# Mathematical Tools for the Understanding of Life

Winfried Just

Department Mathematics, Ohio University,
Athens, Ohio, U.S.A.

November 1, 2007

**This talk will**

- Introduce you to the art and science of mathematical modeling.

- Show you how different areas of mathematics enter this process.

- Illustrate the skills needed to become a successful mathematical modeler.

## Models in science

Scientists (biologists, chemists, physicists ...) want to understand how nature works and make predictions.

Mathematical models can aid in this process by reducing a natural systems to its essentials so that predictions can be based on mathematical analysis of the models.

## The art of mathematical modeling

Mathematical modeling is a process of selective ignorance. You always have to make simplifying assumptions.

- What can we remain ignorant about?

- What can we not remain ignorant about?

## Who are we?

Scientists from different disciplines (philosophy, history, sociology, psychology ...) give different answers.

Biologists tell us that we are giant networks of chemical reactions between tens of thousands of biochemicals. These reactions happen in cells, that are organized in tissues, organs, whole organisms. All phenomena at higher levels of biological organization are *emergent properties* of the dynamics at this lowest, biochemical, level.

## The dynamics of a (bio)chemical reaction network

A chemical reaction network consists of chemical compounds (*chemical species*) and a set of chemical reactions during which some compounds (the *reactants*) are consumed and other compounds (the *reaction products*) are formed.

The *dynamics* of the network is the change over time in the amounts of individual compounds. We will model reaction networks as *dynamical systems.*

## Example of a simple network

Compounds: $A, B, C$

Reactions:

$$A + B \rightarrow 2B + C$$

$$C + B \rightarrow A$$

We will remain blissfully ignorant about the names of these compounds.

## Questions about the network

Suppose we start with given amounts of $A, B, C$.

1. Can the system reach a state were all three compounds are present in specified amounts?

2. What will eventually happen to the system? In particular, will the amounts of these chemicals keep fluctuating, or will the system approach a state where these amounts don't change (much)? Such a state would be called a *steady state* or *equilibrium.*

3. If the system does approach a steady state, what amount of each compound is present in the steady state?

## The state of the network

The *state* of the network at any time is given by the vector of amounts of the chemicals in the network.

If amounts are given as actual numbers of molecules of each chemical species, then the state would be a vector of nonnegative integers. For example, for our system a state $[13, 3, 7]$ would mean thirteen molecules of $A$, three molecules of $B$, and seven molecules of $C$.

If the amounts are given as concentrations (most commonly in moles per liter, where one mole consists of $6.022 \times 10^{23}$ molecules), then a state of our system would be a vector $[[A], [B], [C]]$ of nonnegative real numbers.

The *initial state* is the state of our system at time $t = 0$.

## Translating the first question

Suppose we start with given amounts of $A, B, C$. Can the system reach a state were all three compounds are present in specified amounts?

In mathematical language: Let $[x_A(t), x_B(t), x_C(t)]$ be the state of our network at time $t$. Suppose we start in initial state state $[x_A(0), x_B(0), x_C(0)]$, and let $[y_A, y_B, y_C]$ be given. Can here exist a time $t \geq 0$ such that $[x_A(t), x_B(t), x_C(t)] = [y_A, y_B, y_C]$?

If so, then $[y_A, y_B, y_C]$ is said to be *stoichiometrically compatible* with $[x_A(0), x_B(0), x_C(0)]$.

## The stoichiometry matrix

Let us form a matrix whose rows shows the net consumption (negative sign) and net production (positive sign) of molecules of each species by the corresponding reaction.

$$N = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

**Theorem:** A state $[y_A, y_B, y_C]$ is stoichiometrically compatible with a state $[x_A, x_B, x_C]$ only if there are nonnegative real numbers $\lambda_1, \lambda_2$ such that

$$[y_A - x_A, y_B - x_B, y_C - x_C] = \lambda_1 \vec{r}_1 + \lambda_2 \vec{r}_2,$$

where $\vec{r}_1, \vec{r}_2$ are the rows of $N$.

## Conservation laws

$$N = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

In this case, the rows are linearly dependent and $\lambda_1 \vec{r}_1 + \lambda_2 \vec{r}_2 = \lambda[-1, 1, 1]$. This implies the following *conservation laws* for system (I):

$$x_A + x_B = const$$
$$x_A + x_C = const$$

For example, a state $[7, 3, 13]$ could never be reached from initial state $[13, 3, 7]$.

## Our remaining two questions

Suppose we start with an initial state $[x_A, x_B, x_C]$.

1. What *will eventually* happen to the system? In particular, will the amounts of these chemicals keep fluctuating, or will the system approach a steady state or equilibrium $[y_A, y_B, y_C]$ where these amounts don't change (much)?

2. If so, what is the steady state $[y_A, y_B, y_C]$?

**When do chemical reactions happen?**

We will assume for today that the cell is a *well-stirred chemical reactor* in which all relevant molecules move around in random directions and with random speed.

