Introduction to Chemical Reaction Network theory

Seokhwan Moon

May 30, 2024

Abstract

During the semester, we were mainly focused on the population dynamics in mathematical biology in class. In this report, I will focus on molecular-level biochemical reactions for mathematical biology. There are two ways to model molecular-level biochemical interactions - one using ODE to describe concentrations, and other using continuous-time Markov chain to describe number of molecules. The former description is easy to analyze, but hard to describe the randomness of molecular-level interactions. On the other hand, the latter description could describe the randomness of molecular-level interactions, but hard to analyze. In this report, I will introduce both deterministic and stochastic chemical reaction network theory, and highlight the important research topics and results for this area.

1 Introduction

In recent decades, the synthetic biology has made significant strides. From the viewpoint of systems biology, the biological objects have been conceptualized as 'machines', and the technology for constructing and manipulating these 'machines' has gained importance. Today, we could construct artificial living entities with synthetic biochemical reactions. These advancements are applied across various areas such as programming language using bio-computing, creating a new organ, etc. Given these technological backdrop, it is important to construct biochemical reactions in order to realize desired properties.

Chemical Reaction Network (CRN) theory is a subarea of mathematical biology connecting mathematics and systems biology. The main goal of this area is to know about the dynamical properties of biochemical reactions only from its network structure alone. By knowing this, we can easily get how to construct biochemical reactions to obtain typical dynamical properties.

In this report, I will give a brief introduction about CRN. This report will not discuss about the detailed mathematical proofs of the theorem. Besides, we will look for how biochemical reaction can be mathematically modeled, and which research topics are being discussed in recent years.

2 Deterministic chemical reaction network

2.1 Backgrounds

Starting from the toy example, let's think about the following chemical reaction

$$2H_2 + O_2 \to 2H_2O$$

Here, $H_2,\ O_2,\ H_2O$ are molecules, which is the basic unit for biochemical interactions. In CRN, we call them **species**. So, the above chemical reaction can be written as

$$2A + B \to 2C$$

And when we think the reaction as a directed edge, what corresponds to the vertex is called a **complex**. So generally speaking, CRN is described with three sets : **species** $\mathscr{S}$, **complexes** $\mathscr{C}$, and **reactions** $\mathscr{R}$ where

- $\mathscr{S}$: Chemical species that are the basic unit of biochemical interactions (i.e. molecules, proteins, etc)
- $\mathscr{C}$: Non-negative linear combination of species that reactions interact (reactant and products)
- $\mathscr{R}$: Reactions that one complex is converted to another.

For example, if we think about the following set of reactions,

$$\begin{aligned} A + B &\to 2B \\ B &\to A \\ 2B &\to A + B \end{aligned}$$

Then the sets are

$$\begin{aligned} \mathscr{S} &= \{A,\ B\} \\ \mathscr{C} &= \{A + B,\ 2B,\ B,\ A\} \\ \mathscr{R} &= \{A + B \to 2B,\ B \to A,\ 2B \to A + B\} \end{aligned}$$

And when we graphically describe this reactions, we do not make one complex to appear several times. i.e., it can be graphically described as

$$\begin{aligned} A + B &\leftrightharpoons 2B \\ B &\to A \end{aligned}$$

In deterministic CRN, we model the **concentration** of each chemical species. For example, when we denote r_1 the rate of the reaction $A + B \to 2B$,

$$\begin{aligned} \frac{d[A]}{dt} &= -r_1 \\ \frac{d[B]}{dt} &= r_1 \end{aligned}$$

since $A + B \to 2B$ decreases one A and increases one B. If we write the reaction rate for other two reactions as r_2 and r_3, then the full equations becomes

$$\begin{aligned} \frac{d[A]}{dt} &= -r_1 + r_2 + r_3 \\ \frac{d[B]}{dt} &= r_1 - r_2 - r_3 \end{aligned}$$

rewriting this becomes

$$\frac{d}{dt}\begin{bmatrix}[A] \\ [B]\end{bmatrix} = \begin{bmatrix}-1 & 1 & 1 \\ 1 & -1 & -1\end{bmatrix}\begin{bmatrix}r_1 \\ r_2 \\ r_3\end{bmatrix} \tag{1}$$

we can notice that each column of the matrix corresponds to the net change of chemical species for each reactions. We call each vectors **stoichiometric vector**. In general,

$$c_1X_1 + c_2X_2 + \cdots + c_mX_m \to c_1'X_1 + c_2'X_2 + \cdots + c_m'X_m \ \ : \ \ (c_1' - c_1,\ \cdots,\ c_m' - c_m)^T$$

Also, we call the matrix whose column consists of stiochiometric vector as **stoichiometric matrix**.

One might ask how can we determine the rate r_1, r_2, r_3. The most general choice is **mass-action kinetics**, which defines the reaction rate as

$$c_1X_1 + c_2X_2 + \cdots + c_mX_m \to \cdots \ \ : \ k[X_1]^{c_1}[X_2]^{c_2}\cdots[X_m]^{c_m}$$

but this is not the only choice. There are other kinetic laws such as **Michaelis–Menten kinetics**, **Power-law kinetics**, etc. The only rule is that the kinetic law is the function only from its reactants, i.e.,

$$c_1X_1 + c_2X_2 + \cdots + c_mX_m \to \cdots \ \ : \ f(k, [X_1], [X_2], \cdots [X_m])$$

and there is one parameter k, which is determined by many environmental factors such as the activity of enzyme, temperature, volume, other conditions, etc.

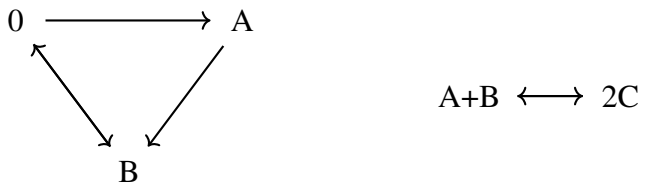
2.2 Deficiency Zero theorem : Existence of stable steady-state from the structure

Definition 1. Suppose that we represent the CRN by the way that no complexes appear twice. We say the CRN is **weakly reversible** when we pick any two complex c_1, c_2 such that there exists reaction $c_1 \to c_2$, then there always exists the path from c_2 to c_1 connected by some reactions.

