)} \leq 0.5.$$

The first criterion directly measures departure along the unstable eigenspace, whereas the second records a later, population-level reduction in genetic variation.

Under the unstable-amplitude criterion, the numerical results closely follow this predicted linear scaling (Fig. 2a). Smaller initial perturbations therefore remain near the interior equilibrium for longer, consistent with the growth rate determined by the unstable eigenvalue.

The heterozygosity-based escape times exhibit the same approximate linear dependence, with a fitted slope also close to one (Fig. 2b). Because this criterion detects a later and more global stage of polymorphism loss, its absolute escape times need not coincide with those obtained from the local unstable-amplitude criterion. Nevertheless, the common scaling indicates that the persistence of heterozygosity is governed primarily by the slow amplification of perturbations along the unstable eigenspace.

The two criteria therefore describe different stages of the same transient process: the unstable-amplitude criterion tests the local saddle dynamics, whereas the heterozygosity criterion captures their eventual effect on genetic variation. Together, they show that the duration of transient polymorphism is controlled by the unstable time scale τ_u and the initial distance from the interior equilibrium.



(a) Escape time based on unstable-mode amplitude.



(b) Escape time based on heterozygosity loss.

Figure 2: Escape-time scaling under unstable perturbations. The equilibrium was perturbed along a fixed direction $\eta_u \in \ker(\mathbf{1}^\top)$ with varying amplitude ϵ . The left panel defines escape by the unstable-mode norm, whereas the right panel defines escape by the heterozygosity criterion $H(t_{\text{esc}})/H(0) \leq 0.5$. Both panels plot the normalised escape time t_{esc}/τ_u against $\log(1/\epsilon)$ and show an approximately linear relationship, supporting the prediction that the transient lifetime is controlled by the slow unstable time scale τ_u . The experiments use the same model parameters as Fig. 1.

3.3. Genetic drift and the limits of transient polymorphism

The present analysis concerns the deterministic infinite-population limit and therefore excludes genetic drift. A simple finite-population extension, under a Wright–Fisher diffusion approximation and the assumption of linkage equilibrium, would take the form

$$dp_i = f_i(\mathbf{p}) dt + \sqrt{\frac{p_i(1-p_i)}{2N_e}} dW_i(t), \quad i = 1, \dots, n,$$

where $f_i(\mathbf{p})$ denotes the deterministic allele-frequency dynamics defined above, N_e is the effective population size, and the W_i are independent Wiener processes. We do not analyse this stochastic system here; it is introduced only to clarify how drift would modify the deterministic picture.

The eigenspace decomposition shows that stabilising selection would not oppose drift equally in all directions. Drift-generated displacements along $E^s = \text{span}\{\mathbf{1}\}$ change the population mean and are opposed by selection at the rate $|\lambda_s| = 1/\tau_s$. By contrast, displacements in $E^u = \ker(\mathbf{1}^\top)$ redistribute allele frequencies among loci and are not restored toward the interior equilibrium. Instead, their deterministic components are amplified at the rate $\lambda_u = 1/\tau_u$. Thus, stabilising selection can restore the population mean after stochastic perturbations, but it does not stabilise the underlying multilocus allele-frequency configuration. Even a population initially located at $\mathbf{p}^*$ would be continually displaced by genetic drift, including along the unstable directions.

The relevance of the two deterministic time scales can be assessed by comparing them with the characteristic drift

time, which is of order N_e generations, up to convention-dependent constants. Equivalently, the relative strengths of deterministic selection and drift along the two eigendirections may be characterised by

$$2N_e|\lambda_s| = \frac{2N_e}{\tau_s}, \quad 2N_e\lambda_u = \frac{2N_e}{\tau_u}.$$

Because $|\lambda_s| = (2n-1)\lambda_u$, there may be a broad range of population sizes for which mean-shifting fluctuations are rapidly restored while locus-redistributing fluctuations remain strongly affected by drift. In particular, when

$$2N_e|\lambda_s| \gg 1 \quad \text{but} \quad 2N_e\lambda_u \lesssim 1,$$

the population mean may remain close to μ^* , even though drift substantially alters allele frequencies and erodes genetic variation. If N_e is much smaller than τ_u , polymorphism may therefore be lost through drift before deterministic growth along the unstable directions becomes appreciable. When N_e and τ_u are comparable, drift and deterministic amplification are expected to act on similar time scales, whereas for $N_e \gg \tau_u$, the deterministic time-scale separation should be more clearly expressed.

