

## Adaptive dynamics for physiologically structured population models

Michel Durinx · J. A. J. (Hans) Metz · Géza Meszéna

Received: 25 April 2006 / Revised: 1 May 2007 / Published online: 18 October 2007  
© Springer-Verlag 2007

**Abstract** We develop a systematic toolbox for analyzing the adaptive dynamics of multidimensional traits in physiologically structured population models with point equilibria (sensu Dieckmann et al. in *Theor. Popul. Biol.* 63:309–338, 2003). Firstly, we show how the canonical equation of adaptive dynamics (Dieckmann and Law in *J. Math. Biol.* 34:579–612, 1996), an approximation for the rate of evolutionary change in characters under directional selection, can be extended so as to apply to general physiologically structured population models with multiple birth states. Secondly, we show that the invasion fitness function (up to and including second order terms, in the distances of the trait vectors to the singularity) for a community of  $N$  coexisting types near an evolutionarily singular point has a rational form, which is model-independent in the following sense: the form depends on the strategies of the residents and the invader, and on the second order partial derivatives of the one-resident fitness function at the singular point. This normal form holds for Lotka–Volterra models as well as for physiologically structured population models with multiple birth states, in discrete as well as continuous time and can thus be considered universal for the evolutionary dynamics in the neighbourhood of singular points. Only in the case of one-dimensional trait spaces or when  $N = 1$  can the normal form be reduced to a Taylor polynomial. Lastly we show, in the form of a stylized recipe, how these results can be combined

---

M. Durinx (✉) · J. A. J. (Hans) Metz  
Institute of Biology, Leiden University, P. O. Box 9516, 2300 RA Leiden, The Netherlands  
e-mail: michel.durinx@gmail.com

J. A. J. (Hans) Metz  
Evolution and Ecology Program, International Institute for Applied Systems Analysis,  
2361 Laxenburg, Austria

G. Meszéna  
Department of Biological Physics, Eötvös University,  
Pázmány Péter sétány 1A, 1117 Budapest, Hungary

into a systematic approach for the analysis of the (large) class of evolutionary models that satisfy the above restrictions.

**Keywords** Adaptive dynamics · Physiologically structured populations · Multivariate evolutionarily singular strategies · Multitype branching processes · Evolutionary modelling

**Mathematics Subject Classification (2000)** 70K45 · 92D15 · 92D25

## 1 Introduction

This paper is concerned with the abstract geometry underlying the process of repeated invasions by novel mutants. Mutation limited near-continuous evolution will be our frame of reference, as we follow the so-called adaptive dynamics approach. Adaptive dynamics studies which rare mutants can establish themselves in an environment inhabited by a large equilibrium population of residents that they closely resemble, which invasions by similar mutants will lead to the demise of the original residents, and what the evolutionary outcome will be of a series of such substitution events. The tricks and tools of this trade are introduced in the following section.

The assumed magnitude of the resident population makes its dynamics deterministic, whereas the rarity of the invading mutant introduces a strong stochastic effect. This complication means that a positive average growth rate is a necessity, but no guarantee for a mutant's invasion success. To ask for the probability of such success is basically to ask what chance a given mutant has of being the ancestor of an unbroken line of descendants. This is analogous to the "surname" problem that led to the theory of branching processes, where the quantity we called for is termed the *establishment probability* of the given mutant (e.g., [32]).

The last major consideration we have in the setup of this enquiry is that we look for general geometric properties and not artifacts generated by specific models. Therefore we must consider as wide a class of models as we can technically handle. To that end, we derive our results within the context of general physiologically structured populations. This class of models is the ultimate generalization of resource competition models, allowing populations structured, e.g., by size, and multiple birth states (think sexes, morphs, or size at birth). The third part of this introduction (1.2) points out the main assumptions and quantities pertaining to such models.

Gathering together the results of perturbation calculations, we are able to extend the so-called canonical equation derived by Dieckmann and Law [15] to general physiologically structured populations. It is *the* adaptive dynamics tool, describing the rate of trait change in the case of directional selection. However, the canonical equation is an approximation that loses its validity in the close proximity of its equilibrium points. At such points, called evolutionarily singular points, a more precise analysis is required.

In this paper we also show that with regard to the invasion fitness function near evolutionarily singular points, all possible models are locally equivalent to Lotka–Volterra models (3.4, Proposition 3). Therefore the fitness function of these well-

known and mathematically relatively tractable models provides a general normal form. Thus the derivation of this property is a step towards classifying the local geometrical properties of invasion functions. Geritz et al. [27] showed that if the trait under evolutionary control is scalar, a full classification of nonexceptional cases consists of eight possibilities. When traits are multidimensional (as in this paper), it is unknown how many classes are needed to cover all nonexceptional cases nor what they would look like, let alone that there is an understanding of the bifurcations between those classes. F.J.A. Jacobs is engaged (together with one of the authors) in analyzing the latter for Lotka–Volterra models with scalar traits; this paper shows that a fair part of his results apply to all models with one-dimensional strategies.

### 1.1 Adaptive dynamics

Adaptive dynamics is concerned with evolutionary outcomes of community-dynamical processes when reproduction is nearly faithful [49]. The main assumptions are rarity of mutations (i.e., the ecological and evolutionary timescales are separated, and hence the community dynamics will settle on an attractor between mutation events), smallness of mutational steps (allowing sensible topological and geometrical inferences) and the initial rareness of mutants (implying a well-mixed resident population of large size).

A key insight of structured population models is given pride of place in adaptive dynamics: the separation of individual and environment, both influencing each other in a feedback loop [48]. The idea is that individuals influence the environment in an additive manner. Given an environment, individuals are independent—any two particular individuals being exceedingly rare as a proportion of the total population, their mutual influence is effectively zero. This decoupling makes the equations linear when the environmental condition is given as a function of time.

The starting point of adaptive dynamics is the *invasion fitness function* [50]. By definition this is the long-term average per capita growth rate of a rare type (the *invader*) in an equilibrium community of a given set of types (the *residents*). Thus a resident type cast in the role of invader always has a zero invasion fitness, since it will on average neither grow nor diminish in abundance. One also sees that a negative fitness for a given type implies the impossibility for such an invader to gain a foothold in the population, whereas a positive fitness means a positive probability of establishment. But as this concerns a stochastic process with an initially very small amount of invaders, even a positive average growth rate will not prevent extinction in a fair amount of cases. However, as we consider gradual, mutation-driven evolution, the relevant invaders are the *mutants*: new types that differ but slightly from one of the residents. When a mutant has a positive invasion fitness, but due to stochasticity its attempt at establishment fails, this is not the end; evolution can bide its time and a later occurring similar mutation may get established due to other chance fluctuations.

Reviewing the technical setup of the framework, we start by considering the parameters under evolutionary control. We refer to this set of parameters as a *strategy* (which gives it a life history flavour), a *trait value* or *trait vector* (which sets the

mind to a more technical, algebraic frame), a *point* in the strategy space (which hints at a graphical representation, or a geometrical argument), or simply the *type* of the individual. We call the set of all possible traits the *trait space* and denote it by  $\mathfrak{X}$ .

The invasion fitness function is also known as the *s*-function, to underline its heritage as a conceptual extension of the selection coefficient of population genetics. The *s*-function for a *monomorphic* community, denoted by  $s_X(\mathbf{Y})$ , describes the invasion fitness of a mutant with trait value  $\mathbf{Y}$  in an environment set by a single resident of type  $\mathbf{X}$ . The *s*-function for a *polymorphic* community, similarly denoted by  $s_{\mathfrak{X}}(\mathbf{Y})$ , gives the invasion fitness of a  $\mathbf{Y}$ -type mutant in an environment set by a community of  $N$  types  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\} =: \mathfrak{X}$ .

That the community can (locally) be identified with the strategies present, comes from the convenient assumption of existence and (local) uniqueness of an attractor for the population dynamics of the community, plus the paucity and small effect of mutations.

The *s*-function generates further functions of central concern, namely the *invasion gradients*, which are the transposes of the derivatives of the fitness in the mutant direction at the trait value of a resident:  $\left(\frac{\partial s_X(\mathbf{Y})}{\partial \mathbf{Y}}\Big|_{\mathbf{Y}=\mathbf{X}}\right)^\top$  for a monomorphic world,  $\left(\frac{\partial s_{\mathfrak{X}}(\mathbf{Y})}{\partial \mathbf{Y}}\Big|_{\mathbf{Y}=\mathbf{X}_i}\right)^\top$  for some  $i$  in the polymorphic case. The trait values where these invasion gradients are zero are called *evolutionarily singular strategies*. The study of evolutionary dynamics can thereby be split into two main parts. First, away from the zeros of the invasion gradient and under the restriction of well-behaved population dynamics, it can be shown that “*invasion implies substitution*” [13, 14]. What well-behaved entails, is considered by Geritz et al. [25, 26], and substitution means that the mutant drives its ancestral resident to extinction if it succeeds in establishing itself. Hence the apparition of a new type, the mutant, does not usually lead to increased diversity—on the contrary, if there are several types coexisting in the resident community, on rare occasions the appearance of a mutant may lead to the demise of not only the resident that spawned it but also of other resident types, thereby actually reducing the diversity of resident types. Close to a singular strategy however, other phenomena come into play. Singularities fall into several categories, one possibility being the classical *ESS*, known i.a. from evolutionary game theory. What makes adaptive dynamics an interesting evolutionary framework, is the existence of other, naturally occurring, types of singularities. Foremost among them is the *branching point*, a singularity that is attracting (for the monomorphic dynamics) but in the proximity of which selection is disruptive. Here selection acts such that a newly established mutant does not drive its progenitor to kingdom come. Subsequent mutants do however wipe out their ancestors, so that after a few mutation events two distinct resident populations will sit on opposite sides of the singularity. Over evolutionary time, these populations form two “branches” of co-viable types, that evolve away from the singularity. Such a splitting of genetic lines through an intrinsic process has an obvious appeal as a model for (the initiation of) speciation.

Research into the mathematical properties of adaptive dynamics models has led to several insights. Foremost there is the canonical equation as formulated by Dieckmann and Law [15], which predicts the speed of evolution as a function of the underlying

individual processes. That formulation so far allows only community dynamics modelled by ODEs. The equation basically predicts evolution under directional selection, at some distance from singularities. In this paper we first extend the applicability of the canonical equation to physiologically structured populations, and later look what happens at those points where the approximation fails to hold true. To this end we devise an expansion near the singular points of the fitness function. The formalism in which we do the calculations is set down in the following subsection.

## 1.2 Physiologically structured population models

As described for example by Diekmann et al. [21], general physiologically structured models assume few restrictions on population dynamical mechanisms other than (local) well-mixedness. We restrict our attention to the special case of structured populations with point equilibria in the resident population dynamics. In that case, the following definitions shape the modelling framework:

- $\mathbf{b}$  is the column vector of *birth rates*, with as components the steady rates at which individuals are born with state-at-birth specified by the component number.
- $\mathbf{I}$  is a vector describing the *environmental conditions* as far as they play a role in the (direct or indirect) interactions between the individuals. The defining requirement is that individuals are independent of one another when  $\mathbf{I}$  is given. In this paper, we restrict our attention to community dynamics with point equilibria, so  $\mathbf{I}$  is time-independent.
- $\mathbf{L}(\mathbf{X}, \mathbf{I})$  is the *next-generation matrix*. The matrix component  $L(\mathbf{X}, \mathbf{I})_{lm}$  is the expected number of offspring with birth state  $l$  born over the lifetime of an individual with trait vector  $\mathbf{X}$  that was born with state  $m$ , given steady environmental conditions as specified by  $\mathbf{I}$ .
- $\mathbf{G}(\mathbf{X}, \mathbf{I})$  is the *feedback matrix*. The matrix component  $G(\mathbf{X}, \mathbf{I})_{tl}$  is the lifetime contribution to the  $t$ th component of  $\mathbf{I}$  by an individual born in state  $l$  with trait vector  $\mathbf{X}$ , given steady environmental conditions as specified by  $\mathbf{I}$ .

The terminology above implies that we are only considering a finite number of possible birth states and of environmental dimensions, although there are no conceptual reasons for this restriction. For example, single celled organisms will inherit their size from their mother (about half her size at the time of division), which implies a continuous range of sizes for the newborns. Similarly, sexual reproduction leads to infinite dimensional environments usually, because each trait can potentially partner with infinitely many other traits to make up a diploid individual.

We restrict ourselves to finite dimensional environments and birth flows, to make sure that our formal calculations make mathematical sense; there is no a priori reason why a generalization would not be possible or desirable. [For modelling work without these limitations, see e.g. Diekmann and Gyllenberg (submitted 2007), Abstract delay equations inspired by population dynamics].

For a community under the above conditions with  $N$  types present, equilibrium means that each generation precisely replaces the previous generation, and that the

feedback is such that it exactly re-creates the environment as experienced by the organisms:

$$\begin{cases} \mathbf{b}_i = \mathbf{L}(X_i, \mathbf{I}) \mathbf{b}_i & (\forall i \in \{1, 2, \dots, N\}) \\ \mathbf{I} = \sum_{j=1}^N \mathbf{G}(X_j, \mathbf{I}) \mathbf{b}_j \end{cases} \quad (1)$$

It is clear that the first equation is equivalent to stating that at equilibrium, a population is either extinct or the expected lifetime offspring production  $R_0$  of its individuals is one, since  $R_0$  is the dominant eigenvalue of  $\mathbf{L}$ . The second equation states nothing more than that all individuals together must contribute to the environment in such a way that it remains unchanged. Diekmann et al. [21] have shown that the equilibrium conditions of most population models in the literature may be cast in the above form (1), a claim hinging on the considerations below about uncoupling the feedback loop that connects populations and individuals. It will however often be an arduous task to rewrite a given model representation into this form while the individual-based recipe for arriving at Eq. (1) is easy.

It should be stressed that Eq. (1) is an equilibrium equation, written in terms of the next-generation operator  $\mathbf{L}$  together with the feedback operator  $\mathbf{G}$ . Discrete time non-overlapping generations models are typically specified by giving matrix valued functions  $\mathbf{L}$  and  $\mathbf{G}$  for all possible environmental conditions, including non-equilibrium population states. Then Eq. (1) is immediately found as the corresponding equilibrium condition. For continuous time models, Diekmann et al. [19,20] have shown how an extension into nonequilibrium conditions can be done through reformulating the dynamics using an integral kernel formulation, which can be a challenging task in concrete cases.

From a biological point of view, the environment  $\mathbf{I}$  is more readily observed as the effect of the community on the world (the *environmental output*  $\mathbf{I}_{\text{out}}$ ) than vice versa (the *environmental input*  $\mathbf{I}_{\text{in}}$ ), as the rest of this subsection will elaborate.

The idea behind physiologically structured population models as put forward by Diekmann et al. [20,21], is to characterize the populations by their birth flow vectors; that is, we register the flux of births  $\mathbf{b}_i$  of the  $i$ th population differentiated according to the possible birth states. The per capita lifetime offspring production depends on the condition of the world,  $\mathbf{I}_{\text{in}}$ , and on the type  $X_i$  of the individual, so that in the special case where the world is constant, a given cohort  $\mathbf{b}_i$  produces  $\mathbf{L}(X_i, \mathbf{I}_{\text{in}})\mathbf{b}_i$  offspring over its lifetime, for some matrix function  $\mathbf{L}$ .

The output  $\mathbf{I}_{\text{out}}$  registers the total influence the individuals have on the environment. This clearly depends on the state of the community; for example, an individual in a virgin (i.e., devoid of competitors) environment may consume more and have far more offspring than an identically born individual that is put in an overcrowded world. It is also clear that this output should scale with the number of individuals there are, as it is an instantaneous output: two individuals will have exactly twice the influence of a single individual if they are kept under exactly the same conditions. Furthermore, this influence depends on the type of the individuals concerned. Therefore we postulate that the output must depend on the input in the following way that accounts for the scaling argument:  $\mathbf{I}_{\text{out}} = \sum_j \mathbf{G}(X_j, \mathbf{I}_{\text{in}})\mathbf{b}_j$ .

All other things being equal, the state of the world must be the result of the compounded influence of all the individuals. Thus the condition  $\mathbf{I}_{\text{in}}$  depends only on the output  $\mathbf{I}_{\text{out}}$  of the population, through some conversion function  $\mathbf{F}$  that accounts for the effect of the environmental dynamics. Hence the feedback loop of the community's influence on itself is closed.

All told, we have the following system to solve, where the last equation is the equilibrium condition:

$$\mathbf{I}_{\text{out}} = \sum_j \mathbf{G}(X_j, \mathbf{I}_{\text{in}}) \mathbf{b}_j \quad \mathbf{I}_{\text{in}} = \mathbf{F}(\mathbf{I}_{\text{out}}) \quad \forall i: \mathbf{b}_i = \mathbf{L}(X_i, \mathbf{I}_{\text{in}}) \mathbf{b}_i$$

Here we see that we can eliminate one equation and have only  $\mathbf{I}_{\text{out}}$  and  $\mathbf{b}$  as unknowns, since

$$\mathbf{I}_{\text{out}} = \sum_j \tilde{\mathbf{G}}(X_j, \mathbf{I}_{\text{out}}) \mathbf{b}_j \quad \forall i: \mathbf{b}_i = \tilde{\mathbf{L}}(X_i, \mathbf{I}_{\text{out}}) \mathbf{b}_i$$

where the matrix functions  $\tilde{\mathbf{G}}$  and  $\tilde{\mathbf{L}}$  are the compositions  $\mathbf{G} \circ (\text{id} \times \mathbf{F})$  and  $\mathbf{L} \circ (\text{id} \times \mathbf{F})$  respectively. We will denote  $\mathbf{I}_{\text{out}}$  simply as  $\mathbf{I}$  and drop the tilde in the notation of  $\mathbf{G}$  and  $\mathbf{L}$ , which gives us the equations introduced at the beginning of this subsection. It is clear that an arbitrarily complicated amount of biological detail can be put in the functions  $\mathbf{G}$  and  $\mathbf{L}$ , justifying the claim that this is a very flexible modelling framework. We do however assume a certain level of smoothness (namely that  $\mathbf{G}$  and  $\mathbf{L}$  are thrice continuously Fréchet differentiable functions), to guarantee the existence of chain rules and to justify our expansion arguments by the implicit function theorem.

### 1.3 Notations

Throughout this paper, we will deal with communities where a finite number of types are present. These are numbered from 1 to  $N$  and denoted by their respective trait vectors  $\mathbf{X}_1$  up to  $\mathbf{X}_N$ . The community as a whole is denoted by  $\mathbb{X}$  and it is interpreted either as a set of trait vectors  $\mathbb{X} := \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$ , or as an  $N$ -column matrix  $\mathbb{X} := [\mathbf{X}_1 \ \mathbf{X}_2 \ \dots \ \mathbf{X}_N]$ , depending on the context. As a convention,

- the indices  $i, j, k$  will exclusively refer to resident types (which were said to range from 1 to  $N$ ),
- the indices  $l, m, n$  are reserved for denoting birth states in a structured population model, and if only a finite number of different birth states exist they are numbered from 1 to  $d$ ,
- the indices  $a, b$  will only be used to indicate the scalar trait components that make up a trait vector, which we take to be  $z$ -dimensional,
- the indices  $s, t$  always relate to environmental components, where the dimension of the environment  $\mathbf{I}$  is  $r$  (cf. Sect. 1.2).

Thanks to these rules, a summation index implicitly has a range attached to it, as for example  $\sum_i$  can be unambiguously read as  $\sum_{i=1}^N$ . Our aim however was not a slight notational simplification, but to make calculations easier to verify.

As far as possible, we adhere to the convention (e.g., [3]) that matrices and tensors are denoted by an upright, sans serif capital like  $\mathbf{M}$ , vectors with a bold Italic letter like  $\mathbf{b}$  or  $\mathbf{V}$ , and scalars with a Greek or Roman letter like  $\lambda$ ,  $\Pi$ ,  $t$  or  $R_0$ . A consequence of this convention is that for example the  $l$ th component of the birth flow vector  $\mathbf{b}$  must be written as  $b_l$ , and one cannot mistake the matrix  $\mathbf{C}_{11}$  for the first diagonal element  $C_{11}$  of another matrix  $\mathbf{C}$ .

To help the reader, brackets around matrix-valued expressions have been made square, where vector- or scalar-valued expressions are signalled by round brackets; thus matrix components are indicated as, e.g.,  $[\mathbf{C}_{11}]_{ab}$ .

Furthermore, column vectors with all entries equal to 1 (resp. zero) will be denoted by  $\mathbf{I}$  (resp.  $\mathbf{0}$ ), where the dimension will be clear from the context. Similarly, the zero matrix is denoted by  $\mathbf{0}$  and the identity matrix by  $\text{id}$ .

Please see Sect. 3.1 for additional notations restricted to Sect. 3.

## 1.4 Assumptions

Here we present an overview of the assumptions scattered throughout this paper. The impact of some of these conditions cannot be meaningfully discussed at this point, as the relevant concepts have not been presented yet. Hence we refer the reader to the subsections where the assumptions are stated as preliminary to specific calculations. One notes that most are stated in the Introduction, and hence are necessarily active from there onwards until the end. Assumptions made in one of Sects. 2 or 3 do not apply to the other section, but are necessarily active in Sect. 4.

First and foremost we abide by the core premises of the adaptive dynamics framework: individuals have heritable traits that influence their life histories and little variation in these traits is present, the resident community is large and well-mixed while both mutants and mutation events are rare (1.1), plus the additional assumption that the community has a global point attractor, or alternatively that it has locally unique point attractors while mutational steps are sufficiently small so as to guarantee that after a successful invasion the community moves to a natural continuation of its earlier attractor (1.1, 1.2, 2.2). The basic process from which the deliberations start is derived in the following manner, as a limit of a fully individual-based community dynamics. Introducing a parameter  $\Omega$  called *system size* that scales inversely with the effects of interactions between the individuals in the community, the number of individuals must be about proportional to  $\Omega$ . The limit to consider is that where  $\Omega$  becomes large while the mutation probability per birth event gets so small that a mutant strategy reaching establishment becomes a rare event on the community dynamical timescale. To compensate for this rarity, time is rescaled so that the number of different established mutants per unit of time stays  $O(1)$ ; this new timescale is called the *evolutionary timescale*. (With increasing  $\Omega$ , the rescaling must be such that the rate of mutations reaching establishment decays sufficiently slowly to guarantee that the rescaled asymptotic rate at which the community goes extinct through demographic



fluctuations, decreases to zero.) On the ecological timescale, the community relaxes to its deterministic attractor before the next mutant comes along. This attractor can be calculated from the equilibrium equations (1) scaled by  $1/\Omega$ , i.e., when  $\mathbf{b}$  is read as a density per time and  $\mathbf{I}$  as a density. The described combination of a limit and a rescaling allows a reduced process description, where at almost all times there is but a small set  $\mathbb{X}$  of trait values around, in densities given by the corresponding deterministic community attractor. Such a process has been variously referred to as adaptive dynamics [49], oligomorphic dynamics [15] and trait substitution process [27]. The validity of the limit has been proven for some specific Markovian models by Champagnat [7]. For general physiologically structured populations there is as yet no proof for the step from the underlying stochastic models to the deterministic models considered by Diekmann et al. [20,21]. In our paper, we take the existence of the limit on faith, and from this vantage point study situations where mutational steps are small and all types present in the population are very similar. All order statements refer to the scale of the differences in the traits under consideration, between mutant and ancestor in Sect. 2 and mutant and residents in Sect. 3. In Sect. 4 however we also consider situations with similar residents and mutational steps that are of an even smaller order.

In addition, we impose regularity conditions that are inherent to our modelling approach: a thrice continuously differentiable dependence of the demographic parameters on trait values and environment (1.2, 3.3), offspring distributions that decay sufficiently quickly to have uniformly bounded third moments (which amounts to the thrice differentiability of the generating function) (2.5), and no birth states with zero birth flow for the sole singular resident (3.4). Finiteness of the number of birth states (1.2) can also be put into this class of requirement, although it is fundamental to our approach only in the sense that it is required by our specific machinery (i.e. vectors and matrices, instead of distributions and operators).

Lastly, we inherit assumptions made by Diekmann and Law [15], as one of our aims is to see how the canonical equation changes when their premise of ODE population regulation is dropped: unbiased mutations (2.6), and a stochastic trait substitution process that becomes deterministic when the mutational steps become small while time is rescaled such that on the new scale the rate of trait change stays  $O(1)$  (2.3).

## 2 The canonical equation of directional adaptive dynamics

### 2.1 Unstructured populations

The *canonical equation of adaptive dynamics* [15] is a first order approximation for the average speed of evolution. The rate of trait change per time of the  $i$ th type in a community is

$$\frac{d\mathbf{X}_i}{dt} \approx \frac{1}{2} \hat{n}_i \mu_i(\mathbf{X}_i) \mathbf{M}(\mathbf{X}_i) \left. \frac{\partial s_{\mathbb{X}}(\mathbf{Y})}{\partial \mathbf{Y}} \right|_{\mathbf{Y}=\mathbf{X}_i} \quad (2)$$

where the *mutational covariance matrix*  $\mathbf{M}$  at trait value  $\mathbf{X}_i$  is defined as  $\mathbf{M}(\mathbf{X}_i) := \int \mathbf{V} \mathbf{V}^T \mathfrak{M}(\mathbf{V}, \mathbf{X}_i) d\mathbf{V}$ , an expression that depends on the multivariate *distribution of mutational steps*  $\mathfrak{M}(\mathbf{V}, \mathbf{X}_i)$  from  $\mathbf{X}_i$  to  $\mathbf{X}_i + \mathbf{V}$ . (here we abuse  $\mathbf{V}$  temporally as

a random variable). The speed of evolution is thus seen to be proportional to the mutation probability per birth event  $\mu_i$ , the equilibrium population size  $\hat{n}_i$  in the given  $N$ -resident community  $\mathbb{X}$ , and the fitness gradient  $\left( \frac{\partial s_{\mathbb{X}}(Y)}{\partial Y} \Big|_{Y=X_i} \right)^{\top}$ .

At the singular strategies the fitness gradient becomes zero. Hence, close to the singular strategies the first and second order terms are of similar size, and the approximation embodied by the canonical equation loses its descriptive power. Champagnat [5,6] has proven that under some additional technical conditions, trait substitution processes that are based on population models with ODE deterministic skeletons sporting globally attracting point equilibria do converge weakly to the deterministic process captured by the canonical equation. His proof applies without change to the general case except for some small changes in the formulas, to be provided in the next subsections. Simulations suggest that away from the singular points, the pictures derived by solving the canonical equation capture the temporal development of the trait composition of the underlying individual-based process rather well (e.g., Fig. 2 in [15]; Fig. 10 in [49]) in a fair-sized parameter volume close to the origin of the three-dimensional parameter space spanned by mutational step size, inverse system size and mutation probability per birth event.

## 2.2 Aims of this section

Where Dieckmann and Law [15] formulated the canonical equation for ODE models, we aim here to relax that limitation by considering the far wider class of physiologically structured population models, and thus to recover a generalized form of Eq. (2). As the canonical equation (in both formulations) fails to capture the trait substitution behaviour of systems near evolutionary singularities, a separate part of this paper will deal with singularities (Sect. 3).

Our goal is to find out how a community (or more precisely, a set of trait values) will evolve, and at what rate. The basic scenario is the following: we start by considering a coalition of  $N$  different trait values that are the strategies of residents, which form a community that is at equilibrium. This fixed point attractor is presumed to exist for the community as a whole, as a unique set of positive equilibrium densities for all  $N$  trait values. When a mutant with positive invasion fitness appears, several things may happen. Usually, it will fail to get established in the community due to stochasticity, and will disappear. However, if it does get established, it will remove its parent from the population through competitive exclusion. Then the  $N - 1$  remaining residents plus the invader will have their densities equilibrate at new values, assumed to be positive and unique to the given set of  $N$  strategies. The first situation means that the community returns to its earlier state, the second that a small evolutionary step has taken place. Mutation events are by assumption so rare, that the community has relaxed to its attractor before the next mutation event takes place. As the cycle of mutation followed by possible invasion and equilibration can occur over and over again, this invasion/replacement dynamics provides a scenario where evolution proceeds through a great number of small trait changes.

