Molecular Systems with Infinite and Finite Degrees of Freedom. Part I: Multi-Scale Analysis

Luca Sbano, Mathematics Institute, University of Warwick
sbano@maths.warwick.ac.uk
and
Markus Kirkilionis Mathematics Institute, University of Warwick
mak@maths.warwick.ac.uk

Abstract

The paper analyses stochastic systems describing reacting molecular systems with a combination of two types of state spaces, a finite-dimensional, and an infinite dimensional part. As a typical situation consider the interaction of larger macro-molecules, finite and small in numbers per cell (like protein complexes), with smaller, very abundant molecules, for example metabolites. We study the construction of the continuum approximation of the associated Master Equation (ME) by using the Trotter approximation [27]. The continuum limit shows regimes where the finite degrees of freedom evolve faster than the infinite ones. Then we develop a rigorous asymptotic adiabatic theory upon the condition that the jump process arising from the finite degrees of freedom of the Markov Chain (MC, typically describing conformational changes of the macro-molecules) occurs with large frequency. In a second part of this work, the theory is applied to derive typical enzyme kinetics in an alternative way and interpretation within this framework.

1 Introduction

Think of a typical situation in Cell Biology, the interaction of macro-molecules in the cell. In most cases there will be a small number of macro-molecular machines, like enzymes, ion-channels, polymerases, ribosomes etc. which are essential for cellular function, but which will not be very abundant in numbers per cell. Moreover this number will typically not change over time of observation. These machines will have different states of operation, like an ion channel can be closed or open. The states of operation of such machines can in general be described by finitely many different discrete states. These discrete states can be associated with meta-stable conformations of proteins (see for example [14]). Smaller molecules like ions, or metabolites like sugars, will interact with these macro-molecules. The most typical and best studied situation are enzymes catalysing metabolic reactions. The classical way to describe the resulting kinetics is given in [24], see also [25]. The number of these smaller molecules clustered in different species will change typically over time of observation. Assuming no inherent spatial
structure of the process, this gives rise to coupled systems of two well studied mathematical objects, Markov chains describing the transitions between the different modes of operation of the macro-molecules, and birth-death processes with reactions describing the change in numbers of the smaller abundant molecules. In this paper we will study both mathematical objects simultaneously as one system, giving a rigorous derivation of the continuum limit. With other interpretations the theory can also applied in various other fields of sciences where interaction of different finite state ‘machines’ will occur, like epidemiology, manufacturing or economy.

Figure 1: Different typical interactions of small and large molecules in a typical cell. (A): Interaction of enzymes with metabolites. The product is a molecule consisting of two elementary species. Enzymes react as catalysts. (B): Genetic interactions. For example a repressor can bind to the DNA only in the case it is in a conformation characterised by the absence of smaller ‘inducing’ molecules. (C): Membrane proteins act in various ways as molecular machines, for example as ion channels.

As a concrete example from Genetics assume mRNA is transcribed depending on whether a specific DNA binding site is bound to a transcription factor $A$. There are two possible limit regimes: either no $A$ molecules are binding causing no mRNA transcription, or molecules $A$ are binding to the DNA implying a maximum transcription rate for the mRNA. Usually such binding/unbinding events occur at large frequencies, and they are proportional to the concentration of the transcription factor. This leads to an effective transcription rate resulting from an ”effective average” of the binding/unbinding event depending on the concentration of $A$. This effect is usually modelled by a Hill-type kinetics (23):

$$K([A]) = \frac{K_1 [A]^n}{1 + K_2 [A]^n}$$

This kinetics describes an effective reaction rate with saturation behaviour for large concentrations $[A]$ of the transcription factor $A$. Here $n$ is a positive integer, an exponent controlling the slope of the sigmoidal $K([A])$. In this paper we consider as an illustrative example the case $n = 1$, in [23] we shall describe the general case. We will follow this particular example throughout this paper (part I) starting from microscopic assumptions, and derive the above deterministic limit in part II.
**Birth-death processes and the density assumption**

We will assume that at time $t$ the state of the subsystem describing small molecules of variable number is fully determined by the collection of these numbers belonging to different species. The time evolution is given through the transition from one possible collection of these molecules to another. The transition itself is prescribed by the reactions, which contain the rates at which the various species form complexes, i.e. other species. Reactions define the dynamics by providing the probability rates at which the elementary transitions occur, and this leads to constructing a stochastic process with independent increments (Markov property). For this part of the system the state at time $t$ is described by a probability distribution, which solves the so-called Master Equation (ME) (see for example [28] and [15]). The ME is a differential equation with respect to time and a difference equation with respect to the various particle numbers. In many situations the number of molecules involved is of the order of the Avogadro number $\sim 10^{23}$, which implies that the concept of density can be introduced. That can be achieved by constructing the so-called continuum approximation of the ME, an operation that transforms the ME into a Fokker-Planck Equation (FPE). This correspondence is a very delicate point and often neglected in applications. Usually the continuum approximation is investigated through the Kramers-Moyal expansion or the Van Kampen size-expansion [15], [28]. Here we employ the Trotter approximation [27, 11, 19] to study multi-scale reactions systems. In fact the stochastic process generated by the FPE is a valid approximation to the one generated by the ME only for a fixed time interval (for further details see [8], [9], [10], and [16] for a recent review). A representation of the ME can be directly studied numerically through Gillespie’s algorithm (see [4]) which provides the construction of the underlying stochastic process. Our focus as outlined in the reminder of this introduction aims at understanding how different scales are affecting the macroscopic dynamics. These are the ingredients of models usually of interest in Mathematical Biology or Systems Biology.

**Molecular subsystems with finite degrees of freedom, mixed systems**

To summarise the system to study will have two kind of state variables (degrees of freedom), first variables which admit a continuum approximation as just described, and secondly variables whose discrete nature is essential. Earlier work to study such systems include [6] and [13]. In this class the state is described by a set of numbers of molecules, and a set of finite discrete states describing all possible molecular conformations, binding/unbinding events, etc. The dynamics is described by a master equation and its solution, a probability distribution generating a Markov process. After taking the continuum approximation it has been formally shown in an appendix of [6] that the whole process is a combination of a continuous process and a discrete Markov chain with finite states. Such systems are sometimes also called random evolutions [5]. They are ubiquitous in the realm of complex systems. It may happen that infinite and finite degrees of freedom evolve on different time scales. Indeed in many applications like [6] it appears that the dynamics associated to the finite degrees of freedom evolves very fast,
and therefore it is important to understand how this affects the dynamics associated to the infinite degrees of freedom. Heuristically this is usually done by introducing \textit{ad hoc} reaction rates (see \cite{3}), which mimic some sort of ”averaged effect”.

Some remarks on this previous work. In \cite{6} such systems are studied by using the ME method and by taking a heuristic continuum approximation. Subsequently formal asymptotic methods are applied to study the large frequency problem. In the context of simulation the analysis of systems with fast and slow dynamics has been addressed in \cite{20}, where the updates of the infinite state variables was optimised. Collective effects of many particles in biological systems have also been investigated in \cite{13} where a many-particle method is used. This approach has its origins in multi-body and field theory, (see \cite{2}). Inspired by the approach presented in an appendix of \cite{6} we develop in this paper a rigorous analysis of the adiabatic theory for systems with infinite and finite degrees of freedoms. The paper presents a general formulation for the ME for such systems whose state space is formed by \(N\) types of particles and \(g\) finite states. Let \(\Sigma\) be the set of all possible discrete states. At time \(t\) the system is in a state \((\mathbf{n},\sigma) \in \mathbb{N}^N \times \Sigma\) with probability \(P_{\sigma}(\mathbf{n},t)\). The \(N\)-tuple \(\mathbf{n} = (n_1, ..., n_N) \in \mathbb{N}^N\) represents the collection of free molecules of different species, and \(\sigma\) is a discrete state in \(\Sigma\) describing the conformational changes of macro-molecules which number does not change during system observation. The ME for the probability \(P_{\sigma}(\mathbf{n},t)\) is then given by

\[
\frac{\partial P_{\sigma}(\mathbf{n},t)}{\partial t} = \sum_{\sigma' \in \Sigma} \mathcal{L}_{\sigma\sigma'}(\mathbf{n})P_{\sigma'}(\mathbf{n},t) + \sum_{\sigma' \in \Sigma} K^T_{\sigma\sigma'}(\mathbf{n})P_{\sigma'}(\mathbf{n},t) \quad \text{with} \quad \sigma \in \Sigma, \quad (1)
\]

where \(\mathcal{L}_{\sigma\sigma'}(\mathbf{n})\) is a collection of difference operators (the \(^*\) indicating the adjoint of an operator, a notation which will become clear later in the paper) and \(K^T(\mathbf{n})\) is the transpose of a generator of a Markov chain on \(S\). We will study the ME with two methods, the \textit{continuum limit} and the \textit{adiabatic approximation}.

The continuum limit

The continuum limit will be formulated by using the so-called \textit{Trotter approximation} (see \cite{27}, \cite{19}). Trotter’s method has been also used in \cite{11} \cite{9} \cite{10}. The equation (1) is derived on the base of the elementary molecular processes that depend on the scale at which they are studied. It is therefore important to reformulate the ME taking into account its dependence on the size and time scales. Let us define two sets of scales:

1. The size scales \(\vec{\delta} = (\delta_1, ..., \delta_N), \delta_i > 0,\)

2. the time scale \(\tau > 0,\)

Let us define the following lattice

\[
\mathbb{L}_{\vec{\delta}} = \{(n_1\delta_1, ..., n_N\delta_N) : (n_1, ..., n_N) \in \mathbb{N}^N\} \quad (2)
\]
The state of the system is now specified on \( \mathbb{L}_\delta \times \Sigma \). The processes on \( \mathbb{L}_\delta \) will be *birth-death* interaction of the form

\[
(n_1 \delta_1, ..., n_i \delta_i, ..., n_N \delta_N) \to (n_1 \delta_1, ..., n_i \delta_i \pm \delta_i, ..., n_N \delta_N),
\]

the processes on \( \Sigma \) are transitions \( \sigma \to \sigma', \)

driven by a finite Markov chain. Note that for fixed \( \delta \) a function \( f \) defined on \( \mathbb{L}_\delta \) is fully determined by its values on \( n \), i.e. when \( \delta \) is fixed we can consider \( f \) defined over \( \mathbb{N}^N \), writing \( f(n) \).

Equation (1) is now rewritten in a form where \( L^* \) and \( K^* \) are operators depending on \( \delta \) and \( \tau \):

\[
\frac{\partial P_\sigma(n,t)}{\partial t} = \sum_{\sigma' \in \Sigma} L_{\sigma\sigma'}(\delta, \tau; n) (P_{\sigma'}(n,t)) + \sum_{\sigma' \in \Sigma} K_{\sigma\sigma'}^T(\delta, \tau; n) P_{\sigma'}(n,t).
\]

The continuum limit is the study of the form of \( L^* \) and \( K^* \) as \( \delta \to 0 \) and \( \tau \to 0 \).

