George William Albert Constable

Fast Variables in Stochastic Population Dynamics

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

George William Albert Constable

Fast Variables in Stochastic Population Dynamics

Doctoral Thesis accepted by
the University of Manchester, UK

🐎 Springer

Author
Dr. George William Albert Constable
Department of Ecology
   and Evolutionary Biology
Princeton University
Princeton, NJ
USA

Supervisor
Prof. Alan McKane
School of Physics and Astronomy
The University of Manchester
Manchester
UK

Printed on acid-free paper

Parts of this thesis have been published in the following journal articles:

G. W. A. Constable, A. J. McKane and T. Rogers
Stochastic dynamics on slow manifolds
J. Phys. A: Math. Theor. Vol 46, 295002 (2013)

G. W. A. Constable and A. J. McKane
Fast-mode elimination in stochastic metapopulation models
Phys. Rev. E Vol 89, 032141 (2014)

G. W. A. Constable and A. J. McKane
Population genetics on islands connected by an arbitrary network: An analytic approach
J. Theor. Biol. Vol 358, 149–165 (2014)

G. W. A. Constable and A. J. McKane
Models of genetic grift as limiting forms of the Lotka-Volterra competition model
Phys. Rev. Lett. Vol 114, 038101 (2015)

G. W. A. Constable and A. J. McKane
Stationary solutions for metapopulation Moran models with mutation and selection
Phys. Rev. E Vol 91, 032711 (2015)

Supervisor's Foreword

It is with the greatest of pleasure that I write the foreword to George Constable's Ph.D. thesis. At one level the thesis is a very clear description of a method of analysing models in population genetics, which I will outline in more detail below. But on a more general level it should also be very useful as an introduction to those wishing to understand the formalism of continuous-time stochastic processes. There are not so many textbooks on this subject, and those which do exist can be technically quite forbidding. This need not be so—the essential ideas and techniques are relatively straightforward to understand—and the opening chapters of this thesis provide an accessible introduction to them.

The main subject of the thesis is the stochastic time-evolution of populations when the individuals making up the population have a very simple genetic make-up: they are haploid and the focus is on a single gene which has only two alleles. The ideas can certainly be extended to more complex systems, but the aim was to develop techniques to allow models to be analysed, and testing them on the simplest situations to start with. Attention was also directed to aspects of the models which have perhaps received less attention than they deserve.

The first of these aspects is the careful specification of the model and its subsequent simplification through the use of a diffusion approximation. Theoretical physicists tend to take great care in distinguishing between microscopic, mesoscopic and macroscopic descriptions of the same system, and the specification of the approximations that are made to go from one level of description to another. There is less of a tradition of doing this in the context of biological systems. In this thesis there is a careful separation of the modelling and approximation processes, so that both the starting point and the nature of the approximations subsequently made are absolutely clear.

The second, and more substantial, contribution described in the thesis is in the development of an additional approximation which makes an intractable equation amenable to analysis. The standard method of specifying a model after the diffusion approximation has been made goes back to the work of Fisher, but was popularised by Kimura in the 1950s. It takes the form of a partial differential equation known as

a Fokker-Planck equation to physicists and as a Kolmogorov forward equation to those in many other disciplines. In simple situations it can be analysed but if, for instance, the population is subdivided into islands (or demes, as they are sometimes called), then this is a partial differential equation in many variables, and as such any in-depth analysis appears to be almost impossible.

As a consequence in some areas of population genetics, the general equations governing the dynamics have not been addressed directly, because of their intractability, and the focus has moved to simpler systems or to those where progress could be made. This is understandable, but what the work in this thesis shows is that this is not necessary; a simple approximation is available which reduces a many-island description to one which is effectively a one-island model, but with effective parameters. The resulting Fokker-Planck equation can now be used to calculate the probability that a particular allele becomes fixed, how long this will take on average or what the nature of the stationary probability distribution of alleles is. There had been a few previous attempts to do this, but the procedure outlined here is both more general and easier to understand than previous studies.

The procedure is based on a fast-mode elimination technique. The idea is very simple and long established in the theory of dynamical systems. Essentially, the variables in the model are decomposed into a set which decay at different rates, $|\lambda^{(i)}|^{-1}$, $i = 1, 2, \ldots$. It turns out that in many cases of interest there is a 'gap' between the smallest $|\lambda^{(i)}|$ (taken to be $|\lambda^{(1)}|$) and all the others. This means that after quite a short time, compared to the timescales of interest to us, only the mode characterised by $|\lambda^{(1)}|$ is left in the model describing the system. This is the 'slow' mode—all the other 'fast' modes have decayed away and dropped out of the theoretical description. What is left is a model with just one degree of freedom, which can then be analysed systematically.

Of course, although the idea just described seems simple enough, finding a concrete procedure which works, and which can be turned into a calculational tool, is not. However, remarkably, a method was found which is both rather straightforward to apply and which also gives results in excellent agreement with computer simulations of the original individual-based (that is, microscopic) model. It results in a 'reduced' model in which the parameters are given in terms of the island sizes, the scale of migration between the islands, or whatever parameters were present in the original (full) model. Although I have used the example of subdivided populations to illustrate the method, it should be more generally applicable to the reduction of complex population genetics models down to much simpler ones with just a few effective parameters which are explicitly given in terms of those of the full model.

I have already mentioned that those looking for an easy-to-understand introduction to the formalism of continuous-time stochastic processes would benefit from reading this thesis. But I would also hope that it would appeal to theoretical biologists seeking to extend the scope of problems it can be applied to, to mathematicians wanting to make the approach more rigorous, and to theoretical physicists looking for an application of the ideas and techniques of non-equilibrium

statistical mechanics. So my hope is that the publication of this thesis will allow a much wider range of people to appreciate the power of the methodology presented here, and also enable them to contribute to extending its range of applicability.

Manchester   Alan McKane
April 2015

Abstract

In this thesis, I present two methods of fast variable elimination in stochastic systems. Their application to models of population dynamics from ecology, epidemiology and population genetics, is explored. In each application, care is taken to develop the models at the microscale, in terms of interactions between individuals. Such an approach leads to well-defined stochastic systems for finite population sizes. These systems are then approximated at the mesoscale, and expressed as stochastic differential equations. It is in this setting that elimination techniques are developed.

In each model a deterministically stable state is assumed to exist, about which the system is linearised. The eigenvalues of the system's Jacobian are used to identify the existence of a separation of timescales. The fast and slow directions are then given locally by the associated eigenvectors. These are used as approximations for the fast and slow directions in the full nonlinear system. The general aim is then to remove these fast degrees of freedom and thus arrive at an approximate, reduced-variable description of the dynamics on a slow subspace of the full system.

In the first of the methods introduced, the conditioning method, the noise of the system is constrained so that it cannot leave the slow subspace. The technique is applied to an ecological model and a susceptible-exposed-infectious-recovered epidemiological model, in both instances providing a reduced system which preserves the behaviour of the full model to high precision.

The second method is referred to as the projection matrix method. It isolates the components of the noise on the slow subspace to provide its reduced description. The method is applied to a generalised Moran model of population genetics on islands, between which there is migration. The model is successfully reduced from a system in as many variables as there are islands, to an effective description in a single variable. The same methodology is later applied to the Lotka-Volterra competition model, which is found under certain conditions to behave as a Moran model. In both cases the agreement between the reduced system and stochastic simulations of the full model is excellent.

It is emphasised that the ideas behind both the conditioning and projection matrix methods are simple, their application systematic, and the results in very good agreement with simulations for a range of parameter values. When the methods are compared, however, the projection matrix method is found in general to provide better results.

Acknowledgments

Writing this thesis has been a surprisingly enjoyable experience. It would not have been so without the help and support of those around me.

I would like to offer my most sincere gratitude to my supervisor, Alan McKane, whose advice on all matters has been invaluable, and whose company I have thoroughly enjoyed.

I thank Tim Rogers, who by virtue of infectious enthusiasm, tempered by bouts of cynicism, is a great guy to work with.

Over the past three years I have been lucky enough to be surrounded by people who are happy to take time out of their day to talk about ideas and clarify thoughts. In this regard I am grateful to Tobias, Diana, Louise and Luis, as well as the many who have shared Office 7.26—Alex, Joe, Tommaso, James, Toby, Peter and Cesar.

The funding for this Ph.D. has been provided by a University of Manchester Engineering and Physical Sciences scholarship, and so I thank the EPS for their generosity.

Finally, I am indebted to my family, which have been unfailingly supportive—David and Julie, Deborah and June.

Contents

Chapter 1
Introduction

Essentially, all models are wrong, but some are useful.
George Box [3]

In what follows I am going to explore how concepts and mathematical tools originally
developed within physics can be applied to a variety of other fields. These can include,
but are not limited to, population genetics, evolution, opinion dynamics, epidemiol-
ogy and ecology. This thesis will focus primarily on models with an interpretation in
population genetics, however models with an ecological and epidemiological flavour
will also be explored. With this in mind, let us begin by discussing the questions,
'what do we mean by a model?' and 'what makes a good model?'. The answers to
these questions are by no means unarguable, but rather serve to give the reader an
impression of the philosophy to which I attempt to adhere.

A model, in its most general form, is a representation of the real world used to help
better comprehend or predict its behaviour. In order to gain any tractability, these
models feature some degree of abstraction from the real world; the most accurate
model would have a one-to-one correspondence with its real world counterpart, but a
replica is no more easy to understand than the original. The degree of abstraction is a
modelling choice which, to some degree, depends both on the aim of the model and
the degree of knowledge one has about the real system. A large degree of abstraction
is preferable if knowledge of the real system is limited, or if one wishes the model to
be very general. For these reasons, I will be concerned with the quantitative analysis
of abstract models which aim to give a qualitative understanding of the behaviour of
certain systems.

This thesis will deal exclusively with dynamic models which describe the evolu-
tion in time of a system which is characterised by a set of variables. If at some time
t_0 the state of the system is described by a vector of variables $\boldsymbol{x}(t_0)$, we ask at some
later time $t_0 + \Delta t$ how the properties of $\boldsymbol{x}(t)$ have changed. The model is said to be
deterministic if the state of the system at $t_0 + \Delta t$ can be determined precisely. If the
system is in state $\boldsymbol{x}(t_0)$ at time t_0, it will be in a calculable state $\boldsymbol{x}(t_0 + \Delta t)$ at time

$t_0 + \Delta t$. Such systems are most frequently described by a set of ordinary differential equations (ODEs) of the following form

$$\frac{d}{dt}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}), \tag{1.1}$$

where the vector-valued function $\boldsymbol{f}(\boldsymbol{x})$, along with initial conditions $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$, defines the model.

Within physics, the principle of Occam's razor is ubiquitous in theoretical work. This philosophy perhaps has its origins in the fact that the physical world has, historically, been found to obey laws which adhere so closely to their model counterparts as to make the line between a model and reality appear blurred. However there is also a more pragmatic factor in constructing simple models, that of gaining an intuitive grasp on how a system behaves. As eluded to earlier, though a model with more variables may provide a more accurate representation of its subject, it will also be harder to understand intuitively. The apotheosis of the interpretability of a mathematical problem comes in the form of an analytic solution. While a far greater range of mathematical problems can be tackled numerically or through simulation than analytically, such methods cannot rival the encompassing power of a single equation that can describe some behaviour of a system.

While physicists in general still hold to an aesthetic desire for simplicity, the last two hundred years has seen an undermining of the belief that simple models are sufficient. Among the issues at the core of this subversion have been the practical intractability and indeterminacy of many variable systems, chaos theory and the realisation that emergent phenomena can defy naïve interpretation. Together, these issues in some sense embody the class of problems at the centre of the discipline known as complex systems.

The first of these problems is encountered while trying to describe the thermodynamic behaviour of gas in terms of its constituent particles, and was tackled by an approach which came to be known as statistical mechanics [10]. If one wishes to understand the behaviour of a gas, one may consider the behaviour of each particle in the gas independently. However, even if we know the exact form of the interaction between the particles and identify the initial conditions of each particle in the system, we are still left with a formidably complicated system of equations to solve. The key to progress is in observing that we are not really interested in all the details that solving the full system would give us; even if we could solve the full set of equations, the velocity of the 13, 500, 303, 304th particle is superfluous information that gives us no understanding as to how the collection of particles behaves. If instead we assume each of the particles is indistinguishable from any other, we can begin to make statistical inferences about the system. We are then interested in the statistics of the mass of particles, about which we can make some analytic progress.

The second of the above problems is chaos. The central, powerful observation of chaos theory is that there exists simple systems (possibly of the form of Eq. (1.1) in three or more variables) whose behaviour is made essentially unpredictable by a limited knowledge of the initial conditions [9]. In a chaotic system, two trajectories

whose initial conditions lie very near one another will evolve in time in an entirely different manner. In order to accurately predict the trajectory a system would follow, one would need to know the initial conditions of the system to infinite precision. Clearly this nullifies the predictive power of the model. In this thesis I will not deal with this class of system. However, the indeterminacy that arises from this phenomena is important in motivating the work that follows, in that it highlights the failing of determinism and points instead towards models where predictability is inherently limited.

A stochastic model is one which incorporates inherent unpredictability. In contrast to a deterministic model, if the state of the system, $x(t)$, is known at some time t_0, we do not know the state of the system at $x(t_0 + \Delta t)$. The variable $x(t)$ is a realisation of a stochastic variable X whose evolution in time cannot be predicted precisely. The exact way in which systems with stochastic components are modelled is once again a matter of choice. In this thesis however, I will concentrate on models which describe the time-evolution of the probability density function (PDF) of the stochastic variable. The PDF gives the probability that a system is in a state $x(t)$. If the system is in state $x(t_0)$ at t_0, rather than ask what state the system is in at time $t + \Delta t$ as in the deterministic description, given by Eq. (1.1), we ask what the probability of being in some state x is at time $t + \Delta t$. However, before we proceed to discuss in full the details of a dynamic-stochastic model, let us review some important results and intuitions from probability theory.

Say a fair, six sided die is thrown in the air. In principle, given enough information about the initial state of the die (its trajectory, weight, alignment etc.) one could calculate on exactly what side the die would fall. However, rather than model all of these parameters in a complicated model, it is common to embody all the dynamical processes in one probabilistic process. With each throw of the die, rather than determine a definite answer, we allow an equal probability of each of the six results occurring. What do these probabilities indicate? If we were to take a frequentist approach, they tells us that if we throw the die an infinite number of times, we would get each side of the die exactly 1/6th of the time [4]. Letting n_x denote the number of times an event x occurs, and N the total number of trials, the probability $p(x)$ of x occurring is then

$$p(x) = \lim_{N \to \infty} \frac{n_x}{N}. \tag{1.2}$$

Equivalently, we could say that if we threw an infinite number of dice, then 1/6th of them would show each side.

What happens if, quite reasonably, we do not have an infinite collection of dice? The expected probability of rolling a 3 is 1/6. However, given a finite number of dice N, one may not (and possibly cannot) achieve this fraction of threes. Instead, one obtains a sample probability,

$$\bar{p}(x) = \frac{n_x}{N}, \tag{1.3}$$

where the bar indicates that the probability is estimated from a finite sample. The sample statistics themselves can be understood as coming from an underlying distribution of potential outcomes. For a large enough sample size, N, one can show via the central limit theorem [8] that $\bar{p}(x)$ is drawn from an approximately normal distribution whose mean is $p(x)$, and whose standard deviation is proportional to $1/\sqrt{N}$. As N tends to infinity, the distribution of samples therefore tends to a delta function centred on the real probability.

This dice example is extremely simple, featuring no statistical behaviour that evolves in time. However it does illustrate in a very intuitive way two important points. The first is that while this modelling approach (replacing mechanics for probabilities) seems like a drastic simplification, it is in fact a rather successful approximation (as anyone who has played Monopoly can attest). The second is that, when dealing with probabilistic systems, an observation or sample of a finite system will only ever be a single realisation of an array of outcomes. In other words, there is a noise inherent in finite systems which scales in size approximately like the inverse square root of the sample size, $N^{-1/2}$. Such noise is termed intrinsic noise and is the focus of this thesis.

Let us now consider a biologically motivated dynamical system in which these statistical properties are exhibited; a growing population. We begin with the natural assumption that the typical birth and death rates are proportional to the size of the population, denoted by the continuous variable x, with the parameters λ and μ denoting the per capita birth and death rates respectively. A seemingly natural way to model this system is deterministically via the ODE

$$\frac{dx}{dt} = (\lambda - \mu)x, \tag{1.4}$$

which has the familiar solution

$$x = x_0 e^{(\lambda - \mu)t}, \tag{1.5}$$

where x_0 is the initial size of the population. The population grows exponentially if the birth rate is larger than the death rate, and decays exponentially if the converse is true. All in all this seems like a very reasonable solution. However, in the real world we know that reproduction is not so simple a process. There may be many additional factors affecting whether an individual reproduces, such as environment or physical condition. In addition, a population of organisms in nature is a discrete quantity whereas the population growth predicted by Eq. (1.5) is continuous. We therefore choose to model the system in a more sophisticated way, replacing this continuous population size x with a discrete population size n and redefining λ and μ as the probability per unit time that a single individual reproduces. In this way, as with the dice example, we bundle all the mechanics and detail of the problem into a probabilistic interpretation. We can now describe the probability at some time of finding the system in a discrete state n $[0, \infty)$. It can be shown to be given by [10]

$$P_1(n, t) = \frac{1}{n!} \left[\frac{d^n}{dz^n} \left(\frac{\mu(z-1)e^{(\lambda-\mu)t} - \lambda z + \mu}{\lambda(z-1)e^{(\lambda-\mu)t} - \lambda z + \mu} \right)^{n_0} \right] \Bigg|_{z=0} . \tag{1.6}$$

How can the difference between the results given by Eqs. (1.5) and (1.6) be resolved? Well, taking the mean of Eq. (1.6) one can show

$$\langle n \rangle = n_0 e^{(\lambda-\mu)t}, \tag{1.7}$$

where n_0 is the initial (discrete) size of the population. The equivalence of Eq. (1.5) and Eq. (1.7) gives a clear interpretation of the deterministic description; the deterministic trajectory is the average trajectory of an infinite number of realisations of an equivalent stochastic process (see Fig. 1.1). We can also calculate the coefficient of variation for the distribution (the standard deviation normalised by the mean), obtaining

$$\frac{\sqrt{\langle n^2 \rangle - \langle n \rangle^2}}{\langle n \rangle} = \frac{1}{\sqrt{n_0}} \sqrt{\frac{\lambda+\mu}{\lambda-\mu}} \sqrt{1 - e^{-(\lambda-\mu)t}}, \tag{1.8}$$

showing that the relative error scales with the inverse root of the initial size of the population sample. This situation is not entirely dissimilar to the dice example, in which we stated that the bigger the sample, the less deviation there was from the mean (deterministic) behaviour. This is illustrated in Fig. 1.1 and will be shown more generally in the body of this thesis.

Let us close the discussion of this model by addressing an important question; what does the stochastic description in Eq. (1.6), with all its complications, really give us over the naïve deterministic description, Eq. (1.5)? The answer is that it allows for fundamentally different behaviour. Take the example $\lambda > \mu$. The deterministic theory predicts that the population strictly increases, however the stochastic
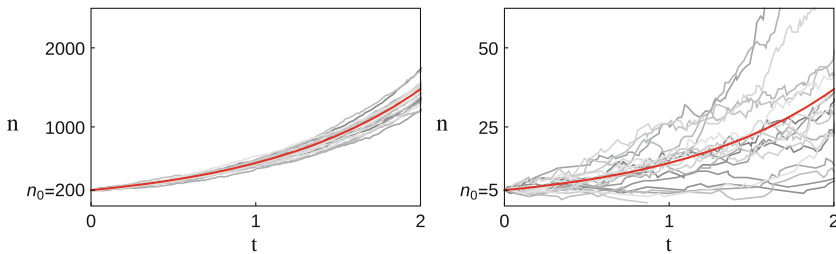

Fig. 1.1 Stochastic simulations of a growing population in *grey* overlaid with the mean deterministic behaviour given by Eq. (1.7) in *red*. In both plots $\lambda = 2$, $\mu = 1$. Comparison of the figures shows that the system appears more noisy with a smaller initial condition, n_0, consistent with Eq. (1.8)

description allows for the possibility that the finite population will become extinct. Setting $n = 0$ in Eq. (1.6), we find that for long times the probability of extinction is approximately $(\mu/\lambda)^{n_0}$. For a small initial population, this is a very relevant effect which the deterministic theory fails to capture.

When developing dynamic models in physics, this noise arising from finite sampling is typically a special case since the systems under consideration are often so large as to be assumed infinite. We note that the preceding model with $\lambda = 0$ can be used to study the decay of a radioactive substance, for which the deterministic description given by Eq. (1.5) is a standard undergraduate result. If one is modelling systems which are relatively small, the effect of intrinsic noise is non-negligible. This is typically true when building models of populations of discrete organisms in a biological or social context. In this setting, the intrinsic noise resulting from the discrete nature of the system's constituents is also referred to as demographic noise or demographic stochasticity [2].

Some of the first stochastic models of populations were constructed by Sewall Wright and Ronald Fisher with the aim of investigating the process of genetic drift [6, 13]. Genetic drift is the process by which the genetic composition of a population changes due solely to random reproduction events [5]. The stochastic model attributed to them, the Wright-Fisher model, in many ways exemplifies how a simple model can lead to a qualitative but informative insights into how a system might behave. The Moran model is very closely related to the Wright-Fisher model and indeed can be understood as the Wright-Fisher model with overlapping generations in continuous time. Since a model of this ilk will be addressed in the body of the thesis, it is perhaps useful to review its general formulation.

The model consists of a finite population of two types of individual, denoted type A and B. Within each type, the individuals are assumed to be indistinguishable from one another. The population is also assumed to be well-mixed; that is, it has no spatial structure and every individual has an equal probability of interacting with every other. In this respect the connection to statistical thermodynamics is clear. The population is fixed such that at all times the total number of individuals of both types is N. If the number of individuals of type A is denoted n, the state of the system is then described by this single state variable since the number of type B is simply $N - n$. In the discrete-time model, at each point in time two individuals are chosen from the population, one to reproduce (an exact copy of this individual is made) and one to die (it is removed from the population) so that the population size is fixed. The probability of either of these being type A or B is simply the proportion of each in the population (see Fig. 1.2). In continuous time, $T(n'|n)$ is defined as the probability per unit time that the system will move from a state n to some other state n'. For this model there are only two processes that can change the constitution of the population;
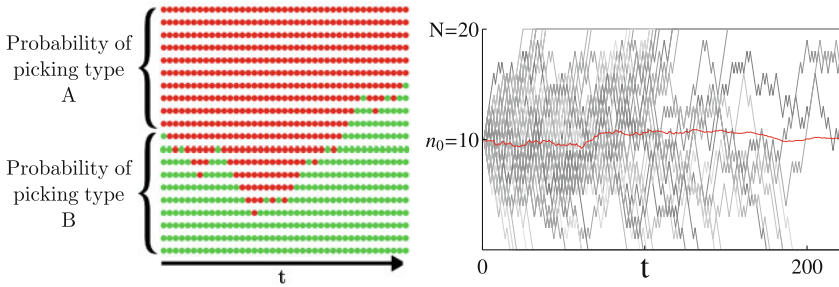
Fig. 1.2 *Left* Realisation of the discrete-time Moran model for individuals of types *A* and *B*. *Right* results from 50 realisations plotted in *grey*, while their mean is plotted in *red*. The mean can be seen to stay approximately at its initial value

$$
\overbrace{T(n+1|n) = \left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)}^{\text{A born/B dies}}
$$
$$
\underbrace{T(n-1|n) = \left(\frac{N-n}{N}\right)\left(\frac{n}{N-1}\right)}_{\text{B born/A dies}} .
\tag{1.9}
$$

Such a model is referred to as a neutral model in population genetics, since neither type is favoured over the other [11]. While this model cannot be easily solved to give a solution of the form of Eq. (1.6), certain properties of the distribution can be straightforwardly calculated. Most interestingly, the mean of the distribution is given by

$$
\langle n \rangle = n_0.
\tag{1.10}
$$

It does not vary in time, but simply maintains its initial value. This is because for any given state, the probability of type *A* increasing is the same as of it decreasing; one picks *A* to reproduce and *B* to die with the same probability of picking *B* to reproduce and *A* to die. The ensemble average thus remains unchanged so that a deterministic description would predict no dynamics at all. Simulation of the stochastic process reveals of course that the system does evolve in time, and one can intuitively see that at infinite times a particular realisation of the system dynamics must end in either state $n = 0$ or $n = N$, from which it cannot return. Here, a deterministic description not only fails to capture some, but *all* of the dynamical behaviour. In problems such as this a stochastic treatment is clearly essential.

The stochastic nature of the Moran model adds to its complexity, however it remains a one-dimensional system over which one can gain some analytic tractability. What happens if the model we wish to deal with is more complicated? It is possibly only a slight exaggeration to say that of all the mathematical models we can dream of,

there are only two kinds which are straightforward to solve: those which are linear, and those which are one-dimensional. This aphorism holds equally for stochastic dynamical systems as it does for their deterministic counterparts.

Given a nonlinear set of coupled ODEs which are not amenable to analysis, we must employ some approximation in order to make analytic progress. If a separation of timescales exists, it is sometimes possible to extricate the dynamics happening on a fast timescale from those occurring on a slow timescale. We can thus achieve a reduction in the dimensionality of the system, which may make an analysis possible or simply improve our intuition as to how the system behaves. Such a method is best illustrated with a simple deterministic example.

Say one wishes to model the motion of a particle in a potential, $U(x) = ax^2 + bx^4$, with some viscous force βv. The potential is a double-well for $a < 0$, $b > 0$. Such a system can be modelled as a set of ODEs of type (1.1);

$$\frac{dx}{dt} = v,$$
$$\frac{dv}{dt} = -\beta v - (2ax + 4bx^3), \tag{1.11}$$

where the mass has been set equal to one. The system is nonlinear and does not appear to be analytically intractable. Rescaling time such that $\tau = \beta t$, the second equation becomes

$$\frac{dv}{d\tau} = -v - \frac{1}{\beta}(2ax + 4bx^3). \tag{1.12}$$

In this form it is clear that if β is large, the second equation evolves on a much faster timescale than that for dx/dt. We therefore assume that v reaches its stationary value long before x, i.e. we set $dv/dt = 0$, solve for v as a function of x and substitute this into the equation dx/dt. We find

$$v = -\frac{1}{\beta}(2ax + 4bx^3), \tag{1.13}$$

and therefore the approximate equation for the time evolution of x is

$$\frac{dx}{dt} = -\frac{1}{\beta}(2ax + 4bx^3). \tag{1.14}$$

Unlike Eq. (1.11), this can be solved analytically. We find the slightly complicated but exact expression

$$x(t) = \pm\sqrt{a\left[\exp(4at\beta^{-1})\left(ax_0^{-2} + 2b\right) - 2b\right]^{-1}}, \tag{1.15}$$
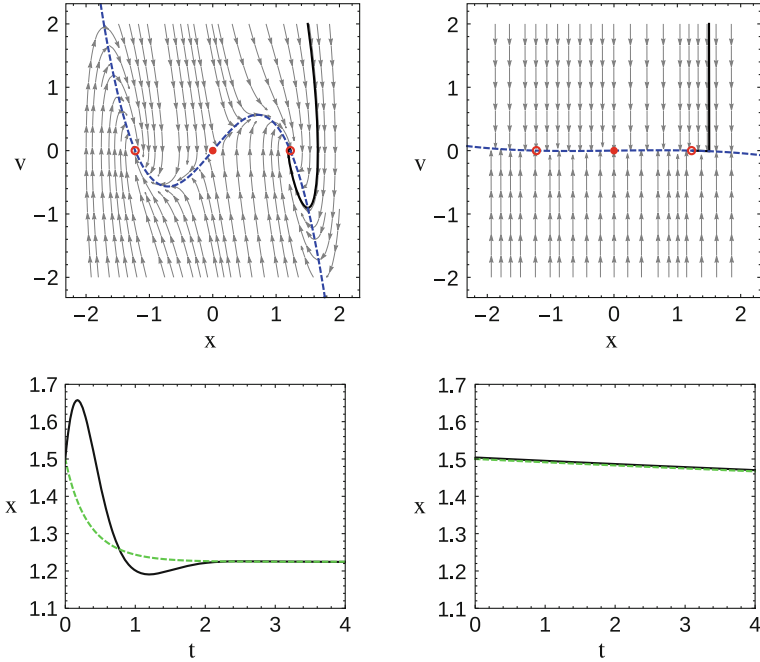
Fig. 1.3 *Top panels* potential trajectories of the system (1.11) plotted in *grey*. A specific trajectory is plotted in *black*, while the slow subspace defined by Eq. (1.13) is plotted as a *blue dashed line*. Filled/empty *circles* represent stable/unstable points in the system. *Bottom panels* Trajectory of Eq. (1.11) plotted in *black*, while the trajectory of the reduced system (1.14) is plotted as a *green dashed line*. In all plots parameters are $a = -3, b = 1$. In *left panels* $\beta = 5$, in *right panels* $\beta = 500$

where $x_0 = x(t_0)$. Some illustrative plots are given in Fig. 1.3. While the reduced system gives a poor approximation of the full dynamics for small values of β (left panels), it provides an excellent approximation for large β (right panels). The trajectory in the later case moves very quickly towards the slow subspace defined by Eq. (1.13), along which it then moves at a much slower rate.

In stochastic systems, we will find that the process of fast-variable elimination is not so straightforward. The goal of removing fast degrees of freedom from stochastic systems has received significant attention and a variety of approximation methods have previously been proposed (see, for example [1, 7, 12]), though a simple and generally applicable theory is yet to emerge.

In this thesis, I will introduce two novel methods for reducing the dimensionality of stochastic systems exhibiting a separation of timescales, as well as discussing their relation to one another. These methods will then be applied to solve some specific problems.

The order of the thesis is then as follows. In Chap. 2, I will introduce the fundamental concepts and mathematical tools which will be employed in the remainder of the thesis. In Chap. 3 the first method for stochastic fast-variable elimination will

be introduced. I will refer to this method as conditioning, for reasons which will become apparent. The method will initially be applied to a toy model in Sect. 3.2 for illustrative purposes, before being generalised and applied to an epidemiological model.

Chapter 4 sees the introduction of a second method of stochastic fast variable elimination, the projection matrix method. Of particular interest is the application of the method to a Moran model with migration between islands. I find that the full model, in as many variables as there are islands, can be reduced to an effective one-dimensional model. The consequences of the population structure are discussed from a biological perspective in Chap. 5.

Finally, in Chap. 6, I report on my most recent work involving stochastic fast-variable elimination. In Sect. 6.1, mutation is incorporated into the Moran model with migration introduced in Chap. 4. In Sect. 6.2, the conditioning and projection matrix methods are compared directly. In Sect. 6.3, a model of genetic drift is considered in which the total population size is not fixed. Once again, the stochastic fast-variable elimination techniques developed in the earlier part of the thesis prove valuable.

References

1. L. Arnold, P. Imkeller, Normal forms for stochastic differential equations. Probab. Theory Relat. Fields **110**, 559–588 (1998)
2. A.J. Black, A.J. McKane, Stochastic formulation of ecological models and their applications. Trends Ecol. Evol. **27**, 337–345 (2012)
3. G.E.P. Box, N.R. Draper, *Empirical Model-Building and Response Surfaces* (Wiley, New York, 1987)
4. M.G. Bulmer, *Principles of Statistics* (Dover, New York, 1979)
5. W.J. Ewens, *Mathematical Population Genetics*, 2nd edn. (Springer, Berlin, 2004)
6. R.A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930)
7. C.W. Gardiner, Adiabatic elimination in stochastic systems. I. Formulation of methods and application to few-variable systems. Phys. Rev. A **29**, 2814–2822 (1984)
8. C.W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2009)
9. A.V. Holden, *Chaos* (Manchester University Press, Manchester, 1986)
10. L.E. Reichl, *A Modern Course in Statistical Physics* (Wiley, New York, 1998)
11. S.H. Rice, *Evolutionary Theory* (Sinauer Associates, Sunderland, 2004)
12. G. Schöner, H. Haken, The slaving principle for Stratanovich stochastic differential equations. Z. Phys. B **63**, 493–504 (1986)
13. S. Wright, Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)

Chapter 2
Technical Background

In this chapter, I will review some of the fundamental ideas and mathematical tools employed in the analysis of stochastic systems. Throughout, the discussion will be concerned with Markov stochastic processes. This class of processes will be introduced more formally in the following section, but for now they are simply described as random processes without memory. We will see that the probability of a system existing in a certain state may be described by a PDF which evolves in time according to an equation known as the master equation. The master equation is a set of partial difference equations and is in general completely intractable. It will be shown that the situation can be greatly simplified however if the state variable is assumed to be continuous.

Formally the approximation introduced is akin to the Kramers-Moyal expansion, which allows the master equation to be approximated by a Fokker-Planck equation (FPE). The FPE is a partial differential equation (PDE) in as many variables as there are distinct types of individual in the original system. In most cases it is far more tractable than the original master equation. Some useful manipulations of the FPE are therefore introduced in Sect. 2.5.

While useful as a calculational tool, the FPE formalism is not very intuitive. Since intuition will play an important role in developing the fast-variable approximation schema in Chaps. 3 and 4, an alternative but equivalent formalism is introduced, that of the stochastic differential equation (SDE). Naïvely, these can be interpreted as ODEs of type Eq. (1.1) but with the addition of some small random noise.

The final sections of this chapter will then concentrate on a review of some essential tools from deterministic dynamical systems theory and their generalisation to stochastic systems. The fundamental problems underlying fast-variable elimination in stochastic systems can then be addressed.

In order to make the discussion less abstract, I will make reference throughout to the concrete example of a Moran model. While a neutral Moran model was discussed in the introduction, for various pedagogical reasons the Moran model with mutation will serve as a preferable illustrative example.

2.1 Stochastic Processes and the Markov Assumption

To begin, let us consider a single stochastic variable X which evolves in time. The measured value of X at some point in time is the non-negative integer n. This defines the state of the one-dimensional system. As stated in the introduction, in a stochastic process the state of the system at some time is not necessarily known with certainty. Therefore rather than consider the time evolution of the state of the system, we must consider the time evolution of the PDF of states in the system.

Say the probability of being in state n at time t is given by the PDF $P_1(n, t)$. The probability of following a specific trajectory in discrete time, beginning at a point n_1 at t_1, moving to point n_2 at t_2 and so on over k steps until point n_k at t_k is a joint PDF which can be denoted

$$P_k(n_k, t_k; n_{k-1}, t_{k-1}; \ldots; n_1, t_1). \tag{2.1}$$

The process can be broken down by introducing the conditional probability distribution, $P_{k|l}(n_{k+l}, t_{k+l}; \ldots; n_{l+1}, t_{l+1}|n_l, t_l; \ldots; n_1, t_1)$. This is the probability of following the trajectory n_{l+1} to n_{k+l} given that the trajectory n_1 to n_l was first followed. The two distributions are related via Bayes' theorem [49], which states

$$\begin{aligned} &P_{k|l}(n_{k+l}, t_{k+l}; \ldots; n_{l+1}, t_{l+1}|n_l, t_l; \ldots; n_1, t_1) \\ &= \frac{P_{k+l}(n_{k+l}, t_{k+l}; \ldots; n_l, t_l; \ldots; n_1, t_1)}{P_l(n_l, t_l; \ldots; n_1, t_1)}. \end{aligned} \tag{2.2}$$

This allows the PDF to be separated into a hierarchy.

A Markov process is one in which the probability of moving from state n_{k-1} to state n_k depends only on the state of the system at t_{k-1} and not on the state of the system at any previous times. This amounts to the conditional PDF having the following property [37],

$$P_{1|k-1}(n_k, t_k|n_{k-1}, t_{k-1}; \ldots; n_1, t_1) = P_{1|1}(n_k, t_k|n_{k-1}, t_{k-1}). \tag{2.3}$$

This is the mathematical meaning of the Markov processes being memoryless. Substituting this into Eq. (2.2) and rearranging, we arrive at

$$P_2(n_k, t_k; n_{k-1}, t_{k-1}) = P_{1|1}(n_k, t_k|n_{k-1}, t_{k-1}) P_1(n_{k-1}, t_{k-1}). \tag{2.4}$$

The Markov process is fully determined by the functions $P_{1|1}(n_k, t_k|n_{k-1}, t_{k-1})$ and $P_1(n_k, t_k)$, since, for any $l > 2$, one can write

$$P_l(n_l, t_l; \ldots; n_1, t_1) = P_1(n_1, t_1) \prod_{k=2}^{l} P_{1|1}(n_k, t_k|n_{k-1}, t_{k-1}). \tag{2.5}$$

The function $P_{1|1}(n_k, t_k | n_{k-1}, t_{k-1})$ will be referred to as the transition probability.

The final step in completing our initial outlook on Markov processes is in noting two summation relations. The first is rather obvious and is simply

$$P_1(n_2, t_2) = \sum_{n_1} P_{1|1}(n_2, t_2 | n_1, t_1) P_1(n_1, t_1). \tag{2.6}$$

The second relation comes from first taking Eq. (2.5) with $l = 3$ and summing over n_2,

$$P_2(n_3, t_3; n_1, t_1) = P_1(n_1, t_1) \sum_{n_2} P_{1|1}(n_3, t_3 | n_2, t_2) P_{1|1}(n_2, t_2 | n_1, t_1). \tag{2.7}$$

Dividing both sides by $P_1(n_1, t_1)$, one can then substitute for the left-hand side using Eq. (2.2) to obtain

$$P_{1|1}(n_3, t_3 | n_1, t_1) = \sum_{n_2} P_{1|1}(n_3, t_3 | n_2, t_2) P_{1|1}(n_2, t_2 | n_1, t_1). \tag{2.8}$$

This equation is a discrete form of the Chapman-Kolmogorov equation [38].

As stated earlier, functions P_1 and $P_{1|1}$ fully determine the Markov stochastic process. However, these functions must obey Eqs. (2.6) and (2.8) in order to describe a Markov process. The fact that Markov processes obey these convenient relationships makes their analysis far more manageable than for their non-Markovian counterparts. In part, this amenability explains their prevalence in stochastic modelling. There are however many examples of processes that are non-Markovian. For instance, successive draws of coins without replacement from a purse containing a finite number of coins in different denominations is a non-Markovian process; selecting a particular denomination at some time decreases the probability of again picking that denomination at each successive time. In this case, the Markovian assumption would clearly be a very bad modelling choice. However, it is Markov processes which we will be concerned with for the rest of the thesis.

2.2 The Master Equation and the Limit of Continuous Time

In the previous section, the fundamental properties of a subset of stochastic processes, Markov processes, were introduced. Let us now restrict our attention slightly further to processes which are also homogeneous. A process is homogeneous if the transition probability between any two given states at any two times is only dependent on the time interval between those times [49]. That is

$$P_{1|1}(n', t + \Delta t | n, t) = W_{\Delta t}(n' | n). \tag{2.9}$$

We now wish to simplify the Chapman-Kolmogorov equation, (2.8), by looking at its infinitesimal time-evolution. This will allow the system's time-evolution to be described by a set of differential equations. Let the system initially be at some state n_0 at t_0. We wish to know the conditional probability that the system is in state n some small time Δt later. This is given by the Chapman-Kolmogorov equation, Eq. (2.8), which in this new notation reads

$$P_{1|1}(n, t + \Delta t | n_0, t_0) = \sum_{n'} P_{1|1}(n, t + \Delta t | n', t) P_{1|1}(n', t | n_0, t_0). \quad (2.10)$$

If the process is homogeneous we can substitute in Eq. (2.9) to arrive at

$$P_{1|1}(n, t + \Delta t | n_0, t_0) = \sum_{n'} W_{\Delta t}(n | n') P_{1|1}(n', t | n_0, t_0). \quad (2.11)$$

The assumption is now made [15] that for small Δt the transition probability $W_{\Delta t}(n | n')$ can be written

$$W_{\Delta t}(n | n') = \begin{cases} T(n | n') \Delta t + \mathcal{O}(\Delta t)^2 & \text{if } n \neq n' \\ 1 - \sum_{n \neq n''} T(n | n'') \Delta t + \mathcal{O}(\Delta t)^2 & \text{if } n = n'. \end{cases} \quad (2.12)$$

Essentially this assumes that in a small time interval, the probability of changing state is to first order proportional to the time interval. Meanwhile the probability of not changing state is much larger, so that as the time interval tends to zero there is no probability that the system can change state. The function $T(n | n')$ is only defined for $n \neq n'$ and is called the probability transition rate, or alternatively just the transition rate. As Eq. (2.12) suggests, $T(n | n')$ can be interpreted as the probability per unit time of a transition from state n' to a state n. Choosing to define the transition probability in this manner ensures that, at least to order Δt, the transition probability remains appropriately normalised; $\sum_n W_{\Delta t}(n | n') = 1$.

Substituting Eq. (2.12) into Eq. (2.11) and rearranging, one finds

$$\frac{P_{1|1}(n, t + \Delta t | n_0, t_0) - P_{1|1}(n, t | n_0, t_0)}{\Delta t} = \quad (2.13)$$

$$\sum_{n \neq n'} \left[T(n | n') P_{1|1}(n', t | n_0, t_0) - T(n' | n) P_{1|1}(n, t | n_0, t_0) \right] + \mathcal{O}(\Delta t). \quad (2.14)$$

Letting $\Delta t \to 0$, the master equation is obtained [15],

$$\frac{dP(n, t)}{dt} = \sum_{n' \neq n} \left[T(n | n') P(n', t) - T(n' | n) P(n, t) \right], \quad (2.15)$$

where for notational convenience the subscript for the conditional distribution has been suppressed, and an initial condition n_0 at t_0 is implicitly assumed. This notation will be used in the remainder of the thesis. The interpretation of the master equation is intuitively clear: the probability that the system is in state n increases with the probability that the system moves into it from one of the surrounding states n', but decreases with the probability that the system is already in state n but transitions to another state. The transition rates fully determine the model.

Thus far no discussion on how a stochastic system might be modelled in this formalism has been presented. Let us rectify this with the aid of a specific example; a Moran model with mutation. The only processes in the neutral model that change the state of the system are a birth of type A followed by a death of type B or conversely a birth of type B followed by a death of type A (see Eq. (1.9)). Adding mutation there are a further two reactions which can occur, mutation from B to A and mutation from A to B.[1] For simplicity, we will here assume that these are independent from birth/death events.

In order to model this process in the formalism described above, it is assumed that the probability of an event in the population occurring is proportional to the probability of picking the individuals involved in instigating the event randomly from the population. The transition probability rate is then equal to this product multiplied by a rate constant, which controls the rate at which the interactions occur. Denoting the rate of birth/death as b, the rate of mutation from B to A as ω_1 and the rate of mutation from A to B as ω_2, the probability transition rates are [32]

$$T(n+1|n) = b\overbrace{\left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)}^{\text{A born/B dies}} + \overbrace{(1-b)\omega_1\left(\frac{N-n}{N}\right)}^{\text{B mutates to A}},$$

$$T(n-1|n) = b\underbrace{\left(\frac{N-n}{N}\right)\left(\frac{n}{N-1}\right)}_{\text{B born/A dies}} + \underbrace{(1-b)\omega_2\left(\frac{n}{N}\right)}_{\text{A mutates to B}}, \qquad (2.16)$$

with all other transitions set to zero. Birth/death events are proportional to the product of frequencies of A and B, since each event involves both types. Mutation events only involve one type and are therefore linear in the relevant frequencies. Setting $\omega_1 = \omega_2 = 0$, $b = 1$, one recovers the neutral Moran model, Eq. (1.9). Substituting Eq. (2.16) into Eq. (2.15), the master equation reads

$$\frac{dP(n,t)}{dt} = b\left[\frac{n+1}{N}\frac{(N-n-1)}{(N-1)}P(n+1,t) + \frac{n-1}{N}\frac{(N-n+1)}{(N-1)}P(n-1,t)\right.$$
$$\left. - \left(2\frac{n}{N}\frac{(N-n)}{(N-1)}\right)P(n,t)\right]$$

---

[1] We choose to add mutation because, as discussed in Chap. 1, the neutral model has no deterministic dynamics. This makes it a poor model with which to illustrate general results.

$$+ (1 - b) \left[\omega_2 \frac{n+1}{N} P(n+1, t) + \omega_1 \frac{N-n+1}{N} P(n-1, t) \right. \qquad (2.17)$$

$$\left. - \left(\omega_2 \frac{n}{N} + \omega_1 \frac{N-n}{N} \right) P(n, t) \right],$$

which together with the initial frequency of type A individuals, n_0, fully describes the time evolution of the PDF.[2] Models constructed in this way are referred to as individual based models (IBMs) since they take account of explicit interactions between individuals.

Despite this simple description, the master equation is very rarely analytically tractable. Though there are methods of making analytic progress for a restricted set of problems (usually with some exploitable symmetry), in general obtaining a solution is a formidable task. The situation is further complicated if the system under consideration is multidimensional. While the derivation of the master equation can be easily extended to a multivariate system [49] described by the state vector $\boldsymbol{n}$,

$$\frac{d P(\boldsymbol{n}, t)}{dt} = \sum_{\boldsymbol{n}' \neq \boldsymbol{n}} \left[T(\boldsymbol{n}|\boldsymbol{n}') P(\boldsymbol{n}', t) - T(\boldsymbol{n}'|\boldsymbol{n}) P(\boldsymbol{n}, t) \right], \qquad (2.18)$$

solutions to the equation become even harder to obtain.

There are ways to make progress however. We will consider two avenues in turn: an analytic approximation (the Fokker-Planck equation) and a method of simulating particular realisations of a process obeying the master equation. For pedagogical reasons, we shall consider the method of simulation first.

2.3 The Gillespie Algorithm

A particular realisation of a process obeying the master equation may be simulated using the Gillespie algorithm [17]. A collection of such realisations can then be used to estimate properties of the underlying distribution, $P(\boldsymbol{n}, t)$. While the algorithm itself is not essential to the work in this thesis, it does provide a nice insight into the behaviour of systems obeying the master equation. As such it is useful to review the salient points of the algorithm here.

We begin by describing a system comprised of m different species, so that the state of the system is given by an m-dimensional state vector $\boldsymbol{n}$. The system is set to $\boldsymbol{n}_0$ at t_0 so that the initial distribution is simply a delta peak $p(\boldsymbol{n}, t_0) = \delta(\boldsymbol{n} - \boldsymbol{n}_0)$. We suppose that there are u different types of reaction which may occur given the system is in this state. In order to make the following discussion more clear, it is useful to introduce the stoichiometric matrix, ν. The stoichiometric matrix is an m by u matrix which gives a concise way of stating which species were transformed

---

[2]Note that while we could have specified the birth/death events and mutation events by separate transition rates in Eq. (2.16), this would have led to no change in Eq. (2.17).

in a given reaction. Each element $\nu_{i\mu}$ ($i = 1 \ldots m$, $\mu = 1 \ldots u$) gives the change in number of the ith species due to the μth reaction. The u reactions then take the system from state $\boldsymbol{n}$ to $\boldsymbol{n}' = \boldsymbol{n} + \boldsymbol{\nu}_\mu$, where the vector $\boldsymbol{\nu}_\mu$ is the μth column of the matrix ν. In this notation, for a given state $\boldsymbol{n}$ at time t, we can describe all the transitions by $T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n})$. The master equation may be rewritten in this new notation as

$$\frac{dP(\boldsymbol{n}, t)}{dt} = \sum_{\mu=1}^{u} \left[T_\mu(\boldsymbol{n} | \boldsymbol{n} - \boldsymbol{\nu}_\mu) P(\boldsymbol{n} - \boldsymbol{\nu}_\mu, t) - T(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n}) P(\boldsymbol{n}, t) \right]. \quad (2.19)$$

In the Moran model with mutation, we simply have $m = 1$ and $u = 2$, while the stoichiometry matrix is

$$\nu = (1, \ -1) \quad (2.20)$$

since the system is described by a single species variable and the transitions either increase or decrease this number by one.

We wish to simulate a realisation of the stochastic process which obeys the master equation. For each reaction realised, this requires evaluating two stochastic quantities. Not only do we need to know which of the u reactions occurs, we also need to know when it occurs. In order to do this in a way consistent with the probability distribution evolving according to Eq. (2.19), we first move to considering an equivalent form of the Eq. (2.19) in terms of a new distribution $P(\tau, \mu | \boldsymbol{n}, t)$. This distribution is defined such that $P(\tau, \mu | \boldsymbol{n}, t)\Delta\tau$ is the probability that the next reaction is the μth and that it occurs in the time interval $[t + \tau, t + \tau + \Delta\tau]$. In other words, it is the probability that no reaction occurs between t and $t + \tau$ and that when the reaction occurs it is the μth reaction in the interval $[t + \tau, t + \tau + \Delta\tau]$. We now note that the conditional distribution, $P(\boldsymbol{n}, \tau | \boldsymbol{n}, t)$, is the probability that no reaction has occurred between t and τ. Since, given some state $\boldsymbol{n}$, the probability that the system transitions to state $\boldsymbol{n}'$ in time $\Delta\tau$ is given by the transition probability $W_{\Delta t}(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n}) \approx T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n})\Delta\tau$ (see Eq. (2.12)), $P(\tau, \mu | \boldsymbol{n}, t)\Delta\tau$ and $P(\boldsymbol{n}, \tau | \boldsymbol{n}, t)$ can be related by

$$P(\tau, \mu | \boldsymbol{n}, t)\Delta\tau = P(\boldsymbol{n}, \tau | \boldsymbol{n}, t) T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n})\Delta\tau. \quad (2.21)$$

To calculate $P(\boldsymbol{n}, \tau | \boldsymbol{n}, t)$ we return to the master equation, Eq. (2.19), and fix the state of the system to $\boldsymbol{n}$. The resulting equation is

$$\frac{dP(\boldsymbol{n}, \tau | \boldsymbol{n}, t)}{d\tau} = -P(\boldsymbol{n}, \tau | \boldsymbol{n}, t) \sum_{\mu=1}^{u} T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n}). \quad (2.22)$$

Given the system is already in state $\boldsymbol{n}$, the master equation can thus be solved for $P(\boldsymbol{n}, \tau | \boldsymbol{n}, t)$ to show

$$P(\boldsymbol{n}, \tau | \boldsymbol{n}, t) = \exp\left(-\sum_{\mu=1}^{u} T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n})\tau \right); \quad (2.23)$$

the time until the next reaction occurs, τ, is exponentially distributed. Substituting this into Eq. (2.21), one obtains

$$P(\tau, \mu|\boldsymbol{n}, t)\Delta\tau = \exp\left[-\sum_{\mu=1}^{u} T_{\mu}(\boldsymbol{n} + \boldsymbol{\nu}_{\mu}|\boldsymbol{n})\tau\right] T_{\mu}(\boldsymbol{n} + \boldsymbol{\nu}_{\mu}|\boldsymbol{n})\Delta\tau. \quad (2.24)$$

Since this can be equivalently expressed

$$P(\tau, \mu|\boldsymbol{n}, t) = \left(\sum_{\mu=1}^{u} T_{\mu}(\boldsymbol{n} + \boldsymbol{\nu}_{\mu}|\boldsymbol{n})\exp\left[-\sum_{\mu=1}^{u} T_{\mu}(\boldsymbol{n} + \boldsymbol{\nu}_{\mu}|\boldsymbol{n})\tau\right]\right)\left(\frac{T_{\mu}(\boldsymbol{n} + \boldsymbol{\nu}_{\mu}|\boldsymbol{n})}{\sum_{\mu=1}^{u} T_{\mu}(\boldsymbol{n} + \boldsymbol{\nu}_{\mu}|\boldsymbol{n})}\right),$$

we see clearly that the time until the next reaction and the specific reaction which takes place are statistically independent. The time until the next reaction and the next reaction itself can therefore be separately specified by two distinct random numbers drawn from the above distributions [16]. This allows us to proceed with an algorithm to evaluate a trajectory for the master equation:

1. Calculate the transition rates based on the initial conditions
2. Use these to determine the distributions of next reaction times and reactions
3. Draw a random time until the next reaction and a type of reaction to occur from these distributions
4. Change the time and state of the system accordingly
5. Recalculate the transition rates based on the new state of the system
6. Repeat.