The above setting assumes that the mutating trait value is not (close to) singular nor close to the boundary of the coexistence region, and that the population dynamics is sufficiently well-behaved, so that the dictum “invasion implies substitution” holds [13, 14, 25, 26, 45]. We stress here that we restrict ourselves to point attractors, as it is not clear yet to which extent the rule holds for more complicated attractors than fixed points and limit cycles. Thus special situations, where either the mutant coexists indefinitely with its parent or where it drives several residents to extinction, are explicitly excluded from this paper’s analysis. Also, in higher dimensional trait spaces there are unavoidable exceptions to the dictum: several selectively neutral mutants (in directions orthogonal to the invasion gradient) may briefly establish a foothold, until the next successful mutant in the direction of the invasion gradient kills off its progenitor along with those recent invaders. But these problematic scenarios are essentially negligible, as they represent a fraction of the total invasion events that vanishes in the limit of infinitesimal mutation steps.

Research shows that the assumption of uniqueness of the community fixed point is merely made for mathematical convenience, as the community attractors before and after successful invasions are arbitrarily close for sufficiently small mutation steps [26, 45, 14]. Thus the invader inherits the attractor of the resident it replaces, as the new attractor lies on the continuation of the older. The existence and (local) uniqueness is therefore guaranteed under the mild restrictions put forward by Geritz et al. [26], which essentially are absence of population dynamical bifurcations and sufficient smoothness of the model ingredients. If several fixed point attractors exist for a given set of trait vectors, they necessarily lie on distinct branches of solutions to the population dynamical equilibrium equations. Distinguishing such multiple attractors is therefore an administrative rather than mathematical problem, as the initial conditions (specifically, the earlier community attractors) determine in which basin of attraction the community finds itself.

The appearance of mutants, governed by the probability per birth event of a mutation and the distribution  $\mathcal{M}$  of mutational steps, and their eventual success or failure at establishment is inherently stochastic. This means that trait values are stochastic and time-dependent variables that we can characterize by the probability of the community being in a given state at a given time. The essential information to determine this probability is the rate at which the community’s state is expected to change from one state to another, an issue we will turn our attention to over the following paragraphs.

### 2.3 The deterministic path

We can now view the change in community composition as a Markov process, with a probability  $\Pi(\mathbb{X}, t)$  that the population is in state  $\mathbb{X}$  at a given time  $t > 0$ . From the interpretation as a Markovian dynamics, there are instantaneous transition rates  $\pi(B, A)$  from any state  $A$  to any  $B$ . The connection between probability distribution and transition rates is found by observing that the rate of change in  $\Pi$  must consist of two terms at any time, a gain in probability mass from other states into  $\mathbb{X}$ , and a loss

from transitions to other states (the Kolmogorov forward equation):

$$\frac{\partial \Pi(\mathbb{X}, t)}{\partial t} = \int (\pi(\mathbb{X}, \mathbb{X}') \Pi(\mathbb{X}', t) - \pi(\mathbb{X}', \mathbb{X}) \Pi(\mathbb{X}, t)) d\mathbb{X}' \tag{3}$$

For any observable  $\psi$  of our dynamical system, the expected value at time  $t$  is defined as the ensemble average

$$\mathbb{E}(\psi(\mathbb{X})) := \int \psi(\mathbb{X}) \Pi(\mathbb{X}, t) d\mathbb{X}$$

Applying the above definition with  $\psi$  the identity and using the Markov property above, we find the following equality:

$$\begin{aligned} \frac{d}{dt} \mathbb{E}(\mathbb{X}) &= \int \mathbb{X} \frac{\partial \Pi(\mathbb{X}, t)}{\partial t} d\mathbb{X} \\ &= \iint \mathbb{X} (\pi(\mathbb{X}, \mathbb{X}') \Pi(\mathbb{X}', t) - \pi(\mathbb{X}', \mathbb{X}) \Pi(\mathbb{X}, t)) d\mathbb{X}' d\mathbb{X} \\ &= \iint (\mathbb{X}' - \mathbb{X}) \pi(\mathbb{X}', \mathbb{X}) \Pi(\mathbb{X}, t) d\mathbb{X}' d\mathbb{X} \\ &= \mathbb{E}(\mathbb{A}_\epsilon(\mathbb{X})) \end{aligned} \tag{4}$$

where we introduce the operator  $\mathbb{A}_\epsilon(\mathbb{X}) := \int (\mathbb{X}' - \mathbb{X}) \pi(\mathbb{X}', \mathbb{X}) d\mathbb{X}'$ , and the parameter  $\epsilon$  that is proportional to the mutation step size (so the distance between a mutant and its ancestor is  $O(\epsilon)$ ). The solution to Eq. (4) is called the *mean path* of  $\mathbb{X}$ . Sadly this equation is not a self-contained equation in  $\mathbb{E}(\mathbb{X})$ , causing much mathematical grief (or joy, depending on one’s disposition). To dodge this issue, the *deterministic path* is introduced, which is the solution to this variation on Eq. (4):

$$\frac{d}{dt} \bar{\mathbb{X}} = \mathbb{A}_\epsilon(\bar{\mathbb{X}}) \tag{5}$$

The mean and deterministic paths would coincide if the distribution of  $\mathbb{X}$  is concentrated in a point or if the integral on the right hand side is linear in  $\mathbb{X}$ , but neither is true in general. Whether the deterministic path is a valid approximation of the mean path clearly depends on whether it is dominated by the first order term of  $\mathbb{A}_\epsilon$  or not. Intuitively one expects this to be true, as the adaptive dynamics modelling approach has evolution proceeding through very many very small steps. Thus with decreasing mutational step size, it takes more and more mutation steps to cover the same distance in trait space and a law-of-large-numbers effect should hold sway in the limit  $\epsilon \rightarrow 0$ . Dieckmann and Law [15] assumed this to be a valid approximation, relying on simulations plus the considerations of Van Kampen [55]. More recently Champagnat [5,6] has proven the weak convergence of the stochastic trait substitution process to the solution of Eq. (6). Apart from a number of more technical assumptions, all papers mentioned assume ODE population dynamics

and the existence of a global point attractor for the deterministic community dynamics. Furthermore, the many-small-steps argument suggests that the error around the deterministic approximation is Gaussian with variance proportional to  $\epsilon$ . This heuristic argument is confirmed by Champagnat [5,6] who derives the full equations for this Gaussian error process as well.

We will simplify the notations  $\mathbb{E}(X)$  and  $\bar{X}$  to  $X$  henceforth, and similarly for the community  $\mathbb{X}$ , so Eq. (5) is rewritten as

$$\frac{d}{dt} \mathbb{X} = \mathbb{A}_\epsilon(\mathbb{X}) \quad (6)$$

One should not lose sight of the fact that for the remainder of this section, any strategy or community not marked by a prime ( $'$ ) should be read as the value predicted by the deterministic limit; hence the mutation step  $V := X'_i - X_i$  is the difference between a potential stochastically realized new strategy and its deterministically calculated originator.

The next step in capturing the dynamics is to divide and conquer the transition probabilities.

#### 2.4 The transition probabilities

Since we consider rare mutations, any transition must be a mutation affecting a single strategy vector. Therefore nontrivial transition rates are of the form  $\pi_o(X'_i, X_i, \mathbb{X})$ , representing the rate at which the  $i$ th resident in a given community  $\mathbb{X}$  switches from state  $X_i$  to  $X'_i$ . Thus if we interpret  $\mathbb{X}$  as the matrix  $[X_1 \ X_2 \ \dots \ X_N]$ , then the  $i$ th column of the matrix equation describing the deterministic path (5) simplifies to

$$\frac{d}{dt} X_i = \mathbb{A}_\epsilon^i(\mathbb{X}) = \int (X'_i - X_i) \pi_o(X'_i, X_i, \mathbb{X}) dX'_i \quad (7)$$

Our next aim must therefore be to derive analytical expressions for the right hand side of Eq. (7). As a first step, we split  $\pi_o$  into separate factors by observing that mutation and selection are independent processes, hence these transition probabilities are the product of the appearance rate of mutants and their probability of establishment:

$$\begin{aligned} & \pi_o(X'_i, X_i, \mathbb{X}) \\ &= (\text{production rate of mutants } X'_i) (\text{establishment chance of } X'_i) \\ &= \overbrace{(\text{birth rate of } X_i \text{ types}) (\text{mutation chance } X_i \rightarrow X'_i)} \overbrace{P(X'_i, \mathbb{X})} \\ &= \overbrace{\lambda(X_i, \mathbb{X})} \overbrace{\hat{n}_i} \overbrace{\mu(X_i)} \overbrace{\mathfrak{M}(X'_i - X_i, X_i)} P(X'_i, \mathbb{X}) \end{aligned} \quad (8)$$

We stress again that the values above are population averages, while  $\hat{n}_i$  stands for the equilibrium density of the  $i$ th type. The probability  $P$  of establishment is the expected outcome of a branching process. This rather complicated beast, which depends heavily on the underlying population model, will be resolved in the next subsection. The

other factors are easy to understand. The appearance rate of mutants (that is,  $X'_i$ -type individuals that have  $X_i$ -type parents) is just the total offspring production by  $X_i$ -type parents, times the mutation rate of  $X_i$  into  $X'_i$ . This comes from the fact that we have assumed the mutational steps to be small, so only the  $i$ th type can be the ancestor of our mutant. The total production of  $X_i$  individuals is (by definition) the instantaneous per capita birth rate of such individuals, times their equilibrium density. The mutation chance  $X_i \rightarrow X'_i$  is the probability per birth event of mutating for an  $X_i$ -type individual, times the mutation distribution around this trait value;  $\mathfrak{M}(V, X_i)$  is the probability density of a mutation from  $X_i$  to  $X_i + V$ .

In a closed system at equilibrium, the per capita birth rate is the inverse of the expected lifespan. This was termed the “microcosm principle” by Mollison [51], and it holds for the stochastic systems we consider. The argument is that in a large population ergodically fluctuating around its attracting density, the density is the product of the influx of new individuals and the time they stay in the population. Since the population is closed, the newborns correspond to the influx of residents, and only death ends a resident’s stay. Hence

$$\begin{aligned} \mathbb{E}(\text{density}) &= \mathbb{E}(\text{influx of individuals per area}) \mathbb{E}(\text{duration of stay}) \\ &= \mathbb{E}(\text{per capita birth rate} \times \text{density}) \mathbb{E}(\text{lifespan}) \\ &= \mathbb{E}(\text{per capita birth rate}) \mathbb{E}(\text{density}) \mathbb{E}(\text{lifespan}) \end{aligned} \tag{9}$$

where the last step follows from our assumptions of large system size and thrice differentiable model ingredients. So we conclude that the expected lifespan  $T_s$  is the inverse of the birth rate:

$$T_s := T_s(X_i, \mathbb{X}) = (\mathbb{E}(\text{per capita birth rate}))^{-1} = \lambda(X_i, \mathbb{X})^{-1} \tag{10}$$

We can substitute this result in our breakdown of  $\pi_o$  (8) and move on to a study of  $P(X'_i, \mathbb{X})$ .

### 2.5 The establishment probability

To determine the establishment probability of a given mutant, we recall from the introduction on adaptive dynamics (Sect. 1.1) a statement about the link between branching processes and adaptive dynamics: under very general conditions, the probability  $P(Y, \mathbb{X})$  of an individual with strategy  $Y$  establishing itself in a given community  $\mathbb{X}$ , is related to that type’s invasion fitness by

$$P(Y, \mathbb{X}) > 0 \Leftrightarrow s_{\mathbb{X}}(Y) > 0 \tag{11}$$

(cf. [32]).

We now require a quantitative relationship between these entities. We will derive this relation in two steps: first we relate  $P$  to the lifetime offspring production  $R_0$ , and then  $R_0$  to the fitness  $s$ . For the first part, we will use some techniques from the theory of branching processes. By assumption we started with the large equilibrium

community  $\mathbb{X}$  and a single mutant. Thus the community resides on its attractor as its size makes deviations from the mean too small to be significant, and a deterministic description is valid. This constitutes the environment of the branching process that describes the demography of the initial mutant and its (still rare) offspring, which are too rare to influence each other. An approximation first heuristically derived (as a generalization of a result of Haldane [33]) by Ewens [24] for single type branching processes (Eq. (12)), and its multitype counterpart (Eq. (14)), gives our first relation as we shall presently see.

If there is only a single possible birth state in our (at this timescale) constant environment, and a small but positive scalar  $\varrho$  so that the lifetime reproductive output is  $R_0 = 1 + \varrho$ , then our single-type process is called *slightly supercritical*. If the probability generating function  $g(z, \varrho)$  of the offspring distribution is three times continuously differentiable in its arguments, then

$$P(\mathbf{X}'_i, \mathbb{X}) = \frac{2\varrho}{\sigma^2} + O(\varrho^2) = \frac{2 \log R_0}{\sigma^2} + O(\varrho^2) \quad (12)$$

where  $R_0$  and  $\sigma^2$  are respectively the mean and variance of the mutant's offspring distribution in the community. For further information see [1, 22, 35].

Unfortunately the above result does not suffice, as we want to include population dynamics where multiple birth states are possible. In cases where there are  $d$  possible birth states, we denote by the stochastic variable  $\xi_{lm}$  the number of offspring born in state  $l$  to a parent that was itself born in state  $m$ . Then  $\mathbb{E}(\xi_{lm}) = [L]_{lm}$  relates these random variables to the reproduction matrix we introduced at the start of Sect. 1.2. Furthermore,  $R_0$  is in such multitype models the dominant eigenvalue of the  $L$  matrix, and we denote by  $\mathbf{u}$  and  $\mathbf{v}^\top$  respectively the right- and left eigenvectors of  $L$  belonging to  $R_0$ :

$$R_0 = \lambda_d(L) = \mathbf{v}^\top L \mathbf{u} \quad (13)$$

where we normalized  $\mathbf{u}$  and  $\mathbf{v}$  by requiring  $\sum_l |u_l| = 1$  and  $\mathbf{v}^\top \mathbf{u} = 1$  (see e.g., [4]).

One should be mindful that this notation for  $\xi_{lm}$  reverses the order of the subscripts with respect to the traditional branching processes notation. The definition of  $\mathbf{u}$  and  $\mathbf{v}$  is similarly reversed, so that in both notations  $\mathbf{u}$  is the stable type distribution, and  $\mathbf{v}$  the vector of the (generationwise) *reproductive values*.

Similarly, in the above  $d$ -type situation for a slightly supercritical process, the chance  $P_l$  for a single mutant born in state  $l$  of establishing itself can be written as

$$P_l(\mathbf{X}'_i, \mathbb{X}) = \frac{2\varrho}{B} v_l + O(\varrho^2) = \frac{2 \log R_0}{B} v_l + O(\varrho^2) \quad (14)$$

with  $B := \sum_{l'mn} u_{l'} v_m v_n \mathbb{E}(\xi_{ml'} (\xi_{nl'} - \delta_{mn}))$  where  $\delta$  is the Kronecker delta (i.e.,  $\delta_{ll} = 1$  and  $\delta_{lm} = 0$  if  $l \neq m$ ) and conditions similar to those of the single state case (12) are assumed to be satisfied (see [2, 23] for further details). Clearly  $B$  and both eigenvectors depend on  $\varrho$ , as does  $R_0$ . It is easily seen that if  $d = 1$ , the earlier version is recovered, as it should be. We have mainly stated the (better known) single-type

result (12) earlier on, to hint at an interpretation of  $B$  as a variance. Bearing in mind that  $\mathbf{u}$  and  $\mathbf{v}^\top$  are the right- and left eigenvectors of  $\mathbf{L}$ , we find

$$\begin{aligned}
 B &= \sum_l u_l \mathbb{E} \left( \sum_{mn} v_m v_n \xi_{ml} \xi_{nl} \right) - \sum_{lm} u_l v_m^2 \mathbb{E}(\xi_{ml}) \\
 &= \sum_l u_l \mathbb{E} \left( \left( \sum_m v_m \xi_{ml} \right)^2 \right) - \sum_{lm} v_m^2 \mathbb{E}(\xi_{ml}) u_l \\
 &= \sum_l u_l \left( \text{Var} \left( \sum_m v_m \xi_{ml} \right) + \left( \mathbb{E} \left( \sum_m v_m \xi_{ml} \right) \right)^2 \right) - \sum_m v_m^2 R_0 u_m \\
 &= \sum_l u_l \text{Var} \left( \sum_m v_m \xi_{ml} \right) + \sum_l u_l v_l^2 R_0^2 - \sum_m v_m^2 R_0 u_m \\
 &= \sum_l u_l \text{Var} \left( \sum_m v_m \xi_{ml} \right) + O(\varrho^2)
 \end{aligned} \tag{15}$$

where the  $O(\varrho^2)$  approximation holds since  $R_0^2 - R_0 = \varrho^2 + \varrho$ . By defining

$$\sigma^2 := \sum_l u_l \text{Var} \left( \sum_m v_m \xi_{ml} \right) \tag{16}$$

we can replace  $B$  with the variance-like quantity  $\sigma^2$  to bring out the close similarity of the multiple birth state case (14) with the simpler case (12):

$$P(\mathbf{X}'_i, \mathbb{X}) = \sum_n P_n(\mathbf{X}'_i, \mathbb{X}) u_n = 2 \frac{\log R_0}{\sigma^2} + O(\varrho^2) \tag{17}$$

since  $\sum_n v_n u_n = 1$ , which concludes the first step in quantifying the relation (11) between establishment chance  $P(\mathbf{Y}, \mathbb{X})$  and invasion fitness  $s_{\mathbb{X}}(\mathbf{Y})$ .

The second step is to determine the relation between  $R_0$  and  $s_{\mathbb{X}}(\mathbf{Y})$ . To derive this, we consider the *birth kernel* notation of a general model. If we denote the environment set by the community  $\mathbb{X}$  as  $\mathbf{I}_{\mathbb{X}} := \mathbf{I}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N)$ , then there exists a matrix function  $\Lambda$  with entries  $[\Lambda(\mathbf{X}'_i, \mathbf{I}_{\mathbb{X}}, a)]_{lm}$  that are the expected number of offspring born in state  $l$  to a  $\mathbf{X}'_i$ -type invader, newly born in state  $m$ , before the invader reaches age  $a$ , in the equilibrium community  $\mathbb{X}$  (cf. [21]). The link with the lifetime offspring production matrix is obviously that  $[\Lambda(\mathbf{X}'_i, \mathbf{I}_{\mathbb{X}}, \infty)]_{lm} = [\mathbf{L}]_{lm} = \mathbb{E}(\xi_{lm})$ . Using this notation, the invasion fitness  $s_{\mathbb{X}}(\mathbf{X}'_i)$  is the (generally unique) solution for  $\rho$  of the Euler–Lotka equation

$$\lambda_d \left( \int_0^\infty e^{-\rho a} \Lambda(\mathbf{X}'_i, \mathbf{I}_{\mathbb{X}}, da) \right) = 1 \tag{18}$$



where  $\lambda_d$  is the dominant eigenvalue operator. In Appendix A we show how to extract from Eq. (18) the following relationship:

$$\log R_0 = T_f(X_i, \mathbb{X}) s_{\mathbb{X}}(X_i + V) + O(\|V\|^2) \tag{19}$$

where  $T_f$  is the *average age at giving birth* (97). If we approximate  $s_{\mathbb{X}}(X'_i)$  using the fitness gradient, we can finally formulate the establishment probability (in both single (12) and multitype (17) cases) as

$$P(X'_i, \mathbb{X}) = \frac{2 T_f s_{\mathbb{X}}(X'_i)}{\sigma^2} + O(\epsilon^2) \tag{20}$$

$$= \frac{2 T_f}{\sigma^2} (X'_i - X_i)^\top \frac{\partial s_{\mathbb{X}}(Y)}{\partial Y} \Big|_{Y=X_i} + O(\epsilon^2) \tag{21}$$

As this last expression contains the factor  $X'_i - X_i$ , we are free to evaluate  $T_f$  and  $\sigma^2$  at  $X_i$  without changing the order of the approximation. Hence the mutant trait value  $X'_i$  only appears in the mutation step  $V := X'_i - X_i$ .

Bear in mind that this result only holds for positive  $P$ , as such is the starting point of the approximation formula (14).

### 2.6 The canonical equation for physiologically structured population models

After this divide-and-conquer campaign, we can substitute the factors that make up the transition rates (8), (10), (21) into the equation describing the deterministic path (7):

$$\begin{aligned} \mathbb{A}_\epsilon^i(\mathbb{X}) &= \int (X'_i - X_i) \pi_o(X'_i, X_i, \mathbb{X}) dX'_i \\ &= \frac{\hat{n}_i \mu}{T_s} \int (X'_i - X_i) \mathfrak{M}(X'_i - X_i, X_i) P(X'_i, \mathbb{X}) dX'_i \\ &= \frac{T_f}{T_s} \frac{2 \hat{n}_i \mu}{\sigma^2} \int V \mathfrak{M}(V, X_i) V^\top \frac{\partial s_{\mathbb{X}}(Y)}{\partial Y} \Big|_{Y=X_i} dV + O(\epsilon^3) \end{aligned} \tag{22}$$

where  $\epsilon$  is the average mutation step size. The estimate of the establishment probability (21) introduces an error term equal to a constant times  $\int V \mathfrak{M}(V, X_i) O(\|V\|^2) dV$ .

Equation (22) allows us finally to formulate the canonical equation for structured population models with unbiased mutation distributions, giving an approximate rate of change under evolutionary selection for traits of the  $i$ th resident in a multitype community  $\mathbb{X}$  in the limit of infinitesimal mutational step size, as

$$\mathbb{A}_0^i(\mathbb{X}) = \frac{T_f}{T_s} \frac{\hat{n}_i \mu}{\sigma^2} M \frac{\partial s_{\mathbb{X}}(Y)}{\partial Y} \Big|_{Y=X_i} \tag{23}$$

We repeat that all factors in Eq. (23) are expected values, and that the canonical equation characterizes the deterministic, not the mean, path. One sees that most of the parameters in the canonical equation (23) depend both on the strategy  $X_i$  and the entire community  $\mathbb{X}$ ; the exceptions are  $\mu$  and  $M$ , which only depend on the strategy.

In the last transition, a factor 2 may seem to be lost. Its disappearance stems from the fact that the approximation formula (14) only holds for supercritical processes, where  $\log R_0 > 0$ , but in other cases we must substitute a zero. As the sign of  $\log R_0$  is that of  $V^T \frac{\partial s_{\mathbb{X}}(Y)}{\partial Y} \Big|_{Y=X_i}$ , we can correctly account for the subcritical cases by integrating over a halfspace. If the mutation distribution is unbiased, this comes down to dividing by two. In case this assumption is not met, one has to stick to Eq. (22). Alternative formulations when mutations are biased are given by Champagnat et al. [6,8].

All the quantities in Eq. (22), including the order estimate, are still on the original community dynamical timescale. The reason for not changing to expressions in evolutionary time, is that doing so lets the biological interpretation of model ingredients disappear from sight. The speeded up timescale necessary for deriving a proper limit process is constructed by equating one unit of evolutionary time to  $1/\epsilon$  units of community dynamical time. The order estimate becomes  $O(\epsilon)$  in evolutionary time, which is higher than the  $O(\sqrt{\epsilon})$  estimate for the approximation to the stochastic process using the deterministic path (cf. the paragraph preceding Eq. (6)). Hence the overall order of the approximation is dominated by the process noise and not by the calculation of the mean speed of change of  $\mathbb{X}$ , and is  $O(\sqrt{\epsilon})$ .

One sees that the only difference between the canonical equation for unstructured (2) and for structured populations (23) is that a factor  $1/2$  becomes a factor  $T_f/(T_s \sigma^2)$ . As an illustration, we now bridge this gap by recovering the canonical equation for unstructured population models from the general result for structured models. The unstructured case deals with ODE models, which implies the absence of any historical dependence of the individual birth and death rates. Hence in such models the initial invasion of a mutant is described by a linear birth-and-death process. If we denote birth and death rates respectively by  $\lambda$  and  $\mu$ , we can calculate the ratio  $T_f/(T_s \sigma^2)$ . First, the ratio of the average age at giving birth to the life expectancy can be computed since

$$T_f = \frac{\int_0^\infty \lambda e^{-\mu t} dt}{\int_0^\infty \lambda e^{-\mu t} dt} = \frac{1}{\mu} = \int_0^\infty \mu e^{-\mu t} dt = T_s$$

Second, the offspring distribution follows from the observation that a lifetime number of  $i$  children means  $i$  successive birth events (each with relative probability  $\lambda/(\lambda + \mu)$ ), followed by a death event (with probability  $\mu/(\lambda + \mu)$ ). All events being independent, the probability of having  $i$  offspring is the product of all these probabilities:

$$P(\xi = i) := p_i = \left( \frac{\lambda}{\lambda + \mu} \right)^i \frac{\mu}{\lambda + \mu}$$

This geometric distribution has variance  $\sigma^2 = \lambda(\lambda + \mu)/\mu^2$ , so the factor we try to calculate is

$$\frac{T_f}{T_s} \frac{1}{\sum_l u_l \text{Var}(\sum_m v_m \xi_{ml})} = \frac{\mu^2}{\lambda(\lambda + \mu)} = \frac{1}{2} + O(\varrho)$$

since  $\lambda = \mu + O(\varrho)$  in a slightly supercritical situation. This completes our recovery of the result of [15].

### 3 The normal form of the invasion function at a singularity

When trying to figure out the nature of the invasion function for a community close to a singularity, the first naive attempts usually fail. A clear example is the formulation of the general form of the  $s$ -function for a community of three or more residents, close to a singular strategy. If one assumes the existence of a Taylor expansion up to quadratic terms and checks some consistency conditions that must surely hold, a single page of calculations (Appendix C) gives the clean-but-nonsensical result that  $s = O(\varepsilon^3)$  at the singularity, no matter what model or parameters.

When we look at a community of two residents that are similar and close to a singular strategy, we can see the root of the problem. At the limit where the residents' strategies are equal to the singular strategy, the population densities show a line of neutrally stable equilibria (Fig. 1); any other combination of trait vectors shows an attracting point equilibrium. Thus a bifurcation that is unusual for general dynamical systems, is generic in the context of invasion analysis. The illustration shows the essential nature of the beast: even though a derivative does not exist, the directional derivatives do. What this suggests, is to blow up singularities by separating the directional components of a strategy from its norm. The notations that follow are natural implementations of this idea.

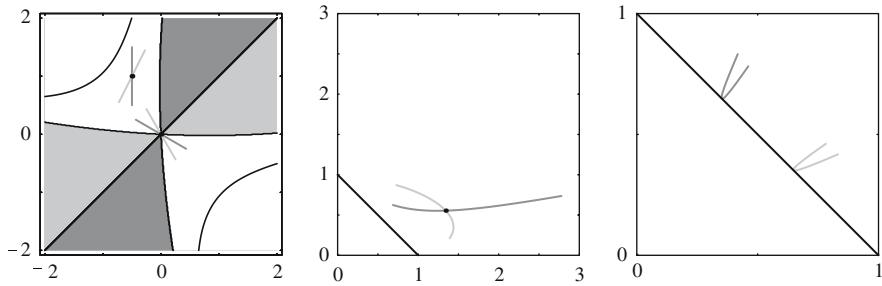
#### 3.1 Additional notations for this section

On top of the notations we presented in Sect. 1.3, we introduce the following conventions.

As we are interested in the form of the fitness function for a community near an evolutionarily singular strategy, we choose a parametrization centered around it. Denoting the singular trait value by  $X^*$ , a resident has strategy vector  $X = X^* + U$ , or  $X_i = X^* + U_i$  if there are several residents. Likewise an invader has trait value  $Y = X^* + V$ .

We introduce the small (bifurcation) parameter  $\varepsilon$  to scale the set of resident traits: for each  $i$  from 1 to  $N$  there is a vector  $\xi_i$  so that the  $i$ th resident has strategy  $X_i = X^* + U_i = X^* + \varepsilon \xi_i$ .

