Protocol 2 – Gene Regulation and Threshold Phenomena in Development

The Central Dogma of Genetics

The Central Dogma of molecular biology describes one of the fundamental processes in all of biology – protein synthesis – and may be represented thusly:

Gene (DNA) — Transcription → **mRNA** — Translation → **Protein**

In other words, a gene is *transcribed* into instructions in the form of a messenger RNA (*m*RNA) molecule. In the cytoplasm of the cell, ribosomes use the information contained in the *m*RNA to direct their synthesis of a protein, a process known as *translation*. This entire process is often referred to as *gene expression*. There's much more to it of course –go here for more on this important topic, or here for LOTS more – but that oversimplified model will suffice for our purpose.

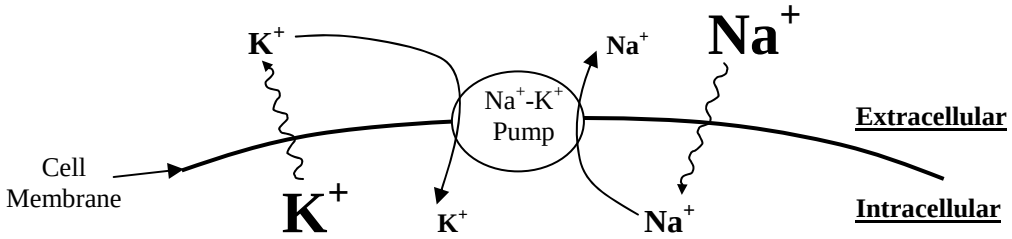
What a cell *is* and what it *does* is determined in large measure by the types of proteins it synthesizes. This in turn means that understanding gene regulation – which genes are being transcribed in a cell, and when – is key to understanding much of biology. And, one area where regulation of gene expression plays a key role is in the development of an organism from a fertilized egg. Recognition of this fact was arguably one of the key advances in biology of the past century.

Gene Regulation and Morphogenetic Gradients

Gradients are ubiquitous in biology. No organism is in complete chemical or thermodynamic equilibrium with its environment, nor are its individual cells in equilibrium with *their* environment (the extracellular, or interstitial, fluid). Gradients are thus a problem for organisms because the exchanges (fluxes) of matter and energy that gradients drive can disrupt homeostasis.

On the other hand, gradients can be highly useful. Nerve cells and muscle cells in the heart could not function without concentration gradients for Na^+, K^+, and Ca^{++} across their plasma membranes. An indication of their importance is evidenced by the fact that roughly 50% of the energy expended by your nervous system does 'nothing' more than fuel the Na^+-K^+ 'pump' that keeps intracellular K^+ concentrations high and intracellular Na^+ concentrations low by pumping Na^+ out of the cell and pulling K^+ back in (at the expense of ATP):

