1 Introduction to chemical reaction networks

This chapter is meant as a gentle introduction to some of the main themes in the course B5.1 *Stochastic modelling of biological processes*. The idea is to build intuition towards some of the detail we will see later.

Many biological models can be formally seen as **chemical reaction networks**, or **CRNs** for short. This includes not just models from biochemistry and physiology, but also models of ecosystems, and epidemiological models for example.

Informally, chemical reactions are processes which cause the **conversion** of some combination of **species** into some other combination of species. We might write

$$\mathrm{A}_1 + \mathrm{A}_2 \to \mathrm{A}_3$$

to mean that a molecule each of species A_1 and A_2 can combine together to produce a molecule of species A_3. A CRN is simply a collection of chemical reactions.

Although we will often refer to the quantities involved as "chemical species", they may equally well refer to:

- Biological species in an ecological model. E.g. $\mathrm{A} + \mathrm{B} \to 2\mathrm{A}$ could caricature the process where A predates on B and this, in turn, allows A to reproduce.
- Subpopulations such as susceptible, infected or recovered individuals in an epidemiological model. E.g., $\mathrm{S} + \mathrm{I} \to 2\mathrm{I}$ could represent a susceptible and infected individual meeting, and the susceptible individual becoming infected.
- Gene expression-levels, protein activation states, and so forth.

To visualise a CRN, we could imagine many different kinds of molecules, or individuals, moving about and colliding; and when this happens there is a chance of a reaction occurring, leading to the creation or destruction of species. If we consider a biological population, the "reaction" could refer to, for example, infection, competition, mutualism, predation, etc.

1.1 Key definitions

Let us introduce some basic concepts which allow us to define and discuss CRNs more formally. Let $\mathcal{S}$ be a set of species which will be, unless stated otherwise, finite. Let $\mathbb{Z}^{\mathcal{S}}_{\geq 0}$ be the set of formal linear combinations of elements of $\mathcal{S}$ with nonnegative integer coefficients. Each element in $\mathbb{Z}^{\mathcal{S}}_{\geq 0}$ is termed a **complex** (on the species of $\mathcal{S}$). The sum of coefficients in a complex is the **molecularity** of the complex. The **zero complex**, denoted 0, is important and will figure frequently. (It is also sometimes called the **empty complex** for obvious reasons.)

Example 1.1 (Complexes). If $\mathcal{S} = \{S_1, S_2, S_3\}$, then $6S_1 + 2S_3$ would be an example of a complex, namely, an element of $\mathbb{Z}^{\mathcal{S}}_{\geq 0}$. The coefficients of this complex would be $(6, 0, 2)$, and their sum, 8, would be its molecularity.

A **reaction** is an ordered pair of complexes $(C_1, C_2) \in \mathbb{Z}^{\mathcal{S}}_{\geq 0} \times \mathbb{Z}^{\mathcal{S}}_{\geq 0}$, which we prefer to write as $C_1 \to C_2$ to remind us that complex C_1 is being converted into the complex C_2. In this reaction, C_1 is the **reactant complex**, and C_2 is the **product complex**. The **order** of the reaction is the molecularity of its reactant complex C_1.

Definition 1.2 (Nonnegative orthant and positive orthant). *The nonnegative orthant in $\mathbb{R}^n$ is defined as $\mathbb{R}^n_{\geq 0} = \{x \in \mathbb{R}^n : x_i \geq 0$ for $i = 1, \ldots, n\}$. The positive orthant in $\mathbb{R}^n$ is defined as $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i > 0$ for $i = 1, \ldots, n\}$. Clearly the nonnegative orthant is the closure of the positive orthant. We define $\mathbb{Z}^n_{\geq 0}$ and $\mathbb{Z}^n_+$ similarly.*

Given a fixed set of n species, it is common practice to identify a complex involving these species with the vector in $\mathbb{Z}^n_{\geq 0}$ of its coefficients, in which case adding and subtracting complexes makes sense: with this convention, $C_2 - C_1$ is termed the **reaction vector** of the reaction. Its components may be positive, negative or zero, and tell us the net *production* or *consumption* of each species in the reaction.