While this simulation procedure gives a useful interpretation of the evolution of the system at short times, it gives no insight into the statistical behaviour of the system. Further, though one can look at the statistics of many realisations of the Gillespie algorithm to infer the properties of $P(\boldsymbol{n}, t)$, this does not provide as deep an understanding of the system as an analytical treatment. While the Gillespie algorithm will prove useful to compare against analytic results, for a deeper understanding we must resort to solving an approximation of the master equation.

2.4 The Expansion of the Master Equation

So far a general formalism for modelling a class of stochastic systems has been presented. It has been stated that the governing equation, the master equation, is in the vast majority of cases very difficult to solve. While simulating the master equation has been shown to be relatively straightforward, we are still no further in making the analytic progress which we initially sought. It will now be shown that the problem can be simplified significantly by using an approximation which resembles the Kramers-Moyal expansion.

The Kramers-Moyal expansion is one of a set of schema which approximate the master equation by assuming a continuous state space. The van-Kampen system size expansion, or linear noise approximation (LNA), is another such scheme [49]. Technically, the approximation used here is not the Kramers-Moyal expansion, which features no explicit small parameter [15]. The approximation presented is instead a hybrid of the two methods, since it follows the methodology of the Kramers-Moyal expansion but using an explicit small parameter determined from the system size.

To illustrate the idea, the procedure will be applied to the specific example of the Moran model with mutation (introduced in Sect. 2.2) before a more general treatment is provided. We begin by taking the master equation (2.17), and introduce a new variable x such that $x = n/N$. The master equation then becomes

$$
\begin{aligned}
\frac{\partial p(x,t)}{\partial t} = & \, b\left[\left(x + \frac{1}{N}\right)\frac{N}{N-1}\left(1 - x - \frac{1}{N}\right)p(x + 1/N, t)\right. \\
& + \left(x - \frac{1}{N}\right)\frac{N}{N-1}\left(1 - x + \frac{1}{N}\right)p(x - 1/N, t) \\
& \left. - \left(2x\frac{N}{N-1}(1-x)\right)p(x,t)\right] \\
& + (1-b)\left[\omega_2\left(x + \frac{1}{N}\right)p(x + 1/N, t)\right. \\
& \left. + \omega_1\left(1 - x + \frac{1}{N}\right)p(x - 1/N, t) - (\omega_2 x + \omega_1(1-x))\,p(x,t)\right],
\end{aligned}
$$

$$(2.25)$$

where we note $p(x,t)$ is a new continuous distribution. The recurrent factors of $1/N$ in this equation give us a clue as how to proceed. If N is large, a Taylor expansion of $p(x,t)$ about x can be conducted. Assuming that the mutation rate is small (of order N^{-1}) and collecting terms order by order in $1/N$, one arrives at a one-dimensional example of the Fokker-Planck equation;

$$
\begin{aligned}
\frac{\partial p(x,t)}{\partial t} = & -\frac{(1-b)}{N}\frac{\partial}{\partial x}\left[(\omega_1 - (\omega_1 + \omega_2)x)p(x,t)\right] \\
& + \frac{b}{2N^2}\frac{\partial^2}{\partial x^2}\left[2x(1-x)p(x,t)\right] + \mathcal{O}(N^3).
\end{aligned}
$$

$$(2.26)$$

This PDE in two variables, x and t, is far more amenable to analysis than the $N + 1$ partial difference equations comprising the master equation. Its physical interpretation is perhaps most evident when read as a convection-diffusion equation; the term preceding $p(x,t)$ in the first spatial derivative governs its 'bulk advection', while that in the second derivative describes diffusion. For this reason they are referred to

as the drift and diffusion terms respectively.[3] We can also write a continuity equation for this system by introducing the probability flux or probability current, $J(x, t)$;

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x}. \tag{2.27}$$

Say we want to calculate how the mean value of x, $\langle x \rangle = \int x p(x, t) dx$, evolves in time. We can derive an equation for its time-evolution by multiplying Eq. (2.26) by x and integrating over all x. The expression can be further simplified by noting that $J(x, t)$ must be zero at the reflecting boundaries $x = 0$ and $x = 1$ (we will discuss meaning of this term and other boundary conditions in Sect. 2.5) and that in this particular model the diffusion term is also zero at the boundaries. Letting $\tau = t/N$, the resulting equation is then

$$\frac{d\langle x \rangle}{d\tau} = (1 - b) [\omega_1 - (\omega_1 + \omega_2)\langle x \rangle]. \tag{2.28}$$

The time-evolution of the mean is governed entirely by the drift term. This behaviour can also be obtained by rescaling time in Eq. (2.26) such that $\tau = t/N$, taking the limit $N \to \infty$ and noting that $p(x', t) = \delta(x(\tau), x')$ is a solution of the resulting equation. We will call this the macroscopic behaviour of the system, or alternatively the deterministic limit. Note that for the neutral model, with $\omega_1 = \omega_2 = 0$, there are no deterministic dynamics, as stated in Eq. (1.10).

The idea is generalised as follows. Let us postulate that there is some large parameter N which is both inversely proportional to the reaction rates and some measure of the typical size of the system. In the case of the Moran model this is specifically the size of the system, though in general it could be the typical size (or volume) of the system which governs the interaction rate, or even the inverse of a naturally small reaction rate. A new set of variables $\boldsymbol{x} = \boldsymbol{n}/N$ is introduced. If N is a measure of system size, the new variables can be naturally interpreted as some measure of the concentration of each species in the population.

Introducing the functions $f_\mu(\boldsymbol{x}) = T_\mu(N\boldsymbol{x} + \boldsymbol{\nu}_\mu | N\boldsymbol{x})$, the master equation Eq. (2.19) can be reexpressed

$$\frac{dp(\boldsymbol{x}, t)}{dt} = \sum_{\mu=1}^{u} \left[f_\mu(\boldsymbol{x} - \boldsymbol{\nu}_\mu N^{-1}) p(\boldsymbol{x} - \boldsymbol{\nu}_\mu N^{-1}, t) - f_\mu(\boldsymbol{x}) p(\boldsymbol{x}, t) \right], \tag{2.29}$$

where we have again changed from a distribution $P(\boldsymbol{n}, t)$ to $p(\boldsymbol{x}, t)$. In this form it is clear that if N is large, one may proceed in the same way as in the Moran model

---

[3] An unfortunate clash of nomenclature appears here between the physics and biology communities. In population genetics, genetic drift is the process by which the composition of a population is changed by noise. Drift in the context of population genetics therefore refers to the noisy component of a system's behaviour, rather than the deterministic component, as in physics.

example by assuming $\boldsymbol{x}$ is continuous and implementing a Taylor expansion about $\boldsymbol{x}$. This is the essence of the master equation expansion.

While the expansion contains results in decreasing orders of N, it can be shown that the resulting expression only leads to a strictly positive PDF if truncated at order N^{-2}, or alternatively keeping the infinite series of terms. This is the Pawula theorem [38]. Since in going to infinite precision the expression is as complicated as the master equation, we truncate the expansion at $\mathcal{O}(N^{-2})$ to achieve our desired simplification. The resulting equation is a multivariate FPE [38] which takes the form

$$
\frac{\partial p(\boldsymbol{x}, t)}{\partial t} = -\frac{1}{N} \sum_{i=1}^{m} \frac{\partial}{\partial x_i} [A_i(\boldsymbol{x}) p(\boldsymbol{x}, t)] + \frac{1}{2N^2} \sum_{i,j=1}^{m} \frac{\partial^2}{\partial x_i \partial x_j} [B_{ij}(\boldsymbol{x}) p(\boldsymbol{x}, t)],
$$

(2.30)

where $\boldsymbol{A}(\boldsymbol{x})$ is is now the drift vector, while $\boldsymbol{B}(\boldsymbol{x})$ is the diffusion matrix. Their forms are governed by the stoichiometry matrix and the transition rates [33] such that

$$
A_i(\boldsymbol{x}) = \lim_{N \to \infty} \sum_{\mu=1}^{u} \nu_{i\mu} f_\mu(\boldsymbol{x})
$$

(2.31)

and

$$
B_{ij}(\boldsymbol{x}) = \lim_{N \to \infty} \sum_{\mu=1}^{u} \nu_{i\mu} \nu_{j\mu} f_\mu(\boldsymbol{x}).
$$

(2.32)

Once again, rescaling time such that $\tau = t/N$ and taking the limit $N \to \infty$, the resulting equation admits $P(\boldsymbol{x}', t) = \delta(\boldsymbol{x}(\tau), \boldsymbol{x}')$ as a solution such that

$$
\frac{d\boldsymbol{x}}{d\tau} = \boldsymbol{A}(\boldsymbol{x}),
$$

(2.33)

describes the deterministic, macroscopic dynamics. We note that it is also possible to obtain an analogous macroscopic description directly from the master equation (2.29) in terms of the mean value $\langle \boldsymbol{x} \rangle$. However, for a general nonlinear system, this requires making the further assumption that the decomposition $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$ can be made in order to close the resulting equations. This is only generally true if the distributions for x_i are delta functions, or as we have seen to be equivalent, if the system size N is infinite.

Since it is possible to calculate $\boldsymbol{A}(\boldsymbol{x})$ and $\boldsymbol{B}(\boldsymbol{x})$ entirely from the stoichiometry matrix and reaction rates, models are sometimes expressed in the notation of chemical reactions, in terms of reactants and products (in fact the term 'stoichiometric matrix' is borrowed from chemistry). For an arbitrary m-dimensional IBM, whose dynamics are fully described by a set of u reaction rates, the model can be expressed in chemical reaction notation as

$$\sum_{i=1}^{m} a_{\mu i} X_i \xrightarrow{r_\mu} \sum_{i=1}^{m} b_{\mu i} X_i, \quad \forall \mu = 1, \ldots, u, \tag{2.34}$$

where $a_{\mu i}$ and $b_{\mu i}$ respectively specify the reactants and products of the μth reaction, and r_μ are the reaction rate constants. The elements of the stoichiometric matrix are then given by $\nu_{i\mu} = b_{\mu i} - a_{\mu i}$, while the reaction rate constants r_μ are related to the transition rates by an equation of the form

$$T_\mu(\boldsymbol{n}'|\boldsymbol{n}) \propto r_\mu \prod_{i=1}^{m} a_{\mu i} \frac{n_i}{N}. \tag{2.35}$$

Implicit in this notation is the assumption that the probability of a reaction occurring is proportional to the product of the reactant concentrations. This is also known as the law of mass action. Most often this is the case, however situations may arise in which we wish to incorporate additional state dependence. An example of such a system will be given in Sect. 2.9.1. For clarity we will not describe such reactions in this notation. Additionally we have introduced a measure of system size N. The determination of what this parameter should be is highly dependent on the system under consideration.

In expanding the master equation, we have moved from the microscopic description of an IBM involving as many equations as there are states (the master equation) to one in which there are only as many variables as there are species (the FPE). The crux of the approximation is in changing to a new set of variables $\boldsymbol{x} = \boldsymbol{n}/N$ which are approximately continuous (the diffusion approximation) before applying a Taylor expansion to the master equation and neglecting terms of order N^{-3}. When modelling populations, the parameter N is usually identified as the size of the population. In the limit $N \to \infty$, we have seen that the FPE describes a deterministic dynamic which we have termed the macroscopic dynamic. In this spirit, since the FPE lies between the master equation and deterministic equation in terms of detail, it is often referred to as a mesoscopic description.

The Fokker-Planck equation is clearly more amenable to analysis than the master equation with which we began. Of particular interest is the one-dimensional Fokker-Planck equation, from which many properties of interest can be calculated analytically.

2.5 The Fokker-Planck Equation and Some Useful Manipulations

In this section we will discuss the way in which the one-dimensional FPE can be used to obtain three particular quantities of interest; the stationary probability distribution, the first passage probability and the first passage time. However, let us first discuss some particular boundary conditions of the one-dimensional FPE which will come in useful later.

We begin by imagining the system existing on an interval $[a_1, a_2]$ in state space, from which it cannot leave. The FPE (2.26) describes such a system, with x lying on the interval $[0, 1]$. In this case there should be zero flow of probability across the boundary. Therefore the probability current must be zero when evaluated at the boundaries; $J(a_1, t) = J(a_2, t) = 0$. Such boundaries are called reflecting.

Now let us consider a system in which there exist states where there are no dynamics. Such states are called absorbing. In order to account for this, we define the barriers as existing outside of the interval, so that when the system reaches the boundary it is removed. The probability of being at either of the boundaries is then zero, $p(a_1, t) = p(a_2, t) = 0$. If the barriers a_1 and a_2 of the system are absorbing, it is clear that given an infinite amount of time, the probability of the system remaining on the interval will tend to zero as probability 'leaks out' of the interval. An example of such a system is the neutral Moran model with transition rates (1.9) discussed in the introduction. Finally, if the boundaries are at infinity, we expect $p(x, t)$, along with $J(x, t)$, to vanish at $x = \pm\infty$.

2.5.1 The Stationary Probability Distribution

First consider a system with two reflecting boundaries at a_1 and a_2. In the limit of long times one might expect the PDF $p(x, t)$ in Eq. (2.30) to become independent of time. We call this distribution the stationary distribution, and it is defined by

$$p_{st}(x) = \lim_{t \to \infty} p(x, t). \tag{2.36}$$

Since $p_{st}(x)$ is independent of time, we find it must satisfy the equation [15]

$$-\frac{1}{N}\frac{d}{dx}[A(x)p_{st}(x)] + \frac{1}{2N^2}\frac{d^2}{dx^2}[B(x)p_{st}(x)] = 0. \tag{2.37}$$

The solution to this equation is

$$p_{st}(x) = \exp\left(\int_{a_1}^{x} dy \frac{2NA(y) - dB(y)/dy}{B(y)}\right)\left[\int_{a_1}^{a_2} dx \exp\left(\int_{a_1}^{x} dy \frac{2NA(y) - dB(y)/dy}{B(y)}\right)\right]^{-1},$$

or

$$p_{st}(x) = \frac{1}{B(x)}\exp\left(2N\int_{a_1}^{x} dy \frac{A(y)}{B(y)}\right)\left[\int_{a_1}^{a_2} dx \frac{1}{B(x)}\exp\left(2N\int_{a_1}^{x} dy \frac{A(y)}{B(y)}\right)\right]^{-1} \tag{2.38}$$

where we have appropriately normalised so that $\int_{a_1}^{a_2} dx\, p_{st}(x) = 1$. This can yield interesting results about the system's long time behaviour. For instance, we can look at the FPE for the Moran model with mutation, Eq. (2.26), in which there are no

absorbing states for both ω_1 and ω_2 greater than zero and b less than one. Comparing Eq. (2.26) with Eq. (2.30), we see for this system that

$$A(x) = (1 - b)(\omega_1 - (\omega_1 + \omega_2)x), \qquad B(x) = 2bx(1 - x). \qquad (2.39)$$

Substituting these terms into Eq. (2.38), and noting that x lies on the interval [0, 1], the stationary distribution takes the form

$$p_{st}(x) = \frac{x^c(1 - x)^d}{\int_0^1 dx x^c(1 - x)^d}, \qquad (2.40)$$

with

$$c = N\omega_1(b^{-1} - 1) - 1, \quad d = N\omega_2(b^{-1} - 1) - 1.$$

Some of the different behaviour the system can exhibit is demonstrated in Fig. 2.1.

If the system instead has absorbing boundaries, there is no stationary distribution on the interval $a_1 < x < a_2$, since all the stochastic trajectories eventually leave the interval. In such cases, there are other measures of the system's behaviour which are more illuminating.
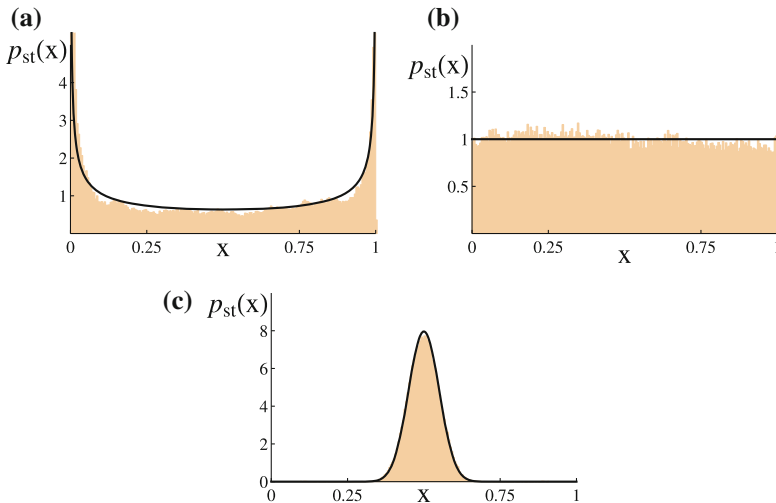

Fig. 2.1 Different regimes of Eq. (2.40) with different c and d parameters. The histograms are obtained from Gillespie simulations using the model defined by Eq. (2.16). **a** Parameters $c < 0, d < 0$. **b** Parameters $c = 0, d = 0$. **c** Parameters $c > 0, d > 0$

2.5.2 First Passage Problems

First passage problems are those which ask 'what is the probability the system reaches a particular final condition?', or 'what are the statistics of the time for the system to reach this final condition?'. In order to calculate these first passage properties, we find it useful to work with the backward Fokker-Planck equation (BFPE). This can be expressed in one dimension as [15, 38]

$$-\frac{\partial q(x, t|x_0, t_0)}{\partial t_0} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [q(x, t|x_0, t_0)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [q(x, t|x_0, t_0)].$$
(2.41)

If the process under consideration is homogeneous, the evolution of the system depends only on the difference between the initial and final time $t - t_0$, and the BFPE can be rewritten in terms of the derivative with respect to t;

$$\frac{\partial q(x, t|x_0, t_0)}{\partial t} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [q(x, t|x_0, t_0)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [q(x, t|x_0, t_0)].$$
(2.42)

The key difference between the forward FPE (2.30) and the BFPE, is which set of variables are kept fixed and which vary. In the forward FPE the initial conditions x_0 at t_0 are kept fixed, and one finds for solutions for $t \geq t_0$. In the BFPE we keep the final condition x at t fixed and calculate for solutions with $t_0 \leq t$. Since in the BFPE we fix the final condition, it is clearly more useful when dealing with first passage problems. For simplicity, we restrict ourselves to a one-dimensional homogeneous system.

We wish to know the time until a system first escapes the region between two points, $x = a_1$ and $x = a_2$, given some initial condition $a_2 > x_0 > a_1$. This time is clearly a stochastic variable and so it will be described by a PDF indicating the probability of a certain first passage time t given initial condition x_0. It is denoted here by $\mathcal{T}(x_0, t)$. We begin by defining the probability $G(x_0, t)$ that, at some time t, the system is still on the interval;

$$G(x_0, t) = \int_{a_1}^{a_2} dx\, q(x, t|x_0, t_0),$$
(2.43)

where the dependence on the initial time has been suppressed. Integrating Eq. (2.42) over x between a_1 and a_2, we find in fact the equation for $G(x_0, t)$ obeys the same BFPE as $q(x, t|x_0, 0)$;

$$\frac{\partial G(x_0, t)}{\partial t} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} [G(x_0, t)] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} [G(x_0, t)],$$
(2.44)

with initial condition

$$G(x_0, t_0) = 1 \quad \text{if } a_1 < x_0 < a_2, \tag{2.45}$$

$$G(x_0, t_0) = 0 \quad \text{elsewhere}, \tag{2.46}$$

and boundary conditions

$$G(a_1, t) = G(a_2, t) = 0. \tag{2.47}$$

Since $G(x_0, t)$ is the probability that at time t the system is still on the interval $a_1 < x < a_2$, the quantity $G(x_0, t) - G(x_0, t + \Delta t)$ is the probability that the system has reached one of the boundaries during t to $t + \Delta t$. This can be related to $\mathcal{T}(x_0, t)$ quite simply, since

$$\mathcal{T}(x_0, t)\Delta t = G(x_0, t) - G(x_0, t + \Delta t), \tag{2.48}$$

or, rearranging and sending $\Delta t \to 0$,

$$\mathcal{T}(x_0, t) = -\frac{\partial G(x_0, t)}{\partial t}. \tag{2.49}$$

The mean time for the system to leave the interval, denoted $T(x_0)$, can be calculated directly from the distribution $\mathcal{T}(x_0, t)$ by $T(x_0) = \int_{t_0}^{\infty} t\mathcal{T}(x_0, t)\, dt$. Using the above equality, this can be expressed

$$T(x_0) = -\int_{t_0}^{\infty} t\frac{\partial G(x_0, t)}{\partial t}dt. \tag{2.50}$$

This equation can be integrated by parts; letting $t_0 = 0$ and assuming that $tG(x_0, t)$ tends to zero as $t \to \infty$, one arrives at

$$T(x_0) = \int_0^{\infty} G(x_0, t)dt. \tag{2.51}$$

Integrating Eq. (2.44) over time and noting that $G(x_0, t)$ tends to zero as $t \to \infty$, we then arrive at an equation for the mean time to reach either of the boundaries

$$-1 = \frac{A(x_0)}{N}\frac{d}{dx_0}T(x_0) + \frac{B(x_0)}{2N^2}\frac{d^2}{dx_0^2}T(x_0) \tag{2.52}$$

with the boundary conditions,

$$T(0) = 0, \quad T(1) = 0. \tag{2.53}$$

In this case the boundary conditions follow by noting that they are the time to reach *either* $x = a_1$ or $x = a_2$, so that at both extremes the system has already fixed on the boundaries.
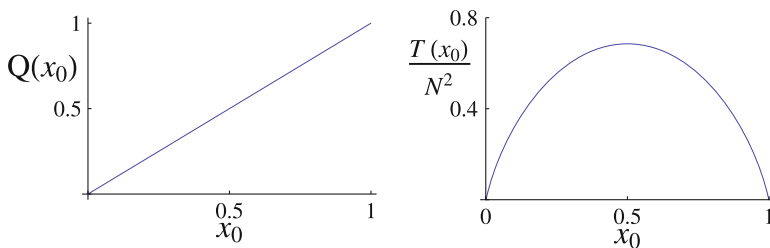
Fig. 2.2 Graphical summary of results from the neutral Moran model. *Left panel* probability of fixation of type A, $Q(x_0)$, in a neutral system as a function of initial A type concentration, x_0. *Right panel* time to fixation of either A or B type, $T(x_0)$, scaled by the system size squared, N^2, as a function of the initial concentration of type A

Let us go back to the neutral Moran model described by Eq. (1.9). In this case the points $x = 0$ and $x = 1$ are absorbing boundaries from which the system cannot leave. At these points the population is said to have fixated. Using Eq. (2.52) and Eq. (2.53), with $A(x) = 0$ and $B(x)$ taken from Eq. (2.39), we can calculate the time this would take to happen. One finds [10]

$$T(x_0) = -N^2 \left[(1 - x_0) \ln (1 - x_0) + x_0 \ln (x_0) \right], \tag{2.54}$$

which is plotted in the right panel of Fig. 2.2. In the nomenclature of population genetics this is called the mean unconditional fixation time.

What about the probability that the system hits one of these boundaries before the other? Let us return to considering a function of the same form as $G(x_0, t)$ in Eq. (2.43), but this time introduce two slightly different functions to account for different integration limits;

$$G_{a_1}(x_0, t) = 1 - \int_{a_1}^{\infty} dx \, q(x, t | x_0, t_0) = \int_{-\infty}^{a_1} dx \, q(x, t | x_0, t_0), \tag{2.55}$$

$$G_{a_2}(x_0, t) = \int_{a_2}^{\infty} dx \, q(x, t | x_0, t_0) = 1 - \int_{-\infty}^{a_2} dx \, q(x, t | x_0, t_0). \tag{2.56}$$

The first of these functions gives the probability that, at time t, the system is at some point $x < a_1$ and the second that it is at some point $x > a_2$. Of course, this gives us no information about which of these regions the system ended up in *first*. To do this would require a consideration of the trajectories conditioned such that the time to one boundary was less that the time to the other. Our task is significantly simplified however if we force the boundaries a_1 and a_2 to be absorbing. Then once the system hits state a_1 or a_2 it is immediately removed and we do not have to worry about time-ordering. The functions $G_{a_1}(x_0, t)$ and $G_{a_2}(x_0, t)$ then tell us respectively whether at time t the system has hit either a_1 or a_2. Both functions obey the equations

$$\frac{\partial G_{a_1/a_2}(x_0, t)}{\partial t_0} = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} \left[G_{a_1/a_2}(x_0, t) \right] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} \left[G_{a_1/a_2}(x_0, t) \right],$$
(2.57)

but with different boundary conditions. As $t \to \infty$, the probability of having hit either a_1 or a_2 tends to one. Introducing

$$Q_{a_1/a_2}(x_0) = \lim_{t \to \infty} G_{a_1/a_2}(x_0, t),$$
(2.58)

we see the equation for both functions is

$$0 = \frac{A(x_0)}{N} \frac{\partial}{\partial x_0} \left[Q_{a_1/a_2}(x_0) \right] + \frac{B(x_0)}{2N^2} \frac{\partial^2}{\partial x_0^2} \left[Q_{a_1/a_2}(x_0) \right],$$
(2.59)

albeit with different boundary conditions. For $Q_{a_1}(x_0)$ we have

$$Q_{a_1}(a_1) = 1, \qquad Q_{a_1}(a_2) = 0,$$
(2.60)

and for $Q_{a_2}(x_0)$ instead

$$Q_{a_2}(a_1) = 0, \qquad Q_{a_2}(a_2) = 1.$$
(2.61)

Once again the neutral Moran model, Eq. (1.9), may be used to illustrate the method. We ask the question, what is the probability of the system reaching the point $x = 1$ given some initial condition x_0? Since at $x = 1$ the system is composed entirely of the A type individuals, this is called the fixation probability. In the neutral case $A(x_0) = 0$ and therefore we obtain

$$Q_1(x_0) = x_0, \qquad Q_0(x_0) = 1 - x_0.$$
(2.62)

The probability of either type fixating is thus simply proportional to their respective initial frequencies in the neutral model (see Fig. 2.2). We note that in population genetics it is most common to simply discuss the probability of fixation of the A type and for simplicity we will often write this probability $Q(x_0) \equiv Q_1(x_0)$.

2.6 Stochastic Differential Equations

Stochastic differential equations (SDEs) are perhaps the earliest way in which dynamical stochastic processes were formalised. Whereas the FPE is a deterministic description of the time evolution of a PDF, SDEs are differential equations describing the evolution of a stochastic variable. Say we wished to postulate the form of a general

nonlinear SDE describing a Markov process. Motivated purely by physical reasoning and a desire for simplicity, we choose to express it

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{\zeta}(t), \tag{2.63}$$

where $g(\boldsymbol{x})$ is a matrix. The vector $\boldsymbol{A}(\boldsymbol{x})$ drives the deterministic dynamics, while the $\boldsymbol{\zeta}(t)$ terms are noise terms arising from some stochastic source. We choose the noise to have zero mean and be delta-correlated in time, with a strength $g(\boldsymbol{x})$ which may be proportional to the state of the system:

$$\langle \zeta_i(t) \rangle = 0, \tag{2.64}$$

$$\langle \zeta_i(t)\zeta_j(t') \rangle = \delta_{ij}\delta(t - t'). \tag{2.65}$$

Now let us fix the entire distribution of $\boldsymbol{\zeta}(t)$ by stating that the noise is Gaussian. Thus each odd moment is zero and each even moment may be calculated from Eq. (2.65). If the noise strength $g(\boldsymbol{x})$ is state dependent the noise is said to be multiplicative [49]. Otherwise it is said to be additive. The choice that the noise is delta correlated in time is a reflection of our wish to model a Markov process. We could obviously have chosen the noise to have a non-zero mean, however the result of this would be to simply add some effective deterministic drift to the system, obfuscating matters.

The SDE perspective offers a much more physically intuitive interpretation of stochastic processes than the Fokker-Planck equation; the system evolves in time according to a deterministic contribution $\boldsymbol{A}(\boldsymbol{x})$ plus some noise $\boldsymbol{\zeta}(t)$. Ultimately however, it provides an inferior starting point for modelling systems with demographic noise. This inferiority is twofold. First, it is a mesoscopic description which has been arrived at in an ad-hoc manner. While the form of the deterministic contributions could conceivably be derived from the interactions in the system, it is unclear what form the noise should take. Second, the interpretation of Eq. (2.63) is ambiguous. To see this, it is perhaps best to ask the question, how would one integrate Eq. (2.63)? The problem here is essentially where the term $g(\boldsymbol{x})$ should be evaluated, since $\boldsymbol{x}$ is undefined at the time when a delta function 'kick' arrives. The two interpretations that are most common are that of Stratanovich and Itō. Respectively they are [38]

$$\boldsymbol{x}(t+\Delta t) - \boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{x}(t))\Delta t + g\left(\frac{\boldsymbol{x}(t) + \boldsymbol{x}(t+\Delta t)}{2}\right)\int_t^{t+\Delta t}\zeta(t')dt' \quad \text{(Stratanovich)},$$

$$\boldsymbol{x}(t+\Delta t) - \boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{x}(t))\Delta t + g(\boldsymbol{x}(t))\int_t^{t+\Delta t}\zeta(t')dt' \quad \text{(Itō)}.$$

Importantly, not only do these two interpretations describe different stochastic processes, they also each lead to different rules of stochastic calculus. While it can be shown that Eq. (2.63) is equivalent to an FPE, these different rules of calculus lead to different forms of the FPE. If one were to use the SDEs as a starting point in the modelling process, one would not only have to postulate the form of the matrix $g(\boldsymbol{x})$, but also the interpretation. If, conversely, one begins with an FPE, the associated SDEs do not suffer from this ambiguity. The FPE is a well defined PDE which

is equivalent to one set of SDEs in the Stratanovich scheme and another in the Itō scheme. Given an FPE, which of the two schemes we work in is thus a matter of choice. We choose to work in the Itō setting, and therefore ignore the Stratanovich interpretation for the rest of the thesis. In the Itō setting, different rules of calculus apply when conducting a non-linear change of stochastic variables, Itō calculus [15]. However, since only linear transformations are utilised in this thesis (for which the rules of calculus are unchanged) the details of Itō calculus will not be discussed here.

Defining the matrix

$$B(x) = Ng(x)g^T(x), \tag{2.66}$$

the SDE (2.63) can be rewritten as

$$\frac{dx}{d\tau} = A(x) + \frac{1}{\sqrt{N}}\eta(\tau), \tag{2.67}$$

with the Gaussian white noise now characterised by

$$\langle \eta_i(\tau) \rangle = 0, \tag{2.68}$$

$$\langle \eta_i(\tau)\eta_j(\tau') \rangle = B_{ij}(x)\delta(\tau - \tau'). \tag{2.69}$$

Interpreting Eq. (2.67) in the Itō sense, it can be shown to be entirely equivalent to the FPE (2.30) with the same drift vector $A(x)$ and diffusion matrix $B(x)$, and time scaled such that $\tau = t/N$. As a result of this alternate form, the diffusion matrix is sometimes also referred to as the noise covariance matrix or noise correlation matrix. In this form some insights about the FPE become very clear. Firstly, in the $N \to \infty$ limit, the system follows a deterministic trajectory governed by $A(x)$. Secondly, the size of the noise is governed by $N^{-1/2}$. Increasing the system size decreases the relative magnitude of the noise.

We have here shown once again that in the limit of infinite system size the system follows a deterministic trajectory. Further, this trajectory is also the mean path of all stochastic trajectories for large N. Both these ideas are very intuitive when posed in the SDE formalism. It is for this reason that we will often choose to work in this setting. Since it is clear that the deterministic element of the equation is of great importance in governing the dynamics, we will now review some basic properties of dynamical systems theory, as well as their extensions to certain stochastic problems.

2.7 Dynamical Systems Theory

In the deterministic limit, the systems under consideration in this thesis will take the form of a set of ODEs

$$\frac{dx}{dt} = A(x). \tag{2.70}$$

We wish to understand the dynamics of such a system. In general, a solution to Eq. (2.70) can only be obtained if the system is linear. In this case the system may be described by

$$\frac{d\boldsymbol{x}}{dt} = H\boldsymbol{x},\tag{2.71}$$

which has solutions

$$\boldsymbol{x}(t) = \sum_{i=1}^{m} c_i \boldsymbol{v}^{(i)} e^{\lambda^{(i)}t},\tag{2.72}$$

where c_i are a set of constants determined from the initial conditions of the problem, $\boldsymbol{v}^{(i)}$ are the right-eigenvectors of H and $\lambda^{(i)}$ their associated eigenvalues. If all the eigenvalues have a negative real part the system will collapse onto the point $x_i = 0$ as $t \to \infty$. If one of the eigenvalues is positive, $\boldsymbol{x}$ will grow exponentially in the direction described by the associated right-eigenvector. If the largest eigenvalue of the system is zero, say $\lambda^{(1)} = 0$, the final state of the system will be dependent on the initial conditions via the constant c_1 [46].

If the system under consideration is non-linear, while one may be able to solve the equations in some special instances, it is more than likely we will not be able to make analytic progress. This is particularly true for non-linear systems in many dimensions. We can however answer questions about the potential long-time behaviour of the system.

The fixed points are points at which there are no dynamics. They shall be denoted $\boldsymbol{x}^*$ and are the solutions to the equations

$$\boldsymbol{A}(\boldsymbol{x}^*) = 0.\tag{2.73}$$

Crucially we want to know what happens to perturbations around these points; do they shrink or do they grow? In other words are the fixed points stable or unstable? To do this we conduct a Taylor expansion in a variable $\boldsymbol{\xi}$ which perturbs $\boldsymbol{x}$ by a small amount ϵ about the fixed point such that $\boldsymbol{x} = \boldsymbol{x}^* + \epsilon\boldsymbol{\xi}$. Conducting the expansion and neglecting nonlinear terms in ϵ, we find

$$\dot{\boldsymbol{\xi}} = J\boldsymbol{\xi},\tag{2.74}$$

where J is the Jacobian of $\boldsymbol{A}$ evaluated on the fixed point $\boldsymbol{x}^*$ about which we linearise; $J_{ij} = (dA_i/dx_j)|_{\boldsymbol{x}=\boldsymbol{x}^*}$. Since Eq. (2.74) has the same form as Eq. (2.71), the solution for $\boldsymbol{\xi}(t)$ is analogous to Eq. (2.72) and determining the stability of $\boldsymbol{x}^*$ therefore comes down to determining the eigenvalues of J. If all eigenvalues of J are less than zero, perturbations away from the fixed point will shrink back to the fixed point which is stable. If any eigenvalues are greater than zero, any perturbation from the fixed point will blow up, and so the fixed point is unstable.

If the system contains noise (such as Eq. (2.67)) then we can still linearise about a fixed point. Setting $\epsilon = 1/\sqrt{N}$, the typical fluctuation size, one obtains

$$\dot{\boldsymbol{\xi}} = J\boldsymbol{\xi} + \boldsymbol{\zeta}(t), \tag{2.75}$$

where

$$\langle \boldsymbol{\zeta}(t) \rangle = 0 \qquad \left\langle \zeta_i(t)\zeta_j(t') \right\rangle = \delta(t - t')B_{ij}(\boldsymbol{x}^*). \tag{2.76}$$

This is a linear system with additive noise, since to linear order $B(\boldsymbol{x})$ is evaluated at the fixed point. If the fixed point is stable this provides a good approximation for the fluctuations of the system (2.67) in the region of the fixed point.[4] The expression is made even more useful under the assumption that the boundary conditions for this problem lie at $\pm\infty$. In this case the associated linear FPE can be shown to describe a Gaussian distribution [30], and thus the entire PDF of the system can be characterised by its first two moments. The equations for these moments are

$$\frac{d}{dt}\langle \xi_k \rangle = \sum_{j=1}^{m} J_{kj}\langle \xi_j \rangle \tag{2.77}$$

and

$$\frac{d}{dt}\langle \xi_k \xi_l \rangle = \sum_{j=1}^{m} J_{kj}\langle \xi_j \xi_l \rangle + \sum_{j=1}^{m} J_{lj}\langle \xi_j \xi_k \rangle + B_{kl}(\boldsymbol{x}^*). \tag{2.78}$$

The first moment has a solution of the same form as Eq. (2.72). Rather than solve the equation for the second moment, it is more convenient to combine Eqs. (2.77) and (2.78) to give an equation of the covariance matrix Ξ of the PDF [49];

$$\dot{\Xi} = J\Xi + \Xi J^T + B(\boldsymbol{x}^*). \tag{2.79}$$

While this equation for Ξ is the same as that for the second moments, Eq. (2.78), its initial conditions are particularly simple. Given a certain initial condition x_{i0}, (a delta peak $p(\boldsymbol{x}, 0) = \prod_{i=1}^{m} \delta(x_i - x_{i0})$ in the associated linear FPE), Ξ vanishes at $t = 0$. The solution to Eq. (2.79) can then be shown to be [49]

$$\Xi(t) = \int_0^t e^{(t-t')J} B e^{(t-t')J^T} dt'. \tag{2.80}$$

---

[4]Of course we could have conducted this linearisation in the FPE setting if we desired. We note that if the SDE (2.67) had been derived from a master equation (via the expansion detailed in Sect. 2.4), the resulting linearised FPE is the same as that which would be obtained if the LNA had been applied directly to the master equation.

If all the eigenvalues of J are negative, the system eventually relaxes to a stationary distribution with zero mean (see Eqs. (2.72) and (2.77)) and a covariance matrix which solves the equation

$$J \Xi + \Xi J + B(x^*) = 0. \tag{2.81}$$

More generally this equation is known as the Lyapunov equation [22].

2.8 Fast-Variable Elimination: The Origins of a Complicated Problem

In Chap. 1, a simple deterministic system featuring a separation of timescales was presented, along with a method of solution. Such a solution goes by many names such as adiabatic elimination, a quasi-steady state approximation or simply fast-variable elimination [19]. In this thesis the term fast-variable elimination will be used. Let us elaborate on the example discussed in the introduction slightly. We may describe a generalisation of a system such as (1.11), in m variables, as [19]

$$\frac{dx_i}{dt} = f_i(x, y), \quad i = 1 \ldots r \tag{2.82}$$
$$\epsilon_j \frac{dy_j}{dt} = -y_j + \epsilon_j h_j(x, y), \quad j = 1 \ldots m - r.$$

If the parameters ϵ_j are small and the functions $f(x, y)$ and $h(x, y)$ are of the same order, we can see that the dynamics in y act on a much faster timescale than those in x. Now, let us assume that a trajectory of the system in the directions y_i tends to some finite value. This is somewhat implied by the form of the equations we have chosen. The system reaches this value much faster than the dynamics in the x_i directions have time to act. We make the approximation that the system reaches its static y value instantaneously by setting $dy/dt = 0$. The quasi-steady state value of this system, $y_{st}(x)$, is termed the slow manifold, and is the solution to the equations

$$- [y_{st}]_j + \epsilon_j h_j(x, y_{st}) = 0. \tag{2.83}$$

The reduced system in r variables is then given by

$$\frac{dx_i}{dt} = f_i(x, y_{st}), i = 1 \ldots r. \tag{2.84}$$

In the course of this thesis, the term 'slow subspace' will be used to refer to a subspace in which the deterministic system moves at a slower rate than in the other subspaces. The term centre manifold will be used to refer to a subspace upon which there are strictly no deterministic dynamics. The introduction of the term 'slow subspace' is

motivated by the fact that within the mathematical literature the slow manifold has a much more strict definition[5] [1]. The field of deterministic dimensional reduction is a vast one however, and an attempt will not be made to fully adhere to the rigorous definitions detailed in such work. Instead, physical intuition will be used to help guide the analysis and interpretation.

Much of the original work that follows this chapter is concerned with how to appropriately remove fast degrees of freedom from stochastic systems. In the introduction it was merely stated that the generalisation to stochastic systems was not straightforward. Now, with the appropriate mathematical tools and theory in hand, we will briefly see why this is the case, before an outline of some of the existing theory in this area is given.

Since the formulation of the problem has first been discussed in the deterministic setting, an obvious starting point is to consider the stochastic system in terms of SDEs. Indeed this is particularly useful since it is in this setting we will later work. Let us begin, as many authors do, with a consideration of the problem of a particle at position x in a potential $U(x)$, which is subject to noise [15, 42]. A deterministic example of this system, with a quartic potential, was given in the introduction (see Eq. (1.11)). Setting the mass equal to one, this system can be expressed as the SDEs

$$\frac{dx}{dt} = v, \tag{2.85}$$
$$\frac{dv}{dt} = -\beta v - U'(x) + \eta(t),$$

where $U'(x)$ is the potential gradient, β is the coefficient of friction and $\eta(t)$ is zero mean Gaussian white noise with correlations $\langle \eta(t)\eta(t') \rangle = \delta(t - t')\beta D$. If β is sufficiently large, the SDEs have an analogous form to Eq. (2.82). Mirroring the deterministic approach, one can argue that after a comparatively short time, $dv/dt = 0$. Solving for v we find

$$v = \beta^{-1}\left(-U'(x) + \eta(t)\right), \tag{2.86}$$

and therefore the reduced system is described by

$$\frac{dx}{dt} = -\beta^{-1}U'(x) + \beta^{-1}\eta(t). \tag{2.87}$$

The corresponding FPE is given by

$$\frac{\partial p(x, t)}{\partial t} = \frac{1}{\beta}\left\{-\frac{\partial}{\partial x}\left[-U'(x)p(x, t)\right] + D\frac{\partial^2}{\partial x^2}\left[p(x, t)\right]\right\}, \tag{2.88}$$

---

[5]Formally, the slow manifold is defined by the collection of trajectories which are tangent to the slow right-eigenvector at the fixed point, although in practice there is unlikely to be a closed analytic expression for this surface.

which is known as the Smoluchowski equation. We note that since the noise term is additive, both the Itō and Stratonovich calculus lead to the same FPE. This method will be referred to as direct adiabatic elimination, in the style of [42]. While in this case we have been successful in reducing the dimensionality of the problem, difficulties can be encountered when using this method. Among them are cases where the equation for the time-evolution of the slow variable contains nonlinear contributions. Say for instance, that we replace $dx/dt = v$ with $dx/dt = v^2$ in Eq. (2.85). Substituting for the value of the fast-variable v from Eq. (2.86), one would obtain an equation containing $\eta(t)^2$. However it is very hard to attach either a physical or a mathematical meaning to this object since the products of delta functions are ill-defined. Therefore, while the method can be usefully applied in systems where the entire system is linear in the fast-variables, in general other methods of fast-variable elimination must be found. The existing literature on the subject can be coarsely split according to the framework within which the stochastic system is represented; the intuitive SDE formalism or the less mathematically problematic FPE formalism.

The more rigorously derived and well-known Haken slaving principle [20, 40, 41], and several other related methods [29, 50] are developed along similar lines of reasoning as the direct adiabatic elimination method and suffer from the same complications which arise from specifying the slow manifold in a stochastic sense. A notable exception is [31], which deals with a particular model in which there is a true centre manifold (that is, a surface on which there is no deterministic flow), and applies a novel method in which stochastic perturbations away from the manifold are assumed to instantaneously relax along the deterministic trajectories.

More mathematically rigorous work on SDE fast-variable elimination has been conducted for stochastic analogues of normal form coordinate transformations. In general, these seek to determine a nonlinear transformation which simultaneously identifies the fast and slow directions in a controlled way, as well as guaranteeing the absence of ill-defined noise terms. While perhaps the earliest example in the SDE setting was [8], work has been significantly extended in the intervening years [1–3, 44]. However, many of these transformations result in noise convolutions which involve anticipating future unknown noise terms. Further work has proposed the use of additive noise terms to emulate these convolution terms in the limit of long times [6, 23, 43], though these methods are arguably less formal than the theory they rest within. Perhaps the most significant advance in this area therefore came in [39], with the construction of a methodology for stochastic normal form transform that avoids such anticipating memory integrals in many cases, though even here there remain many situations where a long time additive noise substitution must be invoked. A body of work also exists on averaging and homogenisation techniques [4, 36], although both have a more limited range of applicability than stochastic normal forms [39] and the former has been shown in certain cases to be equivalent to a stochastic normal form [43]. One of the biggest drawbacks of the work on normal forms however is that it almost exclusively deals in SDE systems with uncorrelated noise terms, whereas SDEs derived from underlying microscopic IBM models often exhibit strong noise correlations.

Other authors have focused on master or Fokker-Planck type equations as the basis for fast-variable elimination. The main advantage to working in this formalism is that one avoids encountering the ill-defined noise terms which can present themselves in the SDE setting. Much of the work in this formalism is heavily influenced by the work of Zwanzig [53, 54], who proposed that a reduction of dimension could be achieved through the application of projection operators, as illustrated by Gardiner [14], and developed by others [47]. While other variants in the FPE setting have been constructed [7, 28, 38], the projection operator technique remains the most popular. The essential idea is to define a projection operator used to separate the system into two spaces, those of the fast and slow directions. It is most instructive to give a cursory description of the method with reference to the system described by Eq. (2.85). The FPE for this system is known as Kramers equation [15]

$$\frac{\partial p(x, v, t)}{\partial t} = - v \frac{\partial}{\partial x} [p(x, v, t)] + U'(x) \frac{\partial}{\partial v} [p(x, v, t)] \qquad (2.89)$$
$$+ \beta \frac{\partial}{\partial v} [(v) \, p(x, v, t)] + D\beta \frac{\partial^2}{\partial v^2} [p(x, v, t)],$$

which we can express in the more simple notation

$$\frac{\partial p(x, v, t)}{\partial t} = (L_1 + \beta L_2) \, p(x, v, t), \qquad (2.90)$$

where the operators L_1 and L_2 can be read off from the previous equation. The aim is to find an equation for just x by solving an equation for v as a function of x in the limit of large β. Essentially it is assumed that an approximate solution to Eq. (2.90) will be formed from the marginal distribution $p(x, t) = \int p(x, v, t) dv$ and the stationary distribution in the limit of large β, $L_2 p_{\text{eq}}(v) = 0$. This idea is formalised by introducing a projection operator, $R(x, v)$, such that when applied to a function $f(x, v)$ one obtains

$$(Rf)(x, v) = p_{\text{eq}}(v) \int f(x, v) dv. \qquad (2.91)$$

Applying $R(x, v)$ to $p(x, v, t)$ one then obtains

$$(RP)(x, v) = p_{\text{eq}}(v) p(x, t), \qquad (2.92)$$

where $p(x, t)$ is the marginal distribution. Applying $R(x, v)$ to Eq. (2.90) and then separately applying the operator $(1 - R)$ one obtains two separate PDEs for the system. The aim is then to solve the second PDE in such a way as to obtain an equation for $p(x, t)$. In this particular case it can be shown that the reduced FPE does indeed converge to Eq. (2.88) in the limit of infinite β [42]. However, the procedure is by no means straightforward; the resulting reduced FPE has a non-Markovian character which can only be eliminated by a careful consideration of the limiting

behaviour with respect to β. A correct treatment only begins to become clear by either introducing a highly non-linear change of variables [42] or by utilising a Laplace transform. It is perhaps testament to the difficulty of working in this formalism, that the first attempt to derive the Smoluchowski equation as the limiting form of the Kramers equation was given (albeit incorrectly) by Brinkman in 1956, while the first correct treatment was given by Stratanovich in 1963 [45] and again independently by Wilemski in 1976 [51] and Titulaer in 1978 (the reader is referred to [15] for a complete discussion). The complexity of working in this methodology increases massively as one encounters systems with multiplicative noise and correlated noise. Additionally, many systems exist in which the large or small factor is not factorisable in as convenient form as Eq. (2.90), leading to further complications. Unfortunately, the details of these methods are cumbersome, especially in cases where the fast degrees of freedom are distinct from the natural basis of the problem.

The difficulties encountered in two of the main areas of research into fast-variable elimination have been discussed. In the SDE formalism one has to work very hard to avoid introducing ill-defined noise terms and systematically eliminate non-Markovian effects. In the FPE formalism, the mathematical manipulations are better defined, but one loses all sense of physical intuition so that progress becomes difficult. In both settings, the reduced system often ends up taking a very complicated form for most cases of interest. Moreover, it could be argued that in both streams of the literature a fundamental and practical question is often avoided; what advantage is gained by replacing a moderately complicated system in many variables with a more complicated system in fewer variables?

2.9 The Moran Model

So far this chapter has been primarily concerned with the more technical aspects of stochastic theory and fast-variable elimination. Chapters 4 and 5 will focus on the application of fast-variable elimination to a model inspired by the Moran model. A slightly more in depth discussion of the Moran model and its variants in thus called for, along with an historical overview of the relevant areas of mathematical population genetics.

It is often said that there are four main processes which drive the evolution of a population: genetic drift, mutation, selection, and migration [21]. A model which incorporates stochastic effects address the first of these issues, as was first illustrated by Fisher and Wright who considered simple stochastic processes in systems where the population size, N, was finite—the Wright-Fisher model [13, 52]. Subsequent work tended to follow their original approach, assuming discrete generations and discrete state variables (corresponding to the number of individuals in the population of a particular type) [10]. However as more details are added into these models, their discrete-time nature can make an analytic treatment difficult [11].

Progress can be made however if one considers an analogous model in continuous time with individual birth-death events, the Moran model. In the Introduction, a verbal description of the neutral Moran model was presented, while throughout this chapter, the specific example of a Moran model with mutation (see Eq. (2.16)) was used to used to illustrate some of the methods and ideas of stochastic theory. In Sect. 2.4 we saw what makes this model more amenable to analysis; it is formulated as a master equation with small 'jumps' from one state to another, allowing the expansion of the master equation to be conducted, which results in an FPE for the system dynamics. This approximation, although originally suggested by Fisher [12], was popularised by Kimura [9, 26], and proved to be a powerful tool and the starting point for many studies of more complex processes in population genetics [24, 25, 27].

We recall that the neutral Moran model is a model for pure genetic drift described by probability transition rates (1.9). As we have defined it, it consists of two types of individual, labelled A and B. In the nomenclature of population genetics these types are called alleles, which are loosely traits which an individual can carry. If one wished to model the Mendellian genetics of a human population, each individual could be modelled as having two alleles, reflecting the two sets of genetic data available from two chromosome sets [18]. Throughout this thesis we will deal exclusively with models in which each individual carries only one allele. These are referred to as haploid (as opposed to diploid) models. Mutation can be added to the Moran model, as in Eq. (2.16).[6] We now turn our attention to the third evolutionary process on our list, selection.

Before proceeding, we note that the designations A and B for the allele types may appear slightly confusing when presented simultaneously with the drift and diffusion terms of the FPE, $A(x)$ and $B(x)$. This notation has been chosen for consistency with the population genetics literature in the first instance and the physics literature in the second. In order to avoid confusion, the drift and diffusion terms will always be presented with their functional dependence.

2.9.1 The Moran Model with Selection

Selection is the process by which the constitution of a population changes according to some bias. The fitness of an allele within the population is a measure of this bias. Of course, this leaves the term fitness as something of an umbrella label, embodying many potential mechanisms, some of which will be explored in Sect. 6.3. Throughout this thesis, when dealing with the Moran model, frequency-independent (or absolute) fitness will be considered, that is fitness that is independent of the constitution of the population [34].