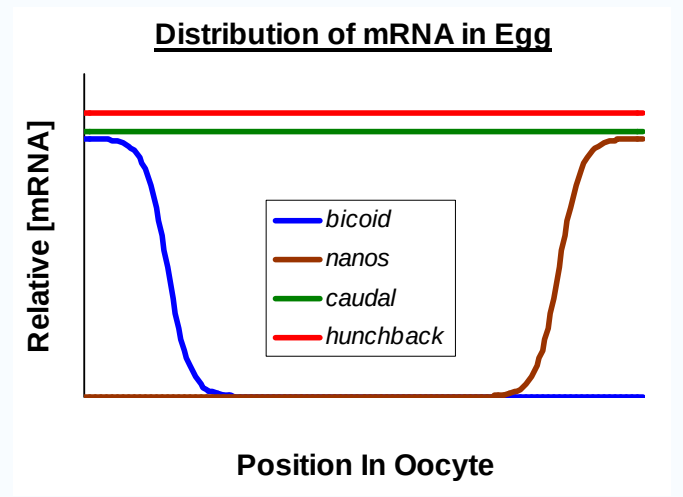
The Na^+-K+ Pump



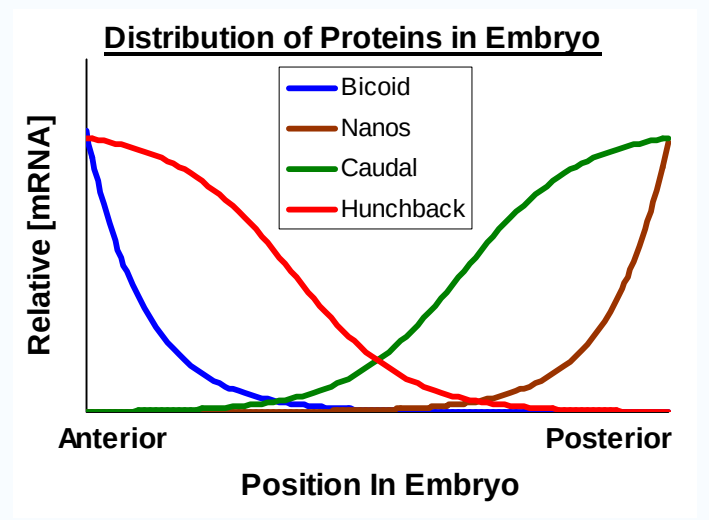
Morphogenetic Gradients and Development

OK, back to embryology. Among the most fascinating and 'useful' of gradients are those involved in regulating embryological development. These so-called *morphogenetic gradients* are an important part of the answer to the question of how a single cell – the fertilized egg – gives rise to the myriad types of cells we see in the adult organism. When the fertilized egg commences mitotic division, some daughter cells develop into 'head' cells, while others develop into 'tail' cells. Later on, some cells develop into heart cells, others into the lens of your eye, others into liver cells, and so on, all under the direction of morphogenetic gradients.

Consider the embryological development of the fruit fly, *Drosophila melanogaster*. Four genes play a crucial role in directing the earliest stages of development in the larva: *bicoid*, *hunchback*, *nanos*, and *caudal*. As the yet-to-be-fertilized egg develops inside a female's ovary, *m*RNA transcripts of *bicoid*, *nanos* and a number of other so-called *maternal genes* are produced and distributed non-uniformly throughout the cytoplasm of the egg, where translation is suppressed until the egg is fertilized. Following fertilization, these non-uniformly distributed *m*RNA transcripts direct protein synthesis, resulting in a non-uniform distribution of the resultant proteins. The consequence is that certain daughter cells receive high concentrations of *bicoid*, others receive less, while some receive essentially none at all; likewise for the *nanos* protein. On the other hand, many *m*RNA's, such as *caudal* and *hunchback* are distributed uniformly throughout the egg's cytoplasm, and all cells in the embryo contain comparable amounts of those *m*RNA's and their corresponding proteins, as illustrated in this cartoon representation:



There is often significant interaction between genes and their products (*m*RNA), and functional networks involving positive and negative feedback have been well documented and continue to be the focus of theoretical and empirical research. In this case, Bicoid protein blocks translation of *caudal* mRNA, with the result that Caudal protein is produced only by cells that are not exposed to Bicoid. Similarly, Nanos protein binds to the *hunchback* mRNA and blocks its translation in what will become the posterior (caudal) end of *Drosophila* embryos, while Bicoid protein blocks production of Nanos and Caudal proteins, but stimulates production of Hunchback protein. The end result is a set of gradients in concentration for each of the four proteins in the developing embryo:



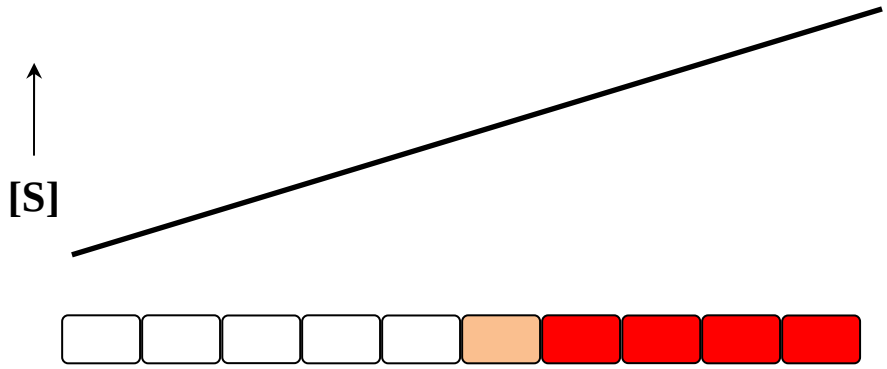
As development proceeds, differences in the relative amounts of the four proteins determine the developmental fate of the cells, leading to features we associate with the head, with the posterior region, with the thorax or the abdomen. In the same vein, other gradients determine what part of the embryo will develop into the dorsal region (the "back") and the ventral region.

It turns out that gradients play roles, some of them surprising, in most aspects of development. Again in *D. melanogaster*, cells in a certain region produce a protein that triggers adjacent cells to develop into a heart. A mutation in the *tinman** gene results in failure of the heart to develop and early death of the embryo. Development of your eye started approximately 22 days post-fertilization in response to gradients involving substances released by cells overlying a particular part of your developing 'brain'. Before that, 14-16 days post-fertilization, your nervous system itself commenced development in response to gradients produced by nearby cell layers.

Threshold Responses In Development

On the face of it, however, using gradients to direct development is problematic. Your eye, for example, is quite distinct from surrounding tissues…and it *needs* to be. Likewise, your heart doesn't blend into your lungs, there's a sharp demarcation between your brain tissue and your skull, and so forth. This distinctness in adjacent structures suggests that developmental 'switches' were flipped somewhere along the line, so that while some cells proceed to develop into, say, an eyeball, other nearby cells took a dramatically different path and produced the bone that comprises the eye socket.

The problem is that morphogenetic gradients are more-or-less linear, at least locally, and it's not apparent how sharply demarcated structures such as eyeballs, teeth, or hearts can develop in response to gradually-changing, linear gradients. The problem is illustrated in the following cartoon, in which the gradient in the concentration of regulatory substance *S* is indicated by the solid line, and the two different developmental fates of the cells – pigmented vs. unpigmented, say – in the layer exposed to the gradient are indicated by their coloring:

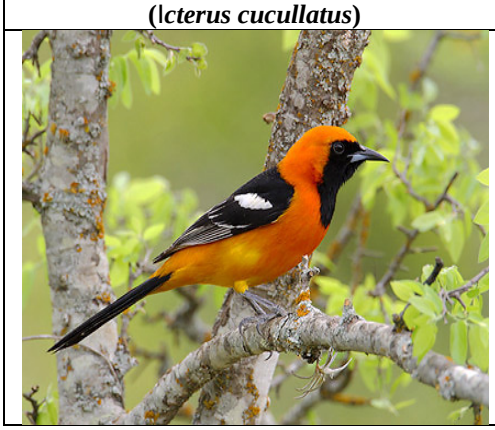
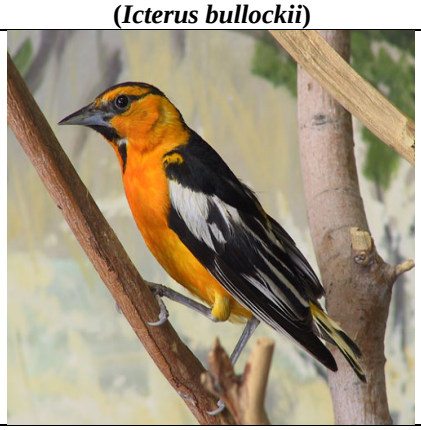


