AP3162D: Lecture 2 - Gene-regulatory circuits: Deterministic dynamics

Hyun Youk
Delft University of Technology

(Dated: February 22, 2018)

In this lecture, we will discuss how to model standard gene-regulatory schemes - constitutive gene expression, positive feedback, and negative feedback. We will analyze the deterministic aspects and begin to study their stochastic behaviors.

I. BASICS OF GENE REGULATION

Gene expression is the process of making RNA and protein from a gene (Fig. 1). 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. 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 multiple proteins), govern transcription and translation. One of these macromolecular machines, called **RNA polymerase** lands on DNA and slides over the sequence that encodes a gene to transcribe mRNA. Specifically, as the RNA polymerase slides over the gene, letter-by-letter (A, T, C, G), it grabs a single base (i.e., a single letter), called a **nucleotide**, 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 encoded by the gene. Once this transcription of mRNA is complete, another macromolecular machine, called the **ribosome**, lands on the mRNA and slides over it, from start to end, to synthesize a protein that the mRNA encodes (and thus the gene encodes). As the ribosome runs through the mRNA, nucleotide by nucleotide, 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 *three* consecutive nucleotides of mRNA - called a **codon**. 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 yet 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 is a dynamic process in which the cell modulates its gene expression - the cell dynamically modifies the amounts of mRNA and protein according to some rule. As we will see, different rules can regulate gene expression. We are usually interested in the concentration of the gene's final product, the protein, as a function of time. Most regulations of a single gene lead to an equation of the following type:

$$\frac{dx}{dt} = F - G \tag{1}$$

where x is the concentration of a protein-X (made from gene X), F and G are non-negative-valued functions that can depend on x and the concentrations of proteins translated from other genes. F is a "**production-rate function**" that describes how much protein is made per unit time and G is a "**Degradation-rate function**" that describes how much protein is degraded per unit time. If we are thus interested in genes X, Y, and Z, for example, we would

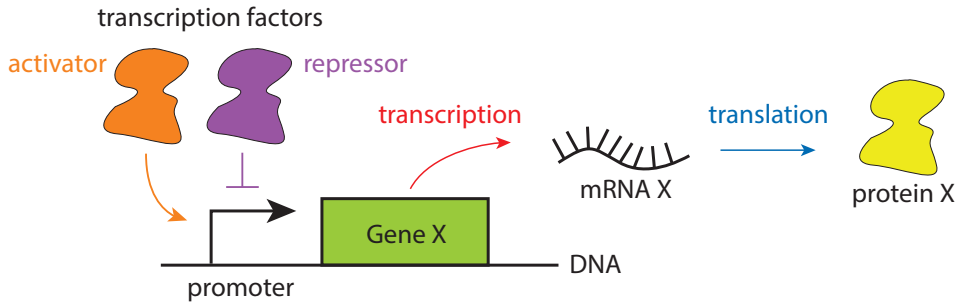

Figure 1. **Central dogma of molecular biology**: A gene (green box) is a sequence of DNA (black horizontal line) and a promoter is the DNA sequence in front of the gene. The promoter is a "molecular dial" that controls how much mRNA (and thus protein) is transcribed from the gene. During translation of a protein from an mRNA, the mRNA is *not* destroyed. Thus, the cell can translate multiple copies of a protein from a single mRNA molecule. Broadly speaking, two types of proteins (transcription factors), called activators and repressors, regulate the transcription of a gene. An activator binds to the promoter to increase the transcription rate (more mRNA produced per unit time) while a repressor binds to the promoter to decrease the transcription rate (less mRNA produced per unit time). A regular arrow (orange arrow) indicates an activation of a gene (the action of the activator) while a blunt-ended arrow (purple arrow) indicates a repression of a gene (the action of the repressor).

write down an equation of the following form:

$$\frac{dx}{dt} = F_1 - G_1 \tag{2a}$$
$$\frac{dy}{dt} = F_2 - G_2 \tag{2b}$$
$$\frac{dz}{dt} = F_3 - G_3 \tag{2c}$$
$$\tag{2d}$$

where F_1, F_2, and F_3 are production-rate functions for proteins X, Y, and Z, respectively, and G_1, G_2, and G_3 are degradation-rate functions for proteins X, Y, and Z, respectively. Note that Eqs. 2a - 2c can be coupled amongst each other. Moreover, note that if we have N genes, then we would have N equations that may be potentially coupled to each other. This already tells us why analyzing more than a few genes is difficult. As we will see, even for a regulation scheme of a single gene, the equation (Eq. 1) is complicated enough that we cannot analytically solve it. Yet, we We now turn to analyzing equations of this type for some commonly occurring gene regulatory schemes.