For example, consider the following CRN.

$$\begin{aligned} 0 &\xrightarrow{k_1} A \\ 0 &\xrightarrow{k_2} B \\ A &\xrightarrow{k_3} B \\ B &\xrightarrow{k_4} 0 \\ A + B &\xrightarrow{k_5} 2C \\ 2C &\xrightarrow{k_6} A + B \end{aligned} \tag{2}$$

then we can represent the CRN as the following



Then, we can check that this CRN is weakly reversible. For example, there is a reaction $0 \to A$, then there is a path of reactions $A \to B \to 0$.

Definition 2. The **stiochiometry space** S is the vector space spanned by the stoichiometric vectors. And the **stoichoimetric compatibility class** is $S + s_0$ where s_0 denotes the initial concentrations.

Definition 3. The **deficiency** δ of the CRN is defined as

$$\delta = n - \ell - s$$

where n is the number of complexes, ℓ is the number of linkage class, and $s = dim(S)$

For example, when we calculate the deficiency of the previous example (2),

- n = 5
- ℓ = 2
- $s = dim(span\{(1,0,0),(0,1,0),(0,-1,0),(-1,1,0),(-1,-1,2),(1,1,-2)\})$ = 3

therefore $\delta = 0$.

Theorem 1. [1, 2] Consider a chemical reaction system following mass-action kinetics. Then, for all choice of rate constants, the system has exactly one equilibrium concentration in each positive stoichiometric compatibility class and that equilibrium concentration is locally asymptotically stable.

By applying the theorem 1, we can know that when we assume CRN (2) follows the mass-action kinetics, then it has the unique equilibrium which is locally asymptotically stable. This deficiency zero theorem is a fundamental theorem for constructing arbitrary CRN, because the convergence for the equilibrium is important for every biochemical reaction systems. We could easily imagine that when the equilibrium is unstable or does not exist, then the concentration of some chemical species will blow up or becomes zero, which will cause serious problem for biological objects.

One may ask, since the equilibrium point unique, *is it globally asymptotically stable*? This question is called **Global attractor conjecture**, and still left unknown. Prof. Gheorghe Craciun from the University of Wisconsin-Madison claims that he proved it, but it is not published yet [3].

2.3 Absolute Concentration Robustness : Properties of steady-state from the structure

The story for this topic is quite interesting. Someday, a group of biologist were doing some experiment using *E. coli*. They put some amount of chemical species, and measured the concentration after very long time. They repeated this experiment by changing the initial amount of chemical species. But the problem was that, even though they change the initial amount, the concentration of some species after long time does not changed. Biologists thought that they were doing something wrong with the experiment, but it turns out that it was not the problem of experiment.

Starting with the toy example, let's think about the following reaction network

$$A + B \xrightarrow{\alpha} 2B$$
$$B \xrightarrow{\beta} A$$

then,

$$\frac{d[A]}{dt} = -\alpha[A][B] + \beta[B]$$
$$\frac{d[B]}{dt} = \alpha[A][B] - \beta[B]$$

Then at the steady-state, $\frac{d[A]}{dt} = \frac{d[B]}{dt} = 0$. therefore, the unique steady-state is

$$[\bar{A}] = \frac{\beta}{\alpha}$$
$$[\bar{B}] = \Theta - [\bar{A}]$$

where $\Theta = [A]_0 + [B]_0$. i.e., no matter what initial values are, the concentration of A converges to the same value. For this case, we say that this systems has **absolute concentration robustness** for A.

Theorem 2. [4] Consider a chemical reaction system following mass-action kinetics and admits a positive steady-state. Suppose the deficiency of this reaction network is one. If, in the network, there are two non-terminal nodes that differ only in species S, then the system has absolute concentration robustness in S.

The biological example to apply this theorem is the *E. Coli* EnvZ-OmpR system.

$$\begin{aligned} X \underset{k_{-1}}{\overset{k_1[T]}{\rightleftarrows}} XT \xrightarrow{k_2} X_p \\ X_p + Y \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} X_pY \xrightarrow{k_4} X + X_p \\ XT + Y_p \underset{k_{-5}}{\overset{k_5}{\rightleftarrows}} XTY_p \xrightarrow{k_6} XT + Y \end{aligned} \tag{3}$$

Calculating the deficiency of CRN 3 gives

- n = 9
- ℓ = 3
- s = 5

gives $\delta = 9 - 3 - 5 = 1$. Also, the blue complexes are non-terminal nodes and the red complexes are terminal nodes. The node XT and $XT + Y_p$ are the non-terminal nodes that differ only one species Y_p. Therefore, by the theorem 2, the system has absolute concentration robustness in Y_p, meaning that the steady-state concentration of Y_p is not affected by the initial concentrations.

Going back to the first story, it turns out that the biologist were not doing anything wrong. Instead, the structure of biochemical reaction network itself has such properties. This is the background story of how the reference paper [4] could be published in *Science*.

2.4 Law of localization : Biological meaning from the structure

Starting from the toy example, let's think about the following dynamical system

$$\begin{aligned}\frac{dx_1(t)}{dt} &= x_2(t) + u(t)\\ \frac{dx_2(t)}{dt} &= -2x_1(t) - 3x_2(t)\\ y(t) &= 2x_2(t)\end{aligned} \tag{4}$$

Let's think that $u(t)$ is the environmental factor for this dynamical system, and $y(t)$ is the output of the system. Calculating the steady-state, we get

$$\begin{aligned}\bar{x_2} &= -u(t)\\ \bar{x_1} &= -\frac{3}{2}\bar{x_2} = \frac{3}{2}u(t)\\ \bar{y} &= -2u(t)\end{aligned}$$

i.e., the steady-state value of y is affected by the environmental factor $u(t)$. But when thinking about the following system,

$$\begin{aligned}\frac{dx_1(t)}{dt} &= -2x_1(t) + 2u(t)\\ \frac{dx_2(t)}{dt} &= -x_2(t) + 2u(t)\\ y(t) &= 2x_1(t) - x_2(t)\end{aligned} \tag{5}$$

then the steady-state values are

$$\begin{aligned}\bar{x_2} &= 2u(t)\\ \bar{x_1} &= u(t)\\ \bar{y} &= 2 \cdot u(t) - 2u(t) = 0\end{aligned}$$

i.e., the steady-state value of y is independent from the environmental perturbation $u(t)$. But notice that this independence is highly dependent on the parameters. i.e., when the dynamic equation for y changes as $y(t) = 3x_1(t) - x_2(t)$, then the independence is no longer satisfied.

