

For high school students: Example of math in biology

Hyun Youk
(Dated: January 2018)

I. MATHEMATICS: DIFFERENTIAL EQUATIONS

In high school, a typical math problem that you encounter is:

Solve for x in the following equation:

$$x^2 + 3x + 2 = 0. \quad (1)$$

Another typical question may look like:

Solve for y in terms of x in the following equation:

$$x^2 + y^2 = 1 \quad (2)$$

Both are examples of **algebraic equations**. A question that you may be less familiar with is of the following sort:

Solve for y in the following equation:

$$\frac{dy(x)}{dx} + 2 = 0 \quad (3)$$

Above is an example of a **differential equation**. It is an equation that involves the unknown function $y(x)$ and its derivatives. y is a function of x , and the question above is asking us to find a function $y(x)$ whose derivative with respect to x satisfies Eq. 3. In Eq. 2, y is a function of x as well, and we recognise it as an equation describing a unit circle (i.e., a circle with radius 1). We can solve algebraic equations such as Eqs. 1 and 2, in principle, by using algebraic manipulations such as multiplications, divisions, taking square roots, factoring, etc. For example, rearranging Eq. 2 and taking the square root of both sides of the resulting equation yields $y(x) = \pm(1 - x^2)^{1/2}$. As for Eq. 1, factoring the equations lets us see that the equation has two solutions: $x = -2$ and $x = -1$. But can we use the same techniques to solve for $y(x)$ in the differential equation (Eq. 3)? At the end of the day, we would like to say $y(x) =$ some function of x , just as we did when we solved Eq. 2. But how do we even introduce x into Eq. 3? As it stands, x does not even appear anywhere in Eq. 3. Surely, we need to have it eventually appear so that we can write $y(x) =$ something involving x .

There are usually two methods of solving differential equations:

- Method 1: Guess the solution based on your intuition of the system being studied, and then check if it's indeed a solution by plugging your guessed function in the equation. For example, if we're studying a pendulum, we would expect that the solution $y(x)$ would be a periodic (oscillating) function such as $\cos(x)$ and $\sin(x)$. So we would try $y(x) = \cos(x)$ and $y(x) = \sin(x)$, plug it into the differential equation, and then see if the equation is satisfied (i.e., lefthand side = righthand side).
- Method 2: Systematic way of solving. This usually involves integrating the differential equation or some other techniques (e.g., power series expansion of $y(x)$).

Method 2 does not always work, mainly because many differential equations are just too difficult to solve by hand (or no systematic method of obtaining exact solution is known for that equation yet). In such cases, we rely on computer algorithms to obtain *approximate* solutions to the equation. Method 1, in principle, should always work since if there is a solution, and you can guess it, then you can check if it's a solution to the equation or not. Of course, the problem is that it is incredibly difficult to guess the correct solution without spending days or weeks of

trial and error in many situations. Also, there is often more than one solution to an equation, just as Eq. 1 has two solutions. So even if we guess a correct solution, we need to ensure that we have not left out any other solutions for the equation. For example, say that we did not know how to solve Eq. 1 systematically (e.g., by factoring, etc.) but that we instead decided to guess what the x should be. And let us say that we guessed $x = -1$, plugged it into Eq 1 and found that indeed, $x = -1$ satisfied Eq. 1, which makes it a solution. Great! Now, how do we know that there are no other solutions besides $x = -1$? After all, $x = -2$ is also a solution. We will address this important question for differential equations after some examples. The subtleties with method 1 aside, we will encounter equations that we can usually solve using both methods. Let us now apply the two methods in solving Eq. 3:

Method 1 applied to solving Eq. 3: Rearranging Eq. 3, we get

$$\frac{dy}{dx} = -2 \tag{4}$$

Browsing through functions that we know of, can we think of a function $y(x)$ whose derivative with respect to x gives us -2? How about $y(x) = -2x$? Let's check this guess by substituting it into Eq. 3. This involves taking a derivative of $-2x$, after which Eq. 3 becomes

$$\frac{dy}{dx} = -2 \tag{5}$$

Therefore, we found a solution to Eq. 3: $y(x) = -2x$. But are there other solutions to above equation? How about $y(x) = -2x + 1$? Plugging the derivative of this into Eq. 3 lets us see that this is indeed also a solution. In fact we can see that

$$y(x) = -2x + c, \tag{6}$$

where c is any constant (i.e., some number, not a function of x) is a solution to Eq. 3. Thus, there is an infinite number of solutions to Eq. 3 (one for each value of c), but all of them have the *same form* (Eq. 6). In fact, it turns out that there are no other functions that satisfy Eq. 3. Eq. 6 describes all *class of solutions to Eq. 3*.

Method 2 applied to solving Eq. 3: Stepping aside from our original problem for a bit, suppose we are told the following:

$$\frac{\Delta y}{\Delta x} = -2 \tag{7}$$

This is just telling us that the slope is -2, and by multiplying both sides by Δx , we get

$$\Delta y = -2\Delta x \tag{8}$$

Suppose that we are standing at $x = 2$. If we walk over to $x = 5$ (thus $\Delta x = 3$), then y changes by $-2\Delta x = -6$. But changing from what initial value? If we say $y(x = 2) = c$, then we would say $y(x = 5) = c - 2\Delta x = c - 6$. Keeping this example in mind, let us go back to our differential equation (Eq. 3). Just as we did above, we *multiply* both sides of Eq. 3 by dx and get:

$$dy = -2dx \tag{9}$$

Think back to what a derivative means. It is the slope of a tangent line at a particular point on the curve. Thus, what Eq. 9 is saying is that if we are initially standing at $x = x_0$ (the curve has value $y(x_0)$ there), and we walk by dx (so we are now at $x_0 + dx$), then the change in the value of y is now dy . More specifically, $y(x_0 + dx) = y(x_0) + dy$ where $dy = -2dx$ as we found in Eq. 9. Now, say that we moved to a new x-position: $x = x_f$, which is quite far away

from where we were standing ($x = x_0$). But we can always take baby steps (each with an *infinitesimal* step size dx) and eventually go from $x = x_0$ to $x = x_f$, after N steps, where $dx = (x_f - x_0)/N$. So we have

$$y(x_f) = y(x_0) + (-2)dx + (-2)dx + \dots + (-2)dx \quad (10a)$$

$$= y(x_0) + (-2)(dx + dx + \dots + dx) \quad (10b)$$

$$= y(x_0) - 2 \sum_{x_0}^{x_f} dx \quad (10c)$$

$$= y(x_0) - 2Ndx \quad (10d)$$

$$= y(x_0) - 2(x_f - x_0) \quad (10e)$$

Now, we can choose x_f to be any position that we want (i.e., we can choose our final position x_f to be any value and the procedure from Eq. 10a to Eq. 10e still works). So, instead of writing the subscript "f" in the x_f , we drop it and just write it as x (i.e., x_f is an *independent variable*, so we write it as x). Hence we have

$$y(x) = y(x_0) - 2x + 2x_0 \quad (11)$$

Finally, while we vary the final destination x_f to be anything (so we called it "x", a variable), we have fixed the initial position x_0 . So $y(x_0) - 2x_0$ is a constant, which we can call c . So, writing $c = y(x_0) - 2x_0$, Eq. 12 becomes

$$y(x) = -2x + c \quad (12)$$

This is the solution to the differential equation (Eq. 3) that matches the one we found with Method 1 (Eq. 6).

We have just solved Eq. 3 with two different methods. Now, let us go back to the beginning of method 2. In particular, let us look at the procedure from Eq. 10a to Eq. 10e. We have been a bit cavalier about writing and summing up dx as if it is nothing special. But in fact, it is very special because, by *definition*, it turns out that

$$\sum_{x_0}^{x_f} dx = \int_{x_0}^{x_f} dx \quad (13)$$

Indeed, if we do the integral, we get

$$\int_{x_0}^{x_f} dx = x \Big|_{x_0}^{x_f} \quad (14a)$$

$$= x_f - x_0 \quad (14b)$$

which is exactly the same as Eq. 10e. The sum of the infinitesimal quantity dx (Eq. 10a) is called the **Riemann sum**, named after the mathematician Bernhard Riemann (1826 - 1866) (see <http://en.wikipedia.org/wiki/Riemann>), and is **THE definition of integration**. Let us now do another example where we really need to use integration, instead of the straightforward summation as in Eqs. 10a - 10e.

Example: Solve the following differential equation:

$$\frac{dy}{dt} - at - v = 0, \quad (15)$$

where a and v are constants (Note: unless stated otherwise, any letters that you see in an equation are constants, not a function of some variable).

Solution: In this problem, y is a function of t : $y(t)$. Our goal is to find what $y(t)$ is. Let us solve this problem with method 2. Mimicking the previous example, we rearrange the equation to get

$$\frac{dy}{dt} = at + v, \quad (16)$$

Again, mimicking the previous example, we multiply both sides by dt and get

$$dy = (at + v)dt \quad (17)$$

Again, we imagine two values of the independent variable (in this case, t): $t = t_0$ is our starting position, and $t = t_f$ is our final position. $y(t_0)$ and $y(t_f)$ are the values of y at t_0 and t_f respectively. Then using the same logic as in the previous example, imagining uniformly dividing up the interval $[t_0, t_f]$ into a huge number of segments - say there are N such segments - each with an infinitesimal length $dt = (t_f - t_0)/N$. Now, if we just write

$$y(t_f) - y(t_0) = \sum_{t_0}^{t_f} (at + v)dt. \quad (18)$$

How would we evaluate this sum? It is unclear how we would do this off the top of our head because the $(at + v)$ is a function that *continuously* changes as the t changes. But by making N large, we can make $dt = (t_f - t_0)/N$ to be as small as we desire (that is what we mean by "infinitesimal" after all). And if we do so, then within a given infinitesimal interval dt (say in the interval $[t_5, t_5 + dt]$, there is not much room for the function $at + v$ to change much. That is, the difference between $at_5 + v$ and $a(t_5 + dt) + v$ is infinitesimal:

$$a(t_5 + dt) + v - (at_5 + v) = a(dt). \quad (19)$$

Since dt is infinitesimal, $a \cdot dt$ too is an infinitesimal quantity as well (i.e., no matter how large a may be, we can always pick N to be sufficiently large enough so that $dt = (t_f - t_0)/N$ is made small enough to compensate for the large a , thus ensuring that the product $a \cdot dt$ is indeed very small). Then, the sum in Eq. 18 is accurately approximated by

$$y(t_f) - y(t_0) = \sum_{t_0}^{t_f} (at + v)dt \quad (20a)$$

$$= (at_0 + v)dt + (at_1 + v)dt + \dots + (at_N + v)dt \quad (20b)$$

$$= \int_{t_0}^{t_f} (at + v)dt \quad (20c)$$

$$= \left(\frac{at^2}{2} + vt \right) \Big|_{t_0}^{t_f} \quad (20d)$$

$$= \frac{a(t_f^2 - t_0^2)}{2} + v(t_f - t_0) \quad (20e)$$

As before, by redefining t_f to be t , above equation becomes

$$y(t) = \frac{at^2}{2} + vt + y(t_0) - \frac{at_0^2}{2} - vt_0 \quad (21)$$

And again, mimicking our solution to the previous example, we lump together the constant $y(t_0) - \frac{at_0^2}{2} - vt_0$ and write it as c . Hence, above equation (and our solution $y(t)$ to the differential equation 15 is

$$y(t) = \frac{at^2}{2} + vt + c \quad (22)$$

Note that this is actually an equation familiar to us from kinematics in high school physics course. There, you may have seen it as $x(t) = x_0 + v_0t + \frac{at^2}{2}$, the equation describing the position $x(t)$ of a particle at time t , where x_0 is the initial position, v_0 is the initial velocity, and a is the constant acceleration. So looking at Eq. 15 again, we see that it is actually describing the position $y(t)$ of a particle as a function of time t and that v and a are respectively the initial velocity and acceleration of the particle.

