The adaptive dynamics of life histories: From fitness-returns to selection gradients and Pontryagin's maximum principle

Metz, J.A.J. and Johansson, J.

Interim Report   **IR-12-054**

The adaptive dynamics of life histories: From fitness-returns to selection gradients and Pontryagin's maximum principle

Johan A.J. (Hans) Metz (metz@iiasa.ac.at)
Jacob Johansson

Approved by

Ulf Dieckmann
Director, Evolution and Ecology Program

February 2015

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

Johan A. Jacob Metz[1,3,4], Jacob Johansson[2,3]

[1] (= Hans)
  Mathematical Institute & Institute of Biology, Leiden University, NL-2333CA
  Leiden, Netherlands
  email: j.a.j.metz@biology.leidenuniv.nl
[2] Theoretical Population Ecology and Evolution Group, Department of Biology, Lund
  University, SE-22362 Lund, Sweden
  email: Jacob.Johansson@biol.lu.se
[3] Evolution and Ecology Program, International Institute for Applied Systems
  Analysis, A-2361 Laxenburg, Austria
[4] Naturalis Biodiversity Center, NL-2333CR Leiden, Netherlands

Abstract Using a fitness-returns argument we derive an expression for the selection gradient for the age dependent allocation strategy in a common class of state variable based life-history models. By setting the selection gradient equal to zero as part of the calculation of the ESS-es for such models, we get a marginal value argument and through this recover the local version of Pontryagin's maximum principle. This fills in a minor gap in a recent paper by Parvinen, Heino and Dieckmann (2012; DOI 10.1007/s00285-012-0549-2), who treated the calculation of the selection gradient and of the ESS-es as separate issues. As bonuses we (i) provide an extension of the framework of these authors that can handle also the more complicated evolutionary dynamics of the life histories that we consider, and (ii) derive also the full Pontryagin's maximum principle from a fitness-returns argument.

1 Introduction

In their recent paper "Function-valued adaptive dynamics and optimal control theory", Parvinen et al. (2012) 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 our fellow evolutionary ecologists, we restrict our calculations from the start to the most commonly encountered class of life history models, and use old-fashioned mathematical arguments (in keeping with our own training) rather than a more rigorous functional analytical approach. (Mathematicians with a modern schooling can anyway easily translate everything into their language, while the opposite does not hold true for evolutionary ecologists.)

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

Below we consider a life history model in which an individual is characterised by two dynamical variables, in addition to an inherited strategy u influencing their dynamics. u 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 $E_{res} = E_{attr}(u_{res})$, with u_{res} the strategy currently in residence. This assumption of a non-fluctuating environment allows us to make use of the fitness proxy $R_0(u_{mut}; E_{res})$, the average lifetime offspring production of a mutant in the environment E_{res}, calculated e.g. by integrating the average rate of producing kids over the age of an individual. Consistency requires that $R_0(u_{res}; E_{res}) = 1$. If its R_0 is larger than 1, a mutant has a positive probability to invade, else it cannot invade.

The invasion fitness F 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

$$F(u_{mut}; E_{res}) = \frac{\ln\left(R_0(u_{mut}; E_{res})\right)}{T_r(u_{res})} + O\left(\ln^2\left(R_0(u_{mut}; E_{res})\right)\right), \qquad (1)$$

$T_r(u_{res})$ the average age at which the residents give birth in the environment E_{res} (e.g. Metz and Diekmann 1986, Durinx et al. 2008).

Remark Parvinen et al. (loc. cit.) 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. (loc. cit.), 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 $x = u_{mut} - u_{res}$. Mathematically it is the derivative of the invasion fitness for u_{mut} evaluated at $u_{mut} = u_{res} = u$. From the previous approximation formula for the invasion fitness it follows that we can calculate G as

$$G(u) = \frac{1}{T_r(u)} \frac{dR_0}{du_{mut}}(u; E_{attr}(u)) \qquad (2)$$

(Durinx et al. 2008, Metz 2008). (The derivative here is a linear operator providing a first order approximation for the dependence of R_0 on changes in u_{mut}.)

The assumption of a non-fluctuating resident environment moreover makes that we can represent the strategy as a function of age, $u : \mathbb{R}_+ \to [0,1] : a \mapsto u(a)$, and write for a function $x : \mathbb{R}_+ \to \mathbb{R}$

$$G(u)x = \frac{1}{T_r(u)} \int_0^\infty g(a; u)x(a)da \qquad (3)$$

(c.f. Parvinen et al. 2012). 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, 2012; Durinx et al. 2008; Méléard and Tran 2009; Champagnat and Méléard 2011; Gupta et al. in prep.)

$$\frac{\mathrm{d}u}{\mathrm{d}t}(a) = \frac{\sigma^{-2}(u)}{T_{\mathrm{s}}(u)}\bar{n}(u)\mu\int\limits_{0}^{\infty}c(a,\alpha)g(\alpha;u)\mathrm{d}\alpha \qquad \text{if } \int\limits_{0}^{\infty}c(a,\alpha)g(\alpha;u)\mathrm{d}\alpha > 0 \text{ and } u(a)\in[0,1)$$

$$\text{or } \int\limits_{0}^{\infty}c(a,\alpha)g(\alpha;u)\mathrm{d}\alpha < 0 \text{ and } u(a)\in(0,1]$$

$$= 0 \qquad\qquad\qquad\qquad \text{otherwise}, \tag{4}$$

with T_{s} the average age at which the residents die, σ^2 the between individual variance of their offspring numbers, $\bar{n}$ 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

$$\mathrm{cov}\left(\int\limits_{a_1}^{a_2}\underline{x}(a)\mathrm{d}a, \int\limits_{a_3}^{a_4}\underline{x}(a)\mathrm{d}a\right) = \int\limits_{a_1}^{a_2}\int\limits_{a_3}^{a_4}c(a,\alpha)\mathrm{d}a\mathrm{d}\alpha. \tag{5}$$

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 A we briefly consider its extension to Mendelian diploids.