For the last toy example, let's consider the following dynamical system.

$$\begin{aligned}\frac{dx_1(t)}{dt} &= x_2(t)\\ \frac{dx_2(t)}{dt} &= -2x_1(t) - 3x_2(t) + u(t)\\ y(t) &= 2x_2(t)\end{aligned} \tag{6}$$

then the steady-state values are

$$\begin{aligned}\bar{x_2} &= 0\\ \bar{x_1} &= \frac{1}{2}u(t)\\ \bar{y} &= 0\end{aligned}$$

i.e., the steady-state value of y is independent from $u(t)$. Notice that this independence is also independent from the choice of parameters.

So, for the (4), we say that the dynamical system has **no perfect adaptation**. For the (5), we say the dynamical system has **Finely-tuned perfect adaptation**, meaning that such perfect adaptation can be only achieved when the parameters are finely tuned. And for the (6), we say the dynamical system has **Robust perfect adaptation (RPA)**, meaning that such perfect adaptation is robust from the choice of parameters. Notice that the absolute concentration robustness and RPA are different mathematical topics. Absolute concentration

robustness is about maintaining same steady-state value regardless of the different initial concentrations, and RPA is about maintaining same steady-state value from the environmental perturbation.

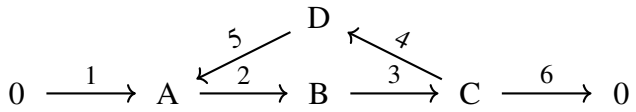
RPA is not only mathematically interesting, but also biologically important. As mentioned before, the parameters for each reaction depends on environmental conditions such as the activity of enzyme or the temperature. Despite those environmental perturbation, biological object maintains the homeostasis of certain important factors. So, the question that 'When does the RPA take place?' could discover the origin of homeostasis in biology.

Theorem 3. [5] Suppose we are representing the CRN by hypergraph. i.e., each chemical species are the nodes for the hypergraph. From given CRN, if we pick subnetwork γ which satisfies the following two conditions,

- If the subnetwork γ contains some chemical species X, then all reactions whose reactant contains X are contained in γ. This is called **output-completeness**.
- - (number of species in γ) + (number of reaction in γ) - (number of closed cycle contained in γ) = 0. We call this value as **influence index**.

then the steady-state concentration of all chemical species outside γ exhibits RPA from the reaction parameters inside γ. In other words, the affect of the parametric perturbation is localized inside γ. And this holds for any choice of kinetic law and parameters.

We call such γ as **Buffering structure**, which is a structual buffer that localizes the effect of perturbation inside it. For example, consider the following CRN



and we pick $\gamma = \{A, B, D, 2, 3, 4, 5\}$. Then,

- Since we pick A, B, D, we contain every reaction whose reactant contains them (reaction $2, 3, 5$)
- number of species = $|\{A, B, D\}|$ = 3, number of reactions = $|\{2, 3, 4, 5\}|$ = 4,
  number of cycle = cycle consisted of reaction 2, 3, 4, 5 = 1. Therefore the influence index is $-3+4-1 = 0$.

Therefore, by theorem 3, the parameter perturbation of reaction $2, 3, 4, 5$ does not affect the steady-state concentration of C, the chemical species outside γ. The counter-intuitive fact is that the parameter of reaction 3 and 4, which affects the concentration of C, does not affects the steady-state concentrations. We could also check this property by numerically solving the ODE system.

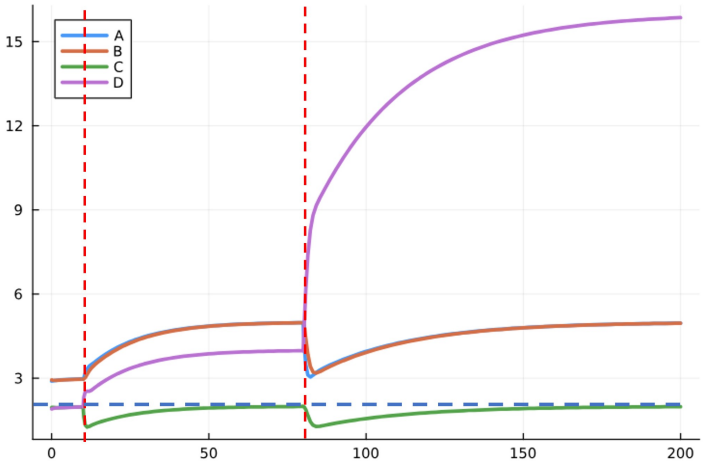


Figure 1: Time-varying concentration of each chemical species within parameter perturbations

Fig, 1 shows the numerical solution of ODE of the example. We made parametric perturbation by multiplying 2 to the parameter of reaction 4 at $t = 10$, and dividing 5 to the parameter of reaction 5 at $t = 75$

(red vertical dotted lines). As we can observe, the steady-state concentrations of A, B and D changes after perturbation, but the steady-state concentration of C does not changed (blue horizontal dotted lines).

One important thing is that the theorem 3 holds for any kinetic law and parameters. In real biology, no one knows whether some reaction follows specific kinetic law such as mass-action kinetics, Michaelis–Menten kinetics, etc. However, for many researches in CRN, we assume specific kinetic law to analyze mathematically. However, this theorem is independent from the choice of kinetic law, which makes it to be widely used.