In the absence of mutation, $p_i = 0$ and $p_i = 1$ are absorbing states in the finite-population model. Drift would therefore turn the deterministic escape trajectory into a distribution of stochastic trajectories and escape times, with fixation or loss occurring in finite populations. As a neutral reference, expected heterozygosity decays approximately as $\exp[-t/(2N_e)]$, although this expression is not an exact prediction for the present multilocus system under selection.

These considerations clarify the scope of our conclusions. The slow deterministic loss of polymorphism demonstrated here does not constitute a mechanism that actively counteracts genetic drift and therefore does not establish long-term maintenance of polymorphism in finite populations. Rather, the spectral separation explains how the population mean can be rapidly restored while the underlying genetic variation remains only transient. A quantitative treatment of the interaction between drift and the saddle dynamics would require analysis or simulation of the stochastic system and lies beyond the scope of the present study [14].

4. Discussion

The central finding of this work is the presence of a *saddle-point* structure at the fully polymorphic equilibrium $\mathbf{p}^*$ whenever the feasibility condition holds in a polygenic allele-frequency dynamical system under pure stabilising selection. This equilibrium has one strongly stable direction, corresponding to collective shifts of allele frequencies that change the population mean μ and restore it toward the equilibrium mean μ^* , and $n - 1$ weakly unstable directions, corresponding to redistributions of allele frequencies among loci. Under equal allelic effects, the ratio between the stable and unstable rates is

$$\frac{|\lambda_s|}{\lambda_u} = 2n - 1.$$

Thus, for $n = 100$, the magnitude of the local stable contraction rate is 199 times the common unstable expansion rate. This is a relative rate comparison only; the absolute unstable time is $\tau_u = 1/[sa^2p^*(1 - p^*)]$, and the local residence time also depends on the initial unstable displacement.

The same decomposition gives a precise connection to adaptation after a shift of the optimum^{15,16}. If θ changes to $\theta + \Delta\theta$ and both equilibrium candidates remain feasible, then

$$\mathbf{p}^*(\theta + \Delta\theta) - \mathbf{p}^*(\theta) = \frac{\Delta\theta}{2n - 1} \left(\frac{1}{a_1}, \dots, \frac{1}{a_n} \right)^\top.$$

For equal effects this displacement lies in $E^s = \text{span}\{\mathbf{1}\}$, so a pure optimum shift excites the fast mean-restoring mode at linear order. Slow allele-frequency redistribution corresponds to any accompanying projection onto the redistributive modes; in the present mutation-free model those modes are unstable, so the comparison is geometric rather than a claim of equilibration at a stable polymorphic state.

We emphasise that this equilibrium does not provide permanent maintenance of polymorphism. A nonzero unstable component causes departure from a sufficiently small neighbourhood of the fully polymorphic saddle. The present local analysis does not classify the global limiting state and does not prove that every locus ultimately fixes or is lost. The contribution of the present framework is therefore not to claim that stabilising selection alone maintains a stable polymorphic equilibrium. Rather, it reveals a geometric mechanism by which pure stabilising selection can generate

a prolonged transient regime: deviations of the population mean are rapidly corrected, whereas redistributive perturbations among loci grow only slowly. As a result, the population mean can remain close to μ^* during a potentially long deterministic finite-time interval. This conclusion does not establish finite-population maintenance under genetic drift.