Our formula for the CE is slightly more complicated than the one in Parvinen et al. (2012). The reason is that these authors did not consider local constraints on the strategy, whereas in our case $0 \le u(a) \le 1$. See Appendix B for further information.

Another difference is that Parvinen et al. (loc. cit.) have set the factor σ^2 equal to 2, in keeping with the idea that for the i-models underlying the standard ODE models the distribution of the lifetime offspring number is geometric. Moreover, for standard ODE models $T_{\mathrm{r}} = T_{\mathrm{s}}$ and since the g of Parvinen et al. corresponds to our $g/T_{\mathrm{r}} =: f$, the T_{s} in (4) cancels. Appendix C treats the corresponding considerations for the periodic ODE case.

The equilibria of the CE are the so-called evolutionarily singular strategies (esses). 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) Evolutionarily Steady Strategies (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 (on good grounds: Geritz et al. 2002; Geritz 2005; Dercole and Rinaldi 2008, Appendix B).)

Another way to calculate ESS-es is to maximise the invasion fitness, or alternatively R_0, over u_{mut} followed by setting $u = u_{\mathrm{mut}} = u_{\mathrm{res}}$. It is here that Pontryagin's maximum principle is encountered (e.g. 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. In Section 6 we give explicit expressions for the life history models described in the next section.

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 (i) $[\ ](a)$ means that all functions written without argument between the $[\ ]$ have the argument a. (ii) To avoid clutter we shall usually without special announcement hide the argument E_{res}. (iii) Similarly we shall often hide the argument u_{mut} in expressions like $P(a;u_{\text{mut}})$ for the probability that an individual survives till age a, or $m(a;u_{\text{mut}})$ for its body size at that age. (iv) When we use the argument u, then u stands either for u_{mut} or u_{res}, or, usually, for $u_{\text{mut}} = u_{\text{res}} = u$, with the context making clear which is the case. (v) For a function of a single scalar variable we use a prime to indicate its derivative; a superscript dot indicates a derivative for age.

The two dynamical variables characterising an individual are (i) one i-state variable, to wit its body mass m, starting from a fixed birth mass m_0, and (ii) its probability P to be still alive, starting from 1. The energy intake by an individual with body mass m equals $e(m)$. The strategy u tells which fraction of this intake is used for reproduction while the remains are used for growth. The body mass just increases as $(1-u)e(m)$, while the birth rate is assumed to be some monotone function b of the available energy, $(u,e) \mapsto b(ue)$. Finally the energy allocation is assumed also to affect the death rate $d : (u,m) \mapsto d(u,m)$. All three functions, e, b and d, may depend in addition on the hidden argument E_{res}. For this model the average lifetime offspring number of a mutant strategy u_{mut} equals

$$R_0 = \int_0^\infty \big[Pb(u_{\text{mut}}e(m)) \big](a)\mathrm{d}a, \qquad \text{with}$$

$$\dot{m} = (1 - u_{\text{mut}})e(m), \quad m(0) = m_0, \quad \dot{P} = -d(u_{\text{mut}},m)P, \quad P(0) = 1. \tag{6}$$

When we set $u_{\text{mut}} = u_{\text{res}}$ we should have that $R_0 = 1$ due to the special value of the hidden argument $E_{\text{res}} = E_{\text{attr}}(u_{\text{res}})$ of e, b and d.

4 Calculating the selection gradient from a fitness-returns argument

Now think of how in the definition of the integral the function x in (3) is approximated as the sum of a large number of short blocks, i.e., functions that are constant over a short time interval and zero elsewhere. To calculate g we consider the fitness returns corresponding to such blocks. The fitness return $r(a;u)$ at age a for $u_{\text{mut}} = u_{\text{res}} = u$ is calculated conditional on the individual surviving till a. It should be interpreted as $(\varepsilon\delta)^{-1}$ times the effect on an individual's expected present and future reproduction of increasing u with a block of height ε and duration δ starting at a. To calculate the effect on R_0 of such a block we have to take into account that only a fraction $P(a;u)$ of the individuals reaches that age and hence contributes to the effect. Hence

$$g = Pr. \tag{7}$$

We calculate r from first biological principles, in the spirit of the marginal value theorem. Imagine an increase of the relative allocation to reproduction with an amount ε for a time δ, $0 < \varepsilon, \delta \ll 1$ starting at age a. Let $\tilde{m}$ be the difference that this makes in m, and $\tilde{P}$ the difference that this makes in P. Between the ages a to $a+\delta$ (with $u = u_{\text{res}}$ the allocation function to which we make the change)

$$\dot{\tilde{m}} = (1-u)e'(m)\tilde{m} - \varepsilon e(m) + \mathrm{O}(\varepsilon^2), \quad \tilde{m}(a) = 0,$$

$$\dot{\tilde{P}} = -d(u)\tilde{P} - d'(u)P\varepsilon + \mathrm{O}(\varepsilon^2), \quad \tilde{P}(a) = 0. \tag{8}$$