If that example seems contrived, consider the following images:

Harlequin Rasbora (*Trigonostigma heteromorpha*)	**Royal Angelfish (*Centropyge bicolor*)**	**Royal Gramma (*Gramma loreto*)**
		

In addition, considerable evidence demonstrates that it's not only *what* genes are expressed during development that matters. The importance of the *timing* of gene expression – i.e., when certain genes

are turned on or off during development – cannot be overstated. A prime example of this is the developmental biology of humans and chimpanzees, where work has revealed *perfect identity* in at least 98.7% of our DNA sequences. Why, then, are we so different-looking? The answer lies in different temporal patterns of gene expression during development in chimps and humans. In other words, it's not only *which* genes are expressed and *what* proteins they code for, but *where* they're expressed and *when* they're expressed in the developing embryo that's important. If you want more information, here's an interesting report of **recent findings in comparative genomics** of humans and chimpanzees, but striking evidence of this can be seen in the head pattern of U. S.orioles:

Hooded oriole (*Icterus cucullatus*)	**Northern oriole (*Icterus galbula*)**	**Bullock's Oriole (*Icterus bullockii*)**
		

flickers (ant-eating woodpecker relatives):

Northern Flicker (*Colaptes auratus*) – Yellow-shafted (Eastern) subspecies

Note the yellow tail feathers and a black 'moustache' below the eye.

Northern Flicker (*Colaptes auratus*) – Red-shafted (Western) subspecies

Note the combination of *red* tail feathers and a red 'moustache' below the eye.

Gilded Flicker (*Colaptes chrysoides*)

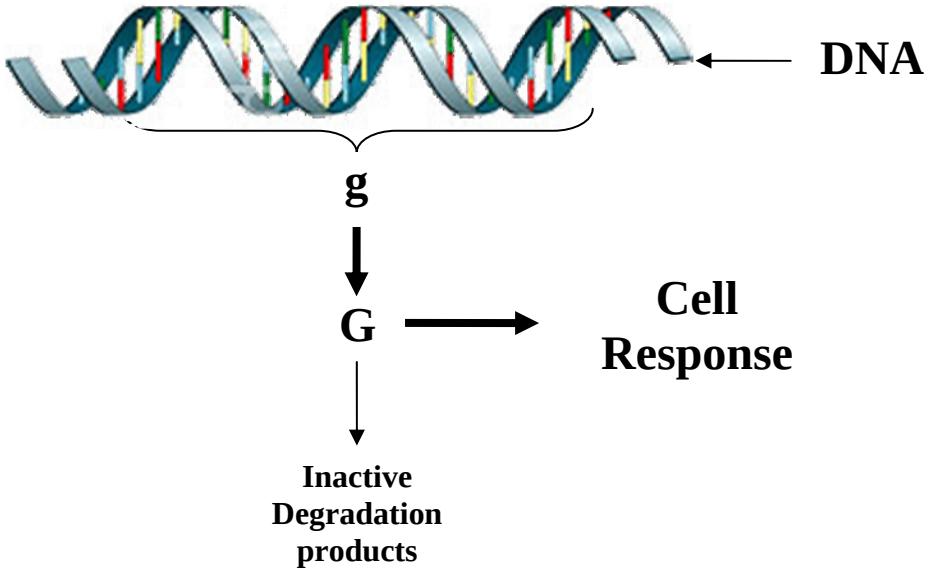
Note the yellow tail feathers, but a red 'moustache' below the eye.

Clearly, abrupt switches in the developmental fate of cells underlie the color patterns in these birds, but the location of the source of (assumed) morphogenetic gradients must have differed among the species, and the timing of when the gradient developed and induced the switches in pigment deposition likely differed as well. Moreover, moustache and tail feather color (and underwing feather color, which you can't see in the above photos) in the flickers involves production of different pigments.