This perspective also clarifies the relationship between the present model and the classical analysis of Barton (1986). A key difference lies in the assignment of phenotypic values to genotypes. In Barton's formulation, for a diploid locus with alleles B/b , the three genotypes BB , Bb , and bb are assigned symmetric phenotypic effects $-a$, 0 , and a , respectively. In the present formulation, the corresponding effects are written in an uncentred form as 0 , a , and $2a$, with an additional parameter m representing the baseline phenotypic value. Accordingly, the population mean is written as $\mu = m + 2 \sum_i a_i p_i$.

The two descriptions are related by a shift of the phenotypic origin. Let

$$A_0 := \sum_{i=1}^n a_i, \quad \tilde{\mu} := \mu - (m + A_0), \quad \tilde{\theta} := \theta - (m + A_0).$$

Then $\mu - \theta = \tilde{\mu} - \tilde{\theta}$, so the selection dynamics are invariant under this simultaneous change of origin. In the equal-effect case, $A_0 = na$ and

$$\begin{aligned} p^* &= \frac{1}{2} + \frac{\tilde{\theta}}{a(2n - 1)} \\ &= \frac{1}{2} + \frac{\theta - m - na}{a(2n - 1)} \\ &= -\frac{1}{4n - 2} + \frac{\theta - m}{a(2n - 1)}. \end{aligned}$$

The corresponding equilibrium mean is

$$\mu^* = \theta + \frac{\theta - m - na}{2n - 1}.$$

Thus m and θ enter the uncentred dynamics only through $\theta - m$; m does not introduce an additional independent dynamical degree of freedom. Whenever the feasibility condition holds, this equilibrium is a saddle rather than an asymptotically stable state.

The present work focuses on Gaussian stabilising selection, for which the fitness function has the standard form determined by the squared deviation between the population phenotype and the optimum. An important open question is whether the same saddle geometry and time-scale separation persist under other mathematical forms of stabilising selection. If the fitness function is no longer an exponential function of the squared deviation, the allele-frequency dynamics may no longer take the same simple form. In that case, one must rederive the within-generation change in allele frequency and determine whether the Jacobian at the interior equilibrium still decomposes into a fast mean-restoring direction and slow redistributive directions. This

remains an interesting direction for further theoretical analysis.

Several extensions of the model are also natural. First, mutation could be introduced to examine how mutation–selection balance interacts with the transient saddle dynamics identified here. Mutation may convert the finite-time transient maintenance of polymorphism into a genuine long-term balance, depending on its rate and structure. Second, random genetic drift could be incorporated to study how stochastic fluctuations affect escape from the interior neighbourhood, especially in finite populations. Because t is measured in generations, a concise comparison with the drift scale is $\Gamma := 2N_e\lambda_u$, for which the deterministic estimate gives $t_{\text{esc}}/(2N_e) \simeq \Gamma^{-1} \log[\delta_{\text{esc}}/(\alpha\epsilon)]$. This comparison does not turn the interior saddle into a finite-population maintenance mechanism. Third, the current analysis assumes linkage equilibrium and additive locus effects. Relaxing these assumptions would allow the model to include linkage disequilibrium, recombination, dominance, and epistasis. These extensions would help determine how robust the time-scale separation mechanism is in more realistic genetic systems.

Overall, the results suggest that pure stabilising selection has a richer deterministic structure than is apparent from single-locus intuition. It does not create an asymptotically stable polymorphic equilibrium in the absence of mutation. Nevertheless, in a multilocus system, stabilising selection can rapidly restore the population mean while allowing only slow redistribution among loci. This can produce a prolonged local deterministic transient near the fully polymorphic saddle, but it does not establish stable or finite-population maintenance of polygenic variation.

Acknowledgements

This work is supported by Tsinghua University Education Foundation fund and National Natural Science Foundation of China (Grant No. 12201015).