Within the scope of the Moran model, matters are complicated slightly by the rather artificial nature of the model. In order to maintain a fixed system size N,

---

[6]The way in which mutation is incorporated is of course by no means unique. Mutation could for instance be coupled to birth and death. Some of these ideas are explored in [5].

birth and death events are coupled. Therefore increasing the birth rate necessarily increases the death rate. Let us introduce the parameters b_A as the birth rate of allele A and b_B as the birth rate of allele B. The transitions rates are then

$$T(n+1|n) = b_A \frac{n}{N} \frac{(N-n)}{N-1}, \tag{2.93}$$

$$T(n-1|n) = b_B \frac{(N-n)}{N} \frac{n}{N-1}. \tag{2.94}$$

We now ask, what is the total birth/death rate of the population? Here we must bear in mind that in our formulation of the master equation, we have only taken (and need only take) account transitions which change the state of the system. However there are of course two other processes implicitly assumed to be taking place; allele A reproducing to replace another allele A and likewise for allele B. The time at which *any* reaction occurs is of course a stochastic variable, but as was shown in Sect. 2.3, it is exponentially distributed and characterised by the sum of all possible reactions (see Eq. (2.23)). The sum of all transitions in this instance can be expressed as

$$\sum_{\mu=1}^{4} T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n}) = \frac{1}{N(N-1)} \left\{ b_A n \left[(n-1) + (N-n) \right] \right.$$

$$\left. + b_B(N-n) \left[(N-n-1) + n \right] \right\},$$

$$= \frac{1}{N} \left[b_A n + b_B(N-n) \right]. \tag{2.95}$$

The birth/death rate changes according to the constitution of the population. If all the population is comprised of individuals carrying allele A, $n = N$ and the mean birth/death rate is b_A while if the entire population is B the mean birth/death rate is b_B. While at first glance this seems like a reasonable result, it suggests that the 'fittest type' would reproduce quickly, but also die quickly. Essentially this is an unnatural pathology of a model which requires that the system size is fixed. We can avoid this behaviour if instead we ask that b_A and b_B are weighted such that the mean birth rate of the population is fixed to be one;

$$b_A = \frac{Nw_A}{nw_A + (N-n)w_B} \qquad b_B = \frac{Nw_B}{nw_A + (N-n)w_B}. \tag{2.96}$$

In this way the fitter type in some sense has a 'greater share' of a fixed birth rate.

It is in this fashion we choose to work, with the parameters w_A and w_B introduced to represent the fitness weightings of alleles A and B respectively [35]. The transition rates are then

$$T(n+1|n) = \frac{nw_A}{nw_A + (N-n)w_B} \frac{(N-n)}{N-1}, \tag{2.97}$$

$$T(n-1|n) = \frac{(N-n)w_B}{nw_A + (N-n)w_B} \frac{n}{N-1}. \tag{2.98}$$

The appearance of n in the denominator complicates the expansion of the master equation slightly. This is usually addressed by rewriting the fitness parameters $w_A = 1+s$ and $w_B = 1$, and expanding in powers of s under the very reasonable assumption that s is small. Positive s indicates a small fitness advantage for individuals with allele A. The expansion is carried out for a more general case in Appendix C. One can obtain the results required in this instance by setting $\mathcal{D} = 1$ in the appendix.

An expansion of the master equation leads to the FPE (2.30) with $m = 1$ and $A(x)$ and $B(x)$ given respectively by

$$A(x) = sx(1-x) \quad \text{and} \quad B(x) = 2x(1-x), \tag{2.99}$$

for small s. Here we have included only the lowest order contribution in s to $A(x)$, and omitted the order s correction in B altogether, since it will be negligible compared with $2x(1-x)$. We note that in the case $s = 0$ we obtain the neutral model. In the same manner as in Sect. 2.5.2, we find equations Eqs. (2.59) and (2.52), with $A(x)$ and $B(x)$ given by Eq. (2.99), for the fixation probability of type A, $Q(x_0) = Q_{a_2}(x_0)$, and the unconditional fixation time $T(x_0)$ as a function of x_0, the initial concentration of allele A.

In this case, solving Eq. (2.59), the familiar expression for the probability of fixation, $Q(x_0)$, is found [11]:

$$Q(x_0) = \frac{1 - \exp(-Nsx_0)}{1 - \exp(-Ns)}. \tag{2.100}$$

While the mean time to fixation can also be obtained analytically (see Eqs. (5.12) and (5.13) with M set to N and σ set to s, and Appendix E), it is sufficient to determine it numerically. Illustrative plots are shown in Fig. 2.3.
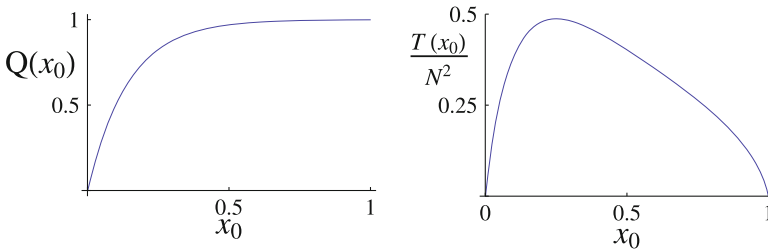

Fig. 2.3 Probability of fixation of allele X, $Q(x_0)$, and mean time to fixation of either allele, $T(x_0)$, in a system where allele A has a selective advantage s over allele B, as a function of initial A concentration. While $Q(x_0)$ is obtained from Eq. (2.100), $T(x_0)$ has been obtained numerically

2.10 Technical Background Overview

In this chapter, the fundamental mathematical tools which are necessary to understand the following research have been discussed. Let us briefly review how they relate to the modelling process.

We began by reviewing the definition of Markov process, which we verbally described as one which has no memory. Clearly however many real-world systems are non-Markovian. In Sect. 2.1 a simple example of such a system was given, however there are many other examples; when considering reproduction in a population, it could cogently be argued that there is a correlation in time between reproduction events, that a single reproduction event at time t makes another at time $t + \Delta t$ less likely. However, if this time is sufficiently small it can be argued that the effect is negligible (effectively another fast-timescale approximation). It may be true that "Non-Markov is the rule, Markov is the exception" [48], but the benefits of simplicity and tractability garnered from treating the process as Markov often outweigh those gained from a more detailed treatment.

In Sect. 2.2, we went on to treat a Markov model in continuous time. This required defining a set of time-independent probability transition rates describing the probability per unit time of moving state. While the master equation provides a total description of the system, we noted that it was in most cases very difficult to solve. Two methods of making progress were then proposed. In Sect. 2.3, the Gillespie algorithm was introduced. While this provides stochastic realisations whose statistics obey the master equation, it does not provide any meaningful understanding of an underlying model. It therefore best serves as a comparative tool for analytic approximations and solutions of the master equation. To make analytic progress a system-size expansion of the master equation was detailed, an approximation which allows the system to be described as an FPE in as many variables as there are species in the system. The equivalent SDE form of the FPE was then given in Sect. 2.6, since this form will often be used in this thesis for intuitive reasons.

In the penultimate section, Sect. 2.8, we arrived at the crux of the thesis; the removal of fast-variables in stochastic systems. Here the origins of the problem in a mathematical sense were discussed, along with the limitations of existing work. In particular, attention was drawn to one of the fundamental problems many existing methods fail to address; a simplification of a problem via a reduction in the number of variables which comes at the cost of its over-complication in other regards, is often little simplification at all.

Throughout this chapter, reference was made to the Moran model and one of its many variants, the Moran model with mutation. This is because a model of its ilk is the subject of much of the thesis. In order to place this work in context, the Moran model was discussed from a more historical standpoint in Sect. 2.9. In this section an additional variant of the model was also introduced which includes the effect of selection.

In what follows, I will introduce two methods of simplifying stochastic problems with a separation of timescales. Both procedures are mathematically explicit,

straightforward to apply, and address the effect of correlated noise terms. In Chap. 3
the first of these approximations will be introduced, the conditioning method. To
begin I develop the method with the aid of a simple illustrative example with an
ecological interpretation in Sect. 3.2, before providing a general formulation. The
results from this example demonstrate the success of the approximation scheme
even in regimes where the fixed point is weakly unstable. In Sect. 3.4 the method
is applied to an epidemiological model with seasonal forcing. This model has been
identified as suffering from the technical numerical difficulties associated with a
large separation between eigenvalues. It will be shown how the method may be used
in tandem with the LNA to provide a very good approximation to results coming
from stochastic simulations.

In Chap. 4 the second of the methods will be introduced, which shall be referred
to as the projection matrix method. In particular, I investigate the method applied to a
generalisation of the Moran model which incorporates migration between a number
of well-mixed populations. Birth and death events are later moderated by a weak
selection pressure. The technique allows the equations for the system (in as many
variables as there are islands) to be reduced so that they resemble those for a single
island, which are amenable to analysis. Once again, the idea behind the method is
simple, its application is systematic, and the results are in very good agreement with
simulations of the full model for a range of parameter values.

Following the reduction of the metapopulation Moran model, Chap. 5 is devoted to
analysing the reduced model, and comparing it to the results obtained from stochastic
simulation. A rich array of behaviour is found, all of which is well predicted by the
reduced system.

Chapter 6 contains work which extends and connects that in the previous chapters.
This includes incorporating mutation into the Moran model with migration, drawing
links between the conditioning method and the projection matrix method and looking
at the possibility of relaxing the fixed system size assumption inherent in the Moran
model.

References

1. L. Arnold, *Random Dynamical Systems*, Springer Monographs in Mathematics (Springer, Berlin, 2003)
2. L. Arnold, P. Imkeller, Normal forms for stochastic differential equations. Probab. Theory Relat. Fields **110**, 559–588 (1998)
3. L. Arnold, X. Kedai, Simultaneous normal form and center manifold reduction for random differential equations, in *Equadiff 91 : International Conference on Differential Equations*, ed. by C. Perello, C. Sim, J. Sola-Morales (1991)
4. N. Berglund, B. Gentz, Geometric singular pertubation theory for stochastic differential equations. J. Differ. Equ. **191**, 1–54 (2003)
5. R.A. Blythe, A.J. McKane, Stochastic models of evolution in genetics, ecology and linguistics. J. Stat. Mech. **P07018** (2007)
6. X. Chao, A.J. Roberts, On the low dimensional modelling of Stratanovich stochastic differential equations. Phys. A **225**, 62–80 (1996)

7. M.N. Contou-Carrere, V. Sotiropoulos, Y.N. Kaznessis, P. Daoutidis, Model reduction of multi-scale chemical Langevin equations. Syst. Control Lett. **60**, 75–86 (2011)

8. P.H. Coullet, C. Elphick, E. Tirapegui, Normal form of Hopf bifurcation with noise. Phys. Lett. A **111**, 277–282 (1985)

9. J.F. Crow, M. Kimura, Some genetic problems in natural populations, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, University of California Press, Berkeley (1956), pp. 1–22

10. W.J. Ewens, *Population Genetics* (Wilmer Brother Limited, Birkenhead, 1969)

11. W.J. Ewens, *Mathematical Population Genetics*, 2nd edn. (Springer, Berlin, 2004)

12. R.A. Fisher, On the dominance ratio. Proc. Roy. Soc. Edinb. **42**, 321–341 (1922)

13. R.A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930)

14. C.W. Gardiner, Adiabatic elimination in stochastic systems. I. Formulation of methods and application to few-variable systems. Phys. Rev. A **29**, 2814–2822 (1984)

15. C.W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2009)

16. D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. **22**, 403–434 (1976)

17. D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)

18. A.J.F. Griffiths, S.R. Wessler, R.C. Lewontin, S.B. Carroll, *Introduction to Genetic Analysis*, 9th edn. (W. H. Freeman and Company, New York, 2007)

19. H. Haken, *Synergetics* (Springer, Berlin, 1983)

20. H. Haken, A. Wunderlin, Slaving principle for stochastic differential equations with additive and multiplicative noise and for discrete noisy maps. Z. Phys. B **47**, 179–187 (1982)

21. D.L. Hartl, A.G. Clark, *Principles of Population Genetics*, 4th edn. Sinauer Associates Inc., Sunderland, Mass. (2007)

22. R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1991)

23. R.Z. Khasminskii, A limit theorem for the solutions of differential equations with random right-hand sides. Theor. Probab. Appl. **11**, 390–406 (1966)

24. M. Kimura, Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harb. Symp. Quant. Biol. **20**, 33–53 (1955)

25. M. Kimura, Diffusion models in population genetics. J. Appl. Probab. **1**, 177–232 (1964)

26. M. Kimura, *Population Genetics, Molecular Evolution and the Neutral Theory* (Chicago University Press, Chicago, 1994)

27. M. Kimura, G.H. Weiss, The stepping stone model of population structure and the decrease in genetic correlation with distance. Genetics **49**, 561–576 (1964)

28. E. Knobloch, K.A. Wiesenfeld, Bifurcations in fluctuating systems: the center manifold approach. J. Stat. Phys. **33**, 611–637 (1983)

29. Y. Lan, T.C. Elston, G.A. Papoian, Elimination of fast variables in chemical Langevin equations. J. Chem. Phys. **129**, 214115 (2008)

30. M. Lax, Fluctuations from the nonequilibrium steady state. Rev. Mod. Phys. **32**, 25–64 (1960)

31. Y.T. Lin, H. Kim, C.R. Doering, Features of fast living: on the weak selection for longevity in degenerate birth-death processes. J. Stat. Phys. **148**, 646–662 (2012)

32. T. Maruyama, *Stochastic Problems in Population Genetics* (Springer, Berlin, 1977)

33. A.J. McKane, T. Biancalani, T. Rogers, Stochastic pattern formation and spontaneous polarisation: the linear noise approximation and beyond. Bull. Math. Biol. **76**, 895–921 (2014)

34. R.E. Michod, *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality* (Princeton University Press, New Jersey, 2000)

35. M.A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, 2006)

36. G.A. Pavliotis, A.M. Stuart, *Averaging and Homogenization*, vol. 53, Multiscale Methods (Springer, New York, 2008)

37. L.E. Reichl, *A Modern Course in Statistical Physics* (Wiley VCH, New York, 1998)

38. H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1989)

39. A.J. Roberts, Normal form transforms seperate slow and fast modes in stochastic dynamics systems. Phys. A **387**, 12–38 (2008)
40. G. Schöner, H. Haken, The slaving principle for Stratanovich stochastic differential equations. Z. Phys. B **63**, 493–504 (1986)
41. G. Schöner, H. Haken, A systematic elimination procedure for Itō stochastic differential equations and the adiabatic approximation. Z. Phys. B **68**, 89–103 (1987)
42. R. Serra, M. Andretta, M. Compiani, G. Zanarini, *Introduction to the Physics of Complex Systems* (Pergamon Press, Oxford, 1986)
43. N. Sri Namachchivaya, Equivalence of stochastic averaging and stochastic normal forms. J. Appl. Mech. **57**, 1011–1017 (1990)
44. N. Sri Namachchivaya, Y.K. Lin, Method of stochastic normal forms. Int. J. Nonlinear Mech. **26**, 931–943 (1991)
45. R.L. Stratanovich, *Introduction to the Theory of Random Noise* (Gordon and Breach, New York, 1963)
46. S.H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview Press, New York, 2001)
47. P. Thomas, R. Grima, A.V. Straube, Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators. Phys. Rev. E **86**, 041110 (2012)
48. N.G. van Kampen, Remarks on non-Markov processes. Braz. J. Phys. **28**, 90–96 (1998)
49. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007)
50. W. Wang, A.J. Roberts, Slow manifold and averaging for slow-fast stochastic differential system. J. Math. Anal. Appl. **398**, 822–839 (2013)
51. G. Wilemski, On the derivation of Smoluchowski equations with corrections in the classical theory of Brownian motion. J. Stat. Phys. **14**, 153–170 (1976)
52. S. Wright, Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)
53. R. Zwanzig, Ensemble method in the theory of irreversibility. J. Chem. Phys. **33**, 1338–1341 (1960)
54. R. Zwanzig, Memory effects in irreversible thermodynamics. Phys. Rev. **124**, 983–992 (1961)

Chapter 3
The Conditioning Method

In this chapter the conditioning method is introduced. The core of the approach is to examine the behaviour of a stochastic system in the SDE framework under the condition that its trajectories are *confined* to the slow subspace of the deterministic version of the system. A similar procedure has been applied in previous works in the analysis of noise induced speciation [11, 12]. Inspired by this success, the method is generalised here for applications to a broader range of systems. I also highlight how the method can be especially useful in some specific instances. Because a static description of the slow subspace is used (as opposed to a stochastic description, as in Eq. (2.86)), the method is applicable to a broader range of systems than the direct elimination procedure or the Haken slaving principle. Moreover, the procedure is mathematically explicit, straightforward to apply, and addresses the effect of correlated noise terms. One also gains a sense of physical intuition as to the behaviour of the system, which is arguably not present in the FPE setting.

I will also show how the conditioning method can be effectively combined with other stochastic approximation techniques, such as the linear noise approximation (LNA) [18]. In the context of the LNA, slow subspaces can be a malign presence. This is because the effect of the slow subspace on the LNA is often to introduce small eigenvalues in the resulting linear SDE. If some eigenvalues of the SDE are close to zero, then the fixed point about which the LNA linearises is only weakly stable, and it may be possible for stochastic fluctuations to carry the system very far from this fixed point, perhaps even into another steady state. In situations such as this, the true non-linear nature of the model is important. Additionally, a large separation between eigenvalues can in some situations lead to theoretical solutions becoming numerically ill-conditioned. Both of these effects can lead to a very poor agreement between stochastic simulations and the LNA theory. It will be seen that the conditioning method can help address both these issues.

In the following section, a method for determining the existence of slow subspaces and their approximate analytic form is introduced. Next, in Sect. 3.2, the conditioning method is developed with the aid of a simple illustrative example with an ecological interpretation, before a general formulation is provided. The results from this example demonstrate the success of the approximation scheme even in regimes where the fixed

point is weakly unstable. In Sect. 3.4, the general formulation of the method is applied to an epidemiological model with seasonal forcing. This model has been identified as suffering from the technical numerical difficulties associated with a large separation between eigenvalues [13]. It will be shown that the conditioning method may be used in tandem with the LNA to provide a very good approximation to results coming from stochastic simulations.

3.1 Identifying the Deterministic Slow Subspace

Even in a deterministic system, it is not necessarily the case that models featuring a separation of timescales will present themselves in such a convenient form as Eq. (2.82). In these equations, the fast-variables were explicitly given by y, but in general the fast directions may not be parallel to the basis of the problem. One may then ask, how can one easily identify a separation of timescales in an arbitrary problem? Let us consider problems in which the system has a known stable fixed point, but a separation in timescales is not necessarily evident on first appraisal:

$$\frac{d\boldsymbol{x}}{dt} = A(\boldsymbol{x}).\tag{3.1}$$

We may linearise this equation about the fixed point $\boldsymbol{x}^*$ to obtain an equation of the form of Eq. (2.74). The solution to the linearised equation is of the same form as Eq. (2.72), and gives an approximate solution to the full system close to the fixed point [17]. From this, we learn that if $\boldsymbol{v}^{(i)}$ is a right-eigenvector of the Jacobian J, with eigenvalue $\lambda^{(i)}$, then perturbations in the direction of $\boldsymbol{v}^{(i)}$ will grow exponentially if $\mathrm{Re}[\lambda^{(i)}] > 0$ and shrink exponentially if $\mathrm{Re}[\lambda^{(i)}] < 0$. If however, the eigenvalues satisfy

$$\mathrm{Re}[\lambda^{(m)}] < \cdots < \mathrm{Re}[\lambda^{(r+1)}] \ll \mathrm{Re}[\lambda^{(r)}] < \cdots < \mathrm{Re}[\lambda^{(1)}] \leq 0\tag{3.2}$$

a separation of time-scales exists; perturbations in the direction of the right-eigenvectors $\boldsymbol{v}^{(r+1)}, \ldots, \boldsymbol{v}^{(m)}$ will decay extremely rapidly in comparison with those in the directions of $\boldsymbol{v}^{(r)}, \ldots, \boldsymbol{v}^{(1)}$. We have thus identified a set of fast directions in the linear system. Let us introduce the r new slow variables $\boldsymbol{\xi}_z$ and the $m - r$ new fast-variables $\boldsymbol{\xi}_w$. For simplicity, let us begin by supposing that all the eigenvalues and eigenvectors are real. We can transform into the new fast-slow basis of the linear system's right-eigenvectors via the matrix $V = [\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)}]$;

$$\begin{pmatrix} \boldsymbol{\xi}_z \\ \boldsymbol{\xi}_w \end{pmatrix} = V^{-1}\boldsymbol{\xi}.\tag{3.3}$$

Applying this to the linear system Eq. (2.74) and utilising the eigendecompostion relation $\Lambda = V^{-1}JV$, where Λ is the diagonal matrix whose ith entry is $\lambda^{(i)}$, we find the new linear system is described by

$$\frac{d}{dt} \begin{pmatrix} \boldsymbol{\xi}_z \\ \boldsymbol{\xi}_w \end{pmatrix} = \Lambda \begin{pmatrix} \boldsymbol{\xi}_z \\ \boldsymbol{\xi}_w \end{pmatrix}. \tag{3.4}$$

Given the eigenvalue relations in Eq. (3.2), we see clearly that this linear system has the same form as Eq. (2.82).

How can one proceed with the full nonlinear system? We begin by introducing r variables z and $m - r$ variables w such that

$$\begin{pmatrix} z \\ w \end{pmatrix} = V^{-1}x, \tag{3.5}$$

to arrive at the set of nonlinear ODEs

$$\frac{d}{dt} \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} A_z(z, w) \\ A_w(z, w) \end{pmatrix}, \tag{3.6}$$

with

$$\begin{pmatrix} A_z(z, w) \\ A_w(z, w) \end{pmatrix} = V^{-1}A(V(z, w)^T). \tag{3.7}$$

In the region of the fixed point, the bases of z and w are parallel to the slow and fast directions respectively. To make progress, we approximate the slow subspace [1] by the surface on which the rate of change in the direction of the fast eigenvector is zero. Formally this is known as the nullcline for w [17]. These equations may then be treated in a way analogous to Eq. (2.82), with the nullcline solution for $w(z)$,

$$A_w(z, w(z)) = 0, \tag{3.8}$$

used as an approximation for the slow subspace in the new coordinates. Indeed, we will find that it often provides an excellent representation, even far from the fixed point.

If the system contains complex eigenvalues and eigenvectors, one may proceed as above with only a few minor changes. The transformation matrix V is redefined such that each column is an eigenvector only if that eigenvector is real. If a complex eigenvector pair $\boldsymbol{v}^{(k)}$ and $\boldsymbol{v}^{(k+1)} = [\boldsymbol{v}^{(k)}]^*$ is encountered, one introduces the new real eigenvectors $\boldsymbol{h}^{(k)} = (\boldsymbol{v}^{(k)} + \boldsymbol{v}^{(k+1)})$ and $\boldsymbol{h}^{(k+1)} = i(\boldsymbol{v}^{(k)} - \boldsymbol{v}^{(k+1)})$, and uses these as the columns of V. Applying the transformation to the linear system no longer results in a system governed by the diagonal Λ in Eq. (3.4), but rather a block diagonal matrix [16]. An example of such a case will be given in Sect. 3.4.

3.2 Illustrative Example

To illustrate the method, we explore the behaviour of a simple ecological model of two interacting populations, labelled X and Y. Individuals of both populations reproduce with rate one, and there is a small probability μ of the offspring mutating from one type to the other. The organisms also prey on each other with rate ε and have a slight preference p for prey of the opposite type. The model is most intuitively expressed in the notation of chemical reaction systems (see Eq. (2.34)):

$$\text{Reproduction:} \quad X \xrightarrow{1-\mu} X + X, \quad Y \xrightarrow{1-\mu} Y + Y$$

$$\text{Mutation:} \quad X \xrightarrow{\mu} X + Y, \quad Y \xrightarrow{\mu} X + Y$$

$$\text{Predation:} \ X + Y \xrightarrow{\varepsilon(1/2+p)} X, \quad X + Y \xrightarrow{\varepsilon(1/2+p)} Y$$

$$\text{Cannibalism:} \ X + X \xrightarrow{\varepsilon(1/2-p)} X, \quad Y + Y \xrightarrow{\varepsilon(1/2-p)} Y. \tag{3.9}$$
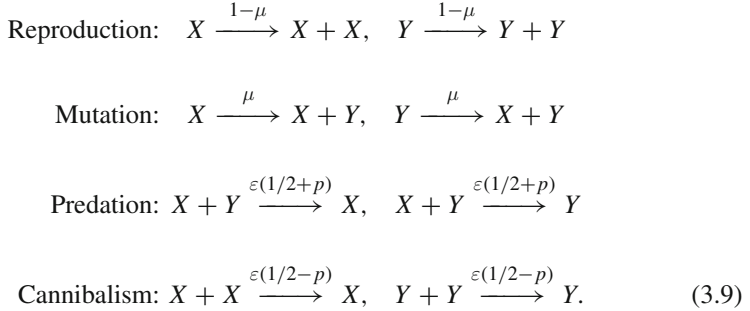
We recall that arrows denote possible reactions and the values above them are the rate constants. Writing n_X and n_Y for the number of individuals in each population, the transition rates $T(\boldsymbol{n}|\boldsymbol{n}')$ can be derived from the reactions (3.9) using Eqs. (2.34), (2.35). They are

$$
\begin{aligned}
T(n_X + 1, n_Y | n_X, n_Y) &= (1 - \mu)n_X + \mu n_Y \\
T(n_X, n_Y + 1 | n_X, n_Y) &= \mu n_X + (1 - \mu)n_Y \\
T(n_X - 1, n_Y | n_X, n_Y) &= \varepsilon(1/2 - p)n_X^2 + \varepsilon(1/2 + p)n_X n_Y \\
T(n_X, n_Y - 1 | n_X, n_Y) &= \varepsilon(1/2 + p)n_X n_Y + \varepsilon(1/2 - p)n_Y^2. \tag{3.10}
\end{aligned}
$$

The model may now be mathematically formulated as the master equation (2.18), which describes the time evolution of the probability distribution $P(n_X, n_Y, t)$. Stochastic simulations of the model can be performed using the Gillespie algorithm [7] described in Sect. 2.3.

When the predation rate ε is small, the population may grow very large. The parameter ε is thus analogous to the parameter N^{-1} utilised in the expansion of the master equation in Sect. 2.4. One can therefore introduce the scaled variables $x = \varepsilon n_X$ and $y = \varepsilon n_Y$ and perform this expansion to yield an effective description of the system in terms of an FPE. In order to derive the precise form of the FPE, one first expresses the transition rates given in Eq. (3.10) in terms of the continuous x and y variables. This yields a set of functions $f_\mu(x, y)$ for $\mu = 1, \ldots, 4$. Obtaining the stoichiometry matrix from Eq. (3.9), the drift and diffusion terms in the FPE can then be read off from Eqs. (2.31) and (2.32) respectively. As discussed in Sect. 2.6,

this FPE is entirely equivalent to a set of SDEs. For the present model, we find the following pair of equations:

$$\frac{dx}{dt} = x - \mu(x - y) - x\left[\frac{1}{2}(x + y) - p(x - y)\right] + \sqrt{\varepsilon}\eta_x(t),$$

$$\frac{dy}{dt} = y + \mu(x - y) - y\left[\frac{1}{2}(x + y) + p(x - y)\right] + \sqrt{\varepsilon}\eta_y(t),$$

(3.11)

where η_x and η_y have the correlation structure specified in Eq. (2.69), with

$$B = \begin{pmatrix} x + \frac{1}{2}x(x + y) - (px + \mu)(x - y) & 0 \\ 0 & y + \frac{1}{2}y(x + y) + (py + \mu)(x - y) \end{pmatrix}.$$

(3.12)

We begin by examining the deterministic system found by putting $\varepsilon = 0$. There is a trivial fixed point at $x^* = 0$, $y^* = 0$, representing the extinct state, which is always unstable. There is a second fixed point at $x^* = 1$, $y^* = 1$, representing equal coexistence of the two populations. This state is stable when $p < \mu$. If p is raised above μ, a bifurcation occurs, with the equal coexistence fixed point becoming unstable and giving rise to a symmetric pair of stable fixed points in which one species dominates the other. The new fixed points have coordinates

$$x^* = \frac{1 - 2\mu \pm \sqrt{(1 - \mu/p)(1 - 2\mu)}}{1 - 2p}, \quad y^* = \frac{1 - 2\mu \mp \sqrt{(1 - \mu/p)(1 - 2\mu)}}{1 - 2p}.$$

(3.13)

We are interested in examining the effect of noise near this transition.

The eigenvalues of the Jacobian at the coexistence state are $\lambda^{(1)} = 2(p - \mu)$ and $\lambda^{(2)} = -1$, with corresponding right-eigenvectors $\boldsymbol{v}^{(1)} = (1, -1)$ and $\boldsymbol{v}^{(2)} = (1, 1)$. If $|\lambda^{(1)}| \ll |\lambda^{(2)}|$ then we have a slow subspace in the direction of $(x - y)$, meaning that perturbations to the balance of populations evolve very slowly. As described in Sect. 3.1, we approximate the slow subspace by the surface on which the rate of change in the direction of the fast eigenvector is zero. In the present model, the slow subspace is approximated by $dx/dt + dy/dt = 0$ which yields the hyperbola

$$(x + y) - \frac{1}{2}(x + y)^2 + p(x - y)^2 = 0.$$

(3.14)

The two plots in Fig. 3.1 capture the typical behaviour of the model for parameters either side of the bifurcation.

The SDE system (3.11) is two-dimensional, non-linear and has noise correlations which depend on the state of the system. These factors combine to make the theoretical analysis of the model very difficult. The situation is not hopeless, however, as it
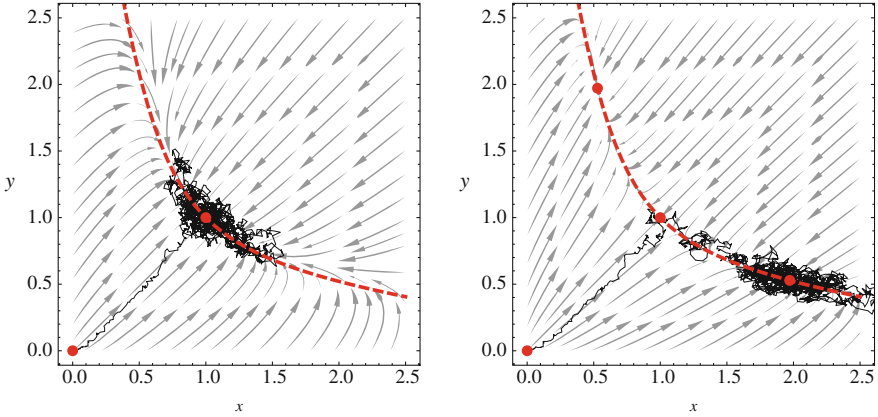
Fig. 3.1 These plots show the behaviour of the illustrative ecological model on either side of the bifurcation. The fixed points are shown as *red circles*, and the *dashed red line* is the nullcline $dx/dt + dy/dt = 0$, given in Eq. (3.14). Qualitatively, it can be seen to give a good approximation to the observed slow subspace. The *grey arrows* show trajectories of the deterministic system, while the *black line* traces out the trajectory of a single (short) stochastic simulation of the individual-based model, starting close to the origin. The parameters are $\varepsilon = 0.005$ and $p = 0.3$ in both plots, while $\mu = 0.35$ on the *left* and $\mu = 0.25$ on the *right*

is clearly visible from Fig. 3.1 that the system state does not typically stray very far from the slow subspace. It is this fact which we aim to exploit in order to produce an 'effective' one-dimensional description of the model. The plan of attack is as follows: first we will make a coordinate transform to separate the fast and slow variables; then we will examine the behaviour of the slow variable under the assumption that the fast variable relaxes instantaneously to its nullcline value.

We introduce $z = (x - y)/2$ and $w = (x + y)/2$, so that

$$\begin{pmatrix} z \\ w \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \equiv V^{-1} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{3.15}$$

where the matrix V is the matrix whose ith column is the ith right-eigenvector of the system's Jacobian, as discussed in Sect. 3.1. In the new coordinates Eq. (3.11) becomes

$$\begin{aligned} \frac{dz}{dt} &= z \left[1 - 2\mu - (1 - 2p)\, w \right] + \sqrt{\varepsilon} \kappa_z(t), \\ \frac{dw}{dt} &= w - w^2 + 2pz^2 + \sqrt{\varepsilon} \kappa_w(t), \end{aligned} \tag{3.16}$$

and Eq. (3.14) can be rewritten $w - w^2 + 2pz^2 = 0$.

To determine the correlation structure of the new noise variables κ_z and κ_w, a general result on Gaussian random variables is applied. Suppose that a vector of

Gaussian random variables η has correlation matrix B, and that $\kappa = V^{-1}\eta$ for some matrix V^{-1}. What is the correlation matrix of κ? Well,

$$\langle \kappa_i \kappa_j \rangle = \left\langle \sum_{k,l} V_{ik}^{-1}\eta_k V_{jl}^{-1}\eta_l \right\rangle = \sum_{k,l} V_{ik}^{-1}\langle \eta_k \eta_l \rangle V_{lj}^{-1} = \tilde{B}_{ij}, \tag{3.17}$$

where $\tilde{B} = [V^{-1}]B[V^{-1}]^T$. In the present case, the matrices B and V^{-1} are given in Eqs. (3.12) and (3.15), respectively. The following correlation matrix for κ_z and κ_w is thus found:

$$\tilde{B} = \frac{1}{2}\begin{pmatrix} (w + w^2 - 2pz^2) & z\,[1 - 2\mu + (1 - 2p)w] \\ z\,[1 - 2\mu + (1 - 2p)w] & (w + w^2 - 2pz^2) \end{pmatrix}. \tag{3.18}$$

Notice that whilst the original noise variables η_x and η_y were independent (the off-diagonal entries of B were zero), the noise variables coordinates are correlated with each other.

To enforce the assumed separation of time-scales between z and w, the following conditions are imposed:

$$w(z) = \frac{1}{2}\left(1 + \sqrt{1 + 8pz^2}\right) \quad \text{and} \quad \kappa_w(t) \equiv 0. \tag{3.19}$$

The first of these sets w to its value on the nullcline for a given z, whilst the second removes the possibility of any noise-induced fluctuations. Since κ_w and κ_z are correlated, imposing $\kappa_w = 0$ will alter the statistical distribution of κ_z. What effect do these constraints have on the evolution of z? Substituting $w(z) = (1+\sqrt{1 + 8pz^2})/2$ into (3.16) to remove the dependence on w, and denoting noise drawn from the altered distribution by ζ, we have

$$\frac{dz}{dt} = \bar{A}(z) + \sqrt{\varepsilon}\zeta(t), \tag{3.20}$$

where

$$\bar{A}(z) = z\left(1 - 2\mu - \left(\frac{1}{2} - p\right)\left(1 + \sqrt{1 + 8pz^2}\right)\right), \tag{3.21}$$

and

$$\langle \zeta(t)\zeta(t') \rangle = \bar{B}(z)\delta(t - t'). \tag{3.22}$$

We must now determine the effect of the conditions (3.19) on the correlation structure of ζ, $\bar{B}(z)$.

Again a general result on correlated Gaussian random variables is applied (see Appendix A). Suppose that a collection of Gaussian random variables κ has correlation matrix $\tilde{B}$. Let $\bar{B}$ be the correlation matrix of κ_z conditioned on the event that $\kappa_w = 0$. Then utilising Eq. (A.28) we find

$$\bar{B} = \tilde{B}_{11} - \frac{\tilde{B}_{21}\tilde{B}_{12}}{\tilde{B}_{22}}. \qquad \text{(Conditioning method)}. \qquad (3.23)$$

This is the correlation structure of ζ, which we can describe as the noise term κ_z conditioned on $\kappa_w = 0$. Applying formula (3.23) to the correlation matrix found in (3.18), we obtain

$$\bar{B}(z) = \frac{1}{2}\left(w(z) + w(z)^2 - 2pz^2\right)\left[1 - z^2\left(\frac{1 - 2\mu + w(z)(1 - 2p)}{w(z) + w(z)^2 - 2pz^2}\right)^2\right],$$
$$(3.24)$$

and where of course $w(z) = (1 + \sqrt{1 + 8pz^2})/2$.

Equations (3.20) and (3.22) together define a one-dimensional stochastic differential equation. Although it may look a little complicated, being one-dimensional means that we can make analytic progress. The single-variable SDE can be mapped back onto a one-dimensional FPE, and the methods discussed in Sect. 2.5 can be used to answer questions about the behaviour of the system.

The full system has an absorbing boundary at $(x, y) = (0, 0)$ (or equivalently $(z, w) = (0, 0)$), while the remainder of the lines $x = 0$ and $y = 0$ ($z = -w$ and $y = w$) are reflecting. The system is otherwise unbounded. In the reduced description, the system is restricted to the slow subspace. The boundaries of the reduced problem are thus the points at which the slow subspace defined by Eq. (3.19) intersects the boundaries of the full problem. In this case, the slow subspace intersects the reflecting boundaries $z = -w$ and $z = w$. Denoting the lower bound of z by a_1, we find that it must satisfy

$$\frac{1}{2}(1 + \sqrt{1 + 8pa_1^2}) = -a_1,$$
$$\implies a_1 = \frac{1}{2p - 1}. \qquad (3.25)$$

The upper bound a_2 may be calculated in a similar way, resulting in $a_2 = -a_1$. We now have all the information we need to calculate the stationary distribution of z using Eq. (2.38).

In Fig. 3.2, the analytical prediction of the equation for the stationary distribution is compared with a histogram of z-coordinates obtained from Gillespie simulations using the same parameters as in Fig. 3.1. Clearly, the reduced one-dimensional model provides a very good fit to the data coming from the individual-based simulation. In Fig. 3.3, we use similar parameters to those in Figs. 3.1 and 3.2, albeit with the parameter ε increased to $\varepsilon = 0.08$, to show that the approximation continues to work well in systems with large noise. It should also be pointed out that although the theory has been developed based on the local behaviour around the coexistence fixed point $(1, 1)$, the approximation remains successful even in the unstable regime.
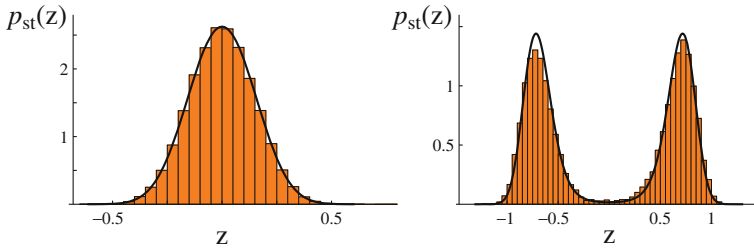
Fig. 3.2 The stationary distribution of $z = x - y$ as predicted by the reduced-dimension model (*black curve*) and measured from a single long simulation run of the individual-based model (*orange* histogram). The parameters are the same as those in Fig. 3.1. The theoretical prediction was obtained by numerically integrating Eq. (2.38), with drift term (3.21), diffusion term (3.24), lower boundary (3.25) and parameters taken from Fig. 3.1. Data points were collected from 10,000 simulations run at time intervals of 100
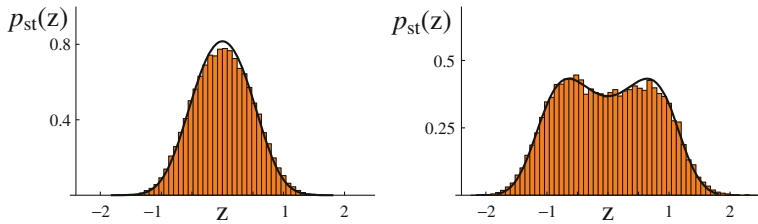

Fig. 3.3 The stationary distribution of z in the illustrative example. Once more the *black curve* is calculated from the reduced model while the *orange* histogram is plotted with data obtained from Gillespie simulation. The parameters are the same as those in Fig. 3.2 with the exception of ε which is now given by $\varepsilon = 0.08$. The approximation continues to work well even in this regime of larger noise

3.3 General Formulation

In this section, a description of the method for an arbitrary m-species IBM is provided. The model is first formulated as an m-dimensional master equation which describes the evolution of the distribution $P(n_1, \ldots, n_m, t)$ (see Sect. 2.2). If the total number of individuals $N = \sum_{i=1}^{m} n_i$ is large, it is natural to expand in the small parameter $\varepsilon = 1/N$ as detailed in Sect. 2.4, to arrive at an FPE in variables $x_i = n_i/N$. This FPE can alternatively be expressed as an SDE of type Eq. (2.67). It is in this setting that the reduction technique is applied.

Suppose we are interested in behaviour around a deterministic fixed point x^*. Let J be the Jacobian of $A(x)$ at that point and write $\lambda^{(1)}, \ldots, \lambda^{(r)}, \ldots, \lambda^{(m)}$ for its eigenvalues. Suppose further that $\lambda^{(r+1)}, \ldots, \lambda^{(m)}$ are non-degenerate, real, negative and large, so that the eigenvalues obey the inequalities given in Eq. (3.2). The associated right-eigenvectors $v^{(r+1)}, \ldots, v^{(m)}$ represent very stable directions. We aim to eliminate fluctuations in these directions to produce a reduced-dimension model. We proceed as in the deterministic case (Sect. 3.1) by first making a change of variables

from the m-vector x to an r-vector z and an $(m - r)$-vector w, via the coordinate transformation Eq. (3.5). In the new coordinates the SDE becomes

$$\frac{d}{dt}\begin{pmatrix} z \\ w \end{pmatrix} = V^{-1}A\left(V\begin{pmatrix} z \\ w \end{pmatrix}\right) + \frac{1}{\sqrt{N}}\kappa(t), \tag{3.26}$$

where

$$\left\langle \kappa_i(t)\kappa_j(t')\right\rangle = \delta(t - t')\left[\tilde{B}(z, w)\right]_{ij}, \tag{3.27}$$

and $\tilde{B} = [V^{-1}]B[V^{-1}]^T$. We wish to constrain w to the slow subspace which we approximate by its nullcline, $w(z)$, defined by Eq. (3.8). To enforce the assumed separation of time-scales between w and the other variables, we impose the following conditions:

$$w = w(z) \quad \text{and} \quad \kappa_i(t) \equiv 0 \quad i = r + 1, \ldots m. \tag{3.28}$$

In order to implement the method, $w(z)$ must be obtained analytically. However, beginning from an IBM of the type described in Sect. 2.2, the resulting SDE system will be polynomial in the state variables, typically of low degree. When the solution for $w(z)$ is not unique one may frequently eliminate alternative solutions based on stability arguments. Notice that we are taking a static description of the slow subspace. While it has been shown that a deterministic slow subspace does not, in general, converge to the expectation of the stochastic slow subspace of generic SDEs [10], the discrepancy is typically of the same order as the noise. Since the noise in the IBM derived SDE model is small by construction (see Sect. 2.6), the deviation from the deterministic slow subspace will be negligible.

For the remaining variables, we have

$$\frac{dz}{dt} = \bar{A}(z) + \frac{1}{\sqrt{N}}\zeta(t), \tag{3.29}$$

where

$$\left\langle \zeta_i(t)\zeta_j(t')\right\rangle = \delta(t - t')\left[\bar{B}(z)\right]_{ij}, \quad i, j = 1, \ldots r. \tag{3.30}$$

The drift vector $\bar{A}(z)$ and diffusion matrix $\bar{B}(z)$ are derived from A and $\tilde{B}$ as follows. For $i = 1, \ldots, r$

$$\left[\bar{A}(z)\right]_i = \left[V^{-1}A\left(V\begin{pmatrix} z \\ w(z) \end{pmatrix}\right)\right]_i. \tag{3.31}$$

The general solution for the conditioned noise covariance matrix can be shown to be

$$\bar{B}(z) = \mathcal{B}_{11}(z) - \mathcal{B}_{12}(z)\mathcal{B}_{22}^{-1}(z)\mathcal{B}_{21}(z), \tag{3.32}$$

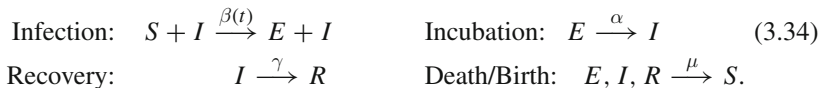where $\mathcal{B}$ are matrices of the partitioned $\tilde{B}$ matrix;

$$\tilde{B} = \begin{pmatrix} \mathcal{B}_{11}(z) & \mathcal{B}_{12}(z) \\ \mathcal{B}_{21}(z) & \mathcal{B}_{22}(z) \end{pmatrix}, \tag{3.33}$$

with $\mathcal{B}_{11}(z)$ an $r \times r$ matrix and $\mathcal{B}_{22}(z)$ an $(m - r) \times (m - r)$ matrix (see Appendix A). Equations (3.29) and (3.30) describe a reduced-dimension stochastic system in which the fast directions associated with the eigenvalues $\lambda^{(r+1)}, \ldots \lambda^{(m)}$ have been eliminated.

3.4 Application: Seasonally Forced Epidemics

3.4.1 Model Definition and Deterministic Treatment

The SEIR model is a simplified epidemiological model describing the spread of a disease through a population [9]. Members of the population may be in one of four states: susceptible (S), exposed (E), infectious (I) and recovered (R). The susceptible individuals come into contact with the infected and become exposed with infection rate $\beta(t)$, which may vary with time. The choice of a time-dependent infection rate is often made in epidemiological models, in order to incorporate seasonal changes in the contact rate between individuals. This is known as seasonal forcing. For instance, in childhood diseases, the forcing can be representative of periods of school attendance and holidays, while in general it may be brought about by the variation in time spent indoors in crowded environments between winter and summer. Those exposed to the disease then become infectious with a rate of disease onset α. Finally, the infectious recover with an average rate of γ. In addition to these disease dynamics, there is a constant birth and death rate μ; it is traditional to hold the population size constant by treating death and birth as a single process whereby an individual returns to the susceptible state. As in the earlier illustrative model, the dynamics may be conveniently summarised using the notation of chemical reactions:

$$
\begin{array}{llll}
\text{Infection:} & S + I \xrightarrow{\beta(t)} E + I & \text{Incubation:} & E \xrightarrow{\alpha} I \qquad (3.34) \\
\text{Recovery:} & I \xrightarrow{\gamma} R & \text{Death/Birth:} & E, I, R \xrightarrow{\mu} S.
\end{array}
$$

The variables n_S, n_E, n_I, n_R are used to denote the number of individuals in states S, E, I and R, respectively. The total population size is then given by $N = n_S + n_E + n_I + n_R$, which does not vary, meaning that there are three degrees of freedom, $m = 3$. With just a slight abuse of notation we introduce variables $S = n_S/N$, $E = n_E/N$ and $I = n_I/N$ which describe the population density of individuals in each disease state. Note that there is no need for a variable associated to the recovered state, since the conservation of total population makes it a dependent variable. Applying the

same master equation expansion as before (see Sect. 2.4 for details), the following
SDE system is obtained:

$$
\frac{dS}{dt} = \mu(1 - S) - \beta(t)SI + \frac{1}{\sqrt{N}}\eta_1(t),
$$
$$
\frac{dE}{dt} = \beta(t)SI - (\mu + \alpha)E + \frac{1}{\sqrt{N}}\eta_2(t),
$$
$$
\frac{dI}{dt} = \alpha E - (\mu + \gamma)I + \frac{1}{\sqrt{N}}\eta_3(t),
$$

(3.35)

where $\eta_{1,2,3}$ are Gaussian white noise variables with correlations

$$
\left\langle \eta_i(t)\eta_j(t') \right\rangle = \delta(t - t')B_{ij},
$$

$$
B = \begin{pmatrix}
\mu(1 - S) + \beta(t)SI & -\mu E + \beta(t)SI & -\mu I \\
-\mu E + \beta(t)SI & \beta(t)SI + (\mu + \alpha)E & -\alpha E \\
-\mu I & -\alpha E & \alpha E + (\mu + \gamma)I
\end{pmatrix}.
$$

(3.36)

In this section, the behaviour of the model in the deterministic limit $N \to \infty$ is
discussed, before the full stochastic system is considered in Sect. 3.4.2. When the
infection rate is not seasonally forced, so $\beta(t) \equiv \beta$, there are a pair of fixed points.
The first of these represents the extinction of the disease: $S = 1$, $E = 0$, $I = 0$. The
second fixed point is

$$
S^* = \frac{(\alpha + \mu)(\gamma + \mu)}{\alpha\beta}, \quad E^* = \frac{\mu(1 - S^*)}{\alpha + \mu}, \quad I^* = \frac{\alpha\mu(1 - S^*)}{(\alpha + \mu)(\gamma + \mu)}, \quad (3.37)
$$

which is referred to as the endemic state. We are concerned with the regime in which
the endemic state is stable and the extinct state is unstable, which holds for a range
of epidemiologically realistic parameter values. In general, the rate parameter μ
controlling birth and death will be much smaller than the remaining rate parameters
β, α and γ, since this takes place on a much longer timescale than the disease
dynamics. The parameter μ can therefore be utilised as an expansion parameter to
simplify some of the expressions in the analysis. To first order in μ, the eigenvalues
of the Jacobian at the endemic state are the complex-conjugate pair

$$
\lambda^{(1,2)} = -\mu\frac{\alpha^2\beta + \beta\gamma^2 + \alpha\gamma(\beta + \gamma)}{2\gamma(\alpha + \gamma)^2} \pm i\sqrt{\mu\frac{\alpha(\beta - \gamma)}{\alpha + \gamma}}, \quad (3.38)
$$

and

$$
\lambda^{(3)} = -(\alpha + \gamma) - \mu\frac{\alpha(2\alpha + \beta) + 3\alpha\gamma + 2\gamma^2}{(\alpha + \gamma)^2}. \quad (3.39)
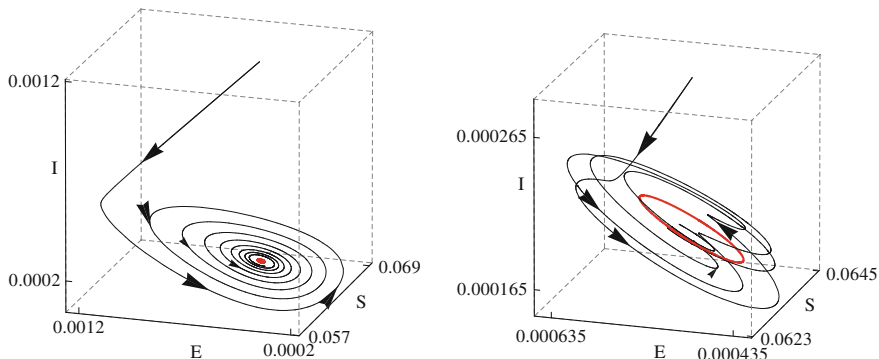$$

Fig. 3.4 Deterministic trajectories for the SEIR model. *Left* No seasonal forcing ($\delta = 0$), fixed point shown in *red*. *Right* System in the presence of forcing, ($\delta = 0.02$), limit cycle highlighted in *red*. Remaining parameters in both plots are $\beta_0 = 1575$ year^{-1}, $\alpha = 35.84$ year^{-1}, $\gamma = 100$ year^{-1} and $\mu = 0.02$ year^{-1}, which are standard choices for measles [13, 14]

Notice that a separation of time-scales exists: $\mathrm{Re}[\lambda^{(3)}] \ll \mathrm{Re}[\lambda^{(1,2)}]$, meaning that $\lambda^{(3)}$ corresponds to a highly stable direction. In addition, since μ is small, the imaginary parts of $\lambda^{(1,2)}$ are typically larger than the real parts, meaning that we may expect highly oscillatory trajectories in the neighbourhood of the endemic state. We can thus expect the system to first collapse rapidly in the direction of the third eigenvector, followed by a slow, almost-planar, spiralling decay to the endemic state, as shown in Fig. 3.4. This separation of timescales has been previously noted and exploited in the deterministic setting [15] and the stochastic setting using normal form techniques [6]. However the stochastic analysis in [6] considered a quite different system (additive noise on an unforced system which was not derived from a microscopic IBM), with very different objectives (namely the replication of stochastic trajectories).

