J. Math. Biol. (2016) 72:1125–1152
DOI 10.1007/s00285-015-0938-4

Mathematical Biology

The canonical equation of adaptive dynamics for life histories: from fitness-returns to selection gradients and Pontryagin's maximum principle

Johan A. Jacob Metz[1,2,3] · **Kateřina Staňková**[4,5] · **Jacob Johansson**[2,6]

Received: 30 December 2014 / Revised: 3 August 2015 / Published online: 19 November 2015

Abstract This paper should be read as addendum to Dieckmann et al. (J Theor Biol 241:370–389, 2006) and Parvinen et al. (J Math Biol 67: 509–533, 2013). Our goal is, using little more than high-school calculus, to (1) exhibit the form of the canonical equation of adaptive dynamics for classical life history problems, where the examples in Dieckmann et al. (J Theor Biol 241:370–389, 2006) and Parvinen et al. (J Math Biol 67: 509–533, 2013) are chosen such that they avoid a number of the problems that one gets in this most relevant of applications, (2) derive the fitness gradient occurring in the CE from simple fitness return arguments, (3) show explicitly that setting said

We dedicate this paper to Mats Gyllenberg's 60th anniversary and 61st birthday.

✉ Kateřina Staňková
k.stankova@maastrichtuniversity.nl

Johan A. Jacob Metz
j.a.j.metz@biology.leidenuniv.nl

Jacob Johansson
Jacob.Johansson@biol.lu.se

1 Mathematical Institute and Institute of Biology, Leiden University, 2333 CA Leiden, The Netherlands

2 Evolution and Ecology Program, International Institute for Applied Systems Analysis, 2361 Laxenburg, Austria

3 Department of Marine Zoology, Naturalis Biodiversity Center, 2333 CR Leiden, The Netherlands

4 Department of Knowledge Engineering, Maastricht University, 6211 LH Maastricht, The Netherlands

5 Delft Institute of Applied Mathematics, Delft University of Technology, 2628 CD Delft, The Netherlands

6 Theoretical Population Ecology and Evolution Group, Department of Biology, Lund University, 22362 Lund, Sweden

fitness gradient equal to zero results in the classical marginal value principle from evolutionary ecology, (4) show that the latter in turn is equivalent to Pontryagin's maximum principle, a well known equivalence that however in the literature is given either ex cathedra or is proven with more advanced tools, (5) connect the classical optimisation arguments of life history theory a little better to real biology (Mendelian populations with separate sexes subject to an environmental feedback loop), (6) make a minor improvement to the form of the CE for the examples in Dieckmann et al. and Parvinen et al.

Keywords Canonical equation of adaptive dynamics · Function valued traits · Pontryagin's maximum principle · Age-dependent resource allocation · Mendelian take on life history theory · Evolution in periodic environments

Mathematics Subject Classification 92D15 · 92D40 · 37N25 · 49K15 · 49N90

1 Introduction

In their recent paper "Function-valued adaptive dynamics and optimal control theory", Parvinen et al. (2013) give (i) an abstract recipe for calculating the selection gradient for function valued traits affecting the i(ndividual)-dynamics of physiologically structured populations for use in the canonical equation of adaptive dynamics (in the terminology of Metz and Diekmann (1986); Parvinen et al. refer to these models as process-mediated) and (ii) a recipe for calculating the corresponding evolutionarily steady strategies (ESS-es) by using Pontryagin's maximum principle (c.q. evolutionarily singular strategies (ess-es) if we confine ourselves to the first order condition derived from this principle). They subsequently apply these recipes to derive concrete expressions for three sample models. However, they do not explicitly consider the relationship between (i) and (ii) but for numerically demonstrating that for their special models the adaptive trajectories approach the ess. In this note we (i) demonstrate how the selection gradient can be calculated from a concrete starting point by using the idea of fitness returns, which gives an interpretation to the components of the resulting formulas, and (ii) show that setting the selection gradient equal to zero leads to a classical marginal value argument which turns out to be equivalent to the local version of Pontryagin's maximum principle.

Terminology We employ the term fitness return here for a concept that is widely used in evolutionary ecology, often also under this name, but for which we failed to find an explicit definition. If some fitness proxy can be decomposed as the sum of a number of terms that supposedly stand for the contributions of different pathways by which fitness can accrue, we call the effect of a strategy change on the contribution of a pathway the fitness return through that pathway. For a global ESS the sum of all fitness returns is non-positive whatever the strategy change. For local ESS-es we consider only the fitness returns of infinitesimal strategy changes. To accord with common usage these should be called marginal returns. However, as these are the only returns that we consider we shall drop the epithet. When the attention is confined to an infinitesimal neighbourhood of a reference strategy far more fitness proxies allow a conceptually

useful additive decomposition thanks to the rules of differential calculus. All that is needed is a biologically interpretable way in which the proxy can be decomposed as a differentiable function of a number of differentiable functions of the strategy. The (marginal) fitness return through one of these functions is then defined as the sensitivity of the proxy to the strategy change in a thought experiment in which we keep the argument of all other functions unchanged. The fitness returns from state dependent decisions are usually determined from first principles conditional on the state under consideration. The epithet conditional is customarily dropped in this case. The (marginal) fitness return from a compound decision involving more than one state is calculated by summing the fitness returns for the separate states weighted with their lifetime occurrence frequencies or duration.

To keep the arguments accessible for evolutionary ecologists, we restrict our calculations from the start to the most commonly encountered class of life history models and use the simplest possible mathematical arguments rather than a more advanced functional analytical approach. In the appendices we will sketch how the same arguments can be obtained more rigorously. Basically we assume our readers to be knowledgeable only about demography and the attendant elementary results from probability theory, but not about systems theory or dynamic optimisation.

2 On selection gradients, canonical equations, and evolutionarily singular strategies, a summary

Below we consider a life history model in which individuals are characterised by two dynamical variables, a physiological state, assumed to move deterministically, and a probability of still being alive, in addition to an inherited strategy u influencing their dynamics. The strategy u (or u_{res} if we talk specifically about the resident's strategy, or u_{mut} if we talk about a mutant strategy) is supposed to be a function of the state of the individual taking values in $[0, 1]$. To make our life simple we assume that on the population dynamical time scale the community dynamics converges to an equilibrium, which generates the non-fluctuating environment $E_{\mathrm{res}} = E_{\mathrm{attr}}(u_{\mathrm{res}})$, with u_{res} the strategy currently in residence. This assumption of a non-fluctuating environment allows us to make use of the fitness proxy $R_0(u_{\mathrm{mut}}; E_{\mathrm{res}})$, the average lifetime offspring production of a mutant in the environment E_{res}, calculated e.g. by integrating the average rate of producing offspring over the age of an individual. Consistency requires that $R_0(u_{\mathrm{res}}; E_{\mathrm{res}}) = 1$. If $R_0(u_{\mathrm{mut}}; E_{\mathrm{res}}) > 1$, a mutant has a positive probability to invade, else it cannot invade. The invasion fitness $F(u_{\mathrm{mut}}; E_{\mathrm{res}})$ of a mutant is by definition equal to the asymptotic exponential growth rate of a mutant population in the environment E_{res} (Metz et al. 1992; Metz 2008). For R_0 close to 1 this invasion fitness is well approximated by

$$F(u_{\mathrm{mut}}; E_{\mathrm{res}}) = \frac{\ln(R_0(u_{\mathrm{mut}}; E_{\mathrm{res}}))}{T_{\mathrm{r}}(u_{\mathrm{res}})} + \mathrm{O}(\ln^2(R_0(u_{\mathrm{mut}}; E_{\mathrm{res}}))), \tag{1}$$

where $T_{\mathrm{r}}(u_{\mathrm{res}})$ is the average age at which the residents give birth in the environment E_{res} (Metz and Diekmann 1986; Durinx et al. 2008).

Remark 1 Dieckmann et al. (2006) and Parvinen et al. (2013) consider seasonal differential-equation-based models where it is possible to calculate the invasion fitness directly by subtracting the time-averaged death rate from the time-averaged birth rate. For such models fitness takes the explicit form of an integral over the year cycle, and there is no need to fall back on an approximation. However, in the usual continuous time life history models only R_0 can be expressed explicitly as an integral. The availability of such an integral-based expression formed the basis for the developments in Parvinen et al. (2013), and will also be the cornerstone for our calculations.

The so-called selection gradient G tells how the invasion fitness of a u_{mut} close to u_{res} depends on the difference $u_{\mathrm{mut}} - u_{\mathrm{res}}$. Mathematically, the selection gradient is the derivative of the invasion fitness for u_{mut} evaluated at $u_{\mathrm{mut}} = u_{\mathrm{res}} = u$. From the previous approximation formula for the invasion fitness it follows that we can calculate G as

$$G(u) = \frac{1}{T_{\mathrm{r}}(u)} \frac{\mathrm{d}R_0}{\mathrm{d}u_{\mathrm{mut}}}(u;\, E_{\mathrm{attr}}(u)) \tag{2}$$

(Durinx et al. 2008; Metz 2008). In this formula $\frac{\mathrm{d}R_0}{\mathrm{d}u_{\mathrm{mut}}}$ is an abstract differential quotient, i.e. a linear map transforming functions of the physiological state into a number that linearly approximates the nonlinear dependence of R_0 on u_{mut}.