In the example $A_1 + A_2 \to A_3$ above, $A_1 + A_2$ is the reactant complex, A_3 is the product complex; and the reaction vector is $(-1, -1, 1)^t$, assuming that A_1, A_2 and A_3 are the only species involved in this CRN. The reaction is second order.

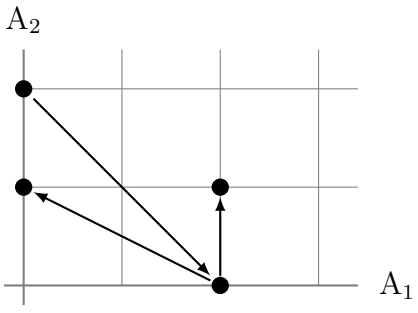
Definition 1.3 (Chemical reaction network). *A CRN is a pair $(\mathcal{S}, \mathcal{R})$ where $\mathcal{S}$ is a finite set of species; and $\mathcal{R} \subseteq \mathbb{Z}^{\mathcal{S}}_{\geq 0} \times \mathbb{Z}^{\mathcal{S}}_{\geq 0}$ is a set of reactions on these species.*

A useful way to visualise a CRN is via its **Euclidean embedded graph**. Given a CRN involving n species, we consider each complex as a node in $\mathbb{Z}^n_{\geq 0}$, and each reaction as a vector between nodes. Picturing an abstract CRN as a set of vectors, each pushing us in a different direction, is immensely useful in both deterministic and stochastic settings.

Example 1.4 **(The Euclidean embedded graph)**. The following CRN has Euclidean embedded graph shown to the right

$$2A_2 \longrightarrow 2A_1, \quad 2A_1 \longrightarrow A_2, \quad 2A_1 \longrightarrow 2A_1 + A_2\,.$$

The reaction vectors tell us in what direction each reaction "pushes" the state of the system when it fires.



Consider a CRN with n species and m reactions, and suppose we choose an ordering on the species and reactions. Let $\zeta_1, \ldots, \zeta_n$ be the reaction vectors of the CRN. Then the matrix $\Gamma = [\zeta_1 \,|\, \zeta_2 \,|\, \cdots \,|\, \zeta_m]$ with reaction vectors as columns is the **stoichiometric matrix** of the network. The **stoichiometric classes** of the CRN are the intersections of cosets of $\operatorname{im}\Gamma$ and $\mathbb{R}^n_{\geq 0}$.

It is often useful to treat processes as chemical reactions, even when they are, strictly speaking, not. Let's think of some examples:

- We've already seen that in an epidemiological model we may write $\mathrm{S}+\mathrm{I}\to 2\mathrm{I}$ for the process where a susceptible individual S interacts with an infected individual I and (with some probability) the susceptible individual becomes infected.
- We may treat a single type of molecule as two different species in two different compartments; and then treat its movement from one compartment to another as a chemical reaction. In this case we might write $\mathrm{A}_\mathrm{i}\rightleftharpoons\mathrm{A}_\mathrm{j}$ where A_i refers to the species A in compartment i and A_j refers to the same species but in compartment j.
- If a chemical species A flows into the compartment of interest to us, or is created from some other species which is abundant and we do not wish to model, then we might write $0\to\mathrm{A}$ for this process. Similarly, we can write $\mathrm{A}\to 0$ for the degradation of A into species we are uninterested in; or for the outflow of A from the compartment of interest.
- We will often be interested in reactions such as $\mathrm{A}\to 2\mathrm{A}$; again, we appear to be creating a molecule of A from nowhere; but we should think of this reaction as involving a *hidden* species which we are omitting from our model.

Reaction networks are often associated with **conserved quantities**. For example, if we have the single reaction $\mathrm{A}\rightleftharpoons\mathrm{B}$, then we see that the total number of molecules of A and B is conserved. Given a CRN with stoichiometric matrix Γ, we will see that for each element of $\ker\Gamma^\mathrm{t}$ we have a linear conservation law associated with any deterministic or stochastic model of the network.