**Mass-action kinetics:**

*A chemical reaction happens if the reactants bump into each other at sufficiently high speed.*

*Compounds disintegrate spontaneously once in a while.*

## Rates of change

Assume the state of a system is described as a concentration vector. The reaction

$$A + B \rightarrow 2B + C$$

will happen at a rate $k_1[A][B]$ and will cause a change in $[A]$ at the rate $-k_1[A][B]$, a change in $[B]$ at the rate $k_1[A][B]$, and a change in $[C]$ at the rate $k_1[A][B]$.

The reaction

$$C + B \rightarrow A$$

will happen at a rate $k_2[C][B]$ and will cause a change in $[A]$ at the rate $k_2[C][B]$, a change in $[B]$ at the rate $-k_2[C][B]$, and a change in $[C]$ at the rate $-k_2[C][B]$.

## An ODE model for system (I)

These observations lead to the following ODE (ordinary differential equations) model of the network (I).

$$\begin{aligned}
\frac{d[A]}{dt} &= -k_1[A][B] + k_2[C][B] \\
\frac{d[B]}{dt} &= k_1[A][B] - k_2[C][B] \\
\frac{d[C]}{dt} &= k_1[A][B] - k_2[C][B]
\end{aligned} \tag{1}$$

Note that these differential equations are *non-linear.*

## Steady states for system (I)

In a steady state $[[A], [B], [C]]$ we must have

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = \frac{d[C]}{dt} = 0.$$

For the ODE model (1) this translates into:

$$\begin{aligned} 0 &= (-k_1[A] + k_2[C])[B] \\ 0 &= (k_1[A] - k_2[C])[B] \\ 0 &= (k_1[A] - k_2[C])[B]. \end{aligned} \tag{2}$$

## Two steady states for system (I)

Equations (2) are equivalent to

$$k_1[A] = k_2[C] \tag{3}$$

or

$$[B] = 0. \tag{4}$$

If we assume for simplicity that $k_1 = k_2$ and if our initial state is $[13, 3, 7]$, then using the conservation laws

$$x_A + x_B = const$$
$$x_A + x_C = const$$

we find two possible steady states $[10, 6, 10]$ and $[16, 0, 4]$.

**Which equilibrium will the system approach?**

Suppose $k_1 = k_2$ and we have $[A] > [C]$ and $[B] > 0$. Then

$$\begin{aligned} \frac{d[A]}{dt} &= (-[A] + [C])[B] < 0 \\ \frac{d[B]}{dt} &= ([A] - [C])[B] > 0 \\ \frac{d[C]}{dt} &= ([A] - [C])[B] > 0 \end{aligned} \tag{5}$$

These inequalities will hold along the *trajectory* that starts in initial state $[13, 3, 7]$, and therefore we conclude that our system will approach the steady equilibrium $[10, 6, 10]$.

**Does this agree with empirical results?**

You report back to the biochemist and are told: "This is not what I observe in the lab. In my system, $B$ gets totally depleted. I also asked a computer scientist, and he wrote me great software that simulates individual molecules bumping into each other, and in these simulations, $B$ gets always totally depleted after some time."

Now what do you do?

**Carefully examine the assumptions!**

You look at the simulations and notice two things:

- It treats reactions as *discrete events* that happen between *individual molecules.*

- The number of molecules in the simulations is very small.

If we are talking about a small number of molecules, then differential equations may not be the right modeling framework, because the notion of a derivative makes sense only if concentrations can change in arbitrarily small increments.

## How to build a discrete model

Let us assume that the state of the system at any given time $t$ is a vector of nonnegative integers $[x_A(t), x_B(t), x_C(t)]$ that represent the actual numbers of molecules of each species. For any given initial state, the stoichiometrically compatible state space is finite.

Reactions happen one at a time. If reaction (i) happens, then we get

$$\begin{aligned}&[x_A(t+1), x_B(t+1), x_C(t+1)] = \\ &[x_A(t)-1, x_B(t)+1, x_C(t)+1].\end{aligned} \tag{6}$$

If reaction (ii) happens, then we get

$$\begin{aligned}&[x_A(t+1), x_B(t+1), x_C(t+1)] = \\ &[x_A(t)+1, x_B(t)-1, x_C(t)-1].\end{aligned} \tag{7}$$

## Incorporating randomness

Which of the two reactions happens next is a *random event.*

The *odds* of reaction (i) occurring next can be calculated as

$$\frac{k_1 x_A x_B}{k_2 x_C x_B} = \frac{k_1 x_A}{k_2 x_C}.$$

This allows us to calculate the probability that the system goes from state $[x_A(t), x_B(t), x_C(t)]$ into state $[x_A(t) - 1, x_B(t) + 1, x_C(t) + 1]$ as $\frac{k_1 x_A}{k_1 x_A + k_2 x_C}$.