Any quantity with an asterisk will refer to a community at equilibrium with only the singular strategy present: e.g.,  $b^*$  is the equilibrium birth flow and  $I^*$  the equilibrium environment when only  $X^*$  is present. Furthermore, all derivatives in this section will



**Fig. 1** The nature of the beast: we consider an  $N$ -resident Lotka–Volterra system with scalar strategies. The population dynamics for the  $i$ th type is given by  $d \log n_i/dt = 1 - \sum_j a(X_i, X_j)n_j - a(X_i, Y)m$  and similarly for the mutant’s density  $d \log m/dt = 1 - \sum_j a(Y, X_j)n_j - a(Y, Y)m$ , where the interaction function was chosen as  $a(X, X') := 1 + (X - X')(0.05X + 1.00X' - 0.03X^2 - 0.02XX' + 0.1X'^2)$ . In the first plot, strategy  $X_1$  is plotted against strategy  $X_2$ , the dark gray area is defined by  $s_{X_1}(X_2) < 0$ , the light gray one by  $s_{X_2}(X_1) < 0$ . In the white zone the equilibrium densities of both residents have the same sign, positive on the origin’s side of the black curve and negative on the other. Thus all points on the four straight lines drawn in gray represent strategy combinations that can coexist in a protected manner (since they are mutually invadable). The second and third graph plot the equilibrium density of  $X_1$  strategists against that of  $X_2$  strategists. The black dot in the second plot corresponds to the coalition  $(-0.5, 1)$  indicated on the first plot, and the gray curves on the second plot correspond to the identically colored lines through  $(-0.5, 1)$  in the first plot. The same correspondence holds between the two lines through the singularity at  $(0, 0)$  in the first plot, and the curves in the third plot. The aim of these figures is to point out what happens as the community approaches the singularity: one sees that there exists no limit for the densities when both strategies converge to the singular trait value, although in each direction this limit exists. Hence the black point on the second plot is the normal situation where the density equations have a stable fixed point solution, but in the third plot we see that this point degenerates into a line of neutrally stable equilibria when both populations are at the singular trait value. Note that the system is scaled such that the equilibrium density is always 1 for a monomorphic population. As all the curves in the second and third plot are above the line  $\hat{n}_1 + \hat{n}_2 = 1$ , the total density in a community with two residents is always higher than in one with a single resident. From the third plot, we expect that the total density in a community “close” to the singularity in terms of some distance measure, will have a zero linear part when expanded in terms of this distance; the analysis we present will show that this holds true in general

be evaluated for exactly that community. Thus a very substantial notational simplification is the systematic suppression of variable names and the location of evaluation: we see that without ambiguity, we can denote, e.g., the average of the lifetime reproductive output  $L = L(Y, I)$ , derived first for its second argument then for its first and evaluated at the singular strategy and environment, as the  $r \times z$  matrix

$$\frac{\partial^2 \lambda_d(L)}{\partial I \partial Y} := \frac{\partial}{\partial Y} \left( \frac{\partial \lambda_d(L(Y, I))}{\partial I} \right)^T \Bigg|_{\substack{Y = Y^* \\ I = I^*}} \tag{24}$$

where  $\lambda_d$  is the dominant eigenvalue operator.

Since no third order derivatives occur in this paper, all partial derivatives of scalar functions ( $s$ ,  $r$  and  $\lambda_d$ ) are either row vectors or matrices. A minor complication is however the occurrence of tensors of rank 3 as derivatives of matrix functions ( $G$  and  $L$ ). Instead of solving this issue by treating them componentwise and thus cluttering the notation, we interpret these tensors as matrices with row vectors as

elements by introducing an additional notation: to take the derivative of  $\mathbf{L}$  in the mutant direction as an example, we define it componentwise as

$$\left[ \frac{\partial \mathbf{L}}{\partial \mathbf{Y}} \right]_{lm} := \frac{\partial [\mathbf{L}]_{lm}}{\partial Y} \tag{25}$$

Whenever this symbol occurs, it will always be in an expansion and acting on an appropriately dimensioned vector like  $\mathbf{U}$ , so that we have a  $d \times d$  matrix  $\frac{\partial \mathbf{L}}{\partial \mathbf{Y}}(\mathbf{U})$  that gives no further complications. The slightly different layout serves as a reminder that the vector-and-matrix notation cannot be used when the tensor is separated from its argument in parentheses. Whenever possible, we opt not to use this unfamiliar notation: e.g., since  $\mathbf{b}^*$  is a constant vector,  $\frac{\partial \mathbf{L}}{\partial \mathbf{Y}}(\mathbf{U}) \mathbf{b}^*$  may be replaced by  $\frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{Y}} \mathbf{U}$ .

In the case of a double subscript, parentheses are added to remove ambiguity: e.g.,  $(b_i)_l$  is the  $l$ th component of the  $i$ th resident’s birth flow. Without parentheses,  $b_{il}$  might just as well be a component of some matrix  $\mathbf{b}$ .

### 3.2 Aims of this section

In the introduction we have defined the invasion fitness of type  $\mathbf{Y}$  in an  $N$ -resident community  $\mathbb{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$  as the long-term average per capita growth rate of a rare  $\mathbf{Y}$ -type individual in a large equilibrium community made up of all the resident types,  $\mathbf{X}_1$  to  $\mathbf{X}_N$ . In this section we show that for such an  $N$ -resident community, the invasion fitness function  $s_{\mathbb{X}}(\mathbf{Y})$  up to quadratic terms can be constructed using only the trait values present plus the second order derivatives at the singularity of the simpler fitness function  $s_{\mathbf{X}}(\mathbf{Y})$ .

The effect is that the task of formulating the fitness function for a polymorphic community in the neighbourhood of an evolutionarily singular strategy for an arbitrarily complicated structured population model, is reduced to formulating the one-resident  $s$ -function, and either fitting the corresponding Lotka–Volterra model (Proposition 1) or substituting the simple  $s$ -functions into the normal form (73) that we will present below. Both procedures yield an invasion fitness function  $s_{\mathbb{X}}(\mathbf{Y})$  which is correct up to quadratic terms in the small parameter  $\varepsilon$ .

For example, assume one knows the simple fitness function  $s_{\mathbf{X}}(\mathbf{Y})$  for some model and one has resident strategies  $\mathbf{X}_1$  and  $\mathbf{X}_2$  (with  $N = 2$ ). First we calculate the second order partial derivatives of  $s_{\mathbf{X}}(\mathbf{Y})$  at the singularity:

$$C_{11} := \frac{1}{2} \frac{\partial^2 s_{\mathbf{X}}(\mathbf{Y})}{\partial X^2}, \quad C_{10} := \frac{1}{2} \frac{\partial^2 s_{\mathbf{X}}(\mathbf{Y})}{\partial X \partial Y}, \quad C_{00} := \frac{1}{2} \frac{\partial^2 s_{\mathbf{X}}(\mathbf{Y})}{\partial Y^2} \tag{26}$$

Using the additional notations  $\bar{\mathbf{U}} := \frac{U_1 + U_2}{2}$  and  $\Delta := \frac{U_1 - U_2}{2}$  where the deviations  $U_1$ ,  $U_2$  and  $V$  are  $O(\varepsilon)$ , we will show in Sect. 3.5 that the invasion fitness of any mutant  $\mathbf{Y}$  is

$$\begin{aligned}
 s_{X_1 X_2}(Y) &= V^T C_{00} V + 2\bar{U}^T C_{10} V + \bar{U}^T C_{11} \bar{U} - \Delta^T C_{00} \Delta \\
 &\quad + 2\Delta^T C_{10} (\bar{U} - V) \frac{\Delta^T [C_{00} + C_{10}^T] \bar{U}}{\Delta^T C_{10} \Delta} + O(\varepsilon^3) \tag{27}
 \end{aligned}$$

Therefore we can consider the equation above to be a normal form. It immediately shows that a Taylor expansion of  $s_{X_1 X_2}$  does not exist and explains why calculations like those in Appendix C are doomed to fail, with the exception of the case where strategies are scalar so that the equation above simplifies to  $s_{X_1 X_2}(Y) = (X_1 - Y)(X_2 - Y)C_{00} + O(\varepsilon^3)$ .

One available route for deriving the normal form for general  $N$ -resident population dynamics close to a singular strategy and showing the mentioned niceties, is to first prove the general case, then cast a general Lotka–Volterra system in that form and show what it reduces to, and lastly demonstrate that this form only depends on the mentioned strategies and derivatives. The unpleasant reality however is, that casting Lotka–Volterra models into the form of physiologically structured population models requires us in general to introduce an infinite dimensional vector as description of the environmental conditions  $I$  (one environmental dimension for every possible trait value). The proof for the infinite dimensional case requires more sophisticated mathematical tools than we use here, like operators and distributions instead of finite dimensional matrices and vectors. We fully expect, though, that the same techniques as used in this paper still hold for any model on a space supporting a chain rule and an inverse function theorem.

For clarity’s sake and given our own more limited mathematical expertise, we have opted for another route: we restrict ourselves to the case of structured populations with a finite dimensional environment, and show that the same normal form is found as derived separately for Lotka–Volterra systems. We will start with a detailed exposition of the Lotka–Volterra case in view of its familiarity, followed by the corresponding calculations for the structured case.

### 3.3 The normal form for Lotka–Volterra systems

The following is a general form for Lotka–Volterra systems, where  $r(Y)$  is the per capita growth rate in a virgin environment (i.e., the growth rate in the absence of competitors), and the interaction is fully determined by the interaction function  $a(Y, X)$  plus the trait value and the densities of the interacting types. We assume that  $r$  and  $a$  are  $C^3$  functions, to guarantee the existence of an expansion of the fitness function up to order  $O(\varepsilon^3)$ . If the community has  $N$  residents plus an invading type, the equations that govern growth can be formulated as

$$\left\{ \begin{aligned}
 \forall j: \quad \frac{1}{n_j} \frac{dn_j}{dt} &= r(X_j) \left( 1 - \sum_i a(X_j, X_i) n_i - a(X_j, Y) m \right) \\
 \frac{1}{m} \frac{dm}{dt} &= r(Y) \left( 1 - \sum_i a(Y, X_i) n_i - a(Y, Y) m \right)
 \end{aligned} \right. \tag{28}$$

We will first perform a trait-dependent rescaling and some calculations pertaining to monomorphic communities.

We first add a tilde to indicate rescaled quantities, and later drop the tilde once convinced that rescaling has no effect on the fitness value. We multiply the density of any type with the strength of its self-competition and similarly divide the interaction function:

$$\begin{aligned}\tilde{n}_i &:= a(\mathbf{X}_i, \mathbf{X}_i) n_i & \tilde{a}(\mathbf{X}_i, \mathbf{X}_j) &:= \frac{a(\mathbf{X}_i, \mathbf{X}_j)}{a(\mathbf{X}_j, \mathbf{X}_j)} \\ \tilde{m} &:= a(\mathbf{Y}, \mathbf{Y}) m\end{aligned}\quad (29)$$

Thus for any strategy  $\mathbf{X}$  we have that  $\tilde{a}(\mathbf{X}, \mathbf{X}) = 1$  and consequently the equilibrium density in a monomorphic world is always  $\hat{\tilde{n}} = 1$ , as seen from the equilibrium equation  $0 = r(\mathbf{X})(1 - \tilde{a}(\mathbf{X}, \mathbf{X})\hat{\tilde{n}})$ . We see that for example  $a(\mathbf{X}_i, \mathbf{X}_j)n_j$  equals  $\tilde{a}(\mathbf{X}_i, \mathbf{X}_j)\tilde{n}_j$ , so that the per capita growth rate, and therefore the invasion fitness  $s_{\mathbf{X}}(\mathbf{Y})$ , is independent of this rescaling. So without loss of generality, we assume from here onwards that  $a(\mathbf{X}, \mathbf{X}) = 1$  for any  $\mathbf{X}$  and hence that  $\hat{n} = 1$  if there is a sole resident type.

By a literal translation of the definition of the  $s$ -function (see 1.1) into symbols, we calculate the invasion fitness for a monomorphic community as

$$s_{\mathbf{X}}(\mathbf{Y}) = \lim_{T \rightarrow \infty} \lim_{m \rightarrow 0} \frac{1}{T} \int_0^T \frac{1}{m} \frac{dm}{dt} \Big|_{n=\hat{n}} dt = r(\mathbf{Y}) (1 - a(\mathbf{Y}, \mathbf{X})) \quad (30)$$

**Proposition 1** *For every single-resident fitness function  $s_{\mathbf{X}}(\mathbf{Y})$  and every strictly positive growth rate in a virgin environment  $r(\mathbf{Y})$ , there exists an interaction function  $a(\mathbf{Y}, \mathbf{X})$  such that the resultant Lotka–Volterra model (28) has the same single-resident  $s$ -function.*

*Proof* As we comply to the rescaling (29), the suitable interaction function can be found from the formula for the invasion fitness in a Lotka–Volterra model (30) as  $a(\mathbf{Y}, \mathbf{X}) := 1 - s_{\mathbf{X}}(\mathbf{Y})/r(\mathbf{Y})$ .  $\square$

In practice, a constant growth rate  $r(\mathbf{Y}) := 1$  is usually preferable as it tends to simplify calculations.

Once we have fitted an interaction function to a simple fitness function and growth rate, the corresponding fitness for a mutant of type  $\mathbf{Y}$  invading in a polymorphic Lotka–Volterra community  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$  is found as in Eq. (30), by combining the definitions of its dynamics (28) and of  $s$ -functions:

$$s_{\mathbf{X}}(\mathbf{Y}) = r(\mathbf{Y}) \left( 1 - \sum_i a(\mathbf{Y}, \mathbf{X}_i) \hat{n}_i \right) \quad (31)$$

Then we simply solve the equilibrium densities  $\hat{n}_i$  from the growth equations and find that

$$s_{\mathbb{X}}(\mathbf{Y}) = r(\mathbf{Y}) \left( 1 - (a(\mathbf{Y}, \mathbf{X}_1) \ a(\mathbf{Y}, \mathbf{X}_2) \ \cdots \ a(\mathbf{Y}, \mathbf{X}_N)) \mathbf{A}^{-1} \mathbf{1} \right) \tag{32}$$

where  $\mathbf{A}$  is the interaction matrix for the given community, with entries  $[\mathbf{A}]_{ij} := a(\mathbf{X}_i, \mathbf{X}_j)$ , and we recall that  $\mathbf{1}$  is a column vector of 1's (cf. 1.3).

From Eq. (32) we see that except for the non-Lotka–Volterra case, there will in general not exist a well-defined interaction function  $a(\mathbf{Y}, \mathbf{X})$  that satisfies this equation for all communities and invaders:

**Proposition 2** *Proposition 1 does not hold if the words single-resident are replaced by  $N$ -resident.*

*Proof* Equation (32) shows that Lotka–Volterra systems only allow pairwise interactions (that are scaled by a specific type of density regulation). Any multiresident  $s$ -function that fails these requirements can therefore serve as a counterexample. In principle, the only constraint on  $s$ -functions is that they have to satisfy the following consistency conditions [49]: zero fitness for each of the residents (i.e.,  $s_{\mathbb{X}}(\mathbf{X}_i) = 0$  for all  $i$ ) and invariance under the renaming of residents (i.e.,  $s_{\mathbf{X}_i \mathbf{X}_j}(\mathbf{Y}) = s_{\mathbf{X}_j \mathbf{X}_i}(\mathbf{Y})$  for all  $i, j$ ). The simplest example would be

$$s_{\mathbf{X}_1 \mathbf{X}_2}(\mathbf{Y}) := (\mathbf{X}_1 - \mathbf{Y})(\mathbf{X}_2 - \mathbf{Y})$$

where the reader can verify that no choice of growth rate and interaction function will lead to a Lotka–Volterra model with this two-resident  $s$ -function. A slightly less caricatural example starts from the fitness function of an  $N$ -resident Lotka–Volterra model (31), and adds interaction terms between triples of strategies

$$s_{\mathbb{X}}(\mathbf{Y}) := r(\mathbf{Y}) \left( 1 - \sum_i a(\mathbf{Y}, \mathbf{X}_i) \hat{n}_i - \sum_{ij} b(\mathbf{Y}, \mathbf{X}_i, \mathbf{X}_j) \hat{n}_i \hat{n}_j \right)$$

through an appropriate function  $b(\mathbf{Y}, \mathbf{X}, \mathbf{X}')$ . For nontrivial choices of  $b$ , it is clearly impossible to account for the above fitness function by using a Lotka–Volterra model. □

How to relate  $N$ -resident Lotka–Volterra and physiologically structured population models instead, will be the central question of this section. To address it we return our attention to the simple fitness function (30) we found, which can be expanded in the small parameter  $\varepsilon$  as



$$\begin{aligned}
s_X(\mathbf{Y}) &= r(\mathbf{X}^* + \mathbf{V}) (1 - a(\mathbf{X}^* + \mathbf{V}, \mathbf{X}^* + \mathbf{U})) \\
&= \left( r(\mathbf{X}^*) + r'(\mathbf{X}^*)\mathbf{V} + \frac{1}{2}\mathbf{V}^\top r''(\mathbf{X}^*)\mathbf{V} + O(\varepsilon^3) \right) \\
&\quad \times \left( 1 - \alpha - \beta_1\mathbf{U} - \beta_0\mathbf{V} - \mathbf{U}^\top\Gamma_{11}\mathbf{U} - 2\mathbf{U}^\top\Gamma_{10}\mathbf{V} - \mathbf{V}^\top\Gamma_{00}\mathbf{V} + O(\varepsilon^3) \right) \\
&= r(\mathbf{X}^*)(1 - \alpha) - (r(\mathbf{X}^*) (\beta_1\mathbf{U} + \beta_0\mathbf{V}) + r'(\mathbf{X}^*)\mathbf{V}(1 - \alpha)) \\
&\quad - r(\mathbf{X}^*) (\mathbf{U}^\top\Gamma_{11}\mathbf{U} + 2\mathbf{U}^\top\Gamma_{10}\mathbf{V} + \mathbf{V}^\top\Gamma_{00}\mathbf{V}) + r'(\mathbf{X}^*)\mathbf{V} (\beta_1\mathbf{U} + \beta_0\mathbf{V}) \\
&\quad + \frac{1}{2}\mathbf{V}^\top r''(\mathbf{X}^*)\mathbf{V}(1 - \alpha) + O(\varepsilon^3) \tag{33}
\end{aligned}$$

were all terms of the same order in  $\varepsilon$  are grouped together.

As  $\Gamma_{11}$  and  $\Gamma_{00}$  are always pre- and postmultiplied by the same vector, their anti-symmetric parts are irrelevant. Thus there is an equivalence class of matrix choices for which the evaluation of Expansion (33) is the same, and from this class we choose a unique element by demanding that  $\Gamma_{11}$  and  $\Gamma_{00}$  are symmetric. As an aside we note that while it is highly nongeneric for  $\Gamma_{10}$  to be symmetric as well, this phenomenon happens often in simple models: either as a result of special symmetries (cf. our example, Sect. 4.6), or since the model is formulated so that the environmental input is effectively one-dimensional, and monotonically influences the invasion fitness (cf. [49]).

Several consistency conditions can be used to simplify Eq. (33). As a result of its definition,  $s_X(\mathbf{X})$  is zero for any value of  $\mathbf{X}$ . So for any  $\mathbf{U} = \mathbf{V}$ , the four parts of the right hand side of (33)—constant, linear, quadratic and higher order in  $\varepsilon$ —must be separately zero. Without loss of generality we may assume that  $r(\mathbf{X}^*)$  is strictly positive, as else the singular type would not be viable. The constant, linear and quadratic parts of the equation then respectively imply that  $\alpha = 1$ ,  $\beta_1 = -\beta_0$  and  $\Gamma_{11} + \Gamma_{10} + \Gamma_{10}^\top + \Gamma_{00} = \mathbf{0}$ .

Since  $\mathbf{X}^*$  is singular, by definition  $\boldsymbol{\theta}^\top = \frac{\partial s_{\mathbf{X}^*}(\mathbf{Y})}{\partial \mathbf{Y}} \Big|_{\mathbf{Y}=\mathbf{X}^*} = -r(\mathbf{X}^*)\beta_0$ , so  $-\beta_1 = \beta_0 = \boldsymbol{\theta}^\top$ . We rename the matrices using  $\mathbf{C} := -r(\mathbf{X}^*)\boldsymbol{\Gamma}$  so that the expansion (33) simplifies to

$$s_X(\mathbf{Y}) = \mathbf{U}^\top\mathbf{C}_{11}\mathbf{U} + 2\mathbf{U}^\top\mathbf{C}_{10}\mathbf{V} + \mathbf{V}^\top\mathbf{C}_{00}\mathbf{V} + O(\varepsilon^3) \tag{34}$$

From this we see that renaming and rescaling the  $\boldsymbol{\Gamma}$ -matrices into the  $\mathbf{C}$ -matrices was consistent with the earlier definition (26) of those as second order partial derivatives at the singularity.

We can now start considering  $N$ -resident invasion fitness functions close to singular points. Starting from Eq. (31), we see that we can express much of the multiresident

$s$ -function immediately in terms of single-resident  $s$ -functions:

$$\begin{aligned}
 s_{\mathbb{X}}(\mathbf{Y}) &= r(\mathbf{Y}) \left( 1 - \sum_i a(\mathbf{Y}, \mathbf{X}_i) \hat{n}_i \right) \\
 &= r(\mathbf{Y}) \left( 1 - \sum_i \left( 1 - \frac{s_{\mathbf{X}_i}(\mathbf{Y})}{r(\mathbf{Y})} \right) \hat{n}_i \right) \\
 &= r(\mathbf{Y}) \left( 1 - \sum_i \hat{n}_i \right) + \sum_i s_{\mathbf{X}_i}(\mathbf{Y}) \hat{n}_i \tag{35}
 \end{aligned}$$

We will now expand this last equality up to but not including  $O(\varepsilon^3)$ -terms. In view of the considerations at the start of this section, we change our coordinates from densities  $\hat{n}_i$  to fractional densities  $p_i$  plus the difference in total density from the monomorphic equilibrium density:

$$p_i := \frac{\hat{n}_i}{\sum_j \hat{n}_j} \quad \Delta n := \sum_i \hat{n}_i - 1 \tag{36}$$

Note that the constant term of  $\Delta n$  is zero since  $\varepsilon = 0$  corresponds to a monomorphic community  $\mathbb{X} = \{\mathbf{X}^*\}$ . Introducing a shorthand notation,

$$c(\mathbf{U}, \mathbf{V}) := \mathbf{U}^\top \mathbf{C}_{11} \mathbf{U} + 2\mathbf{U}^\top \mathbf{C}_{10} \mathbf{V} + \mathbf{V}^\top \mathbf{C}_{00} \mathbf{V} \tag{37}$$

we see that terms like  $c(\mathbf{U}_i, \mathbf{V}) \Delta n$  will be discarded, since  $c(\mathbf{U}, \mathbf{V})$  itself is already purely second order in  $\varepsilon$ . Using the new coordinates, we see that

$$s_{\mathbb{X}}(\mathbf{Y}) = - (r(\mathbf{X}^*) + \mathbf{r}'(\mathbf{X}^*) \mathbf{V}) \Delta n + \sum_i c(\mathbf{U}_i, \mathbf{V}) p_i + O(\varepsilon^3)$$

From the above we also note that only the constant part of the fractions  $p_i$  matters in the calculation of  $s_{\mathbb{X}}(\mathbf{Y})$  up to the given order. We expand the density difference as  $\Delta n = e_1 \varepsilon + e_2 \varepsilon^2 + O(\varepsilon^3)$ . Since  $s_{\mathbb{X}}(\mathbf{X}_i)$  is zero for each resident, we have for each  $i \in \{1, 2, \dots, N\}$  that

$$0 = -r(\mathbf{X}^*)(e_1 \varepsilon + e_2 \varepsilon^2) - \mathbf{r}'(\mathbf{X}^*) \mathbf{U}_i e_1 \varepsilon + \sum_j c(\mathbf{U}_j, \mathbf{U}_i) p_j + O(\varepsilon^3) \tag{38}$$

From the part that is linear in  $\varepsilon$ , we see that  $e_1$  too is zero, and from the quadratic part we have that  $r(\mathbf{X}^*)e_2 \varepsilon^2 = \sum_j c(\mathbf{U}_j, \mathbf{U}_i) p_j$ . Thus  $N + 1$  unknowns  $(p_1, p_2, \dots, p_N$

and  $e_2$ ) have to be solved using the consistency condition  $\sum_i p_i = 1$  plus the requirement that for each  $i$  from 1 to  $N$

$$\sum_j \underbrace{2\mathbf{U}_j^T \mathbf{C}_{10} \mathbf{U}_i}_{[\mathbf{E}]_{ij}} \underbrace{p_j}_{(\mathbf{P})_j} + \underbrace{\sum_j \mathbf{U}_j^T \mathbf{C}_{11} \mathbf{U}_j p_j}_{\theta} - r(\mathbf{X}^*) e_2 \varepsilon^2 = \underbrace{-\mathbf{U}_i^T \mathbf{C}_{00} \mathbf{U}_i}_{(\mathbf{T})_i} \quad (39)$$

Together these equations contain the componentwise definitions of the scalar  $\theta$ , the column vectors  $\mathbf{T}$  and  $\mathbf{P}$ , and the matrix  $\mathbf{E}$ . We can also gather together all  $N$  equations into a single vectorial one, using the vector  $\mathbf{I}$  that has all its components equal to one (cf. 1.3 Notations). The fact that the proportions necessarily sum up to 1 gives us an additional (scalar) equation, so we have altogether  $N + 1$  equations in  $N + 1$  unknowns:

$$\begin{cases} \mathbf{E} \mathbf{P} + \theta \mathbf{I} = \mathbf{T} \\ \mathbf{I}^T \mathbf{P} = 1 \end{cases} \quad (40)$$

If we treat  $\theta$  as an unknown (equivalent to the unknown  $e_2$  once  $\mathbf{P}$  is solved), these are linear equations. Hence we extend  $\mathbf{E}$ ,  $\mathbf{P}$  and  $\mathbf{T}$  to

$$\mathbf{E}^* := \begin{bmatrix} \mathbf{E} & \mathbf{I} \\ \mathbf{I}^T & 0 \end{bmatrix} \quad \mathbf{P}^* := \begin{pmatrix} \mathbf{P} \\ \theta \end{pmatrix} \quad \mathbf{T}^* := \begin{pmatrix} \mathbf{T} \\ 1 \end{pmatrix}$$

so that we can straightforwardly solve  $\theta$  and the proportions  $p_i$  in terms of second order derivatives of simple  $s$ -functions from

$$\mathbf{P}^* = \mathbf{E}^{*-1} \mathbf{T}^* \quad (41)$$

to come to the final conclusion that

$$\begin{aligned} s_{\mathbb{X}}(\mathbf{Y}) &= -r(\mathbf{X}^*) \Delta n + \sum_i c(\mathbf{U}_i, \mathbf{V}) p_i + O(\varepsilon^3) \\ &= \theta + 2 \left( \sum_i p_i \mathbf{U}_i^T \right) \mathbf{C}_{10} \mathbf{V} + \mathbf{V}^T \mathbf{C}_{00} \mathbf{V} + O(\varepsilon^3) \end{aligned} \quad (42)$$

where each term or factor is expressed in second order partial derivatives of the simple  $s$ -function, or a strategy difference vector ( $\mathbf{U}_i$  or  $\mathbf{V}$ , of respectively a resident or the invader), since  $\theta$  and the proportions are solved from

$$\begin{pmatrix} p_1 \\ \vdots \\ p_N \\ \theta \end{pmatrix} = \begin{bmatrix} 2\mathbf{U}_1^T \mathbf{C}_{10} \mathbf{U}_1 & \cdots & 2\mathbf{U}_N^T \mathbf{C}_{10} \mathbf{U}_1 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 2\mathbf{U}_1^T \mathbf{C}_{10} \mathbf{U}_N & \cdots & 2\mathbf{U}_N^T \mathbf{C}_{10} \mathbf{U}_N & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{pmatrix} -\mathbf{U}_1^T \mathbf{C}_{00} \mathbf{U}_1 \\ \vdots \\ -\mathbf{U}_N^T \mathbf{C}_{00} \mathbf{U}_N \\ 1 \end{pmatrix} \quad (43)$$

The invertibility of the matrix  $E^*$  is clearly an important issue here. It will be treated in Sect. 3.6 (and touched upon in 3.5), but the gist is that generically  $E^*$  is invertible if the community  $\{X_1, X_2, \dots, X_N\}$  exists.

### 3.4 The normal form for physiologically structured population models

As explained in Sect. 1.2, the equilibrium equations for a physiologically structured community are

$$\begin{cases} b_i = L(X_i, I)b_i & (\forall i) \\ I = \sum_i G(X_i, I)b_i \end{cases} \tag{44}$$

In Appendix B we show that if the residents and the invader are near a singularity, the invasion fitness is

$$s_{\mathbb{X}}(Y) = \frac{\log R_0(Y, I)}{T_f(Y, I)} + O(\varepsilon^3) \tag{45}$$

where  $R_0$  is the dominant eigenvalue  $\lambda_d(L)$  of the next-generation matrix  $L$ ,  $I$  the equilibrium environment set by the community  $\mathbb{X} := \{X_1, \dots, X_N\}$ , and  $T_f$  the average age at giving birth (cf. Eq. (97)).