Example: Solve the following differential equation:

$$\frac{dy}{dt} - f(t) = 0 \quad (23)$$

Solution: Applying method 2 to this equation and mimicking the solutions of the previous two examples, we get

$$\frac{dy}{dt} = f(t) \quad (24a)$$

$$\implies dy = f(t)dt \quad (24b)$$

$$\implies \int_0^{y_f} dy = \int_0^{t_f} f(t)dt \quad (24c)$$

$$\implies y(t_f) - y(0) = \int_0^{t_f} f(t)dt \quad (24d)$$

$$\implies y(t_f) = \int_0^{t_f} f(t)dt + y(0) \quad (24e)$$

Note that here, we just picked $t_0 = 0$. This is fine. We can pick the starting position t_0 to be anything that we want. Often, $t_0 = 0$ is the easiest choice since then many terms usually become zero, making us write less. Since we do not want to keep writing the subscript "f" in the t_f (which is an independent variable anyway), we drop the subscript and write

$$y(t) = y(0) + \int_0^t f(t')dt' \quad (25)$$

This is the general solution to the differential equation (Eq. 23). Since we do not know the actual function $f(t)$, we cannot go ahead and actually do the integral. So this is as far as we can go. In fact, for many of the equations that we encounter in science, the $f(t')$ is *non-analytic*, meaning that we cannot do the integral by hand. We solve such integrals numerically (i.e., approximately) with computer programs. Also, we picked $t = 0$ here as a reference point in doing the integral above. But we can always pick another point, say $t = 2.5$ or $t = \pi$ and many other possible choices. So we could have written above equation as

$$y(t) = c + \int_a^t f(t')dt', \quad (26)$$

where $y(a) = c$. What value we pick for the reference point a (and thus $y(a) = c$) is up to us; it is arbitrary.

Exercises for this section

Solve the following differential equations. For these equations, use method 2 but you may also try method 1 if you can guess the right solution. You'll see that this guessing game is not so easy. But guessing is easier if the equations are based on some physical phenomena that you are modelling. Remember that the key in solving these equations (using method 2) is isolating the two variables by themselves (one variable on the lefthand side, the other variable on the righthand side of the equal sign).

$$\begin{aligned}
1. \frac{dz}{dx} + 5x &= 0 \\
2. \frac{dy}{dt} + at^2 + b &= 0 \\
3. \frac{dx}{dz} + e^z &= z && \text{Hint : } \int e^t dt = e^t + c \\
4. \frac{dx}{dt} + \frac{1}{t} &= 3 && \text{Hint : } \int \frac{1}{y} dy = \ln(y) + c \\
5. x \frac{dx}{dt} + bt &= 0 \\
6. \frac{dk}{dt} + k &= 3
\end{aligned}$$

For the last question, the following properties of the exponential e and **natural logarithm** \ln (i.e., \log_e) should be helpful.

Properties of the natural logarithm $\ln(x)$ and the exponential e (also written as " $\exp(x)$ "):

- (i) $\ln(xy) = \ln(x) + \ln(y)$
 - (ii) $\ln(x/y) = \ln(x) - \ln(y)$
 - (iii) $\ln(y^n) = n\ln(y)$ This follows from (i) and (ii). Can you prove this relation using (i) and (ii)?
 - (iv) $e^x e^y = e^{x+y}$
 - (v) $\exp(\ln(x)) = x$ and $\ln(\exp(x)) = x$
- (i.e., $\exp(x)$ is another way of writing e^x . e^x and $\ln(x)$ are inverse functions of each other).

See solutions to the above exercises at the end of this document.

II. CHEMISTRY: ENZYME-SUBSTRATE KINETICS

The basis of every chemical reaction is converting one group of atoms into another group of atoms. The "group of atoms" could be a molecule like H_2O or a protein, which consists of many molecules. A protein could be small, consisting of a few amino acids, or it could be a large machine consisting of many subproteins, such as the ribosome. Whatever the case, a key in chemical reactions is the **reaction rate** - this measures how fast one group of atoms is converted into another group of atoms, be they simple molecules or large proteins. The math that describes these rates are differential equations. What determines the rate of the chemical reactions? Many things, such as the temperature, concentration, etc. In this section, we will find an equation that describes all these effects into one equation.

We write chemical reactions like this:



Here, this means that we have a chemical specie A , which meets with a chemical specie, B , and then forms another chemical specie C . This is what we call a **unidirectional reaction**. But we can also have the following type of reaction:



which we call a **bidirectional reaction** because A and B can form C and at the same time, C can spontaneously form A and B . A good example of this occurs for a very important type of reaction in chemistry, including in cells, called the **enzyme-substrate reaction**. We write it as follows:



Here, E is an **enzyme**, S is a **substrate**, ES is the enzyme and the substrate that are bound together (i.e., "stuck" to each other), and P is the **product**.

We now turn to mathematically describing rates of chemical reactions. Let's start with the simplest one, the unidirectional reaction (Eq. 28). In fact, once we have an equation for the unidirectional reaction, then we are set because the bidirectional reaction is the unidirectional reaction occurring twice. What we mean by a rate of chemical reaction is how fast (i.e., rate) each chemical specie changes over time. Rate of change over time is just the derivative with respect to time, $\frac{d}{dt}$. So $\frac{d[A]}{dt}$, for example, would be the rate of change of the concentration of A .

We can derive the equation for $\frac{d[A]}{dt}$ from a pure thought. Suppose that you have N_A particles of A and N_B particles of B floating in a "room" of volume V . For example, if we have both an enzyme and a substrate in a bath of water, then V is the volume of the water inside some beaker (i.e., we assume that there are many water molecules and that they do not participate in the reactions). V can also be the volume of a living cell, if E and S are swimming inside the cell. We assume that the volume of each chemical specie is much smaller than V .

Let us first consider several specific situations first. Suppose we have one particle of A and ten particles of B (i.e., $N_A = 1$, $N_B = 10$). In how many ways can the one particle of A collide with a particle of B ? Limiting ourselves to collisions involving just two particles (i.e., ignoring three particles simultaneously colliding), there are ten possible such **bimolecular collisions**. To see this, we can number each of the ten particles of B as B_1, B_2, \dots, B_{10} . Then when the one particle of A collides with B_1 , we can write AB_1 to denote the instantaneous complex that A and B_1 form. This complex may fall apart only a short time later or may stay glued together permanently but we do not care. The main point is that you see the A touching B_1 at the moment that they collide. So that is why we can call it a complex, AB_1 , the same way that we have ES when we have an enzyme E bound to a substrate S - E and S are touching each other. If we write the ten possible complexes, they are: $AB_1, AB_2, \dots, AB_{10}$. Thus there are ten bimolecular complexes formed by A colliding with B . Note that $N_A N_B = 10$

Let us consider a more complex situation. Suppose now that we have $N_A = 10$ and $N_B = 120$. How many possible bimolecular complexes formed by a particle of A colliding with a particle of B ? This time, let's label the particles of A as well as the particles of B . So we write, A_1, A_2, \dots, A_{10} , and B_1, B_2, \dots, B_{120} . Then the bimolecular complexes involving A_1 are: $A_1 B_1, A_1 B_2, A_1 B_3, \dots, A_1 B_{120}$ - there are 120 of these. The bimolecular complexes involving A_2 are: $A_2 B_1, A_2 B_2, A_2 B_3, \dots, A_2 B_{120}$ - there are 120 of these as well. In fact, we can see that for a particle A_i (where $1 \leq i \leq 10$), there are 120 bimolecular complexes of the form $A_i B_j$ (where $1 \leq j \leq 120$). Thus there is a total of 1200 bimolecular complexes. Note that $N_A N_B = 1200$.

From above two specific situations, we can deduce the total number of bimolecular complexes formed by A colliding with B , when there is some arbitrary number of particles of A and B . If N_A and N_B are the number of particles of A and B respectively, then $N_A N_B$ is the total number of bimolecular complexes formed by A together with B . In the unidirectional reaction (Eq. 28), *we assume that the number of collisions per unit time is directly proportional to the number of particles of C (the product) is produced per unit time*. In other words, the more collisions occurring per unit time means more product produced per unit time. Number of collisions between A and B per unit time is directly proportional to the number of the bimolecular complexes that A and B can form per unit time. Thus the change in the number of particles of C , denoted ΔN_C , per time interval Δt is directly proportional to the total number of bimolecular complexes of the form AB that *can* form, which is $N_A N_B$, in that time interval Δt . Mathematically, this means that

$$\frac{\Delta N_C}{\Delta t} \propto N_A N_B, \quad (31)$$

where \propto symbol means "proportional to". In chemical reactions, we measure concentrations in a reaction vessel such as the cell or the water bath, instead of measuring the absolute numbers of particles N_A , N_B and N_C . But concentration is just the number of molecules per unit volume. If we assume that the particles of A , B , and C are all uniformly distributed (i.e., well mixed) inside the reaction vessel, above equation becomes

$$\frac{\Delta[C]}{\Delta t} \propto [A][B], \quad (32)$$

where $[A]$, $[B]$, and $[C]$ are the concentrations of A , B , and C respectively. The concentrations are typically measured in **moles per liter (written as mol/L)**, where one mole is 6.02×10^{23} particles (this number is also called the

Avogadro number. To turn the proportionality above into an equation, we introduce a proportionality constant k , called a **rate constant**, to write

$$\frac{\Delta[C]}{\Delta t} \propto k[A][B], \quad (33)$$

We make one more modification: we convert the $\frac{\Delta[C]}{\Delta t}$ into $d[C]/dt$ by selecting a time interval Δt to be infinitesimal (and thus $\Delta[C]$ too would be infinitesimal because $[C]$ does not have much time to change). Thus the above proportionality turns into the following equation:

$$\frac{d[C]}{dt} = k[A][B], \quad (34)$$

This is the rate equation for C in the unidirectional reaction (Eq. 28). We can write the rate equation for A and B as well. Note that each time that one particle of C is created, one particle of A and one particle of B must disappear (i.e., they are used up to form the C). This what Eq. 28 means. A and B must be disappearing at the same rate as the rate of creation of C . Thus we have

$$\frac{d[A]}{dt} = -k[A][B] \quad (35a)$$

$$\frac{d[B]}{dt} = -k[A][B] \quad (35b)$$

$$\frac{d[C]}{dt} = k[A][B] \quad (35c)$$

Note the negative sign in front of Eqs. 35a and 35b. The negative sign is there because k and concentrations are positive numbers, and thus $k[A][B]$ is positive, whereas $[A]$ and $[B]$ must decrease. The negative sign makes $\frac{d[A]}{dt}$ and $\frac{d[B]}{dt}$ negative in Eqs. 35a and 35b. The complete set of rate equations for the unidirectional reaction (Eq. 28) consists of Eqs. 35a, 35b, and 35c.