Equation (3) is defined on the space of probability measures on \( \mathbb{L}_\delta \times \Sigma \):

\[
X^*_\delta = \left\{ P \mid \sum_{n \in \mathbb{L}_\delta, \sigma \in \Sigma} P_\sigma(n) = 1 \right\}.
\]

The pairing between \( X^*_\delta \) and \( X_\delta \) is

\[
\langle P, u \rangle = \sum_{n \in \mathbb{L}_\delta, \sigma \in \Sigma} P_\sigma(n) u_\sigma(n).
\]

Using the duality, the ME defined on \( X_\delta \) becomes the Kolmogorov equation

\[
\frac{\partial u_\sigma(n,t)}{\partial t} = \sum_{\sigma' \in \Sigma} L_{\sigma\sigma'}(\delta, \tau; n) (u_{\sigma'}(n,t)) + \sum_{\sigma' \in \Sigma} K_{\sigma\sigma'}^T(\delta, \tau; n) u_{\sigma'}(n,t),
\]

where

\[
\langle L^* P, u \rangle = \langle P, Lu \rangle, \quad \langle K^T P, u \rangle = \langle P, Ku \rangle.
\]

For any fixed \( \delta \) and \( \tau \) equation (4) is an infinitesimal generator of a Markov process on \( \mathbb{L}_\delta \times \Sigma \). Now let \( \delta_n \) and \( \tau_n \) two sequences of scales such that

\[
\delta_n \to 0, \quad \tau_n \to 0.
\]
as \( n \to \infty \). Then we can define a sequence of spaces \( X_n = X_{\vec{\delta}_n}, X^*_n = X^*_{\vec{\delta}_n} \), and a sequence of operators \( \mathcal{L}_n, \mathcal{K}_n \) acting on \( X_n \). In [27], a method is introduced to study the limit of \( \mathcal{L}_n, \mathcal{K}_n \) as \( n \to \infty \). The idea is to look at \( \mathcal{L}_n, \mathcal{K}_n \) defined on \( X_n \) as an approximation of infinitesimal generators \( \hat{\mathcal{L}}, \hat{\mathcal{K}} \) defined on a suitable Banach space that, in most applications, is given by \( X = C^0(\mathbb{R}^N, \mathbb{R}^q) \). Following [27] we construct a sequence of projections

\[
P_n : X \to X_n,
\]

and we state that a sequence \( u_n \in X_n \) approximates \( u \in X \) if

\[
\|P_n(u) - u_n\|_n \to 0.
\]

This will be denoted by \( u_n \approx u \). In order to take into account the presence of the scales \( \vec{\delta}_n \) and \( \tau_n \), we modify the criteria to approximate the limit of operators (in an adapted way different from [27, 19]). In fact a sequence of linear operators \( A_n : X_n \to X_n \) is now a function of \( \vec{\delta}_n \) and \( \tau_n \), and in general we cannot expect some limit to exist for any choice of \( \vec{\delta}_n \to 0, \tau_n \to 0 \). For this reason we say that the sequence of operator converges, \( A_n \approx \hat{A} \), if there exists \( \vec{\delta}_n \to 0, \tau_n \to 0, \) and \( \hat{A} \) defined on \( X \) such that

\[
\|A_n(P_n(u)) - P_n(\hat{A}(u))\|_n \to 0 \text{ for all } u \text{ in the domain of } \hat{A},
\]

as \( n \to \infty \). It is important to note that the limit is now depending on the choice of the sequence of scales converging to 0. We shall show that in general operators may have pre-factors which characterise their limit behaviour. The typical case will be \( \mathcal{L}_n \approx \hat{\mathcal{L}} \) and \( \mathcal{K}_n \approx \frac{1}{\epsilon} \hat{\mathcal{K}} \), which yields the limit

\[
\hat{A} = \hat{\mathcal{L}} + \frac{1}{\epsilon} \hat{\mathcal{K}}.
\]

Here the constant \( \epsilon \) will in general be a function of \( \vec{\delta}_n \to 0, \tau_n \to 0 \), so depending on the scales. This approach determines the operator [5] that, for fixed \( \epsilon \), is the infinitesimal generator of a process on \( \mathbb{R}^N \times \Sigma \). In many applications it turns out that \( \epsilon \) is an infinitesimal function in \( \vec{\delta}_n \to 0, \tau_n \to 0 \) and this leads very naturally to the study of the adiabatic approximation for the equation

\[
\frac{\partial u(x,t)}{\partial t} = \hat{A}(u(x,t)).
\]

(6)

In the applications is often considered the adjoint equation, the Fokker Planck equation (FPE) which corresponds directly to the ME. The FPE is defined on the dual of \( X^* \) and reads

\[
\frac{\partial \rho(x,t)}{\partial t} = \hat{\mathcal{L}}^*(\rho(x,t)) + \frac{1}{\epsilon} \hat{\mathcal{K}}^T(\rho(x,t)) \rho(x,t).
\]

(7)

and

\[
X^* = \left\{ \rho : \sum_{\sigma \in \Sigma} \int dx \rho(x,\sigma) u(x,\sigma) < \infty \right\}.
\]
Adiabatic approximation

The adiabatic approximation theory is based on the observation that for sufficiently small $\epsilon$ the dynamics of the finite state Markov chain is faster than the one of the birth-death process. This should be a reasonable assumption for most or all macro-molecular behaviour in a cell. Such an assumption implies that on sufficiently large time scales the Markov chain can be considered at equilibrium. Generalising [6] we assume that the Markov chain has possibly more than one stationary measure:

$$M_K = \{ \mu(x) \in X^\times : K^T(\mu(x) = 0) \}.$$

To avoid trivialities we assume that

$$m_K \doteq \dim(M_K) < g.$$

Any convex combination

$$\mu = \sum_{m=1}^{m_K} \theta_m \mu^{(m)}$$

with

$$\sum_{m=1}^{m_K} \theta_m = 1$$

is in $M_K$ (see [1]). Each such measure describes the possible asymptotic behaviour of the Markov chain which is in general decomposable, i.e. a product of $m_K$ Markov chains. We now take one convex combination $\mu \in M_K$ and construct the adiabatic theory for the FPE obtaining an asymptotic expansion in $\epsilon$ of $\rho$. This expansion has a leading order term, which will be called average dynamics. This dynamics is given by

$$\frac{\partial f(x,t)}{\partial t} = \sum_{m=1}^{m_K} \sum_{\sigma \in \Sigma} \theta_m \hat{L}^\sigma(x)(\mu^{(m)}(x)f(x,t)),$$  \hspace{1cm} (8)

where

$$f(x,t) = \sum_{\sigma \in \Sigma_{\mu}} \rho_\sigma(x,t)$$

is the marginal distribution associated to $\mu$. It is noteworthy that the appearance of an averaged dynamics occurs in the modified Gillespie’s algorithm as presented in [20].

Description in terms of ODEs and SDEs

In the study of [7] for small $\epsilon$ one could make the non-trivial observation that equation (7) up to order $O(\epsilon)$ generates a Markov process described by a stochastic differential equation. This observation which entails to show that (7) up to $O(\epsilon)$ reduces to a parabolic operator will be clarified in another paper. It is important to mention here that such an approximation is valid only on finite time interval as it was shown in [8], [9] and [10]. Under this restricted condition the dynamics can be described by the following Ito stochastic differential equation

$$\frac{\partial f(x,t)}{\partial t} = \sum_{m=1}^{m_K} \sum_{\sigma \in \Sigma} \theta_m \hat{L}^\sigma(x)(\mu^{(m)}(x)f(x,t)),$$  \hspace{1cm} (8)

where

$$f(x,t) = \sum_{\sigma \in \Sigma_{\mu}} \rho_\sigma(x,t)$$

is the marginal distribution associated to $\mu$. It is noteworthy that the appearance of an averaged dynamics occurs in the modified Gillespie’s algorithm as presented in [20].
\[ dx_{\alpha}(t) = A_{\alpha}(x(t)) \, dt + \sqrt{\varepsilon} \sum_{\beta=1}^{N} \sigma_{\alpha\beta}(x(t)) \, dw_{\beta}^{\alpha} \text{ with } \alpha = 1, \ldots, N, \]  
(10)
where \( \{w_{\beta}^{\alpha}\}_{\beta=1}^{N} \) are \( N \)-independent Wiener processes. Here \( \|\sigma(\varepsilon, x)\| \sim \sqrt{\varepsilon} \) and \( A(x) \) is the averaged vector field given by

\[ A_{\alpha}(x) = \sum_{j \in S} \theta_{m_{j}} L_{\alpha}^{j}(x) \mu_{j}^{(m_{j})}(x). \]  
(11)

Here \( L_{\alpha}^{j}(x) \) is the deterministic vector field associated to the finite state \( j \). Moreover \( A(x) \) is the average over the stationary measure \( \mu(x) \) of all vector fields associated to the finite states in \( S \). If \( m_{K} > 1 \) then the Markov chain is equivalent to a product of \( m_{K} \) Markov chains and the vector-field (11) describes the deterministic dynamics averaged over all \( m_{K} \) components of \( S \). We illustrate the theory using equation (11) and derive as applications effective reaction rates related to different macro-molecular machinery. One prominent example is the well known Hill’s kinetics. In a forthcoming paper we apply this theory to derive rigorously the nonlinear macroscopic model used in [12] to study - on a more heuristic basis - the bistability in the Lac-Operon.

The organisation of the paper is as follows. We first define systems with both infinite and finite degrees of freedom. Then we introduce formally the continuum approximation for our setting. Finally the adiabatic approximation is constructed. The second part of this series contains the examples, noteworthy a new approach to enzyme kinetics ([23]). In the appendix we describe the geometrical structure of the Markov chain which is very important to develop the adiabatic theory.

2 Systems with infinite and finite degrees of freedom

As motivated in the introduction typical macro-molecular systems give rise to mixed microscopic dynamics, consisting of birth-death processes where particle or molecule numbers can be arbitrarily large, and a second part where some entities have a fixed number of molecules in the system, but each equipped with finitely many different functional states giving rise to a finite state Markov chain.

The following definition will fix this structure for further investigation, followed by an illustrative and biologically important example.

Definition 2.1. Let us define two sets of scales

1. size scales \( \vec{\delta} = (\delta_{1}, \ldots, \delta_{N}) \), \( \delta_{i} > 0 \),
2. time scale \( \tau > 0 \).

Let \( L_{\vec{\delta}} \) be the following lattice

\[ L_{\vec{\delta}} = \{n \vec{\delta} = (n_{1} \delta_{1}, \ldots, n_{N} \delta_{N}) : \, n = (n_{1}, \ldots, n_{N}) \in \mathbb{N}^{N} \} \]  
(12)
Remark 2.1. On $\mathbb{L}_{\vec{\delta}}$ we shall define functions, now for fixed $\vec{\delta}$ the value of any function $u$ is uniquely determined by the integer vector $n$ therefore whenever $\vec{\delta}$ is fixed we shall omit the $\vec{\delta}$ dependence and write $u(n)$.

Definition 2.2. Let the tuple $(\zeta, R, P)$ determine a stochastic process by specifying the state $\zeta$, a set of reactions $R$, and a vector of probabilities $P$, such that

(i) the state $\zeta$ of the system is fully specified by $n_1, ..., n_N$ infinite degrees of freedom (i.d.f.) and a second variable, the finite degrees of freedom $s$ (f.d.f.). The state $\zeta$ is therefore the composition

$$\zeta = (n_1 \delta_1, ..., n_N \delta_N, \sigma) = (n \delta, \sigma) \in \mathbb{L}_{\vec{\delta}} \times \Sigma,$$

where $\mathbb{L}_{\vec{\delta}} = \vec{\delta}N^N$, $n$ is an $n$-tuple of natural numbers and $\sigma$ runs in a finite set $\Sigma$, with $|\Sigma| = g$ being the number of discrete states.

(ii) the time evolution of the stochastic process is defined via the set of reactions $R$ having three different types:

(a) Processes involving only i.d.f.’s represented by reactions (possibly reversible) of the form

$$(n, \sigma) \rightarrow (n', \sigma).$$

The operator describing these reactions in the master equation (14) is denoted by $L^*_R$ and has the form $L^*_R = \ell_0 \otimes \delta_{\sigma \sigma'}$ where $\ell_0$ is the same operator for each discrete state $\sigma = 1, ..., g$. Here $\delta_{\sigma \sigma'} = 1$ for $\sigma = \sigma'$, and zero otherwise.

(b) Processes involving only f.d.f.’s represented by reactions (possibly reversible) of the form

$$(n, \sigma) \rightarrow (n', \sigma').$$

The operator describing these reactions in the master equation (14) is the transpose $K^T$ of the Markov chain generator of the process governing the transitions among the discrete states $\sigma = 1, ..., g$. The Markov chain is finite dimensional with a space of stationary states $M_K$ of dimension strictly less than $g$.

(c) Processes involving both i.d.f. and f.d.f. represented by reactions (possibly reversible) of the form

$$(n, \sigma) \rightarrow (n', \sigma).$$

The operator describing these reactions in the master equation (14) is denoted by $L^*_E$. This operator is non-trivial only in the discrete states $\sigma$ which affect processes involving i.d.f.
(iii) each realisation of the process is valued in $\mathbb{L}_0^N \times \Sigma$. The state $\zeta$ at time $t$ is given by the vector of probabilities

$$P(t, n) = (P_1(t, n), ..., P_g(t, n)), \quad \text{with} \quad \sum_{n \in \mathbb{N}^N} \sum_{\sigma=1}^g P_\sigma(t, n) = 1. \quad (13)$$

The time evolution of $P$ is given by the master equation (ME)

$$\frac{\partial P(t, n)}{\partial t} = (\mathcal{L}_R + \mathcal{L}_E^*) \circ P(t, n) + \mathcal{K}^T(n) P(t, n), \quad (14)$$

$P, \mathcal{L}_R, \mathcal{L}_E^*$ and $\mathcal{K}^T$ are sufficiently regular such that (14) has a unique solution for all times $t > 0$. Then the tuple $(\zeta, R, P)$ is called a (microscopic) system with infinite and finite degrees of freedom, or short an IFSS (Infinite-Finite State System).