Example 1.5 (**An epidemiological model)**. Consider a *Susceptible-Infectious-Recovered-Vaccinated* model:

$$\mathrm{S}+\mathrm{I}\to 2\mathrm{I},\quad \mathrm{I}\to\mathrm{R},\quad \mathrm{S}\to\mathrm{V}\,.$$

The meanings of the three "reactions" should be clear. The network has stoichiometric matrix

$$\begin{array}{c} \\ \mathrm{S} \\ \mathrm{I} \\ \mathrm{R} \\ \mathrm{V} \end{array}\begin{array}{c} \begin{array}{ccc} \mathrm{S}+\mathrm{I}\to 2\mathrm{I} & \mathrm{I}\to\mathrm{R} & \mathrm{S}\to\mathrm{V} \end{array} \\ \begin{pmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

It is straightforward to see that the total number of individuals $\mathrm{S}+\mathrm{I}+\mathrm{R}+\mathrm{V}$ is a conserved quantity. This is equivalent to the observation that the vector $(1,1,1,1)^\mathrm{t}$ is orthogonal to each reaction vector.

1.2 Modelling CRNs

Models of CRNs explore the evolution of **numbers** or **concentrations** of species in **time** and, for more complicated models, in **space**.

Deterministic and stochastic models. Depending on the assumptions we make, we can get either **deterministic** models involving ordinary differential equations (ODEs) or partial differential equations (PDEs); or **stochastic** models in which the evolution follows probabilistic rules. It is stochastic models which are the focus of this course, but we are often interested in comparing their behaviour to the behaviour of deterministic models, so we need to understand both.

Kinetics. Apart from knowing the actual reactions, in order to build models of CRNs, we need rules telling us the *kinetics*: the rates – or "probabilities per unit time" – of occurrence of each reaction in a network. These rates will, in general, depend on the concentrations or numbers of species present at each particular moment.

Some choices we make for the kinetics follow very naturally from physical assumptions. For example, in the stochastic case, we will often make the **time-homogeneous, Markov** assumption, roughly, that reaction rates depend on the current state of the system, but not its previous history, and not on external factors. It is also often natural to assume that the rate of a given reaction depends only the concentrations/numbers of molecules appearing in its reactant complex, as these are the molecules that need to be present in order for the reaction to occur.

Remark 1.6 (**Differential equations to study stochastic models)**. We will frequently write down and work with differential equations describing the evolution of probabilities or probability density functions (e.g., **master equations**, the **Kolmogorov backward equation**, the **Fokker-Planck equation**, etc.). We can think of studying such equations as using deterministic tools, namely, differential equations, to help us understand a stochastic process. It should not be confused with deterministic modelling.

Remark 1.7 (**Deterministic to stochastic and back again)**. The nature of the models we study depends fundamentally on the scale at which we observe a system. For example, at the very detailed scale, we might consider studying *deterministic* models, where we track each individual molecule in space and time as it moves, collides with other particles, reacts with some, etc. The complexity rapidly becomes immense, and it is natural to simplify the picture by moving to *stochastic* models: we do not try to follow all the particles; rather the state of a particle colliding with millions of others is treated as a collection of random variables – a stochastic process. If we increase the scale still further, we no longer focus on individual particles, but rather on concentrations, gradients, etc., which we hope to study using differential equations: we end up back again with *deterministic* models.

Writing down a model of a CRN involves specifying:

1. The CRN itself, i.e., the set of species and reactions involved;
2. The modelling methodology (e.g. deterministic or stochastic);
3. The kinetics, i.e., the rates or probabilities of reactions occurring;
4. The setting: are we allowing diffusion, or other physical/environmental processes as part of the model.

Time. We also need to specify how we are treating *time*: for example, time could proceed in discrete steps, or could be continuous. In this course, we will focus on continuous time;

but sometimes we will look at "time-discretised" versions of the models we write down for the purposes of simulation.

Analysis and computation. Intelligently approaching CRNs often involves a mixture of analysis and computation. In both deterministic and stochastic cases, there are deep and difficult questions connected with how best to **simulate** models; and we will be thinking of some of these questions in the stochastic case during this course. In fact, sometimes we will do considerable work in order to arrive at formulae we can't hope to solve, but have to approach numerically!