Introducing seasonal forcing of the infection rate creates an additional layer of complexity. A typical choice would be

$$\beta(t) = \beta_0 \Big(1 + \delta \cos(2\pi t) \Big), \tag{3.40}$$

where β_0 describes the basal infection rate, δ is the forcing amplitude, and time is measured in years. The deterministic system will now not settle to the endemic state, but instead exhibit limit cycle behaviour. For the sake of simplicity, we will consider only parameter values which result in a single stable limit cycle of period $T = 1$ year.

Similar to linear stability analysis of fixed points, there is a well-developed theory for analysing perturbations around limit cycles. Writing $(S^*(t), E^*(t), I^*(t))$ for the limit cycle, we introduce the vector

$$\boldsymbol{\xi}(t) = \sqrt{N} \begin{pmatrix} S(t) - S^*(t) \\ E(t) - E^*(t) \\ I(t) - I^*(t) \end{pmatrix}, \tag{3.41}$$

where the factor $\sqrt{N}$ has been introduced for notational consistency with later sections. Since we are effectively linearising the system, it should be noted that this factor has no effect on the following analysis. To first order then, analogous to Eq. (2.74), the dynamics of $\boldsymbol{\xi}$ are governed by

$$\frac{d\boldsymbol{\xi}}{dt} = J(t)\boldsymbol{\xi}, \tag{3.42}$$

where J is the time-dependent Jacobian of Eq. (3.35). This equation may be solved using Floquet theory [8]. The key result of the theory states that solution trajectories may be decomposed into the product of a periodic vector with an exponentially growing/decaying amplitude. General solutions are then of the form

$$\boldsymbol{\xi}(t) = \sum_{i=1}^{m} c_i \boldsymbol{q}^{(i)}(t) e^{\sigma^{(i)} t}, \tag{3.43}$$

where m is the number of degrees of freedom in the model, c_i is a constant, $\boldsymbol{q}^{(i)}(t)$ a periodic vector with the same period T as the limit cycle, and the value $\sigma^{(i)}$ determining the rate of growth/decay is referred to as the Floquet exponent. Akin to the eigenvalues of the fixed point Jacobian, Floquet exponents are indicative of the stability of the limit cycle in the time-varying directions $\boldsymbol{q}^{(i)}(t)$; perturbations to the trajectory will grow if $\text{Re}[\sigma^{(i)}] > 0$ and decay if $\text{Re}[\sigma^{(i)}] < 0$.

Although this formalism may be carried through for much of the calculation analytically, ultimately the Floquet exponents and periodic vectors will be obtained numerically. Of course, other techniques exist to study driven systems [2, 5] and Floquet exponents may be obtained in some cases through asymptotic expansions in small amplitudes, but we wish to align our approach to the previous work on the system which we will investigate below [13]. The procedure (detailed in Appendix B) requires first computing a matrix whose columns are the independent solutions to Eq. (3.42), $X(t) = (\boldsymbol{\xi}_1(t) \cdots \boldsymbol{\xi}_n(t))$, and then determining the eigenvalues of $X(T)$. If there is a rapid collapse along a stable direction, followed by slow dynamics in a subspace, then the columns of $X(T)$ will be almost linearly dependent, since all solution trajectories quickly move towards the subspace. In turn, the real parts of eigenvalues of the matrix $X(T)$ will differ by many orders of magnitude. Obtaining these eigenvalues accurately is crucial to the remaining analysis. However, if the disparity between the eigenvalues is too great, numerical procedures may not maintain sufficient accuracy in calculations involving the eigenvalues. This problem was previously highlighted in [13] within an analysis of the seasonally forced SEIR model. There, using the same epidemiological parameters as used in Fig. 3.4, the eigenvalues of $X(T)$ differed by a factor of 10^{59}. This prompted the authors to implement arbitrary precision numerical methods, at a considerable cost in computing time. It is proposed that the general method detailed in Sect. 3.3 can be used as a technique to remove this fast direction and hence circumnavigate the numerical difficulties encountered in the analysis.

3.4.2 Stochastic Treatment Exploiting the Slow Manifold

Beginning with the unforced case, we let $\lambda^{(1)}$, $\lambda^{(2)}$ and $\lambda^{(3)}$ be as in Eqs. (3.38) and (3.39) and write $\boldsymbol{v}^{(1)}$, $\boldsymbol{v}^{(2)}$, $\boldsymbol{v}^{(3)}$ for the corresponding eigenvectors. We introduce the transformation matrix $V = ((\boldsymbol{v}^{(1)} + \boldsymbol{v}^{(2)}) \quad i(\boldsymbol{v}^{(1)} - \boldsymbol{v}^{(2)}) \quad \boldsymbol{v}^{(3)})$ and new variables

$$\begin{pmatrix} z_1 \\ z_2 \\ w \end{pmatrix} = V^{-1} \begin{pmatrix} S \\ E \\ I \end{pmatrix}. \tag{3.44}$$

The Jacobian of the transformed system at the endemic fixed point takes the form

$$\bar{J} = \begin{pmatrix} L & 0 \\ 0 & \lambda^{(3)} \end{pmatrix} + \mathcal{O}(\mu^{3/2}), \quad \text{where} \quad L = \begin{pmatrix} \mathrm{Re}[\lambda^{(1)}] & \mathrm{Im}[\lambda^{(2)}] \\ \mathrm{Im}[\lambda^{(1)}] & \mathrm{Re}[\lambda^{(2)}] \end{pmatrix}. \tag{3.45}$$

The slow subspace for w is determined by transforming Eq. (3.35) into the form Eq. (3.7) (with A taken from Eq. (3.35)) and applying Eq. (3.8) to obtain $w = w(z_1, z_2)$. An explicit form for $w(z_1, z_2)$ can be found, though the expression is more clearly expressed in S, E and I as

$$I = \frac{\alpha(\alpha + \gamma)E}{\gamma^2 + \alpha\beta_0 S} + \mathcal{O}(\mu),$$

to leading order in μ. To capture the effects of stochasticity, we introduce variables describing the fluctuations in the new coordinates, rescaled by a factor of $\sqrt{N}$,

$$\bar{\boldsymbol{\xi}} = \sqrt{N} \begin{pmatrix} \boldsymbol{z} - \boldsymbol{z}^* \\ w - w^* \end{pmatrix}. \tag{3.46}$$

Making this substitution in (3.35) and keeping only first order terms in $1/N$ and μ, we find that $\bar{\boldsymbol{\xi}}$ obeys

$$\frac{d\bar{\boldsymbol{\xi}}}{dt} = \bar{J}\bar{\boldsymbol{\xi}} + \boldsymbol{\kappa}(t), \tag{3.47}$$

where

$$\langle \kappa_i(t)\kappa_j(t') \rangle = \delta(t - t')\tilde{B}_{ij}, \quad i, j = 1, 2, 3. \tag{3.48}$$

The matrix $\tilde{B}$ is given by

$$\tilde{B} = \left[V^{-1}\right] B \left[V^{-1}\right]^T \Big|_{(S,E,I)=(S^*,E^*,I^*)}, \tag{3.49}$$

where B is taken from Eq. (3.36). Note that applying the constraint $w = w(z_1, z_2)$ induces the relationship

$$w^* + \frac{\bar{\xi}_3}{\sqrt{N}} = w\left(z_1^* + \frac{\bar{\xi}_1}{\sqrt{N}}, z_2^* + \frac{\bar{\xi}_2}{\sqrt{N}}\right). \tag{3.50}$$

Expanding once more in large N, we find

$$\bar{\xi}_3 = \left[\bar{\xi}_1 \frac{\partial w(z)}{\partial z_1} + \bar{\xi}_2 \frac{\partial w(z)}{\partial z_2}\right]\bigg|_{z=z^*}. \tag{3.51}$$

After elimination of the fast direction, Eq. (3.47) becomes

$$\frac{d}{dt}\begin{pmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{pmatrix} = L\begin{pmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{pmatrix} + \begin{pmatrix} \zeta_1(t) \\ \zeta_2(t) \end{pmatrix}, \tag{3.52}$$

where ζ_1 and ζ_2 have correlation matrix $\bar{B}$, which is related to $\tilde{B}$ by Eq. (3.32).

We move on now to study the situation of seasonally forced infection rate. In principle, the calculations above apply only in the limit of small forcing amplitude (that is, $\delta \to 0$ in Eq. (3.40)). We learnt in the illustrative ecological example, however, that although our theory is developed to apply in the locality of a stable fixed point, it can continue to provide a useful approximation if this condition does not strictly hold. Applying that lesson to the present case, we modify Eq. (3.52) to allow L and $\bar{B}$ to become functions of time, as dictated by the replacement $\beta \mapsto \beta(t)$. Essentially, we are approximating the limit cycle by its components on the slow subspace of the endemic fixed point of the unforced model, and then demanding that any stochastic fluctuations remain on this slow subspace. The application of such a coordinate change to the forced system can be further motivated by a consideration of the periodic matrix $J(t)$ in Eq. (3.42) in the limit of small forcing. While our description of the slow subspace is time-independent, the possibility of a dynamic slow subspace has been explored in the deterministic setting [4].

A Floquet analysis of the deterministic part of the reduced system finds two complex conjugate Floquet multipliers, with the third disparate multiplier having been eliminated from the system. This system no longer suffers from the numerical difficulties which plague the full three-dimensional model.

To quantify the effect of stochastic fluctuations in this model, we follow the standard procedure of computing the autocorrelation matrix $\mathcal{C}(\tau)$ of oscillations around the limit cycle, which has entries

$$\left[\mathcal{C}(\tau)\right]_{ij} = \frac{1}{T}\int_0^T \left\langle \bar{\xi}_i(t)\bar{\xi}_j(t+\tau)\right\rangle dt. \tag{3.53}$$
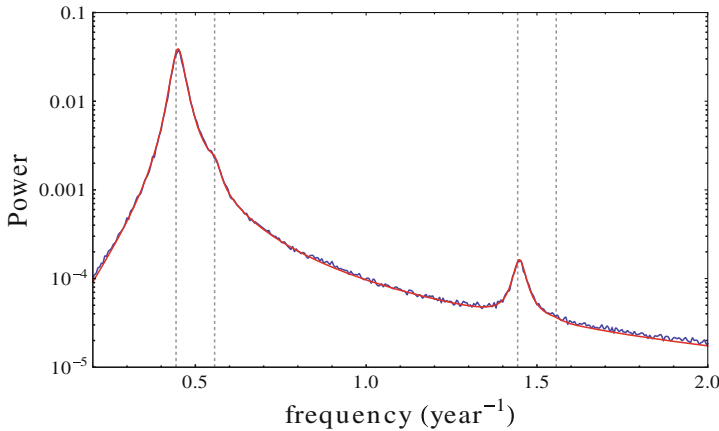
Fig. 3.5 Power spectra for the fluctuations in the number of infected individuals about the limit cycle. The analytical power spectrum calculated using the slow subspace approximation is plotted in *red*, while the power spectrum from stochastic simulations is in *blue*. Agreement is such that the spectra are difficult to distinguish; the spectrum from simulated results is primarily discernible through its stochastic nature relative to the smooth analytical line. *Dotted lines* indicate the position of the peaks in the power spectra given by Eq. (3.54). Epidemiological parameters are $\beta_0 = 1575$ year^{-1}, $\alpha = 35.84$ year^{-1}, $\gamma = 100$ year^{-1} and $\mu = 0.02$ year^{-1}. The simulated spectrum is calculated as the average power spectrum of 1000 stochastic realisations, each lasting 200 years with a system size of $N = 10^8$

Of course our reduced system (3.52) is two-dimensional, meaning that the entries of $\mathcal{C}$ pertaining to $\bar{\xi}_3$ must be deduced from Eq. (3.51). The coordinate transformation applied at the start in (3.44) may then be inverted to give the autocorrelation matrix for the fluctuations in S, E and I.

The Fourier transform of the diagonal entries of the autocorrelation gives the power spectrum of oscillations, which provides a convenient visualisation of the stochastic fluctuations. In Fig. 3.5 we plot the power spectrum of fluctuations in the number of infected individuals around the limit cycle, comparing the stochastic simulations and the theoretical prediction using the reduced-dimension model (3.52). The agreement between the simulation and theory can be seen to be excellent; the spectra lie virtually coincident. The peaks in the theoretical spectrum are found at the same positions as those for the simulated spectrum. These are approximately given by

$$\nu_j = \frac{j}{T} \pm \frac{|\mathrm{Im}[\sigma^{(1,2)}]|}{2\pi}, \tag{3.54}$$

where j is an integer and $\sigma^{(i)}$ are the Floquet exponents as defined in Appendix B. This is in agreement with [3, 13] where peaks at these positions were also found. The overall benefit that was garnered by the procedure however was that of computational efficiency. The computation of the theoretical power spectra approximated from the reduced system takes only a small fraction of the computing time of the full system.

3.5 Discussion

In this chapter I have introduced a systematic and general procedure to eliminate fast degrees of freedom in stochastic dynamical systems derived from individual based models. The method is inspired by the highly successful theory of slow manifolds in deterministic systems, from which the basic notion of restricting attention to trajectories occupying a low-dimensional subspace is borrowed. In the stochastic setting, the conditioning method achieves this by enforcing the condition that the system state remains fixed to the nullcline of the fast direction. This condition is applied to SDE systems with correlated Gaussian white noise, so that a lower-dimensional system is obtained in which the fast degree of freedom has been eliminated. Importantly, the reduced system has the same basic form as the original, meaning that complicating factors such as non-Markovian processes or coloured noise have not been inadvertently introduced.

In the first application, a nonlinear, two-dimensional system with multiplicative noise is reduced to a one-dimensional system which is amenable to analysis. The predictions from the model agree extremely well with simulations, as shown in Figs. 3.2 and 3.3. In the second application, that of the seasonally forced SEIR epidemic model, one of three degrees of freedom is removed. While this does not make the nonlinear system any more tractable, it does aid the analysis of the linearised model. In this case, the same timescale separation exploited in the reduction method is responsible for difficulties encountered when numerically calculating the power spectra of the system linearised about the limit cycle. By removing the fast timescale, these difficulties are avoided, with very little loss of precision (see Fig. 3.5).

While the conditioning method is successful in many cases, it is possible to construct systems for which the conditioned dynamics are not qualitatively similar to those of the full system. In particular, if the noise matrix B is singular, then eliminating the noise away from the manifold can lead to a system which is entirely deterministic. It is instructive to consider a simple example of such a case, to demonstrate how and why the method fails. We consider then the simple (if somewhat artificial) linear SDE system with additive noise,

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix}, \tag{3.55}$$

with $\eta(t)$ as usual a zero mean Gaussian white noise process with noise correlation matrix,

$$B = \begin{pmatrix} 0 & 0 \\ 0 & b \end{pmatrix}, \tag{3.56}$$

which is singular. The noise term η_1 is then identically zero. This system was introduced as an illustrative example in [10]. We proceed through the methodology outlined in Sect. 3.3 by identifying the eigenvalues and eigenvectors of the system. We find in fact that a centre manifold exists, since

$$\lambda^{(1)} = 0, \quad \boldsymbol{v}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \lambda^{(2)} = -1, \quad \boldsymbol{v}^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{3.57}$$

The transformed variables are then given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix} \equiv V \begin{pmatrix} z \\ w \end{pmatrix}, \tag{3.58}$$

$$\begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \equiv V^{-1} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{3.59}$$

This leads to the dynamical equations

$$\frac{d}{dt} \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix} + \begin{pmatrix} \kappa_1(t) \\ \kappa_2(t) \end{pmatrix}, \tag{3.60}$$

with noise correlation matrix

$$\tilde{B} = \begin{pmatrix} b & b \\ b & b \end{pmatrix}. \tag{3.61}$$

The fast direction is clearly w, with the centre manifold being the plane $w = 0$. Conditioning the noise to lie on this manifold leads to an equation for z

$$\frac{dz}{dt} = 0, \tag{3.62}$$

since the conditioned noise correlator is

$$\bar{B} = \tilde{B}_{22} - \frac{\tilde{B}_{12}\tilde{B}_{21}}{\tilde{B}_{11}} = 0. \tag{3.63}$$

The conditioning method therefore suggests that there are no dynamics, deterministic or stochastic, in $x = z - w$. Following this example it is clear why the method fails. The noise is conditioned to have no dynamics parallel to w, but if the noise correlator B is singular, conditioning the noise to one dimension effectively kills off any stochastic dynamics. It is perhaps interesting to note that this result is seen in other stochastic fast-variable elimination techniques, as highlighted in [10].

The problem arising in this example is not usually an issue for SDEs derived from IBMs, as their noise matrices are generally non-singular. To see this, note that the noise matrix B may be decomposed as $B = \nu R \nu^T$ (see Eq. (2.32)), where ν is a stoichiometric matrix and R is a diagonal matrix whose entries are the rate coefficients $f_\mu(\boldsymbol{x})$. Since the rate coefficients are generally positive, $\text{rank}(B) = \text{rank}(\nu)$ and thus B is singular if and only if there exists a vector $\boldsymbol{q}$ such that $\nu \boldsymbol{q} = 0$. This is a linear combination of reactants which is unchanged by any reaction, that is, a conservation relation. Thus the rank of the noise matrix B is typically never less than the number of degrees of freedom in the system.

While this problem should not then be relevant for most IBM derived systems, we would ideally like to use a method of fast-variable elimination for which we do not need to be concerned with these issues. This leads to the second of the two elimination schemes, the projection matrix method.

References

1. L. Arnold, *Random Dynamical Systems*, Springer Monographs in Mathematics (Springer, Berlin, 2003)
2. L. Arnold, P. Imkeller, Normal forms for stochastic differential equations. Probab. Theory Relat. Fields **110**, 559–588 (1998)
3. A.J. Black, A.J. McKane, Stochastic amplification in an epidemic model with seasonal forcing. J. Theory Biol. **267**, 85–94 (2010)
4. C. Chicone, Y. Latushkin, Center manifolds for infinite dimensional nonautonomous differential equations. J. Differ. Equ. **141**, 356–399 (1997)
5. P.H. Coullet, C. Elphick, E. Tirapegui, Normal form of Hopf bifurcation with noise. Phys. Lett. A **111**, 277–282 (1985)
6. E. Forgoston, L. Billings, I.B. Schwartz, Accurate noise projection for reduced stochastic epidemic models. Chaos **19**, 043110 (2009)
7. D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. **22**, 403–434 (1976)
8. R. Grimshaw, *Nonlinear Ordinary Differential Equations* (CRC Press, Oxford, 1990)
9. M.J. Keeling, P. Rohani, *Modelling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, 2007)
10. A.J. Roberts, Normal form transforms separate slow and fast modes in stochastic dynamics systems. Phys. A **387**, 12–38 (2008)
11. T. Rogers, A.J. McKane, A.G. Rossberg, Demographic noise can lead to the spontaneous formation of species. Europhys. Lett. **97**, 40008 (2012)
12. T. Rogers, A.J. McKane, A.G. Rossberg, Spontaneous genetic clustering in populations of competing organisms. Phys. Biol. **9**, 066002 (2012)
13. G. Rozhnova, A. Nunes, Stochastic effects in a seasonally forced epidemic model. Phys. Rev. E **82**, 041906 (2010)
14. I.B. Schwartz, Multiple stable recurrent outbreaks and predictability in seasonally forced nonlinear epidemic models. J. Math. Biol. **21**, 347–361 (1985)
15. I.B. Schwartz, H.L. Smith, Infinite subharmonic bifurcation in an SEIR epidemic model. J. Math. Biol. **18**, 233–253 (1983)
16. G. Strang, *Introduction to Linear Algebra*, 4th edn. (Wellesley-Cambridge Press, Wellesley, 2009)
17. S.H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview Press, New York, 2001)
18. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007)

Chapter 4
The Projection Matrix Method

In this section, a second method of fast variable elimination in stochastic systems will be introduced. This method aims to remove the problems encountered in the conditioning method when the noise matrix is singular (see Sect. 3.5). The essential idea is to explicitly remove any contribution to the dynamics from the fast directions and retain only the slow degrees of freedom. Since the methodology follows in a straightforward way from the conditioning method, a verbal description of the method is given in the following section, along with applications to some of the models encountered in Chap. 3. The method is later applied to a Moran model with migration, however for clarity some of the history of models of population genetics featuring migration is given in Sect. 4.2, while the model pertaining to later work is introduced in Sect. 4.3.

4.1 Introduction to the Projection Matrix Method

In the projection matrix method, as with the conditioning method, a stable fixed point is assumed to exist. The system is then linearised about this point. The eigenvalues of the linear system are used to inform us about the existence of a separation of timescales, while the eigenvectors give the fast/slow directions. The slow subspace is then calculated to be the deterministic nullcline of the fast directions, and we evaluate the SDEs on this subspace. Thus far, the treatment is exactly the same as for the conditioning method. The approach differs in its treatment of the noise. In the projection matrix method, the component of the noise in the slow variables is considered to be the only component relevant to the reduced system dynamics. All other contributions are discarded. The approach is similar to that utilised in [22], and the method can be formalised as the application of a linear projection matrix to the SDE (along with the evaluation of the system on the slow subspace). It is from this that the method takes its name. For clarity however, we avoid constructing the projection matrix here.

The idea is best illustrated on a familiar system, the ecological model of Sect. 3.2. To make things particularly transparent we begin by considering the transformed system in variables z and w (see Eq. (3.16)). Recall that the Jacobian in these variables (evaluated at $(x, y) = (1, 1) \implies (z, w) = (0, 1)$) is diagonal with entries corresponding to the eigenvalues $\lambda^{(1)} = 2(p - \mu)$ and $\lambda^{(2)} = -1$. The slow subspace is given by the deterministic solution to $dw/dt = 0$, Eq. (3.19). Now, as stated previously, in the projection matrix method we only wish to consider noise contributions from the z-direction. In the z and w basis, the component of the noise in the slow z-direction is simply $\tilde{B}_{11}$ from Eq. (3.18). The projection matrix method therefore gives the reduced dimension system

$$\frac{dz}{dt} = \bar{A}(z) + \zeta(t), \tag{4.1}$$

with $\bar{A}(z)$ once again given by Eq. (3.21) and the correlation function

$$\langle \zeta(t)\zeta(t') \rangle \equiv \bar{B}(z)\delta(t - t') \tag{4.2}$$
$$= \tilde{B}_{11}(z, w(z)). \qquad \text{(Projection matrix method)} \tag{4.3}$$

While in some ways this seems a rather drastic simplification, it will be seen to fare surprisingly well. Let us compare the predictions made by this reduced model to those obtained the conditioned system, Eq. (3.20). Comparing the solutions for the stationary distributions (evaluating Eq. (2.38)), very little qualitative difference can be discerned (see Fig. 4.1), and both remain a good approximation for the simulated IBM. A distinct advantage of this new method (in this particular case at least) is that the functional form of the noise correlator (4.3) is far simpler; by coincidence, the new noise correlator obtained using the projection matrix method is $\bar{B}(z) = w(z)$, with $w(z)$ taken from Eq. (3.19). This allows for an analytic solution of the stationary distribution
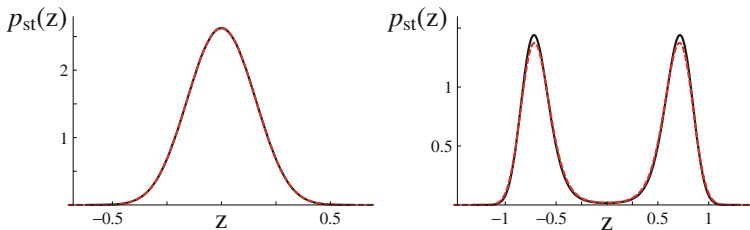
Fig. 4.1 The stationary distribution of the variable z from the illustrative example in Sect. 3.2, using the same parameters as those in Fig. 3.2. The *black line* is obtained from a numerical solution of Eq. (2.38) with drift and diffusion terms taken from Eqs. (3.21) and (3.24), i.e. the reduced system from the conditioning method. The *red dashed line* meanwhile is obtained from Eq. (4.4) (via the projection matrix method) appropriately normalised. In this particular case the results are nearly indistinguishable, however we will see in Sect. 6.2 that this is not always the case

$$p_{\mathrm{st}}(z) \propto [w(z)]^{-((1-2\mu)+2p)/2\varepsilon p} \exp\left[\frac{(1-2\mu)}{\varepsilon p}w(z) - \frac{(1-2p)}{\varepsilon}z^2\right], \quad (4.4)$$

plotted in as the red dashed line in Fig. 4.1 alongside the numerical solution for the stationary solution of the conditioned system (3.20) in black.

Before proceeding to the next section, one more important example is considered; the SDE system (3.55) given in the fast/slow basis by Eq. (3.60). Applying the method discussed above to this system, the reduced variable system

$$\frac{dz}{dt} = \zeta(t) \qquad (4.5)$$

is obtained, with $\langle \zeta(t)\zeta(t)\rangle = b$, a result which retains the qualitative behaviour of the full system (see Sect. 3.5). In many ways it seems surprising that such a straightforward technique would work. However, there are other ways of interpreting the method which are perhaps less jarring. Considering the results of Appendix A it can be seen that the treatment of the noise (essentially ignoring all the components related to the fast directions) is in fact equivalent to taking the marginal distribution of the noise κ_z in Eq. (3.16) or Eq. (3.60), rather than the distribution of κ_z conditioned on $\kappa_w = 0$. Expressed in this way, the success of the method may seem less surprising. However, since this interpretation does not aid the understanding of the mechanics of the method, such a description will not be used in the remainder of this chapter.

In this section an outline of the method has been given, along with two ways in which this arguably more naïve technique can be preferable to the conditioning method. A more quantitative motivation for this approximation will be given in Sect. 6.2, in which the two methods are compared more thoroughly. For now we proceed to apply the projection method to a particular problem, a Moran model with migration. In order to set the model in the appropriate context however, it is best to begin with a brief discussion highlighting the ways in which migration has historically been incorporated into the Moran framework.

4.2 The Moran Model with Migration

Models of population genetics featuring migration were first considered by Wright [27], who looked at what is now often referred to as the standard island model [23]. Instead of the well-mixed population of size N which had been previously studied, he considered a set of $\mathcal{D}$ well-mixed subpopulations. With migratory individuals being chosen from the global population, there was no spatial structure assumed, only interactions between the various subpopulations. These are called demes in the genetic context [13], although we will also use the term islands in this thesis to refer to areas in which there is no spatial structure, but between which interactions can occur. Within the modern nomenclature of ecology, this collection of demes effectively describes a metapopulation [16], or 'population of populations'. The case of one deme therefore should reduce to the well-mixed case.

Subsequently, the formulation of the stepping stone model [15] introduced what was a very simple topology into the description of migration; the islands were ordered, with migration from island i only allowed onto islands $i - 1$ and $i + 1$. Maruyama compared selection in the stepping stone model and in the island model [17]. He concluded that if selection was additive (i.e. frequency-independent) and local, with the same selection pressure in all demes, then deme population structure played no role. In other words, the population behaved approximately as a well-mixed and spatially homogeneous population, with a size equivalent to the sum of the deme population sizes.

The books by Ewens and Moran [9, 18] describe variants of these models, analyses and conclusions, but for our purposes the next result of note is the work of Nagylaki [19] who studied what would be in modern terminology an arbitrary network of demes. He constructed a migratory model with discrete generations (of the Wright-Fisher type) in the limit of strong migration, that is where the probability of a migration event is of the same order as that of a birth or death event. The effect of this assumption was to create a separation of timescales in the Markov chain. Nagylaki then employed his earlier results on Markov chains with timescale separation [8] to derive an equation in the diffusion limit. Starting with a neutral model, it was concluded that in the long-time limit the population behaved as if it were well-mixed, but with an effective population size less than or equal to the total unstructured population. Equality was shown to be achieved if the migration matrix was symmetric.

The analysis which is used to reach these conclusions is however difficult to follow, with some parts of the proof relying on results from the theory of Markov chains and others relying on the nature of the diffusion approximation. Nevertheless, the results of the analysis are widely quoted. The work was extended [19] to the case of different selection strengths on different islands, showing that once again the population behaved as if it were well-mixed with an effective population size, but now also with an effective selection coefficient. The situation where the selection on different islands operates in different directions was not discussed. In this case, within certain parameter ranges a stable fixed point emerges, allowing coexistence of alleles deleterious in some demes, but advantageous in others. The deterministic implications of this have been discussed in [10, 18, 20].

In the wake of this work a number of studies were carried out and a plethora of results obtained, all with a variety of different approximations and objectives. Several of these were concerned with an effort to determine the effective population size, which amounts to a rescaling of time for the structured population. Here the temptation to describe the results obtained in terms of an effective population size will be avoided, due to its amorphous definition, and at times misleading designation. The reader is referred to [6] for a review of such work. In turn, other work has focused instead on the effect of migration on local deme properties [4].

The approach adopted in the following section will be to carefully define the model in terms of individuals (i.e. at the microscale). Migration will be incorporated in the form of a metapopulation model with explicit migration rates between each subpopulation. The master equation for this can be written down in a systematic way

[5], although it is too complicated to allow analytic progress to be made. As has been indicated, the key to further progress is to write down a mesoscopic description which is achieved through a diffusion approximation, itself derived by expanding the master equation in inverse deme size. The resulting FPE is still difficult to solve; a model with migration between $\mathcal{D}$ demes which includes selection, leads to a nonlinear FPE in $\mathcal{D}$ variables [5]. It will be shown that progress can be made however, if a separation of timescales exists.

A related set of questions to those that we ask here have been studied in a model of language evolution [1–3], in which each island is mapped on to a speaker having two different linguemes (different ways of saying the same thing) whose concentrations are modified through interaction events (analogous to migration events). While this model has similar features to the one we discuss here, it is distinct, and the methods of analysis and the final results are also different. We have already mentioned the work of Nagylaki [19]. Once again our model, analysis and conclusions differ. Finally, Houchmandzadeh and Vallade have obtained exact results for the probability of fixation in a variant of the model introduced in the next section [14]. However their approach, based on calculating the stationary probability generating function [21] for the system, does not allow for the calculation of quantities related to fixation times or capture the effect of varying selection strengths, both of which we will address.

If selection is allowed to vary across demes, migration and selection can balance in order to induce a polymorphic equilibrium. Results relating to the fixation probability in such systems have only been obtained previously for standard island models [24] and two deme cases with symmetric migration [12], both of which form a subset of of the cases addressed here. While work by Whitlock [25] allows asymmetric migration and multiple demes, the selection strength may only take on two distinct values in those many demes.

4.3 The Metapopulation Moran Model

The model consists of a series of $\mathcal{D}$ demes, on each of which well-mixed populations of fixed and finite size exist. The number of individuals that island i contains is given by $\beta_i N$, where N is some typical island size, and β_i is a scaling factor such that $\beta_i N$ is an integer. The individuals in the population can carry one of two alleles, A and B. An independent deme, unconnected to any others and with a sufficiently large population size, would then be well described by an FPE with transition rates (1.9). However we are interested in the form of the FPE for the system which comprises the whole set of $\mathcal{D}$ demes, with migration between them.

The process is shown diagrammatically in Fig. 4.2 for the case $\mathcal{D} = 2$. In Fig. 4.2a, a reproduction site is chosen with probability f_j, which corresponds to a total birth rate for deme j; if the demes have an equal birth rate per captia, we simply find $f_j = \beta_j (\sum_{i=1}^{\mathcal{D}} \beta_i)^{-1}$. In Fig. 4.2b either one of the two alleles is chosen to reproduce based on their relative frequencies in that deme. The individual then reproduces and
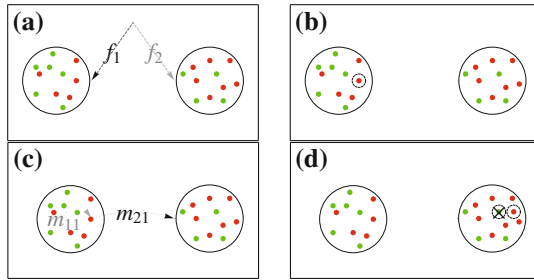
Fig. 4.2 Diagram of the neutral metapopulation Moran model for $\mathcal{D} = 2$. Each large *black circle* is a deme populated by two types of haploid individuals carrying either an allele *A*, in *red*, or allele *B* in *green*. Subfigures **a–d** depict the stages in picking an allele to reproduce and an allele to replace/die

its progeny may either displace an individual in their own deme, or an individual in another deme according to the matrix element m_{ij} (see Fig. 4.2c). The matrix m_{ij} is then the probability that a individual reproducing in j will have offspring which displaces an individual in i. Finally, in Fig. 4.2d the type of individual in i that is displaced is decided, again based on their relative frequencies in i. The vector f_j and the matrix m_{ij} represent probabilities and so satisfy the conditions $\sum_j f_j = 1$ and $\sum_i m_{ij} = 1$ for all j. A more general situation is depicted in Fig. 4.3. The model is essentially the same as that presented in [5], albeit generalised to allow for different deme sizes.

We denote the number of *A* alleles in deme i by n_i. The number of *B* alleles in deme i is then given by $\beta_i N - n_i$, where $\beta_i N$ is the total population of that deme. The equivalent transition rates for the process depicted in Fig. 4.2 can then be calculated using combinatoric arguments. Looking at deme i and summing over the $\mathcal{D}$ demes from which an allele can originate, one obtains

Fig. 4.3 Model extended to $\mathcal{D} = 4$ islands, each containing two types of individual, connected by links whose strength is given by a product of the birth rates, f_j, and the migration probabilities m_{ij}
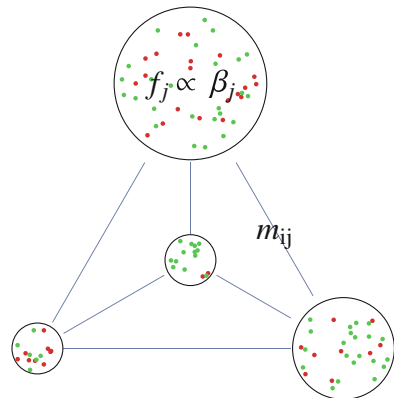
$$T(n_i + 1|n_i) = \sum_{j=1}^{\mathcal{D}} (f_j) \left(\frac{n_j}{\beta_j N} \right) (m_{ij}) \left(\frac{\beta_i N - n_i}{\beta_i N - \delta_{ij}} \right),$$

$$T(n_i - 1|n_i) = \sum_{j=1}^{\mathcal{D}} (f_j) \left(\frac{\beta_j N - n_j}{\beta_j N} \right) (m_{ij}) \left(\frac{n_i}{\beta_i N - \delta_{ij}} \right),$$

where the dependence of $T(\boldsymbol{n}|\boldsymbol{n}')$ on the elements of $\boldsymbol{n}$ that do not change in the transition has been suppressed. Each of the four factors in these expressions for $T(n_i \pm 1|n_i)$ corresponds to one of the four processes displayed in Fig. 4.2. The diagonal elements of m_{ij} do not represent a migration process, but instead the probability that an offspring remains in its parent's deme; they are simply equal to one minus the sum of the other elements in the same column to ensure that $\sum_i m_{ij} = 1$.

Since f_j and m_{ij} always occur together in the combination $m_{ij} f_j$, it is convenient to introduce the matrix G, a generalised migration rate matrix, which is the product of the birth rate, f_j, and the migration probability, m_{ij}:

$$G_{ij} = m_{ij} f_j, \quad \text{where} \quad \sum_{i,j=1}^{\mathcal{D}} G_{ij} = 1; \tag{4.6}$$

the normalisation of G being inherited from the normalisation of m and $\boldsymbol{f}$. Again, the diagonal elements of G_{ij} represent the probability that both deme j is chosen *and* the progeny of the reproduction remains in the parent deme.

The transition rates in terms of G_{ij} then become

$$T(n_i + 1|n_i) = G_{ii} \frac{(\beta_i N - n_i) n_i}{(\beta_i N - 1) \beta_i N} + \frac{\beta_i N - n_i}{\beta_i N} \sum_{j \neq i}^{\mathcal{D}} G_{ij} \frac{n_j}{\beta_j N}, \tag{4.7}$$

$$T(n_i - 1|n_i) = G_{ii} \frac{(n_i)(\beta_i N - n_i)}{(\beta_i N - 1) \beta_i N} + \frac{n_i}{\beta_i N} \sum_{j \neq i}^{\mathcal{D}} G_{ij} \frac{\beta_j N - n_j}{\beta_j N}, \tag{4.8}$$

where we have separated out the contribution from the processes involving two islands (i and j) from those which only involve island i.

The master equation associated with this process (Eq. (2.18) with transition rates (4.7) and (4.8)) is clearly more complicated than that for the well-mixed single-island population and no more tractable. Again however, we can make the diffusion approximation discussed in Sect. 2.4. This time we make the change of variables $x_i = n_i/(\beta_i N)$ and again expand in powers of N^{-1}. For further details, one may consult Appendix C. Truncating at second order we have the FPE (2.30) with

$$A_i(\boldsymbol{x}) = \frac{1}{\beta_i} \left(-x_i \sum_{j \neq i}^{\mathcal{D}} G_{ij} + \sum_{j \neq i}^{\mathcal{D}} G_{ij} x_j \right), \tag{4.9}$$

and

$$B_{ii}(\boldsymbol{x}) = \frac{1}{\beta_i^2}\left(x_i \sum_{j=1}^{\mathcal{D}} G_{ij} + \sum_{j=1}^{\mathcal{D}} G_{ij}x_j - 2x_i \sum_{j=1}^{\mathcal{D}} G_{ij}x_j\right). \qquad (4.10)$$

It is sometimes assumed that the off-diagonal elements of the matrix G_{ij} are such that $G_{ij} = \mathcal{G}_{ij}/N$ for all $i \neq j$, where $\mathcal{G}$ is of order unity [5, 19]. This means that the off-diagonal elements in $B(\boldsymbol{x})$ may be neglected, since they give $\mathcal{O}(N^{-3})$ contributions. Since only the off-diagonal elements of G appear in the vector $A(\boldsymbol{x})$ and only the diagonal elements of G appear in $B(\boldsymbol{x})$, both terms on the right-hand side of Eq. (2.30) are of order N^{-2}, and they effectively balance each other. Asking that the off-diagonal elements of the matrix G are small has a clear biological interpretation. The population is strongly subdivided and it is far more likely that an individual's offspring will remain in the deme of its parents than migrate. This is not the generic case however, and the scaling of the off-diagonal terms with the inverse of the population size is in some cases little more than a mathematical convenience.

Here we will make the choice that the elements of the migration matrix are approximately all of the same order. In doing so we are assuming that once a deme is selected, the probability of allele reproduction-migration is not too dissimilar to that of allele reproduction.

4.3.1 The Metapopulation Moran Model with Selection

We now proceed to incorporate selection into the migration model defined in Sect. 4.3. We begin as before with a set of transition rates, however now the probability of an individual migrating to deme i from deme j is a function of the progenitor's fitness in deme j. In a similar fashion to the case presented in Sect. 2.9.1, the fitness of allele A on each deme is denoted by the vector $\boldsymbol{w}_A$ while the fitness of allele B is $\boldsymbol{w}_B$. The transition rates are

$$T(n_i + 1|n_i) = \sum_{j=1}^{\mathcal{D}} \frac{(\beta_i N - n_i)}{\beta_i N - \delta_{ij}} G_{ij} \frac{[\boldsymbol{w}_A]_j n_j}{[\boldsymbol{w}_A]_j n_j + [\boldsymbol{w}_B]_j(\beta_j N - n_j)},$$

$$T(n_i - 1|n_i) = \sum_{j=1}^{\mathcal{D}} \frac{n_i}{\beta_i N - \delta_{ij}} G_{ij} \frac{[\boldsymbol{w}_B]_j(\beta_j N - n_j)}{[\boldsymbol{w}_A]_j n_j + [\boldsymbol{w}_B]_j(\beta_j N - n_j)}. \quad (4.11)$$

Letting $[\boldsymbol{w}_B]_i = 1$ for every island i, the elements of the fitness term $[\boldsymbol{w}_A]$ are now dependent on both the typical selection strength, s, and a vector $\boldsymbol{\alpha}$, which moderates the typical selection strength in magnitude and direction such that

$$[\boldsymbol{w}_A]_i = 1 + s\alpha_i. \qquad (4.12)$$

Positive α_i therefore corresponds to allele A being advantageous relative to B on island i, while negative α_i means A is deleterious. We assume that the elements of $\boldsymbol{\alpha}$ are of order unity.

As in Sect. 2.9.1, the appearance of n_i terms in the denominator makes the expansion described in Sect. 2.4 difficult. Once again, we assume that s is small and Taylor expand to arrive at a more amenable form. One may then proceed in the same spirit as Sect. 4.3, to arrive at FPE (2.30), valid in the limit of large N and small s. The calculation is carried out in Appendix C in full. The FPE is defined through an $A(x)$ vector and a diagonal $B(x)$ matrix which, when expressed as a series in s, have elements

$$A_i(x) = \frac{1}{\beta_i} \left\{ \sum_{j \neq i}^{\mathcal{D}} G_{ij}(x_j - x_i) + s \sum_{j=1}^{\mathcal{D}} G_{ij}\alpha_j x_j(1 - x_j) - s^2 \sum_{j=1}^{\mathcal{D}} G_{ij}\alpha_j^2 x_j^2(1 - x_j) \right\} + \mathcal{O}(s^3).$$
(4.13)

and

$$B_{ii}(x) = \frac{1}{\beta_i^2} \left\{ x_i \sum_{j=1}^{\mathcal{D}} G_{ij} + \sum_{j=1}^{\mathcal{D}} G_{ij}x_j - 2x_i \sum_{j=1}^{\mathcal{D}} G_{ij}x_j \right\} + \mathcal{O}(s).$$
(4.14)

While it will be seen in our analysis that it is sufficient to work to these orders in s, the appropriate truncation is ultimately dependent on the relative sizes of s and N. The expansions are conducted independently, however as explained in Sect. 2.4 we wish to omit any terms greater than order N^{-2}. For consistency we need to also omit any terms involving s which lead to contributions smaller than N^{-2}.

4.4 The Effective Metapopulation Moran Model

Let us begin by considering the neutral case, with drift and diffusion terms given by Eqs. (4.9) and (4.10). In order to highlight the linearity in $A_i(x)$, we express it as

$$A_i(x) = \sum_{j=1}^{\mathcal{D}} H_{ij}x_j,$$
(4.15)

where the matrix H is defined by

$$H_{ij} = \frac{G_{ij}}{\beta_i} \quad i \neq j, \qquad H_{ii} = -\sum_{j \neq i}^{\mathcal{D}} \frac{G_{ij}}{\beta_i}.$$
(4.16)

As in Chap. 3, we find it more instructive to work in the equivalent SDE setting than that of the FPE. The SDE for the system is

$$\dot{x}_i = \sum_{j=1}^{\mathcal{D}} H_{ij} x_j + \frac{1}{\sqrt{N}} \eta_i(\tau), \quad i = 1, \ldots, \mathcal{D}, \tag{4.17}$$

with noise correlations given by Eq. (4.10).

First, as in Chap. 3 we examine this equation in the deterministic limit, $N \to \infty$. The SDE (4.17) then becomes an ODE which is solvable through a linear analysis. It is clear from the definition of the matrix H in Eq. (4.16) that each row in H sums exactly to 0 for any choice of parameters. We may therefore write $\sum_{j=1}^{\mathcal{D}} H_{ij} = 0$ for all i, or alternatively as the eigenvalue equation $\sum_{j=1}^{\mathcal{D}} H_{ij} v_j^{(1)} = 0$, where $v_j^{(1)} = 1 \, \forall j$ is a right-eigenvector of H with eigenvalue zero. We shall denote this eigenvector as $\boldsymbol{v}^{(1)} = \mathbf{1}$, where $\mathbf{1}$ is the $\mathcal{D}$-dimensional vector

$$\mathbf{1} \equiv \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \text{so that} \quad H\mathbf{1} = 0. \tag{4.18}$$

In addition, it can be shown that all other eigenvalues, $\lambda^{(2)} \ldots \lambda^{(\mathcal{D})}$ have a real part which is negative, under the condition that H is irreducible. In terms of our physical system, this amounts to specifying that no subgroup of demes is isolated from any other.

To prove this, one can transform H into a stochastic matrix. Firstly we introduce a matrix $\tilde{H}$, such that $\tilde{H}_{ij} = \beta_{\min} H_{ij}/(\mathcal{D} - 1)$, where $\beta_{\min}$ is the smallest element of $\boldsymbol{\beta}$. Every off-diagonal element of $\tilde{H}$ then lies in the interval $[0, 1]$, while every diagonal element lies in the interval $[-1, 0)$. We now form the matrix S with entries $S_{ij} = \tilde{H}_{ij} + \delta_{ij}$; since all entries of this matrix are non-negative, and since each row sums to one, S is a stochastic matrix [7, 11]. This implies that the largest eigenvalue of S is 1, and the magnitude of all its other eigenvalues is less than one [7, 11]. Further, by construction, S and H share the same set of eigenvectors. We can use these properties to show that the largest eigenvalue of H is zero, while all other eigenvalues have a negative real part.

The right- and left-eigenvectors corresponding to eigenvalue $\lambda^{(i)}$ will be denoted by $\boldsymbol{v}^{(i)}$ and $\boldsymbol{u}^{(i)}$ respectively. They are orthogonal to each other, and will be normalised so that

$$\sum_{k=1}^{\mathcal{D}} u_k^{(i)} v_k^{(j)} = \delta_{ij}. \tag{4.19}$$

In the special case that H is symmetric the left- and right-eigenvectors coincide and the eigenvalues are real.

Already this tells us a great deal about the system dynamics in the deterministic limit, since the general solution to these equations is given by (2.72). Both $\dot{x}_i$ and $A_i(\boldsymbol{x}) = \sum_j H_{ij} x_j$ have a similar form to Eq. (2.72), which means that after some time:

(i) The vectors $\boldsymbol{x}$, $\dot{\boldsymbol{x}}$ and $\boldsymbol{A}(\boldsymbol{x})$ are all in the direction **1**.
(ii) The vectors $\dot{\boldsymbol{x}}$ and $\boldsymbol{A}(\boldsymbol{x})$ are actually zero in this direction, since $\lambda^{(1)} = 0$.

In other words, the eigenvalues $\lambda^{(i)}$ for $i \geq 2$ control the initial transient dynamics, which decay exponentially along the eigenvectors $\boldsymbol{v}^{(i)}$, $i \geq 2$, to some point on the vector $\boldsymbol{v}^{(1)}$ where the system will stay indefinitely. We recall that in the terminology of dynamical systems, $\boldsymbol{v}^{(1)}$ is coincident with a centre manifold [26]. While this system shares many similarities with the systems discussed in Chap. 3, it is particularly special since the deterministic component is entirely linear. This means that while the fast and slow directions identified in Sects. 3.2 and 3.4 are only valid in the region of the fixed point, the directions identified here are valid far from the line of fixed points which comprise the centre manifold.

Now, suppose we wish to ignore the initial fast behaviour of the system and pick out only the long-term dynamics. We begin by noting that one can decompose the vector $\boldsymbol{A}(\boldsymbol{x})$ in the right-eigenvectors:

$$\boldsymbol{A}(\boldsymbol{x}) = \sum_{k=1}^{m} \boldsymbol{v}^{(k)} a^{(k)}(\boldsymbol{x}). \tag{4.20}$$

The condition that $\boldsymbol{A}(\boldsymbol{x})$ has no components in the fast directions $\boldsymbol{v}^{(j)}$, $j = 2, \ldots, \mathcal{D}$ may then be written in the form

$$\sum_{i=1}^{\mathcal{D}} u_i^{(j)} A_i(\boldsymbol{x}) = 0, \qquad j = 2, \ldots, \mathcal{D}. \tag{4.21}$$

We note that this condition can be shown to be entirely equivalent to Eq. (3.8) with $m = \mathcal{D}$ and $r = 1$. Since the state of the system, $\boldsymbol{x}$, lies on the line **1**, we have $x_1 = x_2 = \cdots = x_{\mathcal{D}}$. We will denote the coordinate along **1** as z, so that the centre manifold in the neutral case is simply

$$x_i = z, \quad i = 1, \ldots, \mathcal{D}. \tag{4.22}$$

However both $\dot{\boldsymbol{x}}$ and $\boldsymbol{A}(\boldsymbol{x})$ are zero on the centre manifold, and so the value of z does not change with time. Although the deterministic dynamics of the neutral model are trivial, the methodology developed here will be applicable to the case with selection, which has non-trivial dynamics.

We now ask, what happens when the population is finite and the stochastic dynamics play a role? We would expect that far from the centre manifold, the deterministic dynamics would dominate over the noise terms, and drag the system to the centre manifold, along which the stochastic dynamics would dominate. In turn any fluctuation that acted to move the system off the centre manifold, would soon be quashed by the deterministic term. This is indeed what we see, as demonstrated in Fig. 4.4 for two and five demes respectively. A clear separation of timescales exists; the deterministic dynamics act quickly to bring the system to the region of the centre manifold, along
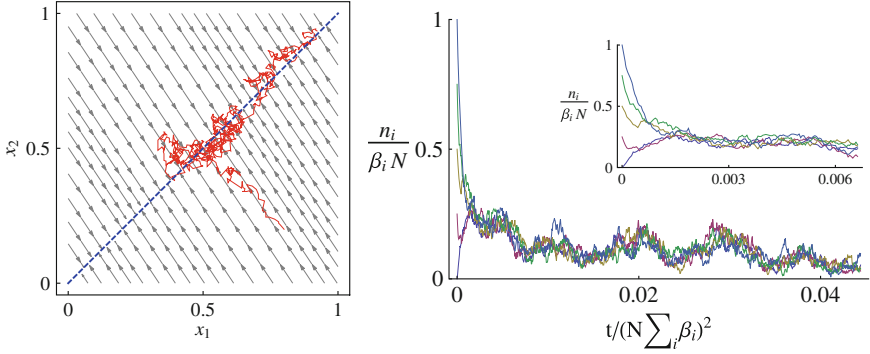
Fig. 4.4 *Left panel* Time series of an individual stochastic trajectory (*red*) and deterministic trajectories (*grey*) for a neutral model with $\mathcal{D} = 2$. The stochastic trajectory can be seen to quickly collapse onto the deterministic centre manifold $x_1 = x_2$, highlighted in *blue*, along which stochastic dynamics are observed. *Right panel* Time series for the populations in a $\mathcal{D} = 5$ system with equal deme sizes. The inset graph shows clearly that the trajectories collapse to the centre manifold, after which they are coupled and move in a stochastic fashion. The system size is $N = 300$ for each deme in both plots

which the system moves stochastically. Our intention is now to extend our treatment of the deterministic system, in which we sought to neglect initial transient dynamics, to the full stochastic system.