As before, we will use an invertible, trait-dependent rescaling. In this case, we do not rescale population densities at equilibrium to 1 (while compensating by rescaling the interaction function, or vice versa) as these do not appear in the equilibrium equations. Instead we rescale the birth flow such that, for the monomorphic equilibrium community set by any strategy  $X$  in the trait space,

$$b = b^* \tag{46}$$

where  $b^*$  is the equilibrium birth flow for a community with only the singular strategy  $X^*$  present. We do this by defining for each strategy  $X$  the rescaled birth flow  $\tilde{b} := D b$  where  $D$  is the diagonal  $d \times d$  matrix with components  $[D]_{ll} := b_l^*/b_l$ , where  $b_l$  is the  $l$ th component of the unscaled equilibrium birth flow in the monomorphic community set by  $X$ . This transformation clearly ensures that Eq. (46) is satisfied. If all components of  $b^*$  are strictly positive, there is a neighbourhood of the singularity in which the birth flow  $b_l$  in each state is nonzero, so the matrix  $D$  is well-defined. The invertibility of the rescaling is guaranteed as well if all components of  $b^*$  are strictly positive. So we assume henceforth that  $b_l^* > 0$ , which we can do essentially without loss of generality since models flouting this assumption should be rare indeed. As in the Lotka–Volterra case (29), we compensate the first rescaling by rescaling the interaction; here by choosing  $\tilde{L} := D L D^{-1}$  and  $\tilde{G} := G D^{-1}$ . The matrices  $\tilde{L}$  and  $L$  necessarily have the same eigenvalues, hence the rescaling does not affect  $s_{\mathbb{X}}(Y)$  while it allows us to greatly simplify the calculations. From here on we revert to the old notations while assuming the rescaling has happened.

To expand a structured population’s invasion fitness function (45) near a singularity, we have to look at the lower orders of dependence on  $\varepsilon$  for all unknowns. To that end, we start by defining  $I_i$  as the monomorphic environment set solely by strategy  $X_i$ , so that  $I_i = G(X_i, I_i)b^*$  (note that the rescaling has been used here). We then expand

respectively the polymorphic environment set by  $\mathbb{X}$  and the monomorphic environment set by  $X_i$  as follows:

$$\begin{aligned}
 I &= I^* + \varepsilon I' + \varepsilon^2 I'' + O(\varepsilon^3) \\
 \forall i: I_i &= I_i^* + \varepsilon I_i' + \varepsilon^2 I_i'' + O(\varepsilon^3)
 \end{aligned}
 \tag{47}$$

In order to establish a relation between the  $N$ -resident environment  $I$  and its  $N$  monomorphic counterparts  $I_1, I_2, \dots, I_N$ , we introduce first some new coordinates, similar to those we used in the Lotka–Volterra case (36). We will need to calculate the relative abundance of each type of resident in the community. But as we now look from a generational perspective, we define this time a vector  $p_i$  that is the proportional abundance at birth of the  $i$ th type in the respective birth states, plus a difference vector  $\Delta b$  that is the proportional change in total births from the monomorphic equilibrium: for each birth state from 1 to  $d$  and for each resident from 1 to  $N$ ,

$$\forall l, \forall i: (p_i)_l := \frac{(b_i)_l}{\sum_j (b_j)_l} \quad 1 + (\Delta b)_l := \frac{\sum_j (b_j)_l}{b_l^*}
 \tag{48}$$

We expand the  $N$  proportion vectors  $p_i$  and  $\Delta b$  with respect to  $\varepsilon$  as

$$\begin{aligned}
 \forall i: p_i &= p_i^0 + q_i \varepsilon + O(\varepsilon^2) \\
 \Delta b &= e_0 + e_1 \varepsilon + O(\varepsilon^2)
 \end{aligned}
 \tag{49}$$

defining vectors  $e_0, e_1, p_1^0, p_2^0, \dots, p_N^0, q_1, q_2, \dots, q_N$  in the process. As happened with the density difference  $\Delta n$  in the Lotka–Volterra case (Eq. (36)), the constant part of the birth flow difference automatically disappears:  $e_0 = \mathbf{0}$  since  $\varepsilon = 0$  corresponds to a monomorphic case. Note that for each birth state  $l$  separately these proportions sum up to one, since  $\sum_i (p_i)_l = \sum_i (b_i)_l / \sum_j (b_j)_l = 1$ . Expanding both sides of these equalities with respect to  $\varepsilon$ , we find two times  $d$  consistency conditions

$$\sum_i p_i^0 = \mathbf{1} \quad \sum_i q_i = \mathbf{0}
 \tag{50}$$

As we only have the equilibrium equations (44) to start from, let’s begin by expanding all parts of the first equation:

$$(b_i)_l = (p_i)_l \sum_j (b_j)_l = ((p_i^0)_l + (q_i)_l \varepsilon) (1 + (e_1)_l \varepsilon) b_l^* + O(\varepsilon^2)
 \tag{51}$$

$$L(X_i, I) = L(X^*, I^*) + \frac{\partial L}{\partial Y}(\varepsilon \xi_i) + \frac{\partial L}{\partial I}(\varepsilon I') + O(\varepsilon^2)
 \tag{52}$$

where, e.g.,  $\frac{\partial L}{\partial I}(\varepsilon I')$  is the  $d \times d$  matrix with entries  $\left( \frac{\partial L(X^*, I)_{lm}}{\partial I} \Big|_{I=I^*} \right) \varepsilon I'$  in accordance with the conventions introduced in Sect. 3.1.

From Eqs. (51) and (52) we see that  $\mathbf{b}_i = \mathbf{L}(X_i, \mathbf{I})\mathbf{b}_i$  can be rewritten for each  $i$  as

$$\begin{aligned}
 & ((\mathbf{p}_i^o)_l + \varepsilon(\mathbf{q}_i)_l + \varepsilon(\mathbf{p}_i^o)_l(\mathbf{e}_1)_l) b_l^* + O(\varepsilon^2) \\
 &= \sum_m \left[ \mathbf{L}(X^*, \mathbf{I}^*) + \frac{\partial \mathbf{L}}{\partial \mathbf{Y}}(\varepsilon \boldsymbol{\xi}_i) + \frac{\partial \mathbf{L}}{\partial \mathbf{I}}(\varepsilon \mathbf{I}') \right]_{lm} \left( (\mathbf{p}_i^o)_m + \varepsilon(\mathbf{q}_i)_m + \varepsilon(\mathbf{p}_i^o)_m(\mathbf{e}_1)_m \right) b_m^* \\
 &= \sum_m [\mathbf{L}(X^*, \mathbf{I}^*)]_{lm} (\mathbf{p}_i^o)_m b_m^* + \sum_m \varepsilon \left[ \frac{\partial \mathbf{L}}{\partial \mathbf{Y}}(\boldsymbol{\xi}_i) + \frac{\partial \mathbf{L}}{\partial \mathbf{I}}(\mathbf{I}') \right]_{lm} (\mathbf{p}_i^o)_m b_m^* \\
 &+ \sum_m [\mathbf{L}(X^*, \mathbf{I}^*)]_{lm} (\varepsilon(\mathbf{q}_i)_m + \varepsilon(\mathbf{p}_i^o)_m(\mathbf{e}_1)_m) b_m^* \tag{53}
 \end{aligned}$$

As this equality has to hold for all  $\varepsilon$ , it has to hold for all orders of  $\varepsilon$  separately. Thus the constant part tells us that for each  $i$  the vector with components  $(\mathbf{p}_i^o)_l b_l^*$  is an eigenvector of  $\mathbf{L}(X^*, \mathbf{I}^*)$  with eigenvalue 1. Since this eigenvalue was assumed to be simple, with corresponding eigenvector  $\mathbf{b}^*$ , necessarily there must exist some scalars  $p_i$  such that

$$\forall i: \mathbf{p}_i^o = p_i \mathbf{I} \tag{54}$$

This fact helps us simplify the part of Eq. (53) that is linear in  $\varepsilon$ . We can transform it further by summing over  $i$ , so that the  $\mathbf{q}_i$ -components disappear (50):

$$\begin{aligned}
 (\mathbf{e}_1)_l b_l^* &= \sum_i \sum_m \left[ \frac{\partial \mathbf{L}}{\partial \mathbf{Y}}(\boldsymbol{\xi}_i) + \frac{\partial \mathbf{L}}{\partial \mathbf{I}}(\mathbf{I}') \right]_{lm} p_i b_m^* \\
 &+ \sum_m [\mathbf{L}(X^*, \mathbf{I}^*)]_{lm} (\mathbf{e}_1)_m b_m^* \tag{55}
 \end{aligned}$$

If we define a ‘‘help’’ vector  $\mathbf{h}_{\text{eb}}$  componentwise as  $(\mathbf{h}_{\text{eb}})_l := (\mathbf{e}_1)_l b_l^*$ , we have

$$\mathbf{h}_{\text{eb}} = \sum_i p_i \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i + \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \mathbf{I}' + \mathbf{L}(X^*, \mathbf{I}^*) \mathbf{h}_{\text{eb}} \tag{56}$$

$$\mathbf{0} = \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i + \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \mathbf{I}'_i \tag{57}$$

where the second equation is the monomorphic case, for which we scaled the equilibrium birth flow to  $\mathbf{b}^*$  so  $\mathbf{e}_1 = \mathbf{h}_{\text{eb}} = \mathbf{0}$ .

Let’s now look at the second equilibrium equation,  $\mathbf{I} = \sum_i \mathbf{G}(X_i, \mathbf{I})\mathbf{b}_i$ . If we here too expand both  $\mathbf{I}$  and  $\mathbf{b}_i$  while using the new coordinates (48), we can combine the per-state summation rules (50) with the fact that  $\mathbf{p}_i^o = p_i \mathbf{I}$ , to find that

$$\begin{aligned}
 & \mathbf{I}^* + \varepsilon \mathbf{I}' + O(\varepsilon^2) \\
 &= \sum_i \left[ \mathbf{G}(\mathbf{X}^*, \mathbf{I}^*) + \frac{\partial \mathbf{G}}{\partial \mathbf{I}}(\varepsilon \mathbf{I}') + \frac{\partial \mathbf{G}}{\partial \mathbf{Y}}(\varepsilon \boldsymbol{\xi}_i) \right] \begin{pmatrix} (p_i + \varepsilon(\mathbf{q}_i)_1 + \varepsilon p_i(\mathbf{e}_1)_1) b_1^* \\ \vdots \\ (p_i + \varepsilon(\mathbf{q}_i)_d + \varepsilon p_i(\mathbf{e}_1)_d) b_d^* \end{pmatrix} \\
 &= \mathbf{G}(\mathbf{X}^*, \mathbf{I}^*) \mathbf{b}^* + \varepsilon \mathbf{G}(\mathbf{X}^*, \mathbf{I}^*) \mathbf{h}_{\text{eb}} + \varepsilon \frac{\partial \mathbf{G}}{\partial \mathbf{I}}(\mathbf{I}') \mathbf{b}^* + \sum_i p_i \varepsilon \frac{\partial \mathbf{G}}{\partial \mathbf{Y}}(\boldsymbol{\xi}_i) \mathbf{b}^* \quad (58)
 \end{aligned}$$

The part of this equation that is independent of  $\varepsilon$  does not tell us anything new, but the part that is linear in  $\varepsilon$  gives the following relation:

$$\mathbf{I}' = \mathbf{G}(\mathbf{X}^*, \mathbf{I}^*) \mathbf{h}_{\text{eb}} + \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \mathbf{I}' + \sum_i p_i \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i$$

From this we solve  $\mathbf{I}'$  and  $\mathbf{I}'_i$  (where again  $\mathbf{e}_1 = \mathbf{0}$  so  $\mathbf{h}_{\text{eb}}$  disappears), as

$$\mathbf{I}' = \left[ \text{id} - \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \right]^{-1} \left( \mathbf{G}(\mathbf{X}^*, \mathbf{I}^*) \mathbf{h}_{\text{eb}} + \sum_i p_i \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i \right) \quad (59)$$

$$\mathbf{I}'_i = \left[ \text{id} - \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \right]^{-1} \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i \quad (60)$$

where  $\text{id}$  represents the identity matrix, and taking the inverse is allowed as this transformation is nonsingular provided we stay away from bifurcation points of the population dynamics.

With the shorthand notations  $\mathbf{L}^* := \mathbf{L}(\mathbf{X}^*, \mathbf{I}^*)$  and  $\mathbf{G}^* := \mathbf{G}(\mathbf{X}^*, \mathbf{I}^*)$ , we have as a consequence of Eqs. (59) and (60) that

$$\mathbf{I}' = \sum_i p_i \mathbf{I}'_i + \left[ \text{id} - \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \right]^{-1} \mathbf{G}^* \mathbf{h}_{\text{eb}} \quad (61)$$

which we substitute into Eq. (56) so that we can use Equality (57):

$$\begin{aligned}
 [\text{id} - \mathbf{L}^*] \mathbf{h}_{\text{eb}} &= \sum_i p_i \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i + \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \mathbf{I}' \\
 &= \sum_i p_i \left( \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{Y}} \boldsymbol{\xi}_i + \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \mathbf{I}'_i \right) + \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \left[ \text{id} - \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \right]^{-1} \mathbf{G}^* \mathbf{h}_{\text{eb}} \\
 &= \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \left[ \text{id} - \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \right]^{-1} \mathbf{G}^* \mathbf{h}_{\text{eb}} \quad (62)
 \end{aligned}$$

After a slight rewrite we find that

$$\mathbf{0} = \left[ [\mathbf{L}^* - \text{id}] + \frac{\partial \mathbf{L} \mathbf{b}^*}{\partial \mathbf{I}} \left[ \text{id} - \frac{\partial \mathbf{G} \mathbf{b}^*}{\partial \mathbf{I}} \right]^{-1} \mathbf{G}^* \right] \mathbf{h}_{\text{eb}} \tag{63}$$

As there is no a priori, fixed connection between  $\mathbf{L}$  and  $\mathbf{G}$ —changing one without changing the other (while not violating consistency conditions) results in an equally valid population model—we see that the matrix in this equation generically (within the set of local equivalence classes of models characterized by  $\mathbf{L}$ ,  $\mathbf{G}$ ,  $\partial \mathbf{L} / \partial \mathbf{I}$  and  $\partial \mathbf{G} / \partial \mathbf{I}$ ) has full rank. In the special case of a single birth state model, clearly  $\text{id} = \mathbf{L}^* = 1$ , and hence  $\mathbf{e}_1 = \mathbf{e}_1 = 0$ . In general,  $\mathbf{L}^* - \text{id}$  always has rank  $d - 1$  as  $[\mathbf{L}^* - \text{id}] \mathbf{b}^* = \mathbf{0}$  and this eigenvalue is simple. Other examples corroborating the intuition that the matrix in Eq. (63) generically is invertible, are models with  $\mathbf{G}(X_i, \mathbf{I})$  independent of  $\mathbf{I}$ .

Therefore  $\mathbf{e}_1 = \mathbf{h}_{\text{eb}} = \mathbf{0}$  is generically the only possible solution, since all entries of  $\mathbf{b}^*$  are strictly positive as argued in the justification of the birth flow rescaling (46). Hence Eq. (61) shows that the relation we sought between the linear parts of the environments  $\mathbf{I}_i$  and  $\mathbf{I}$  is simply

$$\mathbf{I}' = \sum_i \mathbf{I}'_i p_i \tag{64}$$

With this, we can formulate an expansion of the multitype  $s$ -function at the singularity  $\mathbf{X}^*$ , up to  $O(\varepsilon^3)$ . For that, we start by recalling the single-resident invasion fitness for structured population models (45),

$$s_{X_i}(\mathbf{Y}) = \frac{\log \lambda_d(\mathbf{L}(\mathbf{X}^* + \mathbf{V}, \mathbf{I}^* + \varepsilon \mathbf{I}'_i + O(\varepsilon^2)))}{T_f(\mathbf{Y}, \mathbf{I}_i)} + O(\varepsilon^3)$$

First we note that the dominant eigenvalue of  $\mathbf{L}(\mathbf{X}^*, \mathbf{I}^*)$  is one, so that the numerator has no constant part (with respect to  $\varepsilon$ ). Therefore, only the constant part  $T_f^* := T_f(\mathbf{X}^*, \mathbf{I}^*)$  of the denominator appears in a first order expansion of  $s_X(\mathbf{Y})$ :

$$\begin{aligned} s_{X_i}(\mathbf{Y}) &= \frac{\lambda_d(\mathbf{L}(\mathbf{X}^* + \mathbf{V}, \mathbf{I}^* + \varepsilon \mathbf{I}'_i)) - 1}{T_f(\mathbf{Y}, \mathbf{I}_i)} + O(\varepsilon^2) \\ &= \frac{1}{T_f^*} \left( \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{Y}} \mathbf{V} + \varepsilon \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{I}} \mathbf{I}'_i \right) + O(\varepsilon^2) \end{aligned} \tag{65}$$

As we are expanding at a singular point, we have  $\partial s / \partial \mathbf{Y} = \partial \lambda_d(\mathbf{L}) / \partial \mathbf{Y} = \mathbf{0}^\top$ . But because  $s_X(\mathbf{X}) = 0$  for any  $\mathbf{X}$ , the linear terms of (65) must add up to zero when choosing  $\mathbf{V} = \mathbf{U}_i$ —or if you prefer, they add up to zero as a consequence of Eq. (57), since for simple eigenvalues  $\partial \lambda_d(\mathbf{L}) / \partial f = (\mathbf{v}^{*\top} \mathbf{b}^*)^{-1} \mathbf{v}^{*\top} [\partial \mathbf{L} / \partial f] \mathbf{b}^*$  for any  $f$ , where  $\mathbf{v}^{*\top}$  is a left eigenvector of  $\mathbf{L}$  with eigenvalue 1 (cf. Eq. (96)). Hence we also have  $0 = (\partial \lambda_d(\mathbf{L}) / \partial \mathbf{I}) \mathbf{I}'_i$ , and because of (64) also the more general  $0 = (\partial \lambda_d(\mathbf{L}) / \partial \mathbf{I}) \mathbf{I}'$ . This shows that invasion fitness functions for  $N$ -resident communities, expanded at singularities, do not have linear terms either.



As we have just shown that in an expansion of  $\lambda_d(\mathbf{L})$  no linear terms appear, we see that in a quadratic expansion of the  $N$ -resident fitness function, only the constant part of the denominator  $T_f$  and the first order part of the logarithm will have to appear:

$$\begin{aligned}
 s_{\mathbb{X}}(\mathbf{Y}) &= \frac{\log \lambda_d(\mathbf{L}(\mathbf{X}^* + \mathbf{V}, \mathbf{I}^* + \varepsilon \mathbf{I}' + \varepsilon^2 \mathbf{I}''))}{T_f(\mathbf{Y}, \mathbf{I})} + O(\varepsilon^3) \\
 &= \frac{\lambda_d(\mathbf{L}(\mathbf{X}^* + \mathbf{V}, \mathbf{I}^* + \varepsilon \mathbf{I}' + \varepsilon^2 \mathbf{I}'')) - 1}{T_f^*} + O(\varepsilon^3) \\
 &= \frac{1}{T_f^*} \left( \varepsilon^2 \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{I}} \mathbf{I}'' + \frac{1}{2} \varepsilon^2 \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I}^2} \mathbf{I}' \right. \\
 &\quad \left. + \varepsilon \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I} \partial \mathbf{Y}} \mathbf{V} + \frac{1}{2} \mathbf{V}^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{Y}^2} \mathbf{V} \right) + O(\varepsilon^3) \tag{66}
 \end{aligned}$$

If we compare this expansion, in the case of a single resident  $X_i$ , with the Taylor series of the simple  $s$ -function at a singularity,  $s_{X_i}(\mathbf{Y}) = \frac{1}{2} \mathbf{U}_i^\top \frac{\partial^2 s}{\partial \mathbf{X}^2} \mathbf{U}_i + \mathbf{U}_i^\top \frac{\partial^2 s}{\partial \mathbf{X} \partial \mathbf{Y}} \mathbf{V} + \frac{1}{2} \mathbf{V}^\top \frac{\partial^2 s}{\partial \mathbf{Y}^2} \mathbf{V} + O(\varepsilon^3)$ , we can associate the partial derivatives of simple  $s$ -functions with the terms just found:

$$\begin{aligned}
 \mathbf{U}_i^\top \frac{\partial^2 s}{\partial \mathbf{X}^2} \mathbf{U}_i &= \frac{1}{T_f^*} \left( 2\varepsilon^2 \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{I}} \mathbf{I}'' + \varepsilon^2 \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I}^2} \mathbf{I}' \right) \\
 \mathbf{U}_i^\top \frac{\partial^2 s}{\partial \mathbf{X} \partial \mathbf{Y}} \mathbf{V} &= \frac{1}{T_f^*} \varepsilon \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I} \partial \mathbf{Y}} \mathbf{V} \\
 \mathbf{V}^\top \frac{\partial^2 s}{\partial \mathbf{Y}^2} \mathbf{V} &= \frac{1}{T_f^*} \mathbf{V}^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{Y}^2} \mathbf{V}
 \end{aligned} \tag{67}$$

To simplify the notation and to bring out the similarity to the Lotka–Volterra case (34), we use the matrices  $\mathbf{C}_{11}$ ,  $\mathbf{C}_{00}$  and  $\mathbf{C}_{10}$  introduced before (cf. Eq. (26)), which consist of the second order partial derivatives of  $s_{\mathbf{X}}(\mathbf{Y})$  at  $\mathbf{X} = \mathbf{Y} = \mathbf{X}^*$ . In addition, we use a shorthand notation

$$\theta := \frac{\varepsilon^2}{T_f^*} \left( \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{I}} \mathbf{I}'' + \frac{1}{2} \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I}^2} \mathbf{I}' \right) \tag{68}$$

Since for each resident necessarily  $s_{\mathbb{X}}(\mathbf{X}_i) = 0$ , we see from combining the expansion of  $s_{\mathbb{X}}(\mathbf{Y})$  (66) with Equality (64) that for each  $i$

$$\begin{aligned}
 0 &= \varepsilon^2 \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{I}} \mathbf{I}'' + \frac{1}{2} \varepsilon^2 \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I}^2} \mathbf{I}' + \varepsilon^2 \mathbf{I}'^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I} \partial \mathbf{Y}} \boldsymbol{\xi}_i + \frac{1}{2} \varepsilon^2 \boldsymbol{\xi}_i^\top \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{Y}^2} \boldsymbol{\xi}_i \\
 &= \mathbf{T}_f^* \left( \theta + 2 \sum_j \mathbf{U}_j^\top \mathbf{C}_{10} \mathbf{U}_i p_j + \mathbf{U}_i^\top \mathbf{C}_{00} \mathbf{U}_i \right) \tag{69}
 \end{aligned}$$

As for the Lotka–Volterra case (39), we can formulate these  $N$  equalities as

$$\forall i: \quad \sum_j \underbrace{2\mathbf{U}_j^\top \mathbf{C}_{10} \mathbf{U}_i}_{[\mathbf{E}]_{ij}} \underbrace{p_j}_{(\mathbf{P})_j} + \theta = \underbrace{-\mathbf{U}_i^\top \mathbf{C}_{00} \mathbf{U}_i}_{(\mathbf{T})_i} \tag{70}$$

From here onwards, we can paraphrase all steps that led us to the result in the Lotka–Volterra case, since we have the same set of equations and all terms and factors have exactly the same meaning.

To reiterate succinctly, our set of  $N$  equations (70) can be used to define componentwise an  $N \times N$  matrix  $\mathbf{E}$  and vectors  $\mathbf{P}$  and  $\mathbf{T}$ , so that we can write the  $N$  equations in a vectorial form:  $\mathbf{E} \mathbf{P} + \theta \mathbf{I} = \mathbf{T}$ . Treating  $\theta$  as an independent unknown, we have  $N + 1$  linear equations (since additionally we know  $\mathbf{I}^\top \mathbf{P} = \sum_i p_i = 1$ ) in the  $N + 1$  unknowns  $\theta$  and  $p_1, p_2, \dots, p_N$ . We then gather the vectorial and the scalar equation together by extending  $\mathbf{E}$ ,  $\mathbf{P}$  and  $\mathbf{T}$  as

$$\mathbf{E}^* := \begin{bmatrix} \mathbf{E} & \mathbf{I} \\ \mathbf{I}^\top & 0 \end{bmatrix} \quad \mathbf{P}^* := \begin{pmatrix} \mathbf{P} \\ \theta \end{pmatrix} \quad \mathbf{T}^* := \begin{pmatrix} \mathbf{T} \\ 1 \end{pmatrix} \tag{71}$$

so that  $\theta$  and the proportions  $p_i$  are solved from  $\mathbf{P}^* = \mathbf{E}^{*-1} \mathbf{T}^*$  using only the  $\mathbf{C}$ -matrices and the strategy differences  $\mathbf{U}_i$ . Componentwise this gives

$$\begin{pmatrix} p_1 \\ \vdots \\ p_N \\ \theta \end{pmatrix} = \begin{bmatrix} 2\mathbf{U}_1^\top \mathbf{C}_{10} \mathbf{U}_1 & \cdots & 2\mathbf{U}_N^\top \mathbf{C}_{10} \mathbf{U}_1 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 2\mathbf{U}_1^\top \mathbf{C}_{10} \mathbf{U}_N & \cdots & 2\mathbf{U}_N^\top \mathbf{C}_{10} \mathbf{U}_N & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{pmatrix} -\mathbf{U}_1^\top \mathbf{C}_{00} \mathbf{U}_1 \\ \vdots \\ -\mathbf{U}_N^\top \mathbf{C}_{00} \mathbf{U}_N \\ 1 \end{pmatrix} \tag{72}$$

The issue of the invertibility of  $\mathbf{E}^*$  will be explored in Sect. 3.6. Note however, that a full rank of  $\mathbf{E}^*$  is a necessity for the structurally stable existence of the community  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$ .

Finally, we have to cast the second order approximation of  $s_{\mathbb{X}}(\mathbf{Y})$  close to  $\mathbf{X}^*$  (45) in the form we found for Lotka–Volterra systems (42). To that end, we use Equality (66), the relationship  $\mathbf{I}' = \sum_i p_i \mathbf{I}'_i$  (64), and the definitions of  $\theta$  and the  $\mathbf{C}$ -matrices. Then we find the following form for the invasion fitness function of any structured

population model with  $N$  resident types:

$$\begin{aligned}
 s_{\mathbb{X}}(\mathbf{Y}) &= \frac{\log(\lambda_d(\mathbf{L}(\mathbf{Y}, \mathbf{I})))}{T_f(\mathbf{Y}, \mathbf{I})} + O(\varepsilon^3) \\
 &= \frac{1}{T_f^*} \left( \varepsilon^2 \frac{\partial \lambda_d(\mathbf{L})}{\partial \mathbf{I}} \mathbf{I}'' + \frac{1}{2} \varepsilon^2 \mathbf{I}'^T \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I}^2} \mathbf{I}' \right) \\
 &\quad + \frac{1}{T_f^*} \left( \varepsilon \mathbf{I}'^T \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{I} \partial \mathbf{Y}} \mathbf{V} + \frac{1}{2} \mathbf{V}^T \frac{\partial^2 \lambda_d(\mathbf{L})}{\partial \mathbf{Y}^2} \mathbf{V} \right) + O(\varepsilon^3) \\
 &= \theta + 2 \left( \sum_i p_i \mathbf{U}_i^T \right) \mathbf{C}_{10} \mathbf{V} + \mathbf{V}^T \mathbf{C}_{00} \mathbf{V} + O(\varepsilon^3) \tag{73}
 \end{aligned}$$

This is exactly the same equation as we found for the invasion fitness in Lotka–Volterra models (42) and leads us to the following proposition:

**Proposition 3** *The invasion fitness function of a general physiologically structured population model with  $N$  resident strategies near an evolutionarily singular strategy, can be approximated by that of an  $N$ -resident Lotka–Volterra system, up to terms of order  $O(\varepsilon^3)$  for distances between residents and singularity of order  $O(\varepsilon)$ .*

*Proof* By comparing the multiresident  $s$ -functions of the Lotka–Volterra case (42–43) with the physiologically structured population case (72–73), it is seen that both depend in exactly the same way on the second order derivatives of the single-resident  $s$ -function at  $\mathbf{X}^*$  (i.e., the  $\mathbf{C}$ -matrices) and the deviations  $\mathbf{U}_i$  of the resident strategies  $\mathbf{X}^* + \mathbf{U}_i$  from the singular strategy. Applying Proposition 1, we can fit a Lotka–Volterra model with the same  $N$ -resident fitness function as the given physiologically structured population model, up to quadratic terms in  $\varepsilon$ . □

As remarked before, adding the requirement that the fitted Lotka–Volterra model has a trait-independent growth rate  $r$  in virgin environments, makes the approximating system unique since  $a(\mathbf{Y}, \mathbf{X}) = 1 - s_{\mathbf{X}}(\mathbf{Y})/r$ .

