# The Canonical Equation of Adaptive Dynamics:
# A Mathematical View

N. CHAMPAGNAT[1*], R. FERRIÈRE[1,2,3] and G. BEN AROUS[4]

[1]*Department of Ecology, Unit of Mathematical Eco-Evolutionary Biology,*
*Ecole Normale Supérieure, Paris, France,*
[2]*Adaptive Dynamics Network, IIASA, Laxenburg, Austria,*
[3]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA,*
[4]*Department of Mathematics and Applications, EPFL, Lausanne, Switzerland*

The Darwinian evolution of a quantitative adaptive character is described as a jump process. As the variance of the distribution of mutation steps goes to zero, this process converges in law to the solution of an ordinary differential equation. In the case where the mutation step distribution is symmetrical, this establishes rigorously the so-called canonical equation first proposed by Dieckmann and Law (1996). Our mathematical approach naturally leads to extend the canonical equation to the case of biased mutations, and to seek ecological and genetic conditions under which evolution proceeds either through punctualism or through radiation.

## 1. Introduction

The Darwinian evolution of a quantitative trait involves three biological ingredients: reproduction passes the trait through generations; mutation generates variations in the trait value; selection, resulting from ecological interactions between individuals and their environment, acts upon trait variations. In the early 90s, Hofbauer and Sigmund (1990), Marrow et al. (1992) and Metz et al. (1992) have cast a new class of models that incorporate these three processes to describe the evolutionary dynamics of adaptive traits. One important advance of this theory due to Dieckmann and Law (1996) is the so-called canonical equation of adap-

tive dynamics: an ordinary differential equation that gives the rate of change over time of the expected trait value in a monomorphic population (Dieckmann and Law, 1996). For a scalar trait $s$, the canonical equation takes on the following form

$$\frac{d\hat{s}}{dt} = \mu(\hat{s}) \frac{\sigma_0^2(\hat{s})}{2} n(\hat{s}) \partial_1 f(\hat{s}, \hat{s}) \qquad (1)$$

where $\hat{s}$ denotes the mean of the distribution of the trait value $s$ at time $t$; $\mu(s)$ is the probability that a birth from an individual with trait $s$ gives rise to a mutation; $\sigma_0^2(s)$ denotes the variance of the distribution of a mutant trait $s'$ born from an individual with trait $s$, whose probability density function is denoted by $M(s, s' - s)$ and supposed to be symmetrical with respect to $s' - s$; $n(s)$ is the equilibrium population size (which is supposed to

exist) in a population composed only of individuals with trait $s$; $\partial_1 f(s,s)$ denotes the partial derivative with respect to the first variable at point $(s,s)$, of the fitness function $f(s',s)$ of an individual with trait $s'$ in a population composed only of individuals with trait $s$ (this fitness function can be calculated from birth and death individual parameters as influenced by interactions between individuals and their environment). Fixed points of the canonical equation are the points where the fitness gradient $\partial_1 f(s,s)$ nullifies, and are called 'evolutionary singularities'.

Deriving the canonical equation requires to make appropriate approximations on the underlying mutation-selection process, but so far no complete mathematical argument has been expounded to show (i) that the processes involved are well defined, and (ii) in which sense the 'real' process does converge towards the approximate process. The purpose of this paper is to fill the gap, thereby establishing a firm mathematical basis for the canonical equation of adaptive dynamics. The mathematical techniques we use are fairly demanding (at least they may look so to those readers who are not so keen to probability theory!), yet they pay off eventually by naturally pointing out new directions for modeling the evolutionary dynamics of adaptive traits. Thus, our study demonstrates in general that pure mathematics motivated by important biological issues can feedback into our apprehension of biological phenomena with new valuable questions – and in favorable instances with novel answers to them.

The paper is organized as follows. In section 2, we explain that the trait substitution sequence (Metz et al., 1992, 1996) involved in adaptive dynamics can be seen in the framework of probability theory as a jump process. In section 3, we outline how to recover the canonical equation from this jump process by taking the appropriate limit on parameters that characterize the mutation process. We shall be referring mostly to methods for Markov processes and their convergence that can be found in the celebrated book by Ethier and Kurtz (1986). Notice that complete proofs are rather technical and will be presented in full detail elsewhere. Section 4 paves the way for new developments: extending the canonical equation to biased mutation, and to polymorphic populations. The biological insights that we earn from this mathematical endeavor are recaped and discussed in the last section.

## 2. Adaptive dynamics as a jump process

The derivation of the canonical equation (1) by Dieckmann and Law (1996) stems from modeling the dynamics of a population as a Markov process that accounts for reproduction, mutation and selection. This approach assumes asexual reproduction, which entails that mutation is the only source of trait variation (no recombination). There are two further critical assumptions. First, mutations are supposed to be rare, so that the fate of a mutant is entirely settled by selection before a new mutant enters the scene. Second, the following principle of mutual exclusion applies: a mutant either invades and replaces the wild type, or it is eliminated (due to bad luck or selective inferiority) – no long-term coexistence is allowed unless the inferior type is renewed by mutation. These two assumptions make it possible to define an evolutionary time-scale over which the population is *monomorphic* at any time $t$, which means that there is only one trait value present in the population. The dynamics take place in some trait space denoted by $S$; at any (evolutionary) time $t$, if $s$ is the only trait in the population, the population size is at its equilibrium $n(s)$, which is determined by the ecology of the system. Thus, evolution proceeds as a trait substitution sequence (Metz et al., 1992, 1996) in which substitutions occur at random times. In other words, it is a *jump process* in the trait space $S$.