Model outputs. What are we interested in getting out from models of CRNs? We may want to consider short-term or **transient** behaviour; or focus only on long-term or **steady state** behaviour. In deterministic models, long-term behaviour corresponds to **limit sets** of the associated dynamical systems. In stochastic models, long-term behaviour can correspond, for example, to **stationary measures (or distributions)** (if these exist).

In stochastic models, only in the simplest cases can we follow all probabilities for all time. Often we have to make do with finding stationary distributions; or tracking moments of distributions over time; or constructing algorithms to approximate the quantities we want to calculate.

1.3 Deterministic models of reaction networks

In these models, chemical concentrations are taken to be nonnegative real numbers which vary in time and, possibly, space. **ODE models** arise from the assumption that the system is "well-mixed", and so chemical concentrations do not vary in space. **PDE models** allow chemical concentrations to vary in space, for example via diffusion or advection, and so are inherently more complicated than ODE models.

In a CRN involving n species, the **state of the system** is a vector in $\mathbb{R}^n_{\geq 0}$: its kth component is the concentration of the kth species. We expect this concentration vector to vary in time (ODE case); or time and/or space (PDE case). The basic assumption behind these deterministic models is that there are so many molecules of every species in any given "region" that we can effectively ignore small fluctuations, and stochastic effects. We will explore this assumption in some detail in this course, and see how it can break down.

Remark 1.8. Even in the case of ODE models of CRNs, there is still a *lot* that is not known. Many exciting theorems have been proved about ODE models of CRNs in the last twenty years, but there are still some big open conjectures.

One common choice of kinetics is **mass action kinetics** which arises, roughly speaking, from the assumption that the probability of a reaction occurring is proportional to the probability of the molecules in its reactant complex "meeting" each other. Both deterministic mass action kinetics and the stochastic version, **stochastic mass action kinetics**, will be discussed further below.

Remark 1.9 (**Kinetics other than mass action)**. Mass action is by no means the only possible choice of reaction kinetics for either deterministic or stochastic models. Even in the deterministic case, there are sometimes very good reasons why we might expect reaction rates to deviate from mass action.

Autonomous and nonautonomous models. While the simplest models give rise to autonomous systems of differential equations (i.e., there is no explicit dependence on time), we could also allow reaction rates to depend explicitly on time. You can think, for example, of a container which is gradually being heated over time, or whose volume is changing periodically, altering the rate at which reactions occur.

1.3.1 ODE models of CRNs

Consider a system of m chemical reactions on n species $X_1, \ldots, X_n$, evolving in a well-mixed container. Let x_i denote the concentration of species X_i, and let $\mathbf{x}$ denote the vector $(x_1, \ldots, x_n)^t$, dependent on time t. Let $v_j(\mathbf{x})$ be the rate of the jth reaction, and let $\boldsymbol{v}(\mathbf{x})$ denote the vector of reaction rates. Finally, let $\zeta_1, \ldots, \zeta_m$ be the reaction vectors of the reactions involved, and let

$$\Gamma = [\zeta_1 \,|\, \zeta_2 \,|\, \cdots \,|\, \zeta_m]$$

be the $n \times m$ stoichiometric matrix obtained by writing the reaction vectors as the columns of a matrix. Then we can write down the ODE governing the evolution of the species concentrations (on $\mathbb{R}^n_{\geq 0}$ or $\mathbb{R}^n_{>0}$ for example) as

$$\dot{\mathbf{x}} = v_1(\mathbf{x})\zeta_1 + \ldots + v_m(\mathbf{x})\zeta_m \,,$$

or, more briefly, as

$$\dot{\mathbf{x}} = \Gamma \boldsymbol{v}(\mathbf{x}) \,. \tag{1}$$

This form is quite general and makes no assumptions about kinetics. But we have assumed that reaction rates do not explicitly depend on time. We also always assume that $\boldsymbol{v}$ is sufficiently well-behaved to guarantee existence and uniqueness of solutions, for example, locally Lipschitz.