References

- [1] Sewall Wright. “The analysis of variance and the correlations between relatives with respect to deviations from an optimum”. In: *Journal of Genetics* 30 (1935), pp. 243–256.
- [2] Alan Robertson. “The effect of selection against extreme deviants based on deviation or on homozygosis”. In: *Journal of Genetics* 54 (1956), pp. 236–248.
- [3] B. D. H. Latter. “Natural Selection for an Intermediate Optimum”. In: *Australian Journal of Biological Sciences* 13.1 (1960), pp. 30–35.
- [4] M. G. Bulmer. “The Genetic Variability of Polygenic Characters under Optimizing Selection, Mutation and Drift”. In: *Genetical Research* 19.1 (1972), pp. 17–25.
- [5] N. H. Barton. “The maintenance of polygenic variation through a balance between mutation and stabilizing selection”. In: *Genetical Research* 47.3 (1986), pp. 209–216.
- [6] Michael Turelli. “Heritable genetic variation via mutation-selection balance: Lerch’s zeta meets the abdominal bristle”. In: *Theoretical Population Biology* 25.2 (1984), pp. 138–193.
- [7] Ken-ichi Kojima. “STABLE EQUILIBRIA FOR THE OPTIMUM MODEL”. In: *Proceedings of the National Academy of Sciences* 45.7 (1959), pp. 989–993.
- [8] Thomas Nagylaki. “The maintenance of genetic variability in two-locus models of stabilizing selection”. In: *Genetics* 122.1 (1989), pp. 235–248.
- [9] Reinhard Bürger and Alexander Gimelfarb. “Genetic Variation Maintained in Multilocus Models of Additive Quantitative Traits Under Stabilizing Selection”. In: *Genetics* 152.2 (1999), pp. 807–820. ISSN: 1943-2631.
- [10] J. Gale and M. Kearsey. “Stable equilibria under stabilising selection in the absence of dominance”. In: *Heredity* 23 (1968), pp. 553–561.
- [11] Harold P. de Vladar and Nick Barton. “Stability and response of polygenic traits to stabilizing selection and mutation.” In: *Genetics* 197.2 (June 2014), pp. 749–767. ISSN: 1943-2631 0016-6731.
- [12] Sergey Gavrilets and Alan Hastings. “Maintenance of multilocus variability under strong stabilizing selection”. In: *Journal of Mathematical Biology* 32.4 (1994), pp. 287–302.
- [13] David Waxman. “A model of population genetics and its mathematical relation to quantum theory”. In: *Contemporary Physics* 43.1 (2002), pp. 13–20.
- [14] Harold P. de Vladar and Nicholas H. Barton. “The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift”. In: *Journal of the Royal Society Interface* 8.58 (2011), pp. 720–739.
- [15] Kevin R. Thornton. “Polygenic Adaptation to an Environmental Shift: Temporal Dynamics of Variation under Gaussian Stabilizing Selection and Additive Effects on a Single Trait”. In: *Genetics* 213.4 (2019), pp. 1513–1530.
- [16] Laura Katharine Hayward and Guy Sella. “Polygenic Adaptation after a Sudden Change in Environment”. In: *eLife* 11 (2022), e66697.

A. Proofs

Detailed derivation of the classical model

If $Z \sim \mathcal{N}(\mu, \sigma^2)$, Gaussian integration gives

$$\overline{W} = \frac{1}{\sqrt{1+s\sigma^2}} \exp\left[-\frac{s(\mu-\theta)^2}{2(1+s\sigma^2)}\right], \quad s = w^{-2}.$$

Consequently,

$$\log \overline{W} = -\frac{1}{2} \log(1+s\sigma^2) - \frac{s(\mu-\theta)^2}{2(1+s\sigma^2)}.$$

For bounded μ and σ^2 , expansion as $s \rightarrow 0$ yields

$$\begin{aligned} \log \overline{W} = & -\frac{s}{2} [\sigma^2 + (\mu - \theta)^2] \\ & + s^2 \left[\frac{\sigma^4}{4} + \frac{\sigma^2(\mu - \theta)^2}{2} \right] \\ & + O(s^3 \{ \sigma^6 + \sigma^4(\mu - \theta)^2 \}). \end{aligned}$$

The allele-frequency ODE analysed below retains only the leading $O(s)$ term, namely

$$\ell_0(\mathbf{p}) = -\frac{s}{2} [\sigma^2(\mathbf{p}) + \{\mu(\mathbf{p}) - \theta\}^2].$$

Because σ_E^2 is independent of the allele frequencies, it contributes an additive constant to ℓ_0 and disappears after differentiation with respect to p_i .