Another important aspect of this theorem is the role of buffering structure. When we use machines such as calculator, cell phone, etc, we give some input to the machine and expect some specific output, no matter what the internal voltage, current, and resistance values are. Similarly, buffering structures gets some input and makes output. And those steady-state outputs values exhibit RPA with respect to the internal parametric perturbation such as the activity of enzyme, genetic mutation, etc. So, buffering structure works as some machine.

When applying this theorem to the realistic biological example, something interesting happens.

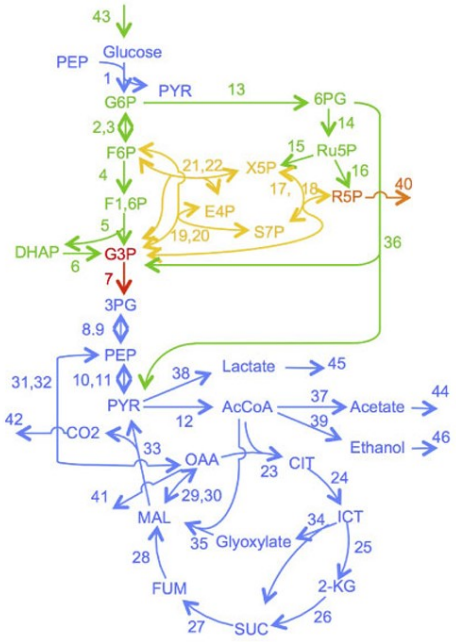


Figure 2: Buffering structures in the central metabolic pathway of *E. coli*.

Fig. 2 is the central metabolic pathway of *E. coli*. And the colored subnetworks are the buffering structures in the central metabolic pathway. The yellow subnetwork is the minimal buffering structure, the union of yellow and red subnetwork is another buffering structure. Also, the blue subnetwork (except reaction 1) is another buffering structure, and the union of yellow, red, green and blue subnetwork is the last buffering structure.

Interestingly, those buffering structures already have a biological name. We call blue subnetwork as *tricarboxylic acid cycle*, the yellow subnetwork as *pentose phosphate pathway*, and the green subnetwork as *glycolysis*. Those names were given by biologists, who named those subnetworks by their distinct functions. **So, before knowing anything about buffering structures, the biologists distinguished some subnetworks by their biological functions. And it turns out that they were actually buffering structures that works as a machine.**

This theorem shows how CRN could give impact to the systems biology, and tells us about the structural origin of biochemical reactions in real biological systems.

3 Stochastic chemical reaction network

3.1 Backgrounds

In the deterministic CRN, we modeled the *concentration* of each chemical species. But modeling with deterministic CRN has some problems. One major problem is that deterministic CRN does not capture the

randomness of biochemical reactions, so that the deterministic CRN becomes unrealistic when it is applied to small systems.

For example, we can deterministically model the SIR model as the following

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

and the solution for this ODE will be some smooth function. But imagine we apply SIR model to our MATH 495 classmates. SIR model will predict, for example, $(S, I, R) = (1.2,\ 4.3,\ 10.5)$ for some timepoint. But *how can 1.2 number of people exist*? The number of population should be non-negative integer. What really happens is that sometimes the population becomes $(S, I, R) = (5,\ 5,\ 6)$, sometimes becomes $(S, I, R) = (6,\ 4,\ 6)$, and sometimes becomes $(S, I, R) = (16,\ 0,\ 0)$.

In this sense, it is more natural to think about the probability of each state, and describing the dynamics of CRN on the state-space. In general, the state space I will be

$$I = \mathbb{Z}_{\geq 0}^{|\mathscr{S}|}$$

and the jump between states will be the stiochiometric vector. Also, we consider a slightly different form of mass-action kinetics, which is defined as

$$\begin{aligned}c_1X_1 + c_2X_2 + \cdots + c_mX_m \to \cdots \ : \ & kn_1(n_1-1)\cdots(n_1-c_1+1)\cdots n_m(n_m-1)\cdots(n_m-c_m+1) \\ &= k\frac{n_1!}{(n_1-c_1)!}\frac{n_2!}{(n_2-c_2)!}\cdots\frac{n_m!}{(n_m-c_m)!}\end{aligned}$$

where n_i denotes the number of X_i. We can notice that if n_i is less than c_i, then the reaction rate becomes zero, which means that the reaction does not fires.

In these setting, we could model the dynamics of system by continuous-time Markov chain. The state for the number of each species corresponds to the state, and the reaction rate in each state corresponds to the transition rate (especially, the Q-matrix of the continuous-time Markov chain). The steady-state corresponds to the stationary distribution of continuous-time Markov chain, and the stability of steady-state corresponds to the ergodicity (i.e., positive recurrent) of the continuous-time Markov chain.

For example, when we construct the continuous-time Markov chain for the example 2, then the state space becomes

$$I = \{(a, b, c) \mid a, b, c \in \mathbb{Z}_{\geq 0}\} = \mathbb{Z}_{\geq 0}^3$$

and the Q-matrix becomes

$$Q_{(a_1,b_1,c_1),(a_2,b_2,c_2)} = \begin{cases} k_1 & \text{if } a_2 = a_1 + 1,\ b_2 = b_1,\ c_2 = c_1 \\ k_2 & \text{if } b_2 = b_1 + 1,\ a_2 = a_1,\ c_2 = c_1 \\ k_3a_1 & \text{if } a_2 = a_1 - 1,\ b_2 = b_1 + 1,\ c_2 = c_1 \\ k_4b_1 & \text{if } a_2 = a_1,\ b_2 = b_1 - 1,\ c_2 = c_1 \\ k_5a_1b_1 & \text{if } a_2 = a_1 - 1,\ b_2 = b_1 - 1,\ c_2 = c_1 + 2 \\ k_6c_1(c_1-1) & \text{if } a_2 = a_1 + 1,\ b_2 = b_1 + 1,\ c_2 = c_1 - 2 \\ -(k_1 + k_2 + k_3a_1 + k_4b_1 & \text{if } a_2 = a_1,\ b_2 = b_1,\ c_2 = c_1 \\ \quad +k_5a_1b_1 + k_6c_1(c_1-1)) & \\ 0 & \text{otherwise} \end{cases}$$

Here, I would like to emphasize two things.