The continuous-time stochastic process in the trait space $S$ (here $S$ is an open subset of $\mathbf{R}^d$) considered by Dieckmann and Law (1996) is described by the "transition probabilities per unit time for the trait substitution $s \to s'$", denoted by $w(s',s)$, which means that the rate of transition from a given population state (trait value) $s$ to any state $s'$ in a measurable set $\Gamma$ of $S$ is $\int_\Gamma w(s',s)ds'$. We now introduce a more precise characterization of such a Markov process in terms of its (infinitesimal) generator. We recall that to any Markov process $\{X(t), t \geq 0\}$ with values in some set $E$ can be associated an operator $A$ on the set of bounded real-valued functions on $E$, $\mathbf{B}(E)$,

which generally characterizes the Markov process $X$. A possible interpretation of this generator is that it tells us how the mean of any *observable* $\varphi \in \boldsymbol{B}(E)$ of the Markov process $X$ evolves across time; namely, for any $x \in E$, let us define the quantity $E[\varphi(X(t))]$ for the process starting at $X(0) = x$. Then, as a function of $x$, $E[\varphi(X(t))]$ must obey the following differential equation:

$$\frac{d}{dt} E[\varphi(X(t))] = A E[\varphi(X(t))]$$

$$= E[A\varphi(X(t))]. \tag{2}$$

The study of the generator of a Markov process can yield much information about the process, especially about convergence.

The intuitive meaning of the transition rate $w(s',s)$ introduced by Dieckmann and Law (1996) translates heuristically into the convergence of the measures $t^{-1}P(t,s,\cdot)$ toward the measure with density $w(\cdot,s)$ as $t \to 0$, where for any measurable subset $\Gamma$ of $\boldsymbol{S}$, $P(t,s,\Gamma)$ is the transition probability from state $s$ at time 0 toward any state in $\Gamma$ at time $t$. Then, formally:

$$A\varphi(s) = \lim_{t\to 0} t^{-1} E[\varphi(X(t)) - \varphi(s) \mid X(0) = s]$$

$$= \lim_{t\to 0} \int_{S} [\varphi(s') - \varphi(s)] t^{-1} P(t,s,ds')$$

$$= \int_{\boldsymbol{S}} [\varphi(s') - \varphi(s)] w(s',s) ds', \tag{3}$$

and this operator takes the classical form of *the generator of a jump process*.

Thus, the process described by Dieckmann and Law (1996) is a jump process with generator $A$, defined on the domain $\boldsymbol{B}(\boldsymbol{S})$, and given by

$$A\varphi(s) = \int_{\boldsymbol{S}} [\varphi(s') - \varphi(s)] w(s',s) ds'. \tag{4}$$

It is shown in Dieckmann and Law (1996) that

$$w(s',s) = \mu(s)b(s)n(s)M(s,s'-s)\frac{(f(s',s))_+}{b(s',s)} \tag{5}$$

where $\mu(s)$, $n(s)$, $M(s,s'-s)$ and $f(s',s)$ have the same interpretation as in the canonical equation (1); $b(s',s)$ is the per capita birth rate of a single individual with trait $s'$ in a population comprising $n(s)$ individuals with trait $s$; and $b(s) = b(s,s)$. The so-called fitness function $f(s',s)$ is equal to $b(s',s) - d(s',s)$, where $d(s',s)$ is the per capita death rate of an individual with trait $s'$ in a population of trait $s$, and $(\cdot)_+$ denotes the positive part, i.e. $h_+ = h \wedge 0$.

To obtain the canonical equation (1), Dieckmann and Law (1996) assume that the deviations of the jump process from the mean path are small, i.e. the jumps of the process are small. This assumption allows one to write

$$E\left[\int_{S} (s' - X(t))w(s',X(t))ds'\right]$$

$$\approx \int_{S} (s' - E[X(t)])w(s',E[X(t)])ds'. \tag{6}$$

If we suppose that the variance $\sigma_0^2(s)$ of $M(s,s'-s)ds'$ is finite for all $s$, this assumption of small deviations amounts to the *hypothesis that $\sigma_0$ is small*.

Considering the case $\boldsymbol{S} = \boldsymbol{R}$, we are going to show that as $\sigma_0 \to 0$ the jump process converges in law towards the deterministic process that is solution to the equation

$$\frac{ds}{dt} = \mu(s)\frac{\sigma_0^2(s)}{2}n(s)\partial_1 f(s,s) \tag{7a}$$

$$= \frac{\sigma_0^2(s)}{2}\partial_2 g(s,s) \tag{7b}$$

where we have set $g(s,s') = \mu(s)b(s)n(s)\frac{f(s',s)}{b(s',s)}$, so that $w(s',s) = g_+(s,s')M(s,s'-s)$, and $g(s,s) = 0$ since $f(s,s) = 0$ for all $s \in \boldsymbol{S}$. The equation (7b) is more precise than (1) for it gives the exact dynamics of the process, not the dynamics of its mean. The bottom line is that, on a bounded interval of time, under the assumption of sufficiently small mutations, the behaviour of the system is nearly deterministic.