We could discretise (1) to get

$$\mathbf{x}(t + \Delta t) \simeq \mathbf{x}(t) + \Gamma \boldsymbol{v}(\mathbf{x}(t))\Delta t$$

which would form the natural starting point for numerical schemes to integrate the ODE. Given an initial condition at time 0, we could also naturally recast (1) as an integral equation

$$\mathbf{x}(t) = \mathbf{x}(0) + \Gamma \left[\int_0^t \boldsymbol{v}(\mathbf{x}(s)) ds \right] , \tag{2}$$

which could also be written

$$\mathbf{x}(t) = \mathbf{x}(0) + \left[\int_0^t v_1(\mathbf{x}(s)) ds \right] \zeta_1 + \left[\int_0^t v_2(\mathbf{x}(s)) ds \right] \zeta_2 + \cdots + \left[\int_0^t v_m(\mathbf{x}(s)) ds \right] \zeta_m \,. \tag{3}$$

Later on, we will see stochastic analogues of all of these equations, and it will be helpful to look back to compare and contrast the determistic and stochastic evolutions equations.

From (1) we immediately see, for example, that if $\mathbf{p}$ is any vector in $\ker \Gamma^t$, then along any trajectory of the system

$$\frac{d}{dt} \mathbf{p}^t \mathbf{x} = \mathbf{p}^t \Gamma \boldsymbol{v}(\mathbf{x}) = 0,$$

i.e., $\mathbf{p}^t\mathbf{x}$ is a conserved (linear) quantity. This justifies our earlier claim that each element of $\ker \Gamma^t$ corresponds to a linear conservation law; we will see that this is also true in the stochastic case.

Exercise 1.10 (**Necessary and sufficient conditions for bounded stoichiometric classes)**. We may assume the following result which is a straightforward corollary of Farkas' Lemma in convex geometry (it is sometimes called "Gordan's alternative"):

> Let S be a linear subspace of $\mathbb{R}^n$, and $S^\perp$ the orthogonal complement of S. Then one of the following two occurs:
>
> - $S \cap \mathbb{R}^n_{\geq 0} = \{0\}$; or
> - $S^\perp \cap \mathbb{R}^n_+ \neq \emptyset$.

With the help of this result, prove the following claim:

Claim. The following are equivalent for a CRN with stoichiometric matrix Γ. (i) There exists a positive vector $\mathbf{p} \in \ker \Gamma^t$; (ii) all stoichiometric classes of the CRN are bounded. (It may help to note from basic linear algebra that $\operatorname{im} \Gamma$ and $\ker \Gamma^t$ are orthogonal subspaces of $\mathbb{R}^n$.)

Except in the simplest cases, we would not expect to be able to solve (1) explicitly. However, the goal would often be to carry out a qualitative analysis of its solutions and find out, for example, the number and stability of steady states; whether periodic solutions are possible; whether solutions can become unbounded or converge to the boundary of the state space, etc.

1.3.2 Deterministic mass action kinetics

Let C_1 and C_2 be complexes and consider the reaction $C_1 \to C_2$. Deterministic mass action kinetics is the assumption that the rate of this reaction is proportional to the concentration of each species in C_1 to the power of its stoichiometry in C_1. The constant of proportionality is termed the **rate constant** of the reaction. For example, a reaction with reactant complex

$$A + B$$

would proceed at rate kab where k is the rate constant, a is the concentration of A, and b is the concentration of B. Similarly, a reaction with reactant complex

$$2A + B$$

would proceed at rate ka^2b. More generally, a reaction with reactant complex

$$\alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_n A_n\,,$$

with mass action kinetics has rate

$$k\, a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_n^{\alpha_n}\,,$$

where a_i is the concentration of A_i, and k is the rate constant of the reaction.

The assumption of mass action comes from considering the probabilities of molecules meeting in a "well-mixed" environment. We will see this in more detail when we examine stochastic mass action kinetics later.

1.3.3 PDE models of CRNs

ODE models of CRNs arise from the assumption that the reactions are occurring in a well-mixed compartment – the concentrations of species depend on time, but do not vary in space. If we cannot assume this, then we need to consider concentrations depending on both time and *space*. We might, for example, introduce diffusion into a CRN model, to get a **reaction-diffusion** model. In this case we obtain, for the evolution of the concentration vector $\boldsymbol{x}$ in time and space:

$$\frac{\partial \boldsymbol{x}}{\partial t} = \Gamma \boldsymbol{v}(\boldsymbol{x}) + \mathbf{D}\nabla^2 \boldsymbol{x}\,, \tag{4}$$

where D is a diagonal matrix of diffusion coefficients, and $\nabla^2 \boldsymbol{x}$ is the vector obtained by the operation of the Laplacian on each component of $\boldsymbol{x}$, namely $(\nabla^2 x_1, \ldots, \nabla^2 x_n)^t$. Note that in (4) $\boldsymbol{x}$ is the vector of chemical concentrations which depends on time and a spatial variable whose name we have not written down.