Why not write $d[A]/dt = k[A][B]/2$ and $d[B]/dt = k[A][B]/2$? It might be tempting to write $d[A]/dt = k[A][B]/2$ because the creation of C at rate $k[A][B]$ destroys A and B together. Moreover, A and B combine in one to one ratio to yield a molecule of C . But this would be a mistake. The reason is that one particle of C is created from one particle of A and one particle of B combined. So there is a one-to-one ratio between C and A . There is also the same one-to-one ratio between C and B . Thus both A and B must be created at the same rate as C . Another way to resolve a potential confusion here is by noting that there are two particles on the left side and one particle on the right side of the reaction in Eq. 28. Thus one particle of C creates two particles - one is A and the other is B . So the total particle number is *not* conserved in this reaction scheme: If $N_A = 50$, $N_B = 50$, and all of A and B are converted into C , then we would have gone from a total of 100 particles ($=N_A + N_B$) to a final count of 50 particles ($=N_C$). Thus the total number of particles goes down by half, a loss in the total particle number. Writing $d[A]/dt = k[A][B]/2$ and $d[B]/dt = k[A][B]/2$ would mean that the total particle number in the system would stay the same at all times, which is evidently not the case.

Bidirectional reactions: Let's write down the rate equations for bidirectional reactions (Eq. 29). Bidirectional reactions are just the unidirectional reactions occurring twice, in opposite directions. In fact, we already know how to write the rate equations for *any* unidirectional reaction because our derivation above did not require us to know whether A or B or C represents carbon, oxygen, protein, etc. The only additional item here is that an arrow can go forward and reverse, meaning that C is created (by the forward arrow) and also destroyed (by the reverse arrow). A and B to are created (by the reverse arrow) and destroyed (by the forward arrow). Let's start with the rate equation for C . For the forward reaction, the **rate of creation** is same as in Eq. 34: $k_f[A][B]$. Here, we defined a new variable k_f , which has the same meaning as in Eq. 34 but just a different name. The **rate of destruction** is: $k_r[C]$, where k_r is the rate constant for the reverse reaction. The total change in C is due to the net effect of the creation competing with the destruction. Thus we can write

$$\frac{d[C]}{dt} = k_f[A][B] - k_r[C], \quad (36)$$

If you are confused by why the destruction term is $-k_r[C]$, ask yourself how we derived the rate equation for our first unidirectional reaction (Eqs. 35a - 35c). There, we counted the number of bimolecular complexes. In other words, we counted the number of molecules that can lead to creation of C . The logic behind writing $-k_r[C]$ for the degradation is the same: If there are N_C particles of C , then each one of them may degrade in a unit time. Degradation, like creation, is a stochastic event. Not every one of the N_C molecules will degrade in the next second, but the more particles of C we have, the higher the chance that any one of them will degrade. This is reflected by the degradation rate being proportional to $[C]$. Another point to note is that k_r does not have to equal k_f since the creation and destruction are independent of each other.

Similarly, we can write down the rate equations for A and B for the bidirectional reaction. They are:

$$\frac{d[A]}{dt} = k_r[C] - k_f[A][B] \quad (37a)$$

$$\frac{d[B]}{dt} = k_r[C] - k_f[A][B] \quad (37b)$$

$$(37c)$$

In summary, the complete set of rate equations for the bidirectional reaction (Eq. 29) is

$$\frac{d[A]}{dt} = k_r[C] - k_f[A][B] \quad (38a)$$

$$\frac{d[B]}{dt} = k_r[C] - k_f[A][B] \quad (38b)$$

$$\frac{d[C]}{dt} = k_f[A][B] - k_r[C], \quad (38c)$$

Equilibrium condition: Ideally, we would solve above equations to determine $[A]$, $[B]$, and $[C]$ as functions of time. We will turn to such a situation in the next example, when we discuss the Enzyme-substrate reaction. For now, let's ask what happens if we wait a long time. Intuitively, for most chemical reactions, we would expect that the reaction will stop because some of the chemical species would have all been depleted. In the bidirectional reaction that we consider here (Eq. 29), nothing runs out because A and B can be generated from C through the reverse arrow while C can be generated from A and B through the forward arrow. But it may be possible that these two processes - the reverse and forward reactions - would occur at equal rates, in which case no more changes in $[A]$, $[B]$, and $[C]$ would be possible. Indeed, almost all chemical reactions reach such an **equilibrium state**. For the bidirectional reaction (Eq. 29), this means that

$$\frac{d[A]}{dt} = 0 \quad (39a)$$

$$\frac{d[B]}{dt} = 0 \quad (39b)$$

$$\frac{d[C]}{dt} = 0, \quad (39c)$$

Taking $\frac{d[A]}{dt} = 0$, we see that this means that

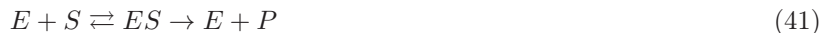
$$k_r[C] = k_f[A][B] \quad (40a)$$

$$\implies \frac{k_f}{k_r} = \frac{[C]}{[A][B]} \quad (40b)$$

The last line is simply saying that an equilibrium state is reached when $[C]$ is a certain multiple of $[A][B]$. Once this multiple is reached, $[A]$, $[B]$, and $[C]$ no longer changes in subsequent times. The constant k_f/k_r is so important in chemistry that we give it a special name, **equilibrium constant** K . $K = \frac{k_f}{k_r}$. For many chemical reactions, we cannot measure k_f and k_r separately. We can only measure K , through measuring the concentrations of each chemical species (in this case, $[A]$, $[B]$, and $[C]$). We started this discussion by asking what happens to the concentrations

after a long time. Without any math, we can convince ourselves that the bidirectional reaction (Eq. 29) reaches an equilibrium state after a sufficiently long time. To see this, let's suppose that initially (i.e., $t = 0$), the forward rate is "higher" than the reverse rate. Then $[C]$ would increase while $[A]$ and $[B]$ would decrease. But as the $[C]$ increases, the reverse rate is increasing too because that rate is proportional to $[C]$. Meanwhile, as the $[A]$ and $[B]$ decrease, the forward rate also decreases because that rate is proportional to these two concentrations. Thus as forward rate, which was initially higher than the reverse rate, decreases while the reverse rate, which was initially lower than the forward rate, increases. These two rates must then eventually meet. Where they meet is defined by the equilibrium condition. Indeed, note that Eq. 40a is precisely the mathematical statement of this. This argument also applies if the forward rate initially is lower than the reverse rate. Thus, in summary, equilibrium condition is reached for the bidirectional reaction after a sufficiently enough time has passed.

Enzyme-substrate reaction: We now turn to the enzyme-substrate reaction. Recall that it is:



We want to write the rate equations for each chemical specie that appears here: E is the enzyme, S is the substrate, ES is the Enzyme bound to the substrate, and P is the final product that the enzyme converts the substrate to. We can think of above reactions as composed of two reactions: $E + S \rightleftharpoons ES$ and $ES \rightarrow E + P$. From the previous two examples, we can write the rate equations for each of the two:

$$\frac{d[E]}{dt} = -k_f[E][S] + k_r[ES] + k[ES] \quad (42a)$$

$$\frac{d[S]}{dt} = -k_f[E][S] + k_r[ES] \quad (42b)$$

$$\frac{d[ES]}{dt} = k_f[E][S] - k_r[ES] - k[ES] \quad (42c)$$

$$\frac{d[P]}{dt} = k[ES] \quad (42d)$$

These are the rate equations for the enzyme-substrate reaction scheme shown in Eq. 41. Our goal now is to solve these equations to find how $[P]$ changes over time. In other words, we want to determine $[P]$ as a function of time t by solving the differential equation 42d. Since Eq. 42d depends on $[ES]$ which itself is determined by a differential equation (Eq. 42c), we must determine $[ES]$ as a function of time by solving Eq. 42c. But Eq. 42c itself depends on other variables, $[E]$ and $[S]$, which themselves have their own differential equations. In fact, Eqs. 42a - 42d form a **system of coupled differential equations**. This like in high school, where you learned to solve a **system of coupled algebraic equations** like

$$x + y = 3 \quad (43a)$$

$$2x + y = 5, \quad (43b)$$

where you need both equations to solve for the unknowns, x and y . We have a similar situation here: we need all four equations (Eqs. 42a - 42d) to solve for $[E]$, $[S]$, $[ES]$, and $[P]$ as functions of time t . To solve for these four variables, we need to play around with Eqs. 42a - 42d. There is no algorithm for solving a system of coupled differential equations that works every time. Let's find $[E]$, $[S]$, $[ES]$, and $[P]$ as a function of time. First, note that E never gets created or destroyed in any of the reactions in Eq. 41. E is either alone or is stuck to S (in the form ES), but it does not disappear when we look from one side to the other side of any of the three arrows in Eq. 41. This means that $[E] + [ES]$, which is the total amount of enzyme (free and bound to S) should not change over time. This is indeed the case. If we add Eq. 42a to Eq. 42c (i.e., add lefthand side of Eq. 42a to the lefthand side of Eq. 42c, and the righthand side of one equation to the righthand side of the other equation), we get

$$\frac{d[E]}{dt} + \frac{d[ES]}{dt} = 0 \quad (44a)$$

$$\implies \frac{d([E] + [ES])}{dt} = 0 \quad (44b)$$

$$\implies [E] + [ES] = \text{constant} \quad (44c)$$

Let's call the constant, E_0 (i.e., $[E] + [ES] = E_0$). E_0 is the total amount of enzyme, including those that are bound to S in the form ES . It is the concentration of the enzyme that we measure and insert by hand to begin the enzyme-substrate reaction. Next, let's make an assumption about the enzyme-substrate reaction: We assume that an equilibrium state is reached for the bidirectional reaction in Eq. 41. Then, with K being the equilibrium constant ($K = k_f/k_r$), we have

$$[ES] = K[E][S] \quad (45a)$$

$$\implies E_0 - [E] = K[E][S] \quad (45b)$$

$$\implies E_0 = [E](1 + K[S]) \quad (45c)$$

$$\implies \frac{E_0}{1 + K[S]} = [E] \quad (45d)$$

Now, note that we can substitute $[ES]$ with $E_0 - [E]$ in Eq. 42d. Thus we have,

$$\frac{d[P]}{dt} = k(E_0 - [E]) \quad (46)$$

Substituting Eq. 45d into above equation, we have

$$\frac{d[P]}{dt} = k\left(E_0 - \frac{E_0}{1 + K[S]}\right) \quad (47a)$$

$$\implies \frac{d[P]}{dt} = k\left(E_0 - \frac{E_0}{1 + K[S]}\right) \quad (47b)$$

$$\implies \frac{d[P]}{dt} = kE_0 \frac{K[S]}{1 + K[S]} \quad (47c)$$

$$\implies \frac{d[P]}{dt} = \frac{kE_0[S]}{\frac{1}{K} + [S]} \quad (47d)$$

We can redefine new constants to make Eq. 47d appear simpler. We define V_m to be kE_0 and $1/K$ to be K_m . Then Eq. 47d becomes

$$\frac{d[P]}{dt} = \frac{V_m[S]}{K_m + [S]} \quad (48)$$

Above is a famous result in chemistry, called **Michaelis-Menten kinetics**. Solving this equation tells us $[P]$ as a function of time t . But this equation is very difficult to solve with a pen and paper. But in fact, what we care more about is the rate, $d[P]/dt$, for a given $[S]$. This is what the Michaelis-Menten equation (Eq. 48) tells us.