In view of our stress on life history models, let us moreover assume that u is an allocation, so that u takes values in $[0, 1]$. The assumptions of a non-fluctuating resident environment and a deterministically moving physiological state moreover allow us to represent the strategy u as a function of the age a of an individual, i.e., $u : \mathbb{R}_+ \to [0, 1]$. In that case we can write for a function $x : \mathbb{R}_+ \to \mathbb{R}$:

$$G(u)x = \frac{1}{T_{\mathrm{r}}(u)} \int_0^\infty g(a; u)x(a)\, \mathrm{d}a \tag{3}$$

(c.f., Parvinen et al. 2013). Hence the problem of calculating G reduces to that of calculating the function g. On the assumption that mutations are rare and mutational steps small the dynamics of u can on the evolutionary time scale be described by the so-called canonical equation (CE) of adaptive dynamics (Dieckmann and Law 1996; Champagnat 2003; Dieckmann et al. 2006; Parvinen et al. 2006, 2013; Durinx et al. 2008; Méléard and Tran 2009; Champagnat and Méléard 2011; Gupta et al. 2014)

$$\frac{\mathrm{d}u}{\mathrm{d}t}(a) \begin{cases} = \frac{\sigma^{-2}(u)}{T_{\mathrm{s}}(u)} \overline{n}\,(u)\ \mu \int_0^\infty c(a, \alpha)\, g(\alpha; u)\, \mathrm{d}\alpha, \\ \quad \text{if} \quad \int_0^\infty c(a, \alpha)\, g(\alpha; u)\, \mathrm{d}\alpha > 0 \quad \text{and} \quad u(a) \in [0, 1), \\ \quad \text{or} \quad \int_0^\infty c(a, \alpha)\, g(\alpha; u)\, \mathrm{d}\alpha < 0 \quad \text{and} \quad u(a) \in (0, 1], \\ = 0 \quad \text{otherwise}, \end{cases} \tag{4}$$

with T_{s} the average age at which the residents die, σ^2 the between individual variance of their offspring numbers (i.e., if $\underline{m}_i$ is a lifetime offspring number of the i-th individual, $\sigma^2 = \mathrm{Var}\,(\underline{m}_i)$), $\overline{n}$ their equilibrium population size, μ the (small) probability at a birth

event of a mutation affecting u, and c the (small) covariance kernel of the mutational steps, i.e., if $\underline{x}$ denotes a mutational step in u, then

$$\mathrm{Cov}\left(\int_{a_1}^{a_2} \underline{x}(\alpha)\,\mathrm{d}\alpha, \int_{a_3}^{a_4} \underline{x}(\zeta)\,\mathrm{d}\zeta\right) = \int_{a_1}^{a_2}\int_{a_3}^{a_4} c(\alpha,\zeta)\,\mathrm{d}\zeta\,\mathrm{d}\alpha. \tag{5}$$

The form of the CE given above is the one for clonally reproducing organisms (the customary assumption in most of life history theory which, however, usually is left implicit). In Appendix 1 we briefly consider its extension to Mendelian diploids.

Our formula for the CE is slightly more complicated than the one in Dieckmann et al. (2006) and Parvinen et al. (2013). The reason is that these authors did not consider local constraints on the strategy, whereas in our case $u(a) \in [0, 1]$, for each possible age $a > 0$. See Appendix 2 for further information. Another difference is that Dieckmann et al. (2006) and Parvinen et al. (2013) have set the factor σ^2 equal to 2, in keeping with the idea that for the i-models underlying the standard ordinary differential equation (ODE) models the distribution of the lifetime offspring number is geometric. Moreover, for the standard ODE models $T_\mathrm{r} = T_\mathrm{s}$ and since the g in Dieckmann et al. (2006) and Parvinen et al. (2013) corresponds to our $f \overset{\mathrm{def}}{=} g/T_\mathrm{r}$, the T_s in (4) cancels. Appendix 3 treats the corresponding considerations for the periodic ODE case considered by Dieckmann et al. (2006) and Parvinen et al. (2013), with as outcome that in this case their $\overline{n}$ should be interpreted as a harmonic death-rate weighted mean of the population sizes over a cycle.

The equilibria of the CE are the so-called ess-es. If these strategies are moreover (local) fitness maxima for the corresponding E_res then they are also evolutionary equilibria, to which we refer as (local) ESS-es. (An alternative is that at an attracting ess the population starts to accumulate variation, so that it no longer stays quasi-monomorphic as is supposed in the derivation of the CE. (The latter on good grounds: see Geritz et al. 2002; Geritz 2005; Dercole and Rinaldi 2008, Appendix 2).) Another way to calculate ESS-es is to maximise the invasion fitness, or alternatively R_0, over u_mut, leading to a function-valued map[1] $u^*_\mathrm{mut}(u_\mathrm{res})$, followed by solving the equation $u^*_\mathrm{mut}(u_\mathrm{res}) = u_\mathrm{res}$. It is here that Pontryagin's maximum principle is encountered (e.g., Pontryagin et al. 1964; Intrilligator 1971). This principle is derived by considering the differential equations for the i-states as constraints on their time development, and to extend the idea of Lagrange multipliers as encountered in finite dimensional optimisation problems to this case. The Lagrange multipliers then become functions of time, which can be shown to satisfy a set of differential equations, and for this reason are referred to as co-states (or adjoints). In Sect. 6 we give explicit expressions for the life history models described in the next section. Appendix 5 shows how Pontryagin's maximum principle can be derived directly from a weak variant of Bellman's principle of optimality, which is rather better known among ecologists.

[1] Here star denotes optimality.

3 Model ingredients

Before we get to the specifics we first introduce some notational conventions in order to keep our formulas from becoming too unwieldy.

Conventions

1. The argument $E_{\rm res}$ will be usually hidden.
2. Similarly we shall often hide the argument u in expressions like $P(a;u)$ for the probability that an individual survives till age a, or $m(a;u)$ for its body size at that age.
3. When we use the argument u, then u stands either for $u_{\rm mut}$, $u_{\rm res}$, or $u_{\rm mut}=u_{\rm res}$, with the context making clear which is the case.
4. For a function of a single scalar variable we use a prime to indicate its derivative. A superscript dot indicates a derivative for age, also when a function has other arguments as well.

The two dynamical variables characterising an individual are (i) one i-state variable, to wit its body mass m, increasing from a fixed birth mass $m(0)=m_0$, and (ii) its probability P to be still alive, starting from $P(0)=1$. The energy intake by an individual with body mass m will be denoted by $e(m)$. The strategy of an individual will be denoted by $u:\mathbb{R}_+\to[0,1]:a\mapsto u(a)$, where $u(a)$ determines which fraction of the energy intake at age a is used for reproduction while the remains are used for growth. The body mass just increases with $(1-u)e(m)$, while the birth rate is assumed to depend monotonically on the available energy $u\,e(m)$ as $b:\mathbb{R}_+\to\mathbb{R}_+:u(a)\,e\,(m(a))\mapsto b\,(u(a)\,e\,(m(a)))$. Finally, the energy allocation is assumed also to affect the death rate $d:[0,1]\to\mathbb{R}_+:u(a)\mapsto d\,(u(a))$. All three functions e, b, and d also implicitly depend on $E_{\rm res}$. In this model the average lifetime offspring number of a mutant strategy $u_{\rm mut}$ equals

$$R_0=\int_0^\infty P(a)b\,(u_{\rm mut}(a)e\,(m(a)))\;{\rm d}a,$$

with m and P solving

$$\begin{aligned}\dot m&=(1-u_{\rm mut})\,e(m),\quad m(0)=m_0,\\ \dot P&=-d(u_{\rm mut})P,\quad P(0)=1.\end{aligned}\tag{6}$$

Note that if $u_{\rm mut}=u_{\rm res}$, due to the value of $E_{\rm res}$ necessarily $R_0=1$. Moreover, we assume that the tail of P is bounded by a negative exponential and that b and e are bounded. These assumptions derive from the biology behind the example and imply that the improper integral in (6) exists.

4 Calculating the selection gradient from a fitness-returns argument

We shall express g from Eq. (3) in terms of the fitness returns r, that is, the proportional effects of small local changes in u on the total future reproduction. To calculate $r(a;u)$

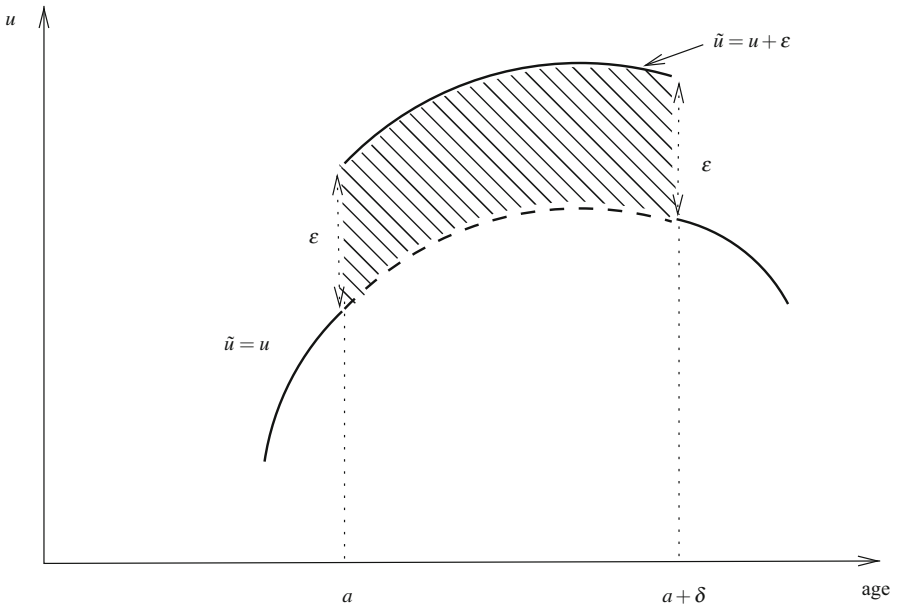

Fig. 1 Illustration of the fitness-returns argument where u is any control path. The idea is to do a thought experiment in which we increase u between a and $a + \delta$ with a block B of height ε, and observe its effect on fitness

we proceed by means of a thought experiment. For a living individual aged a we increase u between a and $a+\delta$ by an amount ε, i.e., we construct a function $\tilde{u} = u+B$, $B : \mathbb{R}_+ \to \mathbb{R}$, $B(\alpha) = \varepsilon$ for $a \le \alpha < a + \delta$ and 0 elsewhere (see Fig. 1),

Calculate the resulting expected change in the expected life-time offspring number, multiply this number with $(\varepsilon\delta)^{-1}$ and let both ε and δ go to zero. Since the fitness return by definition is calculated conditional on an individual surviving to a, only the fraction $P(a; u)$ surviving till age a contributes in this manner to R_0. Hence

$$g = P\, r. \tag{7}$$

To calculate those expected additional offspring numbers we proceed in the spirit of the marginal value theorem, that is, we calculate and then add the components of r contributed by different routes. These components include the immediate additional number of offspring coming from the temporary increase in energy allocation to reproduction and the decreases in future offspring numbers caused by the future smaller size and lesser survival caused by the temporary decrease in allocation to growth and to staving off death. We start with the calculation of the second and third components. Let Δ_m and Δ_P denote the differences $m(\tilde{u}) - m(u)$ and $P(\tilde{u}) - P(u)$, respectively. Then, by a first-order Taylor expansion of $P(\alpha)$ and $m(\alpha)$ with respect to $u(\alpha)$, we obtain for $a \le \alpha < a + \delta$

$$\begin{aligned} \dot{\Delta}_m &\approx (1-u)e'(m)\,\Delta_m - \varepsilon\, e(m), \quad \Delta_m(a) = 0, \\ \dot{\Delta}_P &\approx -d(u)\,\Delta_P - d'(u)\,P\,\varepsilon, \quad \Delta_P(a) = 0, \end{aligned} \tag{8}$$

which implies

$$\Delta_m(a+\delta) \approx -e(m(a))\varepsilon\,\delta, \quad \Delta_P(a+\delta) \approx -d'(u(a))P(a)\,\varepsilon\,\delta. \tag{9}$$

The immediate offspring gain from this strategy change over the time interval $[a, a+\delta)$ (for an individual that survived till a) is

$$\varepsilon\delta[b'(u(a)e(m(a)))e(m(a))]. \tag{10}$$

From $a+\delta$ onwards Eq. (8) apply with ε set to zero and with initial condition (9).