We now make the assumption, as in the introduction to this chapter, that there is no noise in the fast directions. The only remaining contribution to the noise is then in the direction $\mathbf{1}$: $\boldsymbol{\eta}(\tau) = \zeta(\tau)\mathbf{1}$. Since, $\boldsymbol{x}$ is restricted to the centre manifold, and there is no deterministic drift along it, the SDE (4.17) simply reduces to

$$\dot{z} = \frac{1}{\sqrt{N}}\zeta(\tau). \tag{4.23}$$

The noise ζ can be characterised by using Eq. (4.19) to write it in terms of $\boldsymbol{\eta}$ as $\zeta(\tau) = \sum_i u_i^{(1)} \eta_i(\tau)$. Then we see that ζ is a Gaussian white noise with zero mean and correlator

$$\langle \zeta(\tau)\zeta(\tau') \rangle = \sum_{i,j=1}^{\mathcal{D}} u_i^{(1)} \, B_{ij}(\boldsymbol{x})\big|_{\boldsymbol{x}=z\mathbf{1}} \, u_j^{(1)} \delta(\tau - \tau'), \tag{4.24}$$

where $B_{ij}(\boldsymbol{x})$ has been evaluated on the centre manifold $\boldsymbol{x} = z\mathbf{1}$. Setting $\boldsymbol{x} = z\mathbf{1}$ in Eq. (4.10) we find

$$\bar{B}(z) \equiv \sum_{i,j=1}^{\mathcal{D}} u_i^{(1)} \, B_{ij}(\boldsymbol{x})\big|_{\boldsymbol{x}=z\mathbf{1}} \, u_j^{(1)}$$

$$= 2z(1-z) \sum_{i,k=1}^{\mathcal{D}} [u_i^{(1)}]^2 G_{ik} \beta_i^{-2}$$

$$\equiv 2b_1 z(1-z), \tag{4.25}$$

where we have introduced the constant

$$b_1 = \sum_{i,k=1}^{\mathcal{D}} [u_i^{(1)}]^2 G_{ik} \beta_i^{-2}, \tag{4.26}$$

and use a bar to indicate the slow component evaluated on the centre manifold. Extending this notation, Eq. (4.23) may be more generally expressed as

$$\dot{z} = \bar{A}(z) + \frac{1}{\sqrt{N}} \zeta(\tau), \tag{4.27}$$

where the drift term evaluated on the centre manifold is zero, $\bar{A}(z) = 0$, and where $\zeta(t)$ is Gaussian correlated white noise with zero mean and correlation function

$$\langle \zeta(\tau)\zeta(\tau') \rangle = \bar{B}(z)\delta(\tau - \tau'). \tag{4.28}$$

Our aim in this section has been to characterise the stochastic dynamics along the centre manifold, and so to develop a one-dimensional, reduced theory. We have also observed that given some set of initial conditions, the stochastic system relaxes to the centre manifold on a much faster timescale than that on which the stochastic dynamics act. The reduced model is therefore ideally suited to answering questions related to global fixation, such as the fixation probability and the mean fixation time, which are long-time properties of the system.

In order to approximate the initial value of the system on the centre manifold, we assume that the trajectory to the centre manifold is essentially deterministic. Then the initial condition on the centre manifold, z_0, is simply the component of the full initial condition $\mathbf{x}_0$, along $\mathbf{v}^{(1)} = \mathbf{1}$:

$$z_0 = \sum_{i=1}^{\mathcal{D}} u_i^{(1)} x_{0i}. \tag{4.29}$$

Together with Eqs. (4.27) and (4.28), this fully defines the reduced model. We note that being able to determine the initial condition in z is possible here because a centre manifold is present. Without this centre manifold, the system would eventually reach a fixed point and there would be no clean way to determine the point at which the fast transient would die out. Further complications would of course arise if the system was nonlinear.

4.4.1 The Case with Selection

Having considered the neutral model with migration, in which neither allele A nor B has any fitness relative to the other, we now consider the case where there is a relative fitness.

To reiterate, this model is now described by the SDE Eq. (2.67), with drift and diffusion terms given by Eqs. (4.13) and (4.14) to leading order. We note that deterministically the system may now admit a non-trivial fixed point in the region between $0 < x_i < 1$, $i = 1, \ldots, \mathcal{D}$; the analysis of the consequences of this will be discussed in Chap. 5.

We begin by noting that since s has been defined as a small parameter in relation to the migration matrix (and hence also the matrix H), we would still expect the system to exhibit a separation of timescales. Now however there is no centre manifold; there is no line along which the deterministic dynamics vanish, as the nonlinear s terms cause deterministic drift. Instead, a slow subspace exists onto which the system quickly relaxes, depicted in Fig. 4.5. The existence of such a subspace is clearly seen in deterministic and stochastic simulations. It is a curved line in the present context, although because s is small it only has a slight curvature.

How do we mathematically specify the slow subspace? Although Eq. (4.13) shows that the $s \neq 0$ deterministic theory is inherently nonlinear, since s is typically very small, we will continue to use the $s = 0$ left-eigenvectors $\boldsymbol{u}^{(j)}$ to approximate the
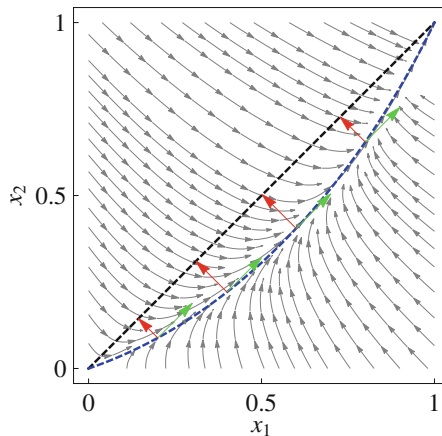

Fig. 4.5 Deterministic trajectories for $s \neq 0$ plotted in *grey*. Here a large value of s has been used ($s = 0.1$), in order to emphasise the nature of the dynamics. The order s and s^2 non-linear terms in Eq. (4.13) result in trajectories that are curved relative to the neutral case depicted in Fig. 4.4. The variable z is measured along the straight *black dashed line*, while the approximation of the slow subspace is plotted as a curved *blue line*. The distance between the slow subspace and the line $\boldsymbol{x} = z\boldsymbol{v}^{(1)}$ is a function of s. The key elements of the approximation are that the system lies on this slow subspace, can move in the direction $\boldsymbol{v}^{(1)}$ (*green arrows*), but cannot move along the direction $\boldsymbol{v}^{(2)}$ (*red arrows*)

slow subspace through Eq. (4.21). Solving these equations numerically, we find that they provide a very good approximation to the observed slow subspace. In Appendix F we solve the $(\mathcal{D} - 1)$ equations (4.21) to find the slow subspace analytically. To do this we transform to a co-ordinate system as in Eq. (3.5); in the present context this is $\boldsymbol{x} = z\mathbf{1} + \sum_{a=1}^{\mathcal{D}-1} w_a \boldsymbol{v}^{(a+1)}$. We find that the equation of the slow subspace takes the form

$$w_a(z) = c_a s z(1 - z) + \mathcal{O}(s^2),\tag{4.30}$$

where the c_a are constants which are calculated in the Appendix F. As $s \to 0$, $w_a(z) \to 0$, and the slow subspace becomes the centre manifold, $\mathbf{1}$, of the neutral case.

The condition given by Eq. (4.21) still restricts $\boldsymbol{A}$ to lie in the $\mathbf{1}$ direction. Therefore once again the approximation has been made that the deterministic dynamics are only in this direction. This inconsistency (the system existing on the curved slow subspace, yet only evolving in the direction $\mathbf{1}$) is a necessary element of the approximation scheme which holds to leading order in s. If we ask that the noise is only in this direction too, then we again write $\boldsymbol{\eta}(\tau) = \zeta(\tau)\mathbf{1}$, just as we did in the neutral case, to find Eqs. (4.27) and (4.28), but now with $\bar{A}(z) \neq 0$. To find $\bar{A}(z)$ we use Eq. (4.19) to pick out the component of $\boldsymbol{A}$ along $\mathbf{1}$, and evaluate it in the slow subspace given by Eq. (4.30):

$$\bar{A}(z) \equiv \sum_{i=1}^{\mathcal{D}} u_i^{(1)} A_i(z, \boldsymbol{w}(z)).\tag{4.31}$$

It is important to note that the dynamics are only in the direction $\mathbf{1}$, even though the system relaxes to the slow subspace given by Eq. (4.30). Thus the approximation requires that the w_a have no dynamics; they are simply mapped on to a value of z using Eq. (4.30).

In fact from a mathematical point of view, the whole procedure may be specified in terms of a linear projection onto the direction $\mathbf{1}$ from any point in state space, together with an understanding that the drift should be evaluated on Eq. (4.30). To do this, a matrix P is constructed, such that when it is applied to any vector wipes out the fast directions $\boldsymbol{v}^{(a)}$ for $a = 2, \ldots, \mathcal{D}$, but leaves the component along the direction $\boldsymbol{v}^{(1)} = \mathbf{1}$, untouched. Using the vector $\boldsymbol{u}^{(1)}$, which is perpendicular to the fast directions, we can construct the projection matrix as

$$P_{ij} = \frac{v_i^{(1)} u_j^{(1)}}{\sum_{k=1}^{\mathcal{D}} v_k^{(1)} u_k^{(1)}}.\tag{4.32}$$

Since $\boldsymbol{v}^{(1)}$ is simply $\mathbf{1}$,

$$P_{ij} = u_j^{(1)}, \qquad i = 1, \ldots, \mathcal{D},\tag{4.33}$$

using the orthonormality condition (4.19). We have already implicitly used this to define $\bar{A}$ and $\bar{B}$, but it can be directly applied to Eq. (2.67) with drift vector (4.13) and diffusion matrix (4.14) to obtain Eq. (4.27). The requirement that the drift vanishes in the fast directions, Eq. (4.21), is still required to find the slow subspace Eq. (4.30).

Let us now apply this approximation procedure. We begin by noting that the drift vector Eq. (4.13) truncated at second order in s can be alternatively expressed by

$$
A_i(x) = \sum_{j=1}^{\mathcal{D}} H_{ij} x_j + \beta_i^{-1} \left\{ s \sum_{j=1}^{\mathcal{D}} G_{ij} \alpha_j x_j (1 - x_j) \right.
$$
$$
\left. - s^2 \sum_{j=1}^{\mathcal{D}} G_{ij} \alpha_j^2 x_j^2 (1 - x_j) \right\}, \tag{4.34}
$$

while the diffusion matrix is left unchanged from the neutral case to leading order (see Eq. (4.14)). In Appendix F, we obtain the expression Eq. (F.8) for the drift vector evaluated on the slow subspace ($A_i(z, w(z))$ in Eq. (4.31)) in terms of the projected variable z. The elements of $A(z)$ to this order take the form

$$
A_i(z) = - s q_i^{(0)} z (1 - z) + s q_i^{(1)} z (1 - z) - s^2 q_i^{(2)} z^2 (1 - z)
$$
$$
- s^2 q_i^{(3)} z (1 - z)(1 - 2z),
$$

where the vectors of parameters, $q^{(i)}$, $i = 0, \ldots, 3$, are defined in Appendix F. The diffusion matrix evaluated on the slow subspace is given in Eq. (F.10).

As with the reduction of the neutral model, we apply the projection Eq. (4.33) to the SDE Eq. (2.67), with the above drift vector and diffusion matrix. This again leads to the reduced SDE of type Eq. (4.27), but now with $\bar{A}(z)$ given by Eq. (4.31). The term $-s q_i^{(0)} z (1 - z)$ which appears in $A_i(z)$ does not appear in $\bar{A}(z)$ because $\sum_i u_i^{(1)} q_i^{(0)} = 0$, which follows from $\sum_i u_i^{(1)} H_{ij} = 0$. The remainder of the expression is given by

$$
\bar{A}(z) = s a_1 z (1 - z) + s^2 a_2 z^2 (1 - z) + s^2 a_3 z (1 - z)(1 - 2z), \tag{4.35}
$$

and $\bar{B}(z)$ retains the form obtained in the neutral case, Eq. (4.25). The parameters a_1 and a_2 are found to be only dependent on the parameters of the problem (m, f, β, α) and on the left-eigenvector $u^{(1)}$:

$$
a_1 = \sum_{i=1}^{\mathcal{D}} P_{ki} q_i^{(1)} = \sum_{i,j=1}^{\mathcal{D}} u_i^{(1)} \frac{G_{ij} \alpha_j}{\beta_i} \tag{4.36}
$$

and

$$
a_2 = - \sum_{i=1}^{\mathcal{D}} P_{ki} q_i^{(2)} = - \sum_{i,j=1}^{\mathcal{D}} u_i^{(1)} \frac{G_{ij} \alpha_j^2}{\beta_i}. \tag{4.37}
$$

The parameter a_3 meanwhile is found to be dependent on the full set of left- and right-eigenvectors and their corresponding eigenvalues:

$$a_3 = -\sum_{i=1}^{\mathcal{D}} P_{ki} q_i^{(3)} \tag{4.38}$$

$$= -\sum_{a=1}^{\mathcal{D}-1} \left[\sum_{i,j=1}^{\mathcal{D}} \frac{u_i^{(1)} G_{ij} \alpha_j}{\beta_i} \sum_{k,l=1}^{\mathcal{D}} \frac{v_j^{(a+1)} u_k^{(a+1)}}{\lambda^{(a+1)}} \frac{G_{kl} \alpha_l}{\beta_k} \right].$$

Its more complicated form is a consequence of the curvature of the slow subspace.

In order to complete our description, we must determine the initial condition in the reduced system. Once again we rely on s being small, and approximate the initial condition on the slow manifold projected onto $v^{(1)}$ by Eq. (4.29).

For both the neutral and the selective metapopulation Moran model we have obtained a one dimensional effective description of the global dynamics, Eq. (4.27), which we expect to be valid in the limit of long times. In the next chapter we will begin to analyse these equations and compare the results against those from Gillespie simulations.

References

1. G.J. Baxter, R.A. Blythe, W. Croft, A.J. McKane, Utterance selection model of language change. Phys. Rev. E **73**, 046118 (2006)
2. G.J. Baxter, R.A. Blythe, A.J. McKane, Fixation and consensus times on a network: a unified approach. Phys. Rev. Lett. **101**, 258701 (2008)
3. G.J. Baxter, R.A. Blythe, A.J. McKane, Fast fixation with a generic network structure. Phys. Rev. E **86**, 031142 (2012)
4. F. Blanquart, S. Gandon, S.L. Nuismer, The effects of migration and drift on local adaptation to a heterogeneous environment. J. Evol. Biol. **25**, 1351–1363 (2012)
5. R.A. Blythe, A.J. McKane, Stochastic models of evolution in genetics, ecology and linguistics. J. Stat. Mech. **P07018**, (2007)
6. B. Charlesworth, Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. **10**, 195–205 (2009)
7. D.R. Cox, H.D. Miller, *The Theory of Stochastic Processes* (Chapman and Hall, London, 1965)
8. S.N. Ethier, T. Nagylaki, Diffusion approximations of Markov chains with two time scales and applications to population genetics. Adv. Appl. Probab. **12**, 14–49 (1980)
9. W.J. Ewens, *Mathematical Population Genetics*, 2nd edn. (Springer, Berlin, 2004)
10. E.A. Eyland, Moran's island migration model. Genetics **69**, 399–403 (1971)
11. F.R. Gantmacher, *Applications of the Theory of Matrices* (Interscience, New York, 1959)
12. S. Gavrilets, N. Gibson, Fixation probabilities in a spatially heterogeneous environment. Popul. Ecol. **44**, 51–58 (2002)
13. D.L. Hartl, A.G. Clark, *Principles of Population Genetics*, 4th edn. (Sinauer Associates Inc., Sunderland, Mass, 2007)
14. B. Houchmandzadeh, M. Vallade, The fixation probability of a beneficial mutation in a geographically structured population. New J. Phys. **13**, 073020 (2011)
15. M. Kimura, G.H. Weiss, The stepping stone model of population structure and the decrease in genetic correlation with distance. Genetics **49**, 561–576 (1964)

16. R. Levins, Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull. Entomol. Soc. Am. **15**, 237–240 (1969)
17. T. Maruyama, On the fixation probability of mutant genes in a subdivided population. Genet. Res. Camb. **15**, 221–225 (1969)
18. P.A.P. Moran, *The Statistical Processes of Evolutionary Theory* (Clarendon Press, Oxford, 1962)
19. T. Nagylaki, The strong migration limit in geographically structured populations. J. Math. Biol. **9**, 101–114 (1980)
20. T. Nagylaki, Y. Lou, The Dynamics of Migration-Selection Models, in *Tutorials in Mathematical Biosciences IV*, vol. 1922, Lecture Notes in Mathematics, ed. by A. Friedman (Springer, Berlin, 2008), pp. 117–170
21. L.E. Reichl, *A Modern Course in Statistical Physics* (Wiley VCH, New York, 1998)
22. T. Rogers, T. Gross, Consensus time and conformity in the adaptive voter model. Phys. Rev. E **88**, 030102(R) (2013)
23. F. Rousset, *Genetic Structure and Selection in Subdivided Populations* (Princeton University Press, Oxford, 2004)
24. H. Tachida, M. Iizuka, Fixation probability in spatially changing environment. Genet. Res. Camb. **58**, 243–245 (1991)
25. M.C. Whitlock, R. Gomulkiewicz, Probability of fixation in a heterogeneous environment. Genetics **171**, 1407–1417 (2005)
26. S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Springer, New York, 2003)
27. S. Wright, Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)

Chapter 5
Analysis of the Reduced Metapopulation Moran Model

In the last chapter, the projection matrix method was used to simplify the migration model to a one-dimensional SDE with the form of Eq. (4.27). The drift term $\bar{A}(z)$ is zero for the case $s = 0$ and given by Eq. (4.35) to second order in s. The diffusion term $\bar{B}(z)$ is given by Eq. (4.25) in each case.

Having started from an IBM, the validity of the approximation can now be tested by comparing predictions of the model against Gillespie simulations [8] (see Sect. 2.3). As measures to test the validity of the approximation, the fixation probability and mean unconditional fixation time are used. Note that the one-dimensional Itō SDE Eq. (4.27) is equivalent to the one-dimensional FPE

$$\frac{\partial p(z,t)}{\partial t} = -\frac{1}{N}\frac{\partial}{\partial z}\left[\bar{A}(z)p(z,t)\right] + \frac{1}{2N^2}\frac{\partial^2}{\partial z^2}\left[\bar{B}(z)p(z,t)\right]. \tag{5.1}$$

The fixation probability of allele A, $Q(z_0)$, and time to fixation, $T(z_0)$, as a function of the initial condition projected onto the manifold,

$$P\boldsymbol{x}_0 = z_0\boldsymbol{1}, \tag{5.2}$$

can be calculated from the backward Fokker-Planck equation [6], as illustrated in Sect. 2.5.2. Solutions to these equations will now be described for the reduced dimension model in the cases $s = 0$ and $s \neq 0$, and the predictions from the reduced systems compared to simulation.

5.1 Analysis—Neutral Case

In the neutral case, the reduced Fokker-Planck equation takes on the same functional form as that which is derived for the neutral Moran model on a single island [10] (Eq. (2.99), with $s = 0$). Scaling N^2 by b_1 and setting $z = x_1$, the results are identical.

For a general system with known parameters, the calculation of b_1 depends only on obtaining the left-eigenvector of H, $\boldsymbol{u}^{(1)}$; $b_1 = \sum_{i,j=1}^{\mathcal{D}} \left[u_i^{(1)} \right]^2 G_{ij}/\beta_i^2$. For small values of $\mathcal{D}$, or alternatively, H matrices with some exploitable symmetries, it may be possible to obtain this analytically. Numerically however, expressions for b_1 are easily obtainable. One could thus proceed in an almost algorithmic way to obtain the reduced FPE given a migration matrix, m, island sizes, $\boldsymbol{\beta}$ and the island birth rates $\boldsymbol{f}$.

We now note some special cases. Firstly, let us take the case where the matrix H is symmetric. Then the left- and right-eigenvectors, $\boldsymbol{u}^{(1)}$ and $\boldsymbol{v}^{(1)}$ coincide (up to an overall constant). Since we take $\boldsymbol{v}^{(1)} = \boldsymbol{1}$, $u_i^{(1)} = $ constant for all i. Using the normalisation condition (4.19) we find that $u_i^{(1)} = \mathcal{D}^{-1} \, \forall i$, and so from Eq. (4.26)

$$b_1 = \sum_{i,j=1}^{\mathcal{D}} \frac{G_{ij}}{(\beta_i \mathcal{D})^2}. \tag{5.3}$$

Perhaps more interesting is the case where the matrix G is symmetric. We begin by looking at the quantity $\sum_{i=1}^{\mathcal{D}} \beta_i H_{ij}$ and expressing H in terms of G using Eq. (4.16):

$$\sum_{i=1}^{\mathcal{D}} \beta_i H_{ij} = \sum_{i \neq j}^{\mathcal{D}} \beta_i \frac{G_{ij}}{\beta_i} + \beta_j \left(-\sum_{k \neq j}^{\mathcal{D}} \frac{G_{jk}}{\beta_j} \right)$$

$$= \sum_{i \neq j}^{\mathcal{D}} G_{ij} - \sum_{k \neq j}^{\mathcal{D}} G_{jk} = \sum_{i \neq j}^{\mathcal{D}} \left[G_{ij} - G_{ji} \right] = 0, \tag{5.4}$$

if G is symmetric. So in this case β_i is the left-eigenvector of H with zero eigenvalue, that is, it is equal to $u_i^{(1)}$, up to an overall constant. Since from Eq. (4.19), $\sum_{i=1}^{\mathcal{D}} u_i^{(1)} = 1$, we have that

$$u_i^{(1)} = \frac{\beta_i}{\sum_{j=1}^{\mathcal{D}} \beta_j}. \tag{5.5}$$

This leads to

$$b_1 = \left(\sum_{j=1}^{\mathcal{D}} \beta_j \right)^{-2}, \tag{5.6}$$

using Eqs. (4.6) and (4.26). This shows that, for a symmetric G matrix, the reduced FPE for the metapopulation model is identical to the full FPE for a well-mixed model with the same total number of individuals, $N \sum_{i=1}^{\mathcal{D}} \beta_i$.
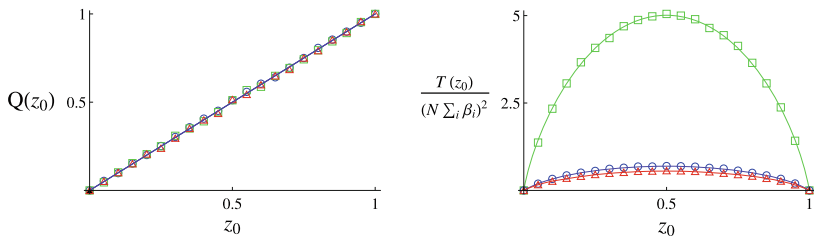
Fig. 5.1 *Left panel* Probability of fixation as a function of the projected initial conditions for neutral systems, $s = 0$. *Right panel* Mean time to fixation as a function of the projected initial conditions, again for $s = 0$ neutral systems. Continuous lines show the analytic predictions from Eqs. (5.7) and (5.8) while the values of symbols are obtained as the mean of 5000 stochastic simulations. Here the model with three demes is studied, each with $N = 200$. The various colours/symbols are obtained from differing migration matrices, given in Appendix D. The system indicted by a *blue line* (the central data in the *right panel*) has a symmetric migration matrix with $b_1 = \left(\sum_{j=1}^{\mathcal{D}} \beta_j \right)^{-2}$

In the neutral case Eq. (2.59) can be solved as in the single island case (see Eq. (2.62)) to give

$$Q(z_0) = z_0, \tag{5.7}$$

where we recall that the projected initial condition, z_0 is found from Eq. (5.2). The probability of fixation depends only on the structure and form of the metapopulations through the determination of the initial conditions.

The mean time to fixation can be calculated from Eq. (2.52), and also resembles the standard result for one island, Eq. (2.54):

$$T(z_0) = -\frac{N^2}{b_1} \left[(1 - z_0) \ln (1 - z_0) + z_0 \ln (z_0) \right]. \tag{5.8}$$

In order to test the predictions of the reduced model, Eqs. (5.7) and (5.8), we compare them against stochastic Gillespie simulations of the underlying microscopic model (specified by the transition rates (4.8)) for a range of systems. We find excellent agreement. In particular, Fig. 5.1 illustrates the results obtained from three different migration matrices, m, for the case $\mathcal{D} = 3$. As predicted the probability of fixation, $Q(z_0)$ is only dependent on the structure of m though the projected initial condition, z_0. The mean time to fixation meanwhile is highly dependent on the structure of m. It is clear that the reduced model reflects this dependency very well.

A natural interpretation is that the population behaves as a well-mixed population with a new 'effective population size'. This terminology will be avoided here however, since the variable z is not directly equivalent to the allele frequency within the global population and, in addition, since it is frequently used in other situations in which its meaning differs from that ascribed to it here [5, 12]. The choice is now made to rewrite Eq. (5.8) in order to compare the results obtained here more easily

with those of Nagylaki [11] (whom we recall considered a Wright-Fisher migration model with non-overlapping generations):

$$T(z_0) = -(N_{\text{Tot}} r_N)^2 \left[(1 - z_0) \ln (1 - z_0) + z_0 \ln (z_0) \right], \tag{5.9}$$

where

$$N_{\text{Tot}} = N \sum_{k=1}^{\mathcal{D}} \beta_k \quad \text{and} \quad r_N = \left(\sqrt{b_1} \sum_{k=1}^{\mathcal{D}} \beta_k \right)^{-1}. \tag{5.10}$$

The parameter r_N thus serves as a clearer measure of the effect of population subdivision relative to an unstructured population of size N_{Tot}. Substituting Eq. (5.6) into the above, we find that the quantity r_N is equal to one in the case of symmetric migration. This agrees with Nagylaki's results for symmetric migration. The results diverge from those of Nagylaki outside this limit however. In his model and analysis, it was found that $r_N \leq 1$, whereas we find no strict upper bound on the value of r_N. Indeed, in Fig. 5.1 we see that r_N may be significantly higher than one in some particular situations (see green line/squares).

To demonstrate the range of values r_N can take, a numerical study can be conducted. An ensemble of random migration matrices, m, is first generated. We have to be careful to pick the elements of m such that the normalisation condition $\sum_{i=1}^{\mathcal{D}} m_{ij} = 1$ holds. Additionally, from a modelling perspective, we would like to see the diagonal elements of m larger than $1/2$ at least, $m_{ii} > 1/2$, so that the probability of an offspring not migrating is greater than the probability it migrates. For each random migration matrix generated, an r_N may then be calculated to give an indication of a potential distribution of r_N values. Since we expect the reduction technique to become unreliable if any of the real parts of the non-zero eigenvalues of H are smaller in magnitude than $N^{-1/2}$ (this will be discussed in more detail in Sect. 5.3), we discard any m matrices that yield such values.

Initially we consider systems with $\mathcal{D} = 4$ and $\beta_i = 1$ for all i. The values for r_N are plotted in a histogram in Fig. 5.2; while no strict upper value for r_N exists, the distribution in this parameter regime does not show $r_N > 1$. We note however, that this is a feature of the modelling choice; if we remove the restriction $m_{ii} > 1/2$, r_N can take a range of values around one (see, for instance, the plot in green/squares in Fig. 5.1, which has a migration matrix given by Eq. (D.3). Further, if we allow the island sizes to vary (as in Fig. 5.2, inset) the distribution of r_N values is altered to allow $r_N > 1$. Testing the theoretical predictions against simulation we once again find excellent agreement across a range of parameters, as shown in the right panel of Fig. 5.2.
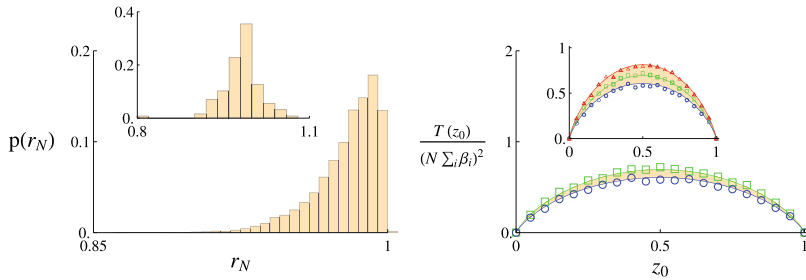
Fig. 5.2 *Left panel* Histograms of values of r_N obtained for a $\mathcal{D} = 4$ system with randomly generated and appropriately normalised migration matrices, m, and $s = 0$. Main histogram shows results obtained when all island sizes are the same, $\beta_i = 1$ for all i. The inset histogram is obtained when the elements β_i are themselves random integers. We have taken $N = 250$ in both cases, and discounted any of the random systems that yield a non-zero eigenvalues with a real part greater than $-N^{-1/2}$. *Right panel* Plots of the mean time to fixation as a function of the initial condition z_0. Analytic predictions from the reduced model, Eq. (5.8), are plotted as continuous lines, while the results from simulation are plotted as symbols. Plots in *blue/circles* correspond to the smallest r_N value obtained in the histograms (*left panel*), those in *green/squares* are obtained from systems with symmetric G matrices ($r_N = 1$), and those in *red/triangles* correspond to the largest r_N value obtained in the histograms. Once again, in the main graph all islands are of the same size, while in the inset plot β_i is allowed to vary. The precise parameters for each plot are given in Appendix D

5.2 Analysis—Case with Selection

In the following analysis we shall consider the case in which $\bar{A}(z)$ given by Eq. (4.35) is truncated at first and second order in s separately. There are two reasons for doing this. First, the case where only the linear term in s is retained can again be mapped onto the full FPE for a one-island Moran model with selection (see Eq. (2.99)). Now not only is N^2 scaled by b_1, but s is scaled by a_1 and $\sqrt{b_1}$. Second, the order at which we truncate $\bar{A}(z)$ can be viewed as an assumption about the relative size of the parameters s and N. In the master equation expansion, which was used to obtain Eq. (2.30), terms smaller than order N^{-2} were neglected. For consistency we would also like to neglect any terms smaller than this arising from s contributions. If $s \approx \mathcal{O}(N^{-1})$, neglecting $\mathcal{O}(N^{-3})$ terms in Eq. (2.30) results in an expression for $\bar{A}(\boldsymbol{x})$ which is first order in s. However, if $s \approx \mathcal{O}(N^{-1/2})$, neglecting $\mathcal{O}(N^{-5/2})$ terms results in $\bar{A}(x)$ which is second order in s. Finally, we will present results for rather small values of N and rather large values of s, as compared with those commonly found in population genetics. We would expect the approximation scheme to become better as N increases, and as a consequence of the above argument, give similar results for proportionally smaller values of s.

5.2.1 First Order in s

We begin by investigating the case where $\bar{A}(z)$ is truncated at first order in s. Solving Eq. (2.59), with $\bar{B}(z)$ given by Eq. (4.25), leads to the probability of fixation being given by

$$Q(z_0) = \frac{1 - \exp\left(-Nsa_1z_0/b_1\right)}{1 - \exp\left(-Nsa_1/b_1\right)}. \tag{5.11}$$

Scaling N by a factor a_1/b_1 gives the one island result of Eq. (2.100) with $z_0 = x_0$.

The form of $T(z_0)$ is found by solving Eq. (2.59) with the same choice of $\bar{A}(z)$ and $\bar{B}(z)$. There are singular points of the differential equation at the boundaries, and care is required when imposing boundary conditions. These aspects are discussed in Appendix E, where expressions for $T(z_0)$ are found in terms of well-defined integrals at various order of s. For instance, to first order in s it is found that

$$T(z_0) = c_2 \left[1 - e^{-M\sigma z_0}\right] - M^2 e^{-M\sigma z_0} \int_0^{z_0} dx\, e^{M\sigma x} \left[\ln x - \ln(1 - x)\right], \tag{5.12}$$

where

$$c_2 = \frac{M^2 e^{-M\sigma}}{1 - e^{-M\sigma}} \int_0^1 dx\, e^{M\sigma x} \left[\ln x - \ln(1 - x)\right]. \tag{5.13}$$

Here $M = N/\sqrt{b_1}$ and $\sigma = a_1 s/\sqrt{b_1}$. The integrals in Eqs. (5.12) and (5.13) may be expressed as combinations of the exponential integral function [1] and logarithms, but they may also be easily evaluated numerically.

The agreement between the simulations and the one-dimensional approximation is excellent for a wide range of parameters, as demonstrated in Fig. 5.3. For systems in which the form of a_1 and b_1 result in a large selective advantage for one or the other of the alleles, the time to fixation can be observed to clearly lose the symmetric form observed in the neutral case. This can be seen in Fig. 5.3 for the results represented in green/squares and those in blue/circles.

5.2.2 Second Order in s

Once again we seek to solve Eqs. (2.59) and (2.52) with x replaced by z and $\bar{A}(z)$ and $\bar{B}(z)$ given by Eqs. (4.35) and (4.25), but now taking $\bar{A}(z)$ to second order in s. We begin by noting that Eq. (4.35) can be written more compactly as$s$

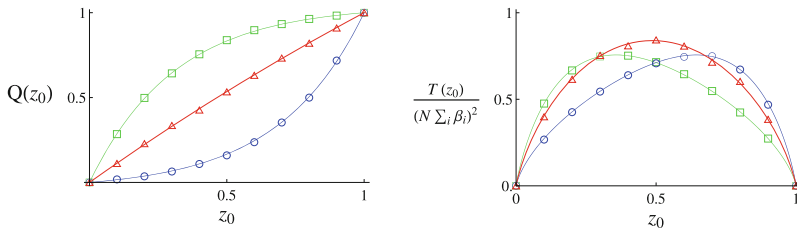$$\bar{A}(z_0) = sz_0(1 - z_0)(k_1 - sk_2z_0), \tag{5.14}$$

Fig. 5.3 Plots for the probability of fixation (*left panel*) and mean time to fixation (*right panel*) as a function of the projected initial conditions for the first order in s case. Continuous lines are obtained from Eq. (5.11) and by evaluating Eq. (5.12) numerically with $\bar{A}(z)$ to order s and $\bar{B}(z)$ given by Eq. (4.25). Various symbols indicate the results obtained from simulations. For each different colour/symbol different $\boldsymbol{\alpha}$ vectors are used; *green squares*, $\boldsymbol{\alpha} = (1, 1, -1)$, *red triangles*, $\boldsymbol{\alpha} = (1, -2, -2)$, and *blue circles*, $\boldsymbol{\alpha} = (-1, -1, 1)$. All other parameters are kept constant; $s = 0.005$, $N = 200$, $\boldsymbol{\beta} = (3, 2, 1)$ and the migration matrix m is given by Eq. (D.8). Simulation results are the average of 5000 runs

with

$$k_1 = a_1 + sa_3 \quad \text{and} \quad k_2 = 2a_3 - a_2. \tag{5.15}$$

We can now solve Eq. (2.59) to obtain an expression for the probability of fixation. We shall merely state the result here; full details of the calculation are given in Appendix G. Defining the function

$$l(z_0) = \sqrt{\frac{N}{(2b_1|k_2|)}}(sk_2z_0 - k_1), \tag{5.16}$$

the probability of fixation, given initial weighted frequency of A allele z_0, is given by

$$Q(z_0) = \frac{1 - \chi(z_0)}{1 - \chi(1)}; \quad \chi(z_0) = \frac{f(l(z_0))}{f(l(0))}, \tag{5.17}$$

where the form of the function f depends on the sign of k_2. If $k_2 < 0$ the function f is the complementary error function [1], if $k_2 > 0$, it is the imaginary error function [4].

The form of $Q(z_0)$ is more complex as compared to the neutral case and the case to first order in s, for which we found that the metapopulation model behaved analogously to the well-mixed model (see Eqs. (5.7), (2.62), (5.8) and (2.54) respectively). However, we can gain further insight into the model by considering certain limits. We examine the situation in which the advantageous allele is the same on each of the demes. In this case we find that the parameters a_1, a_2 and b_1 are all of order 1, while we have taken N large. Given that the function $l(z)$ is then relatively large, we can perform an asymptotic expansion of the error function (see Appendix G, Eqs. (G.10) and (G.11)) to find that for both $k_2 < 0$ and $k_2 > 0$,
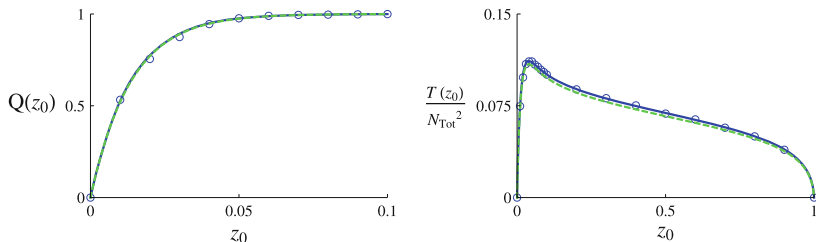
Fig. 5.4 Plots for the probability of fixation, $Q(z_0)$, at low values of z_0, and the mean time to fixation, $T(z_0)$, in a system where s is of order $N^{-1/2}$. *Continuous blue lines* are obtained from the reduced model (using Eq. (5.17) with $k_2 > 0$ for $Q(z_0)$ and solving Eq. (2.52) with $\bar{A}(z)$ given by Eq. (4.35) and $\bar{B}(z)$ by Eq. (4.25), numerically for $T(z_0)$). Parameters used here are $\mathcal{D} = 4$, $s = 0.05$, $N = 400$, $\boldsymbol{\beta} = (1, 1, 1, 2)$, $\boldsymbol{\alpha} = (1, 0.1, 0.5, 1)$. The explicit form of the migration matrix given by Eq. (D.9). Since there are no demes in which selection acts in a contrary direction to any of the others, the asymptotic expansion for $Q(z)$, Eq. (5.18), can be used. The asymptotic expression is plotted by a *green dashed line*; it is indistinguishable from the full order s^2 solution in this regime. For $T(z_0)$ we also plot the first order in s solution as a *green dashed line*; while qualitatively similar to the full solution, there is some numerical discrepancy

$$\chi(z_0) \approx \left(1 - \frac{k_2}{k_1} s z_0\right)^{-1} \exp\left(-\frac{k_1}{b_1} s N z_0 + \frac{k_2}{2 b_1} s^2 N z_0^2\right).$$

Having obtained this expression, valid for large $l(z_0)$, we can make a further approximation for small s. Taking only linear s terms from the above equation, we obtain

$$\chi(z_0) \approx \exp\left(-\frac{a_1}{b_1} s N z_0\right), \tag{5.18}$$

which is the form given in Eq. (2.100), but with a selection strength (or system size) weighted by the ratio a_1/b_1. This is the same result obtained in Sect. 5.2.1, however here we note that it is dependent on the direction of selection being the same in each deme. We find that this provides an excellent approximation in this regime, as demonstrated in Fig. 5.4.

If the direction of selection varies from deme to deme however, then from a consideration of the forms of a_1 and a_3 one can see that there may be some cancellations. This reduces the size of these parameters and invalidates the use of the asymptotic expansion; one must therefore resort to evaluating the expressions given in Eq. (5.17) numerically. We find very good agreement across a wide range of parameters, as shown in Fig. 5.5.
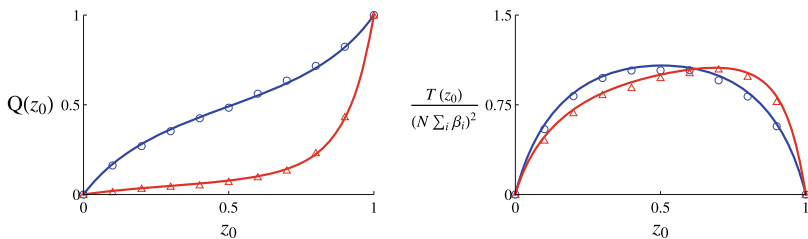
Fig. 5.5 Plots for the probability of fixation, $Q(z_0)$, and mean time to fixation, $T(z_0)$, to second order in s. Continuous lines are obtained by evaluating Eq. (5.17) and solving Eq. (2.52) numerically. Parameters used for the results in *blue* are $\mathcal{D} = 2$, $N = 400$, $s = 1/\sqrt{N}$, $\boldsymbol{\beta} = (1, 1)$, $\boldsymbol{\alpha} = (1, -1)$. Parameters used for the results in *red* are $\mathcal{D} = 4$, $N = 300$, $s = 1/\sqrt{N}$, $\boldsymbol{\beta} = (1, 1, 1, 2)$, $\boldsymbol{\alpha} = (1, 1, 0.3, -1)$. In both cases the migration matrices are non-symmetric; they are stated explicitly in Appendix D. Simulation results are the mean of 10^4 runs

5.3 Estimating the Range of Validity of the Method

Having discussed the approximation method and results, we now turn to considering the range of validity of the expressions for the reduced system, providing a heuristic argument along with a numerical analysis.

As stated in Sect. 4.4 the quantities that govern the separation of timescales are the eigenvalues of H. While the first eigenvalue, $\lambda^{(1)}$, is always zero, we require that the remaining eigenvalues are sufficiently less than zero so that the collapse of the system onto the slow subspace or centre manifold happens on a much faster timescale than that of fixation. What, then, is a sufficient separation of eigenvalues? To investigate this we consider a $\mathcal{D} = 3$ system where each deme is of equal size, whose migration matrix is characterised by a single number $0 < \theta < 1$:

$$m = \begin{pmatrix} \theta & (1-\theta)/2 & (1-\theta)/2 \\ (1-\theta)/2 & \theta & (1-\theta)/2 \\ (1-\theta)/2 & (1-\theta)/2 & \theta \end{pmatrix}. \tag{5.19}$$

With these properties, the H matrix for the system can be simply constructed. We find a degenerate system with only two eigenvalues, the first, zero, and the other two given by $\lambda^{(2)} = (\theta - 1)/2$. We can then plot how the predictions of the reduced system, Eq. (4.27), compare against the results of simulation for some fixed initial condition as $|\lambda^{(2)}|$ decreases.

The results in the neutral case are shown in Fig. 5.6. We recall that since the matrix G is symmetric, b_1 is given by Eq. (5.6). One can see that the reduced system agrees well with the probability of fixation across a remarkably large range of eigenvalues and for initial conditions far from the centre manifold. While the prediction for the time to fixation fares slightly less well, results from simulation still agree over a very large range of parameters, only beginning to diverge at approximately $\lambda_2 = -0.05$, at which point one begins to see a rapid increase in fixation time of the simulations.
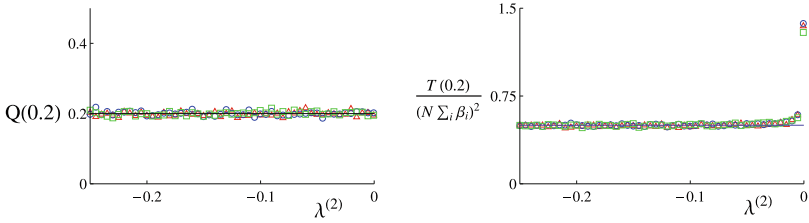
Fig. 5.6 Plots of the probability of fixation and normalised time to fixation as a function of increasing eigenvalue (equivalent to decreasing the strength of collapse onto the centre manifold). The system is prescribed by the migration matrix given in Eq. (5.19), with $\mathcal{D} = 3$, $\boldsymbol{\beta} = (1, 1, 1)$, $s = 0$, and $N = 200$, and the initial condition for both plots is $z_0 = 0.2$. Mean values from 10^4 stochastic simulations are plotted by symbols, whereas continuous lines represent theoretical predictions. The different symbols are obtained from different initial conditions $\boldsymbol{x}_0$ which lead to the same projected initial condition $z_0 = 0.2$ via Eq. (4.29). These are $\boldsymbol{x}_0 = (0.2, 0.2, 0.2)$ (*green squares*), $\boldsymbol{x}_0 = (0, 0.3, 0.3)$ (*blue circles*) and $\boldsymbol{x}_0 = (0, 0.4, 0.2)$ (*red triangles*). The final point on both plots is $\lambda^{(2)} = -5 \times 10^{-4}$, and each point is obtained from the mean of 5000 simulations
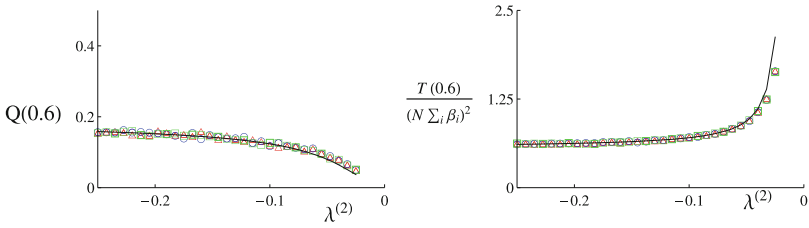

Fig. 5.7 Plots of the probability of fixation and normalised time to fixation as a function of increasing eigenvalue in a system with $s = 0.035$. The system is prescribed by the migration matrix given in Eq. (5.19), with $\mathcal{D} = 3$, $\boldsymbol{\beta} = (1, 1, 1)$, $N = 200$ and $\boldsymbol{\alpha} = (1, -2, 0.5)$, and the initial condition for both plots is $z_0 = 0.6$. Mean values from 6×10^3 stochastic simulations are plotted as symbols, whereas continuous lines represent theoretical predictions. Each symbol once again represents a different initial condition in $\boldsymbol{x}$ with the same effective initial condition in z. Here they are $\boldsymbol{x}_0 = (0.6, 0.6, 0.6)$ (*green squares*), $\boldsymbol{x}_0 = (0.4, 0.7, 0.7)$ (*blue circles*) and $\boldsymbol{x}_0 = (0, 0.9, 0.9)$ (*red triangles*). The final point on both plots is now $\lambda^{(2)} = -0.025$ while once again each point is obtained from the mean of 5000 simulations

This is also the point at which the magnitude of the noise, moderated by $1/\sqrt{N}$, is of the order of the deterministic term (see Eq. (2.67)).

In the case where selection is present, $s \neq 0$, we can conduct a study of the same system for a fixed set of selection parameters. We recall that here the reduction techniques relies on the term $\sum_{j=1}^{\mathcal{D}} H_{ij} x_j$ in Eq. (4.34) inducing a near-deterministic, linear collapse and restriction of the system to the slow subspace. In this situation, where $s \neq 0$, this assumption is not only broken by the noise but also the order s nonlinear terms in Eq. (4.34). We might threfore expect the reduced system to perform less well with decreasing $\lambda^{(2)}$ than the neutral case. While we find this is the case (see Fig. 5.7), the approximation still works very well up to $\lambda^{(2)} \approx -0.05$. At this point the reduced system under-predicts the probability of fixation and over-predicts

the time to fixation, growing rapidly, much faster than the results from simulation would suggest. The physical reasoning behind this behaviour will be discussed in the next section.

5.4 Migration-Selection Balance

So far, a one-dimensional FPE which captures the dynamics of the metapopulation has been used to calculate the fixation probability and mean unconditional fixation time in varying selection regimes. Having obtained these general results, it is now both interesting and instructive to consider a specific system. Of particular note is the case in which the deterministic system, $N \to \infty$, predicts a migration-selection balance of the two alleles; both alleles A and B can coexist in a stable polymorphic equilibrium. We note that a system reminiscent of this has been studied in [2], however there the selection is unchanged across demes and the polymorphic equilibria instead appear as consquences of the diploid nature of individuals in that model.

Let us begin by considering the deterministic Eq. (2.67) with $A(x)$ given by Eq. (4.34) and $N \to \infty$. For clarity we restrict our attention to a two island system with equal island sizes, $\mathcal{D} = 2$, $\boldsymbol{\beta} = (1, 1)$, and a symmetric migration matrix

$$m = \begin{pmatrix} \theta & (1 - \theta) \\ (1 - \theta) & \theta \end{pmatrix}, \tag{5.20}$$

parametrised by the appropriately normalised probability of offspring *not* migrating, $0 < \theta < 1$. While the behaviour of the linear neutral system was straightforward, the introduction of the non-linear s terms in Eq. (4.35) allows for more complicated behaviour. One finds that for $s \neq 0$, nine fixed points emerge. Two of these are at the points of fixation $x_A^* = (1, 1)$ and $x_B^* = (0, 0)$. While a numerical analysis finds that six of the remaining seven fixed points have values outside the physical range, a final fixed point, x_{PE}^*, may arise in between $x = (0, 0)$ and $x = (1, 1)$, under the condition that the selective pressure works in opposite directions on each of the demes. Further, one can observe that only one of the fixed points x_A^*, x_B^*, or x_{PE}^* is stable for a given set of parameters. This behaviour is not observed in [9], in which the selection pressure is fixed across demes. An overview of the situation is given in Fig. 5.8 (left panel); while the region of stable polymorphic equilibrium may appear large in this highly symmetric regime, we note that in general it only occurs for a very restricted range of parameters.

Let us now further restrict our attention to a perfectly symmetric set of parameters by setting $\boldsymbol{\alpha} = (1, -1)$. A phase diagram for this system is shown in Fig. 5.8 (right panel). A fixed point exists, the stability of which increases with increasing s. It is interesting to note the position at which this fixed point x_{PE}^* is found. One might expect, given the highly symmetric nature of the system, that it would be located equidistant between the points of fixation of allele A and allele B, x_A^* and x_B^*. While this is true at first order in s, at second order x_{PE}^* is shifted closer to x_B^*. This break
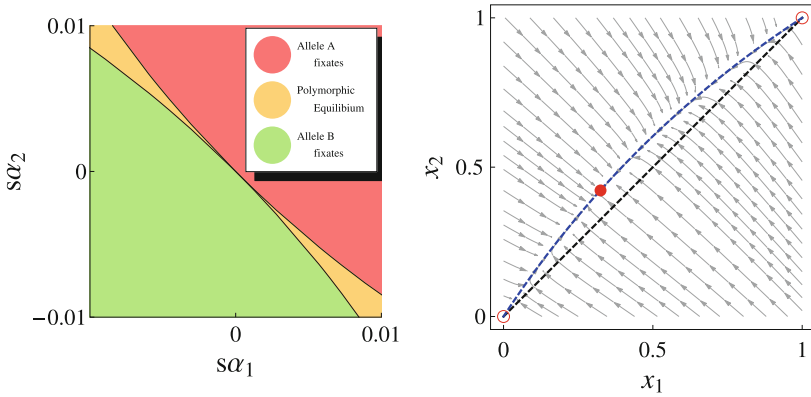
Fig. 5.8 *Left panel* Plot of regions of stability for the fixed points x_A^* (*red, upper-right region*), x_B^* (*green, lower-left region*) and x_{PE}^* (*orange, central region*), for a system $\mathcal{D} = 2$, $\boldsymbol{\beta} = (1, 1)$ and m given by Eq. (5.20) with $\theta = 0.95$. *Right panel* Deterministic trajectories (*grey*) for the same system with $\theta = 0.8$, $\boldsymbol{\alpha} = (1, -1)$ and $s = 0.14$. The stable fixed point x_{PE}^* is indicated by a *red disc*, while unstable fixed points, x_A^* and x_B^*, are *red circles*. The *straight line* $\boldsymbol{x} = z\boldsymbol{v}^{(1)}$ is plotted as a *black dashed line*, while the analytic approximation of the curved slow subspace is plotted as a *blue dashed line*. The location of x_{PE}^* directly on the approximate slow subspace serves to further emphasise the quality of the approximation

in symmetry in fact has its origins in the construction of the model with selection. In Sect. 4.3.1, the elements of the fitness weightings $\boldsymbol{w}_A$ and $\boldsymbol{w}_B$ were chosen to have the form $[w_A]_i = 1 + s\alpha_i$, $[w_B]_i = 1$. This results in a fitness ratio between the two alleles which has a nonlinear character; an island with a fitness weightings of $[w_A]_i = 1 + s\alpha_i$, $[w_B]_i = 1$ does not correspond to an island with $[w_A]_i = 1$, $[w_B]_i = 1 - s\alpha_i$. While a more symmetric choice for the fitness weightings may be possible, this must be ultimately left for further investigation. Instead we now ask, how does this deterministic behaviour in such a regime impact the predictions of the reduced stochastic system, Eq. (4.27)?