This leads to

$$\tilde{m}(a+\delta) = -e(m(a))\varepsilon\delta + \mathrm{O}(\varepsilon)\mathrm{O}(\delta^2) + \mathrm{O}(\varepsilon^2)\mathrm{O}(\delta)$$

$$\tilde{P}(a+\delta) = -d'(u(a))P(a)\varepsilon\delta + \mathrm{O}(\varepsilon)\mathrm{O}(\delta^2) + \mathrm{O}(\varepsilon^2)\mathrm{O}(\delta). \tag{9}$$

From then onwards, where from now on we neglect the higher order terms,

$$\dot{\tilde{m}} = (1-u)e'(m)\tilde{m}, \quad \tilde{m}(a+\delta) \approx -e(m(a))\varepsilon\delta,$$

$$\dot{\tilde{P}} = -d(u)\tilde{P}, \quad \tilde{P}(a+\delta) \approx -d'(u(a)P(a)\varepsilon\delta. \tag{10}$$

The immediate gain in offspring from this strategy change for an individual that already has survived till a is

$$\varepsilon\delta\big[b'(ue(m)e(m)\big](a).$$

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

$$-\frac{1}{P(a)}\int_a^\infty \big[\tilde{P}b(ue(m)) + Pb'(ue(m))ue'(m)\tilde{m}\big](\alpha)\mathrm{d}\alpha.$$

Since the equations for $\tilde{P}$ and $\tilde{m}$ are linear, this future offspring loss is also proportional to $\varepsilon\delta$. So for the calculation of r we can already at this stage divide both by ε and δ.

Below we use that for $\alpha > a + \delta$ the differential equation for $P(\alpha)/P(a+\delta)$ is the same as that for $\tilde{P}$ except for a difference in initial conditions. Let $\hat{P}(\alpha;a)$ and $\hat{m}(\alpha;a)$ be defined by

$$\frac{\mathrm{d}\hat{P}}{\mathrm{d}\alpha} = -d(u)\hat{P}, \quad \hat{P}(a;a) = 1, \quad \frac{\mathrm{d}\hat{m}}{\mathrm{d}\alpha} = (1-u)e'(m)\hat{m}, \quad \hat{m}(a;a) = 1. \tag{11}$$

$\hat{P}$ and $\hat{m}$ are functions of both α and a. However, we shall from now on hide the latter argument. Using that $P(\alpha) = P(a)\hat{P}(\alpha)$, $\tilde{P}(\alpha) \approx -[d'(u)P](a)\varepsilon\delta\hat{P}(\alpha)$ and $\tilde{m}(\alpha) \approx -[e(m)](a)\varepsilon\delta\hat{m}(\alpha)$ we find, on letting ε and δ go to zero, that

$$r(a;u) = \big[b'(ue(m))e(m)\big](a)$$

$$-\int_a^\infty \hat{P}(\alpha)\big[e(m(a))b'(ue(m))ue'(m)\hat{m} + d'(u(a))b(ue(m))\big](\alpha)\mathrm{d}\alpha$$

$$= \big[b'(ue(m))e(m)\big](a)$$

$$-e(m(a))\int_a^\infty \big[\hat{P}b'(ue(m))ue'(m)\hat{m}\big](\alpha)\mathrm{d}\alpha - d'(u(a))\int_a^\infty \big[\hat{P}b(ue(m))\big](\alpha)\mathrm{d}\alpha$$

$$\tag{12}$$

At an optimal u (more in particular, an optimal u_{mut} with in addition $u_{mut} = u_{res} = u$, as is the case at an ESS) the return r is 0 when $0 < u(a) < 1$, non-positive when $u(a) = 0$ and nonnegative when $u(a) = 1$. In the Section 6 we consider the determination of an optimal u using Pontryagin's maximum principle. It turns out that there is a detailed match between the results from the two approaches. However, where the Pontryagin argument applies only to an optimal u, the fitness return argument applies to any $u = u_{mut} = u_{res}$.

5 The other ingredients of the canonical equation

To complete the CE we still need T_r, T_s and σ^2. The first two come easy.

$$T_r(u) = \int_0^\infty aP(a;u)b\big(u(a)e(m(a;u))\big)\mathrm{d}a, \tag{13}$$

$$T_s(u) = \int_0^\infty -a\dot{P}(a;u)\mathrm{d}a = \int_0^\infty P(a;u)\mathrm{d}a, \tag{14}$$

where we used that $R_0(u; E_{attr}(u)) = 1$ and that the probability density of the age at death $\underline{a}$ is $-\dot{P}(a;u)$.