Proposition 3 may be read as follows: *The  $s$ -function for  $N$ -resident Lotka–Volterra models (31) is a second order normal form for multiresident fitness functions, since for any given structured population model for which we can write down the simple invasion function  $s_{\mathbf{X}}(\mathbf{Y})$ , we can easily fit a Lotka–Volterra model by defining the growth rate in a virgin environment and the interaction function as*

$$\forall \mathbf{X}, \mathbf{Y}: \quad r(\mathbf{Y}) := 1, \quad a(\mathbf{Y}, \mathbf{X}) := 1 - s_{\mathbf{X}}(\mathbf{Y}) \tag{74}$$

The single-resident fitness function of this Lotka–Volterra model is exactly the same as that of the given model, and the multiresident  $s$ -function for Lotka–Volterra models (31) was found to be

$$s_{\mathbb{X}}(\mathbf{Y}) = 1 - (a(\mathbf{Y}, \mathbf{X}_1) \cdots a(\mathbf{Y}, \mathbf{X}_N)) \mathbf{A}^{-1} \mathbf{I}$$

where the interaction matrix  $A$  was defined as

$$A := \begin{bmatrix} a(X_1, X_1) & a(X_1, X_2) & \cdots & a(X_1, X_N) \\ a(X_2, X_1) & a(X_2, X_2) & \cdots & a(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ a(X_N, X_1) & a(X_N, X_2) & \cdots & a(X_N, X_N) \end{bmatrix} \tag{75}$$

with necessarily each diagonal element  $a(X_i, X_i)$  equal to one.

Proposition 3 guarantees that the difference between this polymorphic fitness function and the correct function for the given population model is of order  $O(\varepsilon^3)$ .

### 3.5 The case of two residents

We found a normal form for fitness functions that is generally applicable to any  $N$ -resident Lotka–Volterra (42) or physiologically structured (73) community near a singularity. It is however not a very perspicacious relation between the model ingredients.

If there are only two residents, we can come to a more insightful form by choosing as coordinates  $\bar{U} := (U_1 + U_2)/2$  and  $\Delta := (U_1 - U_2)/2$  (so conversely  $U_1 = \bar{U} + \Delta$  and  $U_2 = \bar{U} - \Delta$ ). Translating  $T^*$  and  $E^*$ , we have

$$T^* := \begin{pmatrix} -U_1^T C_{00} U_1 \\ -U_2^T C_{00} U_2 \\ 1 \end{pmatrix} = \begin{pmatrix} -\bar{U}^T C_{00} \bar{U} - 2\Delta^T C_{00} \bar{U} - \Delta^T C_{00} \Delta \\ -\bar{U}^T C_{00} \bar{U} + 2\Delta^T C_{00} \bar{U} - \Delta^T C_{00} \Delta \\ 1 \end{pmatrix}$$

and (after some computing)

$$E^{*-1} = -\frac{1}{8\Delta^T C_{10} \Delta} \begin{bmatrix} -1 & 1 & 4\bar{U}^T C_{10} \Delta - 4\Delta^T C_{10} \Delta \\ 1 & -1 & -4\bar{U}^T C_{10} \Delta - 4\Delta^T C_{10} \Delta \\ 4\Delta^T C_{10} \bar{U} & -4\Delta^T C_{10} \bar{U} & 16\Delta^T C_{10} \Delta \bar{U}^T C_{10} \bar{U} \\ -4\Delta^T C_{10} \Delta & -4\Delta^T C_{10} \Delta & -16\bar{U}^T C_{10} \Delta \Delta^T C_{10} \bar{U} \end{bmatrix}$$

By adding the second row or column to the first, we can calculate that  $\det E^{*-1} = -8\Delta^T C_{10} \Delta$ . Thus we find  $p_1$ ,  $p_2$  and  $\theta$  from  $E^{*-1} T^*$  to be

$$\begin{pmatrix} p_1 \\ p_2 \\ \theta \end{pmatrix} = \begin{pmatrix} \frac{1}{2} - \frac{1}{2} \frac{\bar{U}^T C_{00} \Delta + \bar{U}^T C_{10} \Delta}{\Delta^T C_{10} \Delta} \\ \frac{1}{2} + \frac{1}{2} \frac{\bar{U}^T C_{00} \Delta + \bar{U}^T C_{10} \Delta}{\Delta^T C_{10} \Delta} \\ -\Delta^T C_{00} \Delta - \bar{U}^T C_{00} \bar{U} - 2\bar{U}^T C_{10} \bar{U} \\ + 2 \frac{\Delta^T C_{10} \bar{U} \Delta^T [C_{00} + C_{10}^T] \bar{U}}{\Delta^T C_{10} \Delta} \end{pmatrix} \tag{76}$$

Substituting these three in the normal form (73) presented before, we find

$$\begin{aligned} s_{X_1 X_2}(Y) &= V^T C_{00} V + 2\bar{U}^T C_{10} V - \bar{U}^T [C_{00} + 2C_{10}] \bar{U} - \Delta^T C_{00} \Delta \\ &\quad + 2\Delta^T C_{10} (\bar{U} - V) \frac{\Delta^T [C_{00} + C_{10}^T] \bar{U}}{\Delta^T C_{10} \Delta} + O(\varepsilon^3) \end{aligned} \tag{77}$$

which is Eq. (27) as presented in Sect. 3.2, ‘‘Aims’’. As an aid to the reader, we remark that most of the correspondence between the general (73) and the two-resident normal form (77) is seen by observing that

$$\sum_i p_i U_i^T = \bar{U}^T - \frac{\Delta^T [C_{00} + C_{10}^T] \bar{U}}{\Delta^T C_{10} \Delta} \Delta^T$$

For two strategies close to  $X^*$ , to be mutually invadable (a requirement for stable coexistence) implies that  $\Delta^T C_{10} \Delta$  is negative:

$$\begin{aligned} 0 &< s_{X_1}(X_2) + s_{X_2}(X_1) \\ &= U_1^T C_{11} U_1 + 2U_1^T C_{10} U_2 + U_2^T C_{00} U_2 + O(\varepsilon^3) \\ &\quad + U_2^T C_{11} U_2 + 2U_2^T C_{10} U_1 + U_1^T C_{00} U_1 + O(\varepsilon^3) \\ &= 2\bar{U}^T [C_{11} + C_{00}] \bar{U} + 2\Delta^T [C_{11} + C_{00}] \Delta \\ &\quad + 2\bar{U}^T [C_{10} + C_{10}^T] \bar{U} - 2\Delta^T [C_{10} + C_{10}^T] \Delta + O(\varepsilon^3) \\ &= -8\Delta^T C_{10} \Delta + O(\varepsilon^3) \end{aligned} \tag{78}$$

Equation (77) shows that the relative densities can at least be calculated, as long as the denominator  $\Delta^T C_{10} \Delta$  is nonzero. However, this does not amount to coexistence when one of the proportions  $p_i$  is negative. From (76) we see that coexistence (the positiveness of both  $p_1$  and  $p_2$ ) is equivalent to

$$\left| \frac{\bar{U}^T C_{00} \Delta + \bar{U}^T C_{10} \Delta}{\Delta^T C_{10} \Delta} \right| < 1 \tag{79}$$

To see how likely it is that this inequality is fulfilled, consider the case where  $X_1, X_2$  and  $X^*$  are collinear, so  $U_2 = \alpha U_1$  for some value of  $\alpha$ . Typically residents will be on opposite sides of the singularity (corresponding to a negative  $\alpha$ ), although shortly after a branching event we may find them on the same side (positive  $\alpha$ ). We see that

$$E = 2U_1^T C_{10} U_1 \begin{bmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{bmatrix} \quad \det E^* = -2U_1^T C_{10} U_1 (\alpha - 1)^2 \tag{80}$$

so there is no unique solution if  $\alpha$  is one;  $T^*$  then lies in the range of  $E^*$  and the linear system  $E^* P^* = T^*$  is underdetermined. This much was expected (cf. Fig. 1) since the residents are indistinguishable in this case and coexist at a neutrally stable equilibrium, their relative abundances dependent on initial conditions.

Thus for a nonsingular  $E^*$ , the condition for true coexistence is

$$\left| \frac{1 + \alpha}{1 - \alpha} \right| < \left| \frac{U_1^T C_{10} U_1}{U_1^T C_{00} U_1 + U_1^T C_{10} U_1} \right| = \left| \frac{1 + \frac{U_1^T C_{11} U_1}{U_1^T C_{00} U_1}}{1 - \frac{U_1^T C_{11} U_1}{U_1^T C_{00} U_1}} \right| \tag{81}$$

By plotting the left hand side of the inequality, we can draw some conclusions. For two residents to lie on the same side of a singularity ( $\alpha > 0$ ), it is necessary that the right hand side of the inequality is larger than one. The pole at  $\alpha = 1$  shows that the closer two such residents are, the less likely it is that the condition is satisfied. On the other hand, we see that  $\alpha = -1$  is always a solution, and the closer two residents are to being each other’s opposite ( $U_1 \approx -U_2$ ), the likelier it is that the condition is fulfilled.

If  $X^*$  is invadable,  $C_{00}$  has positive eigenvalues. We can then choose  $U_1$  such that  $U_1^T C_{00} U_1 > 0$ . Moreover, we concluded that  $U_1^T C_{10} U_1$  is negative (cf. Eq. (78)), so necessarily the right hand side of Inequality (81) is larger than one and any  $\alpha < 0$  suffices. If however  $X^*$  is uninvadable, the right hand side of the inequality is smaller than one and only a narrow interval around  $\alpha = -1$  will lead to coexistence.

### 3.6 Limits to the level of local polymorphism

We have solved  $\theta$  and the proportions  $P$  by inverting  $E^*$ . To justify this approach, we show first that by excluding singular matrices  $E^*$ , we have only excluded structurally unstable communities.

Let us first point out that the frame of reference in Propositions 4 and 5 and Lemmata 3–6 below, is the set of all systems in the product of the space of strategy deviations  $\{U_1, U_2, \dots, U_N\}$  and the space of local equivalence classes of communities specified by the matrices  $C_{11}, C_{10}$  and  $C_{00}$ ; within this context the qualification “generically” must be understood.

**Proposition 4** *Generically, if the strategies  $\{X^* + \varepsilon \xi_i \mid i = 1, 2, \dots, N\}$  can coexist in the limit  $\varepsilon \rightarrow 0$ , then  $E^*$  is nonsingular.*

*Proof* If  $E^*$  is singular, there exists a nontrivial vector  $\mu$  such that

$$(\mu_1 \mu_2 \cdots \mu_N \mu_{N+1})E^* = \mathbf{0}^T$$

or in another form

$$\begin{cases} (\mu_1 \mu_2 \cdots \mu_N)E = -\mu_{N+1}I^T \\ (\mu_1 \mu_2 \cdots \mu_N)I = 0 \end{cases}$$

If we now apply the equalities  $1 = I^T P$  and  $E P = T - \theta I$  (which we found for both Lotka–Volterra (40) and physiologically structured models (70)), we see that

$$-\mu_{N+1} = -\mu_{N+1}I^T P = (\mu_1 \cdots \mu_N)E P = (\mu_1 \cdots \mu_N)T$$

which is generically a contradiction since  $\mu$  only relates to  $E^*$  whereas  $T$  has no direct connection to  $E^*$  since they depend on different C-matrices that can be varied independently, by slight changes in the model specification. □

We will now formulate some results about the influence of the system ingredients on the invertibility of  $E^*$  (Lemma 1–4). Proposition 4 then provides a recipe for translating those results into upper bounds to the possible complexity of communities (Proposition 5–Lemma 6). Finally, Proposition 7 relates these results to some very general ideas floating around in the literature about the abstract generalization of the theorem by Levin [42] that  $N$  limiting resources can robustly support the coexistence of at most  $N$  types.

**Lemma 1** *If  $\det E \neq 0$ , then  $\det E^* = -(\det E)I^T E^{-1}I$ .*

**Lemma 2** *If  $\text{rank } E \leq N - 2$ , then  $E^*$  is singular.*

**Lemma 3** *If  $\text{rank } E = N - 1$ , then generically  $E^*$  is invertible.*

**Lemma 4** *If  $\text{rank } E = N$ , then generically  $E^*$  is invertible.*

The first three lemmata are proven in Appendix D, while the last statement is verified by a look at Lemma 1 plus the realization that the sum  $I^T E^{-1}I$  of all the elements of the inverse matrix is typically nonzero.

**Proposition 5** *Generically, the number of residents  $N$  near a singularity is at most one higher than the dimension  $m$  of the trait vectors.*

This proposition generalizes a result of Christiansen and Loeschke [11].

*Proof* First we define an  $m \times N$  trait matrix  $U := [U_1 U_2 \cdots U_N]$  from the trait vectors (or componentwise  $[U]_{ai} := (U_i)_a$ ). We then see that  $E$  is a product of matrices,

$$[E]_{ij} := 2U_j^T C_{10} U_i = [2U^T C_{10}^T U]_{ij}$$

As the rank of a product of matrices is never higher than the rank of any of its constituent matrices, the rank of  $E$  is at least two below maximal if  $N \geq m + 2$ , in which case  $E^*$  is not invertible according to Lemma 2 and the conclusion follows from Proposition 4. □

**Lemma 5** *When two resident strategies are linearly dependent on the other  $N - 2$  in a set of strategies close to a singularity, this set of  $N$  strategies generically cannot coexist.*

*Proof* In this case  $U$  has a rank at least two below  $N$ , and by Lemma 2 this holds for  $E$  too; applying Proposition 4 concludes the proof.  $\square$

**Lemma 6** *Near a singularity, the number of residents is generically at most one higher than the rank of  $C_{10}$ .*

This is a corollary of Lemma 2 and Proposition 4 as well.

**Lemma 7** *Even if both  $U$  and  $C_{10}$  have full rank,  $E^*$  may still be singular.*

This statement is proven in Appendix D. We remark however that a small perturbation of  $U$  and/or  $C_{10}$  will suffice to make  $E^*$  invertible.

We will now point out in Proposition 7 an indirect connection between  $E = 2U^T C_{10} U$  and the dimension of the feedback environment [17,46]. As a corollary, Proposition 5 may be recovered, as at most  $N$  types can stably coexist in an  $N$ -dimensional environment [47,46]. As a preliminary we show how to find the exceptions to the rule that Lotka–Volterra models generate infinite dimensional environments:

**Proposition 6** *A Lotka–Volterra model with interaction function  $a(Y, X)$  has feedback dimension  $e$ , if and only if  $e$  is the minimal number for which we can write  $a(Y, X) = \sum_{q=1}^e b_q(Y) c_q(X)$ , for some functions  $b_q$  and  $c_q$ .*

A proof is given in Appendix E.

**Proposition 7** (local environmental dimension) *Near singularities, the environment is locally at most  $(z + 1)$ -dimensional when the strategies are  $z$ -dimensional.*

*Proof* In this section we showed that an adaptive dynamics model near a singularity can be approximated, up to quadratic terms, by a Lotka–Volterra model with any interaction function of the form

$$a(X^* + V, X^* + U) := 1 - U^T C_{11} U - 2U^T C_{10} V - V^T C_{00} V + O(\varepsilon^3)$$

and growth rate  $r(Y) := 1$ , where the  $C$ -matrices are determined by the model to be approximated. The fitted model then has the same  $s$ -function as the original model, up to quadratic terms in  $\varepsilon$ . Making the specific choice

$$a(X^* + V, X^* + U) := 1 - U^T C_{11} U - 2U^T C_{10} V - V^T C_{00} V + U^T C_{11} U V^T C_{00} V$$

and using the following definitions,

$$\begin{aligned} b_0(Y) &:= 1 - (Y - X^*)^T C_{00} (Y - X^*) \\ c_0(X) &:= 1 - (X - X^*)^T C_{11} (X - X^*) \\ \forall a \in \{1, \dots, z\}: b_a(Y) &:= 2 \sum_b [C_{10}]_{ab} (Y - X^*)_b \\ c_a(X) &:= (X - X^*)_a \end{aligned}$$



we see that  $a(Y, X) = \sum_{a=0}^z b_a(Y)c_a(X)$ . Proposition 6 then shows us that the feedback dimension of the approximating model is at most  $z + 1$ .  $\square$

#### 4 The meaning of it all

What is the relation between the canonical equation (Sect. 2) and the fitness function near singularities (Sect. 3)? How can they aid in interpreting a model? Or in other words, why do these sections appear together in a single paper? We will address these questions here, by describing a recipe for analysing concrete models and illustrating it with an example from the literature.

The starting point should be a model that satisfies the assumptions of adaptive dynamics: a large system size and a population dynamics where the individuals have heritable life history parameters that are under evolutionary control through a low rate of mutations with relatively small effect. In this situation, the ecology typically is dominated by a few, markedly different, strategy vectors. The distribution of the strategies thus has a few distinct peaks, with almost all individuals having a trait very close to the position of one of the peaks. If there are one or two such trait values, for example, the community is termed *quasi-monomorphic* or *quasi-dimorphic*. Interesting evolutionary dynamics are those where the number of peaks increases over time, as such increasing diversity and specialization has an obvious interpretation as (the onset of) (sym- or parapatric) speciation. Of course, it is just as important to know when this buildup of diversity does not and/or cannot occur.

The evolutionary analysis of such a model starts by setting up a resident population at equilibrium, with all individuals of a single type. Typically the trait finds itself under directional selection and evolves as predicted by the canonical equation (23). The community is then quasi-monomorphic at an evolutionary timescale, as the appearance of a successful mutant is immediately followed by the disappearance of the former resident (cf. Sect. 2.2, “invasion implies substitution”). This substitutional regime only comes to an end when the population finds itself near a singular trait value (cf. Introduction 1.1). Near singular trait values the invasion/replacement dynamics acts differently with other behaviours possible: resident traits may stop evolving, or the quasi-monomorphic population may diversify into a quasi-dimorphic community after a brief polymorphic phase. If we find that the population does branch into several populations with differing strategies, these branches must be followed until they go extinct or until they are a safe distance away from the singularity, so that one knows whether several types of residents coexist in a protected manner or whether the polymorphism quickly disappears again. Once the branches have evolved out of the proximity of the singularity, a canonical equation again governs the evolution of the subpopulation associated with each branch, and the analysis can proceed as before.

This process is possibly repeated, with further evolutionary branching, until all branches have reached evolutionary endpoints (i.e., attracting, uninvadable trait values): only then a final evolutionary outcome has been found. In some cases however, evolution does not proceed towards a definite endpoint but ends up on a limit cycle or on a more complicated attractor (e.g., [16]).

A recipe for studying the evolutionary behaviour of a model, from random initial resident until the final outcome(s), can be broken down in the following steps.

### 4.1 Model (re)formulation

The adaptive dynamics approach requires that the basic model assumptions are formulated in terms of the behaviour of individuals. These must possess near-faithfully inherited traits [49] influencing their reproduction, survival, change in spatial and physiological states, and impact on their environment. (The environment was introduced in Sect. 1.1 and described further in 1.2. It is a concept of which the utility lies in the fact that in a given environment the growth dynamics of any clone is linear with an asymptotic relative growth rate depending on both the strategy and the environment.) Such a description requires i.a. postulates for how individuals convert resources into offspring (depending on the state and inherited strategy of the organism, and the state of the resources), but also for the dynamics of the resources consumed by individuals. Therefore the vector of heritable traits will appear as a parameter of the individual dynamics (and hence of the population dynamics), and the population distribution will appear as an input of the environmental dynamics.

### 4.2 Life history parameters

Once the model is formulated in individual-based terms, the necessary life history parameters can be calculated: for a resident with strategy  $X$  in the environment  $I$ , there is the expected lifetime offspring production  $L(X, I)$  and the feedback matrix  $G(X, I)$  (cf. Sect. 1.2), the life expectancy  $T_s(X, I)$  and the average age-at-giving-birth  $T_f(X, I)$  (97), the equilibrium population densities  $T_s(X, I) \|\mathbf{b}\|$  where  $\mathbf{b}$  is the birth flow vector (1), and the offspring variance  $\sigma^2$  (16). Notice that these are all observables, with a clear biological interpretation.

From Appendix A we know that the invasion fitness of a mutant that resembles the  $i$ th resident type has the following form in general:

$$s_{\mathbb{X}}(X_i + V) = \frac{\log \lambda_d(L(X_i + V, I_{\mathbb{X}}))}{T_f(X_i, I_{\mathbb{X}})} + O(\|V\|^2)$$

where  $\lambda_d$  is the dominant eigenvalue operator. Notice how in this context one can avoid solving the characteristic equation to find  $\lambda_d(L)$ : the dominant eigenvalue of  $L$  for any resident  $X_i$  is necessarily 1. Therefore the right- and left-eigenvectors  $\mathbf{b}_i$  and  $\mathbf{v}_i^T$  of  $L$  are the (unique, up to a scalar) solutions to  $[\text{id} - L(X_i, I_{\mathbb{X}})] \mathbf{b}_i = \mathbf{0}$  and  $\mathbf{v}_i^T [\text{id} - L(X_i, I_{\mathbb{X}})] = \mathbf{0}^T$ . This allows us to approximate the  $a$ th component of the fitness gradient  $\left( \frac{\partial s_{\mathbb{X}}(Y)}{\partial Y_a} \Big|_{Y=X_i} \right)^T$  by  $\mathbf{v}_i^T \left[ \frac{\partial L(Y, I_{\mathbb{X}})}{\partial Y_a} \Big|_{Y=X_i} \right] \mathbf{b}_i / (T_f \mathbf{v}_i^T \mathbf{b}_i)$ . Hence we see that for the fitness gradient we only have to solve the next-generation and environmental feedback equations (1) for  $\mathbf{b}_i$  (and  $L(X_i, I_{\mathbb{X}})$ ), and to find  $\mathbf{v}_i$ .

For each  $i$ , the eigenvectors are normalized so that  $\|\mathbf{u}_i\| = 1$  and  $1 = \mathbf{v}_i^\top \mathbf{u}_i$  (13). This allows us to calculate  $T_f$  (97) and  $T_s$ ; the life expectancy for example is

$$T_s := \sum_i u_i \int_0^\infty F_i(\mathbf{X}, \mathbf{I}, a) da \quad (82)$$

with  $F_i(\mathbf{X}, \mathbf{I}, a)$  the probability that an individual with trait value  $\mathbf{X}$  and born in state  $i$ , survives to at least age  $a$ .

Depending on model type, the previously described steps in the analysis may be numerical or analytical. Clearly an analytical approach has many advantages, like showing how life histories depend on strategy parameters. However, even when this treatment is theoretically possible, it may be so cumbersome as to be unfeasible: finding the dominant eigenvalue of  $\mathbf{L}$  where there are three or four possible birth states would be a case in point.

We remark furthermore that this description of the second step of the analysis is written so as to be as general as possible. In practice life tends to be simpler, and often parts of this step may be skipped: e.g., for an ODE or difference equation model, the population dynamical equilibrium and per capita growth rate are found directly, making the explicit formulation of  $\mathbf{G}$  and  $\mathbf{L}$  (and its derivatives and eigenvectors) redundant. Furthermore the community will be starting from a single resident, so that  $1 = i = N$  and the number of equations to solve simultaneously may be low.

### 4.3 Monomorphic dynamics under directional selection

The canonical equation (23) predicts how the trait vectors in a community will change over time. This prediction is valid as long as the fitness gradient stays nonzero, allowing a deterministic approximation of the path a community will follow from any given initial state. Here we assume that the mutational covariance matrix is nonsingular. (It may become singular, for example when the traits under evolutionary control are subject to constraints. For trait vectors on the surface that then forms the boundary of the attainable trait space, the null space of the covariance matrix is locally orthogonal to that surface. The equation shows that at equilibrium the fitness gradient must lie in that null space. This scenario may be treated in the same way as that of a bounded trait space, discussed below). From the canonical equation, we see that a state is transient if the fitness gradient is nonzero. As it is derived from an approximation that fails in the proximity of singular points, the canonical equation does not help the analysis close to those interesting points towards which evolution drives the community. Paradoxically, we will use the (one-resident) canonical equation to find the strategy values where it fails as an approximation, and subsequently use the (multiple-resident) equation to predict how the community will evolve around those strategies.

A major exception to the above scenario occurs when the trait space is bounded in one or more directions. In that case, a distinction must be made between the dynamics tangential and orthogonal to the boundary. If the fitness gradient points outward at the boundary, the dynamics orthogonal to the boundary will trap any approaching

community. Hence the analysis can be continued in a lower-dimensional strategy space, looking for singularities of the dynamics constrained to the boundary. Some care must be taken if the fitness gradient points outward only along part of the boundary, as the community may evolve for a while along the boundary, but later stray away from it.

After locating those points where the fitness gradient is zero—the so-called *singular points* (1.1)—we can use the canonical equation to find out whether they are attracting or not, and if so, to find their basin of attraction. Singular points that are repelling or have inappreciable domains of attraction are clearly not that interesting, as the community will normally not approach such points. In principle, the equation can be as rich in dynamical features as any ODE. As such, the appropriate numerical tools, or theorems about the qualitative behaviour of solutions, will depend on the specifics of the model under consideration.

At this point we still lack one ingredient of the canonical equation: the mutational covariance matrix  $M(X)$  near the trait value  $X$  (cf. Eq. (2), (23)). The influence of  $M$  can be understood in the following way: selection impels traits to change in the direction of the steepest increase in fitness, but the covariances are changing the direction of movement away from this “optimal” direction. Covariance matrices are the great unknowns in evolutionary biology. There is little or no understanding of which choices of  $M$  (or, equivalently, of the mutational distribution  $\mathfrak{M}$ ) are the reasonable ones for each type of biological model, as the covariances are footprints of deeper developmental, physiological and biophysical processes (cf. [52]). The safest conclusion to draw is that any complicated dynamical features found from the canonical equation should be dismissed as biologically irrelevant, unless they are robust against changes in  $M$ . When there is no further information about it, most people in practice choose  $M = \text{id}$ , the identity matrix. However, it is never a bad idea to do simulations for a few other values of  $M$ . The best possible result would then be that one can classify the potential evolutionary outcomes as dependent on  $M$ . Then the theoretical analysis leads to a potentially empirically answerable question: “Is  $M$  expected to lie in this or that domain?”.

If we are satisfied that evolution is towards a singularity, whether of the full trait space or of a constrained subspace, we continue to the next step. If there is no such singularity, the community will stay monomorphic indefinitely while the resident trait keeps changing according to the canonical equation.

#### 4.4 Near singularities

When a resident strategy moves closer and closer to a singularity, the first order approximation of fitness by the selection gradient breaks down, and so do a number of other approximations that were made so far. Around a singularity, several zones can be distinguished in which different refinements come into play.