- When we say some stochastic CRN is ergodic, we are considering the case where both mean and variance of the distribution is finite. There are some cases where mean converges to finite value but the variance diverges. In this case, we don't say that the stochastic CRN is ergodic.
- Sometimes people think that the mean of the distribution of stochastic CRN would be same with the concentration value from the deterministic CRN. But it is not true in general.

3.2 Relationship between Stochastic and Deterministic Models

Let's think again about applying the SIR model. When we apply SIR model for only few people, then there are too many randomness so that it is hard to determine which state will the model would be. However, when we apply SIR model for the entire US, we could imagine that the fraction of each susceptible, infectious, and recovered population will have fewer randomness.

We can ask the same question to the deterministic CRN and stochastic CRN. In some sense, if we increase the total volume, then does the stochastic CRN converges to deterministic CRN? And the answer is yes.

Theorem 4. [6, 7, 8] Let $X(t, x_0)$ be the solution of the initial-value ODE problem of

$$\frac{\partial X(t, x_0)}{\partial t} = F(X(t, x_0)), \quad X(0, x_0) = x_0$$

which corresponds to the dynamics of deterministic CRN (e.g. (1)). And let $X^V(t)$ be the continuous-time Markov chain of corresponding stochastic CRN with V amount of volume. If

$$\lim_{V \to \infty} V^{-1} X^V(0) = x_0$$

then,

$$\lim_{V \to \infty} P\{\sup_{s \leq t} |V^{-1} X^V(s) - X(s, x_0)| > \epsilon\} = 0$$

for every t and $\epsilon > 0$.

We could consider the volume V as the following : for the reaction involving m molecules, the chance that all m molecules react during the certain time interval is proportional to $1/V^{m-1}$. The reason for this assumption can be intuitively seen by considering the probability of m balls placed at random in n boxes all ending up in the same box. Theorem 4 says that as the volume size increases, if the initial number of species per unit volume converges to the initial value of deterministic CRN, then the number of species per unit volume at **any timepoint** converges to the concentration calculated from the deterministic CRN.

3.3 Deficiency Zero theorem

Here, let's look at another theorem which shows the relationship between deterministic and stochastic CRN.

Theorem 5. [9] Suppose that the CRN has deficiency zero and weakly reversible and follows mass-action kinetics. Then the corresponding continuous-time Markov chain is positive recurrent and has a stationary distribution consisting of the product of Poisson distributions,

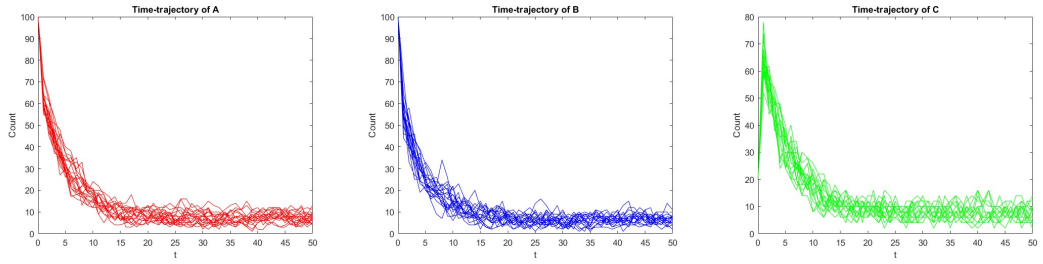
$$\pi(x) = \prod_{i=1}^{m} \frac{c_i^{x_i}}{x_i!} e^{-c_i}, \quad x \in \mathbb{Z}_{\geq 0}^m \tag{7}$$

where c_i's are the equilibrium points determined by modeling the same reactions by deterministic CRN. Also, if $\mathbb{Z}_{\geq 0}^m$ is irreducible, then this is the unique stationary distribution. If $\mathbb{Z}_{\geq 0}^m$ is not irreducible, then the stationary distribution has the form on the state space Γ

$$\pi_\Gamma(x) = M_\Gamma \prod_{i=1}^{m} \frac{c_i^{x_i}}{x_i!} e^{-c_i}, \quad x \in \Gamma \tag{8}$$

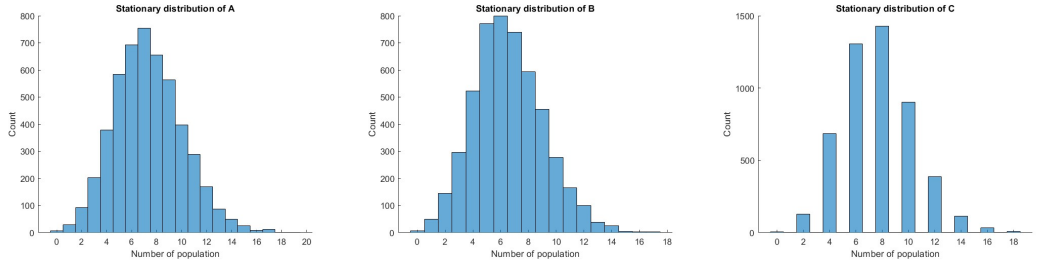
where $\pi_\Gamma(x) = 0$ if $x \notin \Gamma$, and M_Γ is the normalizing constant.

So, by applying theorem 5 to the CRN (2), we could know that it converges to the stationary distribution, which is the product-form of Poisson distributions. Performing the simulation of CRN (2) using the Gillespie's algorithm [10] shows the following trajectories and stationary distributions.



(a) Species A (b) Species B (c) Species C

Figure 3: Time-trajectory of 20 simulations of (2)



(a) Species A (b) Species B (c) Species C

Figure 4: Stationary distribution of 5000 simulations of (2)

We can check from Fig. 4a and 4b that the stationary distribution actually follows the Poisson distribution. Note that the CRN (2) changes the number of C by two units, which makes $\mathbb{Z}^m_{\geq 0}$ not irreducible. That's the reason why Fig. 4c is not fully Poisson-distributed.