The future loss of offspring from this change in strategy for an individual that already has survived till a is

$$-\frac{1}{P(a)}\int_a^\infty \left(\Delta_P(\alpha)\,b\,(u(\alpha)\,e(m(\alpha)))+P(\alpha)\,b'\,(u(\alpha)\,e(m(\alpha)))\,u(\alpha)\,e'(m(\alpha))\Delta_m(\alpha)\right)\mathrm{d}\alpha. \tag{11}$$

The linearity of Eq. (8) with $\varepsilon = 0$ implies that $\Delta_{\mathrm{P}}(\alpha)$ and $\Delta_{\mathrm{m}}(\alpha)$ are linearly dependent on the initial conditions given by (9), and therefore the outcome (11) is proportional to $\varepsilon\delta$. To make the coming calculation more transparent we introduce new functions $\hat{P}(\alpha; a)$, $\hat{\Delta}_m(\alpha; a)$, $\alpha \geq a$, defined by

$$\begin{aligned}\frac{\mathrm{d}\hat{P}}{\mathrm{d}\alpha} &= -d(u)\hat{P}, \quad \hat{P}(a;a) = 1,\\ \frac{\mathrm{d}\hat{\Delta}_m}{\mathrm{d}\alpha} &= (1-u)e'(m)\,\hat{\Delta}_m, \quad \hat{\Delta}_m(a;a) = 1,\end{aligned} \tag{12}$$

where $\hat{P}(\alpha; a)$ can be interpreted as the conditional survival of an individual that has already survived to age a, i.e., $\hat{P}(\alpha; a) = P(\alpha)/P(a)$, given the strategy u, and $\hat{\Delta}_m(\alpha; \alpha+\delta) = \Delta_m(\alpha)/\Delta_m(a+\delta)$ as the relative amount by which a small perturbation in m present at age $a+\delta$ will propagate into the future given u. Similarly $\hat{P}(\alpha; a+\delta)$ allows an alternative interpretation as relative amount by which a small perturbation in m present at age $a+\delta$ will propagate into the future given u. For $\varepsilon \downarrow 0$ and $\delta \downarrow 0$ we can then express the fitness return $r(a; u)$ as follows:

$$\begin{aligned} r(a;u) = {} & b'(u(a)e(m(a)))e(m(a)) \\ & - e(m(a))\int_a^\infty \hat{P}(\alpha;a)b'(u(\alpha)e(m(\alpha)))u(\alpha)e'(m(\alpha))\hat{\Delta}_m(\alpha;a)\mathrm{d}\alpha \\ & - d'(u(a))\int_a^\infty \hat{P}(\alpha;a)\,b(u(\alpha)e(m(\alpha)))\mathrm{d}\alpha. \end{aligned} \tag{13}$$

At an ess u^* the return $r(a)$ should be 0 when $u^*(a) \in (0, 1)$, non-positive when $u^*(a) = 0$ and non-negative when $u^*(a) = 1$.

5 The other ingredients of the canonical equation

To complete the canonical equation we need to find expressions for T_r, T_s, and σ^2. Since $R_0(u; E_{\mathrm{attr}}(u)) = 1$, the expression $P(\alpha; u)b(u(\alpha)e(m(\alpha; u)))$ is a probability density. Furthermore,

$$-\int_0^\infty \dot{P}(\alpha; u)\,\mathrm{d}\alpha = P(0) - \lim_{A\to\infty} P(A) = 1, \tag{14}$$

thus also $-\dot{P}(\alpha; u)$ is a probability density. Therefore, T_r and T_s can be expressed directly:

$$T_\mathrm{r}(u) = \int_0^\infty \alpha\, P(\alpha; u)b(u(\alpha)e(m(\alpha; u)))\mathrm{d}\alpha, \tag{15}$$

$$T_\mathrm{s}(u) = -\int_0^\infty \alpha\dot{P}(\alpha; u)\mathrm{d}\alpha = -\lim_{A\to\infty} \alpha P(\alpha)|_{\alpha=0}^{A} + \int_0^\infty P(\alpha; u)\mathrm{d}\alpha = \int_0^\infty P(\alpha; u)\mathrm{d}\alpha, \tag{16}$$

where the last equation comes from integrating by parts and using $\lim_{A\to\infty}(AP(A)) = 0$.

To calculate σ^2 we have to be more specific about the microstructure of the reproduction process. The assumption that naturally leads to (5) is that for an individual that is still alive the births come in a Poisson process with rate $b(ue(m))$, or, slightly more generally, in clutches of average size $C(u, e(m))$ produced according to a Poisson process with rate $\frac{b(ue(m))}{C(u,e(m))}$. We confine ourselves here to the former option. In such a case, for a given age at death a the total offspring number is Poisson distributed with mean

$$\lambda(a; u) = \int_0^a b\,(u(\alpha)e(m(\alpha; u)))\,\mathrm{d}\alpha. \tag{17}$$

In general, a is a realisation of a random variable $\underline{a}$. Hence, the lifetime offspring number is a mixture of Poisson random variables. The mean of $\underline{\lambda} = \lambda(\underline{a}; u)$ is nothing but the average lifetime offspring number[2]

$$\mathbf{E}(\underline{\lambda}) = R_0\,(u; E_{\mathrm{attr}}(u)) = 1. \tag{18}$$

[2] Using integration by parts, $\mathbf{E}(\lambda) = -\int_0^\infty \dot{P}(\alpha; u)\,\lambda(\alpha; u)\mathrm{d}\alpha = P(0)\lambda(0) - \lim_{T\to\infty} P(T)\lim_{T\to\infty} \lambda(T) + R_0 = R_0$, because $P(0)\lambda(0) = \lim_{T\to\infty} P(T)\lim_{T\to\infty}\lambda(T) = 0$.

Therefore, $\sigma_\lambda^2 = \mathbf{E}(\lambda^2) - (\mathbf{E}(\lambda))^2 = \mathbf{E}(\lambda^2) - 1$, with

$$\mathbf{E}(\underline{\lambda}^2) = -\int_0^\infty \dot{P}(\alpha;u)\,\lambda(\alpha;u)\mathrm{d}\alpha = 2\int_0^\infty \lambda(\alpha;u)P(\alpha;u)b(u(\alpha)e(m(\alpha;u)))\mathrm{d}\alpha \tag{19}$$

(integration by parts).

Finally, from the general rules for mixtures of distributions[3]

$$\sigma^2 = \sigma_\lambda^2 + 1 = \mathbf{E}(\underline{\lambda}^2). \tag{20}$$

6 Locating fitness maxima by means of Pontryagin's maximum principle

The equilibria of the canonical equation are called ess-es. The reason for this from a differential equations viewpoint unusual terminology is that among the ess-es only the ESSes, characterised by the fact that they are also maxima of the current fitness landscapes, are immune to evolutionary change. One way of calculating ESSes for life history problems is to make use of Pontryagin's maximum principle to locate the fitness maxima in u_{mut} that go with a given u_{res} and then to set $u_{\mathrm{mut}} = u_{\mathrm{res}}$.

In contrast to the canonical equation, Pontryagin's maximum principle, at least in its original form and with a number of standard assumptions satisfied, is textbook material. In this section we will just in the wake of Intrilligator (1971) state the conditions that an optimal u has to satisfy. For a discussion of different variants of Pontryagin's maximum principle and its connection to the Bellman's principle of optimality (Bellman 1957), see Appendix 5. In the notation of Intrilligator (1971), Eq. (6) can be rewritten in the following form:

$$\mathbf{x} = \begin{pmatrix} m \\ P \end{pmatrix}, \quad I(\mathbf{x},u) = P\,b(ue(m)), \quad J = R_0, \quad \mathbf{f}(\mathbf{x},u) = \begin{pmatrix} (1-u)e(m) \\ -d(u)P \end{pmatrix}, \tag{21}$$

with $\mathbf{x}$ the state vector, J the quantity to be optimised, calculated as the lifetime integral of I, and $\mathbf{f}$ the right hand side of the differential equation for $\mathbf{x}$. Pontryagin's maximum principle then says that if u^* maximises J, then at each age $a \in [0,\infty)$ it also maximises the so-called Hamiltonian, defined as

$$H \stackrel{\mathrm{def}}{=} I + \mathbf{y}^T\mathbf{f}, \tag{22}$$

with $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ being the so-called co-state (or adjoint) vector, where its components satisfy the differential equations

[3] Let $\underline{m}$ be the lifetime offspring number. Then $1 = R_0 \stackrel{\mathrm{def}}{=} \mathbf{E}(\underline{m}) = \mathbf{E}(\underline{m}|\underline{a}) = \mathbf{E}(\lambda(\underline{a}))$. Since for a Poisson random variable with mean λ its variance also equals λ, $\sigma^2 = \mathrm{Var}(\underline{m}) = \mathbf{E}(\underline{m}^2) - (\mathbf{E}(\underline{m}))^2 = \mathbf{E}(\mathbf{E}(\underline{m}^2|\underline{a})) - 1 = \mathbf{E}(\mathrm{Var}(\underline{m}|\underline{a}) + (\mathbf{E}(\underline{m}|\underline{a}))^2) - 1 = \mathbf{E}(\lambda(\underline{a})) + \mathbf{E}(\lambda^2(\underline{a})) - 1 = \mathbf{E}(\lambda^2(\underline{a})) = \mathbf{E}(\underline{\lambda}^2)$.