**Illustrative example of a typical IFSS: Single enzyme kinetics**

Consider a system with 2 i.d.f. and 1 f.d.f.. The system has state space $(a, m, O_i) \in (\delta \mathbb{N})^2 \times S$, and will be described by the vector probability

$$P(t, m, a) = (P_0(t, m, a), P_1(t, m, a)).$$

Here $a$ and $m$ are the numbers of two small species of molecules (i.d.f.) called $A$ and $M$, and $S = \{O_0, O_1\}$ are the discrete states of a molecular machine (f.d.f.), for example an enzyme. Reactions of type $(a)$ are those independent from the discrete states of the macro-molecular machinery. In our example we assume that degradation of $M$ is of this type:

$$M \rightarrow^{\nu/\tau} \emptyset$$

This reaction takes place at speed $\nu > 0$. This gives the following contributions to the ME

$$+\frac{\nu}{\tau} (m + 1) P_\sigma(t, m + 1, a) - \frac{\nu}{\tau} m P_\sigma(t, m, a) \quad \text{for } \sigma = O_0, O_1.$$ 

Such terms can be rewritten as

$$\frac{\nu}{\tau} (\mathbf{E}^+ - \mathbf{id})(m P_\sigma(t, m, a))$$

where $\mathbf{E}^+$ and $\mathbf{id}$ are difference operator defined as

$$\mathbf{E}^+ f(m) = f(m + 1), \quad \mathbf{id} f(m) = f(m) \quad \text{for every } f : \delta \mathbb{N} \rightarrow \mathbb{R}.$$ 

Using this difference operator notation $\mathcal{L}^*$ is then given by

$$\mathcal{L}_R^* = \frac{1}{\tau} \begin{pmatrix} \nu (\mathbf{E}^+ - \mathbf{id})(m \cdot) & 0 \\ 0 & \nu (\mathbf{E}^+ - \mathbf{id})(m \cdot) \end{pmatrix},$$
Reactions of type (b) are given by

\[ A + O_0 \rightarrow^{\frac{k^+}{\tau}} O_1, \]
\[ O_1 \rightarrow^{\frac{k^-}{\tau}} O_0 + A, \]

they describe the transitions of the discrete states of the macro-molecule, possibly depending on binding of smaller molecules, in this case of molecules of type A. Moreover the transition rates depend on the relative Markov chain 'switching' time scale \( \tau \) and the system size \( \delta \), which is defined as inverse of the largest average number of A molecules in the system. The generator of the Markov chain is then given by

\[ \mathcal{K} = \frac{1}{\tau} \begin{pmatrix} -a & a & 0 \\ k^- & -k^- & 0 \end{pmatrix}. \]

This also implies the time evolution of the system must be described through the vector probability

\[ P(t, m, a) = (P_0(t, m, a), P_1(t, m, a - 1)) \simeq (P_0(t, m, a), P_1(t, m, a)) \]

for \( a \) much larger than \( \delta \).

Finally the only reaction of type (c) is given by

\[ \emptyset \rightarrow^{v/\tau} M \text{ for } \sigma = O_1, \]

with \( O_1 \) interpreted as the active state, the only one at which the enzyme in addition catalyses molecules of type \( M \). This reaction gives the following contribution to the ME

\[ 0 \text{ for } s = O_0 \]
\[ \frac{v}{\tau} P_1(t, m - 1, a) - \frac{v}{\tau} P_1(t, m, a) \text{ for } \sigma = O_1. \]

The second contribution can be rewritten as

\[ \frac{\nu}{\tau} (E^- - \text{id})P_\sigma(t, m, a) \text{ for } \sigma = O_1 \text{ and } m \geq 1. \]

where \( E^- \) is the difference operator defined by

\[ E^- f(m) = f(m - 1) \text{ for every } f: \delta \mathbb{N} \rightarrow \mathbb{R}. \]

The operator \( L_E^s \) is defined as

\[ L_E^s = \frac{1}{\tau} \begin{pmatrix} 0 & 0 \\ 0 & \nu (E^- - \text{id})(\cdot) \end{pmatrix}. \]

The ME can now be written as

\[ \frac{\partial P}{\partial t} = L^* P + K^T P, \]
where \( \mathcal{L}^* = \mathcal{L}_R^* + \mathcal{L}_E^* \). This is the matrix form of

\[
\begin{align*}
\frac{dP_0(t, m, a)}{dt} &= \frac{\nu (m + 1)}{\tau} P_0(t, m + 1, a) - \frac{\nu m}{\tau} P_0(t, m, a) + \\
&\quad - \frac{a k^+}{\tau} P_0(t, m, a) + \frac{k^-}{\tau} P_1(t, m, a) \\
\frac{P_1(t, m, a)}{dt} &= \frac{\nu (m + 1)}{\tau} P_1(t, m + 1, a) - \frac{\nu m}{\tau} P_1(t, m, a) + \\
&\quad + \frac{v}{\tau} P_1(t, m - 1, a) - \frac{v}{\tau} P_1(t, m, a) + \frac{a k^+}{\tau} P_0(t, m, a) - k^- P_1(t, m, a)
\end{align*}
\]

This system will be fully analysed in the second part, see [23]. The boundary conditions are the natural one at \( m = 0 \) (see [28]) and are given by

\[
\begin{align*}
\frac{dP_0(t, 0, a)}{dt} &= \frac{\nu}{\tau} P_0(t, 1, a) - \frac{a k^+}{\tau} P_0(t, 0, a) + \frac{k^-}{\tau} P_1(t, 0, a) \\
\frac{P_1(t, 0, a)}{dt} &= \frac{\nu}{\tau} P_1(t, 1, a) + \frac{a k^+}{\tau} P_0(t, 0, a) - k^- P_1(t, 0, a)
\end{align*}
\]

3 Construction of the continuum approximation

The ME results from the specification of the reactions at a given scales \( \delta, \tau \). The ME describes the evolution of a probability measure \( P_\sigma(t; n) \) according to

\[
\frac{\partial P}{\partial t} = A^*[\delta, \tau]P,
\]

where \( A^*[\delta, \tau] \) is the infinitesimal generator defined on the scales \( \delta, \tau \) by

\[
A^*[\delta, \tau] = \mathcal{L}^*[\delta, \tau] + \mathcal{K}_T[\delta, \tau].
\]

The operator \( A^*[\delta, \tau] \) is defined on the space

\[
X_{\delta, \tau}^* = \left\{ P_\sigma(t; n) : \sum_{n \in \mathbb{N}_0} \sum_{\sigma \in \Sigma} P_\sigma(t; n) = 1 \text{ for all } t \right\}.
\]

Let us now consider a sequence of scales \( \delta_n, \tau_n \) such that \( \delta_n \to 0 \) and \( \tau_n \to 0 \) as \( n \to \infty \). For each index \( n \) we have an operator \( A_n = A[\delta_n, \tau_n] \) defined on \( X_n^* = X_{\delta_n, \tau_n}^* \) where the configuration space can now be denoted by \( \mathbb{I}_n = \mathbb{I}_{\delta_n} \). We ask ourselves what would be the fate of (16) as \( n \to \infty \).

We can think of \( \delta_n \to 0 \) and \( \tau_n \to 0 \) as limit at which space and time step become continuous and the numbers of particles are sufficiently to be accounted as densities and this motivates the name \textit{continuum limit}.
The formulation of the continuum limit can be obtained by using the approximation scheme introduced by Trotter in [27], (see also [19], [11]). To introduce Trotter approach we first need to observe that to each $A^*_n$ defined on $X^*_n$ we can associate a vector space $X_n$ and an adjoint operator $A_n$. The vector space is defined by

$$X_n = \mathbb{X}_{\delta_n, \tau_n} = \left\{ u_\sigma(t, n) : \mathbb{L}_n \times \Sigma \to \mathbb{R}^g : \|u\|_\infty = \sup_{n \in \mathbb{L}_n, \sigma \in \Sigma} |u_\sigma(t, n)| < \infty, \text{ for all } t \right\}. \quad (18)$$

Each $X_n$ is dual to $X^*_n$ according to the pairing:

$$\langle u, P \rangle_n = \sum_{(n, \sigma) \in \mathbb{L}_n \times \Sigma} u_\sigma(n) P_\sigma(n). \quad (19)$$

The adjoint $A_n$ is defined by:

$$\langle A_n u, P \rangle_n = \langle u, A^*_n P \rangle_n. \quad (20)$$

Let us consider

$$u(t, n, \sigma) = \sum_{n', \sigma'} P(t, n, n', \sigma, \sigma') u(n', \sigma')$$

then to equation (15) we now associate

$$\frac{\partial u}{\partial t} = A_n u, \quad (21)$$

defined on each $X_n$. Here

$$A_n = \lim_{t \to 0} \frac{1}{t} (P^t - \text{id}) \quad (22)$$

see [29] for all the details.

For any index $n$ equation (21) is the standard Kolmogorov and $A_n$ is the infinitesimal generator of Markov process on $\mathbb{L}_n \times \Sigma$. The definition of the continuum limit is based on the choice a target space where the limit is attained. We shall consider as target the space of continuous function $\mathbb{X} = C^0(\mathbb{R}_+^N, \mathbb{R}^g)$. The topological dual of $\mathbb{X}$ is formed by signed measures on $\mathbb{R}_+^N \times \Sigma$:

$$\mathbb{X}^* = \{ \rho(\mathbf{x}) : \langle \rho, u \rangle < \infty, \ u \in \mathbb{X} \}, \quad (23)$$

where the pairing is defined by

$$\langle \rho, u \rangle = \int_{\mathbb{R}_+^N} d\mathbf{x} \sum_{\sigma \in \Sigma} \rho_\sigma(\mathbf{x}) u_\sigma(\mathbf{x}).$$

According to [27] we define a sequence of projections

**Definition 3.1.** Let $\mathcal{P}_n : \mathbb{X} \mapsto \mathbb{X}_n$ be the operator that maps $u \in \mathbb{X}$ to $\mathcal{P}_n(u) \in \mathbb{X}_n$ defined as

$$\mathcal{P}_n(u)(k) = u(k \delta_n) = u(k_1 \delta_n^1, ..., k_N \delta_n^N).$$
The following holds true

**Proposition 3.1.** The projections $P_n$ satisfy the following properties

(i) $\|P_n\|_n < 1$,

(ii) $\lim_{n \to \infty} \|P_n(u)\|_n = \|u\|_\infty$ for every $u \in X$.

**Proof.** Part (i) follows from:

$$\|P_n\|_n = \sup_{u \in X} \frac{\|P_n(u)\|_n}{\|u\|_\infty}$$

and from

$$\|P_n(u)\|_n = \sup_{k \in L_n} \|u(k \tilde{\delta}_n)\| < \|u\|_\infty$$

for $L_n \subset L_{n+1}$.

Let us now show (ii). We know that $\|P_n(u)\|_n < \|u\|_\infty$. Since $L_n \subset L_{n+1}$ the following holds

$$\|P_n(u)\|_n \leq \|P_{n+1}(u)\|_{n+1}.$$ 

therefore the sequence $\|P_n(u)\|_n$ is increasing and as $n \to \infty$

$$\|P_n(u)\|_n \to \sup_n \|P_n(u)\|_n = \|u\|_\infty.$$

Following (27) the projectors $P_n$ allow to define in what sense the spaces $X_n$ approximate $X$.

**Definition 3.2.** A sequence $u_n \in X_n$ converges to $u \in X$ if

$$\|P_n(u) - u_n\|_n \to 0$$

as $n \to \infty$.

We denote this by $u_n \approx u$.

**Remark 3.1.** Condition (ii) on $P_n$ guarantees that the limit $\approx$ is unique. In fact if $u_n \approx u$ and $u_n \approx u'$ then by (ii) we can estimate

$$\|u - u'\|_\infty = \lim_{n \to \infty} \|P_n(u - u')\|_n \leq \lim_{n \to \infty} \|P_n(u) - u_n\| + \lim_{n \to \infty} \|P_n(u') - u_n\|$$

which of course goes to 0 as $n \to \infty$.