3.4 Robust perfect adaptation and Antithetic Integral Feedback controller

One problem of stochastic CRN is that it is hard to know the distribution. In deterministic CRN, the concentration is just the solution of ODE, but in stochastic CRN, the solution is probability distribution. Then, what about we can control the mean of the distribution?

Theorem 6. [11, 12] Suppose every reaction follows mass-action kinetics. Consider the original reaction network and assume that for some given values of its parameters, the original reaction network is ergodic. When we add the four additional reactions

$$\begin{aligned} 0 &\xrightarrow{\mu} Z_1 \\ X_\ell &\xrightarrow{\theta} X_\ell + Z_2 \\ Z_1 + Z_2 &\xrightarrow{\eta} 0 \\ Z_1 &\xrightarrow{k} Z_1 + X_1 \end{aligned} \tag{9}$$

where X_ℓ and X_1 are species from original reaction network, and there exists a sequence of reactions from X_1 to X_ℓ. Then, after adding those four reactions

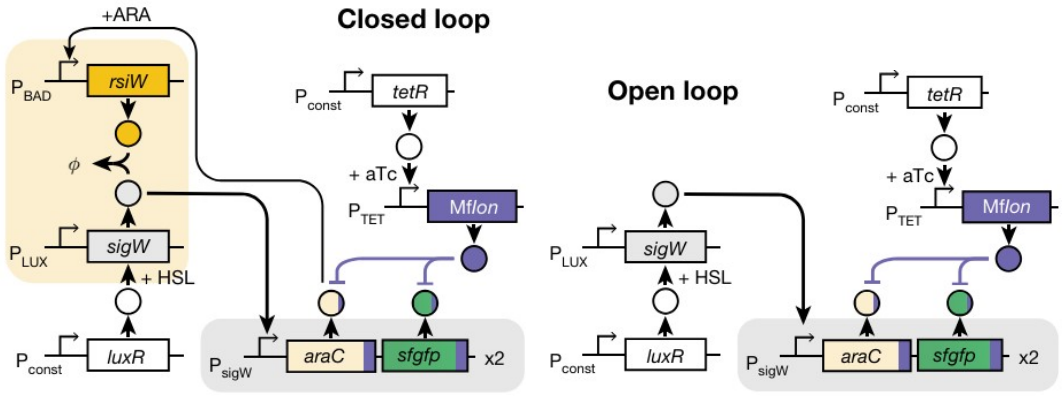
$$\mathbb{E}[X_\ell] = \frac{\mu}{\theta}$$

at the stationary distribution. Also, this value achieves RPA from the parametric perturbation of original reaction network. Moreover, in order to make arbitrary controller that such RPA is satisfied, then the four reactions (9) are necessary.

In other words, no matter what the distribution of original CRN looks like, when we attach four reactions (9), we can control the mean of the stationary distribution of certain chemical species X_ℓ. And we call those

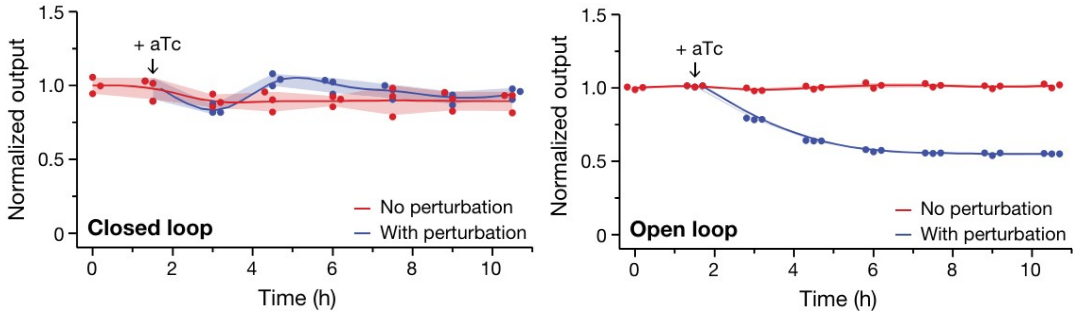
four reactions (9) as **Antithetic Integral Feedback controller**. The statement of this theorem has been slightly modified for better understanding, and the full statement can be seen at Theorem 7

This theorem is also experimentally showed. In the reference paper [12], the authors generated a real biological circuit using the synthetic biological technology.



(a) Reaction network with controller (b) Reaction network without controller

Figure 5: Biochemical reaction with and without controller (9)



(a) Concentration with controller (b) Concentration without controller

Figure 6: Time-trajectory of concentration within environmental perturbation

In [12], the authors measured the number of *araC* by capturing the intensity of fluorescence from *sfgfp*, which is generated simultaneously with *araC*. When the reaction network contains the structure of antithetic integral feedback controller (Fig. 5a), then the concentration of output exhibits RPA (Fig. 6a). However, when the reaction network does not have a controller structure (Fig. 5b), then the concentration of output does not exhibit RPA (Fig. 6b).

4 Conclusion

In this report, we have a slight overview about the chemical reaction network theory. We look through deterministic CRN and stochastic CRN, and the important research topics for each area. I would like to compare deterministic CRN and the stochastic CRN as the following : deterministic CRN is more convenient to check simple properties, so that it could be deeply mathematically studied. These days, the deterministic CRN becomes tightly related with the algebraic geometry. On the other hand, stochstic CRN is more hard to check simple properties, but they are more realistic than deterministic CRN.

There are still many interesting questions about this area. For example, in the deterministic CRN, can we find necessary and sufficient condition for the existence of locally asymptotically stable equilibrium point? What properties does the equilibrium point have? How fast does the dynamical system converges to the equilibrium point? In the stochastic CRN, how can we calculate the exact form of the stationary distribution? How does the mixing time of corresponding continuous-time Markov chain connected to its network structure? How can we check the ergodicity of continuous-time Markov chain from its network structure?