## 3. Recovering the canonical equation: A tale of limits

We are now in a position to prove formally that the canonical equation can be recovered as the variance of the mutation distribution goes to zero. In a first part, we keep assuming that the biological system comprises only one species, and that the trait space is one-dimensional ($S = \mathbf{R}$). Next we extend the results to the case of multispecific communities, as in Dieckmann and Law (1996), and to the case of a multidimensional trait space.

### 3.1. One species, one trait

At this stage we have to construct a family of processes such that the variance of the size of the jumps can be made arbitrarily small. A natural way to do this is to modify the original jump process such that this variance is $\sigma^2 \sigma_0^2$ instead of $\sigma_0^2$, and to let the positive parameter $\sigma$ go to 0. To this end, for all $\sigma > 0$ let

$$A_\sigma \varphi(s) = \frac{1}{\sigma^2} \int_{\mathbf{R}} (\varphi(s') - \varphi(s)) g_+(s,s') M\left(s, \frac{s'-s}{\sigma}\right) \frac{ds'}{\sigma} \quad (8)$$

be an operator on $B(S)$. The variance of the probability measure $M\left(s, \frac{s'-s}{\sigma}\right) \frac{ds'}{\sigma}$ is:

$$\int_{\mathbf{R}} s'^2 M\left(s, \frac{s'-s}{\sigma}\right) \frac{ds'}{\sigma} = \sigma^2 \sigma_0^2, \quad (9)$$

as required. As the variance of the jumps is multiplied by $\sigma^2$, the factor $1/\sigma^2$ is introduced to rescale time so that the variation of trait values in one unit of time is unchanged (without this factor, the process would be constant in the limit). In other words, making the mutation step variance small requires that we track the process over sufficiently long times so that a sufficiently large number of mutations compensates for smaller mutation steps.

In order to construct explicitly a Markov (jump) process $\{X_\sigma(t), t \geq 0\}$ in $\mathbf{R}$ with generator $A_\sigma$ (see Ethier and Kurtz, 1986, pp. 162–164), we have to assume that

$\forall s \in \mathbf{R}$ and $\forall \sigma > 0$,

$$\lambda_\sigma(s) = \int_{\mathbf{R}} g_+(s,s') M\left(s, \frac{s'-s}{\sigma}\right) \frac{ds'}{\sigma} < \infty \quad (10)$$

and $s \mapsto \lambda_\sigma(s)$ is bounded.

We can now state our main result:

**Theorem 1** *Suppose that g is bounded on $\mathbf{R}^2$, $C^1$ with respect to the second variable, and $\partial_2 g$ is bounded and Lipschitz on $\mathbf{R}^2$. Suppose also that the p.d.f. $M(s, s' - s)$ has finite and bounded third moments, and that its variance $\sigma_0^2(s)$ is Lipschitz on $\mathbf{R}$. Suppose finally that $\{X_\sigma(0)\}$ converges in law to a probability measure $v$ on $\mathbf{R}$ as $\sigma$ goes to 0. Let $\{X(t), t \geq 0\}$ be a process defined on the state space $\mathbf{R}$ whose initial state follows distribution $v$, and whose sample paths are differentiable and are solutions to the equation (7). Then $X_\sigma$ converges in law to X when $\sigma$ goes to 0.*

Here the convergence in law of the processes $X_\sigma$ to the process $X$ means that for all finite sequence $\{t_1, \ldots, t_k\}$ of $\mathbf{R}_+$, the random variables $(X_\sigma(t_1), \ldots, X_\sigma(t_k))$ converge in law to $(X(t_1), \ldots, X(t_k))$.

A full proof of theorem 1 will be presented elsewhere (Champagnat et al., in prep.). Here let us examine briefly how the assumptions of theorem 1 are used to achieve the existence of processes $X_\sigma$ and $X$. The boundedness of $g$ ensures that conditions (10) are fulfilled, so that the processes $X_\sigma$ exist. The hypothesis on the third moment of $M$ implies that $\sigma_0$ is bounded; then $g$ and $\sigma_0$ are Lipschitz, bounded functions, and the Cauchy–Lipschitz theorem ensures the existence and unicity of the process $X$ for all $t \geq 0$. The assumptions on $g$ are obviously satisfied if we restrict the trait space to a bounded subset of $\mathbf{R}$, and if we assume that the functions $\mu$, $b$, $n$ and $f$ are $C^2$. Notice that the only condition in theorem 1 that is not necessary to establish the existence of the proc-

esses $X_\sigma$ and $X$ is the hypothesis that $M$ has finite and bounded third moments; thus, our result is nearly optimal. Perhaps the assumption on $M$ third moments could be relaxed, but the methods involved here do not seem to permit such an improvement.