$$\dot{y}_1 = -\frac{\partial H}{\partial m}, \quad \dot{y}_2 = -\frac{\partial H}{\partial P}, \tag{23}$$

with final conditions $\lim_{A\to\infty} y_1(A) = \lim_{A\to\infty} y_2(A) = 0$. If u^* maximises R_0 in (6), then it also maximises

$$\begin{aligned} &H(m(a), P(a), u(a)) = P(a)\, b\,(u(a)e\,(m(a))) \\ &\quad + y_1(a)\,(1-u(a))\,e\,(m(a)) - P(a)\,y_2(a)d\,(u(a))\,, \end{aligned}$$

subject to

$$\begin{aligned} &\dot{y}_1 = -\,(1-u)\,e'\,(m)\,y_1 - Pb'\,(ue\,(m))\,ue'\,(m)\,, \\ &\quad \dot{y}_2 = d\,(u)\,y_2 - b\,(ue\,(m))\,, \\ &\dot{m} = (1-u)\,e\,(m)\,, \quad \dot{P} = -d\,(u)\,P, \\ &\lim_{A\to\infty} y_1(A) = \lim_{A\to\infty} y_2(A) = 0, \quad m(0) = m_0, \quad P(0) = 1. \end{aligned} \tag{24}$$

In other words, if J is maximised by u^*, then

$$u^*(a) = \arg \max_{u(a)\in[0,1]} H\,(m(a), P(a), u(a))\,, \tag{25}$$

at each $a \in [0, \infty)$. This implies that

$$\begin{aligned} &g_H(a) \stackrel{\text{def}}{=} \frac{\partial H(m(a), P(a), u(a))}{\partial u(a)} = P(a)b'\,(u(a)\,e(m(a)))\,e\,(m(a)) \\ &-y_1(a)e\,(m(a)) - y_2(a)d'\,(u(a))\,P(a) = 0 \\ &\quad \text{when} \;\; u^*(a) \in (0,1), \\ &g_h(a) \geq 0 \quad \text{when} \;\; u^*(a) = 1, \\ &g_h(a) \leq 0 \quad \text{when} \;\; u^*(a) = 0. \end{aligned} \tag{26}$$

Obviously, to assure that $u^*(a)$ is a local maximum of J, resp. H, the derivative of $g_H(a)$ with respect to $u(a)$ has to be negative whenever $0 < u(a) < 1$.

The co-states $y_1(a)$ and $y_2(a)$ in (26) can be expressed from (24) as follows:

$$\begin{aligned} y_1(a) &= y_1(0)\exp\left(-\int_0^a (1-u(\alpha))e'(m(\alpha))\mathrm{d}\alpha\right) \\ &\quad - \int_0^a (P(\alpha)\,b'(u(\alpha)\,e(m(\alpha)))u(\alpha)e'(m(\alpha)))\exp\left(-\int_\alpha^a (1-u(\tau))e'(m(\tau))\mathrm{d}\tau\right)\mathrm{d}\,\alpha \\ y_2(a) &= y_2(0)\exp\left(-\int_0^a d(u(\alpha))\mathrm{d}\alpha\right) \\ &\quad - \int_0^a b\,(u(\alpha)\,e\,(m(\alpha)))\,\exp\left(\int_\alpha^a (d(u(\tau)))\mathrm{d}\tau\right)\mathrm{d}\,\alpha, \end{aligned} \tag{27}$$

where $y_1(0)$ and $y_2(0)$ have to be chosen such that $\lim_{A\to\infty} y_1(A) = \lim_{A\to\infty} y_2(A) = 0$. In Appendix 4 we show that:

(i)

$$\frac{y_1(a)}{P(a)} = \int_a^\infty (\hat{P}(\alpha; a)\, b'(u(\alpha)\, e\,(m(\alpha)))u(\alpha)e'(m(\alpha)))\Delta_m(\alpha)\, \mathrm{d}\alpha, \tag{28}$$

which can be interpreted as the marginal loss or gain per unit weight change (sensitivity) of lifetime offspring due to lower subsequent weights, and

(ii)

$$\frac{y_2(a)}{P(a)} = \int_a^\infty (\hat{P}(\alpha; a)\, b(u(\alpha)\, e\,(m(\alpha)))\, \mathrm{d}\alpha, \tag{29}$$

which can be interpreted as the sensitivity of lifetime offspring due to lower subsequent survival.

Moreover, in the same appendix we show that Formulas (12)–(13) for calculating the fitness returns (c.q. the selection gradient) and Formulas (26)–(27) for the derivative of the Hamiltonian with respect to u, are equivalent.

The detailed match between the results from the two approaches more generally follows from the correspondence between Bellman's principle of optimality and Pontryagin's maximum principle that we work out in some detail in Appendix 5.

On the practical side we point at the fact that even when one is only interested in calculating an ESS with the help of Pontryagin's maximum principle, and has no particular interest in the evolutionary trajectories by which this ESS may be reached, running some discretised variant of the canonical equation can still provide an effective computational implementation of that principle as used in ESS calculations.

7 Discussion

The main contribution of this note is that we carefully set up the CE for life history decisions. As it turned out, a few details had to be added to the exposition in Parvinen et al. (2013). In particular, it was necessary to extend the canonical equation so as to be able to handle inequality constraints. In addition, there was the small detail of the appearance of an additional multiplicative factor accounting for the difference in the initial branching process that mutants have to get through before getting established compared to the linear birth and death process that appears in this role for ODE population models (c.f. Durinx et al. 2008).

Given the venerable history of Pontryagin's maximum principle and its applications to life history theory it should raise no wonder that interpreting the co-states is not new. In particular, Jesus Alberto Leon already did so in the nineteen-seventies (Leon 1976); see also Perrin and Sibly (1993). However, in those days there was no canonical equation around and hence no need to make a connection. Moreover, these early authors

put forward the interpretation seemingly ex cathedra, and only post hoc and summarily related it to a marginal value argument, without exhibiting the explicit connection made in our Sects. 4 and 6 and Appendices 4 and 5. In particular, they did not consider "co-state variables" for other u than the optimal one. Precisely these "generalised co-state variables" occur as ingredients of the selection gradient. Although such variables are already used in numerical approaches to Pontryagin's maximum principle (e.g. Näslund et al. 1974), we believe that our explicit calculations add to the biological understanding of the mathematical structure of eco-evolutionary models.

As a final point we note that the argument that we provide in Sect. 5, although this was not spelled out there, is exemplary of a more general principle. When we delve a little more deeply into the stochastic models for individual behavior, as was necessary in order to calculate σ^2, it generally becomes clear how embarrassingly oversimplified such models tend to be. In our case it turned out that it was implicitly assumed that microscopically the production of young is coupled far more loosely to the energy flow to reproduction than seemingly is assumed at the macroscopic level. Real organisms first have to accumulate the necessary energy that then is transformed into the birth of a young, instead of randomly producing young on the basis of the instantaneous availability of resources. Therefore in reality the production of young usually is far more regular than Poisson (so that σ^2 is close to σ_λ^2), and at a given time depends also on past energy availabilities. Hence the idea that the average rate of offspring production at age a is just a function b of $u(a)\, e(m(a))$ is at best only a rough approximation. One possible justification is that most of the time $u\, e(m)$ varies only slowly compared to the rate at which young are produced, and that if reproduction does occur spread out in time, no two individuals will be in the same phase of their reproduction cycle, so that at any one time the effective offspring production of the individuals that have a size close to the scalar $\overline{m}$ may well be on average close to $b(ue(\overline{m}))$. However, the modelling community is still a long way from proving any rigorous approximation theorems of this ilk. (See Heijmans and Metz (1989) for another possible justification, which, however, is less often applicable in a general life history context.) Of course we also made other simplifying assumptions, like neglecting basal metabolism. However, these simplifications were only put in to ease the exposition, raise no deep mathematical issue, and hence can presumably be relaxed without great difficulty.

Acknowledgments We thank Kalle Parvinen for a helpful discussion about the topics of this paper. This work benefitted from the support from the "Chaire Modélisation Mathématique et Biodiversité of Veolia Environnement—Ecole Polytechnique—Museum National d'Histoire Naturelle—Fondation X".

Appendix 1: Mendelian organisms

Most life history models implicitly assume clonal reproduction. Yet, by far the majority of organisms that are supposedly targeted by these modeling efforts are Mendelian

diploids (c.f. Stearns 1976, 1977). To help overcoming this awkward discrepancy we summarise here some results for the Mendelian case (c.f. Metz and de Kovel 2013).

The first difference between the clonal and Mendelian cases is that the homozygote phenotype present after a substitution differs from the heterozygote phenotype that invaded. Since for small mutational steps the genotype to phenotype map is approximately additive, this leads to the appearance of an additional factor two (on the assumption that there are no parental effects) in the right hand side of the canonical equation.