Under Hardy–Weinberg and linkage equilibrium, additive gene action, and no dominance or epistasis,

$$\mu = m+2 \sum_{i=1}^n a_i p_i, \quad \sigma^2 = 2 \sum_{i=1}^n a_i^2 p_i(1-p_i) + \sigma_E^2, \quad a_i > 0.$$

Thus the derivatives required for the leading-order log-mean fitness ℓ_0 are

$$\frac{\partial \mu}{\partial p_i} = 2a_i, \quad \frac{\partial \sigma^2}{\partial p_i} = 2a_i^2(1-2p_i).$$

$$\begin{aligned} \frac{\partial \ell_0}{\partial p_i} &= -\frac{s}{2} \frac{\partial [\sigma^2 + (\mu - \theta)^2]}{\partial p_i} \\ &= -\frac{s}{2} \left[\frac{\partial \sigma^2}{\partial p_i} + 2(\mu - \theta) \frac{\partial \mu}{\partial p_i} \right] \end{aligned}$$

Substituting these derivatives into the leading-order log-mean fitness ℓ_0 and then into Wright's formula gives the one-generation change

$$\begin{aligned} \Delta p_i &= \frac{p_i(1-p_i)}{2} \frac{\partial \ell_0}{\partial p_i} + O(s^2) \\ &= \frac{p_i(1-p_i)}{2} \left(-\frac{s}{2} \right) \left[\frac{\partial \sigma^2}{\partial p_i} + 2(\mu - \theta) \frac{\partial \mu}{\partial p_i} \right] + O(s^2) \\ &= -\frac{s}{2} \frac{p_i(1-p_i)}{2} [2a_i^2(1-2p_i) + 2(\mu - \theta)2a_i] + O(s^2) \\ &= \frac{1}{2} s a_i p_i(1-p_i) [a_i(2p_i-1) + 2(\theta - \mu)] + O(s^2) \end{aligned}$$

Proof of the Existence Theorem for Solutions in Mean-Field Model

The proof of Theorem 1 is given as follows.

Proof. The elements of matrix $\mathbf{A}$ are defined as:

$$\mathbf{A}_{ij} = \begin{cases} -2a_i & \text{if } i = j, \\ -4a_j & \text{if } i \neq j. \end{cases}$$

We decompose $\mathbf{A}$ into the sum of two matrices: $\mathbf{A} = \mathbf{D} + \mathbf{B}$, where:

$$\mathbf{D} = \text{diag}(2a_1, 2a_2, \dots, 2a_n), \quad (12)$$

$$\mathbf{B} = -4\mathbf{e}\mathbf{a}^T, \quad (13)$$

where $\mathbf{e} = (1, 1, \dots, 1)^T$ is the all-ones vector and $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$.

For diagonal elements: $\mathbf{A}_{ii} = \mathbf{D}_{ii} + \mathbf{B}_{ii} = 2a_i + (-4a_i) = -2a_i$.

For off-diagonal elements ($i \neq j$): $\mathbf{A}_{ij} = \mathbf{D}_{ij} + \mathbf{B}_{ij} = 0 + (-4a_j) = -4a_j$.

Since $\mathbf{B} = (-4)\mathbf{e}\mathbf{a}^T$ is a rank-one matrix, we apply the rank-one update formula.