To calculate σ^2 we have to be more specific about the microstructure of the reproduction process. The assumption that naturally leads to Formula (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 $b(ue(m))/C(u,e(m)))$. We stick here to the first possibility. Then for a given age at death a the total offspring number is Poisson distributed with mean $\lambda(a;u) := \int_0^a b\big(u(\alpha)e(m(\alpha;u))\big)\mathrm{d}\alpha$. The first moment of $\underline{\lambda} = \lambda(\underline{a};u)$ therefore is equal to

$$\mathbf{E}\underline{\lambda} = R_0(u; E_{attr}(u)) = 1. \tag{15}$$

Hence $\sigma_\lambda^2 = \mathbf{E}\underline{\lambda}^2 - 1$, with

$$\mathbf{E}\underline{\lambda}^2 = \int_0^\infty -\dot{P}(a;u)\lambda^2(a;u)\mathrm{d}a = 2\int_0^\infty \lambda(a;u)P(a;u)b\big(u(a)e(m(a;u))\big)\mathrm{d}a. \tag{16}$$

Finally, from the general rules for mixtures of distributions,

$$\sigma^2 = \sigma_\lambda^2 + 1 = \mathbf{E}\underline{\lambda}^2. \tag{17}$$

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

In contrast to the CE, Pontryagin's maximum principle is textbook material. Hence we shall not give an explanation, but just proceed in the wake of one such textbook, to wit Intrilligator (1971) Chapter 14. In the notation of Intrilligator our model is represented as,

$$\mathbf{x} = \begin{pmatrix} m \\ P \end{pmatrix}, \quad I\left(\begin{pmatrix} m \\ P \end{pmatrix}, u\right) = Pb(ue(m)), \quad J = R_0, \quad \mathbf{f}\left(\begin{pmatrix} m \\ P \end{pmatrix}, u\right) = \begin{pmatrix} (1-u)e(m) \\ -d(u)P \end{pmatrix}. \quad (18)$$

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 to maximise J we should at each age maximise the so-called Hamiltonian,

$$H = I + \mathbf{y}\mathbf{f}, \quad (19)$$

with $\mathbf{y} = (y_1, y_2)$ the so-called co-state variables, where the latter satisfy the differential equations

$$\dot{y}_1 = -\frac{\partial H}{\partial x_1}, \qquad \dot{y}_2 = -\frac{\partial H}{\partial x_2}, \qquad y_1(\infty) = y_2(\infty) = 0. \quad (20)$$

Written out

$$H(m,P,u,y_1,y_2) = Pb(ue(m)) + y_1(1-u)e(m) - y_2 d(u)P,$$
$$\dot{y}_1 = -(1-u)e'(m)y_1 - Pb'(ue(m))ue'(m), \qquad \dot{y}_2 = d(u)y_2 - b(ue(m)),$$
$$\dot{m} = (1-u)e(m), \qquad \dot{P} = -d(u)P,$$
$$y_1(\infty) = 0, \qquad y_2(\infty) = 0, \qquad m(0) = m_0, \qquad P(0) = 1. \quad (21)$$

J is maximised by u if and only if $u(a)$ maximises $[H(m,P,u,y_1,y_2)](a)$ at any a.

The equations for the co-state variables have the solutions

$$y_1(a) = y_{1,0} \exp\left(-\int_0^a [(1-u)e'(m)](\alpha)d\alpha\right)$$

$$-\int_0^a [Pb'(ue(m))ue'(m)](\alpha)\exp\left(-\int_\alpha^a [(1-u)e'(m)](\tau)d\tau\right)d\alpha$$

$$y_2(a) = y_{2,0} \exp\left(\int_0^a [d(u)](\alpha)d\alpha\right) - \int_0^a [b(ue(m))](\alpha)\exp\left(\int_\alpha^a [d(u)](\tau)d\tau\right)d\alpha$$

$$(22)$$

where $y_{1,0}$ and $y_{2,0}$ have to be chosen such that the terminal conditions $y_1(\infty) = 0$, $y_2(\infty) = 0$ are satisfied:

$$y_{2,0} = \lim_{a\to\infty}\left[\exp\left(-\int_0^a [d(u)](\alpha)d\alpha\right)\int_0^a [b(ue(m))](\alpha)\exp\left(\int_\alpha^a [d(u)](\tau)d\tau\right)d\alpha\right], \quad (23)$$

with a similar expression for $y_{1,0}$.

Differentiating $[H](a)$ for $u(a)$ gives

$$g_{\mathrm{H}}(a) := \frac{\partial [H](a)}{\partial u(a)} = \left[Pb'(ue(m))e(m) - y_1 e(m) - y_2 d'(u)P\right](a). \quad (24)$$

An ESS has $g_H(a)$ equal to 0 for $0 < u(a) < 1$, nonnegative for $u(a) = 1$ and nonpositive for $u(a) = 0$. Moreover, where $g_H(a) = 0$ its derivative for $u(a)$ should be negative.

7 Relating the results of Sections 4 and 6

In this section we show that formulas (11) and (12) for calculating fitness returns (c.q. the selection gradient) and formulas (22) – (24) for the differential of H amount to the same.

To enhance the similarity divide g_H by P and set $\tilde{y}_1 = P^{-1}y_1$, $\tilde{y}_2 = y_2$ to get

$$r_H(a) := P^{-1}g_H(a) = \left[b'(ue(m))e(m) - \tilde{y}_1 e(m) - \tilde{y}_2 d'(u) \right](a)$$

with

$$\dot{\tilde{y}}_1 = -(1-u)e'(m)\tilde{y}_1 - \hat{P}b'(ue(m))ue'(m), \qquad \tilde{y}_1(\infty) = 0,$$

$$\dot{\tilde{y}}_2 = d(u)\tilde{y}_2 - b(ue(m)), \qquad \tilde{y}_2(\infty) = 0, \tag{25}$$

which is to be compared with

$$r(a) = \left[b'(ue(m)e(m) \right](a)$$

$$- e(m(a)) \int_a^\infty \left[\hat{P}b'(ue(m))ue'(m)\hat{m} \right](\alpha)\mathrm{d}\alpha - d'(u(\alpha)) \int_a^\infty \left[\hat{P}b(ue(m)) \right](\alpha)\mathrm{d}\alpha$$

with

$$\frac{\mathrm{d}\hat{P}}{\mathrm{d}\alpha} = -d(u)\hat{P}, \quad \hat{P}(a) = 1, \quad \frac{\mathrm{d}\hat{m}}{\mathrm{d}\alpha} = (1-u)e'(m)\hat{m}, \quad \hat{m}(a) = 1. \tag{26}$$