A more fundamental difference is that as a rule the gametes involved in sexual reproduction come in two types, macro- and micro-gametes. To keep the discussion simple we concentrate on the case where the sexes are separate, for otherwise we have to consider triple allocation targets, to growth, macro-gametes, and micro-gametes. In the case of separate sexes we simply have $u = (u_\mathrm{f}, u_\mathrm{m})$, with u_f the allocation rule of the females, and u_m the one of the males. These allocation rules in general will be evolutionarily coupled through mutational co-variances, but, except for a common time scaling with $T_\mathrm{r}^{-1}(u)$, the selection gradients can be treated separately, as if we were dealing with two coevolving species, with each of the sexes setting part of the environment, which now also includes fertilization opportunities, for the other sex. This independence derives from the additive relation $R_0 = \frac{1}{2}(R_\mathrm{f} + R_\mathrm{m})$, with R_f the average lifetime number of kids of a female and R_m the average lifetime number of kids of a male (e.g. Metz and Leimar 2011; Gyllenberg et al. 2011). Similarly, $T_\mathrm{r} = \frac{1}{2}(T_\mathrm{r,f} + T_\mathrm{r,m})$, where the additional indices f and m mean that the so indexed quantity, in this case the average age of the parent at the birth of its kids, is calculated conditional on the sampled individual being a female or a male. Hence, for $\mathrm{S} \in \{\mathrm{f}, \mathrm{m}\}$,

$$G_\mathrm{S}(u) = \frac{\partial F}{\partial u_\mathrm{S,mut}}(u; E_\mathrm{attr}(u)) = \frac{1}{T_\mathrm{r,f} + T_\mathrm{r,m}} \frac{\mathrm{d}R_\mathrm{S}}{\mathrm{d}u_\mathrm{S,mut}}(u; E_\mathrm{attr}(u)) . \tag{30}$$

The action of the derivative can again be expressed as an integral

$$\frac{\mathrm{d}R_\mathrm{S}(u; E_\mathrm{attr}(u))}{\mathrm{d}u_\mathrm{S,mut}} x = \int_0^\infty g_\mathrm{S}(\alpha; u)\, x(\alpha)\, \mathrm{d}\alpha, \tag{31}$$

with the functions g_S calculated in the same manner as for the clonal model, with the hidden argument E_attr in the functions b_S accounting for any differences in availability of fertilization opportunities at different u_res. Finally, $T_\mathrm{s} = q_\mathrm{f}\, T_\mathrm{s,f} + q_\mathrm{m}\, T_\mathrm{s,m}$, with q_f and q_m the relative frequencies with which the sexes are born into the resident population, and $\sigma^2 = \frac{1}{4}(q_\mathrm{f}\, \sigma_\mathrm{f}^2 + q_\mathrm{m}\sigma_\mathrm{m}^2 + q_\mathrm{f}^{-1} + q_\mathrm{m}^{-1} - 2)$ (the latter formula also takes into account the random sampling of alleles during the offspring production by the heterozygotes).

The upshot is that the males and females in any ESS-es satisfy separate Pontryagin maximum principles, with the coupling between the sexes appearing in the equations only through the influences the resident female and male strategies exert on E_attr. Yet, the fact that the fertilization opportunities come as a component of E_attr inextricably entwines life history evolution with sex ratio evolution.

Appendix 2: How to deal with local constraints?

This appendix emulates the treatment in Dieckmann et al. (2006) with the difference that we go one step further in working out the result, so that we end up with a simple formula that provides the right match for the Pontryagin maximum principle.

For a start we note that in principle the mutational covariance function is not constant over evolutionary time, but depends on the evolutionary history of the population. In particular, the distribution of mutational steps has to change near a constraint boundary so as to preclude overstepping it. There are various ways in which this change may happen. Most of these will make the distribution of the steps asymmetric, with close to the boundary steps towards the interior of the space of feasible strategies becoming more common relative to steps towards the boundary. The CE as given by Dieckmann et al. (2006) and Parvinen et al. (2013) is based on the assumption that the mutation distribution is symmetric, in line with most papers on the CE; (Formulas for the non-symmetric case may be found in Dieckmann and Law (1996), Champagnat (2003), and Champagnat and Méléard (2011).) In our formula we have kept the form of the CE unchanged in the interior of the constraint set and only set the right hand side equal to zero where that movement would lead to the passing of a constraint boundary. The rationale for this ploy is the following. The CE is derived as a limit in which one lets a factor that scales the mutational steps go to zero. This means that at any distance from the constraint boundary eventually the effect of the constraint will no longer be felt, and if the mutation distribution would otherwise be symmetric, this symmetry would eventually be recovered for all resident strategies that are not located on the boundary. At boundary strategies, in the CE limit the movement component in the outward direction has to drop to zero, since there the mutation distribution stays forever asymmetric, with its probability mass all located on the feasible side. In the limit the distribution of this mass contracts towards the boundary. On segments of the boundary where the nearby movement is towards that boundary the movement on the boundary becomes restricted to it by the covariance function abruptly becoming singular. On the natural assumption that the constraint does not affect movement parallel to the boundary, this corresponds to just setting the right hand side to zero at the indicated values of a. (In finite dimensional trait spaces the analogous condition is that on the boundary the movement component orthogonal to the boundary becomes zero whenever close by the movement is towards the boundary, while the movement component parallel to the boundary is a continuous extension of the movement component in that direction in the interior of the constraint set.)

Appendix 3: The canonical equation for periodic ODE population models

The right hand side of the CE equals

[Rate at which mutants are produced] $\times$ average of [the effect of a mutation times a linear approximation for the probability that mutant invades]. (32)

On the assumption of small mutational steps and a symmetric mutation distribution the latter average gives $\frac{1}{2}$ times the mutational covariance operator applied to the selection gradient, where the $\frac{1}{2}$ comes from the fact that the linear approximation only applies in the half space where the invasion fitness is positive and is replaced by 0 where it is negative. The factor σ^{-2} in (4) comes from the lowest order term of the asymptotic expansion for the probability Q that a mutant with a slightly positive fitness ($0<F\ll 1$) invades. When births occur singly the individual-based models underlying ODE population models can for the initial phases of mutant invasion be approximated by a linear birth and death process. For constant environments the corresponding generation process is of Galton–Watson type with a geometric offspring distribution with mean $R_0 = \frac{b}{d}$ with b and d the per capita birth and death rates of the mutant, respectively, while $F = b - d = \frac{R_0-1}{T_s}$. Hence the invasion probability equals $Q = 1 - R_0^{-1} = R_0 - 1 + \mathrm{O}((R_0-1)^2)$. More in general, $Q = 2\,\sigma_e^{-2}\,\ln\,R_0 + \mathrm{O}(\ln^2\,R_0)$, with σ_e^2 a measure for the average variability of the offspring production of the residents (for which $R_0 = 1$), which in the case of a single birth state reduces to the variance of the offspring distribution σ^2 (c.f. Metz and de Kovel 2013; Durinx et al. 2008). (For a geometric distribution with mean 1: $\sigma^2 = 2$.) The rate at which mutants are produced equals the population birth rate times the per birth probability of a mutation. The factor $\overline{n}$ in (4) appears by re-expressing the population birth rate B of the resident as $\frac{\overline{n}}{T_s}$, based on the general consistency relation $\overline{n} = B\,T_s$. Below we consider the extension of these considerations to periodic environments; the further extension to general ergodic environments is treated in Ripa and Dieckmann (2013).

In the case of periodic environments we have to average both the number of births as well as the probability to invade over a cycle, where the first average is a time average and the latter average is over the distribution of births over the cycle.

To calculate the invasion probability in dependence of the phase θ of appearance of a mutant during the environmental cycle, $q(\theta)$, we use the general formula for the invasion probability for linear birth and death processes with time variable parameters derived in Kendall (1948):

$$q(\theta) = \frac{1}{1 + \int_\theta^\infty \mathrm{e}^{-r(\alpha;\theta)}\,d(\alpha)\,\mathrm{d}\alpha} \quad \text{with} \quad r(t;\theta) = \int_\theta^t (b(\tau) - d(\tau))\,\mathrm{d}\tau. \tag{33}$$

With time rescaled so that the period equals $T = 1$, we then get

$$Q = \int_0^1 q(\theta)\,w(\theta)\,\mathrm{d}\theta, \tag{34}$$

with

$$w(\theta) = \frac{b_0(\theta)\mathrm{e}^{r_0(\theta;0)}}{\int_0^1 b_0(\tau)\,\mathrm{e}^{r_0(\theta;0)}\mathrm{d}\tau} \tag{35}$$

the probability distribution of the phase of the environmental cycle at which a mutant may be expected to appear, with b_0 and d_0 the periodic per capita birth and death rates of the residents and r_0 defined as in (33).

The stationarity of the resident population implies that $r_0(t+1;t) = 0$, i.e., $\int_t^{t+1} b_0(\tau)\,\mathrm{d}\tau = \int_t^{t+1} d_0(\tau)\,\mathrm{d}\tau$ (no per capita population growth over a full environmental cycle) as well as $\int_t^{t+1} b_0(\tau)\,\mathrm{e}^{r_0(\tau;t)}\mathrm{d}\tau = \int_t^{t+1} d_0(\tau)\,\mathrm{e}^{r_0(\tau;t)}\,\mathrm{d}\tau$ (the total births over a cycle matches the death toll over the cycle). More in general $F = r(t+1;t) = r(1;0)$ and

$$R_0 = \frac{\overline{b}}{\overline{d}}, \tag{36}$$

with

$$\overline{b} \stackrel{\text{def}}{=} \int_0^1 b(\tau)\,\mathrm{d}\tau = \int_t^{t+1} b(\tau)\,\mathrm{d}\tau \quad \text{and} \quad \overline{d} \stackrel{\text{def}}{=} \int_0^1 d(\tau)\mathrm{d}\tau = \int_t^{t+1} d(\tau)\,\mathrm{d}\tau \tag{37}$$

(Bacaer and Guernaoui 2006), where in the periodic case R_0 is defined as the dominant eigenvalue of the operator that gives the average number of newborns born at different phases of the cycle for mothers born at different phases. To calculate the derivative of Q we introduce a scalar variable x by which we parametrise a curve in the space of strategies passing transversally through the resident value at $x = 0$, and write all the coefficient functions as functions of x, written as an index in the case of b, d, and r. As later on we also need the invasion probability and invasion fitness as a function of any mutant strategy, we will denote the maps from x to these two quantities as $\tilde{Q}$ and $\tilde{F}$. With

$$M(x) \stackrel{\text{def}}{=} \int_0^\infty \mathrm{e}^{-r_x(\alpha;0)}\, d_x(\alpha)\,\mathrm{d}\alpha \tag{38}$$

we can write

$$q(\theta;x) = \frac{1}{1 + \int_\theta^1 \mathrm{e}^{-r_x(\alpha;\theta)}\, d_x(\alpha)\,\mathrm{d}\alpha + \mathrm{e}^{-r_x(1;\theta)}\, M(x)}. \tag{39}$$

From $\lim_{x\to 0} q(\theta;x) = 0$ it follows that $\lim\limits_{x\to 0} M(x) = \infty$. Hence

$$\frac{\partial\, q}{\partial\, x}(\theta;0) = -\mathrm{e}^{r_0(1;\theta)} \lim_{x\to 0} \frac{M'(x)}{M^2(x)} \tag{40}$$

and

$$\tilde{Q}'(0) = -\frac{\int_0^1 b_0(\theta)\,\mathrm{d}\theta}{\int_0^1 b_0(\theta)\,\mathrm{e}^{r_0(\theta;0)}\,\mathrm{d}\theta} \lim_{x\to 0} \frac{M'(x)}{M^2(x)}. \tag{41}$$