Interpreting the Michaelis-Menten kinetics: Let us interpret what the Michaelis-Menten equation (Eq. 48) is saying about rate of the product creation. First, we note that in the limit of $[S]$ approaching infinity, $d[P]/dt$ approaches a constant value, V_m , rather than shooting off to infinity. This means that after there is sufficiently high concentration of the substrate, we cannot increase the creation rate of the product further by adding even more substrate. $d[P]/dt$ does increase as we increase the substrate concentration, but that increase becomes smaller and smaller, and eventually negligible despite adding, say, million times more substrate in the reaction mix. The second major feature of the Michaelis-Menten equation concerns the K_m . note that when $[S] = K_m$, we have $d[P]/dt = V_m/2$. For this reason, K_m is often called the **half-saturation concentration** - it is the substrate concentration at which the production rate is half the **maximum possible production rate**, V_m . The half-saturation concentration K_m is a measure of the **binding affinity** - how well can an enzyme capture the substrate (or how much do the substrate and the enzyme "like" each other). The *lower* the K_m , the *higher* the binding affinity (i.e., the more easily the enzyme can bind a substrate). The *higher* the K_m , the *lower* the binding affinity (i.e., the more difficult it is for an enzyme to bind a substrate). For many enzymes, the maximum possible production rate V_m is independent of the binding affinity. That is, an enzyme with a high binding affinity and another enzyme with a low binding affinity can both have the same high maximal production rate.

To deduce the units of V_m and K_m , we first note that the unit of $d[P]/dt$ is concentration/time (e.g., Molarity per second, M/s) and thus the right side of Eq. 48 must have this unit too. The right side of Eq. 48 has a denominator,

$K_m + [S]$, where $[S]$, which has a unit of concentration, is added to K_m . Since we can only add apples with apples and oranges with oranges (i.e., add two variables only if they have the same units), K_m must also have a unit of concentration (e.g., moles/L, M). Then in $\frac{V_m[S]}{K_m + [S]}$, we see that $\frac{[S]}{K_m + [S]}$ is dimensionless because the numerator's $[S]$ has a unit of concentration that cancels with the denominator's unit, which is also a concentration. Thus we have V_m times a dimensionless number that must give us a unit of concentration/time. We thus conclude that V_m must have a unit of concentration/time (e.g., M/s).

Finally, let us plot $d[P]/dt$ as a function of $[S]$ (i.e., plot Eq. 48). We deliberately left this until the end to show that we could deduce the properties of the Michaelis-Menten kinetics above without looking at the plot.

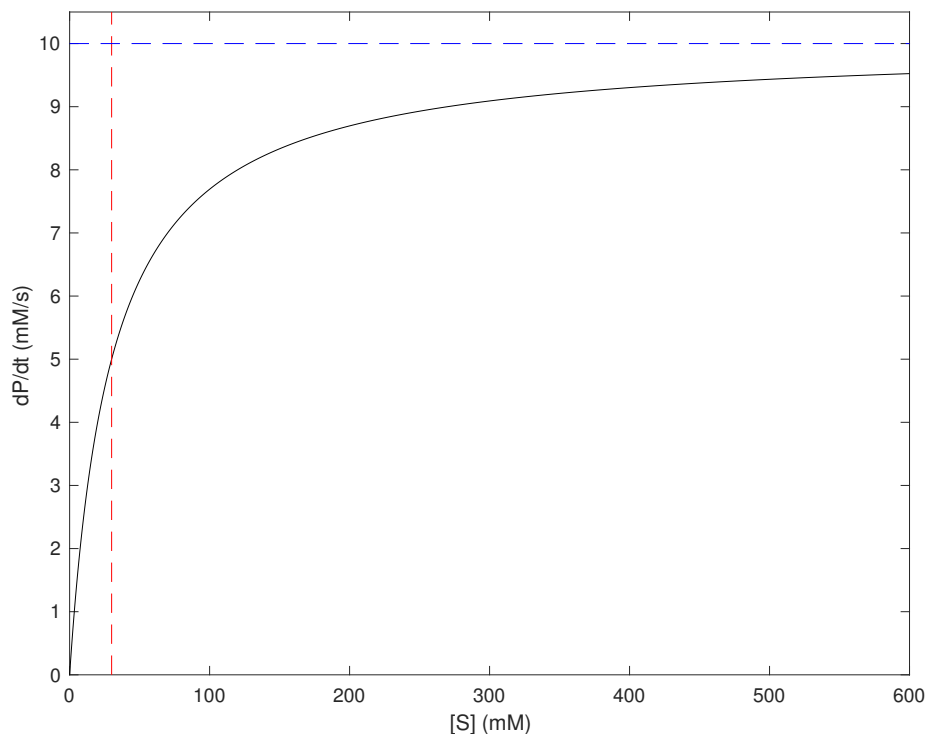


Figure 1. **Michaelis-menten kinetics (Plot of Eq. 48)**: Product-creation rate $d[P]/dt$ as a function of the substrate concentration $[S]$. Red vertical line shows location of the half-saturation concentration K_m , set as 30 mM, on the horizontal axis. Blue horizontal line shows the location of the maximum possible production rate V_m , set as 10 mM/s, on the vertical axis.

As seen in Fig. 1, the interpretations of Eq. 48 that we deduced prior to plotting the equation are indeed true: The black curve in Fig. 1, which is the plot of the right-hand side of Eq. 48, approaches the maximal production rate ever slower and slower as the substrate concentration $[S]$ increases. Moreover, the sharp increase seen in the black curve is because when the substrate concentration equals to K_m , which is 30 mM in the graph, the product-creation rate must be half the maximal production rate.

Exercises for this section

1. Write down the reaction-rate equations (i.e., equations for the rate of change of each chemical specie in the reaction) for the following reaction: **Hydrolysis of sucrose** by an enzyme, **invertase**, into glucose and fructose.



where `sucrose::invertase` is the sucrose bound to invertase. Let k_f be the rate constant for the forward arrow in the bidirectional reaction, k_r be the rate constant for the reverse arrow in the bidirectional reaction, and k be the rate constant for the unidirectional reaction.

2. Re-derive the Michaelis-menten kinetics equation (Eq. 48) by starting with the enzyme-substrate reaction (Eq. 41).
3. Consider the Michaelis-Menten kinetics of the enzyme-substrate reaction with $V = 2$ nM/hr and $K = 0.5$ nM. If the substrate concentration changes from 5 nM to 5000 nM, by how much does the product-creation rate change?

See solutions to the above exercises at the end of this document.

III. BIOLOGY: GENE REGULATION

Gene expression is the process of making RNA and protein from a gene. Genes are particular sequences on DNA. We say that a gene is expressed if the molecules that the gene sequence codes for - RNA and (usually) the protein - are made. More specifically, if a protein is coded by the gene, then the RNA that is made from the gene is called **mRNA (messenger RNA)**. The mRNA in turn contains the sequences that encode the protein, both its content and shape. In a more fancy language, we say that mRNA is **transcribed** from the gene on the DNA - this process is called **transcription** - and that a protein is **translated** from the mRNA - this process is called **translation**. Specifically, particularly big enzymes, which are called macromolecular machines due to their size (because they are usually formed from several proteins), govern the transcription and translation. One of these macromolecular machines, called **RNA polymerase** slides over the gene sequence on a DNA molecule to make (transcribe) an mRNA. Specifically, as the RNA polymerase slides over the gene sequence, letter-by-letter, it grabs a single base (i.e., a single letter) that is floating around inside the cell and joins it to the next letter that it grabs (which corresponds to the next letter of the gene sequence). As a result, once the RNA polymerase has finished running through the entire gene sequence, it will have assembled a string of single-bases (i.e., a string of letters that look like one half of a ladder) - this is the mRNA coded by the gene. Once this transcription of mRNA is completed, another macromolecular machine called the **ribosome** slides over this mRNA to synthesize a protein that the mRNA, and in turn the gene, encodes. As the ribosome runs through the mRNA, letter-by-letter, it grabs onto **amino acids** that are floating inside the cell and then, like the RNA polymerase, attach one amino acid to the next, with each amino acid corresponding to the letters that it is running through. Once assembled, the string of amino acids folds, like an origami, into a three-dimensional shape that we call a **protein**. How this folding occurs is still not fully understood and is an active topic of research in biology. Each protein is a molecular machine that performs a certain function inside the cell. The protein's shape and amino-acid composition determines what it does. For example, the protein can be an enzyme for certain processes or it can form part of the cellular structure (e.g., cell wall) like a brick in a building. Despite the wide variety of proteins that exist inside a cell and thus the genes that code the proteins may be - the human cell roughly has 20,000 genes that code for a protein - only a few types of differential equations dictate the concentrations of proteins and of mRNAs inside cells. The specific form of the equation is determined not by the name of the proteins or the genes, but rather by which type of **gene-regulation scheme** controls those genes. In this section, we derive and analyze the equations for some widely-occurring types of gene-regulations.