The natural guess is that we should try to identify $\tilde{y}_1$ with

$$\hat{y}_1 := \int_a^\infty \left[\hat{P}b'(ue(m))ue'(m)\hat{m} \right](\alpha)\mathrm{d}\alpha \tag{27}$$

and $\tilde{y}_2$ with

$$\hat{y}_2 := \int_a^\infty \left[\hat{P}b(ue(m)) \right](\alpha)\mathrm{d}\alpha. \tag{28}$$

The mathematical structure of both comparisons is

– for fitness returns: $\quad \hat{y}(a) = \int_a^\infty \varphi(s)z(\alpha)\mathrm{d}\alpha, \quad \dot{z}(\alpha) = -\psi(\alpha)z(\alpha), \quad z(a) = 1, \tag{29}$

– for Pontryagin: $\qquad \dot{\tilde{y}}(a) = \psi(a)\tilde{y}(a) - \varphi(a), \quad \tilde{y}(\infty) = 0. \tag{30}$

Working out the integrals gives

$$\hat{y}(a) = \int_a^\infty \varphi(\alpha)\exp\left[-\int_a^\alpha \psi(\tau)\mathrm{d}\tau \right]\mathrm{d}\alpha \tag{31}$$

and

$$\tilde{y}(a) = -\int_0^a \varphi(s)\exp\left[\int_\alpha^a \psi(\tau)d\tau\right]d\alpha + \tilde{y}_0\exp\left[\int_0^a \psi(\tau)d\tau\right]$$

with

$$\tilde{y}_0 = \exp\left[-\int_0^\infty \psi(\tau)d\tau\right]\int_0^\infty \varphi(\alpha)\exp\left[\int_\alpha^\infty \psi(\tau)d\tau\right]d\alpha = \int_0^\infty \varphi(s)\exp\left[-\int_0^\alpha \psi(\tau)d\tau\right]d\alpha. \quad (32)$$

Hence

$$\begin{aligned}\tilde{y}(a) &= -\int_0^a \varphi(\alpha)\exp\left[\int_\alpha^a \psi(\tau)d\tau\right]d\alpha + \int_0^\infty \varphi(\alpha)\exp\left[\int_\alpha^a \psi(\tau)d\tau\right]d\alpha\\[1mm]&= \int_a^\infty \varphi(\alpha)\exp\left[-\int_a^\alpha \psi(\tau)d\tau\right]d\alpha.\end{aligned} \quad (33)$$

Therefore indeed $\tilde{y}_i = \hat{y}_i$.

In Appendix D we show that more generally the fitness return from a general, not necessarily small, short local change in u after age a multiplied with the probability that an individual survives till that age corresponds to H as defined in Formula (21).

8 Discussion

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 CE 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 showing the explicit connection made in our Sections 4 and 7. 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. We therefore believe that our calculations genuinely add to our collective understanding of the mathematical structure of our world.

On the practical side we put forward 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 may provide an effective computational implementation of that principle.

A second 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. (2012). In particular, it was necessary to extend the CE 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).

As a final point we note that the argument in Section 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 behaviour, 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 deterministic 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 rate of offspring production at age a is just a function b of $[ue(m)](a)$ is at best only a rough approximation. One possible justification is that most of the time $ue(m)$ 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 $\hat{m}$ may well be on average close to $b(ue(\hat{m}))$. 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.

Appendix A: Mendelian organisms

Most life history models implicitly assume clonal reproduction. Yet, by far the majority of organisms that are supposedly targeted by these modelling efforts are Mendelian diploids (c.f. Stearns 1976,1977). To help overcoming this awkward discrepancy we summarize here some results for the Mendelian case. (The underlying technicalities are the subject of another paper that the first author hopes to publish with another co-author.)

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 CE.

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 $u = (u_f, u_m)$, 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 $T_r^{-1}(u)$, 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 fertilisation opportunities, for the other sex.

This independence derives from the additive relation $R_0 = \frac{1}{2}\left[R_f + R_m\right]$, 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, $T_r = \frac{1}{2}\left[T_{r,f} + T_{r,m}\right]$, 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 $S = f, m$,

$$G_S(u) = \frac{\partial F}{\partial u_{S,mut}}(u; E_{attr}(u)) = \frac{1}{T_{r,f}(u) + T_{r,m}(u)} \frac{dR_S}{du_{S,mut}}(u; E_{attr}(u)). \tag{A1}$$

The action of the derivative can again be expressed as an integral $\left[dR_S/du_{S,mut}\right](u; E_{attr}(u))x = \int_0^\infty g_S(a; u)x(a)da$, 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 fertilisation opportunities at different u_{res}.