In the most convenient (and hence generally emphasized) case, there exists an outer zone where the second order terms start to dominate the  $s$ -function, while the radius of curvature of the local fitness contours is still large in comparison with the average mutational step length. We note here that the existence of such a region depends on the

separation of two scales: that of the distance of the residents from the singular point and that of the distance of the mutants from their resident progenitors, where the relative effects on the two scales have to be gauged through the  $s$ -function. Whether or not the singular point attracts throughout this outer zone can be determined through recourse to the canonical equation. The only difference with the earlier discussion is that here standard analytic tools from linearized stability theory can be used, as the second order terms of the invasion fitness function fully determine the local linearization of the canonical equation. The one snag is that the canonical equation still contains the mutational covariance matrix. Leimar [39–41] analyzed the extent to which the stability of a singular point can be determined without any knowledge of that matrix. In this analysis, the eigenvalues of the matrix  $\partial^2_{sX}(Y)/\partial Y^2 - \partial^2_{sX}(Y)/\partial X^2$  (or  $C_{00} - C_{11}$  in our earlier notation (26)) play the key role in the classification of the evolutionary possibilities relatively close to a singularity  $X^*$ . The possibilities are as follows:

- If  $C_{00} - C_{11}$  is positive definite (i.e., has only strictly positive eigenvalues),  $X^*$  is repelling. Thus the community can only find itself close to  $X^*$  if the initial resident was close to  $X^*$  as well. In such cases, the community will evolve away from the singularity and the canonical equation will quickly become valid, so that the analysis can proceed in accordance with 4.3.
- If  $C_{00} - C_{11}$  is negative definite (i.e., has only strictly negative eigenvalues), convergence to  $X^*$  is assured. In this case the next step of this recipe, Sect. 4.5, applies. Different outcomes are possible, as  $X^*$  may be an evolutionary endpoint or a branching point.
- If  $C_{00} - C_{11}$  is indefinite (i.e., has both strictly positive and negative eigenvalues), the covariance matrix  $M(X^*)$  is decisive. Firstly, for a given  $M$  the singularity may be a saddlepoint, so that at first the resident can approach  $X^*$  but later grow more and more distinct from  $X^*$ . Thus the regime of directional selection stays intact and the analysis of 4.3 applies. Secondly, for a given  $M$  the singularity may be an attractor so that the analysis can continue at the next step in this analysis. Thirdly, for a given  $M$  the singularity may be a repeller, and hence be disregarded as unattainable.

The model under consideration may constrain  $M$  to one of the three cases and thus simplify the analysis, or several cases may occur depending on parameter values.

Closer to the singular point where the curvature of the fitness contours starts to have its effects, it is also no longer possible to assume permanent quasi-monomorphism, as it may be that mutants are no longer able to oust their progenitor. A discussion of the final convergence to an uninvadable singular point (characterized by  $C_{00}$  being negative definite) under a still mutation limited regime, can be found in Appendix B of [17]. Although the problem of the final convergence is far from solved, in this case it turns out to be both independent of the mutational covariance matrix and an all-or-none property. Hence the problem can be solved in principle by a single extended simulation run, of a type similar to those described in the next subsection. A final problem is that sufficiently close to the singular point, the timescale of selective takeovers will in any concrete case become so slow that the assumption of mutation limitation will break down, thus necessitating an approach along the lines of quantitative genetics (cf. [53]).

#### 4.5 At attracting singularities

Once we know that the singularity  $X^*$  is an attractor, we can start wondering whether selection will be stabilizing (so that the resident distribution will forever have the shape of a single peak close to  $X^*$ ), or disruptive (so that distinct (sub)populations may form near  $X^*$  and evolve away from each other).

When  $C_{00}$  is negative definite then  $X^*$  is uninvadable and selection is stabilizing close to  $X^*$ . Hence the first requirement for branching to occur is that  $X^*$  must be invadable:  $C_{00}$  must have at least one positive eigenvalue. If increasingly differentiating polymorphisms indeed do arise, we once again have to distinguish between the zones discussed in the previous subsection.

For the zone where both the canonical equation and the second order approximation to the  $s$ -function hold, the following conjecture is floating around (Stefan Geritz, Lecture given at the International Conference on Computational and Mathematical Population Dynamics, Trento, 2004): in those cases where  $C_{00}$  has a simple positive maximal eigenvalue, there will remain only a single pair of branches if  $e_M^T [C_{00} + C_{11}] e_M > 0$ , where  $e_M$  is an eigenvector corresponding to the maximal eigenvalue; else all but one branch will eventually go extinct. When there are two co-maximal positive eigenvalues, then in principle three branches can grow away from each other at 120 degree angles, without any two of them driving the remaining third to extinction [56]. But even in the particular case of co-maximality, it appears that three-way splitting happens only under special conditions. (The authors welcome any counterexamples!)

As an aside we note that in finite populations, the largest positive eigenvalue of  $C_{00}$  has to be sufficiently large for branching to really occur. Moreover, several other processes may obstruct diversification. Consider for example a diploid, sexual organism and assume some diversity has arisen. If a male and female of differing types mate, their offspring will be of a type close to the average of the parental types. Hence unless a preference for assortative mating is present already, sexual reproduction and the recombination it engenders will tend to average out strategies and thus prevent the buildup of specialized subpopulations through what could be called the ‘‘Mendelian mixer’’.

Before we can with some confidence rely on the canonical equation, evolution has to get the resident community out of the region where the radius of curvature of the local fitness contours is small relative to the average mutational step length. It is in the analysis of what happens in this region that Sect. 3 is useful in our recipe: since (up to quadratic terms in the mutation step size) all models behave as Lotka–Volterra models (28) near singularities, we can fit such a model and study it instead of the original model. As explained after Proposition 3 and applied in the Example 4.6 below, to construct a Lotka–Volterra model with the same  $N$ -resident fitness function as the model under consideration, we merely have to define the per capita growth rate in a virgin environment and the interaction function as

$$\forall X, Y: \quad r(Y) := 1, \quad a(Y, X) := 1 - s_X(Y) \quad (83)$$

This model is now fitted so as to have the same single-resident fitness function as the original model, for any combination of resident and invader. As mentioned above, Sect. 3 shows that fitness functions of communities close to a singularity are

model-independent in a sense, so we can proceed our analysis with either model. Whatever the nature of the original model, the new one is an ODE model, so a possible advantage is that we can study its dynamics with standard packages. Furthermore Lotka–Volterra models are mathematically rather straightforward (e.g., [34]), as for example the equilibrium population sizes in a community of  $N$  types is easily calculated (Formula 32). Finally, stochastic simulations of their individual-based counterparts are easily performed using the Gillespie algorithm [30] (see e.g., the appendix to [16], and [9]).

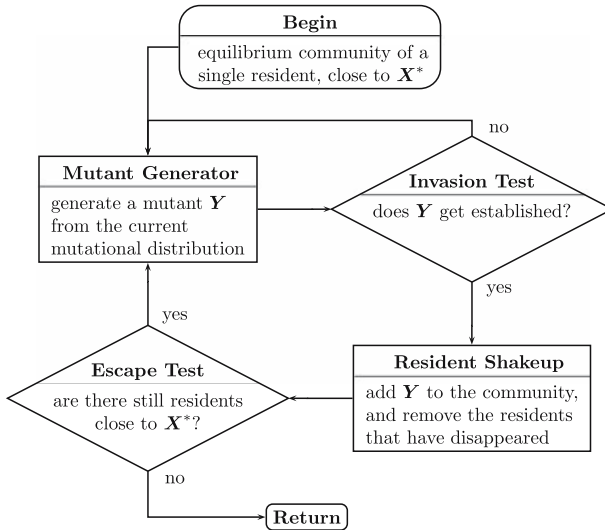
To explore the evolutionary behaviour of a system near an invadable attracting singularity  $\mathbf{X}^*$ , simulations can be run along the following lines:

1. Choose an initial resident near the singularity: for some small  $U_1$ , this resident's strategy is  $\mathbf{X}_1 = \mathbf{X}^* + U_1$ . The resident's density is set at its (nontrivial) equilibrium, and the total number  $N$  of resident types in the community is 1.
2. Draw a mutant from the mutational distribution associated with the community. To do this straightforwardly, we just have to think about the mutant's ancestor. If  $N$  types are present in the community at densities  $\hat{n}_i$ , then with probability  $p_i = \hat{n}_i / \sum_j \hat{n}_j$  the mutant derives from the  $i$ th resident. From the mutational step distribution  $\mathfrak{M}(\mathbf{X}_i)$  around this progenitor we draw a mutation step  $\mathbf{V}$ , generating a mutant of type  $\mathbf{Y} := \mathbf{X}_i + \mathbf{V}$ .

If the invasion fitness of the newly found mutant is negative, we discard it and draw another mutant according to the above recipe. Once we have a mutant with positive fitness, we have to see whether it will get established or not. For this we turn the Wheel of Fortune a second time, where the chance of success scales with the mutant's fitness as required by Equality (21). If our mutant is unlucky, we go back to generating mutants until one successfully invades.

3. See which of the original residents survive the invasion of our mutant. This is done by initializing the Lotka–Volterra ODEs at the community attractor, adding a small number of invaders—enough to avoid disappearance of the mutants by roundoff errors—and then following the community dynamics to its new equilibrium. After the transition from (close to) the equilibrium of the former to that of the new community, we will usually find that the mutant's progenitor has disappeared. In some situations however, several residents may have disappeared, while in the case of disruptive selection it is possible that no residents disappear at all. Thus the community  $\mathbb{X}$  may have lost residents  $\mathbf{X}_{d_1}, \mathbf{X}_{d_2}, \dots, \mathbf{X}_{d_k}$  but gained a new resident  $\mathbf{X}_{N-k+1} := \mathbf{Y}$ , where the equilibrium densities are easily found analytically or numerically, using the interaction function and the resident strategies (cf. Eq. 32).
4. Go back to Step 2, as long as there still are residents in a close neighbourhood of  $\mathbf{X}^*$ .

In some cases however, the singularity is invadable but does not allow a polymorphism to build up. This happens when the zone of mutual invadability (and hence stable coexistence) is so narrow that a mutant will be outside of this area after very few mutation steps, and therefore drive all close by residents to extinction. The trajectory will then hover close to the singularity, while neither really closing in on  $\mathbf{X}^*$  nor branching. Thus, if any polymorphism in the community only remains for a brief period while the trajectory makes no progress towards the nearby sin-



**Fig. 2** Steps in simulating the evolutionary dynamics close to a singularity

gularity within a reasonable time, the algorithm must be interrupted and  $X^*$  be proclaimed an evolutionary endpoint.

At the last step, there is no hard and fast rule to work out what a safe distance is for concluding that all branches present have evolved away from the proximity of the singularity. Out of hand, we would say ten mutation steps. The underlying idea is that away from a singularity, the “invasion implies substitution” dictum holds sway for each of the separate branches, so we can be confident that no buildup of diversity is possible beyond our chosen boundary. Similarly, what constitutes a “reasonable time” at Step 2 is not exactly defined; we would call it a day when a thousand successive mutants fail to invade. Given these inexactitudes, it is clear that our story relies on the common sense of the programmer. A summary of the algorithm as a flow diagram is given in Fig. 2.

If we are convinced there is no diversification occurring while the resident creeps ever closer to  $X^*$ , we label this strategy an evolutionary endpoint. If on the other hand branching has occurred, then the branches come under a regime of directional selection once they have outgrown the influence of  $X^*$ . Thus we find ourselves back at Sect. 4.3 to repeat the entire analysis, this time in a more complicated fashion since equilibria of the canonical equation for several residents have to be found, and equilibria of a community dynamics with several residents. The invasion analysis, however, keeps studying the singular points separately, with the other (faraway) strategists being part of the background that determines part of the environment  $I$  when examining the evolution of a resident.

To illustrate our recipe, we conclude with an example from the literature.



#### 4.6 A fitting example

As an example we consider a one-locus model for intraspecific competition [11]. Diploid additive genetics is assumed in a discrete time, non-overlapping generations analogue to Lotka–Volterra dynamics. There are multiple resources, and the trait under evolutionary control is the modus of the utilization function:

- Resources are distributed as a  $z$ -variate Gaussian with mean 0 and covariance matrix  $\Psi$ , where  $\Psi$  is real, symmetric and positive definite.
- Each existing allele  $A_i \in \{A_1, A_2, \dots, A_N\}$  has a trait value  $X_i$  associated with it, such that for a diploid with genotype  $(A_i, A_j)$ , the optimal resource type is  $\mathbf{D}_{ij} := \mathbf{D} + \mathbf{X}_i + \mathbf{X}_j$  and the diploid has a Gaussian resource utilization function  $U_{ij}(\mathbf{r}) := \alpha \exp\left(-\frac{1}{2}(\mathbf{D}_{ij} - \mathbf{r})^\top \Phi^{-1}(\mathbf{D}_{ij} - \mathbf{r})\right)$ . The scaling constants  $\alpha$  and  $\mathbf{D}$ , and the symmetrical, positive definite covariance matrix  $\Phi$  are shared by all genotypes.
- Diploid individuals  $(A_i, A_j)$  and  $(A_k, A_l)$  interact through the competition coefficient  $\gamma_{ij,kl} := \exp\left(-\frac{1}{4}(\mathbf{D}_{ij} - \mathbf{D}_{kl})^\top \Phi^{-1}(\mathbf{D}_{ij} - \mathbf{D}_{kl})\right)$  while the carrying capacity is given as  $k_{ij} := \exp\left(-\frac{1}{2}\mathbf{D}_{ij}^\top [\Phi + \Psi]^{-1} \mathbf{D}_{ij}\right)$ .

Genotypes are formed by random mating under free recombination. Between generations, a genotype  $(A_i, A_j)$  changes in abundance as

$$n_{ij}(t+1) = n_{ij}(t) \left( 1 + \beta \left( k_{ij} - \sum_{kl} \gamma_{ij,kl} n_{kl}(t) \right) \right) \quad (84)$$

for some scaling constant  $\beta$ .

For an adaptive dynamics analysis of this model, it is important to realize at this point that the alleles, not the diploids, are the individuals to consider. This is seen by contemplating who is faithfully replicating [49] in this community: in the extreme situation where the population is made up of homozygotes  $(A_1, A_1)$  and  $(A_2, A_2)$ , half of the offspring is of a brand new type so that the heterozygote per capita growth rate is infinite.

We can now interpret the model ingredients in an adaptive dynamics context. Firstly, the strategy of an allele  $A_i$  is the allelic trait value  $X_i$  associated with it. Secondly, for a rare allele  $A_j$  that is introduced in this monomorphic community, the number of alleles  $A_j$  is actually  $n_{ij}$  as mutant homozygotes are exceedingly rare. Thus we find the one-resident invasion fitness (i.e., the per capita growth rate of a rare mutant in a monomorphic equilibrium community) as the logarithm of the mutant heterozygote's growth rate. From these considerations plus the model ingredients and Eq. (84), we see that

$$s_X(\mathbf{Y}) = \log \left( 1 + \beta e^{-\frac{1}{2}(\mathbf{D} + \mathbf{X} + \mathbf{Y})^\top [\Phi + \Psi]^{-1} (\mathbf{D} + \mathbf{X} + \mathbf{Y})} - \beta e^{-\frac{1}{2}(\mathbf{D} + 2\mathbf{X})^\top [\Phi + \Psi]^{-1} (\mathbf{D} + 2\mathbf{X})} e^{-\frac{1}{4}(\mathbf{X} - \mathbf{Y})^\top \Phi^{-1} (\mathbf{X} - \mathbf{Y})} \right) \quad (85)$$

From the selection gradient

$$\left( \frac{\partial s_X(\mathbf{Y})}{\partial \mathbf{Y}} \Big|_{\mathbf{Y}=\mathbf{X}} \right)^\top = -\beta e^{-\frac{1}{2}(\mathbf{D}+2\mathbf{X})^\top [\Phi+\Psi]^{-1}(\mathbf{D}+2\mathbf{X})} [\Phi + \Psi]^{-1}(\mathbf{D} + 2\mathbf{X}) \tag{86}$$

we see that the unique singularity lies at  $\mathbf{X}^* := -\mathbf{D}/2$ , so we translate the origin of our coordinate system to  $\mathbf{X}^*$  and use the (not necessarily small) strategy difference vectors  $\mathbf{U} := \mathbf{D}/2 + \mathbf{X}$  and  $\mathbf{V} := \mathbf{D}/2 + \mathbf{Y}$ .

The singularity is globally attracting, as can be seen in the following way. First we note that the inverse of the sum of real, positive definite, symmetric matrices has these three properties as well, so that  $0 < \mathbf{X}^\top [\Phi + \Psi]^{-1} \mathbf{X}$  for any nontrivial vector  $\mathbf{X}$ . If we look at the canonical equation (23), we see that it predicts the resident trait to change according to

$$\frac{d\mathbf{X}}{dt} = \alpha(\mathbf{X})\mathbf{M}(\mathbf{X}) \frac{\partial s_X(\mathbf{Y})}{\partial \mathbf{Y}} \Big|_{\mathbf{Y}=\mathbf{X}} =: f(\mathbf{X}) \tag{87}$$

for some positive function  $\alpha(\mathbf{X})$ . Considering only small mutations in the direction of the singularity (i.e.,  $\mathbf{V} := (1 - \gamma)\mathbf{U}$  with  $\gamma > 0$  and  $\mathbf{V} - \mathbf{U} = O(\varepsilon)$ ), we see that such a mutant has positive fitness, while a step away from the singularity ( $\gamma < 0$ ) implies negative fitness:

$$\begin{aligned} s_{\mathbf{X}^*+\mathbf{U}}(\mathbf{X}^* + \mathbf{V}) &= 0 + \frac{\partial s_X(\mathbf{Y})}{\partial \mathbf{Y}} \Big|_{\mathbf{Y}=\mathbf{X}^*+\mathbf{U}} (-\gamma \mathbf{U}) + O(\varepsilon^2) \\ &= 2\beta \gamma e^{-2\mathbf{U}^\top [\Phi+\Psi]^{-1} \mathbf{U}} \mathbf{U}^\top [\Phi + \Psi]^{-1} \mathbf{U} + O(\varepsilon^2) > 0 \end{aligned} \tag{88}$$

Given this inequality, it is easy to find a Lyapunov function for  $f(\mathbf{X})$ : e.g.,  $L(\mathbf{X}) := \mathbf{X}^\top [\Phi + \Psi]^{-1} \mathbf{X}$  is positive, continuously differentiable and for every nontrivial  $\mathbf{U}$

$$\nabla L(\mathbf{U}) f(\mathbf{X}^* + \mathbf{U}) = -\alpha_2(\mathbf{X}^* + \mathbf{U}) \mathbf{U}^\top [\Phi + \Psi]^{-1} \mathbf{M} [\Phi + \Psi]^{-1} \mathbf{U} < 0 \tag{89}$$

where  $\alpha_2$  is a positive function. Hence the singular point always is a global attractor for the one-resident canonical equation, and there is no separate need to check the conditions described in Sect. 4.4.

We draw attention here to the fact that  $\mathbf{C}_{10}$  is symmetric, as often happens in simple models (cf. the paragraph following Eq. (33)).

As the sign of  $s_X(\mathbf{Y})$  is not influenced by the value of  $\beta > 0$ , we can put this proportionality constant at  $\beta := 1$ . For strategies close to the singularity, where  $\mathbf{U}$  and  $\mathbf{V}$  are  $O(\varepsilon)$ , we approximate the invasion fitness function as

$$\begin{aligned} s_{\mathbf{X}^*+\mathbf{U}}(\mathbf{X}^* + \mathbf{V}) &= \log \left( 1 + e^{-\frac{1}{2}(\mathbf{U}+\mathbf{V})^\top [\Phi+\Psi]^{-1}(\mathbf{U}+\mathbf{V})} - e^{-2\mathbf{U}^\top [\Phi+\Psi]^{-1} \mathbf{U}} e^{-\frac{1}{4}(\mathbf{U}-\mathbf{V})^\top \Phi^{-1}(\mathbf{U}-\mathbf{V})} \right) \\ &= \mathbf{U}^\top \left[ \frac{3}{2} [\Phi + \Psi]^{-1} + \frac{1}{4} \Phi^{-1} \right] \mathbf{U} + 2\mathbf{U}^\top \left[ -\frac{1}{2} [\Phi + \Psi]^{-1} - \frac{1}{4} \Phi^{-1} \right] \mathbf{V} \\ &\quad + \mathbf{V}^\top \left[ -\frac{1}{2} [\Phi + \Psi]^{-1} + \frac{1}{4} \Phi^{-1} \right] \mathbf{V} + O(\varepsilon^3) \end{aligned} \tag{90}$$

from which form we can read off  $\mathbf{C}_{11}$ ,  $\mathbf{C}_{10}$  and  $\mathbf{C}_{00}$ .

In the case of scalar traits, the invadability of  $X^*$  now settles the rest: if  $0 < C_{00}$  we find a branching point, else a CSS. The first case corresponds to  $\Phi < \Psi$ , the second to the reverse. This is the classical result that for branching to occur, the resource utilization kernel must be narrower than the resource abundance spectrum [10].

In the case of vectorial traits,  $X^*$  is an evolutionary endpoint if  $C_{00}$  is negative (semi)definite, and a branching point if  $C_{00}$  is positive (semi)definite. In Appendix G we show that  $C_{00}$  is positive definite if and only if  $\Psi - \Phi$  is positive definite, and that the same holds for indefiniteness, (non)negative and nonpositive definiteness. Hence a similar result holds for the relation between resource utilization kernel and resource abundance spectrum as in the scalar case. However, that  $C_{00}$  has a single positive eigenvalue is not sufficient for branching, as a protected polymorphism is not guaranteed to arise. Even if it does arise, the region of coexistence may be so narrow that the dimorphic population quickly strays out of it.

In the case where  $C_{00}$  has both positive and negative eigenvalues, we should resort to the algorithm suggested in Fig. 2 and use a Lotka–Volterra model to simulate the evolutionary dynamics. This is a valid approach, since the dynamics (up to  $O(\varepsilon^3)$ ) of the proportions  $p_i$  are identical in all models with the same single-resident fitness function, as argued in Appendix F: two different community dynamics that yield the same one-resident fitness functions up to terms of order  $O(\varepsilon^3)$  will arrive at an equilibrium with the same set of types surviving, whenever they start with the same  $N$ -resident community at equilibrium plus the same invader at a low density. Therefore, instead of the original discrete time equations (84) for diploids, we switch to a Lotka–Volterra system that has a sign-equivalent fitness function for any community of alleles, up to terms of order  $O(\varepsilon^3)$ . If alleles with trait values  $X^* + U_1, X^* + U_2, \dots, X^* + U_N$  are present in the community, their respective densities change over time as

$$\frac{d \log n_i}{dt} = 1 - \sum_j (U_j^T C_{11} U_j + 2U_j^T C_{10} U_i + U_i^T C_{00} U_i) n_j$$

from which we can calculate the equilibrium densities of the residents. Consequently, the initial conditions for an invasion event consist of the resident community at this attractor in addition to an invading allele, with associated trait vector  $X^* + V$ , at a very small density. By running the population dynamical equations with  $N + 1$  types present, we get to know the fate of the invader and all the resident types.

However, where do the invaders come from? At this point, we have to postulate a mutational process, as there is none given by Christiansen and Loeschke [11]. Then we can follow the algorithm given in Sect. 4.1, which involves generating a new mutant after each invasion attempt and running a similar invasion experiment, with the surviving community as initial condition. This is repeated until the branches have escaped the close proximity of the singularity and we are satisfied that the polymorphism is either protected or unstable.

As an example, using the notations  $\Delta := (U_1 - U_2)/2$  and  $\bar{U} := (U_1 + U_2)/2$ , we turn to Eq. (27) which approximates the invasion function for dimorphic communities

with both residents near a singularity  $X^*$ :

$$s_{X_1, X_2}(X^* + V) = V^T C_{00} V + 2\bar{U}^T C_{10} V + \bar{U}^T C_{11} \bar{U} - \Delta^T C_{00} \Delta + 2\Delta^T C_{10} (\bar{U} - V) \frac{\Delta^T [C_{00} + C_{10}^T] \bar{U}}{\Delta^T C_{10} \Delta} + O(\varepsilon^2)$$

Note that at the singularity,  $s_{X^*}(X^* + V) = V^T C_{00} V$  so mutants in the direction of the largest (positive) eigenvalue of  $C_{00}$  have the highest probability of invading. When branching indeed occurs, there are almost always two branches on opposite sides of the singularity that move at the same pace in opposite directions, along the steepest fitness gradient (cf. 4.5). Approximating this situation by  $U_1 = -U_2$ , we have  $\bar{U} = 0$  and  $\Delta = U_1$  and find ourselves in the special case

$$s_{X^*+U_1, X^*+U_2}(X^* + V) = V^T C_{00} V - U_1^T C_{00} U_1 + O(\varepsilon^3)$$

in which any more extreme mutant ( $V := \alpha U_1$  with  $|\alpha| > 1$ ) can invade and replace its ancestor, while  $\alpha < 1$  implies negative fitness. This shows that branches initially indeed grow away from the singularity.

If we are convinced that branching indeed does occur (either from a simulation as described above or because  $C_{00}$  is positive definite), one starts by formulating the two-resident fitness function

$$s_{X^*+U_1, X^*+U_2}(X^* + U_3) = \log \left( \max_{j=1,2} (1 + \beta (k_{j3} - \gamma_{j3,11} \hat{n}_{11} - \gamma_{j3,12} \hat{n}_{12} - \gamma_{j3,22} \hat{n}_{22})) \right) \tag{91}$$

where the equilibrium densities are calculated from

$$\begin{pmatrix} k_{11} \\ k_{12} \\ k_{22} \end{pmatrix} = \begin{bmatrix} 1 & \gamma & \gamma^4 \\ \gamma & 1 & \gamma \\ \gamma^4 & \gamma & 1 \end{bmatrix} \begin{pmatrix} \hat{n}_{11} \\ \hat{n}_{12} \\ \hat{n}_{22} \end{pmatrix} \tag{92}$$

with  $\gamma := \exp(-\frac{1}{4}(U_1 - U_2)^T \Phi^{-1}(U_1 - U_2))$ .

If the traits are scalar, we can calculate from the above expression (92) the  $X_1$ -isocline defined by  $\frac{\partial s_{X_1, X_2}(Y)}{\partial Y} \Big|_{Y=X_1} = 0$ , and the similarly defined  $X_2$ -isocline. These allow *trait evolution plots* (TEPs) to be drawn, which are basically *pairwise invasibility plots* (PIPs) with added information related to the  $s$ -function of dimorphic communities. For an easy explanation on how to plot and interpret PIPs consult [18], for examples of TEPs with some explanations including properties of the isoclines see [28]. The usefulness of TEPs lies in the fact that one can tell from them at a glance whether the community will evolve towards the edge of the coexistence region (so that the community reverts to a monomorphic state), towards the edge of the trait space (an evolutionary endpoint) or towards a singularity (with a possibility for secondary branching where the same analysis as before applies again).

## 5 Discussion

While superficially the first two parts of this paper are wildly disparate, we hope that the last part has shown their fundamental connexion by addressing one of the open problems the evolutionary biologist faces, namely how to systematically treat long-term evolutionary behaviour.

To the more mathematically inclined, Sect. 3 shows that classifying the bifurcation patterns that the  $s$ -functions of quadratic Lotka–Volterra models can exhibit is not just a niche hobby, but in the case of codimension-1 singularities amounts to a full classification of these singularities for models of evolution driven by small mutations. Furthermore we have revealed how the local geometry of coexistence (3.5, 3.6) and the residents' proportional density dynamics (Appendix F) are model independent (up to a given order), in the sense that they only depend on the geometry of the one-resident fitness function near the singularity and on the strategies present in the community.

To the more biologically inclined, Sect. 4 gives a handle on the systematic analysis of concrete evolutionary models from scratch. The approach can also prove useful in the interpretation of any odd pattern one has encountered, by showing a straightforward way to explore the geometry close to the location of the oddity and hence to figure out what model features cause the pattern. The approach on the whole allows the reader to focus on the phenomena at hand instead of the specific mathematical problems that are encountered: as models are shown to be interchangeable in some ways, computational difficulties may be avoided, e.g., by switching between continuous and discrete time versions of a model, or a similar sleight of hand.

The biologist will note how similar the canonical equation is to Lande's equation [38] which in turn derives from the breeder's equation [43] in the animal sciences. Except for the population density which does not appear in Lande's equation as a factor, all visible differences with the canonical equation amount to differences in the definition of parameters. Where they differ is in interpretation, as Lande's equation describes changes in the genetical makeup of a population through selection on standing genetic variation. This variation can for example be accumulated in a population near to a (weak) optimal strategy in a stable environment. Some have cast doubts on the sufficiency of the mutation/selection balance for the generation of the observed levels of variation (e.g., [37, 54]), though our feeling is that the associated problems are of greater mathematical than biological interest (see [57] for some mathematical counterarguments). Changing the environmental parameters that a population close to an evolutionary optimum is subjected to, leads to a rapid evolution in the genetical makeup. However Haldane [33] already realized that in the next phase, true innovations and long-term changes must come from mutations. He also made plain that mutation limited evolution is a slower process than naive analytical models would suggest since most advantageous mutants will fail to establish themselves due to stochasticity, an effect that is quantitatively captured by Ewens' approximation (14) of the establishment probability. The canonical equation builds on those ideas to derive a quantitative relation between the factors involved, establishing in particular how the ecology determines the selective pressures. The extended form presented in Sect. 2 is applicable to a very wide variety of discrete and continuous time models, instead of only to ODE models as is the original version by Dieckmann and Law [15].