II. GENE-REGULATION SCHEME 1: CONSTITUTE GENE EXPRESSION

At its simplest, gene expression can be captured by the following process:

$$DNA_X + \text{RNA polymerase} \rightarrow \text{RNA polymerase} + DNA_X + mRNA_X \tag{3a}$$
$$mRNA_X + \text{Ribosome} \rightarrow \text{Ribosome} + mRNA_X + protein_X, \tag{3b}$$

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. 3a) and k_p be the translation rate (i.e., rate associated with the chemical reaction in Eq. 3b). Note that DNA and the RNA polymerase are not consumed in this process shown in Eq. 3a. Thinking of RNA polymerase as an enzyme, this makes sense - like any enzyme, it is not destroyed in any reactions. Note that DNA also is not destroyed 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. 3a 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. 3b, 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

$$mRNA_X \rightarrow \emptyset \tag{4a}$$
$$protein_X \rightarrow \emptyset, \tag{4b}$$

where $\emptyset$ represents "nothingness" (i.e., symbol for null). Let γ_m be the constant degradation rate of the $mRNA_X$ (reaction in Eq. 4a) and γ_p be the constant degradation rate of the $protein_X$ (reaction in Eq. 4b). 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, the above four chemical kinetics equations - Eqs. 3a- 4b - yield the following equations:

$$\frac{dm}{dt} = k_m[DNA_X][\text{RNA polymerase}] - \gamma_m m \tag{5a}$$
$$\frac{dp}{dt} = k_p m[\text{Ribosome}] - \gamma_p p \tag{5b}$$

where $m = [mRNA_X]$ and p $= [protein_X]$. Since $[DNA_X]$, [RNA polymerase] and [Ribosome] do not change over time, we can simplify Eqs. 5a and 5b by defining new constants: $r_m = k_m[DNA_X][\text{RNA polymerase}]$ and $r_p = k_p[\text{Ribosome}]$. Then Eqs. 5a and 5b become

$$\frac{dm}{dt} = r_m - \gamma_m m \tag{6a}$$
$$\frac{dp}{dt} = r_p m - \gamma_p p \tag{6b}$$

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, we can assume that the mRNA concentration quickly reaches a steady-state value before any protein is translated from it. This is called a **quasi steady-state assumption**. Then Eqs. 6a and 6b become

$$\frac{dm}{dt} = r_m - \gamma_m m \tag{7a}$$
$$\frac{dp}{dt} = r_p m_{eq} - \gamma_p p, \tag{7b}$$

where m_{eq} is the constant, equilibrium-state concentration of the mRNA. Now note that Eqs. 7a and 7b have mathematically the same form but with different variables. Let us solve for m_{eq}. By definition of being a steady state, we must have $dm/dt = 0$ when $m = m_{eq}$. Then Eq. 7a gives us

$$m_{eq} = \frac{r_m}{\gamma_m}, \tag{8}$$

and thus Eq. 7b becomes

$$\frac{dp}{dt} = \frac{r_p r_m}{\gamma_m} - \gamma_p p, \tag{9}$$

Eq. 8 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. 7a, we solve for the simpler case in which there is no production. In this case, Eq. 7a becomes

$$\frac{dm_1}{dt} = -\gamma_m m_1, \tag{10}$$

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. 10 yields

$$m_1(t) = Ae^{-\gamma_m t} \tag{11}$$

where A is a constant whose value we will later determine. Next, we look for a particular solution m_2 for the original equation, Eq. 7a:

$$\frac{dm_2}{dt} = r_m - \gamma_m m_2, \tag{12}$$

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. 12. To do so, we note that if $m_2 =$ constant is a solution to Eq. 12). In particular, the following is a solution to Eq. 12):

$$m_2 = \frac{r_m}{\gamma_m} \tag{13}$$

In fact, this m_2 is the steady 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. 10, and m_2 is the solution to the original equation, Eq. 12. If we add dm_1/dt to dm_2/dt, note that we obtain

$$\frac{d(m_1 + m_2)}{dt} = r_m - \gamma_m (m_1 + m_2), \tag{14}$$

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

$$\frac{dm}{dt} = r_m - \gamma_m m, \tag{15}$$

which gives us back the original equation that we wanted to solve, Eq. 7a. $m(t)$ is the most **general solution** to the original equation, Eq. 7a:

$$m(t) = m_1(t) + m_2(t) \tag{16a}$$

$$= \frac{r_m}{\gamma_m} + Ae^{-\gamma_m t} \tag{16b}$$

$m(t)$ contains an undefined constant, A, which is related to the concentration of mRNA at $t = 0$:

$$m(0) = \frac{r_m}{\gamma_m} + A \tag{17}$$

Under the quasi-steady state assumption, dp/dt (Eq. 9) takes the same form as dm/dt (Eq. 15). 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. 16b to obtain the most general solution for $p(t)$:

$$p(t) = \frac{r_p r_m}{\gamma_m} + Be^{-\gamma_p t}, \tag{18}$$

where B is a constant that is determined by the initial condition on p (i.e., $p(0)$). Thus, like Eq. 17, we have

$$p(0) = \frac{r_p r_m}{\gamma_m} + B \tag{19}$$

Now that we have deduced $m(t)$ and $p(t)$, 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. 17 and B through Eq. 19 to convert Eqs. 16b and 18 into

$$m(t) = \frac{r_m}{\gamma_m}\left(1 - e^{-\gamma_m t}\right) \tag{20a}$$

$$p(t) = \frac{r_p r_m}{\gamma_m}\left(1 - e^{-\gamma_p t}\right) \tag{20b}$$

Note that the equlibrium-state of the protein concentration, which we get by setting $dpdt = 0$, is $\frac{r_p k_m}{\gamma_m}$. 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. 20a and 20b 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. 6a and 6b into Eqs. 7a and 7b. So our solutions, Eqs. 20a and 20b, 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. 20a and 20b, 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 values (i.e. when the curves in Fig. 2 flatten out). This is because of the time t in Eqs. 20a and 20b 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. 20a and 20b - would be invalid. The quasi-steady-state assumption mathematically means that γ_m is much larger than γ_p.