Finally, $T_s = q_f T_{s,f} + q_m T_{s,m}$, with q_f and q_m the relative frequencies with which the sexes are born into the resident population, and $\sigma^2 = \frac{1}{4}\left[q_f \sigma_f^2 + q_m \sigma_m^2 + q_f^{-1} + q_m^{-1} - 2\right]$ (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}.

The fact that the fertilisation opportunities come as a component of E_{attr} inextricably entwines life history evolution with sex ratio evolution.

Appendix B: How to deal with local constraints

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. However, the CE as given by Parvinen et al. (2012) is based on the assumption that the mutation distribution is symmetric. (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. Hence, on the natural assumption that the constraint does not affect movement parallel to the boundary, in the CE limit it does not affect the right hand side at a where $u(a)$ runs no risk of crossing its constraint 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. In our case 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 C: The canonical equation for periodic ODE population models

The right hand side of the CE equals

[rate at which mutants are produced] × average of [the effect of a mutation
× a linear approximation for the probability that that mutant invades]

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 $R_0 = b/d$ with b and $d = T_s^{-1}$ the per capita birth and death rates of the mutant, while $F = b - d = (R_0 - 1)/T_s$. Hence the invasion probability equals $Q = 1 - R_0^{-1} = R_0 - 1 + O((R_0 - 1)^2)$. More in general, $Q = 2\sigma_e^{-2}\ln(R_0) + O(\ln^2(R_0))$, 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. Durinx et al. 2008). ($\sigma^2 = 2$ for a geometric distribution with mean 1.) The rate at which mutants are produced equals the population birth rate times the per birth probability of a mutation. The factor $\bar{n}$ in (4) appears by re-expressing the population birth rate B of the resident as $\bar{n}/T_s$, based on the general consistency relation $\bar{n} = BT_s$. Below we consider the extension of these considerations to periodic environments; the further extension to general ergodic environments is treated in Ripa and Dieckmann (manuscript).

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 by Kendall (1948)

$$q(\theta) = \frac{1}{1 + \int_\theta^\infty e^{-r(t;\theta)} d(t) \mathrm{d}t} \quad \text{with} \quad r(t;\theta) = \int_\theta^t [b - d](\tau) \mathrm{d}\tau . \tag{A2}$$

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

$$Q = \int_0^1 q(\theta) w(\theta) \mathrm{d}\theta , \tag{A3}$$

with

$$w(\theta) = \frac{b_0(\theta) e^{r_0(\theta;0)}}{\int_0^1 b_0(\tau) e^{r_0(\tau;0)} \mathrm{d}\tau} \tag{A4}$$

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 (A2).

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 $\int_t^{t+1} b_0(\tau) e^{r_0(\tau;t)} \mathrm{d}\tau = \int_t^{t+1} d_0(\tau) e^{r_0(\tau;t)} \mathrm{d}\tau$ (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

$$R_0 = \frac{\overline{b}}{\overline{d}} \tag{A4}$$

with

$$\overline{b} := \int_0^1 b(\tau) \mathrm{d}\tau = \int_t^{t+1} b(\tau) \mathrm{d}\tau \text{ and } \overline{d} := \int_0^1 d(\tau) \mathrm{d}\tau = \int_t^{t+1} d(\tau) \mathrm{d}\tau \tag{A5}$$

(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

$$M(x) := \int_0^\infty e^{-r_x(t;0)} d_x(t) \mathrm{d}t \tag{A6}$$

we can write

$$q(\theta;x) = \frac{1}{1 + \int_\theta^1 e^{-r_x(t;\theta)} d_x(t) \mathrm{d}t + e^{-r_x(1;\theta)} M(x)} . \tag{A7}$$

From $q(\theta;x) \to 0$ for $x \to 0$ it follows that then $M(x) \to \infty$. Hence

$$\frac{\partial q}{\partial x}(\theta;0) = -e^{r_0(1;\theta)} \lim_{x \to 0} \frac{M'(x)}{M^2(x)} \tag{A8}$$

and

$$\tilde{Q}'(0) = -\frac{\int_0^1 b_0(\theta)\mathrm{d}\theta}{\int_0^1 b_0(\theta)e^{r_0(\theta;0)}\mathrm{d}\theta} \lim_{x \to 0} \frac{M'(x)}{M^2(x)} . \tag{A9}$$

To calculate the term after the limit sign we observe that

$$M(x) := \int_0^1 e^{-r_x(\theta;0)}d_x(\theta)\mathrm{d}\theta + e^{-r_x(1;0)}M(x) . \tag{A10}$$

Hence

$$M(x) = \frac{\int_0^1 e^{-r_x(\theta;0)}d_x(\theta)\mathrm{d}\theta}{1 - e^{-r_x(1;0)}} , \tag{A11}$$

$$\lim_{x \to 0} \frac{M'(x)}{M^2(x)} = \frac{-\tilde{F}'(0)}{\int_0^1 e^{-r_0(\theta;0)}d_0(\theta)\mathrm{d}\theta} , \tag{A12}$$

and

$$\tilde{Q}'(0) = \frac{\int_0^1 b_0(\theta)\mathrm{d}\theta}{\int_0^1 b_0(\theta)e^{r_0(\theta;0)}\mathrm{d}\theta \int_0^1 e^{-r_0(\theta;0)}d_0(\theta)\mathrm{d}\theta} \tilde{F}'(0) . \tag{A13}$$