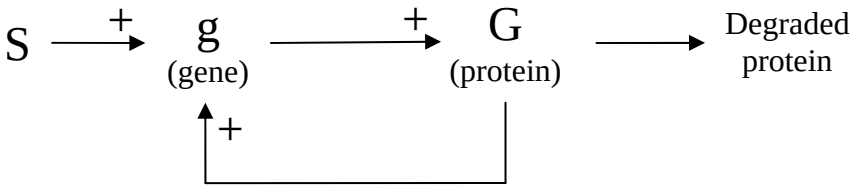
In fact, developmental switches in response to changes in concentration of some 'signal' molecule are a pervasive feature of embryogenesis. Given that the color patterns illustrated above serve as species-recognition and mate-choice cues, it's clear that gene expression patterns during development can have important ecological and evolutionary consequences. Understanding the mechanism(s) responsible for these apparent abrupt changes in gene expression is therefore of considerable interest to developmental biologists. In this lab, you will investigate the behavior of a simple model that has been proposed to account for the phenomenon of developmental thresholds, or switches, induced by continuously-varying morphogenetic gradients.

The Model

The model on which this lab is based involves a gene, *g*, whose product, the protein G, controls some process in a cell such as, say, pigment production. A simple diagram of the system would be:



Lewis et al. (1975) assumed for their model that expression of *g* is stimulated by a substance, *S*, that is present in concentrations that vary linearly as indicated in the previous diagram. As is often observed in real genetic systems, G also upregulates *g*, thereby meaning it increases its own production (= *product activation*). They also assumed that G degrades with time, which is true of essentially all proteins in the cytosol. A schematic representation of the proposed model would then be:



However, when we actually write the equation for the model, we omit *g* and treat both S and G as though they *directly* impacted the amount of G. The postulated structure and connectivity of this model system leads to a single time-dependent differential equation:

$$\frac{dG}{dt} = k_1 S - k_2 G + \frac{KG^2}{k_n + G^2} \qquad \text{Equation 1}$$

The first two terms on the RHS of Equation 1 provide **Michaelis-Menten (saturation) kinetics** for the concentration of G, while the activation of *g* by its product, G, is provided by the last term. As we'll

see, the addition of that (nonlinear) positive feedback term means that the system can exhibit some interesting dynamical behavior.

---

* The *tinman* gene was named for Tin Man, the *Wizard of Oz* character who lacked a heart. Here's a list of some other amusing, sometimes poignant, names that have been given to various mutants that have been discovered in *Drosophila melanogaster*.) back

Working with `morph_switch.m` (download the .m file)

The script you'll be working with during this lab contains many coding features that you've had experience with previously, some you've probably not previously encountered. Some features, new as well as familiar, are worth mentioning:

1. As written, you pass only one argument, the value of S, when calling the function **morph_switch.m**. However, for Exercise 9 below, you may want to rewrite the code so that you can pass different values for the four constant parameters ($k_1, k_2, k_n,$ and K), as well as S.
2. Note the code used in the four calls to **`ode113`** (one of MATLAB's ODE solver engines) in Lines 69-72. Compare that with the code you'll see if you enter **`doc ode45`** at the MATLAB command line. The code in **morph_switch.m** is the only way to pass parameters (as opposed to independent and dependent variables) to any of MATLAB's ODE solvers, but you won't find it even mentioned in MATLAB's help files; you therefore might want to make note of it
3. To highlight the switch-like behavior of the system and emphasize the threshold phenomenon, the numerical solution starts with four different initial values for G(0). These are plotted as colored asterisks in Figure 1. Observing the resulting trajectories should make it easier to understand the consequences of changes in [S], and to answer Questions 1 & 2, below. Feel free to change the values of G(0) (stored as elements of the 1×4 row vector, **`G_init`** in Line 16 of the code) as you see fit.
   a. Note: the MATLAB function **`eps`** returns the smallest number that MATLAB can discriminate from zero, meaning that numbers smaller than that will be treated as zero by MATLAB. This value is computer-specific, although in most computers it is $\approx 2.2\times10^{-16}$. It is used to supply a value for G(0) that is as close to zero as possible on your computer.
4. The MATLAB **`sprintf`** command is used to display the current value of S in the title of Figure 1, and to automatically update the legend entries if you choose to change the initial values of G.
5. This script also illustrates how to embed LaTeX commands into MATLAB's **`text`** command to enhance graphical output of mathematical symbols, formulae, etc. LaTeX (LaTeX in plain text, or simply Latex) is, quoting the **Wikipedia article** on LaTeX,. "a *document markup language* and *document preparation system* for TeX, a typesetting program that is widely used by mathematicians, scientists, philosophers, engineers, and scholars in academia and the commercial world, … because of the quality of typesetting achievable by TeX". An extensive listing of LaTeX commands and conventions is available at:
   http://www.math.wisc.edu/~mitchell/curl/learntex.pdf.
6. Finally, the script also illustrates dynamic positioning of text on a plot, done here to minimize the overlap of graph labels with their corresponding lines and other features of the plot. (yeah, that's probably overkill, and a simple legend would do the same thing, perhaps better, and I probably should be embarrassed for incorporating the feature, but what can I say? I'm a geek.).

Exercises

1. Run **`morph_switch`** using S = 0.01. Familiarize yourself with the features of the plots in Figures 1 and 2. Note that Figure 1 displays three plots:
   - $f(G,S) = (k_1 S - k_2 G)$,
   - $h(G) = \dfrac{KG^2}{k_n + G^2}$, and
   - $\dot{G} = f(G,S) + h(G)$.

Thinking back to our discussion of the spruce budworm model and catastrophe theory late last semester, can you explain why this was done? Hint: note that *f* and *h* are components of the derivative, $\dot{G}$, and that $\dot{G} = 0$ at fixed point(s).

2. Do repeat runs, increasing S in increments of 0.01 up to a value of 0.20. Note carefully changes in the appearance of the curves in Figures 1 and 2.
    - In accordance with convention, stable fixed points are represented by filled circles, while unstable fixed points are represented by unfilled circles.
3. Is the origin stable or unstable? Is the origin a fixed point? Justify your responses.

For the next two exercises, you may need to increase t_max to 300 or greater in Line 16 to get a true picture of the long-term dynamics.

4. For $S = 0.14$, do all trajectories lead to the same final value (G*) as $t \to \infty$? If so, why does it take so long for trajectories with $G(0) < 0.3$ to converge on G*? Hint: what determines how rapidly the point G(t) approaches G*?
5. Do a run with S = 0.078, and note how long it takes for the trajectory with G(0) = 2 to reach the stable fixed point, compared with the other three trajectories. What accounts for the difference?
6. Does the assertion that "The first two terms on the RHS of Equation 1 provide **Michaelis-Menten (saturation) kinetics** for the concentration of G, while the activation of *g* by its product, G, is provided by the last term." hold up under scrutiny? One way to test this is to make appropriate modifications to **morph_switch.m** and **morph_switch_deriv** and, *after renaming the function (*Line 1*) and saving the changed code under a different file name*, observe the (temporal) behavior of the solution.
7. **Hand in.** If your results show evidence of bifurcation(s), construct a bifurcation diagram for Equation 1 and tell me what type(s) of bifurcation you observed. If bifurcations did occur, was the result a limit cycle, or multiple fixed points? Did the result match your *prediction*?
    - To help you with this Exercise, look at **samp_bif_plot_code.m**, a MATLAB script that displays the bifurcation plot for $\dot{x} = -x^3 + rx$, the prototype function for the supercritical pitchfork bifurcation. Feel free to modify this for your purpose.
8. **Hand in.** Is the presence of S above some critical concentration really necessary to cause a switch in the developmental fate of the G-synthesizing cells? For example, suppose a very brief, perhaps stochastic, spasm of expression of *g* produced the tiniest amount of G (equivalent to entering a value of **eps** for **G_init**). Wouldn't the positive feedback loop connecting G and *g* lead inexorably to high levels of G in a cell? Justify your answer.
    - A good reason for pondering this question is that positive feedback is considered by most biologists, certainly by physiologists, to be *destabilizing* for a system, leading to a runaway state that can only be reversed by some external event that interrupts the loop. Autocatalytic activation of digestive enzymes, itching, uterine contractions during parturition, and orgasm exemplify systems whose temporal dynamics are governed by a positive feedback loop that requires some external event to interrupt it and return the system to its 'resting' state.
9. **Hand in.** Do you see evidence for hysteresis in this system? If so, what *is* that evidence? Speculate about the possible consequence(s) for the cells located along the gradient in S.
10. **Hand in.** In keeping with our 'theme' of thinking about mathematical models in different biological contexts, consider this: distributions of many organisms are characterized by rather abrupt boundaries between species, even though habitat conditions such as temperature, precipitation, etc. usually show smooth, clinal changes across the species boundaries. This is exemplified by the often sharply-demarcated 'life zones' seen along altitudinal gradients in the