Gene-regulation scheme I: Constitutive gene expression: At its simplest, gene expression can be captured by the following process:



where X is the name of a gene, DNA_X is the DNA molecule that contains the gene sequence, $mRNA_X$ is the mRNA transcribed from gene X , and $protein_X$ is the protein translated from $mRNA_X$. Here, gene- X produces one type of mRNA and one type of protein, each at some constant rate. Let k_m be the transcription rate (i.e., rate associated with the chemical reaction in Eq. 50a) and k_p be the translation rate (i.e., rate associated with the chemical reaction in Eq. 50b). Note that DNA and the RNA polymerase are not consumed in this process shown in Eq. 50a. Thinking of RNA polymerase as an enzyme, this makes sense. Note that DNA also does not get consumed during transcription. If it did, it would be like burning a book (DNA) after copying all the words in the book (sequence of letters), which wouldn't make for a good library (cell)! Cells indeed do the sensible thing and keeps the DNA intact. Eq. 50a is thus like the enzyme-substrate kinetics except that here, we have two enzymes - RNA polymerase and DNA_X - that

yield a product, $mRNA_X$, without any substrates. Note also, in Eq. 50b, that the mRNA and the ribosome are not consumed during the translation of $protein_X$.

Unlike the RNA polymerase, DNA, and the ribosome, the $mRNA_X$ and the $protein_X$ do degrade. We represent these degradations as



where \emptyset represents "nothingness" (i.e., symbol for null). Let γ_m be the constant degradation rate of the $mRNA_X$ (reaction in Eq. 51a) and γ_p be the constant degradation rate of the $protein_X$ (reaction in Eq. 51b). In addition to not having any degradations of the Ribosome, DNA, and the RNA polymerase, we will assume that their concentrations do not change over time. That is, we assume that their concentrations inside the cell remain constant throughout the life of the cell. This is a good assumption for many situations. One place where this assumption can fail is during the cell cycle in which the cell replicates all of its DNA. This assumption also fails during physical replication of the cell itself (one cell becoming two cells) since the cell's volume doubles in this case (thus the concentration would be halved). With these caveats in mind and using our knowledge from the previous section (section II), the above four chemical kinetics equations - Eqs. 50a- 51b - yield the following equations:

$$\frac{dm}{dt} = k_m[DNA_X][\text{RNA polymerase}] - \gamma_m m \quad (52a)$$

$$\frac{dp}{dt} = k_p m[\text{Ribosome}] - \gamma_p p \quad (52b)$$

where $m = [mRNA_X]$ and $p = [protein_X]$. Since $[DNA_X]$, $[\text{RNA polymerase}]$ and $[\text{Ribosome}]$ do not change over time, we can simplify Eqs. 52a and 52b by defining new constants: $r_m = k_m[DNA_X][\text{RNA polymerase}]$ and $r_p = k_p[\text{Ribosome}]$. Then Eqs. 52a and 52b become

$$\frac{dm}{dt} = r_m - \gamma_m m \quad (53a)$$

$$\frac{dp}{dt} = r_p m - \gamma_p p \quad (53b)$$

Above two are coupled differential equations. They are coupled because the protein concentration p depends on the mRNA concentration m . We would like to determine m and p as functions of time (i.e., determine $m(t)$ and $p(t)$). We can solve for $m(t)$ by hand. The tricky part is finding $p(t)$ because it depends on $m(t)$, which changes over time. Due to this complication, it turns out that we cannot solve for $p(t)$ exactly by hand. So we have two options: (1) Solve the above equations exactly using a computer program (e.g., MATLAB), or (2) Solve above equations approximately using pen and paper (i.e., analytically) by making biologically reasonable assumptions. Let's do option (2) first. Note that we can solve above equations by hand if $p(t)$ does not depend on the m , or if m does not change over time. Indeed our saviour is that for typical genes, the changes to their mRNA concentrations occur much faster than the changes to the concentrations of the proteins that these mRNAs encode. That is, as we did in deriving the Michaelis-Menten equation in the previous section, we assume that the mRNA concentration quickly reaches an equilibrium-state value before any protein is translated from it. This is what we call a **quasi steady-state assumption**. Then Eqs. 53a and 53b become

$$\frac{dm}{dt} = r_m - \gamma_m m \quad (54a)$$

$$\frac{dp}{dt} = r_p m_{eq} - \gamma_p p, \quad (54b)$$

where m_{eq} is the constant, equilibrium-state concentration of the mRNA. Now note that Eqs. 54a and 54b have mathematically the same form but with different variables. Let us solve for m_{eq} . By definition of equilibrium-state, this means that when $m = m_{eq}$, we must have $dm/dt = 0$. Then Eq. 54a gives us

$$m_{eq} = \frac{r_m}{\gamma_m}, \quad (55)$$

and thus Eq. 54b becomes

$$\frac{dp}{dt} = \frac{r_p r_m}{\gamma_m} - \gamma_p p, \quad (56)$$

Eq. 55 makes sense: The higher the production rate of the mRNA, the more of the mRNA there is inside a cell. The higher the degradation rate of the mRNA, the less of the mRNA there is inside the cell. Fully solving for dm/dt tells us how the mRNA concentration reaches this steady-state over time. To solve equations like dm/dt and dp/dt - the later now has the same form as dm/dt - we use the following two-step procedure. First, taking Eq. 54a, we solve for the simpler case in which there is no production. In this case, Eq. 54a becomes

$$\frac{dm_1}{dt} = -\gamma_m m_1, \quad (57)$$

where we relabelled m as m_1 (just a different variable name so that we do not confuse ourselves later, as we will see).

Solving Eq. 57 (using Method 2 in Section I) yields

$$m_1(t) = A e^{-\gamma_m t} \quad (58)$$

where A is a constant whose value we will later determine. Next, we look for a particular solution for the original equation, Eq. 54a:

$$\frac{dm_2}{dt} = r_m - \gamma_m m_2, \quad (59)$$

where we have relabeled m as m_2 (again, just a different variable name so that we do not confuse ourselves later, as we will see). Here, we would succeed in finding a solution that satisfies Eq. 59. To do so, here we use Method 1 from Section I: Namely, we guess a function for $m_2(t)$ and then check that it is a solution by plugging it into Eq. 59 (i.e., if the lefthand side and the righthand side of Eq. 59 match after plugging in our guessed $m_2(t)$, then this $m_2(t)$ is a solution). Since we need to find just one solution that satisfies above equation, we can make our job easier by finding the easiest case. For example, if we guess that m_2 is a constant (i.e., does not depend on time), then its derivative with respect to time would be zero. This seems simple enough. Is it actually a solution (i.e., satisfy Eq. 59)? We check by plugging in $m_2 = \text{constant}$ into Eq. 59. Then we find that m_2 must be the following:

$$m_2 = \frac{r_m}{\gamma_m} \quad (60)$$

In fact, this m_2 is the equilibrium-state value that we found earlier! So far, we found m_1 and m_2 . m_1 is the solution to the simpler equation, Eq. 57, and m_2 is the solution to the original equation, Eq. 59. If we add dm_1/dt to dm_2/dt , note that we obtain

$$\frac{d(m_1 + m_2)}{dt} = r_m - \gamma_m(m_1 + m_2), \quad (61)$$

Thus $m_1 + m_2$ is also a solution to the original equation. To see this, note that if we relabel $m_1 + m_2$ as m , then above equation becomes

$$\frac{dm}{dt} = r_m - \gamma_m m, \quad (62)$$

which gives us back the original equation that we wanted to solve, Eq. 54a. This completes the two-step procedure for solving differential equations of the form, Eq. 54a. To summarize, we first found the solution to a simpler equation - the one which does not have any m on the righthand side, like Eq. 57) - and called it $m_1(t)$. Then in the second step, we guessed *one* solution for the original equation, Eq. 59, and called it $m_2(t)$. The most **general solution** to the original equation, Eq. 54a, is the sum of the two solutions: $m_1(t) + m_2(t)$. Unfortunately, explaining why this is, in a mathematically rigorous manner, is beyond the scope of this reading material. You will learn this in advanced math

courses. Suffice to say here that although $m_2(t)$ is a solution to the original equation, it is just *one* solution. Thus successfully finding this one solution does *not* guarantee that we have found all solutions to the original equation or that there are no other solutions to the original equation. Quite the contrary, we showed above that $m = m_1 + m_2$ is also a solution the original equation. The logical leap that we are making here, without the proof that goes beyond the scope of this reading material, is that $m(t) = m_1(t) + m_2(t)$ is in fact the most general solution to the original equation. We call this the most general solution because

$$m(t) = m_1(t) + m_2(t) \quad (63a)$$

$$= \frac{r_m}{\gamma_m} + Ae^{-\gamma_m t} \quad (63b)$$

contains an undefined constant, A . By setting A to be a particular value (e.g., $A = 3$), we get a **particular solution**. We know that A can indeed be any value because in obtaining Eq. 58, we did not have to assume any particular value for the A to show that it satisfies Eq. 57. In fact, the freedom to choose a value for A comes from the freedom to choose how many molecules of $mRNA_X$ there is initially inside the cell (i.e., $m(t = 0)$). To see this, note that at $t = 0$, Eq. 63b becomes

$$m(0) = \frac{r_m}{\gamma_m} + A \quad (64)$$

since $e^0 = 1$. Thus, $A = m(0) - r_m/\gamma_m - A$ is indeed determined by the initial condition (i.e., condition at $t = 0$). Under the quasi-steady state assumption the equation for dp/dt , which is Eq. 56, takes the same form as that of dm/dt , which is Eq. 62. Thus we can use the same two-step procedure that gave us the $m(t)$. Or more simply, we just relabel the variable names accordingly in Eq. 63b to obtain the most general solution for $p(t)$:

$$p(t) = \frac{r_p r_m}{\gamma_m} + Be^{-\gamma_p t}, \quad (65)$$

where B is a constant that is determined by the initial condition on p (i.e., $p(0)$). Thus, like Eq. 64, we have

$$p(0) = \frac{r_p r_m}{\gamma_m} + B \quad (66)$$

With these general solutions, let's look at the case in which there are initially no proteins or mRNAs (i.e., $p(0) = 0$ and $m(0) = 0$). Then we can solve for A through Eq. 64 and B through Eq. 66 to convert Eqs. 63b and 65 into

$$m(t) = \frac{r_m}{\gamma_m} \left(1 - e^{-\gamma_m t} \right) \quad (67a)$$

$$p(t) = \frac{r_p r_m}{\gamma_m} \left(1 - e^{-\gamma_p t} \right) \quad (67b)$$

Note that the equilibrium-state of the protein concentration, which we get by setting $dp/dt = 0$, is $\frac{r_p k_m}{\gamma_p}$. This intuitively makes sense (run through the argument yourself as we did above for m_{eq} to convince yourself). Curiously, note that the time to reach the equilibrium-state value is determined only by the degradation rates, γ_p and γ_m , because only the degradation rates appear inside the exponentials.