Not only these mathematical questions, CRN could answer the problem from biology [4], and also purpose new biochemical reactions to obtain some properties [12]. It would be interesting to see how does the CRN and systems biology will cooperate and developed in the future.

References

[1] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems. *Chemical Engineering Science*, 42(10):2229–2268, 1987.

[2] Martin Feinberg. *Foundations of Chemical Reaction Network Theory*. Springer, 2019.

[3] Gheorghe. Craciun. Toric Differential Inclusions and a Proof of the Global Attractor Conjecture. *arXiv preprint*, 2016.

[4] Guy Shinar and Martin Feinberg. Structural Sources of Robustness in Biochemical Reaction Networks. *Science*, 327(5971):1389–1391, 2010.

[5] Takashi Okada and Atsushi Mochizuki. Law of Localization in Chemical Reaction Networks. *Phys. Rev. Lett.*, 117:048101, Jul 2016.

[6] Thomas G. Kurtz. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 10 1972.

[7] Thomas G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970.

[8] T. G. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356, 1971.

[9] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-Form Stationary Distributions for Deficiency Zero Chemical Reaction Networks. *Bulletin of Mathematical Biology*, 72(8):1947–1970, 11 2010.

[10] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[11] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. *Cell Systems*, 2(1):15–26, 1 2016.

[12] Stephanie K. Aoki, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570(7762):533–537, 6 2019.

5 Appendix

5.1 Proof of Theorem 3

Let S be the stiochiometric matrix (c.f., (1)). If the CRN is at the steady-state, the concentration of all chemical species does not change. Therefore,

$$Sr(t) = 0 \ \rightarrow \ r(t) \in ker(S)$$

In other words, the rate vector $r(t)$ can be written as

$$r = \sum_{n=1}^{N} \mu^n c_n$$

where $dim(ker(S)) = N$, c_n are the basis of $ker(S)$ and μ^n are coefficients.

Perturbation of the parameter can be written as $k_j \ \rightarrow \ k_j + \delta k_j$. then,

$$\frac{d}{dk_j} r = \sum_{n=1}^{N} \frac{d\mu^n}{dk_j} c_n$$

and we can rewrite this as

$$\frac{d}{dk_j} r_i = \frac{\partial r_i}{\partial k_j} + \sum_{m=1}^{M} \frac{\partial r_i}{\partial x_m} \frac{dx_m}{dk_j}$$

where $\frac{\partial r_i}{\partial k_j}$ denotes the direct effect (nonzero only when $i = j$), and $\frac{\partial r_i}{\partial x_m} \frac{dx_m}{dk_j}$ denotes the indirect effect (parameter perturbation affects chemical species, and then it affects the reaction rates).

So, we can write

$$\frac{d}{dk_j} r_i = \sum_{n=1}^{N} \frac{d\mu_i^n}{dk_j} c_n = \frac{\partial r_i}{\partial k_j} + \sum_{m=1}^{M} \frac{\partial r_i}{\partial x_m} \frac{dx_m}{dk_j}$$

therefore,

$$\sum_{m=1}^{M} \frac{\partial r_i}{\partial x_m} \frac{dx_m}{dk_j} - \sum_{n=1}^{N} \frac{d\mu_i^n}{dk_j} c_n = -\frac{\partial r_i}{\partial k_j}$$

When we write this as a matrix,

$$\begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \dots & \frac{\partial r_1}{\partial x_M} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \dots & \frac{\partial r_2}{\partial x_M} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial r_R}{\partial x_1} & \frac{\partial r_R}{\partial x_2} & \dots & \frac{\partial r_R}{\partial x_M} \end{bmatrix} \begin{matrix} -c_1 & -c_2 & \cdots & -c_N \end{matrix} \begin{bmatrix} \frac{\partial x_1}{\partial k_1} & \frac{\partial x_1}{\partial k_2} & \dots & \frac{\partial x_1}{\partial k_R} \\ \frac{\partial x_2}{\partial k_1} & \frac{\partial x_2}{\partial k_2} & \dots & \frac{\partial x_2}{\partial k_R} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial x_M}{\partial k_1} & \frac{\partial x_M}{\partial k_2} & \dots & \frac{\partial x_M}{\partial k_R} \\ \frac{\partial \mu_1^1}{\partial k_1} & \frac{\partial \mu_2^1}{\partial k_2} & \dots & \frac{\partial \mu_R^1}{\partial k_R} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial \mu_1^n}{\partial k_1} & \frac{\partial \mu_2^n}{\partial k_2} & \dots & \frac{\partial \mu_R^n}{\partial k_R} \end{bmatrix} = \begin{bmatrix} -\frac{\partial r_1}{\partial k_1} & \dots & & \\ \vdots & \frac{\partial r_2}{\partial k_2} & \dots & \\ \vdots & & & \vdots \\ \vdots & \dots & & -\frac{\partial r_R}{\partial k_R} \end{bmatrix}$$

simply writing,

$$\begin{bmatrix} \frac{\partial r_i}{\partial x_m} & -c_n \end{bmatrix} \begin{bmatrix} \frac{dx_m}{dk_j} \\ \frac{d\mu_i^n}{dk_j} \end{bmatrix} = -diag(\frac{\partial r_1}{\partial k_1} \cdots \frac{\partial r_R}{\partial k_R})$$

and the left matrix is called **A**-matrix. Note that we are interested in $\frac{dx_m}{dk_j}$ where m is the species outside buffering structure and j is the reaction inside buffering structure.