western United States (Merriam, 1898), as evidenced by the altitude-associated plant community changes evident in the following photograph of the Henry Mountains of south-central Utah:

Ecologists, especially those working with animals rather than plants, generally assume that biotic interactions, such as interspecific competition, play an important, if not overriding, role in defining species' boundaries (but see Wilson and Agnew, 1992; Malanson, 1997). Does the model you've been working with in this lab suggest any alternative explanation(s)?

- In this context, suggest biologically/ecologically realistic interpretations of the four parameters $k_1, k_2, k_n,$ and K.
- Vary the four parameters and comment on the impact on the dynamical behavior of the model.

11. **Just for fun**. Explore the consequences of using different exponents, say 1 and 3 as well as 2, in the last term of Equation 1.
    - Note that changing *n* will change the equivalent polynomial for Equation 1, and therefore require changes to the part of **morph_switch** that calculates the roots of the equivalent polynomial for $\dot{G}$.
        i. To see how equivalent polynomials are derived, **click here**.
    - I'd recommend that you construct plots of $\frac{KG^n}{k_n + G^n}$ for *n* = 1, 2, 3, so you can get a good feel for how the shape of the curve changes with *n*. Note: you can use **ezplot** to do this, or you could write a short script that would superimpose the three plots.
    - You may recall from last semester that the *n* = 1 form of this term is the same as the Hassel model of regulated population growth.
12. **Also just for fun –** use **dfield7** – or write your own MATLAB program – to check the assertion (1st line after Equation 1, p. 5) that the first two terms of Equation 1 (i.e., $\frac{dG}{dt} = k_1 S - k_2 G$) yield saturation kinetics.
13. **Also just for fun –** try to derive an analytical solution for the fixed point(s) of Equation 1.
14. **Also just for fun –**the stability of Equation 1's fixed points was assessed by looking at the fate of a small negative perturbation applied to the fixed point in Line 32 of **morph_switch**,. Is that sufficient to determine stability? Shouldn't the code also look at the fate of a small *positive* perturbation? Justify your answer.

Literature Cited

Lewis, J., J.M.W. Slack, and L. Wolpert. 1977. Thresholds in development. J. Theor. Biol. 65:579-590.

Malanson, G. P. 1997. Effects of feedbacks and seed rain on ecotone patterns. Landscape Ecology vol. 12 no. l pp 27-38.

Merriam, C. H. 1898 Life-zones and crop-zones of the United States. Department of Agriculture, Biological Survey Bulletin 10: 1-79.

Wilson, J.B. & Agnew, A.D.Q. (1992) Positive-feedback switches in plant communities. Advances in Ecological Research, 23, 263-336.