### 3.2. Multispecific coevolutionary dynamics

Following on the agenda set in Dieckmann and Law (1996), we extend the previous result concerning a monospecific system (theorem 1) to a community comprising $k$ co-evolving species. As before, the trait space of each species is $\mathbf{R}$, and each population is assumed to be monomorphic at any (evolutionary) time. We denote the trait of species $i$ by $s_i$, and we set $S = (s_1, \ldots, s_k)$. Here the jump process considered by Dieckmann and Law (1996) is described by $w_i(s_i', S)$, the "transition probabilities per unit time for the trait substitution $s_i \rightarrow s_i'$ in species $i$ and in the community determined by $S$". These probabilities are given by

$$w_i(s_i', S) = \mu_i(s_i) b_i(s_i, S) n_i(S) \times$$

$$M_i(s_i, s_i' - s_i) \frac{(f(s_i', S))_+}{b_i(s_i', S)} \quad (11)$$

where $\mu_i(s_i)$ is the probability that a birth from an individual of species $i$ with trait $s_i$ gives rise to a mutation, $b_i(s_i', S)$ is the birth rate of a single individual of species $i$ with trait $s_i'$ in the community characterized by $S$, $n_i(S)$ is the equilibrium size of species $i$ in the community characterized by $S$, $M_i(s_i, s_i' - s_i)$ is the p.d.f. of the distribution of a mutant trait $s_i'$ born from an individual of species $i$ with trait $s_i$, and $f_i(s_i', S)$ is the fitness of a single individual of species $i$ with trait $s_i'$ in the community characterized by $S$ (if $s_i' = s_i$, $f_i(s_i, S) = 0$ for all $S \in \mathbf{R}$ and $i = 1, \ldots, k$).

Defining $g_i(s_i', S) = \mu_i(s_i) b_i(s_i, S) n_i(S) \frac{f_i(s_i', S)}{b_i(s_i', S)}$ for all $i = 1, \ldots, k$, the generator of this process writes

$$A_1 \varphi(S) = \sum_{i=1}^{k} \int_{\mathbf{R}} (\varphi(S_i') - \varphi(S)) w_i(s_i', S) ds_i' \quad (12)$$

$$= \sum_{i=1}^{k} \int_{\mathbf{R}} (\varphi(S_i') - \varphi(S)) g_{i+}(s_i', S) M_i(s_i, s_i' - s_i) ds_i' \quad (13)$$

where $S_i' = (s_1, \ldots, s_{i-1}, s_i', s_{i+1}, \ldots, s_k)$ if $S = (s_1, \ldots, s_k)$. This is in fact the generator of a jump process since

$$A_1 \varphi(S) = \int_{\mathbf{R}^k} (\varphi(S') - \varphi(S)) W(dS', S), \quad (14)$$

with

$$W(\Gamma, S) = \sum_{i=1}^{k} \int_{\Gamma \cap \{S + \mathbf{R}e_i\}} w_i(s_i', S) ds_i', \quad (15)$$

where $e_i$ is the $i^{\text{th}}$ unit vector of $\mathbf{R}^k$.

Next we proceed as in the previous section: for all $\sigma > 0$ let $X_\sigma$ be a jump process on $\mathbf{R}^k$ generated by

$$A_\sigma \varphi(S) = \sum_{i=1}^{k} \frac{1}{\sigma^2} \times$$

$$\int_{\mathbf{R}} (\varphi(S_i') - \varphi(S)) g_{i+}(s_i', S) M_i\left(s_i, \frac{s_i' - s_i}{\sigma}\right) \frac{ds_i'}{\sigma}. \quad (16)$$

The existence of $X_\sigma$ will be ensured by the hypothesis of theorem 2.

The limiting process that we obtain in theorem 2 is the solution to the following differential equation:

$$\frac{ds_i}{dt} = \frac{\sigma_i^2(s_i)}{2} \partial_1 g_i(s_i, S) \quad (17)$$

$$= \mu_i(s_i) \frac{\sigma_i^2(s_i)}{2} n_i(S) \partial_1 f(s_i, S), \quad (18)$$

for all $i = 1, \ldots, k$

where $\sigma_i^2$ is the second moment of the jump distribution for species $i$, the p.d.f. of which is $M_i(s_i, s_i' - s_i)$. The following theorem therefore provides a formal statement of the analogous result by Dieckmann and Law (1996).

**Theorem 2** *Suppose that for all $i = 1, \ldots, k$, $g_i$ is bounded on $\mathbf{R}^{k+1}$, $C^1$ with respect to the first variable, and that $\partial_1 g_i$ is bounded and Lipschitz on $\mathbf{R}^{k+1}$. Suppose also that for all $i = 1, \ldots, k$ the p.d.f. $M_i(s_i, s_i' - s_i)$ have finite and bounded third moments, and that their second moments $\sigma_i^2$ are Lipschitz on $\mathbf{R}$. Suppose finally that $\{X_\sigma(0)\}$ converges in law to a probability measure $v$ on $\mathbf{R}^k$ as $\sigma$ goes to 0. Let $\{X(t), t \geq 0\}$ be a process defined on the state space $\mathbf{R}^k$ whose initial state follows distribution $v$, and whose sample paths are differentiable and are solutions to the equations (17). Then $X_\sigma$ converges in distribution to $X$ when $\sigma$ goes to 0.*

The proof of theorem 1 adapts straightforwardly here (Champagnat et al., in prep.). Once again, this theorem stems from nearly minimal assumptions.