Validity of the quasi steady-state approximation: Can we really use the quasi steady-state approximation in real cells? To answer this question, we need to look up "typical" values of degradation and production rates of mRNA and proteins. To look up the typical numbers such as the typical time for a cell to replicate, the number of bases in the human genome, and other useful numbers, you can look at www.BioNumbers.org. Alternatively, you can see: "Key numbers in biology" (Moran, Phillips, and Milo, *Cell* (2010)). The typical life-time of mRNA is indeed shorter than the typical life-time of protein in many cases (see Table I). By "life time", we mean "half life". Half life is the time taken for the concentration of a molecule to decrease to half of its current value.

Now, what about half-life of proteins? Many proteins, including the fluorescent proteins such as the GFP (Green Fluorescent Protein) that researchers often use to measure gene expression in a cell, are very stable. This means that if the cell were not to divide and yet live happily, the GFP does not degrade because there is no active mechanism in a cell that would degrade it. The way for the concentration to decrease to half its current value is if the cell divides. In this process, the mother (original) cell divides into two cells (i.e., the mother gives rise to a daughter cell), and in this process the mother cell must give some of its proteins to the daughter cell. For most biomolecules, the mother

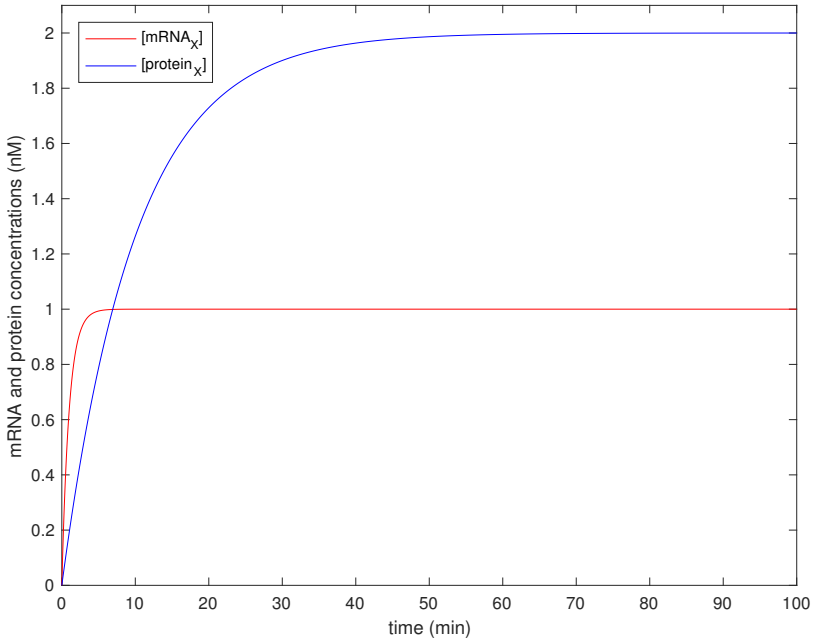

Figure 2. **mRNA and protein concentrations for a constitutive gene expression (Plot of Eqs. 20a and 20b)**: Red curve is [mRNA] as a function of time (i.e., $m(t)$ in Eq. 20a) and blue curve is [protein] as a function of time (i.e., $p(t)$ in Eq. 20b). Values for the parameters are: $r_m = 1$ nM/min, $r_p = 2$ min^{-1}, $\gamma_m = 1$ min^{-1}, and $\gamma_p = 0.1$ min^{-1}. Here nM means nanomolar (nano $= 10^{-9}$).

Organism	half-life of typical mRNA
E. coli	5 minutes
Budding yeast (*Saccharomyces cerevisiae*)	20 minutes
Human cell	600 minutes

Table I. From: Moran, Phillips, and Milo, *Cell* (2010)

cell has no way of counting or regulating many proteins she gives to her daughter. Our differential equations describe the average concentrations inside a cell. On average, we expect that the mother to get the half of the proteins and the daughter to inherit the remaining half of the proteins. Thus the degradation rate γ_p is set by the doubling time of the cell. We can calculate γ_p in terms of the doubling time τ. The doubling time is the average time taken for a cell to grow in size (approximately twice its initial size) and then divide into two, thus going from one cell to two cells. Let $N(t)$ be the integer number of protein X inside a cell at time t and $V(t)$ be the volume of the cell at time t. $t = 0$ is the beginning of a cell cycle (for the daughter cell, its cell cycle is just after it is born). Then the concentration inside the cell is $p(t) = N(t)/V(t)$. Thus, we have

$$\frac{dp}{dt} = \frac{d(N/V)}{dt} \tag{21a}$$

$$= \frac{1}{V}\frac{dN}{dt} - \frac{N}{V}\frac{1}{V}\frac{dV}{dt} \tag{21b}$$

Confining ourselves to the inside of just one cell, we see that the volume changes over time as

$$V(t) = V_0 2^{t/\tau} \tag{22a}$$

$$= V_0 e^{\mu t} \tag{22b}$$

where V_0 is the volume of the cell at $t = 0$ and $\mu = ln(2)/\tau$ is the cell's growth rate. Then computing dV/dt, we find that dp/dt becomes