Similarly, the probability that the system goes from state $[x_A(t), x_B(t), x_C(t)]$ into state $[x_A(t) + 1, x_B(t) - 1, x_C(t) - 1]$ is $\frac{k_2 x_C}{k_1 x_A + k_2 x_C}$.

## Markov Chains

We have modeled our system as *stochastic process,* more precisely, a finite *Markov Chain.*

The state $[16, 0, 4]$ is (the only) *absorbing state:* No reaction can happen in this state, and the process can never leave it. It is a steady state of the system; in fact it is the only state where no change can happen. Moreover, it can be shown that the system will eventually reach this steady state, and this is true even if we are talking of a total of 20,000 or twenty million molecules instead of twenty.

But you may have to wait longer than the lifetime of the universe for this to happen.

**Have we found an adequate model of the experimental results?**

Only if the number of molecules in the network is in fact very small.

In this case the model would in fact predict what is being observed in the lab, and it would give us a plausible explanation why we are observing what we do.

This is the best one can hope for any model of natural phenomena. In contrast to mathematical theorems, all such models are in principle falsifiable by new empirical evidence.

**What if the number of molecules is large?**

We can numerically explore the behavior of our Markov Chain if we start, say, in the initial state $[13000, 3000, 7000]$. This can be done at low cost, for example, by a MatLab program that runs on your PC.

What we would almost certainly observe is that the system moves toward $[10000, 6000, 1000]$ and then fluctuates a little bit, but stays close to this state no matter how long we run the simulation. Thus the state $[10000, 6000, 1000]$ would behave like a steady state in which the amounts don't change very much, and the predictions of our ODE model would still be confirmed.

**But how can we explain the experimental results if the number of molecules is in fact large?**

Check your assumptions again!

Recall that real biochemical networks have tens of thousands of reactions. One typically concentrates on a few reactions and hopes that no major players were missed.

So we go back to our biochemist and ask: "Did we perhaps overlook other species and reactions that influence the outcome?"

Answer: "Well, $B$ likes to form dimers $D$."

## A larger network

Compounds: $A, B, C, D$

Reactions:

$$
\begin{aligned}
A + B &\rightarrow D + C \\
C + B &\rightarrow A \\
2B &\rightarrow D \\
D &\rightarrow 2B
\end{aligned}
$$

## An ODE model for system (II)

These observations lead to the following ODE (ordinary differential equations) model of the network (II).

$$\begin{aligned}
\frac{d[A]}{dt} &= -k_1[A][B] + k_2[C][B] \\
\frac{d[B]}{dt} &= -k_1[A][B] - k_2[C][B] - 2k_3[B]^2 + 2k_4[D] \\
\frac{d[C]}{dt} &= k_1[A][B] - k_2[C][B] \\
\frac{d[D]}{dt} &= k_1[A][B] + k_3[B]^2 - k_4[D]
\end{aligned} \tag{8}$$

Now if $k_4 \ll k_3$ ("$B$ likes to form dimers"), then $\frac{d[B]}{dt} < 0$ until $[B]$ becomes very, very small. In the lab, this may look like moving to an equilibrium were $B$ totally disappears, and it may well be the correct explanation of the experimental results!

**What is really going on in the lab?**

The biochemist used the phrase "$B$ gets totally depleted."

But does this mean that only the concentration of monomers of $B$ goes to zero, or the concentration $[B] + [D]$ of $B$ in both its monomer and dimerized forms tends to zero? In the first case, you may have found the correct explanation, in the second case, something else must be going on.

Always maintain clear communication with the domain expert and insist on resolving such ambiguities!

**This was a toy model**

The study of real biological networks can get a lot more complicated.

- We need to assume *Michaelis-Menton-type kinetics.* This makes the differential equations less tractable.

- We often cannot ignore concentration gradients in the cell and its compartments. For this we may need *PDE's (partial differential equations)* instead of ODE's.

- The differential equations may become too complicated to explore analytically. We may need *numerical analysis* to approximately solve them.

- In many cases it may not even be possible to find a suitable mathematical model, but *computer simulations* still may be used for meaningful predictions.

- We may not know the actual reactions that take place, but may need to infer them from observed concentration changes. This brings us to the field of *reverse-engineering* of biochemical networks.

## Conclusions for artful modeling

- Your permissable level of ignorance depends on the question you want to answer.

- Err on the side of learning as many details about the system as possible. You can always ignore the irrelevant ones later.

- Maintain regular dialogue with the domain experts, communicate clearly, resolve ambiguities.

- Different mathematical frameworks may be appropriate for different problems.

- Keep an open mind!

**What does it take to become a successful mathematical modeler?**

- Familiarity with a wide variety of mathematical tools, including linear algebra, ordinary and partial differential equations, dynamical systems, probability and statistics, stochastic processes, numerical analysis, computer programming. You don't need to be a specialist in any one of these areas, but should know something about all of them and be prepared to talk to the specialists.

- Good communication skills, especially the ability to listen to and explain your mathematics to specialists in other disciplines.

- A commitment to lifelong learning, both of mathematics and other disciplines.