Plotting $m(t)$ and $p(t)$ as functions of time t (Fig. 2), we see that both curves have the same shape, which makes sense since Eqs. 67a and 67b have mathematically the same form with just different variables. In Fig. 2, we also see that the mRNA concentration rapidly rises to and then remains at the equilibrium-state value before the protein concentration does (i.e., the red curve rapidly flattens while the blue curve is still rising in Fig. 2). Recall that we used the quasi-steady-state assumption to simplify Eqs. 53a and 53b into Eqs. 54a and 54b. So our solutions, Eqs. 67a and 67b, are only valid if the mRNA concentration ($m(t)$) indeed reaches an equilibrium-state value much faster than the protein concentration ($p(t)$) does. Looking at Eqs. 67a and 67b, this means that the γ_m and the γ_p must be well-separated in scales because they are the parameters that determine when $m(t)$ and $p(t)$ reach equilibrium-state

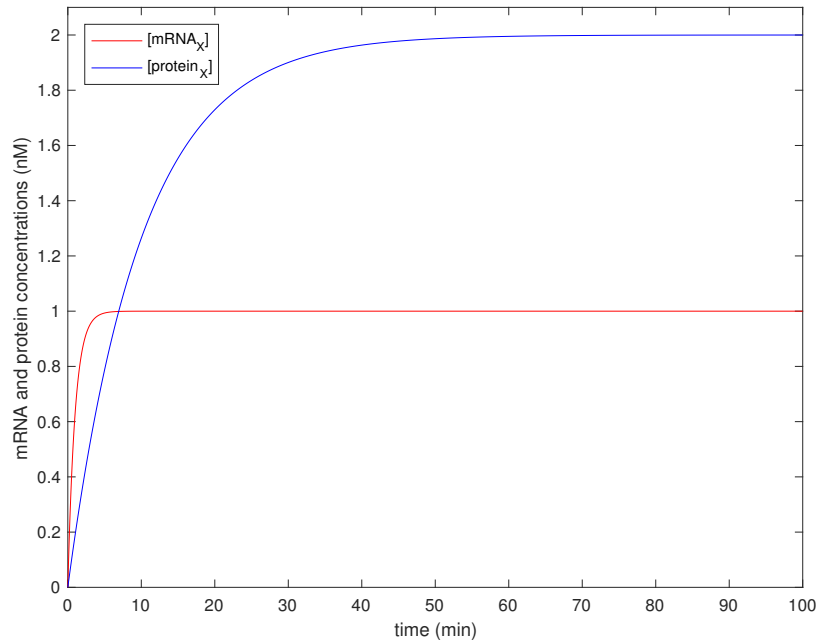


Figure 2. **mRNA and protein concentrations for a constitutive gene expression (Plot of Eqs. 67a and 67b)**: Red curve is [mRNA] as a function of time (i.e., $m(t)$ in Eq. 67a) and blue curve is [protein] as a function of time (i.e., $p(t)$ in Eq. 67b). Values for the parameters are: $r_m = 1$ nM/min, $r_p = 2$ min⁻¹, $\gamma_m = 1$ min⁻¹, and $\gamma_p = 0.1$ min⁻¹. Here nM means nanomolar (nano = 10^{-9}).

values (i.e. when the curves in Fig. 2 flatten out). This is because of the time t in Eqs. 67a and 67b only appears in the exponentials, the $e^{-\gamma_m t}$ and the $e^{-\gamma_p t}$. Both exponentials decay to zero as t increases since γ_m and γ_p are always positive numbers. If γ_m is larger than γ_p , then $e^{-\gamma_m t}$ reaches zero faster than $e^{-\gamma_p t}$. Indeed, in Fig. 2, we have chosen γ_m to be ten times larger than γ_p . If we had assigned a value to γ_m that is similar to or smaller than the value of γ_p , then our quasi-steady-state assumption would be untrue, and thus the equations that we derived based on this assumption - these are Eqs. 67a and 67b - would be invalid. The quasi-steady-state assumption mathematically means that γ_m is much larger than γ_p .

Interpretation of the constitutive gene-expression scheme: At the beginning of this section, we introduced constitutive gene expression as a gene-regulation scheme in which the production rate of mRNA (transcription rate per gene) and protein (translation rate per mRNA) are constant. Yet we found that both the protein and the mRNA concentrations inside a cell reach equilibrium-state values, whereby they do not increase or decrease any further over time. This occurs because the mRNA and the protein degrade over time. From Eq. 53a, we see that as the mRNA concentration increases, the mRNA-degradation rate, $\gamma_m m$, also increases while the mRNA-production rate, r_m , remains constant (this is what the "constitutive" in constitutive gene-expression means). Thus when $m = 0$ initially, the production rate is higher than the degradation rate, which is initially zero. Thus the mRNA concentration starts to increase. But as this happens, the degradation rate gradually increases from zero to a higher value, eventually equalling the production rate r_m . When this happens, the mRNA concentration reaches an equilibrium value m_{eq} (Eq. 55). Looking at Eq. 54b, the same argument works to explain why the protein concentration reaches an equilibrium-state value. The main moral here is that the cell does not burst open because it is getting bigger and bigger with ever increasing concentrations of mRNA and of protein!

Equations that describe general gene-regulation schemes: While we cannot discuss all possible forms of gene-regulation schemes here, we can describe a sort of "universal" equations that describes many forms of gene-regulation. Eqs. 54a and 54b, in which the production rates compete with the degradation rates, give us a hint of such equations. We can describe many forms of gene-regulation scheme via the following equation:

$$\frac{dp}{dt} = f(m, p) - g(p) \quad (68)$$

where $f(m, p)$ is some positive-valued function that depends on the mRNA concentration m and the protein concentration p , $g(p)$ is some positive-valued function that depends on p . Here, m and p are concentrations of mRNA and of protein, both of which are encoded by the same gene. Here, we have limited ourselves only to the protein concentration for convenience. The fact that the production-rate function, f , depends on the mRNA that encodes the protein makes sense - the more mRNA there is, the more likely that ribosomes can produce a protein. The fact that the degradation-rate function, g , does not depend on m also makes sense - degradation of a protein is independent of how many copies of the mRNA there are inside the cell. As mentioned above, biology is full of exceptions. There are exceptions to what we described above. But Eq. 68 still captures a large class of gene-regulation schemes. Finally, note that f and g are functions that yield zero or positive values. Like the positive-valued constants, k_m , k_p , r_m , r_p , γ_m , and γ_p that we used for describing the constitutive gene-regulation, rates are always positive by convention. We placed the negative sign in front of the g in Eq. 68 because, like in Eq. 54b, the degradation term g must decrease the p . Since g is a positive-valued function, the negative sign in front of it is required to represent a degradation of the protein. We will use Eq. 68 in the next section to describe another form of gene-regulation scheme.

Gene-regulation scheme II: Autoregulatory positive feedback: We now consider a gene-regulation scheme in which one gene is linked to itself through a positive feedback. Here, a gene X produces a protein X that is an **activator** of its own production. This means that protein X binds to the **promoter** of gene X, which increases the transcription rate of X, and in turn, the production rate of protein X. A **promoter** is a sequence of DNA that is just in front of the gene. It controls how well RNA polymerase can bind to that region and then run through the gene - it controls the transcription rate of the *mRNA* from that gene. The regulation scheme we are describing here is called an **autoregulatory positive feedback**. Using Eq. 68, the basic equation that describes the protein concentration, denoted p , is

$$\frac{dp}{dt} = f(m, p) - g(p) \quad (69)$$

We make one simplification: The production function f depends only on p , not m . So we have

$$\frac{dp}{dt} = f(p) - g(p) \quad (70)$$

We assume that the same type of protein degradation occurs, so $g(p) = \gamma p$, where γ is a positive constant that signifies the degradation strength. As for the $f(p)$, we use an equation that looks similar to the Michaelis-Menten equation, Eq. 48:

$$f(p) = \frac{Vp^n}{K^n + p^n}, \quad (71)$$

where V , K , and n are positive constants. Note that when p approaches infinity, then f approaches V (take the limit, $\lim_{p \rightarrow \infty} f(p)$, to convince yourself). So V is the maximal production rate, just like in the Michaelis-Menten kinetics (Eq. 48). When $p = K$, f is $V/2$. Thus, like in the Michaelis-Menten kinetics, we call K the **half-saturation constant**, because it is the concentration of protein p at which the production rate is half of the maximum possible value (i.e., half of the saturated rate). The n is called a **Hill coefficient** and is typically between 1 and 2 for most genes. A Hill coefficient of 2 is, in fact, quite rare in cells and is considered to be a very high number. The Hill coefficient characterizes how **switch-like** the positive feedback is, with a higher value being more switch-like and a value closer to one being less switch-like (see Fig. 3). In Fig. 3, we see that indeed, when the Hill coefficient is incredibly high ($n = 10$), the production rate $f(p)$ looks like a step-function (thus the name "switch-like"). Specifically, for $n = 10$, when p is smaller than K , $f(p)$ is nearly zero (no production of protein) whereas when p is larger than K , then $f(p)$ approximately equals the maximum possible production rate, V . Taking a close-up view of the production rate $f(p)$ at low protein concentrations (i.e., low values of p), we see that indeed as the Hill coefficient increases, the closer $f(p)$ approaches zero when p is smaller than K . (see Fig. 4). When $n = 1$, $f(p)$ is exactly the same as the production rate in the Michaelis-Menten reaction scheme (compare Eqs. 71 and 48).

Putting together, Eq. 70 is then

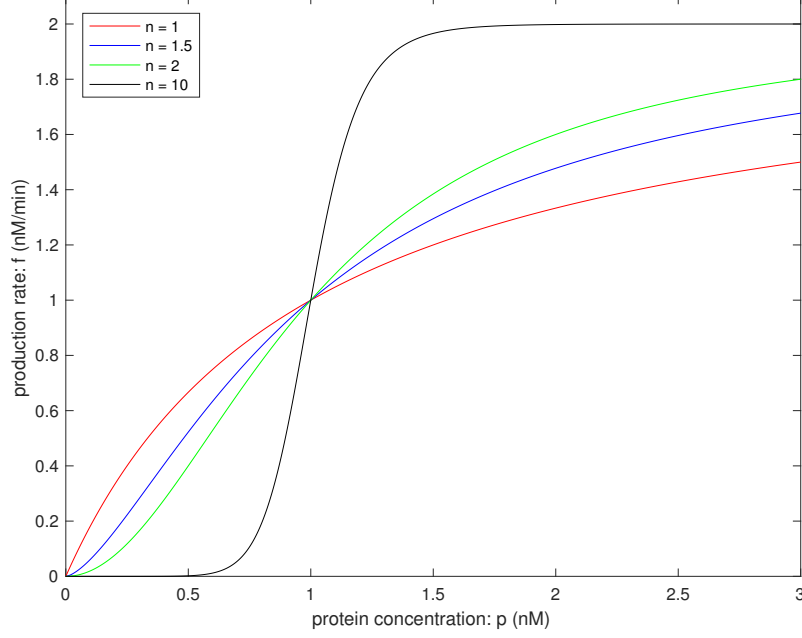


Figure 3. **Production rate for autoregulatory positive feedback (Plot of Eq. 71):** Plot of the production-rate function $f(p)$ given in Eq. 71. Different colored curves denote different values of the Hill coefficient n : Red ($n=1$), blue ($n=1.5$), green ($n=2$) and black ($n=10$). For all curves, $V=2$ nM/min, $K=1$ nM.