Lemma 1 (Rank-One Update Formula). *Let $\mathbf{M}$ be an $n \times n$ invertible matrix, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then*

$$\det(\mathbf{M} + \mathbf{u}\mathbf{v}^T) = \det(\mathbf{M})(1 + \mathbf{v}^T \mathbf{M}^{-1} \mathbf{u}).$$

Proof of Lemma 1. By the matrix determinant identity, we have

$$\det(\mathbf{M} + \mathbf{u}\mathbf{v}^T) = \det(\mathbf{M}) \det(\mathbf{I} + \mathbf{M}^{-1} \mathbf{u}\mathbf{v}^T). \quad (14)$$

Let $\mathbf{w} = \mathbf{M}^{-1} \mathbf{u}$, then the eigenvalues of $\mathbf{I} + \mathbf{w}\mathbf{v}^T$ are 1 and $1 + \mathbf{v}^T \mathbf{w}$ (where 1 is an eigenvalue with multiplicity $(n-1)$). Therefore:

$$\det(\mathbf{I} + \mathbf{w}\mathbf{v}^T) = 1^{n-1} \cdot (1 + \mathbf{v}^T \mathbf{w}) = 1 + \mathbf{v}^T \mathbf{M}^{-1} \mathbf{u}.$$

Thus Lemma 1 holds. $\square$

Now applying this lemma with $\mathbf{M} = \mathbf{D}$, $\mathbf{u} = -4\mathbf{e}$, $\mathbf{v} = \mathbf{a}$:

First, compute $\det(\mathbf{D}) = \prod_{i=1}^n 2a_i = 2^n \prod_{i=1}^n a_i$.

Then compute $\mathbf{v}^T \mathbf{D}^{-1} \mathbf{u}$:

$$\mathbf{D}^{-1} \mathbf{u} = \mathbf{D}^{-1}(-4\mathbf{e}) = -4 \left(\frac{1}{2a_1}, \frac{1}{2a_2}, \dots, \frac{1}{2a_n} \right)^T \quad (15)$$

$$= \left(\frac{-2}{a_1}, \frac{-2}{a_2}, \dots, \frac{-2}{a_n} \right)^T, \quad (16)$$

$$\mathbf{v}^T \mathbf{D}^{-1} \mathbf{u} = \mathbf{a}^T \left(\frac{-2}{a_1}, \frac{-2}{a_2}, \dots, \frac{-2}{a_n} \right)^T = \sum_{i=1}^n a_i \cdot \frac{-2}{a_i} = -2n. \quad (17)$$

Therefore:

$$\det(\mathbf{A}) = \det(\mathbf{D})(1 + \mathbf{v}^T \mathbf{D}^{-1} \mathbf{u}) = 2^n \prod_{i=1}^n a_i \cdot (1 - 2n) = 2^n(1 - 2n) \prod_{i=1}^n a_i.$$

When $n \geq 2$ and $a_i \neq 0$, we have $(1 - 2n) < 0$ and $\prod_{i=1}^n a_i \neq 0$, thus $\det(\mathbf{A}) \neq 0$. Therefore, matrix $\mathbf{A}$ is invertible and the system $\mathbf{A}\mathbf{p} = \mathbf{h}$ has a unique solution. $\square$

The proof of Theorem 2 is given as follows.

Proof. Deriving the Explicit Expression

Since $\mathbf{A} = \mathbf{D} + \mathbf{B}$, where $\mathbf{D} = \text{diag}(2a_1, \dots, 2a_n)$ and $\mathbf{B} = -4\mathbf{e}\mathbf{a}^T$, we use the Sherman-Morrison formula to compute the inverse: Since $\mathbf{B} = \mathbf{u}\mathbf{v}^T$ with $\mathbf{u} = -4\mathbf{e}$ and $\mathbf{v} = \mathbf{a}$, the Sherman-Morrison formula gives

$$\mathbf{A}^{-1} = (\mathbf{D} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{D}^{-1} - \frac{(\mathbf{D}^{-1}\mathbf{u})(\mathbf{v}^T\mathbf{D}^{-1})}{1 + \mathbf{v}^T\mathbf{D}^{-1}\mathbf{u}}.$$

Computing the scalar denominator:

$$\mathbf{v}^T\mathbf{D}^{-1}\mathbf{u} = \mathbf{a}^T \text{diag}\left(\frac{1}{2a_1}, \dots, \frac{1}{2a_n}\right)(-4\mathbf{e}) = -4 \sum_{i=1}^n \frac{a_i}{2a_i} = -2n.$$