### 3.3. Multidimensional trait space

To keep notations as simple as possible, the adaptive dynamics of multidimensional traits (or $d$ distinct real traits) are considered in case of a single species only. The result (theorem 3) can be extended to the multispecific setting without any argument beyond those already expounded.

Let $S = \mathbf{R}^d$. $S = (s_1, \ldots, s_d)$ denotes the trait state of a (monomorphic) population; $s_i$ can be seen as the $i^{\text{th}}$ real coordinate of the trait; $S' = (s_1', \ldots, s_d')$ indicates a mutant trait. The dynamics can be described as follows: consider $w(S', S)$ the "transition probabilities per unit time for the trait substitution $S \to S'$ in a population with trait $S$". The expression for $w(S', S)$ reads

$$w(S', S) = \mu(S)b(S)n(S) \times$$

$$M(S, S' - S)\frac{(f(S', S))_+}{b(S', S)} \qquad (19)$$

where the parameters have the same interpretation as in (5), except that the traits considered here are $d$-dimensional.

Let $g(S, S') = \mu(S)b(S)n(S)\frac{f(S', S)}{b(S', S)}$. Then the generator of the adaptive process is

$$A_1\varphi(S) = \int_{\mathbf{R}^d} (\varphi(S') - \varphi(S))w(S', S)dS' \qquad (20)$$

$$= \int_{\mathbf{R}^d} (\varphi(S') - \varphi(S))g_+(S, S')M(S, S' - S)dS'. \quad (21)$$

We keep following the general layout of section "One species, one trait", and introduce for all $\sigma > 0$ the jump process $X_\sigma$ on $\mathbf{R}^d$ generated by

$$A_\sigma\varphi(S) = \frac{1}{\sigma^2}\int_{\mathbf{R}^d} (\varphi(S') - \varphi(S))g_+(S, S') \times$$

$$M\left(S, \frac{S' - S}{\sigma}\right)\frac{dS'}{\sigma}. \qquad (22)$$

The existence of $X_\sigma$ is ensured by the hypothesis of theorem 3. The limiting process is given by the solution to the system of differential equations

$$\frac{ds_i}{dt} = \frac{\left\langle \nabla_2 g(S, S) \cdot \Sigma_i^2(S) \right\rangle}{2} \qquad (23)$$

for all $i = 1, \ldots, d$,

where $\nabla_2 g$ is the gradient of $g$ with respect to second variable $S'$, $\langle \cdot \rangle$ denotes the scalar product, and $\Sigma_i^2(S) = \int_{\mathbf{R}^d} h_i HM(S, H)dH$. Then the following result is in line with Leimar's (pers. com.):

**Theorem 3** *Suppose that $g(S, S')$ is bounded on $\mathbf{R}^{2d}$, $C^1$ with respect to the second variable $S'$, and that $\nabla_2 g$ is bounded and Lipschitz on $\mathbf{R}^{2d}$. Suppose also that the p.d.f. $M(S, S' - S)$ has finite and bounded third moments, and that for all $S$, $S'$ and $H$ in $\mathbf{R}^d$, $|M(S', H) - M(S, H)| \leq \|S' - S\|m(H)$ where $m$ is defined on $\mathbf{R}^d$, measurable, and such that $\int_{\mathbf{R}^d} |h_i| |h_j| m(H)dH$ is finite for all $i$ and $j$ in $\{1, \ldots, d\}$. Suppose finally that $\{X_\sigma(0)\}$ converges in law to a probability measure $v$ on $\mathbf{R}^d$ as $\sigma$ goes to 0. Let $\{X(t), t \geq 0\}$*

be a process defined on the state space $\mathbf{R}^d$ whose initial state follows distribution $V$, and whose sample paths are differentiable and are solutions to the equations (23). Then $X_\sigma$ converges in law to $X$ when $\sigma$ goes to 0.

The proof is slightly more demanding than for theorem 1 (Champagnat et al., in prep.). First we must show the existence of the process $X$. To this end, we have to prove that $\int_{\mathbf{R}^d} \langle \nabla_2 g(S,S) \cdot H \rangle_+ h_i M(S,H) dH$, as a function of $S$, is bounded and Lipschitz. This is why the hypothesis $|M(S',H) - M(S,H)| \leq \|S' - S\| m(H)$ for all $S$, $S'$ and $H$ in $\mathbf{R}^d$ (that may look odd at first sight) is necessary. Second, proving the convergence of generators turns out to be more technical. However, the assumptions are not so restrictive and should be satisfied in many particular situations.

## 4. Beyond the canonical equation

In this section we delineate two directions for further investigation of adaptive dynamics that goes beyond the canonical model (1). First we envisage the case of biased mutation. We show how the canonical equation is altered when the assumption of a symmetric density of mutation steps is relaxed. The biological scope of this extension will be discussed in the next section. Second we examine how to model adaptive dynamics by making use of diffusion processes, rather than jump processes. Our preliminary result points to a potential way of reconciling the canonical equation paradigm that is valid in monomorphic populations, and the branching phenomenon which begets a transition between monomorphic and polymorphic population states.