To first order in s, the deterministic term $\bar{A}(z)$ in Eq. (4.35), admits no fixed point other than $z = 0$ and $z = 1$. We would expect, however, that the first order in s description would work well for particularly small values of s, say $s \approx 1/N$. Indeed, this is what we find for small s; the deterministic drive towards the polymorphic fixed point is sufficiently weak that its existence has little effect on the probability of fixation or mean time to fixation. The probability of fixation is then well approximated by Eq. (5.11) and the time to fixation by Eq. (5.12), as seen in Fig. 5.9 (green/square plot).

For larger values of s, the stability of the polymorphic fixed point increases in the deterministic limit. To capture the effect on $Q(z_0)$ and $T(z_0)$ one must solve Eqs. (2.59) and (2.52) to second order in s (using Eq. (4.35) in full). One finds that the probability of fixation begins to 'plateau' across a range of initial conditions as s increases, with the fixation of allele B becoming increasingly likely. This counter-
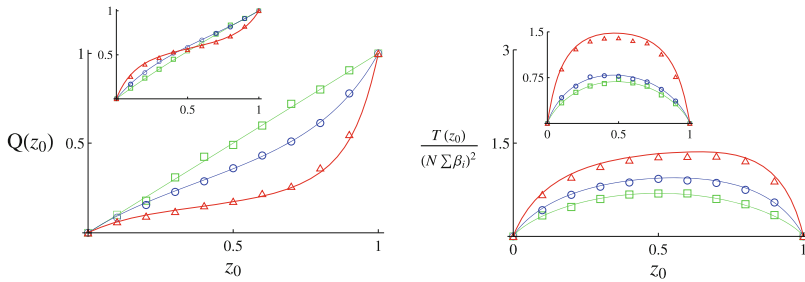
Fig. 5.9 Plots of the probability of fixation, $Q(z_0)$, and mean time to fixation, $T(z_0)$, as a function of the projected initial conditions, for systems featuring symmetric migration balance. Main plots feature $\mathcal{D} = 2$, $\boldsymbol{\beta} = (1, 1)$, $\boldsymbol{\alpha} = (1, -1)$ and m given by Eq. (5.20) with $\theta = 0.8$ for increasing values of s. Continuous lines are obtained from the reduced one-dimensional model, while symbols are obtained from stochastic simulation with $N = 300$. Plots in *green/squares* correspond to $s = 1.66 \times 10^{-3}$ with $Q(z_0)$ and $T(z_0)$ obtained from a first order solution to Eqs. (2.59) and (2.52) (see Eqs. (5.11) and (5.12)). The remaining plots are calculated from second order solutions to Eqs. (2.59) and (2.52) (see Eq. (5.17)) with $s = 5.7 \times 10^{-2}$ for the *blue/circle* plots and $s = 0.11$ for the *red/triangle* plots. Inset plots meanwhile, are obtained from parameters $\mathcal{D} = 5$, $\boldsymbol{\beta} = (1, 1, 1, 1, 2)$, $\boldsymbol{\alpha} = (1, 1, 2.2, -1, -1)$ and an m matrix taken Eq. (D.12)

intuitive break in symmetry can be viewed as a consequence of the skewed fixed point, which biases the system towards fixation at $\boldsymbol{x} = (0, 0)$. The reduced model captures the behaviour extremely well, as observed in Fig. 5.9, left panel. The mean time to fixation meanwhile begins to increase, diverging as the deterministic fixed point holds the system in its vicinity for longer and longer. For these very large, arguably unphysical values of s, the reduced system begins to over-predict the rapidly increasing time to fixation, as seen in Fig. 5.9. This can also be seen as a consequence of s becoming larger than $|\text{Re}[\lambda^{(2)}]|$, as discussed in Sect. 5.3 and shown in Fig. 5.7, in which the parameters selected allow a deterministic polymorphic equilibrium. In situations such as these, alternative approximation schemes may prove useful, such as the Wentzel-Kramers-Brillouin (WKB) expansion of the master equation. When selection is strong and a stable polymorphic equilibrium exists, such an approach has been shown to be superior to the diffusion approximation in certain instances [3]. However, the exploration of this technique lies outside the scope of the current investigation.

The work of Gavrilets and Gibson [7] has already been mentioned and it is worthwhile here relating their approach to our results. They considered a two-deme system with symmetric migration very similar to that described in this section, though with slight differences related to the normalisation of fitness in the transition rates. For systems in which selection acts in the same direction in each deme, our approximations for the fixation probability are analytically equivalent. However, in situations where a polymorphic equilibrium is present, the approximation given in [7] is analytically distinct from that which we derive. Despite this, numerical evaluation of the two results shows them to be very similar. A calculation of the mean fixation time was not conducted in [7].

In this section, a very restricted set of parameters have been considered in order to illustrate the effect of migration-selection balance. While such stable polymorphic equilibria clearly exist for a host of other parameters, including multiple islands of differing sizes and various selection pressures, the parameter range in which they exist becomes increasingly small relative to the full parameter space as $\mathcal{D}$ increases. In addition, while the deterministic analysis of such systems becomes progressively more complex, the reduced system continues to provide a good approximation of the fixation probability and fixation time, as demonstrated in the inset plots of Fig. 5.9. Finally, it should be emphasised that the reduction method has been applied here to an extreme and very particular set of parameters, essentially testing the method to breaking point. This is done to demonstrate the quality of the approximation for large values of s/N.

5.5 Hub

Having discussed the general predictions of the reduced model in both the neutral case and that in which selection is present, we now proceed to apply the results to a specific metapopulation topology, that of the hub or spoke (see Fig. 5.10). The reasons for choosing such a system are twofold. Firstly the system possesses symmetries which make it particularly suitable to an analytic treatment (though it is stressed that the method can also be used for more general systems). Secondly, such a structure allows the behaviour of the model to be systematically investigated as the number of demes increases.

Let us now consider the details of the system. The hub topology is defined as one featuring a main deme which is connected to $\mathcal{D} - 1$ satellite demes. The satellite demes themselves are entirely unconnected to one another. Migration probabilities along the connections are chosen so as to limit the parameter space but still allow for non-trivial behaviour.

Recall the definition of the migration matrix m in Sect. 4.3; it was previously stated that the columns of m were normalised such that the probability the offspring from a reproduction event would not migrate was equal to 1 minus the total probability it would migrate, $m_{jj} = 1 - \sum_{i \neq j}^{\mathcal{D}} m_{ij}$. In this case however, since we have a more restricted geometry, we can instead parametrise the migration probabilities by the probability that the offspring *remains* in the same deme as its parent. Defining ω_1 as the probability that an offspring produced in the central deme does not migrate and ω_2 the probability that an offspring from a satellite deme does not migrate, the normalised migration matrix for $\mathcal{D}$ demes is
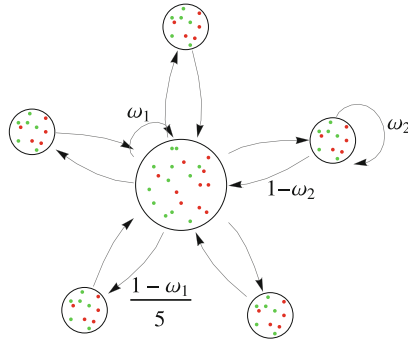
Fig. 5.10 Metapopulation model possessing a hub structure with $\mathcal{D} = 6$. The central deme has a population of $\beta_1 N$ while surrounding demes have a population of $\beta_2 N$. Migration probabilities (conditional on origin island being first selected) in this case can be parametrised by the probability of remaining on a particular deme. The probability of remaining on the central deme is ω_1, while the probability of migrating is dispersed equally over the satellite demes. The probability of remaining on a satellite deme is ω_2, with probability $1 - \omega_2$ of migrating to the central deme

$$
m = \begin{pmatrix}
\omega_1 & 1 - \omega_2 & 1 - \omega_2 & \ldots & 1 - \omega_2 \\
\frac{1-\omega_1}{(\mathcal{D}-1)} & \omega_2 & 0 & \ldots & 0 \\
\frac{1-\omega_1}{(\mathcal{D}-1)} & 0 & \omega_2 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1-\omega_1}{(\mathcal{D}-1)} & 0 & 0 & 0 & \omega_2
\end{pmatrix}. \tag{5.21}
$$

Further, the central deme is taken to have a population of $\beta_1 N$, the satellite demes to have populations $\beta_2 N$, and the birth rate in each deme be proportional to the island size, so that $f_j = \beta_j / \sum_{i=1}^{\mathcal{D}} \beta_i$.

The theory developed in Sect. 4.4 is now applied. We begin by constructing the matrix H from the migration matrix m, and island sizes $\boldsymbol{\beta}$. Before proceeding further, the eigenvalues of H must be calculated, as it is these which define the parameter range over which the approximation is expected to work (see Sect. 5.3). For convenience the quantities

$$
\gamma_1 = \beta_1^2 (1 - \omega_1) \,, \tag{5.22}
$$
$$
\gamma_2 = (\mathcal{D} - 1)\beta_2^2 (1 - \omega_2) \,, \tag{5.23}
$$
$$
\gamma_3 = \gamma_1 + (\mathcal{D} - 1)\gamma_2 \,, \tag{5.24}
$$
$$
\gamma_4 = (\mathcal{D} - 1)\beta_1 \beta_2 \left[\beta_1 + (\mathcal{D} - 1)\beta_2 \right] \,, \tag{5.25}
$$

are introduced. The first two eigenvalues are then given by

$$
\lambda^{(1)} = 0 \,, \qquad \lambda^{(2)} = -\frac{\gamma_3}{\gamma_4} \,, \tag{5.26}
$$

and for the remaining eigenvalues one finds

$$\lambda^{(j)} = -\frac{\gamma_1}{\gamma_4}, \quad j \geq 3. \tag{5.27}$$

By considering these eigenvalues we are already alerted to parameter regimes in which the reduced system could potentially give poor agreement with the full system. For instance, as the number of demes in the system, $\mathcal{D}$, is increased, we find $\lambda^{(2)}$ tends to a finite quantity, $(\omega_2 - 1)/\beta_1$. However, the remaining non-zero eigenvalues tend to zero with increasing deme number. One therefore must be cautious when applying the approximation technique to a hub system with a large number of satellite demes, as the approximation is expected to break down if the magnitude of these eigenvalues approaches $N^{-1/2}$.

To obtain the reduced model in the neutral case, we need only calculate $\boldsymbol{u}^{(1)}$ (see Eq. (4.25)). In the case where we look at second order effects in s, we must also calculate the remaining left- and right-eigenvectors. Since the system contains degenerate eigenvalues (5.27)—this is frequently the case in such highly symmetric systems—the corresponding eigenvectors will not automatically be orthogonal. An orthogonal set must be constructed by taking linear combinations of these vectors, so that the orthonormality condition, Eq. (4.19), holds. The left-eigenvectors can be expressed as

$$\boldsymbol{u}^{(1)} = \frac{1}{\gamma_3} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_2 \end{pmatrix}, \quad \boldsymbol{u}^{(2)} = \begin{pmatrix} -(\mathcal{D} - 1) \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \tag{5.28}$$

and

$$u_i^{(j)} = \delta_{ij} - \frac{1}{j-2} \sum_{l=2}^{j-1} \delta_{li}, \quad j \geq 3. \tag{5.29}$$

The right eigenvectors meanwhile are given by

$$\boldsymbol{v}^{(1)} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \boldsymbol{v}^{(2)} = \frac{1}{(\mathcal{D} - 1)\gamma_3} \begin{pmatrix} -(\mathcal{D} - 1)\gamma_2 \\ \gamma_1 \\ \vdots \\ \gamma_1 \end{pmatrix}, \tag{5.30}$$

and

$$v_i^{(j)} = \frac{j-2}{j-1} \left(\delta_{ij} - \frac{1}{j-2} \sum_{l=2}^{j-1} \delta_{li} \right), \quad j \geq 3. \tag{5.31}$$

With these quantities in hand we can calculate the fixation probability and fixation time as defined in Sects. 5.1 and 5.2.

5.5.1 Hub: $s = 0$

In order to see how the reduced hub model compares with the full system with increasing $\mathcal{D}$, we can look at how the fixation probability $Q(z_0)$ and the fixation time $T(z_0)$, normalised by the total system population, N_{Tot}, changes with the initial condition z_0 kept fixed. For clarity we work in the parameter r_N (introduced in Eq. (5.10)) rather than b_1, since r_N measures the change in timescale relative to that of the well-mixed population. The results are plotted in Fig. 5.11, and we see that as $\mathcal{D}$ increases the approximation continues to provide good agreement with the exact Gillespie simulation of Eq. (2.18), with transition rates given by Eqs. (4.7) and (4.8).

As stated in Sect. 5.1, the probability of fixation is only dependent on network structure through the projected initial condition, z_0. Since z_0 is held constant in this case, the probability of fixation does not change as the network structure is altered. We note however that the behaviour of the fixation time as a function of the number of demes is non-trivial. The results in Fig. 5.11 may be compared with the results for a single island of the same size as the total hub population; increasing the size of the island would give $r_N = 1$, regardless of size. Here we see r_N starts at a significantly higher value and decreases as the deme number (and hence the total population size) is increased.
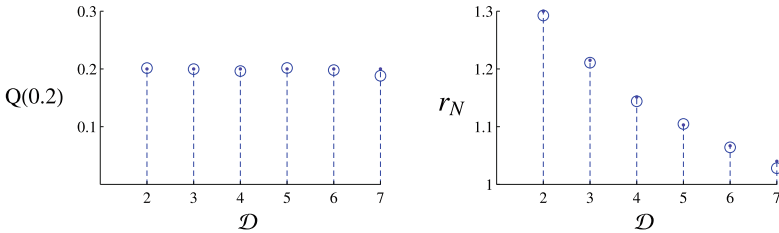

Fig. 5.11 *Left panel* probability of fixation, $Q(z_0)$, with fixed initial condition $z_0 = 0.2$, for the neutral hub model ($s = 0$) plotted as a function of the number of demes, $\mathcal{D}$. *Right panel* parameter r_N as a function of $\mathcal{D}$ for the neutral hub model. Discrete analytic results are plotted as a *dashed blue columns*, and calculated from Eqs. (5.8) and (5.10), and (4.26) calculated using Eq. (5.28). Simulation results, plotted as *circles*, are the mean results from 6000 runs. Parameters used in this example are $\omega_1 = 0.5$, $\omega_2 = 15/16$, $\beta_1 = 6$, $\beta_2 = 1$, and finally $N = 300$

5.5.2 Hub with Selection

Let us now incorporate selection into the general hub model described in Fig. 5.10. Suppose that the selection strength in the central deme, 1, is moderated by α_1 while the selection strength in the satellite demes is moderated by α_2. The $\bar{A}(z)$ term for the system is given by Eq. (4.35) and $\bar{B}(z)$ by Eq. (4.25). The parameters a_1, a_2 and a_3 are given by Eqs. (4.36)–(4.38) which can now be easily calculated since we have the left- and right-eigenvectors of H (Eqs. (5.28)–(5.31)) respectively). Their exact forms are too lengthy to be reproduced here, but are obtained by direct substitution. The results can then be tested against exact Gillespie simulations of the stochastic system defined by Eq. (4.11). As an example, let us compare two systems.

In the first system we fix the number of demes to two with the first deme being defined as the central deme with population $\beta_1 N$ and the second as the satellite deme with population $\beta_2 N$. In deme one, the A alleles experience a selective pressure $s\alpha_1$, while in the deme two, the satellite deme, the alleles experience a selective pressure $s\alpha_2$.

In the second system, we again have a central deme with a population of $\beta_1 N$, but we now fix the total population in each satellite deme to N and vary the total deme number $\mathcal{D}$. Again the fitness in the central deme is equal to $s\alpha_1$ and the fitness in each satellite deme is equal to $s\alpha_2$. We can then say that in both systems, the number of individuals in the selective environments $s\alpha_1$ and $s\alpha_2$ are equal if $\beta_2 = (\mathcal{D} - 1)$.

Naïvely then, one might expect the systems to behave similarly for such metrics as fixation probability and fixation time, as β_2 and $\mathcal{D} - 1$ respectively increase in each system. However, the analytical results predict distinct behaviour. This is supported by simulation; in Fig. 5.12, a particular set of parameters is fixed and the size of the populations (moderated by β_2 in the first case and discretely by $\mathcal{D}$ in the second
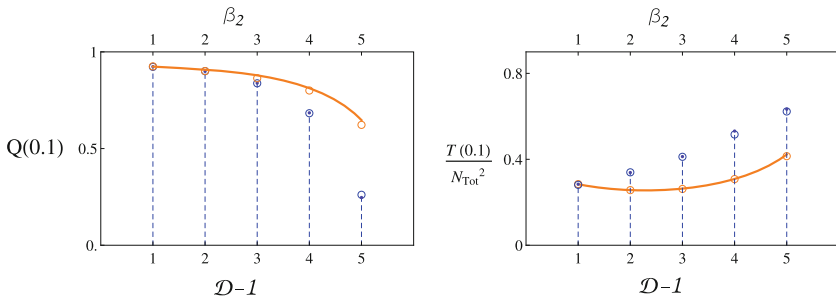

Fig. 5.12 Plots of fixation probability, $Q(z_0)$ and fixation time, $T(z_0)$ for two different models at $z_0 = 0.1$. The *orange lines* are obtained from the reduced model of a two-deme system in which β_2, the relative size of the second deme, is increased. The results from simulation of the system are shown as *orange circles*. The discrete values indicated by the *blue dashed lines* are obtained from a hub model in which the number of $(\mathcal{D} - 1)$ satellite demes is increased. Simulation results are plotted as *blue circles*. Simulation results are the mean of 2000 runs. Parameters used are $s = 0.03$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\beta_1 = 3$, $\beta_2 = 1$, $\omega_1 = 0.625$, $\omega_2 = 0.9375$, and $N_{\text{Tot}} = 300$

case) is increased. The probability of *A* fixating decreases much more rapidly with an increasing number of satellite demes than the two deme system with increasing size of the second island. The time to fixation meanwhile increases more rapidly with the increasing number of satellite demes than that with two islands.

References

 1. M. Abramowitz, I.A. Stegun (eds.), *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1965)
 2. P.M. Altrock, A. Traulsen, F.A. Reed, Stability properties of underdominance in finite subdivided populations. PLoS Comput. Biol. **7**, e1002260 (2011)
 3. M. Assaf, M. Mobilia, Fixation of a deleterious allele under mutation pressure and finite selection intensity. J. Theor. Biol. **275**, 93–103 (2011)
 4. A. Erdélyi (ed.), *Higher Transcendental Functions*, vol. II (McGraw-Hill, New York, 1953)
 5. W.J. Ewens, *Mathematical Population Genetics*, 2nd edn. (Springer, Berlin, 2004)
 6. C.W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2009)
 7. S. Gavrilets, N. Gibson, Fixation probabilities in a spatially heterogeneous environment. Popul. Ecol. **44**, 51–58 (2002)
 8. D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
 9. B. Houchmandzadeh, M. Vallade, The fixation probability of a beneficial mutation in a geographically structured population. New J. Phys. **13**, 073020 (2011)
10. P.A.P. Moran, *The Statistical Processes of Evolutionary Theory* (Clarendon Press, Oxford, 1962)
11. T. Nagylaki, The strong migration limit in geographically structured populations. J. Math. Biol. **9**, 101–114 (1980)
12. M.C. Whitlock, N.H. Barton, The effective size of a subdivided population. Genetics **146**, 427–441 (1997)

Chapter 6
Further Developments

In this chapter further developments of the conditioning and projection matrix method are presented. In the following section, the discussion of the metapopulation Moran model is completed by adding mutation. Following this, a more detailed comparison of the two reduction methods is conducted. Finally, in the third section, the projection matrix method is applied to a new system, the Lotka-Volterra competition model of two interacting populations.

6.1 Mutation and the Metapopulation Moran Model

In this section, mutation will be added to the metapopulation Moran model described in Sect. 4.3. The well-mixed Moran model with mutation has already been discussed in Chap. 2, along with some results relating to its behaviour. It is useful to briefly recap some of its behaviour here, before proceeding to the metapopulation generalisation.

The transition rates defining the well-mixed model are given in Eq. (2.16). The parameters ω_1 and ω_2 are assumed to be small, of the order N^{-1}. In this way terms involving $\omega_1 N^{-2}$ and $\omega_1 N^{-2}$ are neglected in the expansion of the master equation, and the diffusion matrix $B(x)$ in the model's FPE does not feature these parameters (see Eq. (2.39)).

There are no absorbing states in the model, so after some time the system tends towards a stationary distribution given by Eq. (2.40). Some illustrative plots for the stationary distribution are given in Fig. 2.1. Note that while these distributions appear reminiscent of a system which undergoes a bifurcation (the central peak in Fig. 2.40c becoming two peaks in Fig. 2.40a), the model does not predict such dynamics deterministically. The deterministic component in Eq. (2.28) instead predicts a single fixed point at

$$x^* = \frac{\omega_1}{\omega_1 + \omega_2},$$
(6.1)

regardless of the other parameters. The observed peaks at the boundaries of the stationary distribution are a purely stochastic effect. The shape of the stationary distribution is governed by Eq. (2.40), and the different forms it may take can be broadly categorised in terms of the signs of the parameters c and d, defined in Eq. (2.41). If c and d are both greater than 0, the distribution will tend to be peaked in the vicinity of the deterministic fixed point. This is the case if the parameter combinations $N\omega_i(b^{-1} - 1)$ for $i = 1, 2$, are greater than one (see Eq. (2.41)), or alternatively, if system size N is large enough that the deterministic dynamics dominate the system's behaviour. Conversely, if both c and d are less than 0, the distribution will tend to be peaked at the boundaries. For this to be the case, the terms $N\omega_i(b^{-1} - 1)$ for $i = 1, 2$, must be less than one. This can occur if the system size is relatively small, so that the stochastic behaviour dominates the system.

Once again the mechanism can most intuitively understood in the SDE formalism. Expressing the system in the form of Eq. (2.67), with $A(x)$ and $B(x)$ taken from Eq. (2.39), one sees that for large N and ω_1/ω_2, the noise is small relative to the deterministic term, and the system will tend to fluctuate about the fixed point. When N and ω_1/ω_2 are small however, the noise dominates the dynamics. Considering the functional form of $B(x)$ in Eq. (2.39), one sees that the noise becomes small near the boundaries. Therefore when the noise dominates, the system quickly moves to the boundaries; once there the noise is small, and therefore the system remains in this vicinity for a long time, until an unusually large fluctuation pushes it to the opposing boundary, and so on.

With this description in hand, the formulation of the metapopulation Moran model with mutation can be conducted. Since mutation has been included as a separate event to birth and death in the well-mixed case, the metapopulation model is constructed so that mutation is independent of migration. In a similar manner to that in which selection was incorporated in Sect. 4.3.1, the mutation rates will be allowed to vary from deme to deme. The concept of mutation varying with habitat is perhaps less intuitive than that of selective pressure changing according to the environment. However, there have been experimental studies of certain species that suggest that mutation rates can increase as a result of external environmental stress factors (see, for example [12]).

Let us define the vectors $\omega_1 = (\omega_{11}, \omega_{12}, \ldots, \omega_{1\mathcal{D}})$ and $\omega_2 = (\omega_{21}, \omega_{22}, \ldots, \omega_{2\mathcal{D}})$, where ω_{1i} is the mutation rate from B to A in deme i and similarly ω_{2i} is the mutation rate from A to B. The transition rates for the model then read

$$T(n_i + 1|n_i) = b \sum_{j=1}^{\mathcal{D}} \left[(f_j) \left(\frac{n_j}{\beta_j N} \right) (m_{ij}) \left(\frac{\beta_i N - n_i}{\beta_i N - \delta_{ij}} \right) \right] + (1-b)\omega_{1i} \frac{\beta_i N - n_i}{\beta_i N},$$

$$T(n_i - 1|n_i) = b \sum_{j=1}^{\mathcal{D}} \left[(f_j) \left(\frac{\beta_j N - n_j}{\beta_j N} \right) (m_{ij}) \left(\frac{n_i}{\beta_i N - \delta_{ij}} \right) \right] + (1-b)\omega_{2i} \frac{n_i}{\beta_i N},$$

$$(6.2)$$

which, along with the master equation (2.18), defines the model. While potentially the parameter b could be made to vary across demes, it has been assumed constant here for simplicity. As in Sect. 4.3, the approximately continuous variables $x_i = n_i / (\beta_i N)$ are introduced. The expansion of the master equation can then be conducted in a similar manner to that outlined in Appendix C. Assuming once again that the mutation rates are small, and truncating the expansion at second order, the FPE (2.30) is obtained, with drift terms

$$A_i(\boldsymbol{x}) = \frac{1}{\beta_i} \left(b \sum_{j \neq i}^{\mathcal{D}} G_{ij}(x_j - x_i) + (1 - b) \left[-(\omega_{1i} + \omega_{2i})x_i + \omega_{1i} \right] \right) \quad (6.3)$$

and the elements of the diagonal diffusion matrix by

$$B_{ii}(\boldsymbol{x}) = \frac{b}{\beta_i^2} \sum_{j=1}^{\mathcal{D}} G_{ij} \left(x_i + x_j - 2x_i x_j \right) + \mathcal{O}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2). \quad (6.4)$$

Once again G_{ij} is related to m_{ij} and f_j by Eq. (4.6). As in the neutral case and that with selection (see Eqs. (4.15) and (4.34)), it is useful to rewrite the drift term in order to highlight the linearity at leading order:

$$A_i(\boldsymbol{x}) = b \sum_j^{\mathcal{D}} H_{ij} x_j + \frac{(1 - b)}{\beta_i} \left[-(\omega_{1i} + \omega_{2i})x_i + \omega_{1i} \right], \quad (6.5)$$

where the matrix H is given by Eq. (4.16). An illustrative plot is given in Fig. 6.1. The system is nonlinear and has as many variables as there are islands. In this respect it is not entirely dissimilar to the metapopulation Moran model with selection. In Sect. 5.3 it was seen that, under the condition that the nonlinear terms in $\boldsymbol{A}(\boldsymbol{x})$ were of smaller magnitude that the eigenvalues of H, the linear collapse to a slow subspace dominated the dynamics. From Fig. 6.1 we can see that a similar situation arises here. We can therefore follow the methodology developed in Sect. 4.4.1 to attempt to reduce the $\mathcal{D}$-variable system to an effective single variable description.

Let us begin by considering the form of the slow subspace. We recall that the fast and slow directions of the neutral analogue, obtained from the left- and right-eigenvectors of H, $\boldsymbol{u}^{(i)}$ and $\boldsymbol{v}^{(i)}$, are used as approximations for the fast and slow directions in the non-neutral case. The slow subspace is then approximated by the nullcline of the fast directions. This is given by the solution to Eq. (4.21) with the elements $A_i(\boldsymbol{x})$ taken from Eq. (6.5). For the case with selection, the slow subspace was calculated to order s (see Appendix F). This was necessary, since an order s^2 treatment of the reduced system was required in Sect. 5.4. For the case with mutation however, we will find it sufficient to work to first order in ω_1 and ω_2 in the reduced system. The slow subspace need then only be calculated to zeroth order in ω_1 and
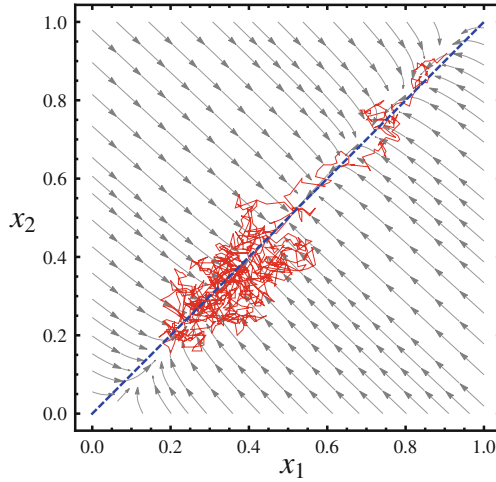
Fig. 6.1 Trajectories of the two-island metapopulation Moran model with mutation. Deterministic trajectories are plotted in *grey* from an ODE of type (2.70) with $A_i(\boldsymbol{x})$ taken from Eq. (6.5). A single stochastic trajectory, obtained from Gillespie simulation of the master equation with transition rates (6.2), is plotted in *red*. The centre manifold, which is equal to the slow subspace at first order in ω_1 and ω_2, is plotted in *blue*. Parameters used here are $\omega_1 = (8 \times 10^{-3}, 7 \times 10^{-4})$, $\omega_2 = (8 \times 10^{-3}, 3 \times 10^{-4})$, $m_{11} = m_{22} = 0.8$, $m_{12} = m_{21} = 0.2$, $\boldsymbol{\beta} = (1, 1)$ and $N = 200$

ω_2, which yields the result from the neutral case, $x_i = z \; \forall i$. The drift vector and diffusion matrix evaluated on the slow subspace are then

$$A_i(z) = \frac{(1 - b)}{\beta_i} \left(-(\omega_{1i} + \omega_{2i})z + \omega_{1i} \right) + \mathcal{O}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2), \tag{6.6}$$

and

$$B_{ii}(x) = 2z(1 - z)\frac{b}{\beta_i^2} \sum_{j=1}^{\mathcal{D}} G_{ij}, \tag{6.7}$$

respectively.

The projection matrix for this system is again defined by Eq. (4.33). Applying this to the above drift and diffusion matrices allows us to pick out the components of the system's dynamics in the slow direction. On applying the projection matrix to Eqs. (6.6) and (6.7), a reduced SDE with the form of Eq. (4.27) is obtained for the slow dynamics in terms of z. The drift vector reads

$$\bar{A}(z) = (1 - b)\left(-(\bar{\omega}_1 + \bar{\omega}_2)z + \bar{\omega}_1 \right), \tag{6.8}$$

where we have introduced the parameters

$$\bar{\omega}_1 = \sum_{i=1}^{\mathcal{D}} \frac{u_i^{(1)}}{\beta_i} \omega_{1i}, \qquad \bar{\omega}_2 = \sum_{i=1}^{\mathcal{D}} \frac{u_i^{(1)}}{\beta_i} \omega_{2i}. \tag{6.9}$$

The diffusion matrix retains the form from the reduced neutral model to leading order, Eq. (4.25).

Since the forms of the drift and diffusion terms for the reduced system are analogous to the one island case, the equation for the stationary distribution in z can be calculated in exactly the same way. The form of the probability of fixation retains the form of Eq. (2.40), but with x replaced by z. The parameters c and d are now given by

$$c = \frac{N}{b_1}\bar{\omega}_1(b^{-1} - 1) - 1, \quad d = \frac{N}{b_1}\bar{\omega}_2(b^{-1} - 1) - 1. \tag{6.10}$$

The effect of structure on the dynamics of the system can now be investigated.

We begin by considering the most simple case, that when all islands are the same size, $\beta_i = 1 \ \forall i$, and migration is symmetric. Under these conditions, the matrix G is also symmetric. As discussed in Sect. 5.1, $\boldsymbol{u}^{(1)}$ is given by Eq. (5.5) and b_1 by Eq. (5.6). Therefore, one obtains $\bar{\omega}_i = \sum_{j=1}^{\mathcal{D}} \omega_{ij}/\mathcal{D}$ for $i = 1, 2$. The effective mutation rates are simply equal to the mean of the mutation rates across demes. The effect of the term N/b_1 in Eq. (6.10) is not so straightforward. Recalling that the total population is given by $N_{\text{Tot}} = N\mathcal{D}$ in this situation, one finds $N/b_1 = \mathcal{D}N_{\text{Tot}}$. The reduced system therefore has a greater effective system size than a well mixed-
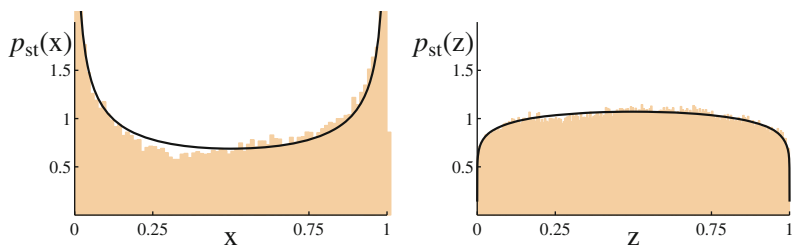

Fig. 6.2 *Left panel* Stationary distribution for a one island model with mutation, with parameters $\omega_1 = 7 \times 10^{-4}$, $\omega_2 = 7 \times 10^{-4}$ and $N = 800$. The *black line* is obtained from the theoretical stationary distribution given by Eq. (2.40) with parameters taken from Eq. (2.41), while the *orange* histogram is obtained from Gillespie simulation. *Right panel* Stationary distribution for the reduced system in terms of the projected variable z for a system with parameters $\mathcal{D} = 2$, $\omega_1 = (7 \times 10^{-4}, 7 \times 10^{-4})$, $\omega_2 = (7 \times 10^{-4}, 7 \times 10^{-4})$, $m_{11} = m_{22} = 0.8$, $m_{12} = m_{21} = 0.2$, $\beta = (1, 1)$ and $N = 400$. The *black line* is obtained from the solution to the stationary distribution of the reduced system, Eq. (2.40), with parameters c and d given in Eq. (6.10), and the *orange* histogram from Gillespie simulations of the full two-dimensional model. Note that while the total populations and mutation rates in both systems are equal, the effect of migration has fundamentally changed the character of the stationary distribution. This change is accurately captured by the reduced theory

system with the same average mutation rates (see Eq. (2.40)). This means that, in the case where the matrix G is symmetric, the effect of the population structure is to reduce the effect of the noise on the stationary distribution. An example of such a case is given in Fig. 6.2.

In general, it is found that the effect of population structure identified above is seen in most other parameter regimes. That is, the effect of population structure is in general to reduce the effect of the noise on the stationary distribution, relative to a well-mixed system with the same total population size and mean mutation rates. However, there do exist some cases where the converse is true, where the population structure *increases* the effect of noise relative to the well-mixed model. In Fig. 6.3, the stationary distributions for a well-mixed model and a two-island system are plotted for a situation in which this is the case. We note that numerically it appears that such
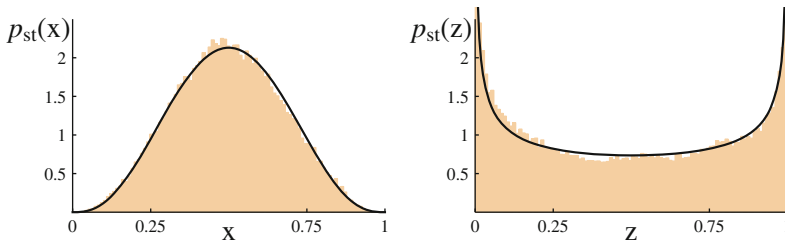

Fig. 6.3 *Left panel* Stationary distribution for a well-mixed model with mutation, with parameters $\omega_1 = 3.7 \times 10^{-3}$, $\omega_2 = 3.7 \times 10^{-3}$ and $N = 1020$. As in Fig. 6.2, the *black line* is obtained from the one-dimensional theory and the *orange* histogram from stochastic simulation. *Right panel* Stationary distribution for a two-island Moran model with mutation in terms of z, the distance along the slow subspace. This system has parameters $\mathcal{D} = 2$, $\omega_1 = \omega_2 = (5.8 \times 10^{-5}, 7.4 \times 10^{-3})$, $m_{11} = 0.7$, $m_{12} = 0.03$ $m_{21} = 0.3$ $m_{22} = 0.97$, $\beta = (3.6, 1.5)$ and $N = 200$. The *black line* is obtained from the predictions of the reduced system, while the *orange* histogram is obtained from stochastic simulation of the full model. The parameters are such that the total system size and the average mutation rates are the same in each case, but again the addition of population structure significantly alters its behaviour
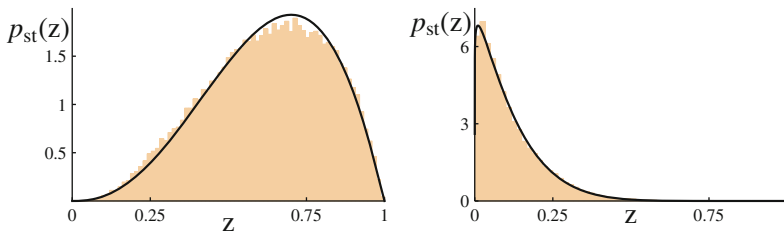

Fig. 6.4 *Left panel* The stationary distribution for z for a metapopulation Moran model with mutation on three islands with asymmetric migration. The *black continuous line* is obtained from the reduced theory and the *orange* histogram from simulations of the full model. *Right panel* A similar plot to that in the *left panel*, but for a system with five islands. Explicit parameters are omitted here for brevity, but are given in Appendix D

behaviour is only possible if the elements of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are allowed to vary across demes. More investigation is clearly needed to explore the full range of behaviour possible in this system. However, as demonstrated in Fig. 6.4, the analytic predictions derived from the reduced model provide remarkably good agreement with the results from Gillespie simulation of the full model, across a range of parameters.

6.2 Comparing the Conditioning and the Projection Matrix Method

In Chap. 3, the conditioning method was introduced, and it was shown that the method provided a good reduced-dimension description of two particular systems. In Chap. 4, another method was introduced, the projection matrix method. This method was used to reduce the dimension of the metapopulation Moran model, and in Chap. 5 it was seen that the effective system successfully captured features of the full model across a range of parameters. Both of the methods have been shown to give good approximations for certain models, but we now ask, 'are there any more mathematical ways we can motivate the implementation of these methods?', and further 'how do they compare?'.

Initially we recall some of the statements made in Sect. 4.1. In both methods we begin with a set of general SDEs with the form of Eq. (2.67). A deterministic fixed point $\boldsymbol{x}^*$ is assumed to exist, about which we linearise. The deterministic analogue of this linear system is then used to identify any separation of timescales which may be present in the system. As described in Sect. 3.1, if the real part of all the eigenvalues of the system's Jacobian are negative but disparate (that is, they satisfy the inequality (3.2)), a separation of timescales exists in the region of the fixed point.

To provide an approximation of the deterministic slow and fast directions, the right-eigenvectors of the system's Jacobian are used. These form a basis into which we can transform (see Eq. (3.5)). In the basis of the fast and slow eigenvectors, the system is described by r slow-variables, denoted $\boldsymbol{z}$ and $m - r$ fast-variables, denoted $\boldsymbol{w}$. The SDEs in these variables can be expressed

$$\frac{dz_i}{dt} = A_i(\boldsymbol{z}, \boldsymbol{w}) + \frac{1}{\sqrt{N}} \kappa_i(t), \quad i = 1, \dots r,$$

$$\frac{dw_j}{dt} = A_j(\boldsymbol{z}, \boldsymbol{w}) + \frac{1}{\sqrt{N}} \kappa_j(t), \quad j = r + 1, \dots m, \quad (6.11)$$

where $\boldsymbol{\kappa}(t)$ is the Gaussian white noise transformed into the new variables, with correlation structure $\langle \kappa_i(t) \kappa_j(t') \rangle = \delta(t - t') B_{ij}(\boldsymbol{z}, \boldsymbol{w})$. The noise-correlation matrix $B(\boldsymbol{z}, \boldsymbol{w})$ can be partitioned as

$$B(\boldsymbol{z}, \boldsymbol{w}) = \begin{pmatrix} \mathcal{B}_{11}(\boldsymbol{z}, \boldsymbol{w}) & \mathcal{B}_{12}(\boldsymbol{z}, \boldsymbol{w}) \\ \mathcal{B}_{21}(\boldsymbol{z}, \boldsymbol{w}) & \mathcal{B}_{22}(\boldsymbol{z}, \boldsymbol{w}) \end{pmatrix}. \quad (6.12)$$

The partitioned matrix $\mathcal{B}_{11}$ is an $r \times r$ matrix and $\mathcal{B}_{22}$ is $(m-r) \times (m-r)$. The exact system has now been separated as much as possible into the fast and slow components.

Let us consider the system about the fixed point $(z^*, w^*) = x^*$. If all the eigenvalues of the system's Jacobian are real, then the linearised system is of the same from as the deterministic Eq. (3.4), but with an extra noise term with the correlation structure $B_{ij}(z^*, w^*)$ (see Eq. (2.76)). The Jacobian in this basis is simply the eigenvalue matrix Λ. In this form it is clear that in the region of the fixed point, a separation of timescales exists in the deterministic equations.

The methodology described above was used to arrive at the transformed systems Eqs. (3.16), (3.47) and (3.60) in Chap. 3. In Chap. 4, when dealing with the metapopulation Moran model, the decision was made to not explicitly give the dynamical equations for z and w (defined in Appendix F). This was because a projection matrix was instead used later in the calculation to isolate the dynamics in the relevant directions. Here we will work in the transformed basis however, so that the two methods may be more easily compared.

The assumption is now made in both methods, that the fast-variables w relax to their static value on a much faster timescale than the slow-variables z. To enforce this deterministically, the variables w are evaluated on their nullcline (see Eq. (3.8)). The deterministic system has now been reduced to r variables. To complete the description of the reduced stochastic model requires a decision on how to treat the noise. It is here that the methods introduced in Chaps. 3 and 4 diverge. For clarity, we begin by noting that in both the conditioning and projection matrix method, the reduced r-variable system has the form

$$\frac{dz}{dt} = \bar{A}(z) + \frac{1}{\sqrt{N}}\zeta(t), \qquad (6.13)$$

where

$$\bar{A}_i(z) = A_i(z, w(z)), \quad i = 1, \ldots r, \qquad (6.14)$$

the vector function $w(z)$ is the w-nullcline, and the r elements of the Gaussian white-noise vector $\zeta(t)$ have an as-yet-unspecified correlation structure, $\bar{B}(z)$. The structure of $\bar{B}(z) = \langle \zeta(t)\zeta^T(t') \rangle$ is dependent on which of the two fast-variable elimination methods is used.

In the conditioning method, the noise covariance matrix $\bar{B}(z)$ is that of Eq. (6.12) conditioned on $\kappa_i(t) = 0 \ \forall i = r+1, \ldots m$. From Eq. (A.28) we see that this is given by

$$\bar{B}_c(z) = \mathcal{B}_{11}(z, w(z)) - \mathcal{B}_{12}(z, w(z))\mathcal{B}_{22}^{-1}(z, w(z))\mathcal{B}_{12}^T(z, w(z)), \quad (6.15)$$

where we have used the subscript c to distinguish the conditioned result.

In the projection matrix method, a projection matrix (defined in terms of the left- and right-eigenvectors of the systems Jacobian, $u^{(i)}$ and $v^{(i)}$) is applied to the

noise, which isolates the component of the noise in the slow-directions. It is this isolated component which forms the correlation matrix for $\zeta(t)$. Constructing the $m \times r$ matrices U_r and V_r, whose ith columns are defined to be the ith left- and right-eigenvectors, $\boldsymbol{u}^{(i)}$ and $\boldsymbol{v}^{(i)}$ for $i = 1, \ldots r$, the equation for the projection matrix is [6]

$$P = V_r \left(U_r^T V_r \right)^{-1} U_r^T. \tag{6.16}$$

This is a generalisation of Eq. (4.33), for which the number of slow variables was one, $r = 1$. In the particular case of the system (6.11), the projection matrix takes a very simple from as a result of the fact that the system is already in the fast/slow basis. The left- and right-eigenvectors are equal (Λ is symmetric) and one finds a general P

$$P = \begin{pmatrix} I_r & 0_{r,m-r} \\ 0_{m-r,r} & 0_{m-r,m-r} \end{pmatrix}, \tag{6.17}$$

where I_r is the $r \times r$ identity matrix and the $0_{k,l}$ are $k \times l$ zero-matrices. Applying this projection to the noise $\boldsymbol{\kappa}(t)$, one finds that the form of the correlations in $\zeta(t)$ is

$$\begin{aligned} \left[\bar{B}_p(z) \right]_{ij} &= \left[P B(z, \boldsymbol{w}) P^T \right]_{ij} \quad i, j = 1, \ldots, r \\ \bar{B}_p(z) &= \mathcal{B}_{11}(z, \boldsymbol{w}(z)), \end{aligned} \tag{6.18}$$

where the subscript p now refers to the result of the projection matrix method. As stated in Sect. 4.1, any noise corresponding to the fast-directions is effectively ignored. Once again it is noted that, perhaps by chance, the projection matrix method predicts a noise covariance matrix in the reduced system, $\bar{B}_p(z)$, which is the same as the covariance matrix of the marginal distribution of the noise in the full system (see Appendix A, Eq. (A.24)).

6.2.1 Applying the Methods to a Linear System with Additive Noise

It is at this point unclear which method in general gives a better approximation to the full system, (though we recall from Sects. 3.5 and 4.1 that if the noise matrix $B(z, \boldsymbol{w})$ is singular, the conditioning method gives a qualitatively worse description of the dynamics than the projection matrix). To make progress we begin by restricting our attention to a system with a linear drift term and a constant diffusion term, so that the associated FPE of the system is linear (see Sect. 2.4). Such a system could also be arrived at if one linearised Eq. (6.11). Further, let us assume that all the eigenvalues of the system are real and negative. The variables in the slow/fast basis are denoted $\boldsymbol{\xi}_z$ and $\boldsymbol{\xi}_w$, to highlight the linearity of the dynamics. In this basis, the system has the same form as Eq. (3.4) but with an additive noise term;

$$\frac{d}{dt}\begin{pmatrix}\xi_z\\\xi_w\end{pmatrix} = \Lambda\begin{pmatrix}\xi_z\\\xi_w\end{pmatrix} + \kappa(t), \tag{6.19}$$

where the correlation structure of κ is given by Eq. (2.69) and the noise covariance matrix can be partitioned as Eq. (6.12), but is independent of the state of the system. The matrix Λ is once again a diagonal matrix of eigenvalues $\lambda^{(i)}$. If we finally assume that the boundaries to the problem lie at $\pm\infty$, the PDF described by the associated FPE is Gaussian [2] (since the FPE is linear) and can thus be described in terms of the time-evolution of its mean and covariance (see Sect. 2.7).

The equations for the mean and covariance are soluble; the solutions for the mean quantities, $\langle\xi_z\rangle$ and $\langle\xi_w\rangle$, have the same form as Eq. (2.72), while the solution for the covariance matrix Ξ^1 is given by Eq. (2.80). Since in this particular example the Jacobian is given by the diagonal matrix Λ, the solution for the components of Ξ can be expressed in the particularly neat form

$$\Xi_{ij} = \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)B_{ij}, \quad i, j = 1, \dots m. \tag{6.20}$$

In the approximation procedures, a reduced form of the SDEs is obtained under the assumption that the inequalities (3.2) hold, that is $|\lambda^{(m)}|, \dots, |\lambda^{(r+1)}| \gg |\lambda^{(r+1)}|, \dots, |\lambda^{(1)}|$. How does such an assumption impact on the form of Eq. (6.20)? To investigate this, a rather extreme limit can be taken. By taking the limit $\lambda^{(j)} \to -\infty \, \forall\, j = r + 1, \dots m$, the inequalities (3.2) are enforced to the greatest possible degree. The elements of the distribution (6.20) then take the limiting form

$$\Xi_{ij} \to \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)B_{ij}, \quad i, j = 1, \dots, r \tag{6.21}$$

$$\Xi_{kl} \to 0, \qquad\qquad\qquad k, l = r + 1, \dots, m$$

$$\Xi_{il} = \Xi_{li} \to 0, \qquad\qquad i = 1, \dots, r \quad l = r + 1, \dots, m.$$

How does Eq. (6.21) compare with the predictions of the reduced systems? The reduced linear systems are obtained in the same manner as Eq. (6.13). Introducing the partitioned $(r \times r)$ Jacobian

$$\bar{\Lambda}_{ij} = \Lambda_{ij}, \qquad \forall\, i, j = 1, \dots r, \tag{6.22}$$

which takes the place of $\bar{A}(z)$ in Eq. (6.13), the reduced system is

$$\frac{d}{dt}\xi_z = \bar{\Lambda}\xi_z + \zeta(t), \tag{6.23}$$

---

[1] The matrix Ξ is the covariance matrix for the Gaussian PDF $p(\xi_z, \xi_w, t)$ described by the linear FPE. It is not to be confused with the diffusion matrix $B(x)$ which describes the covariance of the noise terms, $\kappa_i(t)$, in the SDE.

where $\langle \zeta(t)\zeta^T(t')\rangle = \bar{B}\delta(t - t')$ and the structure of $\bar{B}$ is dependent on the fast-variable elimination method used.

The solution for the time-evolution of the mean of $\boldsymbol{\xi}_z$ is clearly unaltered in both methods, since the system is linear and $\bar{\Lambda}$ is diagonal and partitioned from Λ. In both the conditioning method and the projection matrix method therefore, the evolution of $\langle \boldsymbol{\xi}_z\rangle$ is given precisely by the reduced system, while $\langle \boldsymbol{\xi}_w\rangle$ is assumed to relax instantaneously to its stationary value, 0.

What form does the the covariance matrix in the reduced system take? Denoting the covariance matrix of the reduced system's PDF $\bar{\Xi}$, from Eq. (2.80) we find

$$\bar{\Xi}_{ij} = \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)\bar{B}_{ij}, \quad i, j = 1, \ldots r. \tag{6.24}$$

Employing the conditioning method, $\bar{B}$ is taken to be $\bar{B}_c$ in Eq. (6.15). Substituting this into the equation for $\bar{\Xi}$, on obtains

$$\bar{\Xi}_{ij} = \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)\bar{B}_{c\,ij} \tag{6.25}$$

$$= \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)\left[\mathcal{B}_{11} - \mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{12}^T\right]_{ij}, \quad i, j = 1, \ldots r. \tag{6.26}$$

The projection matrix method meanwhile gives $\bar{B} = \bar{B}_p$, which indicates a $\bar{\Xi}$ of the form

$$\bar{\Xi}_{ij} = \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)\bar{B}_{p\,ij} \tag{6.27}$$

$$= \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)[\mathcal{B}_{11}]_{ij} \tag{6.28}$$

$$= \left(\frac{e^{(\lambda^{(i)}+\lambda^{(j)})t} - 1}{\lambda^{(i)} + \lambda^{(j)}}\right)B_{ij}, \quad i, j = 1, \ldots r. \tag{6.29}$$

Comparing these two results to Eq. (6.21), we see that for a system with a linear FPE, the reduced system obtained via the projection matrix method provides a better approximation to the full variable system than conditioning. We would however expect the two methods to give similar results if the correlation between the fast and slow noise variables, $\mathcal{B}_{12} = \mathcal{B}_{21}^T$, was small. This will be explored in the following section in which we consider the relation to the two methods when applied to the neutral two-deme metapopulation Moran model. While this system is again linear in the drift term, it is distinct from the above example in that it has multiplicative noise, and therefore its associated FPE is non-linear.