$$\frac{dp}{dt} = \frac{Vp^n}{K^n + p^n} - \gamma p \quad (72)$$

There are four constants in Eq. 72 (V , K , n , and γ). Let's reduce this number by making some simplifications. We can do this by choosing to measure the protein concentration p and time t in units other than nanomolar and minutes. Instead, we can measure them in multiples of the parameters that we are trying to eliminate. To see this, we first divide both sides of Eq. 72 by K to obtain

$$\frac{1}{K} \frac{dp}{dt} = \frac{1}{K} \frac{Vp^n}{K^n + p^n} - \gamma(p/K) \quad (73a)$$

$$\implies \frac{d(p/K)}{dt} = \frac{1}{K} \frac{V(p/K)^n}{1 + (p/K)^n} - \gamma(p/K) \quad (73b)$$

Now we see that all instances of p now appear as p/K . This tells us that we can measure p in multiples of K . In other words, it is p/K that is important, not the absolute value of p , for the dynamics of the protein concentration. We define a new variable, $\hat{p} = p/K$. Note that \hat{p} is a **unitless** variable: It does not have any units because both p and K have the same units, the unit of concentration (e.g., nM). But from Eq. 73b, it is not yet clear that we have eliminated the parameter K since it still appears in Eq. 73b even if we replace all appearances of p/K with the new variable \hat{p} . Eliminating the K from our equation was the whole motivation in the first place! Hold your breath, we are not done yet. Remember, we also wanted to measure the time t in a different unit. Motivated by the fact that we almost eliminated K by making p unitless, we attempt to do the same for t - making t unitless as well. To do so, note that γ has units of 1/time. So γt is unitless. The only appearance of t in Eq. 73b is in the dt on the rightside of the equation. Noting that $\gamma dt = d(\gamma t)$ since γ is a constant, we divide both sides of Eq. 73b by γ to obtain

$$\frac{d\hat{p}}{d(\gamma t)} = \frac{1}{\gamma K} \frac{V\hat{p}^n}{1 + \hat{p}^n} - \hat{p} \quad (74)$$

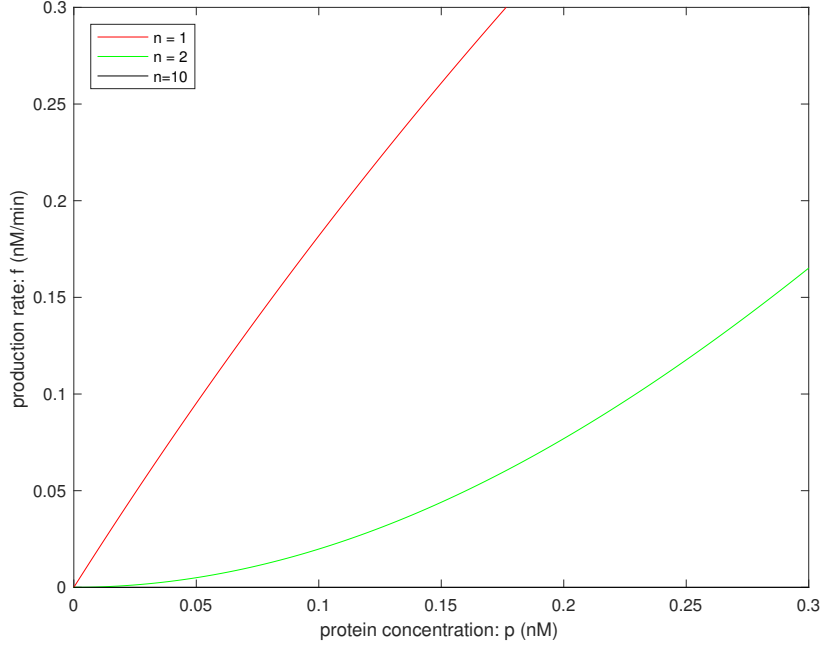


Figure 4. **A close-up view, at low protein concentrations, of the production rate for autoregulatory positive feedback (Plot of Eq. 71):** Plot of the production-rate function $f(p)$ given in Eq. 71 at low values of p (i.e., low protein concentrations). Different colored curves denote different values of the Hill coefficient n : Red ($n=1$), green ($n=2$) and black ($n=10$). For all curves, $V=2$ nM/min, $K=1$ nM. The curve for $n = 10$ is a nearly flat line along the horizontal axis (near $p = 0$).

As we did through introducing \hat{p} , we introduce a new variable $\hat{t} = \gamma t$, which is unitless time, to rewrite Eq. 74 as

$$\frac{d\hat{p}}{d(\hat{t})} = \frac{1}{\gamma K} \frac{V\hat{p}^n}{1 + \hat{p}^n} - \hat{p} \quad (75)$$

Finally, we introduce our final new variable, $\hat{V} = V/(\gamma K)$. \hat{V} is unitless as well because V has unit of concentration/time, γ has unit of 1/time, K has unit of concentration, and thus all the units cancel. We can rewrite Eq. 75 as

$$\frac{d\hat{p}}{d(\hat{t})} = \frac{\hat{V}\hat{p}^n}{1 + \hat{p}^n} - \hat{p} \quad (76)$$

Note that above equation, other than \hat{p} and \hat{t} , there are two constants: \hat{V} , and n . Thus we reduced the number of constants from four in the original equation (Eq. 72) to just two in the simplified equation (Eq. 76). Eq. 76 also contains only unitless parameters: \hat{V} , n , \hat{p} , and \hat{t} . The simplification required us to measure the protein concentration and time in relative units: \hat{p} is the protein concentration in units of K (so $\hat{p} = 1$ means that $p = K$) and \hat{t} is the time measured in units of γ (so $\hat{t} = 1$ means that $t = 1/\gamma$). \hat{V} describes the maximal production rate of the protein in a unitless form.

Let us now analyze Eq. 76. We cannot solve this equation with just a pen and a paper to obtain the protein concentration $\hat{p}(\hat{t})$. But we can still understand how the protein concentration changes over time by plotting the production-rate $f(\hat{p})$ and the degradation-rate $g(\hat{p})$ as functions of \hat{p} . From Eq. 76, we have

$$\frac{d\hat{p}}{d(\hat{t})} = \frac{\hat{V}\hat{p}^n}{1 + \hat{p}^n} - \hat{p} \quad (77a)$$

$$= f(\hat{p}) - g(\hat{p}) \quad (77b)$$

and thus

$$f(\hat{p}) = \frac{\hat{V}\hat{p}^n}{1 + \hat{p}^n} \quad g(\hat{p}) = \hat{p} \tag{78}$$

Note that when $d\hat{p}/dt = 0$, the concentration of the protein does not change over time (i.e., \hat{p} remains constant over time). In other words, this would be the **equilibrium-state gene-expression value**. This occurs if and only if the graphs of $f(\hat{p})$ and $g(\hat{p})$ intersect. There are different ways of these two graphs to intersect depending on the values of the two constants: \hat{V} and n . Let's take a look at these different ways:

Case 1: Monostable (gene expression is stably off) $V=1, n=1$: Plotting the $f(\hat{p})$ and $g(\hat{p})$ (see Fig. 5) we see that the production and degradation rate curves intersect only when $\hat{p}=0$ (note that we relabeled \hat{p} as x in Fig. 5 for convenience). This means that the equilibrium-state of the gene is one in which no protein is produced from it. Note that when $\hat{p} > 0$, then the production rate is lower than the degradation rate (i.e., the $f(\hat{p})$ curve lies below the $g(\hat{p})$ curve in Fig. 5). Thus when the cell has some amount of the protein, the cell will degrade more than it produces the protein, driving the cell's protein concentration towards zero. Thus the gene-expression being zero (i.e., $\hat{p} = 0$) is a **stable equilibrium** - any perturbation away from this equilibrium-state value ($\hat{p} = 0$) returns the protein concentration back to it. Given that there is only one equilibrium-state and that it is stable, we call this a **monostable** case.

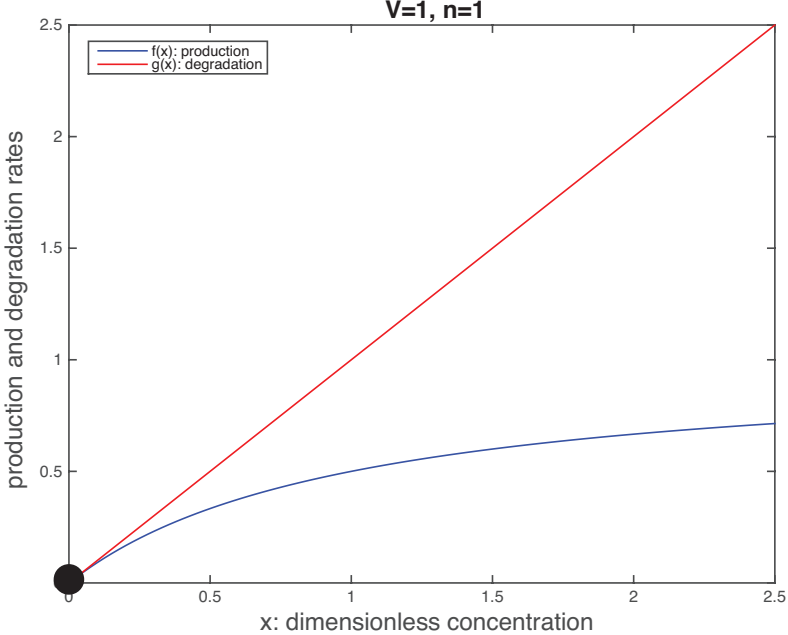


Figure 5. **Monostable (gene expression is stably off) for autoregulatory positive feedback:** Plot of the production-rate function $f(x)$ (blue curve) and degradation-rate function $g(x)$ (red curve) as defined by Eq. 78. For convenience, we have relabeled \hat{p} as x . $\hat{V} = 1$ and $n = 1$. The black circle shows the only point where f and g intersect.