$$\frac{dp}{dt} = \frac{1}{V}\frac{dN}{dt} - \mu p \tag{23}$$

Assuming that a constant translation rate k_N for the absolute (integer) number of proteins (i.e., $dN(t)/dt = kN$), above equation becomes

$$dp/dt = k_N/V - \mu p \tag{24}$$

Eq. 24 is identical to Eq. 9 if the production rate in terms of concentration k_N/V were constant. But note that since the cell volume can double within the doubling time (i.e., cell cycle time), the production rate is not constant over time. The production rate k_N/V can decrease by as much half. But we ignore this and say that the production rate is constant (i.e., replace the k_N/V by a constant) and set $\gamma_p = \mu$. In this sense, we have just derived Eq. 9 in a different way. The important point here is that for very stable proteins, if you know the doubling time of the cell, then you know the dilution rate (i.e., degradation rate) of the highly stable protein in the equation for dp/dt. The typical values for cell division times are shown in Table II.

Organism	Typical cell-division (doubling) time
E. coli	30 minutes
Budding yeast (*Saccharomyces cerevisiae*)	90 minutes
Human cell	3000 minutes

Table II. Typical doubling times of cells

According to Eqs. 20a and 20b, the times taken for the mRNA concentration and the protein concentration to reach their steady-state values is determined only by γ_m and γ_p, respectively. Specifically, the time for the mRNA concentration to reach its steady-state value is roughly $1/\gamma_m$ and the time for the protein concentration to reach its steady-state value is roughly $1/\gamma_p$. Thus the ratio of $1/\gamma_m$ to $1/\gamma_p$ tells us how must faster mRNA concentration reaches its steady-state value compared to the protein concentration. Using Tables I and II, we can compute these values (see Table III). We see that in E. coli, within 5 minutes the mRNA concentration reaches a steady-state value while the protein concentration only reaches its steady-state value in about 30 minutes (so roughly speaking, for the rest of the 25 minutes the protein concentration really is acting on a constant concentration of mRNA) (Table III). Thus, the quasi steady-state approximation is valid for "typical" mRNA and proteins in E. coli. In fact, if $\gamma_p/\gamma_m \ll 1$, then the quasi steady-state approximation is valid.

Organism	γ_p/γ_m
E. coli	0.17
Budding yeast (*Saccharomyces cerevisiae*)	0.22
Human cell	0.20

Table III. If $\gamma_p/\gamma_m \ll 1$, the quasi steady-state approximation is valid.

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. 6a, 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. 8). Looking at Eq. 7b, the same argument works to explain why the protein concentration reaches an equilibrium-state value.

III. GENE-REGULATION SCHEME II: AUTOREGULATORY POSITIVE FEEDBACK

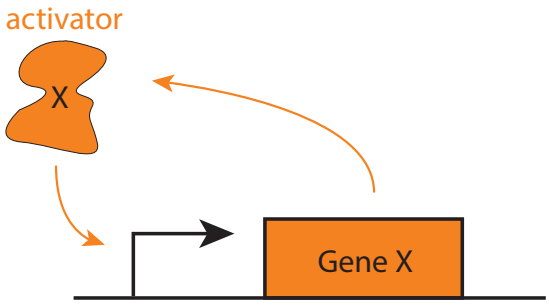

Figure 3. **Auto-regulatory positive feedback loop**: A gene X encodes a protein that is an activator for the transcription of itself. Thus, the more X there is inside the cell, the higher the production rate of X.

We now consider a **positive feedback loop** in which a gene produces a protein that promotes the production of itself (Fig. 3). 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. 1, the basic equation that describes the protein concentration, denoted p, is

$$\frac{dp}{dt} = f(m,p) - g(p) \tag{25}$$

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

$$\frac{dp}{dt} = f(p) - g(p) \tag{26}$$

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. The production-rate function $f(p)$ for a positive feedback loop typically takes the following form:

$$f(p) = \frac{Vp^n}{K^n + p^n}, \tag{27}$$

where V, K, and n are positive constants. $f(p)$ is a phenomenological equation - it is not derived from anything. It simply has the property that f increases as p increases, and that it is non-linear in p. This non-linearity in f as a function of p allows for rich behaviours as we will see. The features that f embodies makes it widely used for modeling positive feedback (and it fits data too in many instances!). For instance, note that when p approaches infinity, f approaches V. So V is the maximal production rate. When $p = K$, f is $V/2$. We thus 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. 4). In Fig. 4, 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. 5). When $n = 1$, $f(p)$ is exactly the same as the production rate in the well-known Michaelis-Menten reaction scheme.