We now give the definition for the limit, in fact the continuum limit, of a sequence of operators $A_n$. This definition is inspired by the one presented in [27]. In fact in the present case we have to consider that the operators are functions of the scales $\tilde{\delta}_n$ and $\tau_n$. Therefore we set
Definition 3.3. Let $\mathcal{A}_n : X_n \mapsto X_n$ be a sequence of linear operators. We say that $\hat{\mathcal{A}} : X \mapsto X$ is the continuum limit of $\mathcal{A}_n$ (denoted by $\mathcal{A}_n \approx \hat{\mathcal{A}}$) if there exists a sequence of scales $\delta_n, \tau_n$ such that

1. $\delta_n \to 0$, $\tau_n \to 0$,

2. the domain of $\mathcal{A}$ is

$$D(\hat{\mathcal{A}}) = \{ u \in X : \mathcal{P}_n(u) \in D(\mathcal{A}_n), \ A_n(\mathcal{P}_n(u)) \text{ converges} \},$$

3. and $\|\mathcal{P}_n(\hat{\mathcal{A}}(u)) - A_n(\mathcal{P}_n(u))\|_n \to 0$ as $n \to \infty$.

Remark 3.2. The dependence on the choice of the scales $\delta_n$ and $\tau_n$ makes the continuum limit non unique. This is very important because with the choice of the scaling we will be able to analyse different type of processes.

![Figure 2: A schematic view of the Continuum Limit](image)

In what follows we shall construct examples $\mathcal{A}_n^*$ that correspond to Master Equations on $X_n^*$. Via the duality we derive the infinitesimal generator $\mathcal{A}_n$, i.e. the Kolmogorov equation on $X_n$. At that point we study the continuum limit of $\mathcal{A}_n$ and finally via the duality between $X^*$ and $X$ we derive the Fokker-Planck equation on $X^*$. This is illustrated in Figure 2.

3.1 Examples

In this section we illustrate the theory by studying the continuum limit of some crucial examples that will be used in the applications to reaction networks.
3.1.1 Difference operator

Let $L_n = \delta_n \mathbb{Z}$ and $X^*_n$ be the space of probability distribution on $L_n$ and $X = C^0(\mathbb{R}, \mathbb{R})$. Consider the following operator

$$\Delta_n^*(P)(k) = \frac{1}{\tau_n}(E^+ - \text{id})(a_n P)(k) = \frac{1}{\tau_n}[a_n(k + 1)P(k + 1) - a_n(k)P(k)]$$

The adjoint $\Delta_n$ is defined by

$$(\Delta_n^* P, u)_n = (P, \Delta_n u).$$

A simple calculation shows that

$$\Delta_n = \frac{a_n(k)}{\tau_n}(E^+ - \text{id}).$$

We now compute the continuum limit. Let $a \in X$ such that $P_n(a(k\delta_n)) = a(k\delta_n) = a_n(k)$. Let $\delta_n, \tau_n$ such that

$$\delta_n \to 0, \tau_n \to 0 \text{ with } \frac{\delta_n}{\tau_n} \to c > 0 \text{ as } n \to \infty.$$

Then

$$\Delta_n \approx -ca(x)\Delta, \text{ where } \Delta = \frac{\partial}{\partial x}.$$ 

In fact consider take $u \in C^2(\mathbb{R}, \mathbb{R}) \subset X$

$$\|\Delta_n(P_n(u)) - P_n(\Delta(u))\|_n = \left\| \frac{a_n(k)}{\tau_n}(E^+ - \text{id}) u(k \delta_n) + ca(k \delta_n) \frac{\partial u}{\partial x}(k \delta_n) \right\|_n.

This can be rewritten as

$$\left\| \frac{a_n(k)}{\tau_n} \left(u((k - 1)\delta_n) - u(k \delta_n)\right) + ca(k \delta_n) \frac{\partial u}{\partial x}(k \delta_n) \right\|_n =
$$

$$= \left\| a_n(k) \left(\frac{\delta_n}{\tau_n} - c\right) \left(u((k - 1)\delta_n) - u(k \delta_n)\right) + c a(k \delta_n) \left(\frac{\partial u}{\partial x}(k \delta_n) + \frac{u((k - 1)\delta_n) - u(k \delta_n)}{\delta_n}\right) \right\|_n
$$

The last term is bounded by

$$\left\| a(k \delta_n) \left(\frac{\delta_n}{\tau_n} - c\right) \left(u((k - 1)\delta_n) - u(k \delta_n)\right) + c a(k \delta_n) \left(\frac{\partial u}{\partial x}(k \delta_n) + \frac{u((k - 1)\delta_n) - u(k \delta_n)}{\delta_n}\right) \right\|_n \leq
$$

$$\leq \left|\frac{\delta_n}{\tau_n} - c\right| \sup_k |a(k \delta_n)| \sup_k \left|\frac{u((k - 1)\delta_n) - u(k \delta_n)}{\delta_n}\right| + c \sup_k |a(k \delta_n)| \sup_k \left|\frac{\partial u}{\partial x}(k \delta_n) + \frac{u((k - 1)\delta_n) - u(k \delta_n)}{\delta_n}\right|$$

Now as $n \to \infty$

$$\left|\frac{\delta_n}{\tau_n} - c\right| \to 0,$$
\[ \sup_k |a(k\delta_n)| \text{ is bounded, the term} \]
\[ \sup_k \left| \frac{u((k-1)\delta_n) - u(k\delta_n)}{\delta_n} \right| = \sup_k \frac{1}{\delta_n} \left| \int_{(k-1)\delta_n}^{k\delta_n} \frac{dx}{\partial x} u(x) \right| \leq \left\| \frac{u(x)}{\partial x} \right\|_{\infty}. \]

For the term
\[ \sup_k \left| \frac{\partial u}{\partial x}(k\delta_n) + \frac{u((k-1)\delta_n) - u(k\delta_n)}{\delta_n} \right| \]
we use
\[ u(y) = u(x) + u'(x)(y-x) + \int_x^y ds(s-x) \frac{\partial^2 u(s)}{\partial s^2} \]
to obtain
\[ \sup_k \left| \frac{\partial u}{\partial x}(k\delta_n) + \frac{u((k-1)\delta_n) - u(k\delta_n)}{\delta_n} \right| = \]
\[ = \sup_k \frac{1}{\delta_n} \left| \int_{(k-1)\delta_n}^{k\delta_n} ds(s-k\delta_n) \frac{\partial^2 u(s)}{\partial s^2} \right| \leq \]
\[ \leq \left\| u'' \right\|_{\infty} \frac{1}{\delta_n} \sup_k \left| \int_{(k-1)\delta_n}^{k\delta_n} ds(s-k\delta_n) \right| \leq \left\| u'' \right\|_{\infty} \delta_n \frac{2}{\delta_n} \to 0 \text{ as } n \to \infty. \]

We therefore conclude \( \mathcal{A}_n \approx \hat{\mathcal{A}}. \) A simple integration by parts shows that
\[ \Delta^*(\rho)(x) = \frac{\partial}{\partial x}(ca(x)\rho(x)) \text{ for all } \rho \in \mathcal{X}^*. \]

### 3.1.2 Multiplication operator

Let \( K_n = K_{\delta_n,\tau_n}(n) \) be defined on \( \mathcal{X}_n. \) We want to find \( K \) is a matrix operator acting by multiplication on \( \mathcal{X} \) such that
\[ \left\| K_n(P_n(u)) - P_n(K(u)) \right\|_n \to 0 \quad \forall u \in \mathcal{X}. \]

This condition holds if we take \( K(x) = \frac{1}{\varepsilon} K(x) \) such that
\[ \left\| K_{\delta_n,\tau_n}(n)u(k\delta_n) - \frac{1}{\varepsilon} K(u(k\delta_n)) \right\|_n \to 0 \quad \forall u \in \mathcal{X}, \]
where \( \varepsilon \) is chosen so that the limits holds.

Let us consider the following example
\[ K_n = \frac{1}{\varepsilon} \begin{pmatrix} -mk^+(\delta_n,\tau_n) & mk^+(\delta_n,\tau_n) \\ k^-(\delta_n,\tau_n) & -k^-(\delta_n,\tau_n) \end{pmatrix}. \]

We seek the continuum limit in the form
\[ K = \frac{1}{\varepsilon} \begin{pmatrix} -xk^+ & xk^+ \\ k^- & -k^- \end{pmatrix} \]

17
with $k^\pm$ positive constants. Now in order to construct the expression

$$||K_n P_n(u) - P_n(K(u))||_n$$

we compute

$$K_n(P_n(u))(m) = \frac{1}{\tau_n} \begin{pmatrix} -m k^+(\delta_n, \tau_n)(u_1(m \delta_n) - u_1(m \delta_n)) \\ k^-(\delta_n, \tau_n)(u_1(m \delta_n) - u_1(m \delta_n)) \end{pmatrix}$$

and

$$P_n(K(u))(m) = \frac{1}{\epsilon} \begin{pmatrix} -m \delta_n k^+(\delta_n, \tau_n)(u_1(m \delta_n) - u_1(m \delta_n)) \\ k^-(\delta_n, \tau_n)(u_1(m \delta_n) - u_1(m \delta_n)) \end{pmatrix}.$$  

Therefore we can have

$$\lim_{n \to \infty} ||K_n P_n(u) - P_n(K(u))||_n = 0$$

for all $u \in X$ if the following relations hold

$$\frac{\delta_n k^+}{\epsilon} = \frac{k^+(\delta_n, \tau_n)}{\tau_n}, \quad \frac{k^-}{\epsilon} = \frac{k^-(\delta_n, \tau_n)}{\tau_n}.$$  

These relations have to satisfy the compatibility condition

$$\frac{\delta_n k^+}{k^-} = \frac{k^+(\delta_n, \tau_n)}{k^-(\delta_n, \tau_n)}.$$  

Here we see that if

$$k^+(\delta_n, \tau_n) = \delta_n k^+, \quad k^-(\delta_n, \tau_n) = k^-$$

then the limit exits for $\epsilon = \tau_n$, which is infinitesimal for large $n$.

**Remark 3.3.** The choice of $\epsilon$ as a function of the scales $\delta_n, \tau_n$ is essentially a form of renormalisation of the time scales associated to respectively the infinitesimal generator $\hat{L}$ and $K$.

### 3.1.3 Diffusion operator

We consider the one dimensional case $L_n = \delta_n \mathbb{Z}$. Let us consider the operator

$$A_n = \frac{1}{\tau_n} [a_n(k) \Delta_n^- + b_n(k) \Delta_n^+]$$

We now take $a, b \in X$ such that $P_n(a) = a_n$ and $P_n(b) = b_n$. In order to construct a continuum limit we compute

$$A_n(P_n(u)) = \frac{1}{\tau_n} a_n \Delta_n^- (u) + \frac{1}{\tau_n} b_n \Delta_n^+ (u) =$$

$$= \frac{a_n(k)}{\tau_n} [u((k-1)\delta_n) - u(k\delta_n)] + \frac{b_n(k)}{\tau_n} [u((k+1)\delta_n) - u(k\delta_n)].$$
We now re-write this operator using \( u \in C^3(\mathbb{R}, \mathbb{R}) \) using:

\[
u(y) = u(x) + u'(x)(y - x) + \frac{(y - x)^2}{2} u''(x) + \frac{1}{6} \int_x^y ds(s - x)^2 \frac{\partial^3 u(s)}{\partial s^3}.
\]

We have

\[
\mathcal{A}_n(\mathcal{P}_n(u)) = a_n(k) \left[ -\frac{\partial_n u(k\delta_n)}{\tau_n} + \frac{\delta_n^2}{2} \frac{\partial^2 u(k\delta_n)}{\partial x^2} \right] + b_n(k) \left[ \frac{\delta_n^2}{\tau_n} \frac{\partial u(k\delta_n)}{\partial x} + \frac{\delta_n^2}{2} \frac{\partial^2 u(k\delta_n)}{\partial x^2} \right] + \frac{a_n(k)}{6\tau_n} \int_{k\delta_n}^{(k-1)\delta_n} ds(s - k\delta_n)^2 \frac{\partial^3 u(s)}{\partial s^3} + \frac{b_n(k)}{6\tau_n} \int_{k\delta_n}^{(k+1)\delta_n} ds(s - k\delta_n)^2 \frac{\partial^3 u(s)}{\partial s^3},
\]

this form can be rewritten as

\[
\mathcal{A}_n(\mathcal{P}_n(u)) = \frac{\delta_n}{\tau_n} \frac{\partial_n u(k\delta_n)}{\partial x} + \frac{\delta_n^2}{2\tau_n} \frac{\partial^2 u(k\delta_n)}{\partial x^2} + \frac{a_n(k)}{6\tau_n} \int_{k\delta_n}^{(k-1)\delta_n} ds(s - k\delta_n)^2 \frac{\partial^3 u(s)}{\partial s^3} + \frac{b_n(k)}{6\tau_n} \int_{k\delta_n}^{(k+1)\delta_n} ds(s - k\delta_n)^2 \frac{\partial^3 u(s)}{\partial s^3},
\]