### 4.1. Asymmetrical mutation density

We will use here the notations of the one species and one trait case. The symmetry of $M(s,\cdot)$ is used in the proof of theorem 1 only when calculating the limiting generator of the $X$ process. It is therefore natural to aim at generalizing this result

to the case of non-symmetrical distributions of mutant traits, i.e. biased mutation. Biased mutation entails that the jump process has a predominant direction of jumps around any trait $s$, given by the sign of the mean of $M(s, s' - s) ds'$. In this case, the rescaling (8) of generators $A_\sigma$ entails that the jumps of all processes $X_\sigma$ have the same predominant direction as the original jump process, but the mean and the standard deviation of these jumps are both multiplied by $\sigma$.

In this case, we introduce a new version of the canonical equation for biased mutation:

$$\frac{ds}{dt} = \mu(s)\sigma_\bullet^2(s)n(s)\big(\partial_1 f(s,s)\big)_+$$

$$+ \mu(s)\sigma_\circ^2(s)n(s)\big(\partial_1 f(s,s)\big)_- \qquad (24)$$

$$= \sigma_\bullet^2(s)\big(\partial_2 g(s,s)\big)_+ + \sigma_\circ^2(s)\big(\partial_2 g(s,s)\big)_- \quad (25)$$

where $h_+ = h \wedge 0$ is the positive part of $h$, $h_- = h \vee 0$ is the negative part of $h$,

$$\sigma_\bullet^2(s) = \int_{\mathbf{R}_+} u^2 M(s,u) du \text{ and}$$

$$\sigma_\circ^2(s) = \int_{\mathbf{R}_-} u^2 M(s,u) du,$$

so that $\sigma_0^2(s) = \sigma_\bullet^2(s) + \sigma_\circ^2(s)$. We call $\sigma_\bullet^2$ and $\sigma_\circ^2$ the *positive* and *negative* second moments of $M$, respectively.

Then our main result becomes:

**Theorem 4** *Suppose that $g$ is bounded on $\mathbf{R}^2$, $C^1$ with respect to the second variable, and that $\partial_2 g$ is bounded and Lipschitz on $\mathbf{R}^2$. Suppose also that the p.d.f. $M(s, s' - s)$ has finite and bounded third moments, and that its positive and negative second moments $\sigma_\bullet^2$ and $\sigma_\circ^2$ are Lipschitz on $\mathbf{R}$. Suppose finally that $\{X_\sigma(0)\}$ converges in law to a probability measure $V$ on $\mathbf{R}$ as $\sigma$ goes to 0. Let $\{X(t), t \geq 0\}$ be a process defined on the state space $\mathbf{R}$ whose initial state follows distribution $V$, and whose sample paths are differentiable and are solutions to the equation (25). Then $X_\sigma$ converges in law to $X$ when $\sigma$ goes to 0.*

The proof straightforwardly imitates that of theorem 1.

The canonical equation (24) obtained with biased mutation extends without any difficulty to multispecific coevolutionary dynamics. In this situation the limiting process $X$ involved in theorem 2 is the solution to the following system of differential equations:

$$\frac{ds_i}{dt} = \sigma_{i\bullet}^2(s_i)(\partial_1 g_i(s_i, S))_+$$

$$+ \sigma_{i\circ}^2(s_i)(\partial_1 g_i(s_i, S))_-, \qquad (26)$$

$$\text{on } \mathbf{R}_+, \text{ for all } i = 1, \ldots, k,$$

where $\sigma_\bullet^2$ and $\sigma_\circ^2$ are, respectively, the positive and negative second moment of the jumps distribution for species $i$, and $S = (s_1, \ldots, s_k)$.

Likewise, the canonical equation extends to a multidimensional trait space with biased mutation. In theorem 3 the limiting process is now given by the solution to the system of differential equations:

$$\frac{ds_i}{dt} = \int_{\mathbf{R}^k} \langle \nabla_2 g(S, S) \cdot H \rangle_+ h_i M(S, H) dH, \qquad (27)$$

$$\text{on } \mathbf{R}_+, \text{ for all } i = 1, \ldots, d,$$

where $\nabla_2 g$ is the gradient of $g$ with respect to the second variable $S'$, $\langle \cdot \rangle$ denotes the scalar product, $S = (s_1, \ldots, s_d)$ and $H = (h_1, \ldots, h_d)$. As before, the integral is taken over a half space of $\mathbf{R}^k$ which depends on $S$.