Reaction-diffusion equations give rise to fundamentally new phenomena, such as spatio-temporal patterns (e.g., travelling waves), and diffusion-driven instability, leading to the formation of complex spatial patterns. Spatial models have been most intensively used in biology to help understand the process of biological pattern formation.

Apart from diffusion, we may also wish to introduce other transport processes, such as advection, into spatial models. In this case we would need further space-dependent terms in (4), involving partial derivatives of the concentration vector $\boldsymbol{x}$ with respect to the spatial variables.

Understanding spatial CRN models in the deterministic setting allows us to write down the stochastic analogues of these models.

1.4 Introduction to stochastic models of CRNs

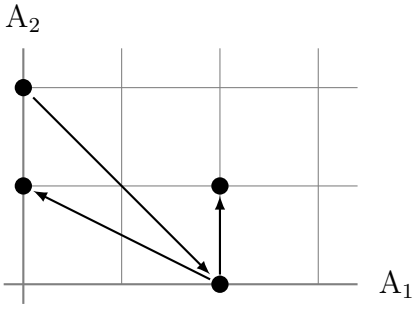
The state space. While we track species *concentrations* in differential equation models of CRNs, we track species *numbers* in stochastic models. Consequently, while the state space of an ODE model of a CRN on n chemical species is some subset of $\mathbb{R}^n_{\geq 0}$, the state space of the corresponding stochastic model will be some subset of $\mathbb{Z}^n_{\geq 0}$. The state of the system at time t, which we will denote by $\mathbf{X}(t)$, is now a random variable taking values in (some subset of) $\mathbb{Z}^n_{\geq 0}$. Its jth component is the number of molecules of species j present at time t.

Stochastic processes. The collection of random variables $\mathbf{X}(t)$ (one for each t in some time domain $T \subseteq \mathbb{R}$) defines a stochastic process. The abstract probability space which forms the common domain of all the $X(t)$ in general isn't specified, but it can be helpful to consider it as the (function) space of all **sample paths**, with some appropriate σ-algebra and probability measure. When we carry out stochastic simulations of CRNs, it is often sample paths that we

compute. Knowing the distribution of $\mathbf{X}(t)$ for each t, conditional on some initial distribution, can be regarded as knowing the "solution" of our stochastic model. But sometimes we may also be interested in features of sample paths not necessarily visible in the distributions of $\mathbf{X}(t)$, e.g., if there is some "periodic" behaviour.

Example 1.11 (**Stochastic models and the Euclidean embedded graph)**. Recall Example 1.4 where we plotted the Euclidean embedded graph of the CRN

$$2A_2 \longrightarrow 2A_1, \quad 2A_1 \longrightarrow A_2, \quad 2A_1 \longrightarrow 2A_1 + A_2\,.$$



This picture helps us understand the evolution: given an initial point in $\mathbb{Z}^2_{\geq 0}$, whenever a reaction fires, we move along the corresponding reaction vector to a new grid point. We can thus visualise the evolution as a walk on $\mathbb{Z}^2_{\geq 0}$ (in fact a proper subset of $\mathbb{Z}^2_{\geq 0}$ in this case – can you see why?). This walk has an element of randomness, because

- We don't know exactly when a reaction will fire.
- We don't know which reaction will fire first.

Nevertheless, the walk is constrained to move in one of at most three possible directions at each step. We already have, in this example, the intuition for how an algorithm to simulate the evolution of a CRN should proceed.