Now we need to guess a suitable operator \( \hat{A} \). Here we take

\[
\hat{A} = \alpha(x) \frac{\partial}{\partial x} + \beta(x) \frac{\partial^2}{\partial x^2}
\]

with \( \alpha, \beta \in \mathcal{X} \). Therefore

\[
\hat{A}(\mathcal{P}_n(u))(k) = \frac{\partial(u(k\delta_n))}{\partial x} + \beta(k\delta_n) \frac{\partial^2 u(k\delta_n)}{\partial x^2}.
\]

If we set \( \delta_n^2/\tau_n \to 1 \) and

\[
\|\alpha(k\delta_n) - \frac{\delta_n}{\tau_n}(b_n(k) - a_n(k))\|_n = o(1/n), \quad \|\beta(k\delta_n) - \frac{\delta_n^2}{2\tau_n}(b_n(k) + a_n(k))\|_n = o(1/n)
\]

then

\[
\|\mathcal{A}_n(\mathcal{P}_n(u)) - \mathcal{P}_n(\hat{A}(u))\|_n \leq o(1/n)(\|u'\|_\infty + \|u''\|_\infty) + \|R(\delta_n, \tau_n)\|_n,
\]

where

\[
R(\delta_n, \tau_n) = a_n(k) \int_{k\delta_n}^{(k-1)\delta_n} ds(s - k\delta_n)^2 \frac{\partial^3 u(s)}{\partial s^3} + \frac{b_n(k)}{6\tau_n} \int_{k\delta_n}^{(k+1)\delta_n} ds(s - k\delta_n)^2 \frac{\partial^3 u(s)}{\partial s^3}.
\]

Now we can estimate

\[
\|R(\delta_n, \tau_n)\|_n \leq \|a_n + b_n\|_n \left\| \frac{\partial^3 u}{\partial s^3} \right\|_\infty \sup_k \left[ \frac{1}{3}(k - 1 - k)^3 + \frac{1}{3}(k + 1 - k)^3 \right] \leq \|a_n + b_n\|_n \left\| \frac{\partial^3 u}{\partial s^3} \right\|_\infty \frac{\delta_n^3}{9\tau_n}
\]

Note that \( \|a_n + b_n\|_n \to \|\alpha + \beta\|_\infty \) and thus have \( \|R(\delta_n, \tau_n)\|_n \to 0 \) as \( n \to \infty \).

Also in this case using the pairing between \( \mathcal{X} \) and \( \mathcal{X}^* \) we can show by an integration by parts that

\[
\hat{A}^*(\rho(x)) = -\frac{\partial}{\partial x}(\alpha(x)\rho(x)) + \frac{\partial^2}{\partial x^2}(\beta(x)\rho(x))
\]

for all \( \rho \in \mathcal{X}^* \).
3.2 Formulation of the multi-scale analysis

We call a multiscale analysis the combination of continuum approximation and an adiabatic approximation for the ME of an IFSS. We construct the continuum approximation of the birth-death processes describing the small molecules first and will get a Fokker-Planck equation (FPE). Next we study the adiabatic approximation of the Markov chain, ending the multi-scale analysis.

3.2.1 Continuum limit of the ME

Definition 3.4 (Continuum approximation). Consider the infinitesimal generator associated to ME of an IFSS given by

\[
\frac{\partial u(n,t)}{\partial t} = L[\delta, \tau](n)u(n,t) + K[\delta, \tau](n)u(n,t),
\]  

(25)

defined on \(X_\delta\). We say that (25) admits as continuum limit

\[
\frac{\partial u(x,t)}{\partial t} = \hat{L}(x)u(x,t) + \frac{1}{\epsilon}K(x)u(x,t),
\]  

(26)

defined on \(X\) if there exist a choice of the scaling \(\delta_n \to 0, \tau_n \to 0\) as \(n \to \infty\) such that

(i) \(L_n = L[\delta_n, \tau_n] \approx \hat{L}\),

(ii) \(K_n = K[\delta_n, \tau_n] \approx \frac{1}{\epsilon}K\) for some choice of \(\epsilon\).

Remark 3.4. Clearly equation (27) depends on the scales \(\delta_n, \tau_n\) through the parameter \(\epsilon\).

Illustrative example: continuum limit

In the illustrative example from enzyme kinetics the matrix \(L^*\) is

\[
L = L_E + L_R = \frac{1}{\tau} \begin{pmatrix}
\nu m (E^+ - \text{id})(\cdot) & 0 \\
0 & \nu m (E^- - \text{id})(\cdot) + \nu (E^+ - \text{id})(\cdot)
\end{pmatrix},
\]

and

\[
K = \frac{1}{\tau} \begin{pmatrix}
-a k^+ & a k^+ \\
k^- & -a k^-
\end{pmatrix}.
\]

Assuming that we can find \(\epsilon = \epsilon(\delta, \tau)\) such that

\[
(\delta \nu(\delta, \tau)/\tau) = \nu, \quad (\delta v(\delta, \tau)/\tau) = v, \quad (a k^+(\delta, \tau)/\tau) = (a \delta k^+)/\epsilon, \quad (k^-(\delta, \tau)/\tau) = k^-/\epsilon.
\]
Taking \( \epsilon(\delta, \tau) = \tau \) with \( \delta/\tau \to 1 \), by recalling section 3.1.1 we obtain

\[
\hat{\mathcal{L}} = \begin{pmatrix}
-\nu x \Delta(\cdot) & 0 \\
0 & -(\nu x - v) \Delta(\cdot)
\end{pmatrix},
\]

where \( x \) is the concentration associated to \( m \) and

\[
\mathcal{K} = \frac{1}{\epsilon} \begin{pmatrix}
-a k^+ & a k^+ \\
-k^- & -k^-
\end{pmatrix}.
\]

The continuum limit provides the construction of the infinitesimal operator \( \hat{A} \)

\[
\hat{A}(x) = \hat{\mathcal{L}}(x) + \frac{1}{\epsilon} \mathcal{K}(x)
\]

densely defined in \( X \) and the associated Kolmogorov equation

\[
\frac{\partial u(x, t)}{\partial t} = \hat{A}(x)u(x, t).
\]

(28)

For this equation there is an adjoint formulation. This is the FPE that is the dynamics read on the the space of probability measures \( X^* \). The FPE is

\[
\frac{\partial \rho(x, \tau)}{\partial \tau} = \hat{A}^*(x)\rho(x, t),
\]

(29)

where \( \hat{A}^* \) is the adjoint of \( \hat{A} \)

\[
\hat{A}^* = \hat{\mathcal{L}}^*(x) + \frac{1}{\epsilon} \mathcal{K}^T(x).
\]

In most of the application the interests is concentrated on the the FPE. For this reason, in the next section we shall study the adiabatic limit for (29). The dynamics on \( X \) is related to the dynamics on \( X^* \) by the following the following result (see [7, 29]):

**Theorem 3.1.** Let \( X \) be the Banach space of continuous functions from \( \mathbb{R}^N \) to \( (\mathbb{R}^g, \langle.,.\rangle) \). For \( \epsilon > 0 \) fixed, assume

(i) \( \hat{\mathcal{L}}(x) \) has a dense domain in \( X \) and

\[
\|(\hat{\mathcal{L}}(x) + \alpha)^{-1}\| < \frac{1}{\alpha} \quad \forall \alpha > 0,
\]

where \( \|\|.\| \) is the standard norm on the space of linear operators in \( C^0(\mathbb{R}^N_+, \mathbb{R}^g) \).

(ii) \( \mathcal{K}(x) \) is a \( g \times g \) matrix with bounded entries, and also the infinitesimal generator of a \( g \)-dimensional Markov chain.
Then equation
\[
\frac{\partial u(x,t)}{\partial t} = \hat{L}(x)(u(x,t)) + \frac{1}{\epsilon} K(x) u(x,t)
\] (30)

admits a solution \( u(x,t) \), and the operator \( \hat{L}(x) + \frac{1}{\epsilon} K(x) \) generates a Markov process on \( \mathbb{R}^N \times S \) whose distribution \( \rho(x,t) \) satisfies equation (29).

As we have seen in the example \( \epsilon \) is small, therefore FPE (29) for \( \rho \) is singular at \( \epsilon = 0 \). For this reason we shall consider a perturbation analysis by means of an asymptotic series in \( \epsilon \) in the spirit of [18]. This will require a time-scale analysis of the FPE which is called adiabatic approximation. We will need the following definitions:

**Definition 3.5.** Let \( U \) be an open and bounded set in \( \mathbb{R}^N_+ \), \( T > 0 \), and let \( C^{r,s} \) be the Banach space of functions from \( U \times [0,T] \subset \mathbb{R}^N_+ \times \mathbb{R} \) which are \( r \)-times continuously differentiable w.r.t. \( x \in U \subset \mathbb{R}^N \) and \( s \)-times continuously differentiable w.r.t. \( t \in [0,T] \subset \mathbb{R} \).

We need to recall the definition of asymptotic series:

**Definition 3.6** (Asymptotic series). Let \( v_\epsilon(x,t) \) be the formal power series
\[
v_\epsilon(x,t) = \sum_{k=0}^{\infty} v^{(k)}(x,t) \epsilon^k.
\]
Then we say \( v_\epsilon \) converges asymptotically to \( v_0 \) for small \( \epsilon > 0 \) if its partial sum
\[
v^m_\epsilon(x,t) = \sum_{k=0}^{m} v^{(k)}(x,t) \epsilon^k
\]
is such that
\[
\sup_{x \in U \subset \mathbb{R}^N, t \in [0,T]} \|v^m_\epsilon(x,t) - v_\epsilon(x,t)\| \leq C(U,T) \epsilon^{m+1}
\]
for some \( C(U,T) > 0 \).

4 Adiabatic theory

In this section we construct the solution of the FPE generated by an IFSS by means of asymptotic expansions. Only the respective result for the Kolmogorov equation will be stated. This choice is not generic from a purely mathematical point of view as a solution of the FPE will require more regularity conditions. Nevertheless the study of the FPE with smooth coefficients and smooth initial data is very relevant in many applications, motivating the subsequent presentation.
4.1 Formulation of the adiabatic problem

An IFSS has by necessity two time scales after taking the continuum limit. One is characterising the dynamics of the continuous degrees of freedom and the other one characterising the evolution of the finite state Markov chain. These time scales are associated to the following semi-groups:

(i) \( \exp(t \hat{\mathcal{L}}(x)) \), the semigroup generated by \( \hat{\mathcal{L}}(x) \),

(ii) \( \exp((t/\epsilon)K(x)) \), which is the semigroup generated by \( K(x) \), the Markov chain generator.

On a time scale of order \( O(\epsilon) \) the Markov chain dynamics should prevail. Without loss of generality we can assume that the Markov chain has at least one invariant measure, and possibly a convex combination of stationary measures. Therefore on long time scales one expects that the Markov chain reaches an equilibrium very fast and the dynamics should essentially be given by the flow associated to \( \hat{\mathcal{L}}(x) \).

4.1.1 Diffusive and deterministic operators

The operators \( \hat{\mathcal{L}}(x) \) and \( \hat{\mathcal{L}}^*(x) \) are differential operators which we investigate in the following, especially giving more details of their structure. The operator \( \hat{\mathcal{L}}^*(x) \) is a diagonal matrix with operator entries. Each non-degenerate nonzero entry is a second order linear parabolic operator.

**Definition 4.1** (Structure of \( \hat{\mathcal{L}}^* \)). Let

\[
\hat{\mathcal{L}}^*(x) \doteq \hat{\mathcal{L}}^1(x) + \hat{\mathcal{L}}^2(x) = \delta_{ij} \otimes (\hat{\mathcal{L}}^1_i(x) + \hat{\mathcal{L}}^2_i(x))
\]

Let each of these operators \( \hat{\mathcal{L}}^1(x) \) and \( \hat{\mathcal{L}}^2(x) \) be given by

\[
\hat{\mathcal{L}}^1_i(x)(f) = \sum_{\alpha} \frac{\partial}{\partial x_{\alpha}}(L^1_{\alpha i}(x) f(x)),
\]

and

\[
\hat{\mathcal{L}}^2_i(x)(f) = \frac{1}{2} \sum_{\alpha,\beta} \frac{\partial^2}{\partial x_{\alpha} \partial x_{\beta}}(C^i_{\alpha \beta}(x) f(x)),
\]

for \( f \) being a sufficiently smooth function. Then \( \hat{\mathcal{L}}^*(x) \) is called a regular Fokker-Planck operator.