Hence away from local constraints the CE becomes

$$\frac{\mathrm{d}s}{\mathrm{d}t}(\theta) = \frac{1}{2}\mu \int_0^1 b_0(\alpha)\bar{n}(\alpha)\mathrm{d}\alpha \frac{\int_0^1 b_0(\alpha)\mathrm{d}\alpha}{\int_0^1 b_0(\alpha)e^{r_0(\alpha;0)}\mathrm{d}\alpha \int_0^1 e^{-r_0(\alpha;0)}d_0(\alpha)\mathrm{d}\alpha} \int_0^1 c(\theta,\alpha)f(\alpha;s)\mathrm{d}\alpha$$

$$= \frac{1}{2}\mu \frac{\int_0^1 d_0(\alpha)\mathrm{d}\alpha}{\int_0^1 \bar{n}^{-1}(\alpha)d_0(\alpha)\mathrm{d}\alpha} \int_0^1 c(\theta,\alpha)f(\alpha;s)\mathrm{d}\alpha, \tag{A14}$$

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

$$f(\theta;s) = \frac{\mathrm{d}[b\text{-}d](\theta)}{\mathrm{d}s_{\mathrm{mut}}(\theta)}\big(s;E_{\mathrm{attr}}(s)\big) . \tag{A15}$$

Hence the $\bar{n}$ in Parvinen et al. (2012) has to be interpreted as

$$\frac{\int_0^1 d_0(\alpha)\mathrm{d}\alpha}{\int_0^1 \bar{n}^{-1}(\alpha)d_0(\alpha)\mathrm{d}\alpha} .$$

To see how (A14) compares with (4) we first observe that for periodically fluctuating populations there is no immediate counterpart for the equality $\bar{n} = BT_s$, so we substitute the latter in (4), while observing that the counterpart of B in (A14) is $\int_0^1 [b_0(\alpha)\bar{n}(\alpha)\mathrm{d}\alpha]$. After substituting $f(\alpha;s) = g(\alpha;s)/T_r(s)$ in (A14) we then end up with the pairing

$$\frac{2T_r}{\sigma_e^2} = \frac{\int_0^1 b_0(\alpha)\mathrm{d}\alpha}{\int_0^1 b_0(\alpha)\mathrm{e}^{r_0(\alpha;0)}\mathrm{d}\alpha \int_0^1 \mathrm{e}^{-r_0(\alpha;0)}d_0(\alpha)\mathrm{d}\alpha} \ . \tag{A16}$$

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

$$T_r = \bar{d}_0^{-1} = \bar{b}_0^{-1} \ . \tag{A17}$$

Therefore

$$\sigma_e^2 == 2\frac{\int_0^1 b_0(\alpha)\mathrm{e}^{r_0(\alpha;0)}\mathrm{d}\alpha}{\int_0^1 b_0(\alpha)\mathrm{d}\alpha} \frac{\int_0^1 \mathrm{e}^{-r_0(\alpha;0)}d_0(\alpha)\mathrm{d}\alpha}{\int_0^1 d_0(\alpha)\mathrm{d}\alpha} = 2\frac{\int_0^1 \mathrm{e}^{r_0(\alpha;0)}d_0(\alpha)\mathrm{d}\alpha \int_0^1 \mathrm{e}^{-r_0(\alpha;0)}d_0(\alpha)\mathrm{d}\alpha}{\left(\int_0^1 d_0(\alpha)\mathrm{d}\alpha\right)^2} \ .$$

$$\tag{A18}$$

Appendix D: Re-deriving Pontryagin's maximum principle from fitness-return arguments

It is also possible to consider the fitness return for just any, not necessarily small, short change of $u(\alpha)$, $\alpha \in [a, a+\delta)$, say to u^*, while leaving the rest of u unchanged. If the duration δ of the change is sufficiently small, its impact in later life will be $\mathrm{O}(\delta)$, and hence its effect in later life can be calculated from the same linearised equations as before. We only do not linearise for the short period over which the perturbation is active. There we use

$$\frac{\mathrm{d}\tilde{m}}{\mathrm{d}a} = \frac{\mathrm{d}m^*}{\mathrm{d}a} - \frac{\mathrm{d}m}{\mathrm{d}a} = (1-u^*)e(m+\tilde{m}) - (1-u)e(m) = -\left(u^*-u\right)e(m) + \mathrm{O}(\delta), \quad \tilde{m}(a) = 0 \ ,$$

$$\frac{\mathrm{d}\tilde{P}}{\mathrm{d}a} = \frac{\mathrm{d}P^*}{\mathrm{d}a} - \frac{\mathrm{d}P}{\mathrm{d}a} = -d(u^*)(P+\tilde{P}) + d(u)P = -\left(d(u^*) - d(u)\right)P + \mathrm{O}(\delta), \quad \tilde{P}(a) = 0 \ ,$$

$$\tag{A19}$$

since, thanks to the differentiability of the solutions of differential equations, in an interval of length δ there can occur only a change of at most $\mathrm{O}(\delta)$. This gives, on the reasonable assumption that u does not change dramatically on $[t, t+\delta)$, more in particular, there exists a K such that $|u(\alpha) - u(a)| < K(\alpha - a)$,

$$\tilde{m}(a+\delta) = -e(m(a))\left(u^* - u(a)\right)\delta + \mathrm{O}(\delta^2),$$

$$\tilde{P}(a+\delta) = -\left(d(u^*) - d(u(a))\right)P(a)\delta + \mathrm{O}(\delta^2) \ . \tag{A20}$$

(This may be compared with the earlier derived expressions for the case that $u^* - u(a) = \varepsilon$: $\tilde{m}(a+\delta) = -e(m(a))\varepsilon\delta + \mathrm{O}(\varepsilon)\mathrm{O}(\delta^2) + \mathrm{O}(\varepsilon^2)\mathrm{O}(\delta)$, $\tilde{P}(a+\delta) = -d^*(u(a)P(a)\varepsilon\delta + \mathrm{O}(\varepsilon)\mathrm{O}(\delta^2) + \mathrm{O}(\varepsilon^2)\mathrm{O}(\delta)$.)