To calculate the term after the limit sign we observe that

$$M(x) \stackrel{\text{def}}{=} \int_0^1 \mathrm{e}^{-r_{\mathrm{x}}(\theta;0)} d_{\mathrm{x}}(\theta)\,\mathrm{d}\theta + \mathrm{e}^{-r_{\mathrm{x}}(1;0)} M(x). \tag{42}$$

Hence

$$M(x) = \frac{\int_0^1 \mathrm{e}^{-r_{\mathrm{x}}(\theta;0)}\, d_{\mathrm{x}}(\theta)\,\mathrm{d}\theta}{1 - \mathrm{e}^{-r_{\mathrm{x}}(1;0)}}, \tag{43}$$

$$\lim_{x\to 0} \frac{M'(x)}{M^2(x)} = \frac{-\tilde{F}'(0)}{\int_0^1 \mathrm{e}^{-r_0(\theta;0)}\, d_0(\theta)\,\mathrm{d}\theta} \tag{44}$$

and

$$\tilde{Q}'(0) = \frac{\int_0^1 b_0(\theta)\,\mathrm{d}\theta}{\int_0^1 b_0(\theta)\,\mathrm{e}^{r_0(\theta;0)}\,\mathrm{d}\theta \int_0^1 \mathrm{e}^{-r_0(\theta;0)} d_0(\theta)\,\mathrm{d}\theta}\tilde{F}'(0). \tag{45}$$

Hence away from local constraints the CE becomes

$$\begin{aligned}\frac{\mathrm{d}s}{\mathrm{d}t} &= \frac{1}{2}\mu \int_0^1 b_0(\alpha)\tilde{n}(\alpha)\,\mathrm{d}\alpha \frac{\int_0^1 b_0(\alpha)\,\mathrm{d}\alpha}{\int_0^1 b_0(\alpha)\,\mathrm{e}^{r_0(\alpha;0)}\,\mathrm{d}\alpha \int_0^1 \mathrm{e}^{-r_0(\alpha;0)}\, d_0(\alpha)\,\mathrm{d}\alpha} \\ &\quad \times \int_0^1 c(\theta,\alpha)\, f(\alpha;s)\,\mathrm{d}\alpha \\ &= \frac{1}{2}\mu \frac{\int_0^1 d_0(\alpha)\,\mathrm{d}\alpha}{\int_0^1 \tilde{n}^{-1}(\alpha) d_0(\alpha)\mathrm{d}\alpha} \int_0^1 c(\theta,\alpha)\, f(\alpha;s)\,\mathrm{d}\alpha \end{aligned} \tag{46}$$

where s now denotes the strategy, which in the seasonal flowering model of Dieckmann et al. (2006) consists of a flowering intensity as a function of θ, and

$$f(\theta;s) = \frac{\mathrm{d}\,(b(\theta) - d(\theta))}{\mathrm{d}\,s_{\mathrm{mut}}(\theta)}\,(s;\, E_{\mathrm{attr}}(s))\,. \tag{47}$$

Hence the $\overline{n}$ in Parvinen et al. (2013) has to be interpreted as

$$\frac{\int_0^1 d_0(\alpha)\,\mathrm{d}\alpha}{\int_0^1 \tilde{n}^{-1}(\alpha)\, d_0(\alpha)\,\mathrm{d}\alpha}. \tag{48}$$

To see how (46) compares with (4) we first observe that for periodically fluctuating populations there is no immediate counterpart for the equality $\overline{n} = B\,T_{\mathrm{s}}$, so we substitute

the latter in (4), while observing that the counterpart of B in (46) is $\int_0^1 b_0(\alpha)\tilde{n}(\alpha)\,\mathrm{d}\alpha$. After substituting $f(\alpha; s) = \frac{g(\alpha;s)}{T_{\mathrm{r}}(s)}$ in (46) we then end up with the pairing

$$\frac{2\,T_{\mathrm{r}}}{\sigma_e^2} = \frac{\int_0^1 b_0(\alpha)\,\mathrm{d}\alpha}{\int_0^1 b_0(\alpha)\,\mathrm{e}^{r_0(\alpha;0)}\,\mathrm{d}\alpha\,\int_0^1 \mathrm{e}^{-r_0(\alpha;0)}\,d_0(\alpha)\,\mathrm{d}\alpha}. \tag{49}$$

To calculate T_{r} we use $F = \overline{b} - \overline{d}$ and $R_0 = \overline{b}/\overline{d}$ together with (1) to find

$$T_{\mathrm{r}} = \overline{d}_0^{-1} = \overline{b}_0^{-1}. \tag{50}$$

Therefore,

$$\begin{aligned}\sigma_e^2 &= \frac{2\int_0^1 b_0(\alpha)\,\mathrm{e}^{r_0(\alpha;0)}\,\mathrm{d}\alpha}{\int_0^1 b_0(\alpha)\,\mathrm{d}\alpha}\cdot\frac{\int_0^1 \mathrm{e}^{-r_0(\alpha;0)}d_0(\alpha)\,\mathrm{d}\alpha}{\int_0^1 d_0(\alpha)\,\mathrm{d}\alpha}\\ &= \frac{2\,\int_0^1 \mathrm{e}^{r_0(\alpha;0)}\,d_0(\alpha)\,\mathrm{d}\alpha\int_0^1 \mathrm{e}^{-r_0(\alpha;0)}\,d_0(\alpha)\,\mathrm{d}\alpha}{\left(\int_0^1 d_0(\alpha)\,\mathrm{d}\alpha\right)^2}.\end{aligned} \tag{51}$$

Appendix 4: Relating the results of Sects. 4 and 6

In this appendix we show that formulas (12)–(13) for calculating the fitness returns (c.q. the selection gradient) and formulas (27)–(26) for the derivative of the Hamiltonian with respect to u are equivalent. To enhance the similarity divide g_H by P and set $\tilde{y}_1(a) \stackrel{\text{def}}{=} P^{-1}(a)\,y_1(a)$, to get $\tilde{y}_2(a) = y_2(a)$ (for each $a \in [0,\infty)$), obtaining

$$r_H(a) \stackrel{\text{def}}{=} \frac{g_H(a)}{P(a)} = b'\left(u(a)\,e(m(a))\right)e(m(a) - \tilde{y}_1(a)\,e(m(a)) - \tilde{y}_2(a)d'(u(a))$$

with

$$\begin{aligned}\dot{\tilde{y}}_1 &= -(1-u)e'(m)\tilde{y}_1 - \hat{P}\,b'(u\,e(m))\,u\,e'(m), \qquad \lim_{A\to\infty}\tilde{y}_1(A) = 0,\\ \dot{\tilde{y}}_2 &= d(u)\tilde{y}_2 - b(u\,e(m)), \qquad \lim_{A\to\infty}\tilde{y}_2(A) = 0,\end{aligned} \tag{52}$$

which is to be compared with

$$\begin{aligned}r(a) &= b'\left(u(a)e(m(a))\right)e(m(a))\\ &\quad - e\left(m(a)\right)\int_a^\infty \hat{P}(\alpha)\,b'\left(u(\alpha)e(m(\alpha))\right)u(\alpha)e'(m(\alpha))\hat{\Delta}_m(\alpha)\,\mathrm{d}\alpha\\ &\quad - d'(u(a))\int_0^\infty \hat{P}(\alpha)b\left(u(\alpha)\,e(m(\alpha))\right)\end{aligned} \tag{53}$$

with

$$\hat{y}_1 \stackrel{\text{def}}{=} \int_a^\infty \hat{P}(\alpha)\, b'\,(u(\alpha)e(m(\alpha)))\; u\, e'(m(\alpha))\hat{\Delta}_m(\alpha)\,\mathrm{d}\alpha, \tag{54}$$

$$\hat{y}_2 \stackrel{\text{def}}{=} \int_a^\infty \hat{P}(\alpha)\, b\,(u(\alpha)\, e(m(\alpha)))\;\mathrm{d}\alpha, \tag{55}$$

where we now dropped the second argument of $\hat{P}$ and $\hat{\Delta}_m$ on the understanding that $\hat{P}(a) = \hat{\Delta}_m(a) = 1$. The structure of the above expressions is

- for fitness returns:

$$\hat{y}(a) = \int_a^\infty \phi(\alpha) z(\alpha)\,\mathrm{d}\alpha, \quad \dot{x} = -\psi\, z, \quad z(a) = 1 \tag{56}$$

- for Pontryagin:

$$\dot{\tilde{y}} = \psi\,\tilde{y} - \phi, \quad \lim_{A\to\infty} \tilde{y}(A) = 0. \tag{57}$$

Expanding the integrals gives

$$\hat{y}(a) = \int_a^\infty \phi(\alpha)\exp\left(-\int_a^\alpha \psi(\tau)\,\mathrm{d}\tau\right)\mathrm{d}\alpha, \tag{58}$$

$$\tilde{y}(a) = -\int_0^a \phi(\alpha)\exp\left(\int_\alpha^a \psi(\tau)\mathrm{d}\tau\right)\,\mathrm{d}\alpha + \tilde{y}(0)\exp\left(\int_0^a \psi(\tau)\,\mathrm{d}\tau\right) \tag{59}$$

with

$$\begin{aligned}\tilde{y}(0) &= \exp\left(-\int_0^\infty \psi(\tau)\mathrm{d}\tau\right)\int_0^\infty \phi(\alpha)\exp\left(\int_\alpha^\infty \psi(\tau)\mathrm{d}\tau\right)\mathrm{d}\,\alpha \\ &= \int_0^\infty \phi(\alpha)\exp\left(\int_0^\alpha \psi(\tau)\,\mathrm{d}\tau\right)\,\mathrm{d}\,\alpha.\end{aligned} \tag{60}$$

Hence,

$$\tilde{y}(a) = -\int_0^a \phi(\alpha)\exp\left(\int_\alpha^a \psi(\tau)\,\mathrm{d}\tau\right)\mathrm{d}\alpha + \int_0^\infty \phi(\alpha)\exp\left(\int_\alpha^a \psi(\tau)\,\mathrm{d}\tau\right)\mathrm{d}\alpha$$
$$= \int_a^\infty \phi(\alpha)\exp\left(-\int_a^\alpha \psi(\tau)\,\mathrm{d}\tau\right)\mathrm{d}\alpha. \tag{61}$$

Therefore, indeed $\tilde{y}_1 = \hat{y}_1$ and $\tilde{y}_2 = \hat{y}_2$.