Case 2: Monostable (gene expression is stably on) $V=2, n=1$: Plotting the $f(\hat{p})$ and $g(\hat{p})$ (see Fig. 6) we see that the production and degradation rate curves intersect at two locations: At $\hat{p}=0$ and at $\hat{p} = 1$ (note that we relabeled \hat{p} as x in Fig. 6 for convenience). In Fig. 6, we see that f and g intersect at two locations. Thus there are two equilibrium states when $V = 2$ and $n = 1$. One of the equilibrium state corresponds to the gene being off (i.e., $\hat{p} = 0$, white circle in Fig. 6). The other equilibrium corresponds to gene-expression at a non-zero value, specifically at $\hat{p} = 1$ (black circle in Fig. 6). To determine the stability of these two equilibrium states, we compare which of the two - either f or g - is greater than the other. When \hat{p} is between zero and one, we see that the production rate f is higher than the degradation rate g (i.e., the blue curve lies above the red curve in Fig. 6). Hence, when the protein concentration is a little above zero (i.e., $0 < \hat{p} < 1$), the more protein is produced than degraded, causing

the cell increase the protein concentration even more (i.e., $d\hat{p}/dt > 0$ in Eq. 77b). Thus the protein concentration does not return to zero once the cell has even a few copies of the protein inside it. Hence we call this equilibrium state to be **unstable equilibrium**. When $\hat{p} > 1$, we see that the degradation rate is higher than the production rate (i.e., the blue curve lies below the red curve in Fig. 6) and thus the protein concentration would decrease towards $\hat{p} = 1$ (i.e., $d\hat{p}/dt < 0$ in Eq. 77b). When $\hat{p} < 1$, we see that the degradation rate is lower than the production rate (i.e., the blue curve lies above the red curve in Fig. 6) and thus the protein concentration would increase towards $\hat{p} = 1$ (i.e., $d\hat{p}/dt < 0$ in Eq. 77b). Thus any perturbation away from this equilibrium state, $\hat{p} = 1$, causes the protein concentration to return to $\hat{p} = 1$. For this reason, like in case 1 above, we say that this equilibrium state is a stable equilibrium and that the gene expression can be stably on. Gene expression being "on" means that a non-zero value of the protein concentration can be stably maintained inside the cell. We still have a monostable case here because there is only one stable equilibrium state even though there are two equilibrium states.

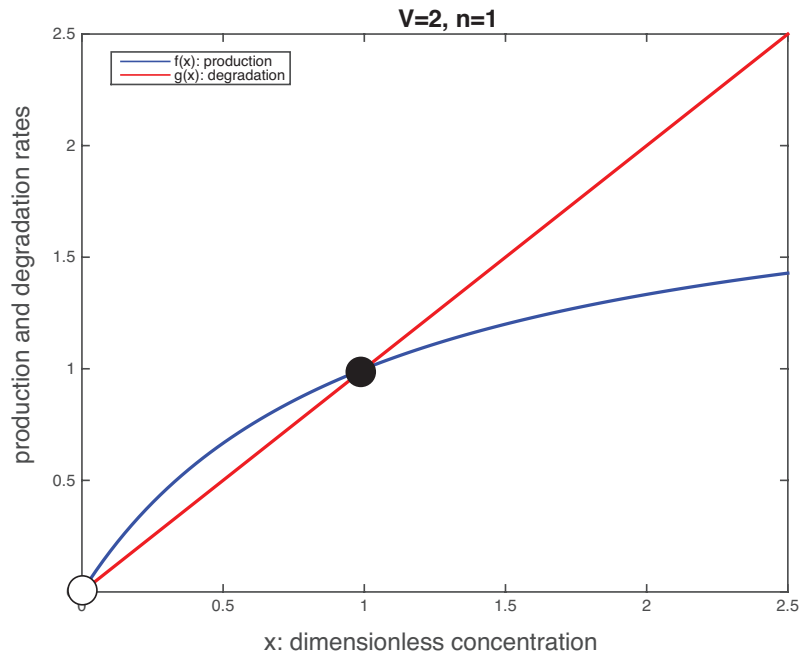


Figure 6. **Monostable (gene expression is stably on) for autoregulatory positive feedback:** Plot of the production-rate function $f(x)$ (blue curve) and degradation-rate function $g(x)$ (red curve) as defined by Eq. 78. For convenience, we have relabeled \hat{p} as x . $\hat{V} = 2$ and $n = 1$. The black point, where f and g intersect, is a stable equilibrium point. The white point, where f and g intersect, is an unstable equilibrium point.

Case 3: Bistable (gene expression can be both stably off and stably on) $V=5, n=2$: Plotting the $f(\hat{p})$ and $g(\hat{p})$ (see Fig. 7) we see that the production and degradation rate curves now intersect at three locations: At $\hat{p}=0$, $\hat{p}=0.5$, and $\hat{p} = 2$ (note that we relabeled \hat{p} as x in Fig. 7 for convenience). Thus there are three equilibrium states here. Using the argument given in case 2, we can deduce that $\hat{p}=0$ and $\hat{p} = 2$ are stable equilibria whereas $\hat{p} = 0.5$ is an unstable equilibrium. Thus, in this case, the gene expression can be stably off (i.e., $\hat{p} = 0$) or stably on (i.e., $\hat{p} = 2$). Using the style of argument given in case 2, we can further deduce that when the protein concentration \hat{p} is above 0.5 (the unstable equilibrium value), then the gene expression turns on (i.e., \hat{p} evolves to and settles down at two) whereas if it is below 0.5, then the gene expression turns off (i.e., \hat{p} evolves to and settles down at zero). For this reason, we call this a **genetic switch**: Like a light switch, the gene can either be off or on. The unstable equilibrium acts as the mid-position in a light switch. It is very difficult to put the light switch in the mid-position and any perturbation (little push) on the light switch at the mid-position will lead to either switching on or off the light. In retrospect, we also see that this is why we mentioned earlier that as the Hill coefficient increases, the more switch-like the gene-regulation becomes. We used $n = 2$ here whereas in cases 1 and 2, the Hill coefficient was at its lowest possible value, $n = 1$.

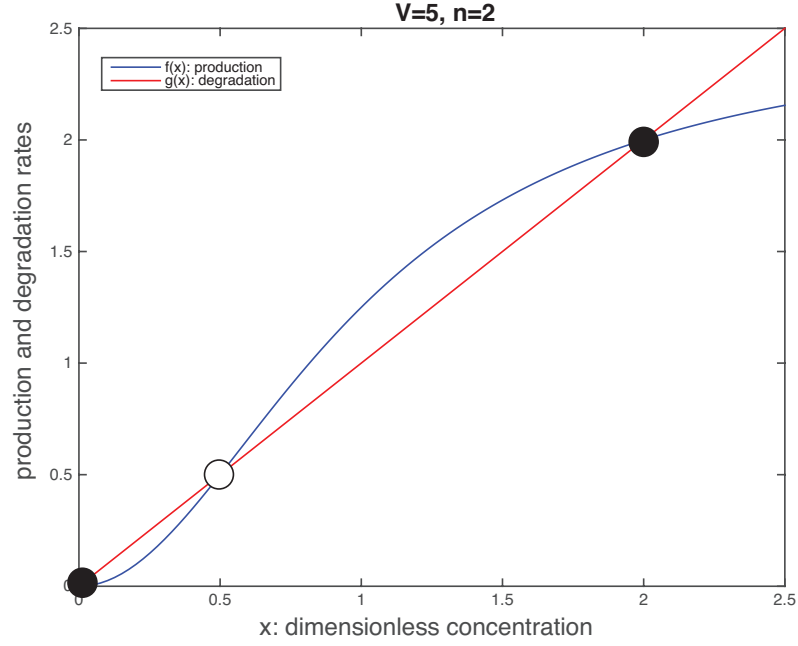


Figure 7. **Bistable gene-expression (gene expression can be stably off and stably on) for autoregulatory positive feedback:** Plot of the production-rate function $f(x)$ (blue curve) and degradation-rate function $g(x)$ (red curve) as defined by Eq. 78. For convenience, we have relabeled \hat{p} as x . $\hat{V} = 5$ and $n = 2$. The two black points, where f and g intersect, are stable equilibrium points. The white point, where f and g intersect, is an unstable equilibrium point.

IV. SOLUTIONS TO THE EXERCISES

A. Solutions to the exercises in section I

Solution to problem 1:

$$\begin{aligned}
 \frac{dz}{dx} + 5x &= 0 \\
 \implies dz &= -5x dx \\
 \implies \int dz &= - \int 5x dx \\
 \implies z - z_0 &= -5 \frac{x^2}{2} - a \quad \text{where } z_0 \text{ and } a \text{ are constants.} \\
 \implies z(x) &= -\frac{5x^2}{2} + c
 \end{aligned}$$

where c is a new constant, $c = z_0 + a$, and it can be any number.

Solution to problem 2: $y(t) = -\frac{at^3}{3} - bt + c$, where c is an arbitrary constant.

Solution to problem 3: $x(z) = -e^z + \frac{z^2}{2} + c$, where c is an arbitrary constant.

Solution to problem 4: $x(t) = 3t - \ln(t) + c$, where c is an arbitrary constant. Note that we must have $t > 0$ since $\ln(t)$ is undefined when $t < 0$.

Solution to problem 5: $x(t) = \pm\sqrt{c - bt^2}$, where c is an arbitrary constant.

Solution to problem 6: $k(t) = 3 - ce^{-t}$, where c is an arbitrary constant.

B. Solutions to the exercises in section II

Solution to problem 1.

$$\frac{d[\text{sucrose}]}{dt} = -k_f[\text{sucrose}][\text{invertase}] + k_r[\text{sucrose}::\text{invertase}] \quad (80a)$$

$$\frac{d[\text{invertase}]}{dt} = -k_f[\text{sucrose}][\text{invertase}] + k_r[\text{sucrose}::\text{invertase}] + k[\text{sucrose}::\text{invertase}] \quad (80b)$$

$$\frac{d[\text{sucrose}::\text{invertase}]}{dt} = k_f[\text{sucrose}][\text{invertase}] - k_r[\text{sucrose}::\text{invertase}] - k[\text{sucrose}::\text{invertase}] \quad (80c)$$

$$\frac{d[\text{glucose}]}{dt} = k[\text{sucrose}::\text{invertase}] \quad (80d)$$

$$\frac{d[\text{fructose}]}{dt} = k[\text{sucrose}::\text{invertase}] \quad (80e)$$

Solution to problem 2. Follow the derivation given in section II step-by-step.

Solution to problem 3. Let f_1 be the production rate when the substrate concentration is $[S_1]$ and f_2 be the production rate when $[S_2]$ is the substrate concentration. Both f_1 and f_2 are given by the Michaelis-Menten kinetics equation (Eq. 48). We let $[S_2] = 5000$ nM and $[S_1] = 5$ nM. Note that $[S_2]/K = 10000$ and $[S_1]/K = 10$. Then,

$$\frac{f_2}{f_1} = \frac{V[S_2]}{K + [S_2]} \frac{K + [S_1]}{V[S_1]} \quad (81a)$$

$$= \frac{[S_2]}{[S_1]} \frac{K + [S_1]}{K + [S_2]} \quad (81b)$$

$$= 1000 \frac{K + 10K}{K + 10000K} \quad (81c)$$

$$= 1000 \frac{11K}{10001K} \quad (81d)$$

$$= \frac{11000}{10001} \quad (81e)$$

$$\approx 1.1 \quad (81f)$$

Thus despite increasing the substrate concentration by 1000 times (from 5 nM to 5000 nM), the production rate only changes by 1.1 times - it barely changes at all. This occurs because the Michaelis-Menten kinetics is only sensitive to substrate concentrations that are close to the half-saturation concentration K , which is 0.5 nM here. (Fig. 1).