6.2.2 Application to Neutral Two-Deme Metapopulation Moran Model

The neutral two-deme metapopulation Moran model is described by the SDEs (2.67), with a drift vector given by Eq. (4.9) and diffusion matrix by Eq. (4.10) with $\mathcal{D} = 2$. In order to align the discussion more closely with that presented in Sect. 6.2.1, we will begin by transforming the system into a new set of fast and slow variables w and z, defined as in Eq. (3.5) with x replaced by x_1 and y by x_2. The transformation matrix can be explicitly calculated to be

$$V = \begin{pmatrix} 1 & -\frac{\beta_2 G_{12}}{\beta_1 G_{21}} \\ 1 & 1 \end{pmatrix}. \tag{6.30}$$

The SDEs in z and w then take the form

$$\begin{aligned} \frac{dz}{dt} &= \qquad\quad 0 \qquad\quad + \frac{1}{\sqrt{N}}\kappa_z(t), \\ \frac{dw}{dt} &= -\frac{(\beta_2 G_{12} + \beta_1 G_{21})}{\beta_1 \beta_2}w + \frac{1}{\sqrt{N}}\kappa_w(t), \end{aligned} \tag{6.31}$$

where $\kappa(t)$ has the transformed correlation structure specified by Eq. (3.27), the exact form of which is too lengthy to be worthwhile reproducing here, though we note that it is now non-diagonal. Applying both reduction methods we obtain an SDE of the form of Eq. (6.13) in terms of the single z variable, with $\bar{A}(z) = 0$. As usual the correlation structure for $\zeta(t)$ is different for the conditioning and projection matrix methods, though we note that in this case the functional form is the same;

$$\bar{B}_c = 2b_c z(1-z), \qquad \bar{B}_p = 2b_p z(1-z), \tag{6.32}$$

where

$$b_c = \frac{(G_{11} + G_{12})(G_{21} + G_{22})}{\beta_2^2(G_{11} + G_{12}) + \beta_1^2(G_{21} + G_{22})} \tag{6.33}$$

and

$$b_p = \frac{G_{12}^2(G_{21} + G_{22}) + G_{21}^2(G_{11} + G_{12})}{(\beta_2 G_{12} + \beta_1 G_{21})^2}. \tag{6.34}$$

Unlike the system examined in Sect. 6.2.1, the full PDF cannot easily be found for the full or reduced systems. Instead, the probability of fixation and unconditional time to fixation predicted from the reduced systems are compared. The predicted probability of fixation is independent of b_c and b_p in the respective cases (see Eqs. (2.59) and (5.7)). The conditioned model therefore provides an equally good approximation
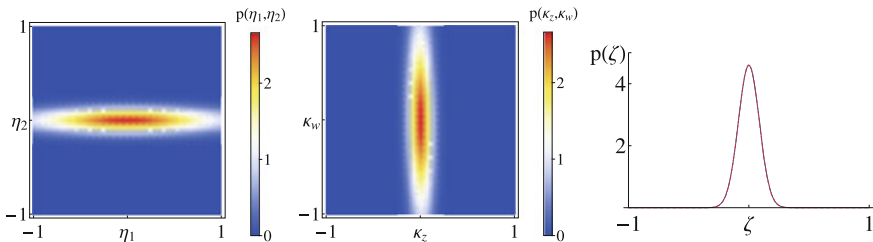
Fig. 6.5 *Left panel* Distribution of noise terms η_1, η_2 with covariance given by $B(\boldsymbol{x})$. *Central panel* Distribution of noise terms κ_1, κ_2 with covariance given by the transformed diffusion matrix $[V^{-1}]B(\boldsymbol{x})[V^{-1}]^T$. *Right panel* Distribution of the single noise term ζ for the reduced model. The *red, dashed line* is obtained from the conditioning method and hence the distribution has a variance equal to that of κ_z conditioned on $\kappa_w = 0$. The *blue line* is obtained from the projection matrix method, and so the distribution has the same variance of the marginal of κ_z. In all plots the diffusion matrix is evaluated at $(x_1, x_2) = (0.5, 0.5)$. The parameters used for this plot are $\boldsymbol{\beta} = (1, 1)$, and the unphysical but illustrative migration rates $m_{11} = 0.975$, $m_{12} = 0.994$, $m_{21} = 0.025$ and $m_{22} = 0.006$

for the probability of fixation. The time to fixation however is dependent on which form of $\bar{B}$ is taken (see Eq. (5.8)).

In Chap. 5, is has already been seen that using the parameter b_p obtained from the projection matrix method gives very good predictions for the time to fixation. What happens if instead the parameter b_c, obtained from the conditioning method, is used? Over many parameter regimes it is found that the conditioned system is comparable to the reduced system obtained from projection, as depicted in Fig. 6.5; in the leftmost panel the original noise distribution is plotted. The transformed noise distribution for κ_z and κ_w is plotted in the central panel. Since these noise terms have a small covariance, the conditional and marginal distributions are nearly identical, as seen in the rightmost panel. In Fig. 6.6, the same quantities are plotted for a system with a more symmetric migration matrix. The covariance of κ_z and κ_w is now slightly larger, and the two method begin to visibly differ in their predictions. This behaviour is magnified drastically in Fig. 6.7. In these figures, unphysical parameters have in some cases been chosen to illustrate the behaviour in each regime.

It is interesting to note that Figs. 6.5, 6.6 and 6.7 are ordered according to the size which they predict b_c/b_p to be. In Fig. 6.5, $b_p < (\sum_i \beta_i)^{-2}$ and therefore the time to fixation is relatively large. In Fig. 6.6, $b_p \approx (\sum_i \beta_i)^{-2}$ and therefore the time to fixation is roughly the same as for the well-mixed population. In Fig. 6.7, $b_p > (\sum_i \beta_i)^{-2}$ and therefore the time to fixation is relatively small. In this latter case, the result from the conditioning method considerably over-predicts the time to fixation (as opposed to the projection method which matches simulations extremely well).

How do we explain this behaviour? The problem essentially lies in the size of the noise in the reduced system. The diagonal terms in the diffusion matrix (6.12) are strictly positive and while the off-diagonal elements may be negative, the diffusion matrix is symmetric. Looking at Eqs. (6.15) and (6.18), we see that this leads to the following inequality, which holds for any system;
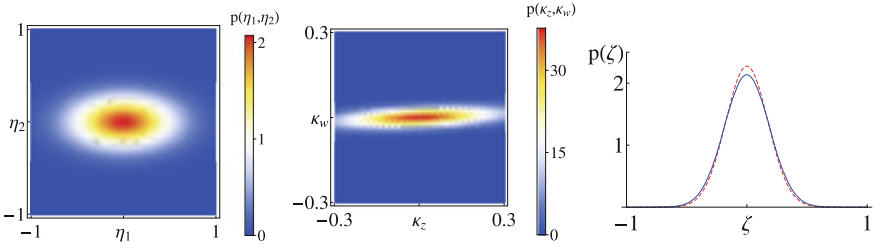
Fig. 6.6 The plots in this figure are the same as those in Fig. 6.5, but evaluated with the migration matrix $m_{11} = 0.95$, $m_{12} = 0.09$, $m_{21} = 0.05$ and $m_{22} = 0.91$. We note that while in Fig. 6.5 the distributions for ζ are virtually coincident, here some disparity can be observed
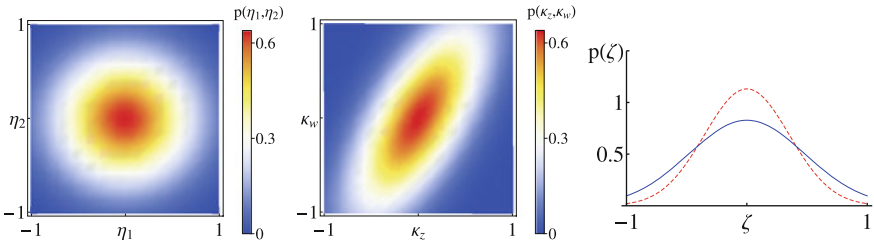

Fig. 6.7 Once again the plots in this figure are the same as those in Fig. 6.5, now evaluated with the migration matrix $m_{11} = 0.9999$, $m_{12} = 0.07$, $m_{21} = 1 \times 10^{-4}$, and $m_{22} = 0.93$. The large covariance of κ_z and κ_w leads to a larger discrepancy between the conditional and marginal distributions, and hence between the distributions for ζ from the conditioning (*red*, *dashed*) and projection matrix methods (*blue*)

$$\bar{B}_p \geq \bar{B}_c. \tag{6.35}$$

The degree to which $\bar{B}_p$ is greater than $\bar{B}_c$ depends on the specific system considered. This is indeed what we see in the case of the Moran model (6.31), most obviously in Fig. 6.7. In this case the reduced system obtained from conditioning under-predicts the magnitude of the noise because it discounts the probability of stochastic events along $\kappa_z = \kappa_w$ by forcing $\kappa_w = 0$. This leads to an over-estimation of the time to fixation relative to results obtained from simulation.

6.3 Reducing the Lotka-Volterra Model

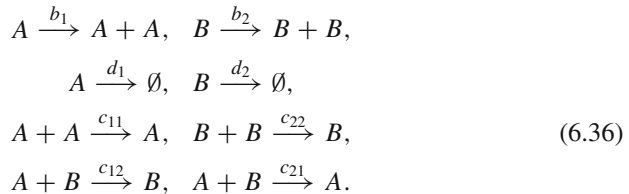In this thesis, the most prominent model considered has been the Moran model, with a special focus on one its variants, the metapopulation Moran model. It is a model which has been used extensively [15], especially in the last decade, where it has featured prominently in the study of stochastic game theory [9, 14]. In almost all of this work, the assumption that the system size N is fixed is hardly ever questioned.

While some authors attempt to address this issue with recourse to an effective population size, replacing variations in population, for instance, by an average [1, 3], such approaches are not satisfying. Excepting these attempts at justifying its formulation, the vast majority of authors simply accept this foundation as a necessary step towards attaining tractability. However, this results in a single rate parameter for the model, making the resulting effects of birth and death impossible to tease apart from each other, as well as other processes. As discussed in Sect. 2.9.1, the ambiguities inherent in this approach are particularly apparent when trying to include the effects of selection in the model.

Since the criticism of over-abstraction can be equally levelled against some of the work in this thesis, a different starting point will be described, which allows some of these questions to be addressed. A more ecologically-orientated approach is adopted by beginning from a population of n_1 haploid individuals which carry allele A and n_2 haploid individuals which carry allele B. They will reproduce at rates b_1 and b_2 respectively and die at rates d_1 and d_2. The types A and B then compete against individuals of their own type at rates c_{11} and c_{22} respectively. In the nomenclature of ecology, this is termed intraspecies competition (meaning competition within each species), though in a population genetics context, the types A and B are simply types of a single species. At rate c_{12}, the two types will compete to the detriment of type A, while at rate c_{21}, the number of B will decrease as a result of competition with type A. The effect of competition will be to regulate the population size, without imposing the condition $n_1 + n_2 = N$.

Although it is quite easy to sketch out this idea, showing how precisely this model relates to the Moran model, is not so straightforward, and there appear to be only a few instances in which this has been attempted. Each of these arguably features some drawbacks, for instance being too focused on deterministic level [7] or not providing a precise correspondence with Moran type models [10]. However it will be shown that a more systematic understanding of the relationship between the two approaches can be obtained via timescale separation arguments, more specifically, by application of the projection matrix method, first introduced in Chap. 4.

Let us begin by considering a well-mixed system, with a state entirely specified by the number of A and B alleles, denoted $\boldsymbol{n} = (n_1, n_2)$ respectively. This state will be able to change because of births, deaths or competition between individuals. In the notation of chemical reactions (see Sect. 2.4) the model is defined by the following rules:

$$A \xrightarrow{b_1} A + A, \quad B \xrightarrow{b_2} B + B,$$

$$A \xrightarrow{d_1} \emptyset, \quad B \xrightarrow{d_2} \emptyset,$$

$$A + A \xrightarrow{c_{11}} A, \quad B + B \xrightarrow{c_{22}} B, \tag{6.36}$$

$$A + B \xrightarrow{c_{12}} B, \quad A + B \xrightarrow{c_{21}} A.$$

To define the dynamics we need to specify the rates at which the allowed changes (6.36) take place. Assuming a law of mass action for the competitive interactions (see Eq. (2.35)) the transition rates are given by

$$
\begin{aligned}
T_1(n_1 + 1, n_2 | n_1, n_2) &= b_1 \frac{n_1}{V}, \\
T_2(n_1, n_2 + 1 | n_1, n_2) &= b_2 \frac{n_2}{V}, \\
T_3(n_1 - 1, n_2 | n_1, n_2) &= d_1 \frac{n_1}{V} + c_{11} \frac{n_1}{V} \frac{n_1}{V} + c_{12} \frac{n_2}{V} \frac{n_1}{V}, \\
T_4(n_1, n_2 - 1 | n_1, n_2) &= d_2 \frac{n_2}{V} + c_{22} \frac{n_2}{V} \frac{n_2}{V} + c_{21} \frac{n_1}{V} \frac{n_2}{V}.
\end{aligned}
\tag{6.37}
$$

The parameter V is not the total number of individuals in the system, which is free to vary. Rather it is a measure of the size of the system. For instance, in a terrestrial ecology it would be the area of land which the individuals inhabit. So typically it would be an area or a volume, but its precise value or even its dimensions can be left unspecified, as they can be absorbed into the rates b_i, d_i and c_{ij}. Together with the master equation (2.18), the transition rates (6.37) fully define the model.

Introducing the variables, $x_i \equiv n_i / V$, the diffusion approximation is made under the assumption that V is large and x_i approximately continuous. An expansion of the master equation can then be conducted as described in Sect. 2.4. This results in an FPE (2.30) with N replaced by V, and drift and diffusion terms which can be calculated using Eqs. (2.31), (2.32) and (6.36). The elements of these terms read

$$
A_i(\boldsymbol{x}) = (b_i - d_i) x_i - c_{ii} x_i^2 - c_{ij} x_i x_j,
\tag{6.38}
$$

and

$$
B_{ii}(\boldsymbol{x}) = (b_i + d_i) x_i + c_{ii} x_i^2 + c_{ij} x_i x_j,
\tag{6.39}
$$

where $i = 1, 2$ and $j \neq i$. The off-diagonal entries of the matrix B are equal to zero. In the limit $V \to \infty$, Eq. (2.67) reduces to the two deterministic differential equations $dx_i/d\tau = A_i(\boldsymbol{x})$, with $A_i(\boldsymbol{x})$ given by Eq. (6.38) and $\tau = t/V$. These are the Lotka-Volterra equations for two competing species [11, 13].

6.3.1 The Neutral Case

To begin the analysis, the individuals of type A and B are assumed to have equal fitness. Thus the theory is neutral, and A and B have equal birth, death and competition rates: $b_i \equiv b_0, d_i \equiv d_0, c_{ij} \equiv c_0$. A simulation of the original IBM defined by Eqs. (6.36) and (6.37) is shown in Fig. 6.8, where it is seen that the trajectories quickly collapse onto a line in the x_1-x_2 plane. A separation of timescales thus exists, which we wish to exploit.

In order to apply the projection matrix method, we begin as described in Chap. 4, by looking for fixed points of the dynamics. Taking the combinations $A_1 \pm A_2$ we find that the fixed points are solutions of the two equations

$$[(b_0 - d_0) - c_0 (x_1 + x_2)] (x_1 \pm x_2) = 0. \tag{6.40}$$

We see that, apart from the trivial fixed point $x_1 = x_2 = 0$, there is a line of fixed points given by

$$x_1 + x_2 = \frac{(b_0 - d_0)}{c_0}. \tag{6.41}$$

The neutral Lotka-Volterra model thus features a centre manifold, plotted as a blue line shown in Fig. 6.8. Further insight can be gained by calculating the Jacobian at points on Eq. (6.41). One finds that it has eigenvalues $\lambda^{(1)} = 0$ and $\lambda^{(2)} = -(b_0 - d_0)$. The time scale for the collapse onto the centre manifold is thus given by $|\lambda^{(2)}|^{-1} = (b_0 - d_0)^{-1}$. The situation is entirely reminiscent of that identified for the metapopulation Moran model in Chap. 4. The stochastic dynamics (shown in red in Fig. 6.8) are dominated by the deterministic dynamics far from the centre manifold, and there is a rapid collapse to its vicinity. Fluctuations taking the system too far away from the centre manifold are similarly countered by the deterministic dynamics dragging the system back. The net result is a drift along the centre manifold until either of the axes are reached and fixation of one of the types is achieved. This effect has also been noted and exploited in [4, 5] under the investigation of the evolution of dispersion.

Having identified this separation in timescales, we can now apply the projection matrix method to the system. Before proceeding however, it is useful to rescale both time and the variables in order to eliminate various constants:

Fig. 6.8 A stochastic simulation of the model specified by Eqs. (6.36) is plotted in *red*, along with the mean deterministic behaviour (given by $A_i(x)$ in Eqs. (6.38)) in *grey*. Parameters used here are for the neutral model, with $b_0 = 2$, $d_0 = 1$, $c_0 = 0.6$ and $V = 300$. The stochastic system follows an approximately deterministic trajectory until it reaches the centre manifold, plotted in *blue* and given by Eq. (6.41)
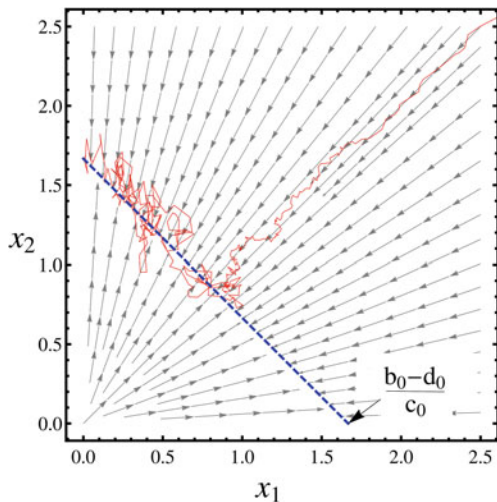
$$y_1 = \frac{c_0}{(b_0 - d_0)} x_1, \quad y_2 = \frac{c_0}{(b_0 - d_0)} x_2, \quad \tilde{\tau} = (b_0 - d_0)\tau. \qquad (6.42)$$

The elements of the drift vector in the new variables are

$$\tilde{A}_1(\boldsymbol{y}) = y_1(1 - y_1 - y_2), \qquad \tilde{A}_2(\boldsymbol{y}) = y_2(1 - y_1 - y_2), \qquad (6.43)$$

with the centre manifold now clearly given by $y_2 = 1 - y_1$. The components of the diagonal noise covariance matrix are now

$$\tilde{B}_{11}(\boldsymbol{y}) = \frac{c_0}{b_0 - d_0} \left(\frac{(b_0 + d_0)}{b_0 - d_0} y_1 + y_1(y_1 + y_2) \right), \qquad (6.44)$$

$$\tilde{B}_{22}(\boldsymbol{y}) = \frac{c_0}{b_0 - d_0} \left(\frac{(b_0 + d_0)}{b_0 - d_0} y_2 + y_2(y_1 + y_2) \right). \qquad (6.45)$$

The rescaled eigenvalues are $\tilde{\lambda}^{(1)} = 0$ and $\tilde{\lambda}^{(2)} = -1$. The corresponding eigenvectors are given by

$$\boldsymbol{v}^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad \boldsymbol{u}^{(1)} = \begin{pmatrix} 1 - y_1 \\ -y_1 \end{pmatrix}, \qquad (6.46)$$

and

$$\boldsymbol{v}^{(2)} = \begin{pmatrix} y_1 \\ 1 - y_1 \end{pmatrix} \qquad \boldsymbol{u}^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \qquad (6.47)$$

normalised such that Eq. (4.19) holds.

The situation is slightly more complicated than that which was found for the neutral metapopulation Moran model. In Sect. 4.4, since the deterministic terms of the model's SDEs were linear, the eigenvectors were simply some combination of constants. In this system, the deterministic term is non-linear, even in the neutral case. The result of this is that, though a centre manifold exists, the fast and slow directions in its vicinity change depending on the position of the system along the line. This can be seen in Fig. 6.8, in which the direction of the dominant fluctuations of the stochastic trajectory change depending on the system's relative position along the centre manifold. Despite this difference in the form of the fast and slow directions, the projection matrix can still be constructed and applied to the system. Note however, that a change of variables of the type defined for the metapopulation Moran model (see Eq. (F.2) is now no longer straightforward, since $\boldsymbol{v}^{(1)}$ is not a constant. The choice is therefore made to denote the coordinate along the centre manifold as z, and choose this to be equal to y_1, although of course many other choices are possible.

As in Chap. 4, a projection matrix, defined by Eqs. (4.33) and (6.46) and (6.46), is applied to the SDEs for the model. This removes the components of the dynamics in the fast directions, while keeping the slow mode intact. Applying the projection operator to the first term on the right-hand side of Eq. (6.43) gives $P\tilde{\boldsymbol{A}} = 0$. Applying

the projection matrix to the noise in the SDE, one sees that the noise in the reduced system must have correlations

$$\langle \zeta(\tilde{\tau})\zeta(\tilde{\tau}')\rangle = \left[P_{11}^2 \tilde{B}_{11}(x, y) + P_{12}^2 \tilde{B}_{22}(x, y) \right] \delta\left(\tilde{\tau} - \tilde{\tau}'\right). \tag{6.48}$$

The reduced system then takes the form

$$\frac{dz}{d\tilde{\tau}} = \bar{A}(z) + \frac{1}{\sqrt{V}}\zeta(\tilde{\tau}), \tag{6.49}$$

where $\bar{A}(z) = 0$ and where $\zeta(\tilde{\tau})$ is a Gaussian noise with zero mean and correlator

$$\langle \zeta(\tilde{\tau})\zeta(\tilde{\tau}')\rangle = \bar{B}(z)\delta\left(\tilde{\tau} - \tilde{\tau}'\right); \quad \bar{B}(z) = 2\frac{b_0 c_0}{(b_0 - d_0)^2}z(1 - z). \tag{6.50}$$

Let a new parameter N be defined such that

$$N = \frac{(b_0 - d_0)}{c_0}V. \tag{6.51}$$

From Eq. (6.41), this can be interpreted as the discrete size of the population on the centre manifold. Substituting this into Eq. (6.50), we find that the reduced model is exactly the Moran model in rescaled time $\bar{t} = V b_0 \tilde{\tau}/c_0$, where z is the fraction of type A alleles and N is the total population size. The neutral form of the Lotka-Volterra model therefore reduces to precisely the Moran model, under the projection matrix method.

As in the case of the neutral metapopulation Moran model, the probability of fixation and mean unconditional time to fixation can be calculated as a function of the initial condition. The results are given by the familiar expressions Eqs. (2.62) and (2.54) with x_0 replaced by z_0 and time now measured in $\bar{t} = [b_0(b_0 - d_0)/c_0] t$. The reduced theory matches the results of Gillespie simulations extremely well. Illustrative cases are given by the green plots in Figs. 6.10 and 6.11.

6.3.2 The Non-neutral Case

It is now natural to ask what model is obtained by the elimination of the fast modes of the non-neutral Lotka-Volterra model. As is usual in population genetics we will work to linear order in the selection strength, s, and so begin by writing

$$b_i = b_0 \left(1 + \epsilon\beta_i\right), \ d_i = d_0 \left(1 + \epsilon\delta_i\right), \ c_{ij} = c_0 \left(1 + \epsilon\gamma_{ij}\right), \tag{6.52}$$

where ϵ is a small parameter which will later be related to s. The constants β_i, δ_i and γ_{ij} are assumed to be of order one. In terms of the rescaled variables y_1 and y_2 defined in Eq. (6.42), the drift vector is given by

$$A_1(\boldsymbol{y}) = y_1(1 - y_1 - y_2) + \varepsilon y_1 \left[\frac{(b_0\beta_1 - d_0\delta_1)}{b_0 - d_0} - \gamma_{11}y_1 - \gamma_{12}y_2 \right],$$

$$A_2(\boldsymbol{y}) = y_2(1 - y_1 - y_2) + \varepsilon y_2 \left[\frac{(b_0\beta_2 - d_0\delta_2)}{b_0 - d_0} - \gamma_{21}y_1 + \gamma_{22}y_2 \right]. \qquad (6.53)$$

The diagonal components of the diffusion matrix are meanwhile unchanged from the neutral case to leading order (see Eq. (6.45)).

Although for $\epsilon \neq 0$ there will not be a centre manifold along which there are no deterministic dynamics, we still expect there to be separation of timescales which will allow us to identify fast and slow variables. This expectation is supported by Gillespie simulation of the stochastic process, as illustrated in Fig. 6.9. We pick out the slow subspace, and so eliminate the fast deterministic dynamics, by setting the product $\boldsymbol{u}^{(2)} \cdot \tilde{\boldsymbol{A}}(\boldsymbol{x})$ equal to zero, as described in Sect. 4.4.1. This leads to an equation of the form $y_2 = 1 - y_1 + \epsilon f(y_1) + \mathcal{O}(\epsilon^2)$, where $f(y_1)$ is quadratic in y_1. In order to make a comparison to the Moran model, it is required that the line passes through the points $(x, y) = (1, 0)$ and $(x, y) = (0, 1)$, which implies that $f(0) = 0$ and
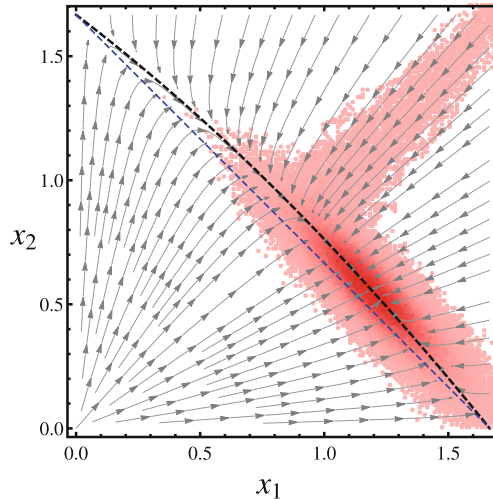

Fig. 6.9 Plots of trajectories for the non-neutral model in the variables x_1 and x_2. In *grey*, potential deterministic trajectories are plotted. A histogram of 50 stochastic trajectories is overlaid in *red*. The centre manifold of the neutral theory is plotted as a *blue line*, while the slow subspace is plotted in *black*. The stochastic trajectories can be seen to quickly collapse to the slow subspace, about which they are confined. Parameters used are $V = 600$, $\epsilon = 0.08$, $b_0 = 2$, $d_0 = 1$, $c_0 = 0.6$, $\delta_1 = \delta_2 = 0$, $\gamma_{11} = \gamma_{22} = 1$, $\gamma_{12} = -1$ and $\gamma_{21} = 0$

$f(1) = 0$. This leads to the two conditions

$$\beta_i = \frac{\gamma_{ii}(b_0 - d_0) + d_0\delta_i}{b_0}, \quad i = 1, 2.$$ (6.54)

With this choice of the birth rates, the slow subspace takes the form

$$y = (1 - x)\left[1 + \epsilon x\,(\gamma_{11} + \gamma_{22} - \gamma_{12} - \gamma_{21}) + \mathcal{O}\left(\epsilon^2\right)\right].$$ (6.55)

It is interesting to note that the conditions in Eq. (6.54) have also eliminated any reference to the death rates δ_i in Eq. (6.53). Additionally, introducing the parameter

$$\Gamma \equiv \gamma_{11} + \gamma_{22} - \gamma_{12} - \gamma_{21},$$ (6.56)

it can be seen that as long as the competition rates do not satisfy $\Gamma = 0$, the slow subspace will be curved. Simulations show that to an excellent approximation, the deterministic system collapses down to a line given by Eq. (6.55), as shown in Fig. 6.9.

The effective stochastic dynamics on the slow subspace are found by applying the same arguments as in the neutral case. Just as in the study of the metapopulation Moran model, it is assumed that the fast and slow directions do not vary significantly from the neutral case, so that the same form of the projection operator can be used when $\epsilon \neq 0$ as was used when $\epsilon = 0$. Therefore applying P_{ij} (given by Eq. (4.33) with Eqs. (6.46) and (6.47)) to Eqs. (6.53) and (6.45), gives Eq. (6.49), but now with

$$\bar{A}(z) = -\epsilon z\,(1 - z)\,[\Gamma z + (\gamma_{12} - \gamma_{11}) + \mathcal{O}\,(\epsilon)].$$ (6.57)

Before discussing the relation between the reduced model (defined by Eqs. (6.49), (6.50) and (6.57)) and the Moran model, the validity of the procedure will be tested by comparing the results for the probability of fixation and the time to fixation found from the reduced model to simulation of the original IBM. Since Eqs. (6.57) and (6.50) have the same functional form as Eqs. (5.14) and (4.25), the solution to the probability of fixation can be calculated in the same manner as in the metapopulation Moran model taken to $\mathcal{O}(s^2)$, outlined in Appendix G. The fixation probability can once again be written

$$Q(z_0) = \frac{1 - \chi(z_0)}{1 - \chi(1)}.$$ (6.58)

In this case the function $\chi(z_0)$ is written in terms of $l(z_0)$ such that

$$l(z_0) = \sqrt{\frac{(b_0 - d_0)N\varepsilon}{2b_0|\Gamma|}}((\gamma_{11} - \gamma_{12}) - \Gamma z_0)$$ (6.59)

and

$$\chi(z_0) = \frac{\text{erfc}\,[l(z_0)]}{\text{erfc}\,[l(0)]}, \quad \text{if } \Gamma < 0, \tag{6.60}$$

$$\chi(z_0) = \frac{\text{erfi}\,[l(z_0)]}{\text{erfi}\,[l(0)]}, \quad \text{if } \Gamma > 0. \tag{6.61}$$

An illustrative plot of the fixation probability is given in red in Fig. 6.10. The agreement between reduced theory and simulation is excellent. The mean unconditional time to fixation can also be calculated using Eq. (2.52) with the drift and diffusion terms of the reduced model. Once again the agreement between the reduced theory and simulation, depicted in red in Fig. 6.11, is very good.

It is interesting to note that the drift term in the reduced Lotka-Volterra model with selection, Eq. (6.57), does not have the same form as the well-mixed Moran model with selection (see Eq. (2.99)). Instead there exists a Γz term, which gives an additional z-dependence. Such a functional form for the drift term can be obtained from a game theoretic formulation of the Moran model [8]; however there is no simple mapping between the payoff matrix in game theory and the matrix γ, related to the competition matrix of the Lotka-Volterra model. The behaviour of the reduced Lotka-Volterra model can of course be mapped onto the standard Moran model with selection, Eq. (2.99), if one final condition on the parameters is added, namely $\Gamma = 0$. In this case, the probability of fixation takes on the same form as Eq. (2.100):

$$Q(z_0) = \frac{1 - \exp\left[-\varepsilon(\gamma_{11} - \gamma_{12})(b_0 - d_0)N z_0/b_0\right]}{1 - \exp\left[-\varepsilon(\gamma_{11} - \gamma_{12})(b_0 - d_0)N/b_0\right]}. \tag{6.62}$$
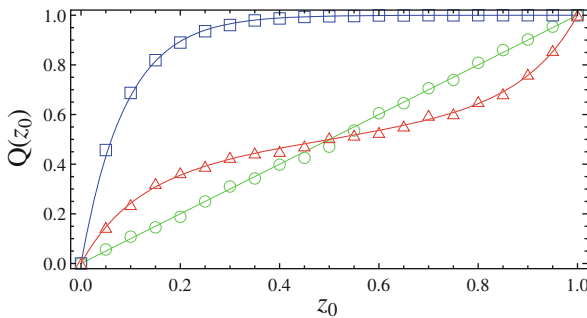

Fig. 6.10 Probability of fixation, $Q(z_0)$ as a function of projected initial condition. *Continuous lines* are obtained from reduced one-dimensional theory, whilst markers have been obtained from Gillespie simulations. *Green circles* are obtained from a neutral system with $V = 150$, $b_0 = 3.1$, $d_0 = 1.1$ and $c_0 = 0.4$. *Blue square* markers are obtained from simulations of the non-neutral model, with parameters $V = 300$, $\epsilon = 0.01$, $b_0 = 2$, $d_0 = 1$, $c_0 = 0.2$, $\gamma_{11} = 1$, $\gamma_{12} = -0.5$, $\gamma_{21} = 2$, $\gamma_{22} = 0.5$. Note that this results in $\Gamma = 0$ (see Eq. (6.56)). *Red triangles* are obtained from a simulation with parameters $V = 500$, $\epsilon = 0.015$, $b_0 = 2$, $d_0 = 1$, $c_0 = 0.8$, $\gamma_{11} = \gamma_{22} = 1$, $\gamma_{12} = \gamma_{21} = -1$
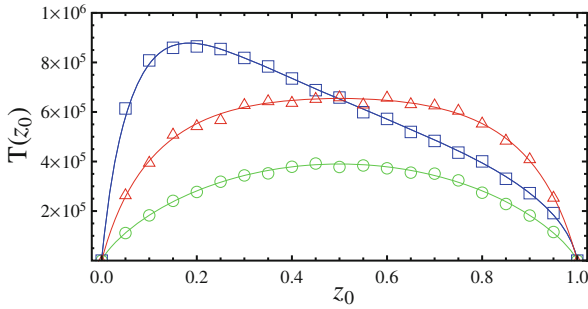
Fig. 6.11 The mean unconditional time to fixation $T(z_0)$, in terms of the time variable $\bar{t}$. *Continuous lines* are obtained from the reduced one-dimensional theory, whilst markers have been obtained from Gillespie simulations of the full model. Colours/symbols are related to sets of parameters as defined in Fig. 6.10

Comparing this with Eq. (2.100), the effective selection pressure can then be seen to be given by

$$s = \frac{\varepsilon(\gamma_{11} - \gamma_{12})(b_0 - d_0)}{b_0} + \mathcal{O}(\varepsilon). \tag{6.63}$$

The probability of fixation for a system in which this condition holds is plotted in blue in Fig. 6.10. The corresponding mean time to fixation is plotted in blue in Fig. 6.11.

While it is possible to obtain a Moran model of the same form as Eq. (2.99) as the limiting form of the Lotka-Volterra competition model, it is difficult to place a precise physical meaning on the condition $\Gamma = 0$. Further, if one were to take a random collection of the matrix elements γ_{ij}, the subset that fulfilled this condition would be vanishingly small. In an ecological context, Γ would generally be taken to be positive. This is by the argument that competition within a species is always more intense than competition between species, as each species tends to inhabit its own specialised niche. In the context of population genetics, it is not entirely clear that this is the case, and careful thought needs given to the interpretation of the results before any strong conclusions can be made. What is clear however, is that allowing a variable population size has a non-trivial effect on the dynamics of the resulting non-neutral model. It would therefore be interesting to extend the Lotka-Volterra competition model to incorporate other evolutionary effects, such as mutation and migration, and thus test the robustness of results derived from the Moran model.

References

1. R. Halliburton, *Introduction to Population Gentetics* (Pearson Press, New Jersey, 2004)
2. M. Lax, Fluctuations from the nonequilibrium steady state. Rev. Mod. Phys. **32**, 25–64 (1960)
3. C.C. Li, *Population Genetics* (The University of Chicago Press, Chicago, 1955)

4. Y.T. Lin, H. Kim, C.R. Doering, Demographic stochasticity and evolution of dispersion I. Spatially homogeneous environments. J. Math. Biol. **70**, 647 (2015)
5. Y.T. Lin, H. Kim, C.R. Doering, Demographic stochasticity and evolution of dispersion II. Spatially inhomogeneous environments. J. Math. Biol. **70**, 679 (2015)
6. C.D. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Philadelphia, 2000)
7. A.E. Noble, A. Hastings, W.F. Fagan, Multivariate Moran process with Lotka-Volterra phenomenology. Phys. Rev. Lett. **107**, 228101 (2011)
8. M.A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, 2006)
9. M.A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, Emergence of cooperation and evolutionary stability in finite populations. Nature **428**, 646–650 (2004)
10. T. Parsons, C. Quince, Fixation in haploid populations exhibiting density dependence I: the non-neutral case. Theor. Pop. Biol. **72**, 121–135 (2007)
11. E.C. Pielou, *Mathematical Ecology* (Wiley, New York, 1977)
12. N. Rohner, D.F. Jarosz, J.E. Kowalko, M. Yoshizawa, W.R. Jeffery, R.L. Borowsky, S. Lindquist, C.J. Tabin, Cryptic variation in morphological evolution: Hsp90 as a capacitor for loss of eyes in cavefish. Science **342**, 1372–1375 (2013)
13. J. Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: An Introduction* (Macmillan, New York, 1979)
14. A. Traulsen, J.C. Claussen, C. Hauert, Coevolutionary dynamics: from finite to infinite populations. Phys. Rev. Lett. **95**, 238701 (2005)
15. D. Waxman, A unified treatment of the probability of fixation when population size and the strength of selection change over time. Genetics **188**, 907–913 (2011)

Chapter 7
Conclusion

In the introduction to this thesis, some time was spent discussing the choices one makes in modelling any system. In some sense we are playing a game by which we wish to incorporate enough detail so as to be realistic and informative, but not so much as to render the model resistant to interpretation. Having struck the balance between these competing considerations, we may be further confounded by the analytic intractability of the resulting problem. While the stochastic nature of the Moran model, makes it difficult to solve in its entirety, its one-dimensional nature makes other quantities, such as the fixation probability and fixation time, obtainable. However many models inspired by nature (especially those which are nonlinear and in many dimensions) stubbornly resist analytic treatment.

Much of theoretical physics and applied mathematics is concerned with deriving methods to solve (or approximately solve) such apparently intractable problems. In this thesis I have explored two methods of fast-variable elimination in stochastic systems. These techniques, described in Chaps. 3 and 4, are by no means the first to tackle such problems [1, 4, 13] (nor, without doubt, the last [3, 8]). However as described in Sect. 2.8, it is my opinion that many existing methods feature a fundamental philosophical drawback; while technically rigorous, they are often guilty of providing effective systems that while of lower dimension than the full system, are almost as complicated.

In contrast, I believe that the methods outlined in this thesis offer some distinct advantages. I would argue that both approaches are relatively simple: working in the SDE setting makes the analogy to deterministic slow-manifold theory clear, and the basic ideas are intuitively easy to grasp. Moreover, the techniques are generally applicable; there is no need for a parameter which controls the fast variable, since the fast direction is determined instead from a linear stability analysis. Nor do the techniques require knowledge of the deterministic trajectories, as is necessary in [9]. The methods are probably most similar to the direct adiabatic elimination method [14] or the Haken slaving method [6] in that they are formulated in the SDE setting and follow intuitively from deterministic fast-variable elimination. The methods introduced in this thesis are however entirely distinct; they do not introduce ill-defined noise terms in the reduced description and therefore can be used successfully on non-linear

SDEs with multiplicative noise. Perhaps what is most attractive about both methods however, is the nature of the resulting system. In general it is simple enough to understand (containing no non-Markovian components for instance) whilst maintaining a high degree of predictive power.

The conditioning method was the first of the two methods to be introduced. Arguably it follows deterministic fast-variable elimination in the most direct way, in that it restricts the dynamics in the fast directions, both deterministic and stochastic, to be zero. This is achieved by setting the fast variable equal to its nullcline value and conditioning the noise correlation matrix on the event that it is zero in the fast direction. Applied first to an illustrative ecological example in Sect. 3.2, the reduced model was found to predict the stationary distribution for the slow variable z well. Particularly impressive was the fact that the reduced description reproduced the behaviour of the model after the bifurcation of the single fixed point. This is in spite of the fact that the fast and slow directions had been identified for the single fixed point case only. The success here is a result of the fact that while the bifurcation introduces two new fixed points, the fast and slow directions remain relatively unchanged. Still, the robustness of the method to these variations is perhaps surprising.

In Sect. 3.4, conditioning was used to simplify a linearised SEIR model. Here the separation in timescale was essentially a product in the disparity between birth/death rates and infection/recovery rates. Using this example, it was shown that timescale separation can provide a computational benefit to the analysis of certain problems. In this particular case, the separation in timescales was the source of a numerical difficulty [12], which the reduced system avoided.

While the conditioning method performed very well overall, a particular area of concern was identified. If the noise covariance matrix of the full system is singular, the conditioning method can predict no noise in some of the remaining directions. Such a case is described in Sect. 3.5. To remedy this, the second of the elimination methods was developed.

The projection matrix method was first introduced in Chap. 4, and is identical to the conditioning method in its treatment of the deterministic component of a set of SDEs. Its treatment of the noise differs however. Rather than restricting the noise in the fast direction to zero (as in the conditioning method), a projection matrix is constructed which isolates the component of the noise in the slow direction. The noise in the reduced system is then simply the component of the noise in the slow direction of the full system. At first glance this seems drastic, however as is mentioned in Sect. 4.1, this is equivalent to taking the marginal distribution of the noise in the full system, as an approximation for the noise in the reduced system. The projection matrix method does not suffer the same drawbacks as the conditioning method when the noise covariance matrix is singular.

Of particular interest was the application to the metapopulation Moran model described in Sect. 4.4. Utilising the projection matrix method, one can move from a description of the dynamics in $\mathcal{D}$-variables (representing each subpopulation in the system) to an effective description in terms of one variable z. The resulting equations are then amenable to analysis, with explicit formulae for the parameters in the final,

simplified equations. These parameters can be straightforwardly calculated from the network structure of the islands.

The reduced, effective system was compared to the results obtained from a direct simulation of the IBM for a range of different networks and parameters in Chap. 5. The method gives excellent results for most networks and parameter values. Where it does not work so well, there are reasons this is to be expected. For instance, the magnitude of the real part of the non-zero eigenvalues of H should be greater than both s and $N^{-1/2}$, so that the separation of timescales is sufficient to apply the approximation, as detailed in Sect. 5.3. Furthermore, parameters modelling similar effects in the model are assumed to be of the same order. Therefore, no island is assumed to be an order or two of magnitude bigger than other islands (all β_i are of order one). Even if these conditions are violated, the elimination of fast variables may still be possible; if it is not, then a different calibration could describe some of these situations. For example, the diagonal elements of H could be made to scale with $N^{-1/2}$, which would result in a different set of formulae. In [2], the diagonal elements were chosen to scale like N^{-1}, which accounts in part for the differences between the results given in that paper and those given in this thesis.

Though the effective metapopulation Moran model has only been expressed to order s^2, the technique is capable of being generalised to higher orders in s. Given the typical size of selection strengths, working to this order is entirely reasonable, and indeed most authors only keep terms of order s. It should be emphasised that although s is frequently compared numerically to $N^{-1/2}$ or N^{-1}, s is not, as in some papers, set equal to $N^{-1/2}$ or N^{-1}; s and N are two independent parameters. One of the reasons for going to order s^2 is to show that the analysis of the stochastic aspects of migration-selection balance could be considerably extended using our results.

The model presented in Sect. 4.3 is slightly different from the previous work cited concerning migration-selection balance. However, I have shown that the projection matrix method, being more general, can be used to investigate a broader range of parameters than previously attempted, including non-symmetric migration, arbitrary deme topology and an arbitrary range of selective pressures across the demes [5, 7, 15, 16]. Further, the mechanism that allows the approximation to work so successfully is the dominance of the large and linear effect of migration (embodied by the matrix H in the metapopulation model) over smaller non-linear and stochastic terms. This allows the method to be extended, in a straightforward way, to other problems. In Sect. 6.1, an example of such an extension was given—the metapopulation Moran model with mutation. The form of the subpopulations and structure of the migration once again led to non-trivial behaviour relative to the well-mixed analogue. This behaviour was once again well-predicted by the reduced model.

In Sect. 6.2 the two methods of fast-variable elimination were compared in more detail. In addition to the case of singular diffusion matrix, it was found that the projection matrix method outperforms the conditioning method in two further ways. The first is that, in the case the system under consideration is equivalent to a linear FPE, the projection matrix method provides a quantitatively better approximation of the full system, as discussed in Sect. 6.2.1. The second is that in general the conditioning method has the potential to under-predict the magnitude of the noise.

As described in Sect. 6.2.2, this is a result of discounting strong correlations between noise in the fast and slow directions, which the projection matrix method picks up.

Despite this, the conditioning method does not do such a terrible job. In both methods we merely seek an approximation for the form of the noise. In the ecological example first explored in Sect. 3.2, it was found that the projection matrix method gave an approximation of the noise (see Eq. (4.3)) with a cleaner functional form than the conditioning method (see Eq. (3.24)). This allowed an analytic solution for the stationary distribution, Eq. (4.4), to be obtained. It is not inconceivable then, that there might arise particular situations in which the converse is true, that the conditioning method provides a reduced description more analytically tractable than the projection matrix method. While the result could still suffer from the drawbacks highlighted in Sects. 3.5 and 6.2, as a proxy for the form of the noise it may prove illuminating.

The relationship between both methods and those discussed in Sect. 2.8 in the SDE formalism have been discussed. It may be asked, quite fairly, 'what is the relation to other methods developed in the FPE formalism?'. The answer to this question is by no means clear and this is certainly an area that it would be interesting to investigate. Bridging this gap may also lead to a more mathematically rigorous underpinning to the theory, which without doubt it would be useful to have. The first step towards a formal proof of the convergence between the reduced and full system is given in Sect. 6.2.1. While this only holds for linear systems, it goes someway to explaining the success of the methods in the region of a fixed point. Ultimately, the full proof of convergence must be left for further work, though a comparison with the methods of Parsons and Quince may prove fruitful [10, 11].

Even though, for the moment, there is not a rigorous mathematical underpinning to the two methods detailed here, it is perhaps worthwhile making a final point. In the applications discussed in this thesis, it is hard to imagine the methods performing better. Further, I believe that even in light of more sophisticated methods, those outlined here provide at the very least a good starting point in the investigation of models featuring a separation of timescales. The models described in Sects. 3.4, 4.3 and 6.3 are canonical models, and yet studies utilising timescale separation arguments are comparatively rare in the literature. Perhaps this is because of the perceived difficulty of removing fast degrees of freedom from stochastic systems. Here however, I have shown that the complex machinery involved in stochastic normal form theory and FPE projection are not always entirely necessary. To a certain extent, the stochastic nature of the problems investigated contributes to the success of the techniques, as unimportant trends which the reduced model may not capture are averaged out when considering the ensemble.

A reasonable question one might ask is, does the behaviour of the metapopulation Moran model analysed in Chap. 5 survive if the population size is not fixed? The first steps toward answering the question have been taken in Sect. 6.3. There the Lotka-Volterra competition model was introduced. While originally developed as a model of interacting populations in ecology, there is nothing conceptually that prevents its use in modelling population genetics. Unlike the Moran model, the population in the Lotka-Volterra model is not held fixed artificially by coupling birth and death. Instead

the size of the population is moderated by competition between individuals. In this sense the Lotka-Volterra model can be seen to be preferable to the Moran model in terms of realism. Using the fast-variable elimination methods developed earlier in this thesis, it was shown that for neutral systems, the Lotka-Volterra competition model for two species behaves at long times like the haploid Moran model for two allele types. Once selection was incorporated, the relationship was not so straightforward. It was seen that, in general, selection operated in a way distinct from that observed in the Moran model.

The extension of the Lotka-Volterra model to many islands would of course be very interesting, and in particular an exploration as to the robustness of the behaviour predicted by the metapopulation Moran model. There are also a host of other extensions which would be interesting to explore, such as the incorporation of multi-allele effects, frequency-dependent selection and diploidy of the individuals. More generally, the application of the method to other models outside population genetics may be possible. For instance, one could investigate a metapopulation susceptible-infectious-recovered model. Here the subpopulations would be representative of cities, and the existence of a separation of timescales dependent on the relative size of the disease and migration parameters. Would it be possible, as in the case of the Moran model, to reduce the full system to an effective, single population model?

Systems exhibiting a separation of timescales are so ubiquitous in the real world, that often their removal is considered part of the modelling process. However, as has been demonstrated in this thesis, there are many systems in which such a separation is woven deeper in to the dynamics, so that it emerges naturally from the model. The techniques developed in this thesis have been useful in making analytic progress in a series of canonical models from epidemiology, population genetics and ecology. I hope that the methodologies expounded here will be taken up by other researchers and will lead to the analysis of more complex and realistic models.

References

1. L. Arnold, P. Imkeller, Normal forms for stochastic differential equations. Probab. Theory Relat. Fields **110**, 559–588 (1998)
2. G.J. Baxter, R.A. Blythe, W. Croft, A.J. McKane, Utterance selection model of language change. Phys. Rev. E **73**, 046118 (2006)
3. M. Bruna, S.J. Chapman, M.J. Smith, Model reduction for slow-fast stochastic systems with metastable behaviour. J. Chem. Phys. **140**, 174107 (2014)
4. C.W. Gardiner, Adiabatic elimination in stochastic systems. I. Formulation of methods and application to few-variable systems. Phys. Rev. A **29**, 2814–2822 (1984)
5. S. Gavrilets, N. Gibson, Fixation probabilities in a spatially heterogeneous environment. Popul. Ecol. **44**, 51–58 (2002)
6. H. Haken, A. Wunderlin, Slaving principle for stochastic differential equations with additive and multiplicative noise and for discrete noisy maps. Z. Phys. B **47**, 179–187 (1982)
7. B. Houchmandzadeh, M. Vallade, The fixation probability of a beneficial mutation in a geographically structured population. New J. Phys. **13**, 073020 (2011)
8. O. Kogan, M. Khasin, B. Meerson, D. Schneider, C.R. Myers, Two-strain competition in quasi-neutral stochastic disease dynamics. Phys. Rev. E **90**, 042149 (2014)

9. Y.T. Lin, H. Kim, C.R. Doering, Features of fast living: on the weak selection for longevity in degenerate birth-death processes. J. Stat. Phys. **148**, 646–662 (2012)
10. T. Parsons, C. Quince, Fixation in haploid populations exhibiting density dependence I: the non-neutral case. Theor. Pop. Biol. **72**, 121–135 (2007)
11. T. Parsons, C. Quince, Fixation in haploid populations exhibiting density dependence II: the quasi-neutral case. Theor. Pop. Biol. **72**, 468–479 (2007)
12. G. Rozhnova, A. Nunes, Stochastic effects in a seasonally forced epidemic model. Phys. Rev. E **82**, 041906 (2010)
13. G. Schöner, H. Haken, The slaving principle for Stratanovich stochastic differential equations. Z. Phys. B **63**, 493–504 (1986)
14. R. Serra, M. Andretta, M. Compiani, G. Zanarini, *Introduction to the Physics of Complex Systems* (Pergamon Press, Oxford, 1986)
15. H. Tachida, M. Iizuka, Fixation probability in spatially changing environment. Genet. Res. Camb. **58**, 243–245 (1991)
16. M.C. Whitlock, R. Gomulkiewicz, Probability of fixation in a heterogeneous environment. Genetics **171**, 1407–1417 (2005)

Appendix A
The Conditional and Marginal of a Multivariate Gaussian Distribution

A multivariate Gaussian distribution for the random m-dimensional vector x takes the form

$$f(x) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} \exp\left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right] \tag{A.1}$$

where μ is the mean value of x, Σ is the covariance matrix and $|\Sigma|$ its determinant. We wish to calculate the conditional distribution for this process, given some subset of x is fixed, and also the marginal distribution given that we integrate out a subset of the x variables. We begin by partitioning the vectors x and μ and the matrices Σ;

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{A.2}$$

where we note that $\Sigma_{21} = \Sigma_{12}^T$, and Σ_{11} and Σ_{22} are symmetric. Further, let the elements of the partitioned inverse of Σ be denoted with upper indices, so that

$$\Sigma^{-1} = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}. \tag{A.3}$$

Since Σ is symmetric, so is Σ^{-1}, so that $\Sigma^{12} = [\Sigma^{21}]^T$. Now, the partitioned elements of Σ and Σ^{-1} can be related by noting that

$$\Sigma\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix} \tag{A.4}$$

$$= \begin{pmatrix} \Sigma_{11}\Sigma^{11} + \Sigma_{12}\Sigma^{21} & \Sigma_{11}\Sigma^{12} + \Sigma_{12}\Sigma^{22} \\ \Sigma_{21}\Sigma^{11} + \Sigma_{22}\Sigma^{21} & \Sigma_{21}\Sigma^{12} + \Sigma_{22}\Sigma^{22} \end{pmatrix} \tag{A.5}$$

$$= \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{A.6}$$

This allows us to express the inverted quantities Σ^{ij} more naturally in terms of the elements of the covariance matrix Σ_{kl}. Solving for Σ^{11} using the first column of (A.5) and for Σ^{22} using the second column, we obtain

$$\Sigma^{11} = \left(\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T \right)^{-1}, \tag{A.7}$$

$$\Sigma^{22} = \left(\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} \right)^{-1}, \tag{A.8}$$

and similarly for $\Sigma^{12} = [\Sigma^{21}]^T$,

$$\Sigma^{12} = -\Sigma_{11}^{-1} \Sigma_{12} \Sigma^{22}. \tag{A.9}$$

For our purposes, a more convenient form for Eq. (A.7) can be obtained. From the first element of Eq. (A.5) we have

$$\Sigma^{11} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1} \Sigma_{12} \Sigma^{21}. \tag{A.10}$$

From the first entry in the second column of Eq. (A.5) we obtain

$$\Sigma^{12} = -\Sigma_{11}^{-1} \Sigma_{12} \Sigma^{22}, \quad \text{or equivalently} \quad \Sigma^{21} = -\Sigma^{22} \Sigma_{12}^T \Sigma_{11}^{-1}. \tag{A.11}$$

Combining these gives an alternative form for Σ^{11},

$$\Sigma^{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12} \Sigma^{22} \Sigma_{12}^T \Sigma_{11}^{-1}. \tag{A.12}$$

With these partitioned quantities in hand, we may now proceed to consider an alternate form of the Gaussian distribution.