Appendix 5: Basic concepts of dynamic optimal control theory

One of the goals in this paper was to elucidate in an accessible manner the connection between the fitness returns of evolutionary ecology and Pontryagin's maximum principle. This appendix considers the more general problem of linking the more intuitive, better known, and more general optimality principle of Bellman with Pontryagin's maximum principle. To this end we derive the latter from the former. While both principles are commonly formulated for optimal strategies, we have reengineered the argument so that in the initial steps we consider just any fixed strategy in order to put in the limelight the direct conceptual link between fitness returns and the local variant of Pontryagin's maximum principle.

Appendix 5.1: The optimal control problem

Consider a dynamic system described by a state equation

$$\dot{x}(t) = f\left(x(t), u(t)\right), \quad x(0) = x_0, \tag{62}$$

where $x(t) \in \mathbb{R}^n$ is a state variable, $u(t) \in \mathbb{R}^m$ is a control variable, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is assumed to be continuously differentiable.

The control aim is to maximise the objective function:

$$J = \int_0^T \Pi\left(x(t), u(t)\right)\,\mathrm{d}t + S\left(x(T)\right), \tag{63}$$

where Π and S are assumed to be continuously differentiable.

The path $x(t)$, $t \in [0, T]$, is called a state trajectory and $u(t)$, $t \in [0, T]$, is called a control (decision, action).

Usually, the control variable $u(t)$ is assumed to be piecewise continuous and constrained as follows:

$$u(t) \in \Omega(t) \subset \mathbb{R}^m, \quad t \in [0, T]. \tag{64}$$

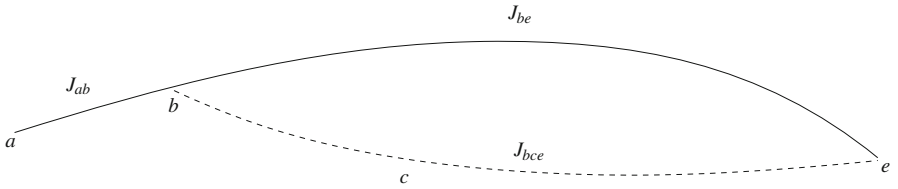

Fig. 2 Illustration of Bellman's optimality principle

If $u(t)$ satisfies Condition (64) for each $t \in [0, T]$, we call u an admissible control.

The optimal control problem is to find an admissible control u^* which maximises the objective function (63) subject to Constraint (62). Such a control u^* is called an optimal control, the corresponding state trajectory is denoted by x^* and is called the optimal trajectory under u^*, $J(u^*)$ or J^* then denotes the optimal value of J.

Appendix 5.2: Bellman's optimality principle

Bellman's optimality principle which has to be satisfied for an optimal control u^* (Bellman 1957) reads: "an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

Observing Fig. 2, we can state Bellman's optimality principle as a proposition:

Proposition *If a b e is the optimal path from a to e, then b e is the optimal path from b to e.*

Proof Suppose it is not. Then there is another path (note that existence is assumed here) $b\,c\,e$ which is optimal from b to e, i.e. $J_{bce} > J_{be}$. But then $J_{abe} = J_{ab} + J_{be} < J_{ab} + J_{bce} = J_{abce}$. This contradicts the hypothesis that $a\,b\,e$ is the optimal path from a to e. □

Appendix 5.3: The value function for any fixed u

We can formulate Bellman's optimality principle using the so-called value function $V : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ for a given (but not necessarily optimal) control $u \in \Omega$. This value function is defined as

$$V(x_t, t; u) \stackrel{\text{def}}{=} \int_t^T \Pi\left(x(s), u(s)\right) \mathrm{d}s + S\left(x(T)\right), \tag{65}$$

where for $s \geq t$

$$\frac{\mathrm{d}x}{\mathrm{d}s} = f\left(x, u\right), \quad x(t) = x_t.$$

In the original work of Bellman (1957) the value function was defined for an optimal strategy u, as the main focus of his work was finding this optimal strategy. We, however, formulate the value function for any u and introduce the Bellman's optimality principle with u being optimal at the further step of this derivation.

In Fig. 3 we illustrate the value function in the (x, t)-space for a given u.

Note that incremental changes in J from t to $t + \Delta_t$ are given by the integral of $\Pi(x, u)$ from t to $t + \Delta_t$. Considering that the change in the objective function consists of the incremental changes in J from t to $t + \Delta_t$ plus the value function $V(x + \Delta_x, t + \Delta_t; u)$ at time $t + \Delta_t$ (for fixed u) we can write

$$V(x_t, t; u) = \int_t^{t+\Delta_t} \Pi\left(x(\tau), u(\tau)\right) \mathrm{d}\tau + V(x(t + \Delta_t), t + \Delta_t; u). \tag{66}$$

Since Π is continuous in t, the integral in (66) can be approximated by $\Pi(x_t, u(t))\, \Delta_t$, so that we can rewrite (66) as

$$V(x_t, t; u) = \Pi(x_t, u(t))\, \Delta_t + V\left(x(t + \Delta_t), t + \Delta_t; u\right) + o(\Delta_t). \tag{67}$$

If V is continuously differentiable, we can use the Taylor expansion of V with respect to Δ_t to obtain

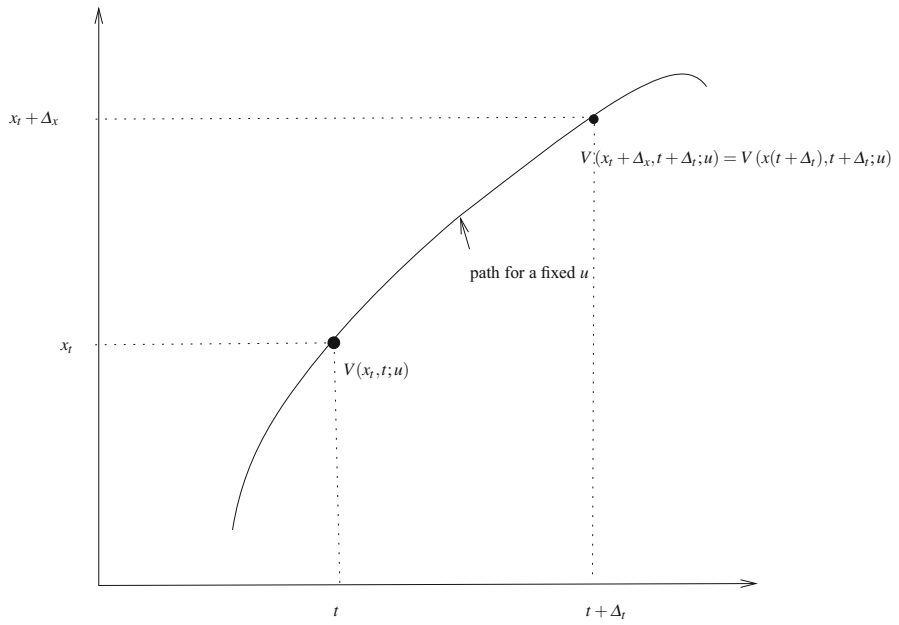

Fig. 3 A path for a specific choice of u in the time-state space

$$V(x(t+\Delta_t), t+\Delta_t; u) = V(x_t, t; u) + \left(\frac{\partial V(x_t, t; u)}{\partial x} \dot{x} + \frac{\partial V(x_t, t; u)}{\partial t} \right) \Delta_t + o(\Delta_t). \tag{68}$$

Substituting x from (62) we obtain

$$V(x_t, t; u) = \Pi(x_t, u(t))\Delta_t + V(x_t, t; u) + \frac{\partial V(x_t, t; u)}{\partial x} f(x_t, u(t))\Delta_t + \frac{\partial V(x_t, t; u)}{\partial t} \Delta_t + o(\Delta_t). \tag{69}$$

By canceling $V(x_t, t; u)$ on both sides and then dividing by Δ_t we get

$$0 = \Pi(x_t, u(t)) + \frac{\partial V(x_t, t; u)}{\partial x} f(x_t, u(t)) + \frac{\partial V(x_t, t; u)}{\partial t} + \frac{o(\Delta_t)}{\Delta_t}. \tag{70}$$

By letting $\Delta_t \downarrow 0$, we obtain

$$0 = \Pi(x_t, u(t)) + \frac{\partial V(x_t, t; u)}{\partial x} f(x_t, u(t)) + \frac{\partial V(x_t, t; u)}{\partial t} \tag{71}$$

with the boundary condition $V(x(T), T; u) = S(x(T))$. Here $\frac{\partial V(x_t, t; u)}{\partial x}$ can be interpreted as the marginal contribution of the state variable x to the objective function for a fixed u. We denote it by $y(t) \in \mathbb{R}^n$ and call it the costate vector (also known as adjoint or auxiliary vector in optimisation and control theory and as a shadow price in economics), thus

$$y(t) \stackrel{\text{def}}{=} \frac{\partial V(x_t, t; u)}{\partial x}, \tag{72}$$

and introduce the so-called Hamiltonian[4]

$$H(x(t), u(t), y(t)) \stackrel{\text{def}}{=} \Pi(x(t), u(t)) + (y(t))^T f(x(t), u(t)). \tag{73}$$

Note that the previous definition for our example identifies y with the marginal fitness return per unit of change in the state variables at age a (here time t). Below we shall focus on optimal controls and trajectories, and in this way derive the equations for y customarily encountered in the literature on Pontryagins maximum principle.

[4] Here we use the notation $H(x(t), u(t), y(t))$ instead of the notation $H(x, u, y, t)$ standard in the optimal control theory in order to stress that the Hamiltonian at time t depends on x, u, and y at the same time t. The former notation is more in sink with the usual mathematical attitude that one should rigorously distinguish between a function and its values for particular values of its argument.