If \( \hat{\mathcal{L}}^2 \equiv 0 \), the operator \( \hat{\mathcal{L}}^* \) becomes first order (i.e. non-regular) and describes the transport in the deterministic dynamics of \( i.d.f. \)’s. We can therefore identify two important regimes: With \( \hat{\mathcal{L}}^2 \neq 0 \) we have diffusive \( i.d.f. \)'s, and with \( \hat{\mathcal{L}}^2 \equiv 0 \) the \( i.d.f. \)'s are deterministic. The second possibility will be treated in part II of this paper series.
4.1.2 Main assumptions

In order to simplify the further analysis we make the following assumption:

(A) The Markov chain on $\Sigma$ has a set of stationary measures $M_K$ with $\dim(M_K) < g$. Each measure $\mu(x)$ is $C^\infty$ on $\mathbb{R}^N$.

The next two assumptions give the explicit conditions for constructing the solution, respectively for the FPE and the Kolmogorov equation.

(B) For a given $\mu(x) \in C_K$, the Cauchy problem

$$\partial_t f(x,t) = \langle 1_\mu, \hat{L}^*(x)(\mu(x)f(x,t)) \rangle + F(x,t), \quad f(x,0) \in C^{r,s}, \quad (31)$$

with $F \in C^{r,s}$ admits a solution which is $C^{r,s}$ w.r.t. $x$ and $t \in [0,T_0] \subset [0,T]$ for any smooth initial data.

(C) For a given $\mu(x) \in C_K$, the Cauchy problem

$$\partial_t \phi(x,t) = \langle \mu(x), \hat{L}(x)(1_\mu \phi(x,t)) \rangle + G(x,t), \quad \phi(x,0) \in C^{r,s}, \quad (32)$$

with $G \in C^{r,s}$, admits a solution which is $C^{r,s}$ w.r.t. $x$ and $t \in [0,T^*_0] \subset [0,T]$ for any smooth initial data.

**Remark 4.1.** In applications we have that the operators

$$\langle 1_\mu, \hat{L}^*(x)(\mu(x)\cdot) \rangle \text{ and } \langle \mu(x), \hat{L}(x)(1_\mu \cdot) \rangle$$

are either parabolic or first order. In the parabolic case note that there is a general result (see [18]), which guarantees that if the differential operator has $C^\infty$ coefficients and the initial condition is also $C^\infty$ then the solution is $C^{1,2}([0,T] \times \mathbb{R}^N) \cap C^\infty([0,T] \times \mathbb{R}^N)$ for some $T > 0$.

**Remark 4.2.** In conditions (B) and (C) respectively the intervals $[0,T_0]$ and $[0,T^*_0]$ are the maximal time intervals where each solution exists. Note that since solution of (B) implies (C), if (B) holds true then $T_0 = T^*_0$.

4.2 Main results

We now state and prove the main results of the adiabatic theory for IFSS. Theorem 4.1 is based on an elaboration of a respective proof presented in [18] and clarifies the construction presented in [6].

**Theorem 4.1.** For fixed $\mu \in C_K$ and under assumptions (A) and (B), equation (29) admits an asymptotic solution in each set of concentrated measures $I_\mu$ (see definition 5.4).
Proof. We start with a few remarks. In the appendix we describe the geometry associated to the Markov chain and in particular to the kernel of $K^T$ (see section 5.1). Note any initial data in $I_\mu$ evolve asymptotically to $\mu$. Next we like to solve the equation

$$\frac{\partial \rho(x, t)}{\partial t} = \hat{L}(\rho(x, t)) + \frac{1}{\epsilon} K(x) \rho(x, t)$$  \hspace{1cm} (33)$$

by using an asymptotic expansion and conditions (A), (B). Fix $\mu \in C_K$ and take an initial condition in $I_\mu$. Let $m^*$ be an integer to be determined later. We take the expansion

$$\rho_\epsilon(x, t) = \sum_{m=0}^{m^*} \epsilon^m \rho^{(m)}(x, t).$$

To determine $\rho^{(m)}(x, t)$ we substitute the expansion for $\rho_\epsilon$ into equation (33) and collect the different orders in $\epsilon$. We obtain a hierarchy of equations

$$O(1/\epsilon) : \quad K^T(x) \rho^{(0)}(x, t) = 0$$

$$O(1) : \quad \frac{\partial \rho^{(0)}(x, t)}{\partial t} - \hat{L}^*(\rho^{(0)}(x, t)) = K^T(x) \rho^{(1)}(x, t)$$

$$O(\epsilon) : \quad \frac{\partial \rho^{(1)}(x, t)}{\partial t} - \hat{L}^*(\rho^{(1)}(x, t)) = K^T(x) \rho^{(2)}(x, t)$$

$$\vdots$$

$$O(\epsilon^{m^*}) : \quad \frac{\partial \rho^{(m^*)}(x, t)}{\partial t} - \hat{L}^*(\rho^{(m^*)}(x, t)) = K^T(x) \rho^{(m^*+1)}(x, t).$$

Note that in the construction of the probability density $\rho_\epsilon$ with its necessary normalisation is not yet fixed. The equation (33) is linear in $\rho_\epsilon$. Therefore the condition

$$\int_{\mathbb{R}^N} dx \ tr (\rho_\epsilon(x, t)) = 1$$

must be imposed on the final form of the expansion. Conditions (A) and (B) guarantee the possibility of solving the first two equations in above hierarchy. Indeed condition (A) implies that the equation

$$K^T(x) \rho^{(0)}(x, t) = 0$$

adopts a solution of the form

$$\rho^{(0)}(x, t) = f^{(0)}(x, t) \mu(x).$$

The second equation becomes

$$\frac{\partial \mu(x)}{\partial t} f^{(0)}(x, t) - \hat{L}^*(\mu(x) f^{(0)}(x, t)) = K^T(x) \rho^{(1)}(x, t).$$  \hspace{1cm} (34)$$

25
By the Fredholm alternative theorem (see [18]) we have that a necessary condition for solving (34) is that the l.h.s. is orthogonal to the kernel of $K(x)$ (see [18]). Condition (A) implies that $\mathbf{1}_\mu$ satisfies

$$K(x) \mathbf{1}_\mu = 0.$$ 

Therefore

$$\left\langle \mathbf{1}_\mu, \frac{\partial \mu(x)}{\partial t} f^{(0)}(x, t) - \hat{L}^*(\mu(x) f^{(0)}(x, t)) \right\rangle = \langle \mathbf{1}_\mu, K^T(x) \rho^{(1)}(x, t) \rangle = \langle K(x) \mathbf{1}_\mu, \rho^{(1)}(x, t) \rangle = 0.$$

To determine a solution one needs to solve

$$\frac{\partial}{\partial t} \langle 1_\mu, \mu(x) f^{(0)}(x, t) \rangle - \langle \mathbf{1}_\mu, \hat{L}^*(\mu(x) f^{(0)}(x, t)) \rangle = 0,$$

which is equal to equation (31) upon noting the condition

$$\langle \mathbf{1}_\mu, \mu(x) \rangle = 1.$$

To proceed further we compute $\rho^{(1)}(x, t)$. Note that any $\rho$ as a vector in $\mathbb{R}^g$ can be decomposed by projection $\Pi_\mu$ (see section 5.1). Let $\rho_i^{(n)}(x, t)$ be the $n$th term of the expansion

$$\rho^{(n)}(x, t) = \Pi_\mu(\rho^{(n)}(x, t)) + (\mathbf{1}_\mu - \Pi_\mu)(\rho^{(n)}(x, t)) = \xi^{(n)}(x, t) + f^{(n)}(x, t) \mu(x).$$

Using that $\Pi_\mu = (K^T_\mu)^D I_\mu K^T(x)$, we have

$$\Pi_\mu(\rho^{(1)}(x, t)) = (K^T_\mu)^D I_\mu \left[ \frac{\partial \rho^{(0)}(x, t)}{\partial t} - \hat{L}^*(\rho^{(0)}(x, t)) \right].$$

Since $\rho^{(0)}(x, t) = f^{(0)}(x, t) \mu(x)$ and $(K^T_\mu)^D \mu(x) = 0$ we get

$$\xi^{(1)}(x, t) = \Pi_\mu(\rho^{(1)}(x, t)) = -(K^T_\mu)^D I_\mu \hat{L}(\mu(x) f^{(0)}(x, t)).$$

To construct $\rho_1(x, t)$ we also need that

$$\langle \mathbf{1}_\mu - \Pi_\mu \rangle(\rho^{(1)}(x, t)) = \langle 1, \rho^{(1)}(x, t) \rangle = \mu(x) f^{(1)}(x, t).$$

Like for $\rho^{(0)}$ we obtain

$$\frac{\partial \rho^{(1)}(x, t)}{\partial t} - \hat{L}^*(\rho^{(1)}(x, t)) = K^T(x) \rho^{(2)}(x, t),$$

which can be projected on $\mathbf{1}_\mu$ leading to

$$\frac{\partial f^{(1)}(x, t)}{\partial t} - \langle \mathbf{1}_\mu, \hat{L}^*(\rho^{(1)}(x, t)) \rangle = 0.$$
Using the fact that
\[ \rho^{(1)}(x,t) = \xi^{(1)}(x,t) + \mu(x)f^{(1)}(x,t) = -(\mathcal{K}^T_\mu)^D I_\mu \hat{\mathcal{L}}^* (\mu(x)f^{(0)}(x,t)) + \mu(x)f^{(0)}(x,t), \]
the equation for \( f^{(1)} \) becomes
\[ \frac{\partial f^{(1)}(x,t)}{\partial t} - \langle 1_\mu, \hat{\mathcal{L}}^* (\mu(x)f^{(1)}(x,t)) \rangle + \hat{\mathcal{L}}^* ((\mathcal{K}^T_\mu)^D I_\mu \hat{\mathcal{L}}^* (\mu(x)f^{(0)}(x,t)))) = 0. \]

The argument can be iterated. Therefore the term \( \rho^{(n)}(x,t) \) is determined by computing its projections
\[ \xi^{(n)}(x,t) : \quad \Pi_\mu (\rho^{(n)}(x,t)) = \xi^{(n)}(x,t) \]
\[ f^{(n)}(x,t) : \quad (I_n - \Pi_\mu)(\rho^{(n)}(x,t)) = \mu(x)f^{(n)}(x,t). \]

Let us assume we know \( \rho^{(k)} \) from \( k = 0 \) up to \( n - 1 \). Then \( \xi^{(n)} \) is obtained by projecting the equation
\[ \frac{\partial \rho^{(n-1)}(x,t)}{\partial t} - \hat{\mathcal{L}}^* (\rho^{(n-1)}(x,t)) = \mathcal{K}^T(x)\rho^{(n)}(x,t), \]
namely
\[ \xi^{(n)}(x,t) = \Pi_\mu (\rho^{(n)}(x,t)) = (\mathcal{K}^T_\mu)^D I_\mu \left[ \frac{\partial \rho^{(n-1)}(x,t)}{\partial t} - \hat{\mathcal{L}}^* (\rho^{(n-1)}(x,t)) \right]. \quad (35) \]

The term \( f^{(n)} \) is determined by projecting the equation
\[ \frac{\partial \rho^{(n)}(x,t)}{\partial t} - \hat{\mathcal{L}}^* (\rho^{(n)}(x,t)) = \mathcal{K}^T(x)\rho^{(n+1)}(x,t), \]
and therefore \( f^{(n)} \) solves
\[ \frac{\partial f^{(n)}(x,t)}{\partial t} - (1_\mu, \hat{\mathcal{L}}^* (\mu(x)f^{(n)}(x,t)) + \hat{\mathcal{L}}^* (\xi^{(n)}(x,t))) = 0. \quad (36) \]

This concludes the construction of the expansion.