### 4.2. Diffusion approximation

A further way of describing adaptive dynamics is to approximate the mutation-selection process in the small $\sigma$ limit by a diffusion process, obtained by expanding the generator $A_\sigma$ with respect to $\sigma$. This yields a generator of the form $A + \sigma L$ where $A$ is the generator of the process $X$ solution of the canonical equation and $L$ is a diffusion operator. In the special case $S = \mathbf{R}$, with only one species, we obtain

$$(A + \sigma L)\varphi(x) = \sigma D(x)\varphi''(x)$$

$$+ (B_1(x) + \sigma B_2(x))\varphi'(x), \qquad (28)$$

where

$$D(x) = \frac{M_3(x)}{2}|\partial_2 g(x, x)|, \qquad (29)$$

$$B_1(x) = \frac{\sigma_0^2(x)}{2}\partial_2 g(x, x), \qquad (30)$$

$$B_2(x) = \frac{M_3(x)}{2}\text{sgn}[\partial_2 g(x, x)]\partial_2^2 g(x, x), \qquad (31)$$

$$M_3(x) = \int_0^\infty h^3 M(x, h) dh. \qquad (32)$$

This formalism suggests to seek a large deviation principle (Wentzel, 1976a, b; Freidlin and Wentzel, 1984) for the sample paths of this diffusion process, when $\sigma$ goes to 0. This perspective raises new questions about the adaptive process: What is the mean time needed to exit the basin of an attractive evolutionary singularity (as predicted by the canonical equation)? When the canonical equation possesses multiple attractive evolutionary singularities, what is the sequence of singularities that the system is more likely to visit? How are these results affected by the nature – evolutionarily stable (i.e. uninvadable), or branching-prone – of the singularities? An interesting property that one can already obtain from (28) is that a branching point of the canonical equation turns out to be a locally unstable equilibrium for the diffusion process, the dynamics of which would thus oscillate around this point. It should be possible to calculate the average time for such an oscillation, that would give an estimate of the typical time needed for significant phenotypic divergence around a branching singularity; and to compare this average time with the mean time required for escaping the basin of attraction of the branching point. Aiming at such a comparison eventually raises a question of biological interest: finding ecological conditions under which evolution is more likely to proceed through punctuated equilibria, rather than through adaptive radiation (Rand and Wilson, 1993).

## 5. Discussion

The purpose of this paper was to recast the adaptive dynamics approach to modeling mutation-selection processes in the rigorous mathematical framework of jump processes. We recall that there are two critical assumptions upon which the jump process interpretation of adaptive dynamics models relies: first, the ecological and evolutionary time-scales are separated, which implies that individuals carrying a mutated value of the trait should appear in a resident population that is at ecological equilibrium; second, an exclusion principle applies: away from special points in the trait space (the so-called branching points), any mutant phenotype goes either to extinction, or to fixation. Our mathematical formalism yields three main conclusions.

First, the family of canonical equations, descended from the case of one species and one trait to multispecific and multidimensional settings, were introduced as giving approximations to the dynamics of the mean trait value in the population over evolutionary time. This probabilistic approach shows that the canonical equations, in fact, determine the *exact* process of mutation-selection as the mutation step variance goes to zero. The approximation (6) made in Dieckmann and Law (1996) can be shown to have a speed of convergence as $\sigma_0$ goes to 0 similar to the speed of convergence as $\sigma$ goes to 0 of the processes $X_\sigma$ that we have considered. Thus, when the mutation step variance is small but not infinitesimal, the exact process governed by (7) keeps performing as well as the process describing the mean trait value.

Second, our approach naturally leads to extend the canonical equations to the case of biased mutation. This may be important for several reasons (Pomiankowski et al., 1991), two of them being structural: the measurement of any mutation bias (or the lack thereof) is scale dependent. If mutation is unbiased given one definition of a character and one scale of measurement then it must be biased for many other definitions and scales. Also, mutation bias may be trait-dependent: there may be little bias away from evolutionary singularities, but strong bias close to singularities that are extreme points of the trait set. There are genetic reasons for which the opposite may happen too. Mukai (1964) predicted that during episodes of directional selection, characters should show a negative correlation between mutation bias and the direction of selection; there may be weak mutation bias close to the adaptive optimum, but strong bias away from the optimum. Mutagenesis studies provide further evidence for the occurrence of mutation bias in quantitative traits. Pomiankowski et al. (1991) reported on the results of Lai and Mackay (1990) who have used the P–M hybrid dysgenesis system of *Drosophila melanogaster* to look at transposable element induced mutations in quantitative characters. They found that average sternopleural and abdominal bristle numbers were lower for P strain chromosomes from dysgenic crosses that suffer high mutation rates because of elevated P-element transposition. Unfortunately, although such reports suggest that mutation bias on traits exist, they do not allow an estimation of its magnitude in natural populations.

Third, our approach raises the possibility of contrasting macroevolutionary patterns dominated either by punctuated equilibria, or by radiation events. Punctuated equilibria refer to rapid transitions between evolutionary states around which a population may dwell for long periods (e.g. Stanley, 1979). Radiation occurs when the population diversifies into two or more phenotypic branches (which may correspond to speciation in sexually reproducing species) (e.g. Schluter, 2000). Our preliminary analysis of diffusion models of adaptive dynamics raises an interesting issue for systems that contain multiple branching points, namely comparing the time needed for phenotypes to diverge around branching points with the time taken by the population to jump between the basins of these branching points. Although branching may be a ubiquitous property of attractive evolutionary singularities (Doebeli and Dieckmann, 2000), there might be ecological and genetic conditions under which evolution would proceed rather through punctuated equilibria and long-term evolutionary cycles in species that yet possess multiple branching points.