The analysis as presented in Sects. 2 and 3 is ready for extension in several directions. In terms of content, the third order terms of the normal form (27), (42) should be worked out and compared between the different model types. While they do show differences, it is not yet clear to the authors whether those are substantial enough to translate into differences in bifurcation patterns.

In terms of rigour, we note that the calculations are presented in a heuristic and biologically slanted manner, at a cost to mathematical precision and exhaustiveness. We have followed most of the biological literature by treating mutations as unbiased, as if tacitly assuming the genotype–phenotype map to be linear and mutations to be unbiased at the genotype level. Since we are treating exceedingly small mutation steps, the mapping is indeed linear, but the possibility of bias remains (cf. [15]). A discussion of this topic can be found in the more mathematical treatment of the canonical equation by Champagnat et al. [7–9]. A far more complicated issue is the one underlying timescale separation and the several latent limits (of system size, mutation step size and mutation probability) which are not commutative, as Metz et al. [49] explained. How quasi-monomorphicity (and by extension, quasi- $N$ -morphicity) is maintained under some not-too-restrictive assumptions is being addressed more recently (e.g., [12,45]). More specifically, the latter authors show for general ODE population models with small differences between the types, that away from evolutionary singularities the dynamics of the relative frequencies  $p_i$  follows (up to the lowest order of approximation) the familiar population genetics equations for the density independent case. Furthermore they show that near singularities the dynamics mimics that of Lotka–Volterra models, with fitnesses approximated along the same lines as in this paper.

In terms of applicability, the collection of models covered should be extended to physiologically structured populations with infinite numbers of birth states. Biologically this is a small step, for example not just allowing a few classes of birth weights but allowing a continuum of sizes at birth. Mathematically however, this means that the matrix operations of Sect. 3 should be rewritten in terms of operators and norms, which we happily leave as a problem for more accomplished nonlinear analysts. Additionally, this requires an extension of Equality (14) to branching processes with infinitely many types, where the  $d$ -type version only is available now [23,2].

In continuous time, the extension to infinite numbers of birth states would automatically remove our restriction to fixed point attractors, as individuals born in a community on a periodic attractor can be assigned the phase of the attractor as (a component of) their birth state. Thus the attractor can be interpreted as a fixed point attractor and all analysis goes through. (In discrete time this trick does not even require the suggested extension.) Analytically, we have no clear idea how to extend our treatment to nonperiodic attractors. Heuristic explorations by Dieckmann (pers. comm.) for ergodically fluctuating environments with linear birth-and-death population dynamics for the invaders, suggest that the canonical equation is robust against such extension: by approximating the fixation probabilities as calculated by Kendall [36], it is found that the establishment probability is still approximately proportional to the fitness (cf. Relation (20)).

The analysis of this paper is valid for models where the population is spread over a finite number of patches, as long as the local resident densities are large enough to ensure local infinite dilution of individual effects and a branching process approxi-

mation for the initial phase of the invasion process. The patch an individual inhabits is then expressed in a component of its (birth) state. More research is badly needed on more complicated spatial models to see under which conditions the probability of establishment scales linearly with changes in strategy, as then an equation similar to the canonical equation will apply.

**Acknowledgments** Michel Durinx is supported by the Dutch national science organization (NWO) through PhD grant 809.34.002. Michel Durinx and Hans Metz received funding from the European Research Training Network ModLife. Géza Meszéna benefited from OTKA grants T049689 and TS049885. All authors received support from the NWO through the Dutch-Hungarian exchange grant 048.011.039.

The authors would like to thank Mats Gyllenberg and both referees for extensive suggestions and very detailed criticism that has benefited this paper immensely. We would also like to thank Frans Jacobs for discussions and a counterexample.

## Appendix A: The relationship between $s_{\mathbb{X}}(Y)$ and $R_0$ . I. Away from singularities

To derive a relationship between the  $N$ -resident invasion fitness  $s_{\mathbb{X}}(\mathbf{X}_i + \mathbf{V})$  and the lifetime offspring production  $R_0$ , we define  $\phi$  as

$$\phi(\rho, \mathbf{V}) := \log \left( \lambda_d \left( \int_0^\infty e^{-\rho a} \Lambda(\mathbf{X}_i + \mathbf{V}, \mathbf{I}_{\mathbb{X}}, da) \right) \right) \quad (93)$$

where  $\Lambda$  is the birth kernel (see Eq. (18) and preceding lines). Then the invasion fitness of a mutant  $\mathbf{X}_i + \mathbf{V}$  in a given community  $\mathbb{X}$  is the (generally unique) solution  $\rho$  to  $\phi(\rho, \mathbf{V}) = 0$  (known as the Euler-Lotka equation (18)).

We can expand  $\phi$  as a function of its first argument,

$$\phi(\rho, \mathbf{V}) = \phi(0, \mathbf{V}) + f(\mathbf{V})\rho + O(\rho^2) \quad (94)$$

As  $\partial\phi(\rho, \mathbf{V})/\partial\rho$  is (generically, in the space of all models that allow a birth kernel notation (93)) nonzero, the implicit function theorem may be applied to  $\rho$ . Hence we can furthermore expand  $s$  as a function of  $\mathbf{V}$ , where the constant term is zero since residents have zero fitness. Thus

$$\begin{aligned} 0 &= \phi(s_{\mathbb{X}}(\mathbf{X}_i + \mathbf{V}), \mathbf{V}) \\ &= \phi(0, \mathbf{V}) + f(\mathbf{V})s_{\mathbb{X}}(\mathbf{X}_i + \mathbf{V}) + O(s_{\mathbb{X}}(\mathbf{X}_i + \mathbf{V})^2) \\ &= \log R_0 + f_0 s_{\mathbb{X}}(\mathbf{X}_i + \mathbf{V}) + O(\|\mathbf{V}\|^2) \end{aligned} \quad (95)$$

where we have used the fact that  $\phi(0, \mathbf{V}) = \log \lambda_d(\mathbf{L}(\mathbf{X}_i + \mathbf{V}, \mathbf{I}_{\mathbb{X}})) = \log R_0$ , and  $f(\mathbf{V})$  was replaced by its lowest order part  $f_0 := f(\mathbf{0})$ . To calculate  $f_0$ , observe that it is a partial derivative of  $\phi$  for  $\rho$  at 0:

$$f_0 = \frac{\partial\phi(0, \mathbf{0})}{\partial\rho} = \frac{1}{\lambda_d \left( \int_0^\infty \Lambda(\mathbf{X}_i, \mathbf{I}_{\mathbb{X}}, da) \right)} \frac{\partial}{\partial\rho} \lambda_d \left( \int_0^\infty e^{-\rho a} \Lambda(\mathbf{X}_i, \mathbf{I}_{\mathbb{X}}, da) \right) \Big|_{\rho=0}$$

Since  $X_i$  is a resident, the first factor is one. The last factor can be resolved, as derivatives of simple eigenvalue  $\lambda(x)$  of a matrix  $M(x)$  are found from

$$\frac{\partial \lambda(x_0)}{\partial x} = \mathbf{v}^\top \frac{\partial M(x_0)}{\partial x} \mathbf{u} \tag{96}$$

where  $\mathbf{u}$  and  $\mathbf{v}^\top$  are respectively right- and left eigenvectors of  $M(x_0)$ , normalized such that  $\mathbf{v}^\top \mathbf{u} = 1$  (e.g., [4, 44]). As  $\int_0^\infty e^{-\rho a} \Lambda(X_i, \mathbb{I}_\mathbb{X}, da)$  is nonnegative and primitive so that  $\lambda_d$  is an isolated eigenvalue, we use this last result to decide that

$$\begin{aligned} f_0 &= \left. \frac{\partial \lambda_d \left( \int_0^\infty e^{-\rho a} \Lambda(X_i, \mathbb{I}_\mathbb{X}, da) \right)}{\partial \rho} \right|_{\rho=0} \\ &= \mathbf{v}^\top \left( \left. \frac{\partial \int_0^\infty e^{-\rho a} \Lambda(X_i, \mathbb{I}_\mathbb{X}, da)}{\partial \rho} \right|_{\rho=0} \right) \mathbf{u} \\ &= -\mathbf{v}^\top \left( \int_0^\infty a \Lambda(X_i, \mathbb{I}_\mathbb{X}, da) \right) \mathbf{u} =: -T_f(X_i, \mathbb{I}_\mathbb{X}) \end{aligned} \tag{97}$$

where  $\mathbf{u}$  and  $\mathbf{v}^\top$  now are normalized right- and left eigenvectors of  $L(X_i, \mathbb{I}_\mathbb{X})$  (cf. Eq. (13)). The (nonzero) quantity  $T_f$  has a natural interpretation as the *average age at giving birth*, since the integral is a lifetime census of the parent’s age at each birth event while the expected lifetime offspring production is one (as  $X_i \in \mathbb{X}$ ). Substituting this value for  $f_0$  into our expansion (95), we conclude that

$$s_\mathbb{X}(X_i + \mathbf{V}) = \frac{\log R_0(X_i + \mathbf{V}, \mathbb{I}_\mathbb{X})}{T_f(X_i, \mathbb{I}_\mathbb{X})} + O(\|\mathbf{V}\|^2)$$

**Appendix B: The relationship between  $s_\mathbb{X}(Y)$  and  $R_0$ . II. Near singularities**

In Appendix A, a relationship between invasion fitness and lifetime reproductive output was formulated, up to  $O(\varepsilon^2) = O(\|\mathbf{V}\|^2)$ . Near a singularity however, we can redo the analysis to show that the approximate relation is correct up to terms of order  $O(\varepsilon^3)$ .

When all residents are close to a singular strategy  $X^*$ , we can describe the community in terms of strategy deviations  $\xi_i$  and a scaling factor  $\varepsilon$  as  $\mathbb{X} := \{X^* + \varepsilon \xi_i \mid i \leq N\}$ . We expand the multiresident fitness function as dependent on the mutation step  $\mathbf{V}$  and the community scaling factor  $\varepsilon$  as

$$s_\mathbb{X}(X^* + \mathbf{V}) = s_{X^*}(X^*) + \beta \mathbf{V} + \varepsilon B(\xi_1, \dots, \xi_N) + O(\|\mathbf{V}\|^2, \|\mathbf{V}\|\varepsilon, \varepsilon^2) \tag{98}$$

The constant term is necessarily zero, and  $\beta$  too since  $s_\mathbb{X}(Y) = s_{X^*}(Y)$  at  $\varepsilon = 0$ . Furthermore, any resident strategy is a zero of the  $N$ -resident fitness function, so we find from any choice  $\mathbf{V} = \varepsilon \xi_i$  that  $B(\xi_1, \dots, \xi_N)$  is zero as well. Thus for any mutant



$\mathbf{X}^* + \mathbf{V}$  where  $\mathbf{V} = O(\varepsilon)$ , we have that

$$s_{\mathbb{X}}(\mathbf{X}^* + \mathbf{V}) = O(\varepsilon^2) \tag{99}$$

When we define

$$\phi(\rho, \mathbf{V}) := \log \left( \lambda_d \left( \int_0^\infty e^{-\rho a} \Lambda(\mathbf{X}^* + \mathbf{V}, \mathbf{I}_{\mathbb{X}}, da) \right) \right) \tag{100}$$

we may expand this  $\phi$  again as a function of  $\rho$  and thus solve the Euler–Lotka equation  $\phi(\rho, \mathbf{V}) = 0$  for  $\mathbf{V} = O(\varepsilon)$  as

$$\begin{aligned} 0 &= \phi(s_{\mathbb{X}}(\mathbf{X}^* + \mathbf{V}), \mathbf{V}) \\ &= \phi(0, \mathbf{V}) + f(\mathbf{V}) s_{\mathbb{X}}(\mathbf{X}^* + \mathbf{V}) + O(s_{\mathbb{X}}(\mathbf{X}^* + \mathbf{V})^2) \\ &= \log R_0(\mathbf{X}^* + \mathbf{V}, \mathbf{I}_{\mathbb{X}}) + f_0 s_{\mathbb{X}}(\mathbf{X}^* + \mathbf{V}) + O(\varepsilon^3) \end{aligned} \tag{101}$$

where  $f_0 := f(\mathbf{0}) = -T_f(\mathbf{X}^*, \mathbf{I}^*) =: -T_f^*$ , as shown at Eq. (97). Hence we conclude that

$$s_{\mathbb{X}}(\mathbf{X}^* + \mathbf{V}) = \frac{\log R_0(\mathbf{X}^* + \mathbf{V}, \mathbf{I}_{\mathbb{X}})}{T_f^*} + O(\varepsilon^3)$$

for communities and mutants near a singularity  $\mathbf{X}^*$ .

### Appendix C: Miscalculating $s_{\mathbb{X}}(\mathbf{Y})$

Let us consider a singular strategy  $\mathbf{X}^*$ , and try to express the  $s$ -function in case there are  $N$  resident strategies close to  $\mathbf{X}^*$ . For each  $i$  from 1 to  $N$ , we can express resident trait values as  $\mathbf{X}_i = \mathbf{X}^* + \mathbf{U}_i$  for some small vector  $\mathbf{U}_i$ , and similarly for invading mutants  $\mathbf{Y} = \mathbf{X}^* + \mathbf{V}$ . We now take the Taylor expansion around  $\mathbf{X}^*$  up to quadratic terms, and can start to figure out the coefficients:

$$\begin{aligned} s_{X_1 \dots X_N}(\mathbf{Y}) &= \alpha + \boldsymbol{\beta} \mathbf{V} + \sum_i \boldsymbol{\beta}_i \mathbf{U}_i + \mathbf{V}^T \mathbf{C}_{00} \mathbf{V} \\ &\quad + 2 \sum_i \mathbf{U}_i^T \mathbf{C}_{i0} \mathbf{V} + \sum_{ij} \mathbf{U}_i^T \mathbf{C}_{ij} \mathbf{U}_j + O(\varepsilon^3) \end{aligned}$$

where  $\mathbf{C}_{00}$  and each matrix  $\mathbf{C}_{ii}$  is taken to be symmetric.

Any resident has zero growth in an equilibrium population, so for each  $k$  we have the consistency condition  $s_{X_1 \dots X_N}(\mathbf{X}_k) = 0$ . Thus for all  $k$  we must have that

$$0 = \alpha + \boldsymbol{\beta} \mathbf{U}_k + \sum_i \boldsymbol{\beta}_i \mathbf{U}_i + \mathbf{U}_k^T \mathbf{C}_{00} \mathbf{U}_k + 2 \sum_i \mathbf{U}_i^T \mathbf{C}_{i0} \mathbf{U}_k + \sum_{ij} \mathbf{U}_i^T \mathbf{C}_{ij} \mathbf{U}_j + O(\varepsilon^3)$$

As this has to hold independently of the strategy deviations  $U_1$  to  $U_N$ , we can split the equation into several equations like

$$0 = \alpha + (\beta + \beta_k) U_k + \sum_{i \neq k} \beta_i U_i \tag{102}$$

$$0 = U_k^T (C_{00} + 2C_{k0} + C_{kk}) U_k \tag{103}$$

$$0 = \sum_{i \neq k} U_i^T (2C_{i0} + C_{ki}^T + C_{ik}) U_k \tag{104}$$

$$0 = \sum_{i, j \neq k} U_i^T C_{ij} U_j \tag{105}$$

Equation (102) shows that  $\alpha = 0$ , and for each  $i \neq k$  we see  $\beta_i = \mathbf{0}^T$  while  $\beta_k = -\beta$ . So if we first choose  $k = 1$  and then  $k = 2$ , we conclude that for *all* the residents  $\beta = \beta_i = \mathbf{0}^T$ . Note that in the monomorphic case, this argumentation does not hold as  $k = 2$  is impossible then.

We deduce from Equality (105) that  $C_{ij} = 0$ , if both  $i$  and  $j$  differ from  $k$ . So taking an initial choice of  $k = 1$  shows that all  $C_{ij}$  are zero except if  $i$  or  $j$  is one, and a further choice of  $k = 2$  shows that all are zero except  $C_{12}$  and  $C_{21}$ . If  $N > 2$ , we can take  $k = 3$  to prove that for any  $i$  and  $j$  the matrix  $C_{ij}$  is zero. The case  $N = 2$  is worked out below, where  $C_{12}$  and  $C_{21}$  may be nonzero.

For  $N > 2$  we substitute our results into Eq. (104) and see in a similar way that  $C_{i0} = 0$  for any  $i$ . From Eq. (103) we deduce that the symmetric matrix  $C_{00}$  is zero as well. Hence we conclude that for any model where  $N > 2$

$$s_{X_1 \dots X_N}(Y) = 0 + O(\varepsilon^3)$$

which is clearly false. (By repeating the argumentation for higher order terms we can “show” that  $s_{X_1 \dots X_N}(Y) = O(\varepsilon^{M+1})$  when  $N > M$ .) The false result can be traced back to one implicit assumption: the existence of partial derivatives, a basic requirement for the applicability of Taylor approximations.

For  $N = 2$  we find that  $C_{10} + C_{10}^T = -C_{00} = C_{20} + C_{20}^T$  and  $2C_{10} = 2C_{20}^T = -[C_{12} + C_{21}^T]$ , so that

$$s_{X_1 X_2}(Y) = (V - U_1)^T [C_{12} + C_{21}^T] (V - U_2) + O(\varepsilon^3)$$

This equation suggests that the  $s$ -function for any two resident model would be locally quadratic at  $X^*$ . If one compares this equation to the correct solution in this paper (27), we see that it is correct only if we are dealing with scalar strategies.

### Appendix D: Proofs of Lemma 1–3 and 7

The following lemmata were mentioned without proof in Sect. 3.6. They are useful in making some points about upper bounds to the number of types that can locally

coexist. The (non)invertibility of  $E^*$  plays a central role in this issue, not just because our route to calculating the multiresident fitness function depends on the invertibility of  $E^*$  (72), but because invertibility is a necessary condition for the population dynamical stability of the community. Since the proofs are of a technical nature, they were moved to this appendix, as they merely divert the attention from the real issues in Sect. 3.6.

**Lemma 1** *If  $\det E \neq 0$ , then  $\det E^* = -(\det E)I^T E^{-1}I$*

*Proof* We start by recalling the general formula for the inverse of a nonsingular  $N \times N$  matrix:

$$(\det E)E^{-1} = \begin{bmatrix} (-1)^{1+1}e_{1,1} & \cdots & (-1)^{1+N}e_{N,1} \\ \vdots & \ddots & \vdots \\ (-1)^{N+1}e_{1,N} & \cdots & (-1)^{N+N}e_{N,N} \end{bmatrix}$$

where the minor  $e_{i,j}$  is defined as the determinant of the matrix obtained by deleting the  $i$ th row and  $j$ th column of  $E$ .

We encounter the same minors, when we calculate the determinant of  $E^*$  by expanding first from the bottom row and then from the rightmost column:

$$\begin{aligned} \det E^* &= \sum_j (-1)^{N+1+j} \det \begin{bmatrix} E_{1,1} & \cdots & E_{1,j-1} & E_{1,j+1} & \cdots & E_{1,N} & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ E_{N,1} & \cdots & E_{N,j-1} & E_{N,j+1} & \cdots & E_{N,N} & 1 \end{bmatrix} \\ &= \sum_{ij} (-1)^{N+1+j} (-1)^{N+i} \det \begin{bmatrix} E_{1,1} & \cdots & E_{1,j-1} & E_{1,j+1} & \cdots & E_{1,N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ E_{i-1,1} & \cdots & E_{i-1,j-1} & E_{i-1,j+1} & \cdots & E_{i-1,N} \\ E_{i+1,1} & \cdots & E_{i+1,j-1} & E_{i+1,j+1} & \cdots & E_{i+1,N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ E_{N,1} & \cdots & E_{N,j-1} & E_{N,j+1} & \cdots & E_{N,N} \end{bmatrix} \\ &= \sum_{ij} (-1)^{i+j+1} e_{i,j} \\ &= -(\det E)I^T E^{-1}I \end{aligned}$$

□

**Lemma 2** *If  $\text{rank } E \leq N - 2$ , then  $E^*$  is singular.*

*Proof* If the rank of  $E$  is  $N - 2$  or less, we can transform it by elementary row operations into an  $N \times N$  matrix with the last two rows equal to zero. After applying the same sequence of elementary operations to  $E^*$  instead, one of its last two rows is a multiple of the other (since only their last coefficients possibly differ from zero). Thus  $E^*$  is singular since the determinant is not affected by elementary row operations. □

**Lemma 3** *If  $\text{rank } E = N - 1$ , then generically  $E^*$  is invertible.*

*Proof (by contradiction)* If  $E^*$  is singular, there exists a nontrivial vector  $\mu$  such that the  $N + 1$  equations  $(\mu_1 \mu_2 \cdots \mu_{N+1}) E^* = \mathbf{0}^T$  are simultaneously satisfied.

If  $\mu_{N+1} = 0$ , then  $\mu$  is the unique (up to a scalar) nonzero vector such that  $(\mu_1 \mu_2 \cdots \mu_N) E = \mathbf{0}^T$ , where uniqueness comes from the rank of  $E$ . This situation is nongeneric as the independent  $(N + 1)^{st}$  equation  $\sum_i \mu_i = 0$  is satisfied as well.

If on the other hand  $\mu_{N+1} \neq 0$ , we rescale  $\mu$  by setting  $\mu_{N+1} := -1$ , thus finding a solution to the  $N$  equations  $(\mu_1 \mu_2 \cdots \mu_N) E = \mathbf{1}^T$ . Generically such a solution does not exist however, as  $E$  has an  $(N - 1)$ -dimensional range. □

**Lemma 7** *Even if both  $U$  and  $C_{10}$  have full rank,  $E^*$  can still be singular. In fact, all four combinations of invertible or singular  $E$  and  $E^*$  can occur.*

*Proof* Only in cases where  $N = m$  we know offhand whether  $E = U^T C_{10} U$  is singular or not, as the determinant of a matrix product is the product of the determinants.

As an example of the case that both  $E$  and  $E^*$  are singular, consider

$$[U_1 \ U_2 \ U_3 \ U_4] := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad C_{10} := \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

We see that for any combination of two residents, both  $U$  and  $C_{10}$  have full rank (respectively two and four). For two out of the six possible combinations (namely  $\{U_1, U_4\}$  and  $\{U_2, U_3\}$ ), both  $E$  and  $E^*$  are invertible. But for the other four pairs of residents (namely  $\{U_1, U_2\}$ ,  $\{U_1, U_3\}$ ,  $\{U_2, U_4\}$  and  $\{U_3, U_4\}$ ), matrix  $E$  has rank zero and  $E^*$  is singular, so that such strategy couples cannot coexist: if  $[C_{00}]_{ii} \neq [C_{00}]_{jj}$ , the system  $E P = T$  (70) is contradictory; if  $[C_{00}]_{ii} = [C_{00}]_{jj}$ , there is a continuum of neutrally stable solutions.

If we add any of the remaining two as a third resident type, coexistence becomes possible again as  $E^*$  is invertible (but  $E$  is singular). Notice that  $T_i := -U_i^T C_{00} U_i = -[C_{00}]_{ii}$ , so that there is no a priori relation whatsoever between the vector  $T$  and the matrix  $E$ . If we consider the community  $\{U_1, U_2, U_3\}$  for example, the proportions  $p_1, p_2, p_3$  will respectively be  $1 + 2T_1 - T_2 - T_3, T_3 - T_1$  and  $T_2 - T_1$ . Values of  $T$  that result in strictly positive proportion vectors are  $(1/4 \ 1/2 \ 1/2)^T$  or  $(-1/2 \ -1/3 \ -1/6)^T$ . Similarly, there are generic solutions with all four given strategies present.

Considering another possible resident,  $U_0^T := (1 \ 0 \ 0 \ 1)$ , we encounter the fourth possibility, as the resident duo  $\{U_1, U_0\}$  has a singular  $E^*$  matrix yet  $E = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$  is invertible. □

**Appendix E: The environmental dimension and finite dimensional Lotka–Volterra environments**

We will first precisely define the environmental dimension, as relevant in an evolutionary context and differing from the dimension concept used in physiologically

structured models. In such models, the environment is used to describe the full population dynamics. In the adaptive dynamics context however, the interest is reduced to a time-averaged growth rate of an individual in that environment (as testified by the definition of invasion fitness). Moreover, it is not the environments themselves that are of central concern here, but the collection of ergodic probability measures on functions that map time to environments.

A closer consideration of this topic reveals there is an observability issue. Whether for reasons of didactical clearness, mathematical manipulability, interpretability or ineptitude, the environment will not always be formulated in a minimal form. In addition one must realize that not the dimension of the set of environments is relevant, but the dimension of the subspace of feasible environments.

To do away with these potential deficiencies, we use the following definitions. For the set of all environments  $\mathcal{J}$  and the trait space  $\mathcal{X}$ , we define the *growth operator* as

$$\rho : \mathcal{J} \rightarrow C(\mathcal{X}) \tag{106}$$

such that  $\rho(\mathbf{I})(\mathbf{Y})$  is the instantaneous per capita growth rate of a given type  $\mathbf{Y}$  in a given environment  $\mathbf{I}$ , as dependent on the model under consideration. In this setting we define the *environmental dimension* of that model as

$$\dim_E := \dim \rho(\mathcal{J}) \tag{107}$$

From the definition we straightforwardly see why Lotka–Volterra type models by default have infinite dimensional environments, as opposed to, e.g., resource dynamics models (cf. Eq. (116)):

**Lemma 8** *Generically, the environment of a Lotka–Volterra model is infinite dimensional.*

*Proof* Let us consider a strategy space  $\mathcal{X}$  with an infinite number of elements. Using the growth operator (106), for an  $N$ -resident Lotka–Volterra model (28) we formally have that

$$\rho(\mathbf{I})(\mathbf{Y}) = r(\mathbf{Y}) \left( 1 - \sum_i a(\mathbf{Y}, \mathbf{X}_i) n_i \right)$$

or for more general Lotka–Volterra models

$$\rho(\mathbf{I})(\mathbf{Y}) = r(\mathbf{Y}) \left( 1 - \int_{\mathcal{X}} a(\mathbf{Y}, \mathbf{X}) \, d\nu(\mathbf{X}) \right)$$

where  $N$ -resident models are made by taking the population distribution  $\nu$  to be a weighted sum of  $N$  Dirac delta distributions  $\nu(\mathbf{X}) := \sum_j n_j \delta(\mathbf{X} - \mathbf{X}_j)$ .

To separate the focal individual  $Y$  from the environment  $I$ , we have to define the environment as (something isomorphic to) the real-valued function

$$I : Y' \mapsto \int_{\mathfrak{X}} a(Y', X) \, d\nu(X) \tag{108}$$

Now the growth rate can indeed be put in the form  $\rho(I)(Y)$ , with  $I$  independent of  $Y$ . Thus we see that the dimension of  $I$  is at most the cardinality of  $\mathfrak{X}$ .

As there is no a priori reason for a smaller set than  $\mathfrak{X}$  to suffice as domain for a function similar to Expression (108), generically  $I$  is infinite dimensional.  $\square$

**Proposition 6** *A Lotka–Volterra model with interaction function  $a(Y, X)$  has finite feedback dimension  $e$ , if and only if  $e$  is the minimal number for which there exist functions  $b_q$  and  $c_q$  such that  $a(Y, X) = \sum_{q=1}^e b_q(Y)c_q(X)$ .*

*Proof* Firstly, by Definition (107) we have

$$e := \dim \rho(\mathfrak{J}) = \dim A(\mathfrak{D}) \tag{109}$$

where  $\mathfrak{D}$  is the space of all Borel measures on  $\mathfrak{X}$  and the operator  $A$  is defined as

$$A(\nu) : Y \mapsto \int_{\mathfrak{X}} a(Y, X) \, d\nu(X) \tag{110}$$

Therefore  $A(\mathfrak{D})$  has a basis  $\{b_1(Y), b_2(Y), \dots, b_e(Y)\}$ , and for any  $\nu \in \mathfrak{D}$  there are coefficients  $\gamma_q$  such that  $A(\nu)(Y) = \sum_{q=1}^e b_q(Y)\gamma_q$ . By choosing Dirac distributions  $\nu(X) := \delta(X - X_0)$ , we see that for any strategy combination  $(X_0, Y)$  the interaction term can be written as

$$a(Y, X_0) = \int_{\mathfrak{X}} a(Y, X) \, d\nu(X) = \sum_{q=1}^e b_q(Y)\gamma_q \tag{111}$$

Thus the  $c_q$ -functions are defined pointwise from Equality (111) as  $c_q(X_0) := \gamma_q$ . That  $e$  is the minimal number of functions  $b_q$  and  $c_q$ , follows from the second part of this proof.