We now prove that the expansion of \( \rho \) converges asymptotically. We extend a similar argument presented in [18]. We show that condition (B) allows us to evaluate the regularity of the asymptotic expansion. Recall that for \( n = 0 \)
\[ \begin{cases} 
  \xi^{(0)}(x,t) = 0 \\
  \frac{\partial f^{(0)}(x,t)}{\partial t} = (1_\mu, \hat{\mathcal{L}}^* (\mu(x)f^{(0)}(x,t))) 
\end{cases} \quad (37) \]
holds. Also for \( n \geq 1 \) we have

\[
\begin{align*}
\xi^{(n)}(x, t) &= (K_\mu^T D) I_\mu \left[ \frac{\partial \xi^{(n-1)}(x, t)}{\partial t} - \hat{L}^*(\xi^{(n-1)}(x, t) + \mu(x) f^{(n-1)}(x, t)) \right] \\
\frac{\partial f^{(n)}(x, t)}{\partial t} &= \langle 1_\mu, \hat{L}^*(\mu(x) f^{(n)}(x, t)) \rangle + \langle 1_\mu, \hat{L}^*(\xi^{(n)}(x, t)) \rangle. 
\end{align*}
\] (38)

Observe that \( f^{(n-1)}, \xi^{(n-1)} \in C^{r,s} \), therefore equations (38) and condition (B) imply that \( f^{(n)}, \xi^{(n)} \in C^{r-2,s-1} \). The solution of (38) defines the map

\[ \Psi : C^{r,s} \mapsto C^{r-2,s-1} \]

as follows:

\[ \Psi(\xi^{(n-1)}(x, t) + \mu(x) f^{(n-1)}(x, t)) = \xi^{(n)}(x, t) + \mu(x) f^{(n)}(x, t). \]

Now for \( n = 0 \) condition (B) implies (37) has a solution. Any initial condition in \( C^{r,s} \) yields \( f^{(0)}(x, t) \in C^{r,s} \). Using the map \( \Psi \) we can write

\[ \rho^{(n)}(x, t) = \Psi^n(\mu(x) f^{(0)}(x, t)), \]

with

\[ \rho^{(n)}(x, t) \in C^{r-2n,s-n} \text{ for } 0 \leq n \leq m^*. \]

Let us now fix the order of the asymptotic expansion to be

\[ m^* = \min \left\{ \frac{r-2}{2}, s+1 \right\}, \]

so that \( \rho^{(m^*)} \in C^{2,1} \). Write \( \rho_{\epsilon} \) as

\[ \rho_{\epsilon}(x, t) = \sum_{k=0}^{m^*} \epsilon^k \rho^{(k)}(x, t) + R(x, t), \]

where \( R(x, t) \) is the error term. Using the equation for \( \rho_{\epsilon} \) we derive an equation for this error term:

\[ \sum_{k=0}^{m^*} \epsilon^k \frac{\partial \rho^{(k)}(x, t)}{\partial t} + \frac{\partial R(x, t)}{\partial t} = \left( \hat{L}_\epsilon^* + \frac{1}{\epsilon} K_\mu^T(x) \right) \left( \sum_{k=0}^{m^*} \epsilon^k \rho^{(k)}(x, t) + R(x, t) \right). \]

Now using the equations for \( \rho^{(k)} \) for \( 1 < k < m^* \), we obtain

\[ \frac{\partial R(x, t)}{\partial t} = \hat{L}_\epsilon^*(R(x, t)) + \epsilon^{m^*} \left( \hat{L}_\epsilon(\rho^{(m^*)}(x, t)) - \frac{\partial \rho^{(m^*)}(x, t)}{\partial t} \right), \]

28
where
\[ \hat{L}^*_\epsilon = \hat{L}^* + \frac{1}{\epsilon} K^T(x). \]
The operator \( \hat{L}^*_\epsilon \) is a generator of a contraction semigroup for \( t \in [0, T_0) \). Therefore we can use its exponential to compute \( R(x, t) \), with
\[ R(x, t) = \exp (t \hat{L}^*_\epsilon) R(x, 0) + e^{m^*} \int_0^t ds \exp (t-s) \hat{L}^*_\epsilon \left( \hat{L}^*(\rho^{(m^*)}(x, s)) - \frac{\partial \rho^{(m^*)}(x, s)}{\partial s} \right). \]
Using the semigroup property the norm of \( \exp(t \hat{L}^*_\epsilon) \) can be bound by
\[ \sup_{x \in \Omega, t \in [0, T_0]} \| \exp(t \hat{L}^*_\epsilon) \| = 1. \]
This implies the estimate
\[ \sup_{x \in \Omega, t \in [0, T_0]} \| R(x, t) \| \leq \sup_{x} \| R(x, 0) \| + e^{m^*} \int_0^{T_0} ds \sup_{x,s} \left\| \hat{L}^*(\rho^{(m^*)}(x, s)) - \frac{\partial \rho^{(m^*)}(x, s)}{\partial s} \right\|. \]
Since \( m^* = (r - 2)/2 \) we have that
\[ \hat{L}^*(\rho^{(m^*)}(x, s)) - \frac{\partial \rho^{(m^*)}(x, s)}{\partial s} \in C_{0,s'}, \]
with \( s' \geq 0 \). Therefore there exists \( C_1(\Omega, T_0) > 0 \) such that
\[ \sup_{x \in \Omega, t \in [0, T_0]} \| R(x, t) \| \leq \sup_{x} \| R(x, 0) \| + e^{m^*} T_0 C_1(\Omega, T_0). \]
Taking an initial condition satisfying
\[ \sup_{x \in \Omega, t \in [0, T_0]} \| R(x, 0) \| = C_2(\Omega, T_0) e^{m^*}, \]
the final estimation on the error is given by
\[ \sup_{x \in \Omega, t \in [0, T_0]} \| R(x, t) \| \leq e^{m^*} (C_2(\Omega, T_0) + T_0 C_1(\Omega, T_0)). \]
Thus we get the estimate
\[ \sup_{x \in \Omega, t \in [0, T_0]} \| \rho(x, t) - \rho^{(m^*-1)}_{\epsilon}(x, t) \| \leq e^{m^*} (C_2(\Omega, T_0) + T_0 C_1(\Omega, T_0)). \]
This concludes the proof. \( \square \)
In case condition (B) does not hold the operator \( \langle 1_\mu, \hat{L}^* \mu(x) \cdot \rangle \) does not yield a probability density which is sufficiently smooth. In such circumstances one can look at a weaker hypothesis such as condition (C).

**Theorem 4.2.** For fixed \( \mu \in C_K \) and under assumptions (A), (C), equation (30) admits an asymptotic solution in the set of concentrated functions \( Y_{\mu} \). This solution gives rise to a solution for the Kolmogorov equation and therefore to a weak solution for (29).

**Proof.** The aim is to construct an asymptotic solution for

\[
\frac{\partial u(x, t)}{\partial t} = \hat{L}(x)(u(x, t)) + \frac{1}{\epsilon} K(x) u(x, t).
\]

(39)

In the following we only outline the proof of theorem 4.2. The method used is close to the proof of theorem 4.1. In summary we have the following steps:

1. Fix \( \mu \in C_K \) and consider initial conditions in \( Y_{\mu} \).
2. Consider an expansion of the form: \( u_\epsilon(x, t) = \sum_{n=0}^{m^*} \epsilon^n u^{(n)}(x, t) \).
3. Construct the equation at each order \( k \).
4. Decompose each \( u^{(n)}(x, t) \) using the projection \( \pi_\mu \)

\[
u^{(n)}(x, t) = \pi_\mu(u^{(n)}(x, t)), \quad \phi^{(n)}(x, t) = \langle 1_\mu, u^{(n)}(x, t) \rangle.
\]

5. Construct the hierarchy of equations. For \( n = 0 \) we have

\[
\begin{align*}
\eta^{(0)}(x, t) & = 0 \\
\frac{\partial \phi^{(0)}(x, t)}{\partial t} & = \langle \mu(x), \hat{L}(1_\mu \phi^{(0)}(x, t)) \rangle.
\end{align*}
\]

(40)

Then for \( n \geq 1 \) we get

\[
\begin{align*}
\eta^{(n)}(x, t) & = K^{(n)} \mu \left[ \frac{\partial \eta^{(n-1)}(x, t)}{\partial t} - \hat{L}(\eta^{(n-1)}(x, t) + 1 \phi^{(n-1)}(x, t)) \right] \\
\frac{\partial \phi^{(n)}(x, t)}{\partial t} & = \langle \mu(x), \hat{L}(1_\mu \phi^{(n)}(x, t)) \rangle + \langle \mu(x), \hat{L}(\eta^{(n)}(x, t)) \rangle.
\end{align*}
\]

(41)

6. The evaluation of the remainder of the asymptotic series is then carried out in the same way as in theorem 4.1.
Remark 4.3. It is worth to mention that in systems where \( \mathcal{L}_E^* \) is not identically zero and therefore \( \mathcal{L}^* \) is not diagonal the higher order corrections play a crucial role. In fact there exist systems with different \( \mathcal{L}^* \) operator but same average dynamics. For such systems it is necessary to study also the higher order terms in the \( \epsilon \)-expansion. This class of system will be investigated in a forthcoming paper.

Illustrative example: adiabatic theory and average dynamics

We illustrate the theory by looking finally at the example from enzyme kinetics following the introduction of an IFSS. We have now obtained two different macroscopic limits due to the nature of the IFSS, in sequential order, first the continuum limit, then the adiabatic limit. The result is

\[
\begin{align*}
\xi^{(n)}(x,t) &= (\mathcal{K}_\mu^T)^D \mathcal{I}_\mu \left[ \frac{\partial \xi^{(n-1)}(x,t)}{\partial t} - \hat{\mathcal{L}}^* (\xi^{(n-1)}(x,t) + \mu(x) f^{(n-1)}(x,t)) \right], \\
\frac{\partial f^{(n)}(x,t)}{\partial t} &= \langle 1, \hat{\mathcal{L}}^* (\mu(x) f^{(n)}(x,t)) \rangle + \langle 1, \hat{\mathcal{L}}^* (\xi^{(n)}(x,t)) \rangle,
\end{align*}
\]

(42)

where the invariant measure is

\[
\mu = \begin{pmatrix}
k^-
\vspace{1em}
k^- + ak^+
\end{pmatrix}.
\]

The matrix \( \mathcal{I}_\mu \) id the identity in \( \mathbb{R}^2 \) and

\[
\begin{align*}
f^{(n)}(x,t) &= \rho^{(n)}_0(x,t) + \rho^{(n)}_1(x,t) \quad \text{and} \quad \xi^{(n)}(x,t) = \begin{pmatrix}
\xi^{(n)}_0(x,t) \\
\xi^{(n)}_1(x,t)
\end{pmatrix}.
\end{align*}
\]

Furthermore \( \hat{\mathcal{L}}^* \) becomes

\[
\hat{\mathcal{L}}^* = \begin{pmatrix}
\Delta(\nu \cdot) & 0 \\
0 & \Delta((\nu x - v) \cdot)
\end{pmatrix},
\]

and the infinitesimal generator has the form

\[
\mathcal{K}^T = \begin{pmatrix}
-a k^+ & k^- \\
\vspace{1em}
\vspace{1em}
\vspace{1em}
\vspace{1em}
\vspace{1em}
\vspace{1em}

ak^+ & -k^-
\end{pmatrix}.
\]

Finally the Drazin inverse is
Remark 4.4. The solution of (42) produces the expansion of the probability distribution $P(t, x, a)$. It is useful to observe that the first two terms of the expansions of $\xi$ and $f$ are given by

\[
\begin{align*}
\xi^{(0)}(0) &= 0, \\
\frac{\partial f^{(0)}(x, t)}{\partial t} &= \langle 1_\mu, \hat{L}^*(\mu(x) f^{(0)}(x, t)) \rangle, \\
\xi^{(1)}(x, t) &= (K_T^\mu)^D I_\mu \left[ \mu(x) f^{(0)}(x, t) \right], \\
\frac{\partial f^{(1)}(x, t)}{\partial t} &= \langle 1_\mu, \hat{L}^*(\mu(x) f^{(1)}(x, t)) \rangle + \langle 1_\mu, \hat{L}^*(\xi^{(1)}(x, t)) \rangle,
\end{align*}
\]

generate the a diffusion process whose diffusion coefficient depend on $\epsilon$ and $\delta$ and therefore the time evolution of the concentration $x$ will be dictated by a stochastic differential equation. The construction of this approximation will the subject of a forthcoming paper and is not further considered here.

The above example will be used to derive the Michaelis-Menten and Hill type kinetics known from enzyme kinetics (but also often used in genetics) as a deterministic limit of the probability distribution $P$.