Since (number of reaction) = (number of chemical species) + (number of closed cycle) for the buffering structure, and by the output-completeness, choosing the proper basis for the $ker(S)$ and the ordering of the basis leds the **A**-matrix as the following structure

$$\mathbf{A} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix}$$

where the upper-left submatrix is the square matrix with size (number of reaction) $\times$ (number of chemical species + number of closed cycle). Note that (1) output-completeness means the chemical species inside the buffering structure does not affect the reaction outside the buffering structure, and (2) closed cycle generates a basis of $ker(S)$ associated with the reactions consisting of the closed cycle. These two factors induces the lower-left submatrix to be zero matrix. Therefore,

$$\begin{bmatrix} \frac{dx_m}{dk_j} \\ \frac{d\mu_i^h}{dk_j} \end{bmatrix} = -\mathbf{A}^{-1} diag(\frac{\partial r_1}{\partial k_1} \cdots \frac{\partial r_R}{\partial k_R})$$

where

$$-\mathbf{A}^{-1} = \begin{bmatrix} A_\Gamma^{-1} & * \\ 0 & A_{\bar{\Gamma}}^{-1} \end{bmatrix}$$

Note that the column index for A_Γ^{-1} contains the reactions inside buffering structure, and the row index for $A_{\bar{\Gamma}}^{-1}$ contains the specied outside buffering structure. Therefore, when we pick m be the species outside buffering structure and j be the reaction inside buffering structure,

$$\frac{dx_m}{dk_j} \propto \mathbf{A}_{mj}^{-1} = 0$$

which completes the proof.

5.2 Ideas for the proof of Theorem 6

For some Markov process X , define

$$T_i f(i) = \mathbb{E}[f(X(t)) \mid X(0) = i]$$

for any real-valued function f, where i denotes the state for continuous-time Markov chain. Then the **infinitesimal generator** A is defined as

$$Af(i) = \lim_{t \to 0} \frac{T_i f(i) - f(i)}{t}$$

When calculating the infinitesimal generator of stochastic CRN, it becomes

$$Af(\mathbf{x}) = \sum_{k=1}^{R} \lambda_k(\mathbf{x})(f(\mathbf{x} + \zeta_k) - f(\mathbf{x}))$$

where $\lambda_k(\mathbf{x})$ denotes the rate for reaction k at state $\mathbf{x}$, and ζ_k denotes the stiochiometric vector of kth reaction. Taking the inner product with the stationary measure π, we get

$$\langle Af(\mathbf{x}), \pi \rangle = 0$$

Let the original CRN is consisted of $X_1 \cdots X_d$ species and R reactions. When we attach four additional reactions (9), the infinitesimal generator of the corresponding continuous-time Markov chain becomes

$$\begin{aligned} Af(\mathbf{x}, z_1, z_2) = &\sum_{k=1}^{R} \lambda_k(x)\{f(\mathbf{x} + \zeta_k, z_1, z_2) - f(\mathbf{x}, z_1, z_2)\} \\ &+ kz_1\{f(\mathbf{x} + e_1 + z_1, z_2) - f(x, z_1, z_2)\} \\ &+ \mu\{f(\mathbf{x}, z_1 + 1, z_2) - f(\mathbf{x}, z_1, z_2)\} \\ &+ \theta x_\ell\{f(\mathbf{x}, z_1, z_2 + 1) - f(\mathbf{x}, z_1, z_2)\} \\ &= \eta z_1 z_2\{f(\mathbf{x}, z_1 - 1, z_2 - 1) - f(x, z_1, z_2)\} \end{aligned}$$

Then, if we pick $f(\mathbf{x}, z_1, z_2) = z_1 - z_2$, then

$$Af(\mathbf{x}, z_1, z_2) = \mu - \theta x_\ell$$

therefore,

$$\langle Af(\mathbf{x}, z_1, z_2), \pi \rangle = \mu - \theta \mathbb{E}_\pi[X_\ell] = 0$$

therefore,

$$\mathbb{E}_\pi[X_\ell] = \frac{\mu}{\theta}$$

Of course, we can choose any real-valued function f to find the properties of the stationary distribution. For example,

$$\begin{aligned}
f(\mathbf{x}, z_1, z_2) = z_1 &\to \mathbb{E}_\pi[Z_1 Z_2] = \frac{\mu}{\eta} \\
f(\mathbf{x}, z_1, z_2) = z_1^2 &\to \mathbb{E}_\pi[Z_1^2 Z_2] = \frac{\mu}{\eta}(1 + \mathbb{E}_\pi[Z_1]) \\
f(\mathbf{x}, z_1, z_2) = z_2^2 &\to \mathbb{E}_\pi[Z_1 Z_2^2] = \frac{\mu + \theta \mathbb{E}_\pi[X_\ell Z_2]}{\eta}
\end{aligned}$$

So the remaining problem is ergodicity. The main idea is to divide whole state space $\mathbb{Z}_{\geq 0}^{|\mathscr{S}|}$ into subareas, and construct Lyapunov function for each subarea. Then check that the Lyapunov criteria is satisfied for 1) state transition inside each subarea and 2) state transition among other subarea. Doing it, the full statement of the Theorem 6 is as following.

Theorem 7. [11] Assume that the state space of the original reaction network is irreducible and that there exists $v \in R_{>0}^d$, $w \in R_{\geq 0}^d$, $w_1,\ w_\ell > 0$, a positive scalar c_2 and nonnegative scalars $c_1,\ c_3,\ c_4,\ c_5,\ c_6$ such that

$$\begin{aligned}
\sum_{k=1}^{R} \lambda_k(\mathbf{x}) \langle v, \zeta_k \rangle &\leq c_1 - c_2 \langle v, x \rangle \\
\sum_{k=1}^{R} \lambda_k(\mathbf{x}) \langle w, \zeta_k \rangle &\geq -c_3 - c_4 \langle e_\ell, x \rangle
\end{aligned}$$

and

$$\begin{aligned}
\sum_{k=1, \langle v, \zeta_k \rangle > 0}^{R} \lambda_k(\mathbf{x}) &\leq c_5 + c_6 \langle v, x \rangle \\
\sum_{k=1, \langle w, \zeta_k \rangle > 0}^{R} \lambda_k(\mathbf{x}) &\leq c_5 + c_6 \langle v, x \rangle
\end{aligned}$$

hold for all $x \in N_{\geq 0}^d$. Then, the reaction network after attaching four reactions (9) is ergodic provided that $\frac{\mu}{\theta} > \frac{c_1}{c_2 v_\ell}$ and, in such a case, we have that $\mathbb{E}[X_\ell(t)] \to \mu/\theta$ as $t \to \infty$.