Reaction intensities. Associated with any state, say x, and any reaction, say the jth reaction, we have a **stochastic reaction rate**, or **intensity**, or **propensity** say $\nu_j(x)$. (You will see all three of these terms in these notes, the course text, and other texts; but please remember that they mean the same thing!) We will give a precise meaning to $\nu_j(x)$ later, but note for now that the probability of reaction j occurring in a time interval of length Δt is equal to $\nu_j(x)\Delta t + o(\Delta t)$, conditional on the system remaining in state x until the reaction occurs. We will justify and more fully discuss this claim later.

Intensities, being probabilities per unit time, are always nonnegative; but they need not be less than 1. We will consider how we might set intensities later on; but for the moment we make only the assumption that:

> If there are sufficient molecules of all the species in the reactant complex of a reaction for the reaction to proceed, then its intensity is positive. If there are insufficient molecules for the reaction to proceed, then its intensity is zero.

It is also useful (to simplify various calculations) to formally set the intensity associated with any unphysical state (e.g., a state with some negative components) to be zero.

The following two examples demonstrate that the state space of a stochastic model of a CRN may depend on the initial state, and may be finite or infinite.

Example 1.12 (**Finite state space)**. Consider the CRN consisting of a single, reversible, **dimerisation** reaction, $2\mathrm{A} \rightleftharpoons \mathrm{B}$ (this is an abbreviation for the two reactions $2\mathrm{A} \to \mathrm{B}$ and $\mathrm{B} \to 2\mathrm{A}$). Let the initial state consist of 3 molecules of B and one molecule of A. With our basic assumptions about intensities above, the state space consists of the four points $\{(1,3),(3,2),(5,1),(7,0)\}$ in $\mathbb{Z}^2_{\geq 0}$, where the first entry in each vector is the number of molecules of A and the second is the number of molecules of B.

Example 1.13 (**Infinite state space)**. Consider the CRN consisting of two reactions involving a single species $0 \rightleftharpoons 2\mathrm{A}$ (namely, $0 \to 2\mathrm{A}$, $2\mathrm{A} \to 0$). With our basic assumptions about intensities,

- if the initial number of molecules is even, the state space consists of all the nonnegative, even integers;
- if the initial number of molecules is odd, then the state space consists of all nonnegative, odd integers.

The definition of reaction vectors and the stoichiometric matrix are exactly as in the deterministic case. As an immediate consequence of the result in Exercise 1.10, we also have the following result:

Exercise 1.14 (**Necessary and sufficient conditions for a finite state space)**. Consider a CRN with stoichiometric matrix Γ, and suppose that there exists a positive vector $\mathbf{p} \in \ker \Gamma^{\mathrm{t}}$. Prove that (regardless of the initial state) the state space of any stochastic model of the CRN is finite. [Hint: Use the result of Exercise 1.10.]

Remark 1.15. Note that the state space of a stochastic model may be finite even if stoichiometric classes of the deterministic system are unbounded. For example, given the single reaction $\mathrm{A} \to 0$, with n molecules of A initially, the state space is $\{0, \ldots, n\}$, where these numbers denote the number of molecules of A.

Remark 1.16 (**Explicit time dependence)**. Although we generally assume time-homogeneity, just as in the deterministic case, we might also want to consider "nonautonomous" evolution, where the reaction intensities depend explicitly on time, and not just on the state of the system. In this case we get stochastic processes which are not time-homogeneous.

1.5 Comparing deterministic and stochastic models

We will frequently be interested in comparing the outputs of deterministic and stochastic models. If we have a fixed volume in which the evolution takes place, we can easily pass back and forth between numbers and concentrations; but we do have to be a bit careful when comparing intensities with deterministic reaction rates. We will discuss more fully the relationship between deterministic and stochastic mass action kinetics later. This will also allow us to examine examples where the deterministic model loses some key information in the stochastic model.

One conclusion we may hope will sometime hold is that the *mean* behaviour of a stochastic model matches the behaviour of the corresponding deterministic model. We will see situations where this is the case, and others where it is not.

We will also see that there are many questions we can ask about stochastic models that we *cannot even ask* about a deterministic model.

Finally, we will also see that there are some behaviours which occur in stochastic models which are quite different to anything we might see in a deterministic model. Thus studying stochastic models opens up a world of interesting behaviours going beyond the – already exotic – range of behaviours seen in deterministic models.