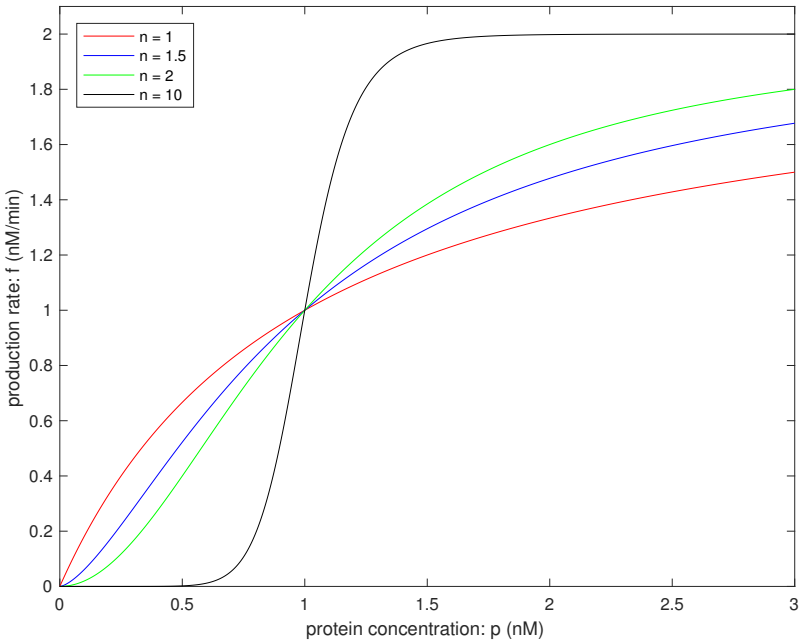

Figure 4. **Production rate for autoregulatory positive feedback (Plot of Eq. 27**: Plot of the production-rate function $f(p)$ given in Eq. 27. Different colored curves denote different values of the Hill coefficient n: Red (n=1), blue (n=1.5), green (n=2) and black (n=10). For all curves, V=2 nM/min, K=1 nM.

Putting together, Eq. 26 is then

$$\frac{dp}{dt} = \frac{Vp^n}{K^n + p^n} - \gamma p \tag{28}$$

There are four constants in Eq. 28 (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. 28 by K to obtain

$$\frac{1}{K}\frac{dp}{dt} = \frac{1}{K}\frac{Vp^n}{K^n + p^n} - \gamma(p/K) \tag{29a}$$

$$\implies \frac{d(p/K)}{dt} = \frac{1}{K}\frac{V(p/K)^n}{1 + (p/K)^n} - \gamma(p/K) \tag{29b}$$

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. 29b, it is not yet clear that we have eliminated the parameter K since it still appears in Eq. 29b 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,

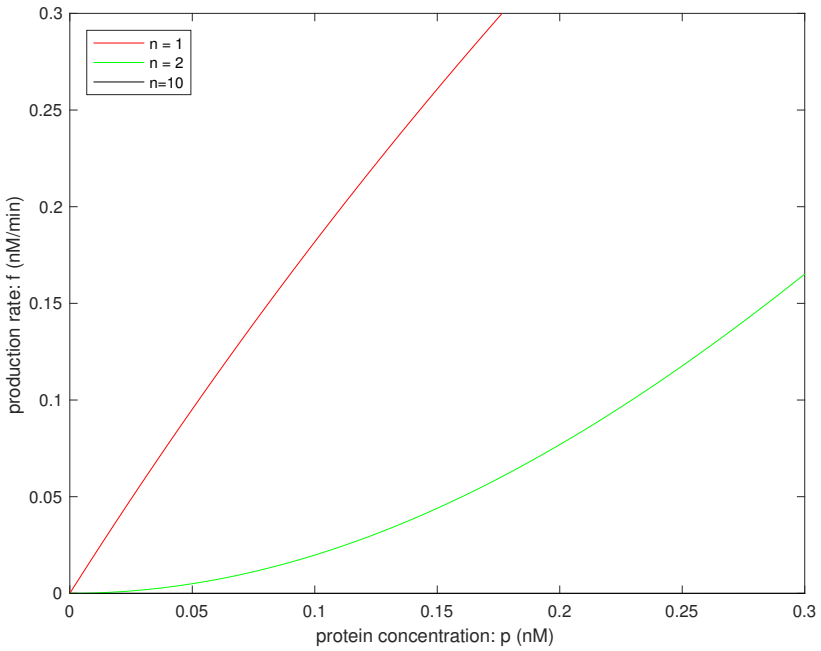

Figure 5. **A close-up view, at low protein concentrations, of the production rate for autoregulatory positive feedback (Plot of Eq. 27**: Plot of the production-rate function $f(p)$ given in Eq. 27 at low values of p (i.e., low protein concentrations). Different colored curves denote different values 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 $p = 0$).

note that γ has units of 1/time. So γt is unitless. The only appearance of t in Eq. 29b 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. 29b by γ to obtain

$$\frac{d\hat{p}}{d(\gamma t)} = \frac{1}{\gamma K}\frac{V\hat{p}^n}{1+\hat{p}^n} - \hat{p} \tag{30}$$

As we did through introducing $\hat{p}$, we introduce a new variable $\hat{t} = \gamma t$, which is unitless time, to rewrite Eq. 30 as

$$\frac{d\hat{p}}{d(\hat{t})} = \frac{1}{\gamma K}\frac{V\hat{p}^n}{1+\hat{p}^n} - \hat{p} \tag{31}$$

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. 31 as

$$\frac{d\hat{p}}{d(\hat{t})} = \frac{\hat{V}\hat{p}^n}{1+\hat{p}^n} - \hat{p} \tag{32}$$