The Gaussian distribution (A.1) may be rewritten

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} \exp\left[-\frac{1}{2} Q(\boldsymbol{x_1}, \boldsymbol{x_2}) \right] \tag{A.13}$$

where $Q(\boldsymbol{x_1}, \boldsymbol{x_2})$ is given by

$$Q(\boldsymbol{x_1}, \boldsymbol{x_2}) = ((\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T, (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T) \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}. \tag{A.14}$$

Expanding out the various partitioned elements, we arrive at

$$\begin{aligned} Q(\boldsymbol{x_1}, \boldsymbol{x_2}) = {} & (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \Sigma^{11} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \\ & + 2(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \Sigma^{12} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2) + (\boldsymbol{x}_2 - \boldsymbol{\mu}_2)^T \Sigma^{22} (\boldsymbol{x}_2 - \boldsymbol{\mu}_2). \end{aligned}$$

Substituting expressions (A.9) and (A.12) into this leads to

$$
\begin{aligned}
Q(x_1, x_2) = {} & (x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) \\
& + \left[(x_1 - \mu_1)^T \Sigma_{11}^{-1} \Sigma_{12} \right] \Sigma^{22} \left[\Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1) \right] \\
& - 2 \left[(x_1 - \mu_1)^T \Sigma_{11}^{-1} \Sigma_{12} \right] \Sigma^{22}(x_2 - \mu_2) \\
& + (x_2 - \mu_2)^T \Sigma^{22}(x_2 - \mu_2)
\end{aligned}
\tag{A.15}
$$

or, by further simplifying, as

$$
\begin{aligned}
Q(x_1, x_2) = {} & (x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1) \\
& + \left\{ (x_2 - \mu_2)^T - \left[\Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1) \right]^T \right\} \\
& \times \Sigma^{22} \left\{ (x_2 - \mu_2) - \left[\Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1) \right] \right\}.
\end{aligned}
$$

To make this expression a little more compact, the vector function $\tilde{\mu}_2(x_1)$ is introduced,

$$
\tilde{\mu}_2(x_1) = \mu_2 + \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1),
\tag{A.16}
$$

so that Eq. (A.15) can be rewritten in terms of the functions $Q_1(x_1)$ and $Q_2(x_1, x_2)$

$$
Q_1(x_1) = (x_1 - \mu_1)^T \Sigma_{11}^{-1}(x_1 - \mu_1),
\tag{A.17}
$$
$$
Q_2(x_1, x_2) = \left[x_2 - \tilde{\mu}_2(x_1) \right]^T \Sigma^{22} \left[x_2 - \tilde{\mu}_2(x_1) \right],
\tag{A.18}
$$

as

$$
Q(x_1, x_2) = Q_1(x_1) + Q_2(x_1, x_2).
\tag{A.19}
$$

We have thus far separated the full Gaussian distribution for x into the product of two functions involving $Q_1(x_1)$ and $Q_2(x_1, x_2)$ respectively. The first of these functions has the same functional form as a Gaussian distribution for only x_1 with covariance matrix Σ_{11}. We would like to separate this distribution out entirely. To do this we must address the $|\Sigma|$ term in denominator of Eq. (A.1).

We wish to factorise the component of $|\Sigma|$ arising from Σ_{11}. To do this we note that Σ may be decomposed as

$$
\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & 0 \\ \Sigma_{12}^T & I \end{pmatrix} \begin{pmatrix} I & \Sigma_{11}^{-1}\Sigma_{12} \\ 0 & \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}.
\tag{A.20}
$$

Since the determinant of the product of two matrices in the product of the determinant of each matrix, $|AB| = |A||B|$, we can write

$$|\Sigma| = |\Sigma_{11}||\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}|$$
$$= |\Sigma_{11}|\left[\Sigma^{22}\right]^{-1}|. \tag{A.21}$$

Substituting Eqs. (A.19) and (A.21) into (A.1),

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^r|\Sigma_{11}|}} \exp\left[-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)\right] \times$$
$$\frac{1}{\sqrt{(2\pi)^{m-r}\left|\left[\Sigma^{22}\right]^{-1}\right|}} \exp\left[-\frac{1}{2}(\mathbf{x}_2 - \tilde{\boldsymbol{\mu}}_2(x_1))^T \Sigma^{22}(\mathbf{x}_2 - \tilde{\boldsymbol{\mu}}_2(x_1))\right] \tag{A.22}$$

where we set $\mathbf{x}_1$ to be a vector of length r and $\mathbf{x}_2$ to be a vector of length $m - r$. This can of course be expressed as the product of two normal distributions;

$$f(\mathbf{x}_1, \mathbf{x}_2) = \mathcal{N}_{(r)}(\boldsymbol{\mu}_1, \Sigma_{11})\mathcal{N}_{(m-r)}\left(\tilde{\boldsymbol{\mu}}_2(x_1), \left[\Sigma^{22}\right]^{-1}\right), \tag{A.23}$$

where $\mathcal{N}_{(r)}$ is a distribution for the r $\mathbf{x}_1$ variables and $\mathcal{N}_{(m-r)}$ is a distribution for the $m - r$ $\mathbf{x}_2$ variables. From this the marginal distribution for $\mathbf{x}_1$ can be simply calculated by integrating over $\mathbf{x}_2$

$$f(\mathbf{x}_1) = \int f(\mathbf{x}_1, \mathbf{x}_2)d\mathbf{x}_2 = \mathcal{N}_{(r)}(\boldsymbol{\mu}_1, \Sigma_{11}). \tag{A.24}$$

The conditional distribution $f(\mathbf{x}_2|\mathbf{x}_1)$ can in turn be calculated from Bayes' theorem, Eq. (2.2);

$$f(\mathbf{x}_2|\mathbf{x}_1) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_1)}$$
$$= \mathcal{N}_{(m-r)}\left(\tilde{\boldsymbol{\mu}}_2(x_1), \left[\Sigma^{22}\right]^{-1}\right). \tag{A.25}$$

Following an analogous proof, the distribution $f(\mathbf{x}_1, \mathbf{x}_2)$ conditioned on $\mathbf{x}_2$ taking a particular value is given by

$$f(\mathbf{x}_1|\mathbf{x}_2) = \mathcal{N}_{(r)}\left(\tilde{\boldsymbol{\mu}}_1(x_2), \left[\Sigma^{11}\right]^{-1}\right). \tag{A.26}$$

where Σ^{11} is given by Eq. (A.7), and $\tilde{\mu}_1(x_2)$ is given by

$$\tilde{\mu}_2(x_1) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2). \tag{A.27}$$

In the case where the mean of the distribution is zero and the conditional variable is zero, $x_2 = 0$, the distribution further simplifies to

$$f(x_1|0) = \mathcal{N}_{(r)}\left(0, \left[\Sigma^{11}\right]^{-1}\right). \tag{A.28}$$

This is the type of conditioning applied to the zero-mean Gaussian noise term $\kappa(t) = (\kappa_z, \kappa_w)$ in Chap. 3.

Appendix B
Floquet Theory

The analogue of a linear stability analysis for systems with periodic components is known as Floquet theory [6]. It can also play an important role in the analysis of stochastic fluctuations about a deterministic trajectory [2, 7]. In this appendix the general formulation of Floquet theory is discussed before the more detailed application to linear stochastic systems is given.

Floquet theory gives the solutions to sets of linear differential equations in the form of Eq. (3.42), where $J(t)$ is periodic with a period T. The general solution can be shown to be

$$\boldsymbol{\xi}(t) = \sum_{i=1}^{m} c_i \boldsymbol{q}^{(i)}(t) e^{\sigma^{(i)} t}, \qquad (B.1)$$

where $\boldsymbol{q}^{(i)}(t)$ is a periodic vector and $\sigma^{(i)}$ are termed the Floquet exponents of the system. Meanwhile the quantities $\rho^{(i)} = e^{\sigma^{(i)} T}$ are called the Floquet multipliers of the system.

In particular one can work in a canonical form for calculational ease, with canonical quantities denoted with a further superscript 0. The canonical form is constructed from m decomposed solutions to Eq. (3.42) such that $\boldsymbol{\xi}^{(0,i)}(t) = \boldsymbol{q}^{(0,i)}(t) e^{\sigma^{(i)} t}$. A fundamental matrix of these solutions may then be introduced along with matrices $Y^{(0)}$ and $Q^{(0)}$. For the case $m = 3$ these may be expressed as

$$X^{(0)} = [\boldsymbol{\xi}^{(0,1)}(t), \boldsymbol{\xi}^{(0,2)}(t), \boldsymbol{\xi}^{(0,3)}(t)], \qquad (B.2)$$

$$X^{(0)} = Q^{(0)} Y^{(0)}, \qquad (B.3)$$

$$Q^{(0)} = [\boldsymbol{q}^{(0,1)}(t), \boldsymbol{q}^{(0,2)}(t), \boldsymbol{q}^{(0,3)}(t)], \qquad (B.4)$$

$$Y^{(0)} = Diag[e^{\mu^{(i)} t}]. \qquad (B.5)$$

A method for obtaining the Floquet multipliers $\mu^{(i)}$ along with the canonical form of the solutions is now required. Obtaining both is dependent on the determination of a matrix known as the monodromy matrix, which we shall now discuss.

The monodromy matrix, D, is defined such that $X(t+T) = X(t)D$, for any fundamental matrix $X(t)$ constructed from linearly independent solutions to Eq. (3.42). It can be shown that while the monodromy matrix is dependent on the fundamental matrix chosen, its eigenvalues are not [6]. The eigenvalues of D are $\rho^{(i)}$, the Floquet multipliers of the system. Further, if a matrix W is constructed from the eigenvectors of D, the canonical fundamental matrix $X^{(0)}(t)$ is related to a general fundamental matrix $X(t)$ via $X^{(0)}(t) = X(t)W$. Therefore, the monodromy matrix allows the canonical fundamental matrix $X^{(0)}(t)$ to be determined from a general fundamental matrix $X(t)$, along with the matrix $Y^{(0)}$. From these the periodic matrix $Q^{(0)}(t)$ can also be deduced.

In general, once the fundamental matrix is obtained it will have to be transformed into canonical form by a numerical determination of the monodromy matrix, $D = X^{-1}(t)X(t+T)$. For a system with initial conditions $t = 0$, $X(0) = I$, this simplifies to $D = X(T)$.

Now the stochastic system can be considered;

$$\frac{d\boldsymbol{\xi}}{dt} = J(t)\boldsymbol{\xi} + \boldsymbol{\eta}(t), \tag{B.6}$$

where $\boldsymbol{\eta}(t)$ is a vector of Gaussian white noise terms defined as in Eq. (2.67), except that now the noise covariance matrix depends explicitly on time through the varying parameter $\beta(t)$; $\langle \eta_i(t)\eta_j(t') \rangle = \varepsilon B_{ij}(t)\delta(t-t')$. The solution may be constructed as a sum of the general solution to Eq. (3.42) along with a particular solution, so that

$$\boldsymbol{\xi}(t) = X^{(0)}(t)\boldsymbol{\xi}^{(0)} + X^{(0)}(t) \int_{t_0}^{t} \left[X^{(0)}(s) \right]^{-1} \boldsymbol{\eta}(s)ds, \tag{B.7}$$

or, setting the initial conditions in the infinite past and making a change of integration variable $s \rightarrow s' = t - s$

$$\boldsymbol{\xi}(t) = Q^{(0)}(t) \int_{t_0}^{t} Y^{(0)}(s') \left[Q^{(0)}(t-s') \right]^{-1} \boldsymbol{\eta}(t-s')ds'. \tag{B.8}$$

In the course of the analysis conducted in Sect. 3.4, $\boldsymbol{\xi}(t)$ represents some stochastic fluctuation around limit cycle behaviour. An obvious quantity of relevance is the power spectrum of such fluctuations. To obtain the power spectrum, one first calculates the two-time correlation function $C(t+\tau, t) = \langle \boldsymbol{\xi}(t+\tau)\boldsymbol{\xi}^T(t) \rangle$; substituting Eq. (B.8) one obtains

$$C_{ij}(t+\tau, t) = Q^{(0)}(t+\tau)Y^{(0)}(\tau)\Lambda(t) \left[Q^{(0)}(t) \right]^T, \tag{B.9}$$

with

$$\Lambda(t) = \int_{t_0}^{\infty} Y^{(0)}(s)\Gamma(t-s)Y^{(0)}(s)ds \tag{B.10}$$

and

$$\Gamma(s) = \left[Q^{(0)}(s)\right]^{-1} B(s) \left[\left[Q^{(0)}(s)\right]^{-1}\right]^T. \tag{B.11}$$

The correlation function, $\mathcal{C}(\tau)$ is then simply related to the two-time correlation function by

$$\mathcal{C}(\tau) = \frac{1}{T} \int_0^T C(t+\tau, t)dt. \tag{B.12}$$

In turn, the Wiener-Khinchin theorem tells us that the power spectrum, $P(\omega)$, is simply the Fourier transform of the correlation function, and so

$$P_i(\omega) = \int \mathcal{C}_{ii}(\tau)e^{i\omega\tau}d\tau. \tag{B.13}$$

The intermediate steps are left to the reader, but full details are found in [3]. A key point to note is that Eqs. (B.7)–(B.11) hold *only* for the canonical matrices $X^{(0)}$, $Q^{(0)}$ and $Y^{(0)}$.

Appendix C
Derivation of the Fokker-Planck Equation for the Metapopulation Moran Model

In this appendix, the details of the master equation expansion (described in Sect. 2.4) are given for the metapopulation Moran model, introduced in Sect. 4.3. For generality, the expansion is described for the model with selection which is defined by the transition rates (4.11) with master Eq. (2.18). The neutral case can be recovered by setting $\boldsymbol{w}_A = \boldsymbol{w}_B = \boldsymbol{1}$.

As in the one-island case described in Sect. 2.9.1, expressions (4.11) are simplified by setting $[\boldsymbol{w}_B]_i = 1$ and $[\boldsymbol{w}_A]_i = 1 + s\alpha_i$ for each island. The parameter s is an indicative selection strength, while the elements of $\boldsymbol{\alpha}$ will be assumed to be of order 1 and will primarily be used to signify the direction of selection. If $\alpha_i > 0$ then $[\boldsymbol{w}_A]_i > [\boldsymbol{w}_B]_i$ and allele A is advantageous on island i, while if $\alpha_i < 0$, allele A will be deleterious on that island. Finally, if we assume that the selection strength s is small, we can express the above transition rates as a Taylor series in s. Suppressing the dependence of $T(\boldsymbol{n}|\boldsymbol{n}')$ on states that do not vary in a particular transition, we obtain

$$T(n_i + 1|n_i) = \sum_{j=1}^{\mathcal{D}} \frac{(\beta_i N - n_i)}{\beta_i N - \delta_{ij}} G_{ij} \times$$
$$\left(\frac{n_j}{\beta_j N} + s\alpha_j \frac{n_j(\beta_j N - n_j)}{(\beta_j N)^2} - s^2\alpha_j^2 \frac{n_j^2(\beta_j N - n_j)}{(\beta_j N)^3} + \mathcal{O}(s^3) \right),$$

$$T(n_i - 1|n_i) = \sum_{j=1}^{\mathcal{D}} \frac{n_i}{\beta_i N - \delta_{ij}} G_{ij} \times$$
$$\left(1 - \frac{n_j}{\beta_j N} - s\alpha_j \frac{n_j(\beta_j N - n_j)}{(\beta_j N)^2} + s^2\alpha_j^2 \frac{n_j^2(\beta_j N - n_j)}{(\beta_j N)^3} + \mathcal{O}(s^3) \right).$$

The dynamics can be seen to be that of a one-step process; any one transition can only move the system from an initial state $\boldsymbol{n}' = (n_1, \ldots n_i, \ldots n_{\mathcal{D}})$ to the adjacent states $\boldsymbol{n} = (n_1, \ldots n_i \pm 1, \ldots n_{\mathcal{D}})$. We can exploit this fact notationally; introducing

new state variables $\boldsymbol{x}$ such that $x_i = n_i/\beta_i N$, we can write $f_i^+(x_i)$ and $f_i^-(x_i)$ as shorthand for the transition rates (in terms of the new variables) for moving up to state $x_i + 1/\beta_i N$ or down in state $x_i - 1/\beta_i N$ from initial state $\boldsymbol{x}'$. This gives

$$
f_i^+(x_i) = \frac{G_{ii}(1-x_i)}{1-(\beta_i N)^{-1}} \left[x_i + s\alpha_i x_i(1-x_i) - s^2\alpha_i^2 x_i^2(1-x_i) \right]
$$

$$
+ (1-x_i) \sum_{j\neq i}^{\mathcal{D}} G_{ij} \left[x_j + s\alpha_j x_j(1-x_j) - s^2\alpha_j^2 x_j^2(1-x_j) \right] + \mathcal{O}(s^3),
$$

$$
f_i^-(x_i) = \frac{G_{ii} x_i}{1-(\beta_i N)^{-1}} \left[(1-x_i) - s\alpha_i x_i(1-x_i) + s^2\alpha_i^2 x_i^2(1-x_i) \right]
$$

$$
+ x_i \sum_{j\neq i}^{\mathcal{D}} G_{ij} \left[(1-x_j) - s\alpha_j x_j(1-x_j) + s^2\alpha_j^2 x_j^2(1-x_j) \right] + \mathcal{O}(s^3).
$$

$$
\text{(C.1)}
$$

For now let us leave the specific form of these transition rate functions alone, pausing only to note that the typical deme size, N, now only appears in the first term of $f_i^+(x_i)$ and $f_i^-(x_i)$.

We now re-express the master equation in terms of the transition rates $f_i^+(x_i)$ and $f_i^-(x_i)$:

$$
\frac{dp}{dt} = \sum_{i=1}^{\mathcal{D}} \left[f_i^+ \left(x_i - \frac{1}{\beta_i N} \right) p \left(x_i - \frac{1}{\beta_i N}, t \right) - f_i^+(x_i) p(x_i, t) \right]
$$

$$
+ \sum_{i=1}^{\mathcal{D}} \left[f^- \left(x_i + \frac{1}{\beta_i N} \right) p \left(x_i + \frac{1}{\beta_i N}, t \right) - f_i^-(x_i) p(x_i, t) \right]. \quad \text{(C.2)}
$$

This is equivalent to Eq. (2.29), albeit with a modified notation. Assuming the typical deme population N to be large, we can carry out a Taylor expansion in N^{-1} as described in Sect. 2.4. The right-hand side of the master Eq. (C.2) becomes

$$
- \sum_{i=1}^{\mathcal{D}} \left\{ \left(\frac{1}{\beta_i N} \right) \frac{\partial}{\partial x_i} \left[f_i^+(x_i) p(x_i, t) \right] \right\} + \frac{1}{2!} \sum_{i=1}^{\mathcal{D}} \left\{ \left(\frac{1}{\beta_i N} \right)^2 \frac{\partial^2}{\partial x_i^2} \left[f_i^+(x_i) p(x_i, t) \right] \right\}
$$

$$
+ \sum_{i=1}^{\mathcal{D}} \left\{ \left(\frac{1}{\beta_i N} \right) \frac{\partial}{\partial x_i} \left[f_i^-(x_i) p(x_i, t) \right] \right\} + \frac{1}{2!} \sum_{i=1}^{\mathcal{D}} \left\{ \left(\frac{1}{\beta_i N} \right)^2 \frac{\partial^2}{\partial x_i^2} \left[f_i^-(x_i) p(x_i, t) \right] \right\},
$$

plus terms in N^{-3} and higher.

We now return to the terms in f_i^+ and f_i^- which involve N. They are identical to lowest order in s, and equal

$$\frac{G_{ii}(1 - x_i)x_i}{1 - (\beta_i N)^{-1}}. \tag{C.3}$$

Now in the master equation f_i^+ and f_i^- appear with different signs in the terms involving the first derivative, and so they cancel. Although their contributions add in the terms involving the second derivative, if we expand the expression (C.3) in powers of N^{-1} we see that these give $\mathcal{O}(N^{-3})$ contributions in the expansion, which we are discarding. By the same argument, the terms in f_i^+ and f_i^- which involve N and powers of s will also give $\mathcal{O}(N^{-3})$ contributions when multiplying the second derivative, and so can also be discarded. Finally, when these s-dependent terms multiply the first derivative, they will give contributions s/N^2 and s^2/N^2, but we will not include such terms in the diffusion matrix B (see below), and so we do not include them in this context either. So, in summary, the N dependence which appears in f_i^+ and f_i^- in Eq. (C.1) may be omitted to the order we are working, and the only N dependence is that shown explicitly in the FPE.

We now define

$$A_i(\boldsymbol{x}) = \frac{1}{\beta_i} \left[f_i^+(\boldsymbol{x}) - f_i^-(\boldsymbol{x}) \right], \qquad B_{ii}(\boldsymbol{x}) = \frac{1}{\beta_i^2} \left[f_i^+(\boldsymbol{x}) + f_i^-(\boldsymbol{x}) \right]. \tag{C.4}$$

With these definitions the expansion of the master equation in inverse powers of N takes the form

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} = -\frac{1}{N} \sum_{i=1}^{\mathcal{D}} \frac{\partial}{\partial x_i} [A_i(\boldsymbol{x})p(\boldsymbol{x}, t)] + \frac{1}{2N^2} \sum_{i=1}^{\mathcal{D}} \frac{\partial^2}{\partial x_i^2} [B_{ii}(\boldsymbol{x})p(\boldsymbol{x}, t)]. \tag{C.5}$$

Substituting the explicit forms for $f_i^{\pm}$ given by Eq. (C.1) into Eq. (C.4) gives the elements of the vector $A(\boldsymbol{x})$ as

$$A_i(\boldsymbol{x}) = \frac{1}{\beta_i} \left\{ \sum_{j \neq i}^{\mathcal{D}} G_{ij}(x_j - x_i) + s \sum_{j=1}^{\mathcal{D}} G_{ij}\alpha_j x_j(1 - x_j) \right.$$
$$\left. - s^2 \sum_{j=1}^{\mathcal{D}} G_{ij}\alpha_j^2 x_j^2(1 - x_j) \right\} + \mathcal{O}(s^3),$$

and a diagonal diffusion matrix with elements given by

$$B_{ii}(\boldsymbol{x}) = \frac{1}{\beta_i^2} \left\{ x_i \sum_{j=1}^{\mathcal{D}} G_{ij} + \sum_{j=1}^{\mathcal{D}} G_{ij}x_j - 2x_i \sum_{j=1}^{\mathcal{D}} G_{ij}x_j \right\} + \mathcal{O}(s). \tag{C.6}$$

The truncation of the series in s, should be chosen to be consistent with the truncation in the expansion in terms of N. This will clearly depend on the assumed size of s. If one sets $s = 0$, the above model reduces to that stated for the neutral case, Eqs. (4.9) and (4.10).

Appendix D
Specification of Parameters Used in Figures

In order to aid the reproducibility of the results in this thesis, this appendix gives sets parameters omitted for brevity from Chaps. 5 and 6.

In Fig. 5.1, results from three different neutral metapopulation Moran systems are given. The results given in red/triangles are obtained from a system with the following parameters;

$$\boldsymbol{\beta} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.892 & 0.082 & 0.253 \\ 0.068 & 0.896 & 0.137 \\ 0.040 & 0.022 & 0.610 \end{pmatrix}. \tag{D.1}$$

The results in blue/circles are obtained from a system with both a symmetric migration matrix and a symmetric H matrix;

$$\boldsymbol{\beta} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.88 & 0.06 & 0.06 \\ 0.06 & 0.88 & 0.06 \\ 0.06 & 0.06 & 0.88 \end{pmatrix}. \tag{D.2}$$

The results in green/squares are obtained from an unusual system in which the probability of remaining on the first island is smaller than the probability that it migrates. The parameters for this system are

$$\boldsymbol{\beta} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.014 & 0.029 & 0.006 \\ 0.847 & 0.932 & 0.077 \\ 0.139 & 0.039 & 0.917 \end{pmatrix}. \tag{D.3}$$

In the right panel of Fig. 5.2, the parameters for the plots are taken from some of the randomly generated systems which yield the r_N values in the histogram in the left panel. In the main plots, the results given in blue/circles correspond to a system with a small r_N value. The parameters used for this system are

$$\beta = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.854 & 0.151 & 0.038 & 0.202 \\ 0.016 & 0.606 & 0.049 & 0.001 \\ 0.096 & 0.094 & 0.903 & 0.103 \\ 0.034 & 0.149 & 0.010 & 0.694 \end{pmatrix}. \tag{D.4}$$

The results in green/squares meanwhile are obtained from a system with a symmetric migration matrix, which yields $r_N = 1$. The parameters are

$$\beta = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.9 & 0.03\ldots & 0.03\ldots & 0.03\ldots \\ 0.03\ldots & 0.9 & 0.03\ldots & 0.03\ldots \\ 0.03\ldots & 0.03\ldots & 0.9 & 0.03\ldots \\ 0.03\ldots & 0.03\ldots & 0.03\ldots & 0.9 \end{pmatrix}. \tag{D.5}$$

The results in the right-inset plots are related the distribution on the left-inset. Once again those in blue/circles correspond to a system with a small r_N value, with the parameters given by

$$\beta = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.557 & 0.023 & 0.029 & 0.041 \\ 0.002 & 0.956 & 0.088 & 0.177 \\ 0.185 & 0.014 & 0.838 & 0.033 \\ 0.256 & 0.007 & 0.045 & 0.749 \end{pmatrix}. \tag{D.6}$$

The results in the inset plot in green/squares are again given by a symmetric migration matrix with islands all of the same size, and the parameters given by Eq. (D.5). The plots in red/triangles correspond to large r_N values in the left-inset histogram, with the following parameters used;

$$\beta = \begin{pmatrix} 1 \\ 1 \\ 5 \\ 1 \end{pmatrix}, \quad m = \begin{pmatrix} 0.803 & 0.079 & 0.128 & 0.033 \\ 0.027 & 0.802 & 0.159 & 0.012 \\ 0.083 & 0.042 & 0.556 & 0.025 \\ 0.087 & 0.077 & 0.157 & 0.930 \end{pmatrix}. \tag{D.7}$$

In Fig. 5.3 the migration matrix for each of the systems analysed is the same. This is

$$m = \begin{pmatrix} 0.513 & 0.056 & 0.208 \\ 0.231 & 0.833 & 0.208 \\ 0.256 & 0.111 & 0.584 \end{pmatrix}. \tag{D.8}$$

All the other parameters of the various systems are given in the caption to this figure.

In Fig. 5.4 the migration matrix for the system under consideration is

$$m = \begin{pmatrix} 0.714 & 0.050 & 0.143 & 0.077 \\ 0.036 & 0.750 & 0.190 & 0.077 \\ 0.071 & 0.050 & 0.619 & 0.077 \\ 0.179 & 0.150 & 0.048 & 0.769 \end{pmatrix}. \tag{D.9}$$

In Fig. 5.5 two sets of data related to two separate systems are compared. Data from the first system is given in blue/circles and relates to a two-deme system with the migration matrix

$$m = \begin{pmatrix} 0.856 & 0.167 \\ 0.154 & 0.833 \end{pmatrix}. \tag{D.10}$$

The second system, with results plotted in red/triangles, is one comprised of four demes with the migration matrix

$$m = \begin{pmatrix} 0.714 & 0.050 & 0.143 & 0.077 \\ 0.036 & 0.750 & 0.190 & 0.077 \\ 0.071 & 0.050 & 0.619 & 0.077 \\ 0.179 & 0.150 & 0.048 & 0.769 \end{pmatrix}. \tag{D.11}$$

All other parameters for these systems are listed in the caption of Fig. 5.5.

In Fig. 5.9 all the parameters for the main plots are given. The migration matrix relating to the inset plots was omitted in the main text however. It is

$$m = \begin{pmatrix} 0.800 & 0.050 & 0.025 & 0.100 & 0.025 \\ 0.050 & 0.800 & 0.025 & 0.025 & 0.100 \\ 0.025 & 0.025 & 0.800 & 0.000 & 0.150 \\ 0.100 & 0.025 & 0.000 & 0.800 & 0.075 \\ 0.025 & 0.100 & 0.150 & 0.075 & 0.650 \end{pmatrix}. \tag{D.12}$$

Finally, in Fig. 6.4, all the parameters relating to the plots have been omitted. The system yielding the stationary distribution in the left panel has parameters

$$\begin{matrix} \mathcal{D} = 3 \\ N = 150, \\ b = 0.5 \end{matrix} \quad \beta = \begin{pmatrix} 2 \\ 1 \\ 1.5 \end{pmatrix}, \quad \omega_1 = \begin{pmatrix} 5 \times 10^{-3} \\ 6 \times 10^{-4} \\ 9 \times 10^{-4} \end{pmatrix}, \quad \omega_2 = \begin{pmatrix} 1 \times 10^{-3} \\ 5 \times 10^{-3} \\ 4 \times 10^{-4} \end{pmatrix} \tag{D.13}$$

and

$$m = \begin{pmatrix} 0.900 & 0.050 & 0.030 \\ 0.100 & 0.900 & 0.070 \\ 0.000 & 0.050 & 0.900 \end{pmatrix}. \tag{D.14}$$

The system yielding the stationary distribution in the right panel meanwhile has parameters

$$
\mathcal{D} = 5 \\
N = 120, \quad \boldsymbol{\beta} = \begin{pmatrix} 2 \\ 1 \\ 1.5 \\ 1 \\ 3 \end{pmatrix}, \quad \boldsymbol{\omega}_1 = \begin{pmatrix} 1 \times 10^{-4} \\ 5 \times 10^{-4} \\ 4 \times 10^{-4} \\ 2 \times 10^{-3} \\ 9 \times 10^{-5} \end{pmatrix}, \quad \boldsymbol{\omega}_2 = \begin{pmatrix} 5 \times 10^{-3} \\ 6 \times 10^{-4} \\ 9 \times 10^{-4} \\ 7 \times 10^{-3} \\ 1 \times 10^{-3} \end{pmatrix}
$$
$$b = 0.5$$

$$\tag{D.15}$$

and the migration matrix

$$
m = \begin{pmatrix}
0.900 & 0.050 & 0.020 & 0.010 & 0.020 \\
0.020 & 0.850 & 0.030 & 0.020 & 0.020 \\
0.020 & 0.050 & 0.920 & 0.010 & 0.030 \\
0.020 & 0.090 & 0.010 & 0.960 & 0.020 \\
0.040 & 0.020 & 0.020 & 0.000 & 0.910
\end{pmatrix}.
\tag{D.16}
$$

Appendix E
Moran Model with Selection: Fixation Time

The mean time to fixation in the reduced metapopulation system, $T(z_0)$, is found from the backward FPE, in exactly the same way as described in Sect. 2.5.2. Therefore, analogous to Eq. (2.52), the equation reads

$$\frac{\bar{A}(z_0)}{N} \frac{dT}{dz_0} + \frac{\bar{B}(z_0)}{2N^2} \frac{d^2T}{dz_0^2} = -1, \tag{E.1}$$

where z_0 is the initial starting point on the centre manifold (or slow subspace). The boundary conditions are as for the single island case, that is, $T(0) = 0$ and $T(1) = 0$. In this appendix we discuss the analytic solution of Eq. (E.1) when $\bar{A}(z_0)$ and $\bar{B}(z_0)$ are given by Eqs. (4.35) and (4.25).

The result for the neutral case is well known [5]. Setting $s = 0$ in Eq. (4.35) gives $\bar{A}(z_0) = 0$, and direct integration of Eq. (E.1) gives Eq. (2.54), albeit divided by a factor of b_1 and with x_0 replaced by z_0. At order s, $\bar{A}(z) = sa_1z(1-z)$, and so the equation for $T(z_0)$ becomes

$$\frac{\sigma z_0(1-z_0)}{M} \frac{dT}{dz_0} + \frac{z_0(1-z_0)}{M^2} \frac{d^2T}{dz_0^2} = -1, \tag{E.2}$$

where we have defined new parameters $M = N/\sqrt{b_1}$ and $\sigma = a_1 s/\sqrt{b_1}$. The reason for introducing these new parameters, other than on grounds of simplicity, is that Eq. (E.2) is exactly the equation found in the single island case with selection.

To solve it we introduce $\phi(z_0) = dT/dz_0$, so that the equation now reads

$$\frac{d\phi}{dz_0} + M\sigma\phi = -\frac{M^2}{z_0(1-z_0)}. \tag{E.3}$$

This equation is difficult to deal with analytically and numerically because of the singularities on the right-hand side at precisely the values of z_0 where we need to impose the boundary conditions. One can avoid this problem by writing $\phi = \phi_0 + \phi_s$,

and choosing ϕ_0 so that the term $d\phi_0/dz_0$ cancels the right-hand side of Eq. (E.3). This choice means that ϕ_0 is simply the $s = 0$ solution, and the equation for ϕ_s is then

$$\frac{d\phi_s}{dz_0} + M\sigma\phi_s = -M\sigma\phi_0 = M^3\sigma\left[\ln z_0 - \ln(1 - z_0)\right], \tag{E.4}$$

which on the left-hand side is exactly the same as the equation for ϕ, but with a right-hand side which is less divergent as $z_0 \to 0$ or $z_0 \to 1$. Although this right-hand side is still divergent, its integral is not, which is all that we need. If we do require a convergent expression we can repeat the process, and write $\phi_s = \phi_1 + \phi_2$, choosing ϕ_1 so that the term $d\phi_1/dz_0$ cancels the right-hand side of Eq. (E.4).

We can now multiply Eq. (E.4) by $e^{M\sigma z_0}$ to find

$$\frac{d}{dz_0}\left[e^{M\sigma z_0}\phi_s\right] = M^3\sigma\left[\ln z_0 - \ln(1 - z_0)\right]e^{M\sigma z_0}, \tag{E.5}$$

which allows the integration to be straightforwardly carried out. One finds

$$T_s(z_0) = c_1 e^{-M\sigma z_0} + c_2$$
$$+ M^3\sigma \int_0^{z_0} dy\, e^{-M\sigma y} \int_0^y dx\, e^{M\sigma x}\left[\ln x - \ln(1 - x)\right], \tag{E.6}$$

where T_s is such that $dT_s/dz_0 = \phi_s$ and c_1 and c_2 are integration constants. Before imposing the boundary conditions, we can simplify the double integral by differentiating the inner integral and integrating by parts. This gives

$$T_s(z_0) = c_1 e^{-M\sigma z_0} + c_2$$
$$- M^2 e^{-M\sigma z_0} \int_0^{z_0} dx\, e^{M\sigma x}\left[\ln x - \ln(1 - x)\right]$$
$$+ M^2 \int_0^{z_0} dy\,\left[\ln y - \ln(1 - y)\right]. \tag{E.7}$$

The last term in Eq. (E.7) is simply the $s = 0$ mean time to fixation, and so applying the boundary conditions one obtains the Eqs. (5.12) and (5.13) given in the main text.

The calculation of $T(z_0)$ when $\bar{A}(z_0)$ is taken to order s^2 can be carried out in a similar way, but the results are more complicated and an integration by parts cannot straightforwardly simplify the double integral down to a single integral. The analogous equation to (E.3) is

$$\frac{d\phi}{dz_0} + M\sigma\left(1 - s\kappa z_0\right)\phi = -\frac{M^2}{z_0(1 - z_0)}, \tag{E.8}$$

where $\kappa = k_2/k_1$ and σ is now given by $\sigma = k_1 s/\sqrt{b_1}$. This is just as singular as Eq. (E.3), and so we perform the same manoeuvre and write $\phi = \phi_0 + \phi_s$, choosing ϕ_0 so that the term $d\phi_0/dz_0$ cancels the right-hand side of Eq. (E.8). The equation for ϕ_s then reads

$$\frac{d\phi_s}{dz_0} + M\sigma \left(1 - s\kappa z_0\right) \phi_s = M^3\sigma \left(1 - s\kappa z_0\right) \left[\ln z_0 - \ln(1 - z_0)\right]. \qquad \text{(E.9)}$$

The right-hand side is now less divergent, and one can proceed as before to multiply this equation by $e^{M\sigma(z_0 - s\kappa z_0^2/2)}$ and integrate twice. We find

$$T(z_0) = -M^2 \left[z_0 \ln(z_0) + (1 - z_0) \ln(1 - z_0)\right]$$
$$+ M^3\sigma \int_0^{z_0} dy\, e^{-M\sigma y(1-s\kappa y/2)} \left\{ \int_0^y dx(1 - s\kappa x) \times \right.$$
$$e^{M\sigma x(1-s\kappa x/2)} \int_0^y \left[\ln x - \ln(1 - x)\right] - c_3 \right\}, \qquad \text{(E.10)}$$

where the constant c_3 is given by

$$c_3 = \left(\int_0^1 dy\, e^{-M\sigma y(1-s\kappa y/2)} \right)^{-1} \int_0^1 dy\, e^{-M\sigma y(1-s\kappa y/2)}$$
$$\times \int_0^y dx\, (1 - s\kappa x)\, e^{M\sigma x(1-s\kappa x/2)} \left[\ln x - \ln(1 - x)\right].$$

Appendix F
Calculation of the Metapopulation Moran Model Dynamics on the Slow Subspace

In this appendix some of the more technical aspects of finding the slow subspace and calculating the dynamics of the reduced system will be set out. So far we have only specified the natural variable which we use in the reduced system, that is $z = \sum_{i=1}^{\mathcal{D}} u_i^{(1)} x_i$.

More generally, we can define a linear transformation to the coordinate z and $\mathcal{D} - 1$ coordinates $\boldsymbol{w}$ such that

$$\begin{pmatrix} z \\ \boldsymbol{w} \end{pmatrix} = T^{-1}\boldsymbol{x}, \quad \boldsymbol{x} = T\begin{pmatrix} z \\ \boldsymbol{w} \end{pmatrix}. \tag{F.1}$$

A convenient choice for x_i is

$$x_i = z + \sum_{a=1}^{\mathcal{D}-1} Q_{ia} w_a. \tag{F.2}$$

Since, from Eq. (4.22), $x_i = z$ on the centre manifold in the neutral case, we ask that the w_a are of order s on the slow subspace in the case with selection. This will simplify our calculation because, as we will see, this means that we will only have to calculate the w_a as functions of z to leading order in s.

In terms of the transformation matrix T, the choices made so far mean that

$$T^{-1} = \begin{pmatrix} [\boldsymbol{u}^{(1)}]^T \\ R \end{pmatrix}, \quad T = (\boldsymbol{1} \quad Q), \tag{F.3}$$

where R is a $\mathcal{D} - 1$ by $\mathcal{D}$ matrix and Q is a $\mathcal{D}$ by $\mathcal{D} - 1$ matrix. The form of the matrices R and Q is restricted through the conditions $TT^{-1} = T^{-1}T = I$, the

identity matrix. The condition relevant if we are trying to express $\boldsymbol{x}$ in terms of z and $\boldsymbol{w}$, is

$$\sum_{i=1}^{\mathcal{D}} u_i^{(1)} Q_{ia} = 0, \quad a = 1, \ldots, \mathcal{D} - 1. \tag{F.4}$$

We will need to check that any choice we make for Q_{ia} satisfies this condition.

We now substitute the transformation (F.2) into Eq. (4.34) for the drift vector in terms of $\boldsymbol{x}$:

$$A_i(z, \boldsymbol{w}) = \sum_{j=1}^{\mathcal{D}} H_{ij} \left(z + \sum_{a=1}^{\mathcal{D}-1} Q_{ja} w_a \right) + sz(1-z) \sum_{j=1}^{\mathcal{D}} \frac{G_{ij}\alpha_j}{\beta_i}$$
$$+ s(1-2z) \sum_{j=1}^{\mathcal{D}} \frac{G_{ij}\alpha_j}{\beta_i} \sum_{a=1}^{\mathcal{D}-1} Q_{ja} w_a - s^2 z^2 (1-z) \sum_{j=1}^{\mathcal{D}} \frac{G_{ij}\alpha_j^2}{\beta_i} + \mathcal{O}(s^2 y, s^3). \tag{F.5}$$

Using (i) $\sum_{j=1}^{\mathcal{D}} H_{ij} z = z \sum_{j=1}^{\mathcal{D}} H_{ij} = 0$, from Eq. (4.16), and (ii) the slow subspace condition $\sum_{i=1}^{\mathcal{D}} u_i^{(a+1)} A_i = 0$, $a = 1, \ldots, \mathcal{D} - 1$ (see Eq. (4.21)), we find

$$0 = \sum_{i,j=1}^{\mathcal{D}} \sum_{a=1}^{\mathcal{D}-1} u_i^{(a+1)} H_{ij} Q_{ja} w_a + sz(1-z) \sum_{i,j=1}^{\mathcal{D}} \frac{u_i^{(a+1)} G_{ij}\alpha_j}{\beta_i}, \tag{F.6}$$

since the slow subspace condition must be satisfied order by order in s and $\boldsymbol{y}$ is assumed to be of order s. Choosing Q_{ja} to be the right-eigenvectors $v_j^{(a+1)}$, $a = 1, \ldots, \mathcal{D} - 1$, which is consistent with the conditions (F.4), we see that the first term on the right-hand side of Eq. (F.6) is simply $\lambda^{(a+1)} w_a$. Therefore

$$w_a(z) = -\frac{sz(1-z)}{\lambda^{(a+1)}} \sum_{i,j=1}^{\mathcal{D}} \frac{u_i^{(a+1)} G_{ij}\alpha_j}{\beta_i} + \mathcal{O}(s^2). \tag{F.7}$$

Substituting Eq. (F.7) into Eq. (F.5), the drift vector evaluated on the slow subspace is found to be

$$A_i(z) = - sq_i^{(0)} z(1-z) + sq_i^{(1)} z(1-z) - s^2 q_i^{(2)} z^2 (1-z)$$
$$- s^2 q_i^{(3)} z(1-z)(1-2z) + \mathcal{O}(s^3), \tag{F.8}$$

where the vectors $\boldsymbol{q}^0$, $\boldsymbol{q}^1$, $\boldsymbol{q}^2$ and $\boldsymbol{q}^3$ are the parameter combinations

$$q_i^{(0)} = \sum_{a=1}^{\mathcal{D}-1} \sum_{j,k,l=1}^{\mathcal{D}} \frac{H_{ij} v_j^{(a+1)} u_k^{(a+1)} G_{kl} \alpha_l}{\beta_k \lambda^{(a+1)}},$$

$$q_i^{(1)} = \sum_{j=1}^{\mathcal{D}} \frac{G_{ij}\alpha_j}{\beta_i}, \quad q_i^{(2)} = \sum_{j=1}^{\mathcal{D}} \frac{G_{ij}\alpha_j^2}{\beta_i},$$

$$q_i^{(3)} = \sum_{a=1}^{\mathcal{D}-1} \sum_{j,k,l=1}^{\mathcal{D}} \frac{G_{ij}\alpha_j}{\beta_i} \frac{v_j^{(a+1)} u_k^{(a+1)}}{\lambda^{(a+1)}} \frac{G_{kl}\alpha_l}{\beta_k}. \tag{F.9}$$

The elements of the diffusion matrix meanwhile, when evaluated on the slow subspace, have the form

$$B_{ii}(z) = 2z(1-z) \sum_{j=1}^{\mathcal{D}} \frac{G_{ij}}{\beta_i^2} + \mathcal{O}(s). \tag{F.10}$$

Since the matrix H is not in general symmetric, then the eigenvalues will not in general be real. However since the entries of H are real, the eigenvalues will occur in complex conjugate pairs, and the eigenvectors associated with an eigenvalue λ^* will be the complex conjugates of those associated with λ. Since the expressions for $q_i^{(0)}$ and $q_i^{(3)}$ in Eq. (F.9) take the form of sums over a, for each term which is not real there will be another term added to it which is its complex conjugate. Thus $q_i^{(0)}$ and $q_i^{(3)}$ are guaranteed to be real. Therefore, the procedure goes through whether the eigenvalues are real or not. Of course, if there are complex conjugate pairs, the corresponding y_a cannot be interpreted as coordinates. However this interpretation is not crucial to the method, and if one wishes, it is always possible to define real coordinates by working with the real and imaginary parts of the eigenvalues and eigenvectors.

Appendix G
The Probability of Fixation in the Metapopulation Moran Model with Selection

The probability of fixation in the reduced system, $Q(z_0)$, is found as the relevant solution to Eq. (2.59). In the notation for the reduced metapopulation this reads

$$\frac{\bar{A}(z_0)}{N} \frac{dQ}{dz_0} + \frac{\bar{B}(z_0)}{2N^2} \frac{d^2Q}{dz_0^2} = 0, \tag{G.1}$$

where z_0 is the initial starting point on the centre manifold (or slow subspace). The boundary conditions are as for the single island case, that is, $Q(0) = 0$ and $Q(1) = 1$. In this appendix we discuss the analytic solution of Eq. (G.1) when $\bar{A}(z_0)$ and $\bar{B}(z_0)$ are given by Eqs. (4.35) and (4.25).

The result for the neutral case and to linear order in s have the same form as in the one-island case, and are well known [5]. When $s = 0$, $\bar{A}(z_0) = 0$, and so the solution of Eq. (G.1) subject to the boundary conditions is simply $Q(z_0) = z_0$. At linear order in s, $\bar{A}(z) = sa_1z(1-z)$, and a straightforward integration of Eq. (G.1) gives Eq. (2.100), albeit with extra factors of a_1 and b_1 and with x_0 replaced by z_0 (see Eqs. (5.17) and (5.11)).

To second order in s, $\bar{A}(z)$ may be written in the form (5.14), while $\bar{B}(z)$ is still given by Eq. (4.25). The equation for the probability of fixation (G.1) now takes the form

$$\frac{s}{N}z_0(1-z_0)(k_1 - sk_2z_0)\frac{dQ}{dz_0} + \frac{1}{N^2}b_1z_0(1-z_0)\frac{d^2Q}{dz_0^2} = 0.$$

Integrating with respect to z_0 we arrive at the equation

$$\frac{dQ}{dz_0} = c_1 \exp\left[-\frac{Ns}{b_1}\left(k_1z_0 - \frac{sk_2}{2}z_0^2\right)\right],$$

where c_1 is a constant of integration yet to be determined and where we note from Eq. (4.26) that $b_1 > 0$.

If $k_2 = 0$, the calculation is identical to that carried out to first order in s, Eq. (5.11), but with a_1 replaced by k_1. If $k_2 \neq 0$, we may complete the square in the exponent to find

$$\frac{dQ}{dz_0} = c_1 \exp\left[-\frac{Nk_1^2}{2b_1k_2}\right] \exp\left[\frac{N}{2b_1k_2}(sk_2z_0 - k_1)^2\right].$$

We now change variables from z_0 to l, where

$$l = \sqrt{\frac{N}{2b_1|k_2|}}(sk_2z_0 - k_1), \tag{G.2}$$

to obtain

$$\frac{dQ}{dl} = \begin{cases} -c_2 \exp\left(-l^2\right), & \text{if } k_2 < 0 \\ c_2 \exp\left(l^2\right), & \text{if } k_2 > 0, \end{cases} \tag{G.3}$$

where

$$c_2 = \frac{c_1}{s}\sqrt{\frac{2b_1}{|k_2|N}} \exp\left\{-\frac{Nk_1^2}{2b_1k_2}\right\}, \tag{G.4}$$

is another constant.

The integrals over the exponentials in Eq. (G.3) can be carried out in terms of functions related to the error function, namely the complementary error function [1]

$$\text{erfc}(y) = 1 - \text{erf}(y) = 1 - \frac{2}{\sqrt{\pi}}\int_0^y e^{-l^2}\, dl, \tag{G.5}$$

and the imaginary error function [4]

$$\text{erfi}(y) = \frac{2}{\sqrt{\pi}}\int_0^y e^{l^2}\, dl. \tag{G.6}$$

Implementing the boundary conditions $Q(l(z_0 = 0)) = 0$ and $Q(l(z_0 = 1)) = 1$, one finds

$$Q(z_0) = \frac{1 - \chi(z_0)}{1 - \chi(1)}, \tag{G.7}$$

where

$$\chi(z_0) = \frac{\text{erfc}(l(z_0))}{\text{erfc}(l(0))}, \quad \text{if } k_2 < 0, \tag{G.8}$$

and

$$\chi(z_0) = \frac{\text{erfi}(l(z_0))}{\text{erfi}(l(0))}, \quad \text{if } k_2 > 0. \tag{G.9}$$

If l is large, then asymptotic forms can be used to simplify both the complementary error function and the imaginary error function [1, 4]:

$$\text{erfc}(l) = \frac{e^{-l^2}}{\sqrt{\pi}l} \left[1 + \mathcal{O}\left(\frac{1}{l^2}\right) \right],$$ (G.10)

and

$$\text{erfi}(l) = \frac{e^{l^2}}{\sqrt{\pi}l} \left[1 + \mathcal{O}\left(\frac{1}{l^2}\right) \right].$$ (G.11)

References

1. M. Abramowitz, I.A. Stegun (eds.), *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1965)
2. A.J. Black, A.J. McKane, Stochastic amplification in an epidemic model with seasonal forcing. J. Theor. Biol. **267**, 85–94 (2010)
3. R.P. Boland, T. Galla, A.J. McKane, Limit cycles, Floquet multipliers, and intrinsic noise. Phys. Rev. E **79**, 051131 (2009)
4. A. Erdélyi (ed.), *Higher Transcendental Functions*, vol. II (McGraw-Hill, New York, 1953)
5. W.J. Ewens, *Mathematical Population Genetics*, 2nd edn. (Springer, Berlin, 2004)
6. R. Grimshaw, *Nonlinear Ordinary Differential Equations* (CRC Press, Oxford, 1990)
7. G. Rozhnova, A. Nunes, Stochastic effects in a seasonally forced epidemic model. Phys. Rev. E **82**, 041906 (2010)

About the Author

George William Albert Constable has a background in non-equilibrium statistical physics, and is especially interested in applying the ideas and techniques of stochastic theory to ecological and biological systems. His first degree was from the University of Glasgow and his Ph.D. was obtained from the University of Manchester. While there, his work was primarily concerned with the derivation of effective theories for microscopically defined individual based models. Currently, he is a postdoctoral researcher looking at the effects of spatial structure on population level dynamics in the Ecology and Evolutionary Biology Department at Princeton University.