Appendix 5.4: From Bellman's optimality principle to Pontryagin's maximum principle

If u is optimal, i.e., if $u = u^*$, the following condition has to hold:

$$0 = \max_{u(t)\in\Omega(t)} \left\{ H\left(x_t, u(t), \frac{\partial V(x_t, t; u)}{\partial x}\right) + \frac{\partial V(x_t, t; u)}{\partial t} \right\}. \tag{74}$$

Equation (74) is called the Hamilton–Jacobi–Bellman (HJB) equation (Bellman 1957) and gives necessary and sufficient conditions that an optimal u has to satisfy, if the value function is differentiable. From Eq. (74) we can get the Hamiltonian maximizing condition, or maximum principle, denoting the costate corresponding to u^* as y^* as

$$H(x^*(t), u^*(t), y^*(t)) + \frac{\partial V(x^*(t), t; u^*)}{\partial t} \geq H(x^*(t), \psi, y^*(t)) + \frac{\partial V(x^*(t), t; u^*)}{\partial t}, \tag{75}$$

for all $\psi \in \Omega(t)$, which leads to inequality

$$H(x^*(t), u^*(t), y^*(t)) \geq H(x^*(t), \psi, y^*(t)). \tag{76}$$

Appendix 5.5: Derivation of the costate (adjoint) equation

We in this narrowed context return to the after effect of infinitesimal changes in the state variables. Let

$$x(t) = x^*(t) + \Delta_x(t) \tag{77}$$

where $\Delta_x(t)$, $\|\Delta_x(t)\| \downarrow 0$. As x^* is the optimal state trajectory, necessarily

$$\begin{aligned} &H\left(x^*(t), u^*(t), \frac{\partial V(x^*(t), t; u^*)}{\partial x}\right) + \frac{\partial V(x^*(t), t; u^*)}{\partial t} \\ &\geq H\left(x(t), u^*(t), \frac{\partial V(x(t), t; u^*)}{\partial x}\right) + \frac{\partial V(x(t), t; u^*)}{\partial t}. \end{aligned} \tag{78}$$

The left hand side of Inequality (78) equals zero, since u^* is the optimal control. In general, the right hand side is less or equal to zero and would be zero for $x = x^*$. Since $x(t)$ is unconstrained, the partial derivative of the right hand side of Inequality (78) with respect to x has to be equal to zero for $x = x^*$, i.e.,

$$\left[\frac{\partial H\left(x_t, u^*(t), \frac{\partial V(x_t, t; u^*)}{\partial x}, t\right)}{\partial x} + \frac{\partial^2 V(x_t, t)}{\partial t\, \partial x} \right]_{x=x^*} = 0. \tag{79}$$

(Note the implicit assumption that V is twice continuously differentiable.) By definition of the Hamiltonian, at $u = u^*$ and $x = x^*$,

$$\frac{\partial \Pi}{\partial x} + \frac{\partial V}{\partial x}\frac{\partial f}{\partial x} + f^T \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial t \partial x} = \frac{\partial \Pi}{\partial x} + \frac{\partial V}{\partial x}\frac{\partial f}{\partial x} + \left(\frac{\partial^2 V}{\partial x^2} f\right)^T + \frac{\partial^2 V}{\partial t \partial x} = 0. \tag{80}$$

By definition (72) of $y(t)$

$$\begin{aligned} \frac{\mathrm{d}\left(\frac{\partial V}{\partial x}\right)}{\mathrm{d}t} &= \left(\frac{\mathrm{d}\left(\frac{\partial V}{\partial x_1}\right)}{\mathrm{d}t}, \ldots \frac{\mathrm{d}\left(\frac{\partial V}{\partial x_n}\right)}{\mathrm{d}t}\right) \\ &= \left(\frac{\partial^2 V}{\partial x^2}\dot{x}\right)^T + \frac{\partial^2 V}{\partial x \partial t} = \left(\frac{\partial^2 V}{\partial x^2} f\right)^T + \frac{\partial^2 V}{\partial x \partial t}. \end{aligned} \tag{81}$$

From Formulas (80) and (81) we have

$$\frac{\mathrm{d}\left(\frac{\partial V}{\partial x}\right)}{\mathrm{d}t} = -\frac{\partial \Pi}{\partial x} - \frac{\partial V}{\partial x}\frac{\partial f}{\partial x}. \tag{82}$$

Substituting the costate from Formula (72), we obtain

$$\dot{y} = -\frac{\partial \Pi}{\partial x} - y\frac{\partial f}{\partial x}.$$

Substituting the Hamiltonian (73) into this expression, we obtain

$$\dot{y} = -\frac{\partial H}{\partial x}. \tag{83}$$

Terminal boundary (or transversality) conditions are defined as

$$y(T) = \frac{\partial S(x(T))}{\partial x}. \tag{84}$$

Equations (83) and (84) determine the adjoint variables From Eq. (73) we can rewrite the state equation as

$$\dot{x} = f = \frac{\partial H}{\partial y}. \tag{85}$$

Combining Formulas (83), (84), (85), and (62) we get

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial y}, & x(0) = x_0, \\ \dot{y} = -\frac{\partial H}{\partial x}, & y(T) = \frac{\partial S(x(T))}{\partial x}. \end{cases} \tag{86}$$

Equation (86) is called a canonical system of equations or canonical adjoints.

Appendix 5.6: Pontryagin's maximum principle

The necessary conditions for u^* to be an optimal control are:

$$\begin{cases} \dot{x}^* = f(x^*, u^*, t), \quad x^*(0) = x_0, \\ \dot{y} = -\frac{\partial H(x^*, u^*, y)}{\partial x}, \quad y(T) = \frac{\partial S(x^*(T))}{\partial x}, \\ H(x^*(t), u^*(t), y(t)) \geq H(x^*(t), u, y(t)), \end{cases} \tag{87}$$

for all $u \in \Omega(t), t \in [0, T]$, where the adjoint (costate) variables $y(t)$ now correspond to the sensitivities of the value function for the optimal u to a state change at t.

References

Bacaer N, Guernaoui S (2006) The epidemic threshold of vector-borne diseases with seasonality. J Math Biol 53:421–436

Bellman R (1957) Dynamic programming. Princeton University Press, Princeton

Champagnat N (2003) Convergence of adaptive dynamics n-morphic jump processes to the canonical equation and degenerate diffusion approximation. Tech. Rep. 03/7, Pr'épublication de l'Université de Nanterre, Paris X

Champagnat N, Méléard S (2011) Polymorphic evolution sequence and evolutionary branching. Probab Theory Relat Fields 151(1–2):45–94

Dercole F, Rinaldi S (2008) Analysis of evolutionary processes: the adaptive dynamics approach and its applications. Princeton, NJ

Dieckmann U, Heino M, Parvinen K (2006) The adaptive dynamics of function-valued traits. J Theor Biol 241:370–389

Dieckmann U, Law R (1996) The dynamical theory of coevolution: a derivation from stochastic ecological processes. J Math Biol 34:579–612

Durinx M, Metz JAJ, Meszéna G (2008) Adaptive dynamics for physiologically structured models. J Math Biol 56:673–742

Geritz SAH (2005) Resident-invader dynamics and the coexistence of similar strategies. J Math Biol 50:67–82

Geritz SAH, Gyllenberg M, Jacobs FJA, Parvinen K (2002) Invasion dynamics and attractor inheritance. J Math Biol 44:548–560

Gupta A, Metz JAJ, Tran VC (2014) A new proof for the convergence of an individual based model to the trait substitution sequence. Acta Applicanda Mathematicae 121(1):1–27

Gyllenberg M, Metz JAJ, Service R (2011) When do optimisation arguments make evolutionary sense? In: Chalub FACC, Rodrigues JF (eds) The mathematics of Darwin's legacy. Birkhauser, Basel, pp 235–269

Heijmans HJAM, Metz JAJ (1989) Small parameters in structured population models and the Trotter–Kato theorem. SIAM J Math Anal 20:870–885

Intrilligator MD (ed) (1971) Mathematical optimization and economic theory. Prentice Hall, Englewood Cliffs

Kendall DG (1948) On the generalized "birth-and-death" process. Ann Math Stat 19:1–15

Leon JL (1976) Life histories as adaptive strategies. J Theor Biol 60:301–335

Méléard S, Tran VC (2009) Trait substitution sequence process and canonical equation for age-structured populations. J Math Biol 58(6):881–921

Metz JAJ (2008) Fitness. In: Jørgensen SE, Fath BD (eds) Evolutionary ecology, encyclopedia of ecology, vol 2. Elsevier, Oxford, pp 1599–1612

Metz JAJ, Diekmann O (eds) (1986) The dynamics of physiologically structured populations. In: Lecture notes in biomathematics, vol. 68. Springer, Berlin

Metz JAJ, de Kovel CGF (2013) The canonical equation of adaptive dynamics for Mendelian diploids and haplo-diploids. Interface Focus 3

Metz JAJ, Leimar O (2011) A simple fitness proxy for ESS calculations in structured populations with continuous traits, with applications to the evolution of haplo-diploids and genetic dimorphisms. J Biol Dyn 5:163–190

Metz JAJ, Nisbet RM, Geritz SAH (1992) How should we define "fitness" for general ecological scenarios? TREE 7:198–202

Näslund B, Hurst EG, Bensoussan A (1974) Management applications of modern control theory. Elsevier, Amsterdam

Parvinen K, Dieckmann U, Heino M (2006) Function-valued adaptive dynamics and the calculus of variations. J Math Biol 52:1–26

Parvinen K, Heino M, Dieckmann U (2013) Function-valued adaptive dynamics and optimal control theory. J Math Biol 67:509–533

Perrin N, Sibly RM (1993) Dynamic models of energy allocation and investment. Annu Rev Ecol Syst 24:379–410

Pontryagin LS, Boltianski VG, Gamkrelidze RV, Mishchenko EF, Brown DE (1964) The mathematical theory of optimal processes. A Pergamon Press, Oxford. http://opac.inria.fr/record=b1122221

Ripa J, Dieckmann U (2013) Mutant invasions and adaptive dynamics in variable environments. Evolution 67(2):1279–1290

Stearns SC (1976) Life history tactics: a review of the ideas. Q Rev Biol 52(2):3–47

Stearns SC (1977) The evolution of life history traits: a critique of the theory and a review of the data. Annu Rev Ecol Syst 8:145–172