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. 28) to just two in the simplified equation (Eq. 32). Eq. 32 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. 32. We cannot solve this equation with just a pen and a paper to obtain the protein concentration $\hat{p}(t)$. 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. 32, we have

$$\frac{d\hat{p}}{d(\hat{t})} = \frac{\hat{V}\hat{p}^n}{1+\hat{p}^n} - \hat{p} \tag{33a}$$

$$= f(\hat{p}) - g(\hat{p}) \tag{33b}$$

and thus

$$f(\hat{p}) = \frac{\hat{V}\hat{p}^n}{1+\hat{p}^n} \qquad g(\hat{p}) = \hat{p} \tag{34}$$

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 a **steady-state gene expression**. 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. 6) 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. 6 for convenience). This means that the steady-state gene-expression 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. 6). 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 steady state value ($\hat{p} = 0$) returns the protein concentration back to it. Given that there is only one steady-state and that it is stable, we call this a **monostable** case.

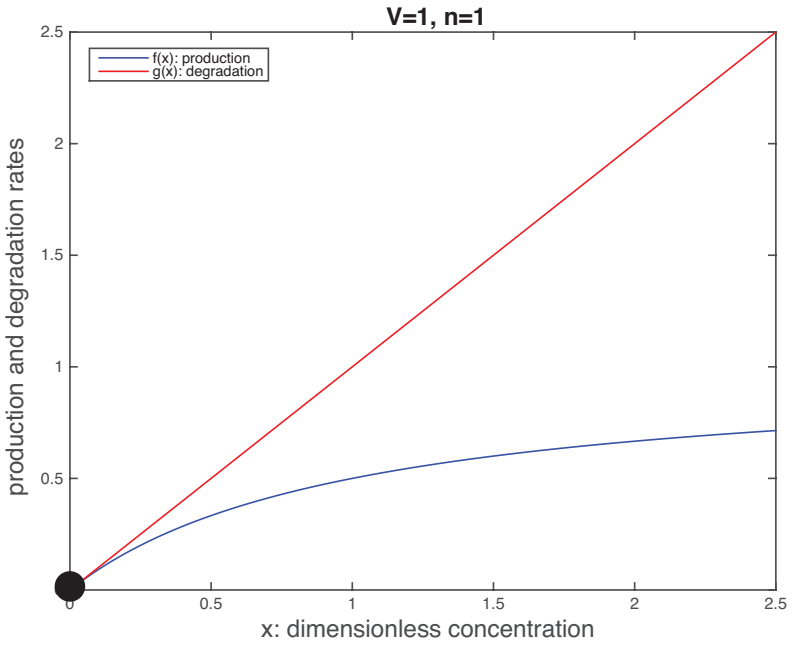

Figure 6. **Monostable (gene expression is stably off) for autoregulatory positive feedback**: Plot of the production-rate function $f(x)$ (blue curve) and degradation-rate function $g(x)$ (red curve) as defined by Eq. 34. For convenience, we have relabeled $\hat{p}$ as x. $\hat{V} = 1$ and $n = 1$. 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. 7) 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. 7 for convenience). In Fig. 7, we see that f and g intersect at two locations. Thus there are two fixed points when $V = 2$ and $n = 1$. One of the steady states corresponds to the gene being off (i.e., $\hat{p} = 0$, white circle in Fig. 7). The other steady state corresponds to gene-expression at a non-zero value, specifically at $\hat{p} = 1$ (black circle in Fig. 7). To determine the stability of these two steady 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. 7). 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. 33b). 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. 7) and thus the protein concentration would decrease towards $\hat{p} = 1$ (i.e., $d\hat{p}/dt < 0$ in Eq. 33b). 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. 7) and thus the protein concentration would increase towards $\hat{p} = 1$ (i.e., $d\hat{p}/dt < 0$ in Eq. 33b). 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 steady 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.