Immediate fitness gain from this strategy change for an individual that already has survived till a:

$$\delta\big[\,b(u^*e(m)) - b(ue(m))\,\big](a).$$

Future fitness loss from this change in strategy for an individual that already has survived till a:

$$-\frac{1}{P(a)}\int_a^\infty \big[\,\tilde{P}b(ue(m)) + Pb^*(ue(m))ue'(m)\tilde{m}\,\big](\alpha)\mathrm{d}\alpha.$$

On dividing by δ we get a fitness return

$$\hat{r}(u^*,a;u) =$$
$$\big[\,b(u^*e(m)) - b(ue(m))\,\big](a)$$
$$-\int_a^\infty \hat{P}(\alpha)\big[\,e(m(t))\big(u^* - u(a)\big)b'(ue(m))ue'(m)\hat{m} + \big(d(u^*) - d(u(a))\big)b(ue(m))\,\big](\alpha)\mathrm{d}\alpha$$
$$= \big[\,b(u^*e(m)) - b(ue(m))\,\big](a) - \big(u^* - u(a)\big)e(m(a))\int_a^\infty \big[\,\hat{P}b'(uE(m))ue'(m)\hat{m}\,\big](\alpha)\mathrm{d}\alpha$$
$$\hspace{3cm} - \big(d(u^*) - d(u(a))\big)\int_a^\infty \big[\,\hat{P}b(ue(m))\,\big](\alpha)\mathrm{d}\alpha$$
$$= \big[\,b(u^*e(m)) - b(ue(m))\,\big](a) - \big(u^* - u(a)\big)e(m(a))\hat{y}_1(a;u) - \big(d(u^*) - d(u(a))\big)\hat{y}_2(a;u)$$
$$= \big[\,b(u^*e(m))\,\big](a) - u^*e(m(a))\hat{y}_1(a;u) - d(u^*)\hat{y}_2(a;u)$$
$$\hspace{2cm} - \big[\,b(ue(m))\,\big](a) + u(a)e(m(a))\hat{y}_1(a;u) + d(u(a))\hat{y}_2(a;u).$$
$$(A21)$$

This suggests introducing

$$H(u^*,a;u) = P(a)\big[\,b(u^*e(m))\,\big](a) + (1 - u^*)e(m(a))\hat{y}_1(a;u) - d(u^*)\hat{y}_2(a;u), \quad (A22)$$

so that

$$\hat{r}(u^*,a;u) = \big(H(u^*,a;u) - H((u(a),a;u)\big)\big/P(a). \quad (A23)$$

H corresponds to the Hamiltonian from Pontryagin's maximum principle. The change in R_0 from adding x, $\int_0^\infty |x(t)|\mathrm{d}t < \varepsilon$, to u equals $P(a)\int_0^\infty \hat{r}\big(u(a) + x(a),a,u\big)\mathrm{d}a + \mathrm{O}(\varepsilon^2)$. This way the Hamiltonian gets a meaning also for u other than the optimal one.

References

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

Champagnat N (2003) 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

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

Dercole F, Rinaldi S (2008) *Analysis of evolutionary processes: the adaptive dynamics approach and its applications.* Princeton UP, 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. DOI 10.1007/s00285-007-0134-2

Gupta A, Metz JAJ, Tran VC (in prep) Adaptive dynamics using averaging techniques.

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 (1971) *Mathematical optimization and economic theory.* Prentice Hall. Englewood Cliffs, NJ

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

Gyllenberg M, Metz JAJ, Service R (2011) When do optimisation arguments make evolutionary sense? P 235-269 in Chalub FACC, Rodrigues JF (ed.) *The Mathematics of Darwin's Legacy.* Birkhauser, Basel

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. Pp. 1599-1612 in Jørgensen SE, Fath BD (Eds) *Evolutionary Ecology.* Vol. 2 of *Encyclopedia of Ecology.* Elsevier, Oxford

Metz JAJ, Diekmann O eds (1986) *The dynamics of physiologically structured populations*, Springer Verlag, *Lecture Notes in Biomathematics* **68**

Metz JAJ, Diekmann O (1986) Age dependence. In: Metz JAJ, Diekmann O (eds) *The dynamics of physiologically structured populations*, Springer Verlag, *Lecture Notes in Biomathematics* **68**: 136-184

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

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 (2012) Function-valued adaptive dynamics and optimal control theory. *J Math Biol* **online first** DOI 10.1007/s00285-012-0549-2

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

Ripa J, Dieckmann U (manuscript) Mutant invasions and adaptive dynamics in variable environments.

Stearns SC (1976) Life history tactics: a review of the ideas. *Q Rev Biol* **52**: 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