## 6. Concluding remarks

Here we offer a mathematical justification of the canonical equation of adaptive dynamics, elaborating on the jump model considered by Dieckmann and Law (1996). Yet this jump process is a simplified description of the biological phenomenon

underlying Darwinian evolution. An exact model including the three main mechanisms (reproduction, mutation and selection) must be *individual-based*: each individual can die or give birth to a new individual with rates depending on the composition of the total (or local) population, and a new-born individual may be a mutant (i.e. has a trait different of its progenitor's) with some probability and according to some distribution. The description of such individual-based models has been expounded by Dieckmann and Law (1996). The challenging problem of the mathematics of adaptive dynamics is to identify the limits that can be applied to the individual-based model to recover the canonical equation, and to contrast predictions obtained from the canonical equation with those resulting from these alternative approximations.

The canonical equation of adaptive dynamics stems from three hierarchical scalings:

- First, the population size is taken to be large. Under generic assumptions on the ecological model, this ensures the existence of a stable equilibrium density $n(s)$ for a population composed only of individuals with trait $s$. By taking this limit, one moves from the individual-based model to an *infinite population model* still allowing for phenotypic diversity, and for dependence of the ecological parameters on the composition of the total population.

- Second, ecological and evolutionary time-scales are separated, which makes deleterious (or unlucky) mutants disappear before a new mutation happens. This is obtained by letting the probability of mutation $\mu$ per birth go to zero. Taking this limit requires to rescale time properly. This *rare mutation model* allows to describe the evolutionary process in the trait space $S$ only (providing that each trait value determines uniquely the equilibrium population state).

- Finally, recovering a model of the adaptive process that takes the form of an ordinary differential equation (i.e. the canonical equation) requires to take the limit where the mutation step variance $\sigma$ goes to zero.

Of course it could be possible to take these limits in different order, or simultaneously. In particular, should we postpone the infinite population limit, we would have to consider the quasi-stationary distribution (stationary distribution conditional to non-extinction of the population) of the finite population dynamics without mutation and introduce a new mutant before the population gets extinct. Considering a large (but finite) population should make it possible to control critical times of the process (i.e. how long it takes for the resident population to converge to its quasi-stationary distribution, and how long it may stay extant). This more general approach would be fruitful for at least two reasons: it would yield a new version of the canonical equation for finite populations and it could pave the way for extending the canonical equation approach to polymorphic populations.

## Acknowledgements

## References

DIECKMANN, U. and LAW, R. (1996): The dynamical theory of coevolution: A derivation from stochastic ecological processes. *J. Math. Biol.* **34**:579–612.

DOEBELI, M. and DIECKMANN, U. (2000): Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Amer. Nat.* **156**:S77–S101.

ETHIER, S. and KURTZ, T. (1986): *Markov Processes, Characterization and Convergence.* John Wiley and Sons, New York.

FREIDLIN, M. I. and WENTZEL, A. D. (1984): *Random Perturbations of Dynamical Systems.* Springer-Verlag, Berlin.

HOFBAUER, J. and SIGMUND, R. (1990): Adaptive dynamics and evolutionary stability. *Appl. Math. Lets* **3**:75–79.

KISDI, E. (1999): Evolutionary branching under asymmetric competition. *J. Theor. Biol.* **198**:149–162.

LAI, C. and MACKAY, T. (1990): Hybrid dysgenesis-induced quantitative variation the X chromosome of *Drosophila melanogaster. Genetics* **124**:627–636.

MARROW, P., LAW, R. and CANNINGS, C. (1992): The coevolution of predator-prey interactions: ESSs and Red Queen dynamics. *Proc. R. Soc. Lond. B* **250**:133–141.

METZ, J. A. J., NISBET, R. M. and GERITZ, S. A. H. (1992): How should we define 'fitness' for general ecological scenarios? *Trends Ecol. Evolut.* **7**:198–202.

METZ, J. A. J., GERITZ, S. A. H., MESZÉNA, G., JACOBS, F. A. J. and VAN HEERWAARDEN, J. S. (1996): Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In VAN STRIEN, S. J., VERDUYN LUNEL, S. M. (eds): *Stochastic and Spatial Structures of Dynamical Systems.* North Holland, Amsterdam, pp. 183–231.

MUKAI, T. (1964): Polygenic mutation affecting quantitative character of *Drosophila melanogaster*. In *Mutations in Quantitative Traits*. Proceedings Gamma Field Symposium 3, Ministry of Agriculture, Japan, pp. 13–29.

POMIANKOWSKI, A., IWASA, Y. and NEE, S. (1991): The evolution of costly mate preferences I. Fisher and biased mutation. *Evolution* **45**:1422–1430.

RAND, D. A. and WILSON, H. B. (1993): Evolutionary catastrophes, punctuated equilibria and gradualism in ecosystem evolution. *Proc. R. Soc. Lond. B* **253**:137–141.

SCHLUTER, D. (2000): *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

STANLEY, S. M. (1979): *Macroevolution: Pattern and Process*. Freeman, San Francisco, Ca.

WENTZEL, A. D. (1976a): Rough limit theorems on large deviations for Markov random processes, I. *Theory Probab. Appl.* **21**:227–242.

WENTZEL, A. D. (1976b): Rough limit theorems on large deviations for Markov random processes, II. *Theory Probab. Appl.* **21**:499–512.