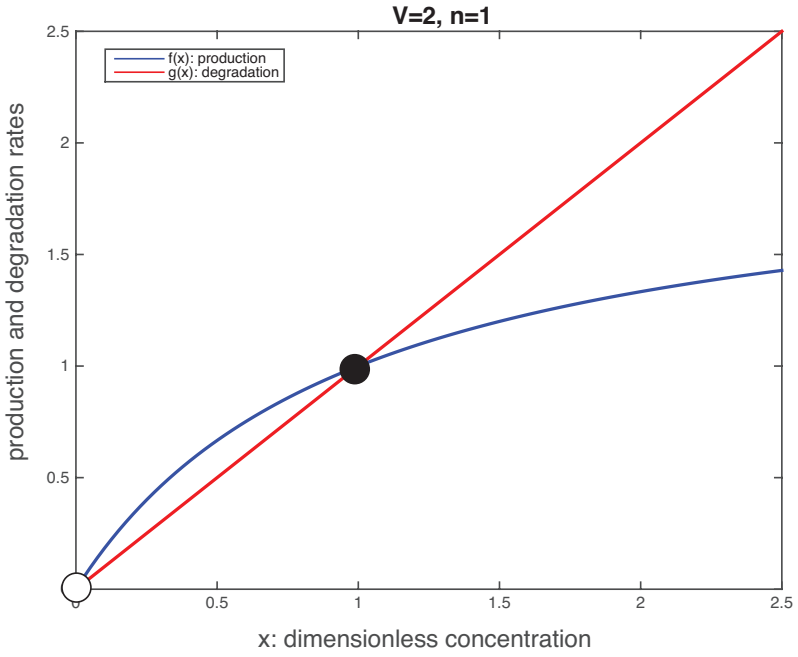

Figure 7. **Monostable (gene expression is stably on) for autoregulatory positive feedback**: Plot of the production-rate function $f(x)$ (blue curve) and degradation-rate function $g(x)$ (red curve) as defined by Eq. 34. For convenience, we have relabeled $\hat{p}$ as x. $\hat{V} = 2$ and $n = 1$. The 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. 8) 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. 8 for convenience). Thus there are three steady states here. Using the argument given in case 2, we can deduce that $\hat{p}$=0 and $\hat{p} = 2$ are stable equilibria where as $\hat{p} = 0.5$ is an unstable fixed point. 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 turns "on" (i.e., $\hat{p}$ evolves to and settles down at zero) whereas if it is below 0.5, then the gene turns "off" (i.e., $\hat{p}$ evolves to and settles down at two). 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, $n = 1$.

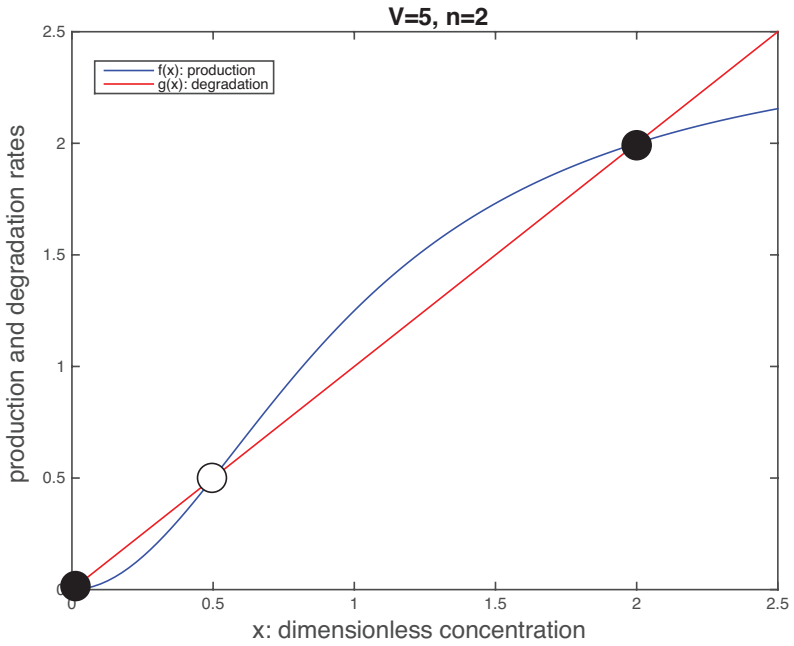

Figure 8. **Bistable gene-expression (gene expression can be stably off and stably on) for autoregulatory positive feedback**: Plot of the production-rate function $f(x)$ (blue curve) and degradation-rate function $g(x)$ (red curve) as defined by Eq. 34. For convenience, we have relabeled $\hat{p}$ as x. $\hat{V} = 5$ and $n = 2$. The two black points, where f and g intersect, are stable fixed points. The white point, where f and g intersect, is an unstable fixed point.

IV. GENE-REGULATION SCHEME III: NEGATIVE FEEDBACK

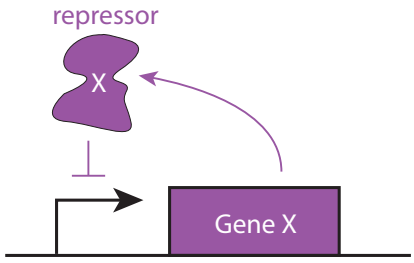

Figure 9. **Auto-regulatory negative feedback loop**: A gene X encodes a protein that is a repressor for the transcription of itself. Thus, the more X there is inside the cell, the smaller the production rate of X.

Just as in the case of positive feedback, we use a phenomenological equation to describe the opposite of positive feedback loop, namely the negative feedback loop. Here, the idea is that a gene X produces a protein X that **represses** its own production (Fig. 9). We call such a protein a **repressor**. Activator (in positive feedback) and repressor (in negative feedback) are both called **transcription factors** because they regulate the transcription of a gene (in their case, they regulate themselves). The idea behind the negative feedback is that the more protein X there is, the more of it will bind to the promoter of gene X. The more protein X is bound to the promoter of gene X, the lower the transcription rate, and thus translation rate of X. The phenomenological equation that we use to model negative feedback is

$$\frac{dp}{dt} = \frac{V}{1+p^n} - p \tag{35}$$

where we have already taken the liberty of using the normalized units of time and concentration that we introduced for the positive feedback scenario (here we have dropped the "hats" on top of t and p for convenience). Note that now, the production rate of the protein is maximum (V) when $p = 0$, which makes sense for the negative feedback scheme described above. As the p increases, we see that the production rate decreases. As p approaches infinity, we see that the production rate goes to zero. Note that Eq. 35 enables only one fixed point and that it is a stable fixed point (you can see this graphically by plotting the production rate and the degradation rate together as functions of p).