Conversely, if  $a(Y, X)$  can be written as a finite sum  $\sum_{q=1}^e b_q(Y)c_q(X)$ , then

$$\rho(Y)(I) = r(Y) \left( 1 - \int_{\mathfrak{X}} \sum_{q=1}^e b_q(Y)c_q(X) \, d\nu(X) \right) = r(Y) \left( 1 - \sum_{q=1}^e b_q(Y)I_q \right)$$

where  $I_q := \int_{\mathfrak{X}} c_q(X) \, d\nu(X)$ . Thus we see that the environment is at most  $e$ -dimensional. In fact,  $I$  is exactly  $e$ -dimensional: if  $I$  were  $e'$ -dimensional with  $e' < e$ , then the first part of the proof shows that  $e$  was not minimal.

To finish the first half of the proof, we note that there cannot exist an interaction operator  $a(\mathbf{Y}, \mathbf{X}) = \sum_{q=1}^{e'} b_q(\mathbf{Y})c_q(\mathbf{X})$  with environments of dimension  $e > e'$ , as the second half of the proof shows that  $e \leq e'$ .  $\square$

## Appendix F: The dynamics of fractions $p_i$

We have seen, through the equality  $\mathbf{P}^* = \mathbf{E}^{*-1}\mathbf{T}^*$  (41), that the equilibrium fractions for all models depend in an identical way on the simple fitness function  $s_{\mathbf{X}}(\mathbf{Y})$  and the strategies of the players, up to but not including terms of order  $O(\varepsilon^3)$ . But there is more than that: in this appendix we argue that for a given fitness function, the dynamics of the fractions is model-independent in the same sense as well.

To show this, we first analyze a general Lotka–Volterra system. Later on we repeat the analysis with a resource dynamics model, as the relevant singular perturbation theory for the dynamical analogues of the equilibrium equations (1) for general structured populations have not yet been developed. (We note that for the single birth state case, an obvious research plan would be to apply the approach in the third chapter of the thesis of Getto [29] to the conjecture of Greiner et al. [31]; see also [19]).

First we consider a Lotka–Volterra community  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{N-1}\}$  at equilibrium, to which we add a small number  $n_N = O(\varepsilon^2)$  of mutants with strategy  $\mathbf{X}_N$ . We recall that for the  $(N - 1)$ -resident community at equilibrium, the density is of the form  $\hat{n} = 1 + O(\varepsilon^2)$  (38), so at least initially  $n = 1 + O(\varepsilon^2)$  for our  $N$ -resident community as well. That this actually holds at all times, is shown in the following way.

Writing  $r_i := r(\mathbf{X}_i)$  and  $a_{ij} := a(\mathbf{X}_i, \mathbf{X}_j) = 1 + \varepsilon^2\alpha_{ij} + O(\varepsilon^3)$  for some constants  $\alpha_{ij}$  (33), the dynamics of each of the  $N$  densities is

$$\frac{dn_i}{dt} = n_i r_i \left( 1 - \sum_j a_{ij} n_j \right) \quad (112)$$

and therefore the dynamics of the total density is

$$\begin{aligned} \frac{d\Delta n}{dt} &= \frac{dn}{dt} = \sum_j \frac{dn_j}{dt} = \sum_j n_j r_j \left( 1 - \sum_k a_{jk} n_k \right) \\ &= n \sum_j p_j r_j \left( 1 - \sum_k (1 + \varepsilon^2\alpha_{jk}) p_k n \right) + O(\varepsilon^3) \\ &= n \sum_j p_j r_j (-\Delta n) + O(\varepsilon^2) \end{aligned} \quad (113)$$

Since the solution to this ODE is continuous, there exists some half open time interval  $[0, \tau)$  during which  $\Delta n = O(\varepsilon)$ . Then the sign of  $d\Delta n/dt$  is the opposite of that of  $\Delta n$ , so  $\Delta n$  cannot escape from an  $O(\varepsilon^2)$ -neighbourhood of zero and  $\tau = \infty$ .

Bearing the above in mind, we expand the dynamics of a fraction  $p_i$  as

$$\begin{aligned} \frac{dp_i}{dt} &= \frac{dn_i}{dt} \frac{1}{n} - \frac{n_i}{n^2} \frac{dn}{dt} \\ &= p_i r_i \left( 1 - \sum_k a_{ik} p_k n \right) - p_i \sum_j p_j r_j \left( -\Delta n - \varepsilon^2 n \sum_k \alpha_{jk} p_k \right) + O(\varepsilon^3) \\ &= p_i r_i \left( -\Delta n - \sum_k \varepsilon^2 \alpha_{ik} p_k n \right) + p_i \sum_j p_j r_j \left( \Delta n + \varepsilon^2 n \sum_k \alpha_{jk} p_k \right) + O(\varepsilon^3) \\ &= -p_i r(X^*) \varepsilon^2 n \sum_k \alpha_{ik} p_k + p_i r(X^*) \varepsilon^2 n \sum_{jk} \alpha_{jk} p_j p_k + O(\varepsilon^3) \end{aligned} \tag{114}$$

where the  $\Delta n$ -terms cancelled each other out because  $r_i, r_j = r(X^*) + O(\varepsilon)$  and thus the difference is absorbed by the order term  $O(\varepsilon^3)$  since  $\Delta n = O(\varepsilon^2)$  at all times. Similarly we can replace  $n$  by 1 in the remaining terms, which both have  $\varepsilon^2$  as a factor. From the calculations following Eq. (33) we know that  $\alpha_{kj} = \xi_j^T \Gamma_{11} \xi_j + 2\xi_j^T \Gamma_{10} \xi_k + \xi_k^T \Gamma_{00} \xi_k$ , so using the renaming  $\mathbf{C} := -r(X^*)\Gamma$  (cf. Eq. 34) we rewrite Eq. (114) as

$$\begin{aligned} \frac{dp_i}{dt} &= -p_i r(X^*) \varepsilon^2 \sum_k (\xi_k^T \Gamma_{11} \xi_k + 2\xi_k^T \Gamma_{10} \xi_i + \xi_i^T \Gamma_{00} \xi_i) p_k \\ &\quad + p_i r(X^*) \varepsilon^2 \sum_{jk} (\xi_k^T \Gamma_{11} \xi_k + 2\xi_k^T \Gamma_{10} \xi_j + \xi_j^T \Gamma_{00} \xi_j) p_j p_k + O(\varepsilon^3) \\ &= p_i \sum_{jk} (2U_k^T \mathbf{C}_{10} (U_i - U_j) + (U_i + U_j)^T \mathbf{C}_{00} (U_i - U_j)) p_j p_k + O(\varepsilon^3) \end{aligned} \tag{115}$$

We will now argue that this dynamics (115) is a shared property of all population dynamical models. However, as mentioned in the first lines of this appendix, there is no dynamical equivalent of the equilibrium equations available for physiologically structured populations (1). Therefore we will show that the dynamics found for Lotka–Volterra models (115), is also found for the most general subclass of the structured population models where the dynamical equivalent of the equilibrium equations (1) is of ODE form. The subclass in question is that of resource dynamics models, generally formulated as follows: for each trait value  $X_i$ , per capita growth is given by

$$\frac{1}{n_i} \frac{dn_i}{dt} = g \left( X_i, \sum_j h_1(X_j) n_j, \sum_j h_2(X_j) n_j, \dots, \sum_j h_r(X_j) n_j \right) \tag{116}$$

for some  $C^3$  functions  $g$  and  $h_1, h_2, \dots, h_r$ . The multiresident invasion fitness is by definition

$$s_{\mathbf{X}}(\mathbf{Y}) := g(\mathbf{Y}, I_1, I_2, \dots, I_r) \tag{117}$$



where the environment  $I$  is defined componentwise as

$$I_s := \sum_j h_s(\mathbf{X}_j)n_j \tag{118}$$

Like we did for Lotka–Volterra (29) and general structured population models (46), we simplify the calculations through a trait-dependent rescaling that does not influence the  $s$ -function:

$$\forall \mathbf{X}: \quad 0 = g(\mathbf{X}, h_1(\mathbf{X}), h_2(\mathbf{X}), \dots, h_r(\mathbf{X})) \tag{119}$$

The effect is that the equilibrium density  $\hat{n}$  is one in every monomorphic community. Once again we need to know that  $\Delta n = O(\varepsilon^2)$  at all times. For that we combine a first-order expansion of the environmental differences,

$$\begin{aligned} I_s - h_s(\mathbf{X}^*) &= \sum_j \left( h_s(\mathbf{X}^*) + \frac{\partial h_s}{\partial \mathbf{X}} U_j \right) p_j (1 + \Delta n) - h_s(\mathbf{X}^*) + O(\varepsilon^2) \\ &= h_s(\mathbf{X}^*) \Delta n + \sum_j \frac{\partial h_s}{\partial \mathbf{X}} U_j p_j (1 + \Delta n) + O(\varepsilon^2) \end{aligned} \tag{120}$$

with a first-order expansion of  $g$

$$\begin{aligned} g(\mathbf{X}^* + \mathbf{V}, I_1, I_2, \dots, I_r) &= g(\mathbf{X}^*, h_1(\mathbf{X}^*), h_2(\mathbf{X}^*), \dots, h_r(\mathbf{X}^*)) + \frac{\partial g}{\partial \mathbf{Y}} \mathbf{V} + \sum_s \frac{\partial g}{\partial I_s} (I_s - h_s(\mathbf{X}^*)) \\ &\quad + O(\varepsilon^2) + \sum_{st} O((I_s - h_s(\mathbf{X}^*))(I_t - h_t(\mathbf{X}^*))) \end{aligned} \tag{121}$$

where the derivatives are taken at  $I_s = h_s(\mathbf{X}^*)$  ( $\forall s$ ) and  $\mathbf{V} = \mathbf{0}$ . Note that the first term is zero because of the rescaling, and the second because  $\mathbf{X}^*$  is a singular strategy.

Because of the rescaling (119),  $\Delta n$  is zero in monomorphic equilibrium communities, and hence  $I_s - h_s(\mathbf{X}^*) = \sum_j (\partial h_s / \partial \mathbf{X}) U_j + O(\varepsilon^2)$  there. If we combine this expansion with that of  $g$  (121), and observe that the linear terms (in  $\varepsilon$ ) are zero in the normal form for  $s$ -functions near singularities (73), we see that

$$\sum_s \frac{\partial g}{\partial I_s} \frac{\partial h_s}{\partial \mathbf{X}} = \mathbf{0}^\top \tag{122}$$

As we have shown in the paragraph preceding Eq. (64),  $\Delta n = O(\varepsilon^2)$  for a community at equilibrium with  $N - 1$  types present. If we then add a small number  $n_N = O(\varepsilon^2)$  of invaders, then there is an open time interval  $[0, \tau)$  during which  $\Delta n = O(\varepsilon)$ . During this interval, by combining Expansions (120) and (121) and Equality (122), we find

$$\begin{aligned} \frac{d\Delta n}{dt} &= \frac{dn}{dt} = n \sum_i p_i g(\mathbf{X}^* + \mathbf{U}_i, \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_r) \\ &= (1 + \Delta n) \sum_s \frac{\partial g}{\partial I_s} h_s(\mathbf{X}^*) \Delta n + O(\varepsilon^2) \end{aligned} \tag{123}$$

From this we can conclude that either  $\Delta n = O(\varepsilon^2)$  or its sign is the opposite of that of  $d\Delta n/dt$ , so the time interval  $[0, \tau)$  is actually unbounded and  $\Delta n = O(\varepsilon^2)$  at all times. The sign difference is easy to show, as the assumption of the existence of a fixed point attractor (1.1) holds for this singular community at equilibrium, so

$$0 > \left. \frac{\partial g(\mathbf{X}^*, \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_r)}{\partial n} \right|_{\substack{I_s = h_s(\mathbf{X}^*)n \\ n=1}} = \sum_s \frac{\partial g}{\partial I_s} h_s(\mathbf{X}^*) \tag{124}$$

To find the dynamics of the fractions  $p_i$ , we need a second-order expansion of the both the environmental differences,

$$I_s - h_s(\mathbf{X}^*) = h_s(\mathbf{X}^*) \Delta n + \sum_j \left( \frac{\partial h_s}{\partial \mathbf{X}} \mathbf{U}_j + \mathbf{U}_j^\top \frac{\partial^2 h_s}{\partial \mathbf{X}^2} \mathbf{U}_j \right) p_j + O(\varepsilon^3) \tag{125}$$

and the community dynamics,

$$\begin{aligned} &g(\mathbf{X}^* + \mathbf{V}, \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_r) \\ &= \sum_s \frac{\partial g}{\partial I_s} (I_s - h_s(\mathbf{X}^*)) + \frac{1}{2} \sum_{st} \frac{\partial^2 g}{\partial I_s \partial I_t} (I_s - h_s(\mathbf{X}^*)) (I_t - h_t(\mathbf{X}^*)) \\ &\quad + \sum_s (I_s - h_s(\mathbf{X}^*)) \frac{\partial^2 g}{\partial I_s \partial \mathbf{Y}} \mathbf{V} + \frac{1}{2} \mathbf{V}^\top \frac{\partial^2 g}{\partial \mathbf{Y}^2} \mathbf{V} + O(\varepsilon^3) \end{aligned} \tag{126}$$

The dynamics of the fractions is then approximated as

$$\begin{aligned} \frac{dp_i}{dt} &= \frac{dn_i}{dt} \frac{1}{n} - \frac{n_i}{n^2} \frac{dn}{dt} \\ &= p_i \left( g(\mathbf{X}_i, \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_r) - \sum_j p_j g(\mathbf{X}_j, \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_r) \right) \\ &= p_i \left( \sum_s \left( \sum_k p_k \frac{\partial h_s}{\partial \mathbf{X}} \mathbf{U}_k \right) \frac{\partial^2 g}{\partial I_s \partial \mathbf{Y}} \mathbf{U}_i + \frac{1}{2} \mathbf{U}_i^\top \frac{\partial^2 g}{\partial \mathbf{Y}^2} \mathbf{U}_i \right) \\ &\quad - p_i \left( \sum_j p_j \left( \sum_s \left( \sum_k p_k \frac{\partial h_s}{\partial \mathbf{X}} \mathbf{U}_k \right) \frac{\partial^2 g}{\partial I_s \partial \mathbf{Y}} \mathbf{U}_j + \frac{1}{2} \mathbf{U}_j^\top \frac{\partial^2 g}{\partial \mathbf{Y}^2} \mathbf{U}_j \right) \right) + O(\varepsilon^3) \\ &= p_i \sum_{jk} p_j p_k (2\mathbf{U}_k^\top \mathbf{C}_{10}(\mathbf{U}_i - \mathbf{U}_j) + (\mathbf{U}_i + \mathbf{U}_j)^\top \mathbf{C}_{00}(\mathbf{U}_i - \mathbf{U}_j)) + O(\varepsilon^3) \end{aligned} \tag{127}$$

where we have used the correspondences

$$C_{00} = \frac{1}{2} \frac{\partial^2 g}{\partial Y^2} \quad C_{10} = \frac{1}{2} \sum_s \frac{\partial}{\partial Y} \left( \frac{\partial g}{\partial I_s} \frac{\partial h_s}{\partial X} \right)^\top = \frac{1}{2} \sum_s \frac{\partial h_s^\top}{\partial X} \frac{\partial^2 g}{\partial I_s \partial Y}$$

which are straightforward to derive, given the definition of  $s_X(Y)$  (117) and the expansion of  $g$  (126). Our last result (127) is identical to that for Lotka–Volterra systems (115) and shows that the dynamics of the fractions is the same, up to and including terms of order  $O(\varepsilon^2)$ , for all systems near evolutionary singularities.

**Appendix G: Resource utilization and abundance, in relation to invadability**

We will show here that in our example (Sect. 4.6), the matrix  $C_{00}$  is positive definite (resp. indefinite, negative, nonnegative or nonpositive definite) if and only if  $\Psi - \Phi$  is positive definite (resp. indefinite, negative, nonnegative or nonpositive definite).

*Proof* We start by decomposing the symmetric, positive definite matrix  $\Phi$  using a matrix  $V$  of normalized, orthogonal eigenvectors:

$$\exists V, \Lambda : \Phi V = V \Lambda \quad \text{for which} \quad \text{id} = V^\top V = V V^\top \tag{128}$$

where  $\text{id}$  is the identity and  $\Lambda$  is a diagonal matrix containing the (strictly positive) eigenvalues of  $\Phi$ . Using the above, we decompose  $\Phi$  and find

$$\Psi + \Phi = V \Lambda^{\frac{1}{2}} \left[ \Lambda^{-\frac{1}{2}} V^\top \Psi V \Lambda^{-\frac{1}{2}} + \text{id} \right] \Lambda^{\frac{1}{2}} V^\top \tag{129}$$

so that

$$\begin{aligned} C_{00} &= \frac{\Phi^{-1}}{4} - \frac{[\Psi + \Phi]^{-1}}{2} \\ &= V \Lambda^{-\frac{1}{2}} \left[ \frac{1}{4} \text{id} - \frac{1}{2} \left[ \text{id} + \Lambda^{-\frac{1}{2}} V^\top \Psi V \Lambda^{-\frac{1}{2}} \right]^{-1} \right] \Lambda^{-\frac{1}{2}} V^\top \end{aligned} \tag{130}$$

Defining  $C_{00}^* := \frac{1}{4} \text{id} - \frac{1}{2} \left[ \text{id} + \Lambda^{-\frac{1}{2}} V^\top \Psi V \Lambda^{-\frac{1}{2}} \right]^{-1}$ , it is easy to see that  $C_{00}$  has a positive (resp. negative, zero) eigenvalue for each positive (resp. negative, zero) eigenvalue of  $C_{00}^*$ : for any vector  $v$ , consider  $w := V^\top \Lambda^{\frac{1}{2}} v$  so that  $w^\top C_{00}^* w = v^\top C_{00} v$ . The same correspondence holds between  $[\Psi - \Phi]^* := \Lambda^{-\frac{1}{2}} V^\top \Psi V \Lambda^{-\frac{1}{2}} - \text{id}$  and  $\Psi - \Phi$ .

To conclude the proof, it suffices to remark that the matrices  $[\Psi - \Phi]^*$  and  $\mathbf{C}_{00}^*$  have the same eigenvectors, and more importantly that the signs of their eigenvalues coincide:

$$\begin{aligned} & \left[ \Lambda^{-\frac{1}{2}} \mathbf{V}^T \Psi \mathbf{V} \Lambda^{-\frac{1}{2}} - \text{id} \right] \mathbf{v} = \lambda \mathbf{v} \\ \Leftrightarrow & \left[ \Lambda^{-\frac{1}{2}} \mathbf{V}^T \Psi \mathbf{V} \Lambda^{-\frac{1}{2}} + \text{id} \right]^{-1} \mathbf{v} = (\lambda + 2)^{-1} \mathbf{v} \\ \Leftrightarrow & \mathbf{C}_{00}^* \mathbf{v} = \left[ \frac{1}{4} - \frac{1}{2}(\lambda + 2)^{-1} \right] \mathbf{v} = \frac{1}{4} \frac{\lambda}{\lambda + 2} \mathbf{v} \end{aligned}$$

where  $\lambda + 2$  is necessarily positive, as it is an eigenvalue of a sum of positive definite matrices.  $\square$

## References

1. Athreya, K.: Rates of decay for the survival probability of a mutant gene. *J. Math. Biol.* **30**, 577–581 (1992)
2. Athreya, K.: Rates of decay for the survival probability of a mutant gene. II. The multitype case. *J. Math. Biol.* **32**, 45–53 (1993)
3. Beccari, C.: Tutorial: typesetting maths for science and tech according to ISO 31/XI. Technical report, T<sub>E</sub>X Users Group (1997)
4. Caswell, H.: *Population models: construction, analysis, and interpretation*, 2nd edn. Sinauer Associates, Sunderland (2001)
5. Champagnat, N.: Convergence of adaptive dynamics n-morphic jump processes to the canonical equation and degenerate diffusion approximation. Prépublication de l'Université de Nanterre (Paris X) no. 03/7 (2003)
6. Champagnat, N.: Etude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptatives. PhD thesis, Paris X, Paris (France) (2004)
7. Champagnat, N.: A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Process. Appl.* **116**, 1127–1160 (2006)
8. Champagnat, N., Ferrière, R., Ben Arous, G.: The canonical equation of adaptive dynamics: a mathematical view. *Selection* **2**, 71–81 (2001)
9. Champagnat, N., Ferrière, R., Méléard, S.: Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theor. Popul. Biol.* **69**, 297–321 (2006)
10. Christiansen, F.: On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* **138**, 37–50 (1991)
11. Christiansen, F., Loeschke, V.: Evolution and intraspecific competition. III. One-locus theory for small additive gene effects and multidimensional resource qualities. *Theor. Popul. Biol.* **31**, 33–46 (1987)
12. Cressman, R., Hofbauer, J.: Measure dynamics on a one-dimensional continuous trait space: Theoretical foundations for adaptive dynamics. *Theor. Popul. Biol.* **67**, 47–59 (2005)
13. Dercole, F.: Evolutionary dynamics through bifurcation analysis: methods and applications. PhD thesis, Department of Electronics and Information, Politecnico di Milano, Milano (Italy) (2002)
14. Dercole, F., Rinaldi, S.: Analysis of evolutionary processes: the adaptive dynamics approach and its applications. Princeton University Press, Princeton (in press)
15. Dieckmann, U., Law, R.: The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**, 579–612 (1996)
16. Dieckmann, U., Marrow, P., Law, R.: Evolutionary cycling in predator–prey interactions: population dynamics and the red queen. *J. Theor. Biol.* **176**, 91–102 (1995)
17. Dieckmann, U., Metz, J.: Surprising evolutionary predictions from enhanced ecological realism. *Theor. Popul. Biol.* **69**, 263–281 (2006)
18. Dieckmann, O.: A beginner's guide to adaptive dynamics. *Math. Model. Popul. Dyn.* **63**, 47–84 (2004)

19. Diekmann, O., Getto, P., Gyllenberg, M.: Stability and bifurcation analysis of Volterra functional equations in the light of suns and stars. *SIAM J. Math. Anal.* (in press)
20. Diekmann, O., Gyllenberg, M., Huang, H., Kirkilionis, M., Metz, J., Thieme, H.: On the formulation and analysis of general deterministic structured population models. II. Nonlinear theory. *J. Math. Biol.* **43**, 157–189 (2001)
21. Diekmann, O., Gyllenberg, M., Metz, J.: Steady state analysis of structured population models. *Theor. Popul. Biol.* **63**, 309–338 (2003)
22. Eshel, I.: On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size—a branching process model. *J. Math. Biol.* **12**, 355–362 (1981)
23. Eshel, I.: On the survival probability of a slightly advantageous mutant gene in a multitype population: a multidimensional branching process model. *J. Math. Biol.* **19**, 201–209 (1984)
24. Ewens, W.: *Population Genetics*. Methuen, London (1969)
25. Geritz, S.: Resident-invader dynamics and the coexistence of similar strategies. *J. Math. Biol.* **50**, 67–82 (2005)
26. Geritz, S., Gyllenberg, M., Jacobs, F., Parvinen, K.: Invasion dynamics and attractor inheritance. *J. Math. Biol.* **44**, 548–560 (2002)
27. Geritz, S., Kisdi, E., Meszena, G., Metz, J.: Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57 (1998)
28. Geritz, S., van der Meijden, E., Metz, J.: Evolutionary dynamics of seed size and seedling competitive ability. *Theor. Popul. Biol.* **55**, 324–343 (1999)
29. Getto, P.: On some quasilinear structured population models. Thesis, Utrecht University, Utrecht (the Netherlands) (2005)
30. Gillespie, D.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976)
31. Greiner, G., Heesterbeek, J., Metz, J.: A singular perturbation theorem for evolution equations and time-scale arguments for structured population models. *Can. Appl. Math. Q.* **2**, 435–459 (1994)
32. Haccou, P., Jagers, P., Vatutin, V.: *Branching Processes—Variation, Growth, and Extinction of Populations*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, Cambridge (2005)
33. Haldane, J.: A mathematical theory of natural and artificial selection. v. selection and mutation. *Proc. Camb. Philos. Soc.* **23**, 838–844 (1927)
34. Hofbauer, J., Sigmund, K.: *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge (1998)
35. Hoppe, F.: Asymptotic rates of growth of the extinction probability of a mutant gene. *J. Math. Biol.* **30**, 547–566 (1992)
36. Kendall, D.: On the generalized “birth-and-death” process. *Ann. Math. Stat.* **19**, 1–15 (1948)
37. Kondrashov, A., Turelli, M.: Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. *Genetics* **132**, 603–618 (1992)
38. Lande, R.: Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402–416 (1979)
39. Leimar, O.: Evolutionary change and darwinian demons. *Selection* **2**(1–2), 65–72 (2001)
40. Leimar, O.: The evolution of phenotypic polymorphism: randomized strategies versus evolutionary branching. *Am. Nat.* **165**, 669–681 (2005)
41. Leimar, O.: Multidimensional convergence stability and the canonical adaptive dynamics. In: Dieckmann, U., Metz, J.A.J. (eds.) *Elements of Adaptive Dynamics*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, Cambridge (UK) (to appear)
42. Levin, S.: Community equilibria and stability, and an extension of the competitive exclusion principle. *Am. Nat.* **104**, 413–423 (1970)
43. Lush, J.: *Animal Breeding Plans*. Iowa State College Press, Ames (1937)
44. Magnus, J., Neudecker, H.: *Matrix Differential Calculus with applications in Statistics and Econometrics*. Wiley, New York (1988)
45. Meszena, G., Gyllenberg, M., Jacobs, F., Metz, J.: Link between population dynamics and dynamics of darwinian evolution. *Phys. Rev. Lett.* **95**(078105), 1–4 (2005)
46. Meszena, G., Gyllenberg, M., Pasztor, L., Metz, J.: Competitive exclusion and limiting similarity: a unified theory. *Theor. Popul. Biol.* **69**, 68–87 (2006)
47. Meszena, G., Metz, J.: Species diversity and population regulation: the importance of environmental feedback dimensionality. IIASA Interim Report IR-99-045 (1999)
48. Metz, J., Diekmann, O.: *The Dynamics of Physiologically Structured Populations*, volume 68 of *Lecture Notes in Biomathematics*. Springer, Berlin (1986)

49. Metz, J., Geritz, S., Meszéna, G., Jacobs, F., Van Heerwaarden, J.: Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In: Van Strien, S.J., Verduyn Lunel, S.M. (eds.) *Stochastic and Spatial Structures of Dynamical Systems*, volume 45 of KNAW Symposium Lectures, Section Science, First Series, pp. 183–231. North-Holland, Amsterdam (1996)
50. Metz, J., Nisbet, R., Geritz, S.: How should we define fitness for general ecological scenarios? *Trends Ecol. Evol.* **7**, 198–202 (1992)
51. Mollison, D.: The structure of epidemic models. In: *Epidemic Models: their Structure and Relation to Data*, pp. 27–28. Cambridge University Press, Cambridge (1995)
52. Pigliucci, M.: Genetic variance–covariance matrices: a critique of the evolutionary quantitative genetics research program. *Biol. Philos.* **21**, 1–23 (2006)
53. Schneider, K.: A multilocus-multiallele analysis of frequency-dependent selection induced by intraspecific competition. *J. Math. Biol.* **52**, 483–523 (2006)
54. Turelli, M.: Heritable genetic variation via mutation-selection balance: Lerch’s zeta meets the abdominal bristle. *Theor. Popul. Biol.* **25**, 138–193 (1984)
55. van Kampen, N.: *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam (1981)
56. Vukics, A., Asbóth, J., Meszéna, G.: Speciation in multidimensional evolutionary space. *Phys. Rev. E* **68**(041903), 1–10 (2003)
57. Zhang, X., Hill, W.: Genetic variability under mutation selection balance. *Trends Ecol. Evol.* **20**, 468–470 (2005)