5 Appendix

In this appendix we collect the main property of the geometrical property associated to the Markov chain generator $K$.

5.1 Geometry of the Markov chain

The adiabatic approximation can be carried out by taking advantage of the geometrical structure associated to the Markov chain, i.e. the occurrence of multiple stationary measures. This will be highly relevant in applications where different parts of the Markov chain will be associated to different distinct molecular machines which will be able to exist in different modes of operation. Such a structure will be preserved by the continuum approximation leading to the most important tool to construct the adiabatic approximation of the FPE. For the construction the following definition is of importance:

Definition 5.1 (Drazin inverse). Let $A : \mathbb{R}^g \mapsto \mathbb{R}^g$ be a linear map with $\ker(A) \neq \emptyset$. The Drazin inverse $A^D$ of $A$ is a linear map defined as

\[
A^D = U_A G_A U_A^{-1},
\]
where

(i) $G_A$ is a diagonal matrix with:

- $(G_A)_{ii} = a_i$ if $a_i$ is a non-zero eigenvalue of $A$, and
- $(G_A)_{ii} = 0$ for a 0 eigenvalue of $A$.

(ii) $U_A$ is the matrix whose columns are the eigenvectors of $A$.

The Drazin inverse satisfies the following proposition:

**Proposition 5.1.** If $v \in \ker(A)$ then $v \in \ker(A^D)$.

**Proof.** Indeed since $v$ is a column on $U_A$ we have that $U_A^{-1}v$ is a vector with all zero entries but one corresponding to $v$ in $U_A$. Therefore the definition of $G_A$ implies $G_AU_A^{-1}v = 0$.

The matrix $K(x)$ is an infinitesimal generator of a finite Markov chain for every $x \in \mathbb{R}^N$, whose transpose is $K^T(x)$. Both $K(x)$ and $K^T(x)$ are linear operators acting on $(\mathbb{R}^g, \langle \cdot, \cdot \rangle)$. The geometric structure we are interested in is based on stationary measures:

**Definition 5.2 (Stationary measures).**

$$M_K = \left\{ \mu(x) : K^T(x)\mu(x) = 0, \sum_{i=1}^{g} \mu_i(x) = 1 \right\}$$

We make the following assumption:

(*) $\dim(\ker(K^T(x))) < g$ uniformly in $x$.

A trivial consequence of the definition 5.2 and assumption (*) is:

**Proposition 5.2.** $M_K$ is a linear subspace of $\mathbb{R}^g$ and $m_K = \dim(M_K) = \ker(K^T(x))$.

Let $\{\theta_i\}_{1}^{g}$ be a sequence of real numbers such that $\sum_{m=1}^{m_K} \theta_m = 1$. Then the vector

$$\mu = \sum_{m=1}^{m_K} \theta_m \mu^{(m)} \in M_K,$$

where $\mu^{(m)} \in M_K$.

This motivates the next definition considering convex combinations of stationary measures:

**Definition 5.3 (Convex combinations).** We denote by

$$C_K = \left\{ \mu \in M_K : \mu = \sum_{m=1}^{m_K} \theta_m \mu^{(m)} \text{ with } \sum_{m=1}^{m_K} \theta_m = 1, \ \theta_m \in \mathbb{R}_+ \right\}.$$

the set of convex combinations of stationary measures if an IFSS.
A normalisation of the combination \( \mu \) can be written as

\[
\langle 1 \mu, \mu(x) \rangle = \text{tr} (\mu(n)) = \sum_{k=1}^{g} \mu_k(x) = 1.
\]

In this context it is useful to make an additional definition. First let us introduce \( 1^T \mu \in \mathbb{R}^g \) is given by

\[
(1^T \mu)_i = \begin{cases} 
0 & \text{if } \mu_i = 0 \\
1 & \text{if } \mu_i \neq 0,
\end{cases}
\]

then we define

**Definition 5.4 (Concentrated measures).** Let \( \mu \in C_K \). Let

\[
\mathcal{I}_\mu \doteq \left\{ \rho : \sum_n \text{tr} (\rho(x)) = 1 \text{ and } \sum_n \langle 1^T \mu, \rho(x) \rangle = 1 \right\}.
\]

We call \( \mathcal{I}_\mu \) the set of concentrated measures.

**Remark 5.1.** Note that the set \( \mathcal{I}_\mu \) contains all probability distributions which have the same support as the chosen convex combination of stationary measures \( \mu \).

Now the vector of probabilities can be decomposed in the following way:

**Proposition 5.3.** Given \( \mu \in C_K \), let \( P \in \mathbb{R}^g \). Then \( P \) can be decomposed into

\[
\rho(x) = \xi(x) + f(x) \mu(x),
\]

where

\[
\xi(x) = \Pi_{\mu}(\rho(x)), \quad f(x, t) = 1^T \mu \rho(n).
\]

The function \( f(x) \) is called marginal distribution.

**Proof.** Let us define the operator

\[
\Pi_{\mu} \doteq I_{\mu} - \mu(n) 1^T_{\mu},
\]

where \( I_{\mu} \) is a diagonal matrix such that \( (I_{\mu})_{\sigma \sigma'} = 1 \) if and only if \( \mu_{\sigma} \neq 0 \) otherwise \( (I_{\mu})_{\sigma \sigma'} = 0 \). One can easily verify that

\[
\Pi_{\mu}^2 = \Pi_{\mu}.
\]

From this relation the decomposition of \( \rho(x) \) follows. \( \square \)

The matrix \( K_T(x) \) cannot be inverted because \( \ker(K_T(x)) \neq \emptyset \). Here we need to use the Drazin inverse. The following result holds true (see [22]):
Proposition 5.4. There exists $\left(K^T_{\mu}\right)^D$ such that

$$K^T(x)\Pi_\mu = K^T(x)I_\mu = I_\mu K^T(x), \quad \left(K^T_{\mu}\right)^D K^T(x) = \Pi_\mu$$ \tag{43}

Proof. From proposition [5.1] follows

$$\left(K^T_{\mu}\right)^D \mu(x) = 0.$$

This proves the first relation of (43). For the second relation the reader is refered to [22]. □

As the matrix $K(x)$ is the transpose of $K^T(x)$ we will show that $K(x)$ provides a splitting of maps from $\mathbb{R}^N$ to $\mathbb{R}^g$. More generally we shall now describe how to decompose any map $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^g$. This decomposition will be useful to study the weak form of the FPE. In order to formulate the decomposition we first observe the following simple implication of assumption (⋆):

Proposition 5.5. $\ker(K(n))$ is generated by $\{1_\mu\}_{\mu \in M_K}$.

Using proposition [5.5] one can show that

Proposition 5.6. Every $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^g$ can be decomposed into

$$\Phi(x) = \eta(x) + \phi(x) 1_\mu,$$

where

$$\eta(n) = \pi_\mu(\Phi(x)), \quad \phi(x) = 1_\mu^T \Phi(x).$$

Proof. Let us define:

$$\pi_\mu \doteq I - 1_\mu 1^T$$

Note that

$$\pi_\mu^2 = \pi_\mu.$$

These relations imply that the decomposition holds true. □

Proposition 5.7. There exits $K^D_\mu$ such that

$$K(n)\pi_\mu = K(x), \quad K^D_\mu K(x) = \pi_\mu.$$ \tag{44}

Proof. The proof proceeds as in proposition [5.4] □
Illustrative example: Invariant measure and Drazin inverse

In our illustrative example from enzyme kinetics the MC has infinitesimal generator $K_\delta$. Its transpose is

$$K^T = \begin{pmatrix} -a k^+ & k^- \\ a k^+ & -k^- \end{pmatrix}.$$ 

The invariant measure $\mu$ that satisfies $K^T \mu = 0$ is

$$\mu = \begin{pmatrix} k^- \\ \frac{k^- + a k^+}{a k^-} \\ \frac{a k^+}{k^- + a k^+} \end{pmatrix}.$$ 

Now the matrices $U_K$ and $U_K^{-1}$ are respectively

$$U_K = \begin{pmatrix} k^- & 1 \\ a k^+ & -1 \end{pmatrix} \quad \text{and} \quad U_K^{-1} = \begin{pmatrix} \frac{1}{a k^+ + k^-} & \frac{1}{a k^+ + k^-} \\ \frac{k^-}{a k^+ + k^-} & -\frac{a k^+}{a k^+ + k^-} \end{pmatrix}.$$ 

Now

$$G_K = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{a k^+ + k^-} \end{pmatrix},$$

so the Drazin inverse $(K^T)^D = U_K G_K U_K^{-1}$ is:

$$(K^T)^D = \frac{1}{(a k^+ + k^-)^2} \begin{pmatrix} -a k^+ & k^- \\ a k^+ & -k^- \end{pmatrix}.$$ 

References

[1] Z. Brezezniak, T.Brezezniak, Basic Stochastic Processes, Springer 2003.

[2] M. Doi, Second quantization representation for classical many-particles system, J. Phys. A 9 (1976).

[3] Fall, Marland, Wagner and Tyson Computational Cell Biology, Springer IAM (2003).

[4] D. Gillespie, The chemical Langevin equation, Journal of Chem. Phys 133 n. 1 (2000).

[5] R. Hersh, Random evolutions: a survey of results and problem, Rocky Mountain J. of Math. 4 (1972).
[6] T.B. Kepler and T.C. Elston, *Stochasticity and Transcriptional Regulation: Origin, Consequences, and Mathematical Representation*, Biophysical Journal **81** (2001).

[7] T. Kato, *Perturbation theory for linear operators*, Springer.

[8] T.G. Kurtz, *Solutions of ordinary differential equations as limits of pure jump Markov process* J. Appl. Prob. **7** (1970).

[9] T.G. Kurtz, *Limit theorems for sequences of jump Markov process approximating ordinary differential processes* J. Appl. Prob. **8** (1971).

[10] T.G. Kurtz, *Relationship between stochastic and deterministic models for chemical reactions* J. Chem. Phys. **7** (1972).

[11] T.G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolutions* J. Funct. Analysis **12** (1973).

[12] E. M. Ozbudak, Mukund Thattai, Han N. Lim, B. I. Shraiman and Alexander van Oudenaarden, *Multistability in the lactose utilization network of Escherichia coli*, Nature **427** 2004.

[13] M. Sasai and P. Wolynes, *Stochastic gene expression as a many body problem* PNAS **100** (2003).

[14] P. Deuflhard, *From Molecular Dynamics to Conformational Dynamics in Drug Design*, in: Trends in Nonlinear Analysis, M. Kirkilionis, S. Kromker, R. Rannacher, F. Toni, (ed.) . Springer Verlag, 2003.

[15] C.W. Gardiner, *Handbook of Stochastic Methods*, Springer 2004.

[16] P. Hänggi, *On derivations and solutions of Master Equations and asymptotic representations*, Z. Physik B **30** (1978).

[17] B. Øksendal, *Stochastic differential equations*, Springer 2003.

[18] G.A. Pavliotis and A.M. Stuart, *An introduction to multiscale methods*, (to appear)

[19] A. Pazy *Semigroups of linear operators and applications to partial differential equation* Springer Verlag

[20] Hana El Samad, M. Khammash, L. Petzold and Dan Gillespie, *Stochastic modelling of gene regulatory networks*, Int. J. Robust Nonlinear, **15**:691711 (2005).

[21] M. Renardy, R. Rogers, *An introduction to partial differential equations* TAM 13, Springer.

[22] U. G. Rothblum, *A Representation of the Drazin Inverse and Characterizations of the Index SIAM Journal on Applied Mathematics*, Vol. 31, No. 4. (Dec., 1976), pp. 646-648.
[23] L. Sbano and M. Kirkilionis *Molecular Systems with Infinite and Finite Degrees of Freedom. Part II: Deterministic Dynamics and Examples*, Warwick pre-print 07/2007.

[24] Siegel, I.H. *Enzyme Kinetics*, Wiley 1975.

[25] Keener, J. and Sneyd, J. *Mathematical Physiology*. Springer-Verlag 1998.

[26] M.I. Freidlin and A.D. Wentzel, *Random perturbation of dynamical systems*, Springer 1979.

[27] H.F. Trotter *Approximation of semi-groups of operators* Pacific J. Math 8 887-919, (1958)

[28] N. G. Van Kampen, *Stochastic processes in Physics and Chemistry* (ed 2005), North Holland.

[29] A. D. Wentzell, *A Course in the theory of stochastic processes*, (1981), McGraw Hill.