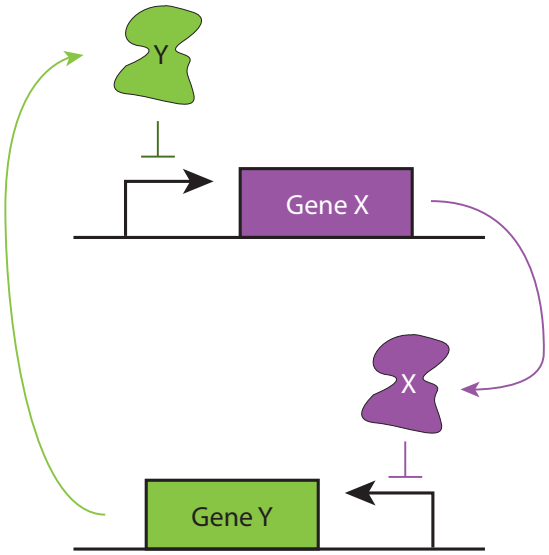

Figure 10. **Genetic circuit with two genes that repress each other**: X is a repressor of gene Y whereas Y is a repressor of gene X. This is sometimes called a **flip-flop switch**.

A more interesting scenario occurs when we have two genes, X and Y, that repress each other (Fig. 10). Specifically, consider a gene-regulation scheme governed by the following equation:

$$\frac{dx}{dt} = \frac{V_1}{1+y^n} - x \tag{36a}$$

$$\frac{dy}{dt} = \frac{V_2}{1+x^n} - y \tag{36b}$$

where x and y are concentrations of protein X and Y respectively. Here, protein X is the repressor of gene Y and protein Y is the repressor of gene X. To find the steady state expression levels of genes X and Y, we can use the idea of **nullclines** from dynamical systems theory. Let $U(x, y) = dx/dt$ (Eq. 36a) and $V(x, y) = dy/dt$ (Eq. 36b). Then at steady state expression levels (x_0, y_0), we must have $U(x_0, y_0) = 0$ and $V(x_0, y_0) = 0$. In other words, if we consider all values (x, y) for which $U(x, y) = 0$, we would obtain a curve $y = A(x)$. By considering all points (x, y) for which $V(x, y) = 0$, we would obtain another curve $y = B(x)$. By plotting A and B as functions of x, if we find that the two curves intersect, then we would conclude that the location of the intersection, (x_0, y_0), corresponds to the fixed point of the system. As an example, let us consider

$$\frac{dx}{dt} = \frac{6}{1+y} - x \tag{37a}$$

$$\frac{dy}{dt} = \frac{3}{1+x} - y \tag{37b}$$

Then setting $U = 0$ and $V = 0$, we find the following nullclines:

$$y = \frac{6}{x} - 1 \qquad \text{(satisfies U(x, y) =0)} \tag{38a}$$

$$x = \frac{3}{y} - 1 \qquad \text{(satisfies V(x, y) =0)} \tag{38b}$$

Sketching by hand Eq. 38a and Eq. 38b together in one place, we graphically see that the two nullclines intersect at only one location, (x_0, y_0) (Fig. 11). In fact, we do not even need to know the specific values of x_0 and y_0. Furthermore, we note that the intersection of the two nullclines divide the xy-plane into four regions. By looking at Eq. 37a and Eq. 37b change at each of the four regions, we deduce that the fixed point (x_0, y_0) is in fact a stable fixed point.

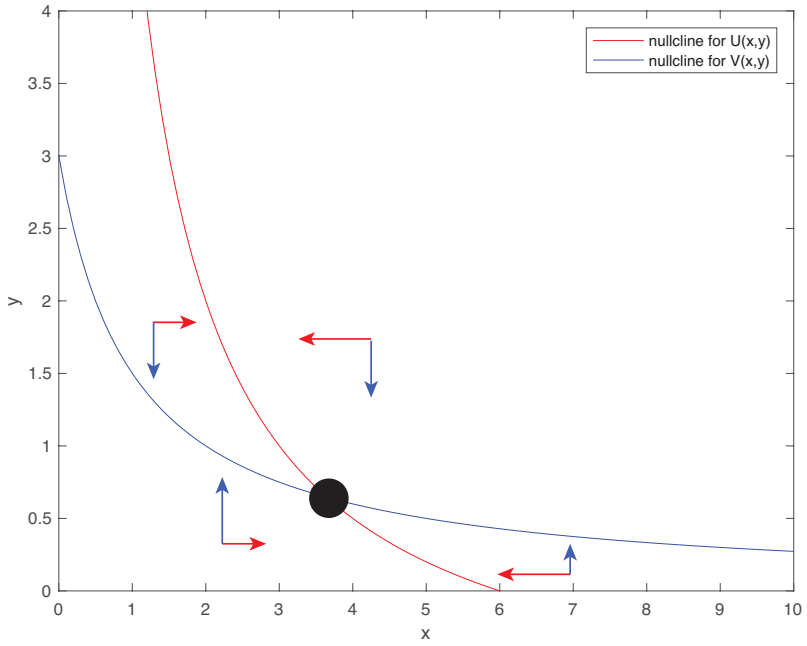

Figure 11. **Nullclines dictated by Eq. 38a and Eq. 38b** Black dot represents a fixed point. By analyzing whether Eq. 37a and Eq. 37b are positive or negative in each of the four regions that are created by the two nullclines, we deduce that the fixed point is stable.