Computing each term:

$$\mathbf{D}^{-1} = \text{diag}\left(\frac{1}{2a_1}, \dots, \frac{1}{2a_n}\right), \quad (18)$$

$$\mathbf{B}\mathbf{D}^{-1} = -4\mathbf{e}\mathbf{a}^T\mathbf{D}^{-1} = -4\mathbf{e}\left(\frac{1}{2}, \dots, \frac{1}{2}\right) = -2\mathbf{e}\mathbf{e}^T, \quad (19)$$

$$\text{tr}(\mathbf{B}\mathbf{D}^{-1}) = \mathbf{a}^T\mathbf{D}^{-1}(-4\mathbf{e}) = -4 \sum_{j=1}^n \frac{a_j}{2a_j} = -2n. \quad (20)$$

Therefore:

$$\mathbf{A}^{-1} = \mathbf{D}^{-1} - \frac{-2\mathbf{D}^{-1}\mathbf{e}\mathbf{e}^T}{1 - 2n} = \mathbf{D}^{-1} + \frac{2\mathbf{D}^{-1}\mathbf{e}\mathbf{e}^T}{1 - 2n}.$$

That is, the elements of $\mathbf{A}^{-1}$ are:

$$(\mathbf{A}^{-1})_{ij} = \begin{cases} \frac{1}{2a_i} - \frac{2}{(2n-1) \cdot 2a_i} = \frac{1}{2a_i} \left(1 - \frac{2}{2n-1}\right) = \frac{2n-3}{2a_i(2n-1)} & \text{if } i = j, \\ -\frac{2}{(2n-1) \cdot 2a_i} = -\frac{1}{a_i(2n-1)} & \text{if } i \neq j. \end{cases}$$

The solution is $\mathbf{p} = \mathbf{A}^{-1}\mathbf{h}$, where $h_i = a_i - 2(\theta - m)$:

$$p_i = \sum_{j=1}^n (\mathbf{A}^{-1})_{ij} h_j \quad (21)$$

$$= \frac{2n-3}{2a_i(2n-1)} h_i - \sum_{j \neq i} \frac{1}{a_i(2n-1)} h_j \quad (22)$$

$$= \frac{1}{2a_i(2n-1)} \left[(2n-3)h_i - 2 \sum_{j \neq i} h_j \right] \quad (23)$$

$$= \frac{1}{2a_i(2n-1)} \left[(2n-3)a_i - 2 \sum_{j \neq i} a_j + 2(\theta - m) \right] \quad (24)$$

$$= \frac{(2n-3)a_i - 2 \sum_{j \neq i} a_j}{2a_i(2n-1)} + \frac{\theta - m}{a_i(2n-1)} \quad (25)$$

$$= \frac{(2n-3)a_i - 2(\sum_{j=1}^n a_j - a_i)}{2a_i(2n-1)} + \frac{\theta - m}{a_i(2n-1)} \quad (26)$$

$$= \frac{(2n-1)a_i - 2 \sum_{j=1}^n a_j}{2a_i(2n-1)} + \frac{\theta - m}{a_i(2n-1)} \quad (27)$$

$$= \frac{1}{2} - \frac{\sum_{j=1}^n a_j}{a_i(2n-1)} + \frac{\theta - m}{a_i(2n-1)}. \quad (28)$$

Simplifying, we obtain:

$$p_i = \frac{1}{2} - \frac{\sum_{j=1}^n a_j}{a_i(2n-1)} + \frac{\theta - m}{a_i(2n-1)}.$$

Finally, because $a_i > 0$,

$$\begin{aligned} 0 < p_i^* < 1 &\iff -\frac{1}{2} < \frac{\theta - m - \sum_{j=1}^n a_j}{a_i(2n-1)} < \frac{1}{2} \\ &\iff \sum_{j=1}^n a_j - \frac{2n-1}{2}a_i < \theta - m < \sum_{j=1}^n a_j + \frac{2n-1}{2}a_i. \end{aligned}$$

Requiring these inequalities for every i gives the necessary and sufficient feasibility condition stated in Theorem 2. $\square$