A Mathematical Model of Gene Regulatory Network Evolution

Josh Schiffman

December 16, 2016

Background

- How do we mathematically model Developmental Systems Drift (DSD) and Gene Regulatory Network (GRN) evolution?

  – First, we need a good analytic description of an organism – something that includes and distinguishes among genotype, GRN, and phenotype.

  – Second, these descriptions should be as precise as possible.

  – Third, we need to describe the evolutionary and population dynamics of this model.

- How do we apply this model?

An organism is described by its molecular time-dynamics.

Molecular (say protein or mRNA) concentrations are represented by $\vec{x}$, the gene regulatory interactions by α, β, γ, and δ, and the environmental/initial inputs by B.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u$$

Each row vector $\begin{bmatrix} \alpha & \beta \end{bmatrix}$ of the GRN matrix is the promoter region of a gene, such that, α describes the how much gene-1 self regulates and β describes how much gene-2 regulates gene-1.

The phenotype is just a subset of an organism's proteins directly visible to selection

$$y = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

For instance, if $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$, the phenotype is only described by the dynamics of gene-1. This means that selection can only "see" what happens to the proteins produced by gene-1, whereas the dynamics of gene-2 are invisible to selection (and thus not formally defined as part of the phenotype).

This does not mean, however, that gene-2 is unnecessary with respect to the phenotype!

For example, consider a GRN with oscillating protein concentrations.

In this 2 gene system, neither gene-1 nor gene-2 self regulate. Gene-2 up-regulates gene-1, and gene-2 is down-regulated by gene-1:
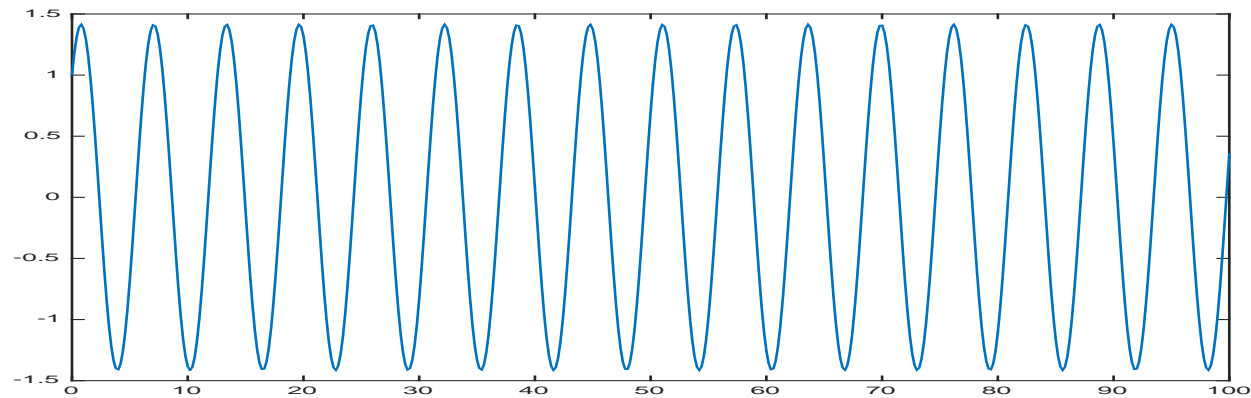
$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u$$

Furthermore, only the dynamics of gene-1 are important:

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

2-Gene GRN Oscillating Expression Dynamics

Figure 1: The dynamics of gene-1 are, $y = \cos(t) + \sin(t)$

A single-gene system cannot oscillate.

Expression dynamic complexity is limited by the number of genes in a network.

$f(t)$ describes protein concentrations at time t, λ_j is the jth eigenvalue of the GRN and ξ is a constant.

$$f(t) = \sum_{j=1}^{n} e^{\lambda_j t} \xi_j$$

The oscillating dynamics can be rewritten in this form too:

$$f(t) = e^{it} \frac{(i-1)}{2i} + e^{-it} \frac{(i+1)}{2i} = \cos(t) + \sin(t)$$

Sometimes it's (visually) easier to study the Laplace transform of the system.

$$\mathcal{L}(f(t)) := G(s) = \sum_{j}^{n} \frac{c_j b_j}{(s - \lambda_j)}$$

Is the GRN architecture for a given phenotype unique? If no, how many GRN topologies yield equivalent dynamics?

Let two systems Σ and $\bar{\Sigma}$ be defined as,

$$\Sigma \begin{cases} \dot{x} = Ax + Bu \\ \\ y = Cx \end{cases} \qquad\qquad \bar{\Sigma} \begin{cases} \dot{\bar{x}} = \bar{A}\bar{x} + \bar{B}u \\ \\ y = \bar{C}\bar{x} \end{cases}$$

Where
$x \in \mathbb{R}^n, \ y \in \mathbb{R}^l, \ u \in \mathbb{R}, \ A, \bar{A} \in \mathbb{R}^{n \times n}, \ B, \bar{B} \in \mathbb{R}^{n \times l}, \ C, \bar{C} \in \mathbb{R}^{l \times n}.$

The systems Σ and $\bar{\Sigma}$ are "algebraically equivalent" (and have identical dynamics) if,

$$
\begin{aligned}
\bar{x} &= Px \\
\bar{A} &= PAP^{-1} \\
\bar{B} &= PB \\
\bar{C} &= CP
\end{aligned}
\tag{1}
$$

Where P is any invertible matrix.

This is only true if both GRNs are composed of the exact same number of genes.

Can two systems have identical dynamics, yet differ in the number of genes?

Two systems, even in different dimensions (!!), exhibit identical dynamics (and are "zero-state equivalent") if and only if their transfer functions are identical $G(s) = \bar{G}(s)$.

Remember, $G(s) = \mathcal{L}(f) = C(sI - A)^{-1}B$

$$G(s) = \bar{G}(s) \iff$$
$$CA^iB = \bar{C}\bar{A}^i\bar{B} \; \forall \; i \geq 0$$

Proof:

$$G(s) = C(sI - A)^{-1}B$$

$$= Cs^{-1}(I - s^{-1}A)^{-1}B$$

$$= Cs^{-1}\left(\sum_{i=0}^{\infty}(s^{-1}A)^i\right)B$$

$$= \sum_{i=0}^{\infty}CA^iBs^{i+1}$$

$$\sum_{i=0}^{\infty}CA^iBs^{-(i+1)} = \sum_{i=0}^{\infty}\bar{C}\bar{A}^i\bar{B}s^{-(i+1)}$$

$$\iff CA^iB = \bar{C}\bar{A}^i\bar{B} \ \forall i$$

Wait... how does any of this make biological sense?

We only care about systems where $\bar{C} = C$ and $\bar{B} = B$.

Any two systems that are "algebraically equivalent" where $PB = B$ and $CP = C$ (P is a stabilizer subgroup of the invertible matrices), have identical inputs and identical phenotypes, but have different *mechanisms*, since it's possible that $A \neq PAP^{-1}$.

The **set of all 2-gene GRNs** with the dynamics $G(s) = \frac{(s-k)}{(s-\lambda_1)(s-\lambda_2)}$ is,

$$
\begin{bmatrix} 1 & 0 \\ r & 1-r \end{bmatrix}
\begin{bmatrix} \lambda_1 & \lambda_2 - k \\ 0 & \lambda_2 \end{bmatrix}
\begin{bmatrix} 1 & 0 \\ \frac{r}{r-1} & \frac{-1}{r-1} \end{bmatrix}
$$

$\forall\, r \neq 1 \in \mathbb{R}$

The set of all GRNs, in any dimension, with identical dynamics is...

$$P \begin{bmatrix} A & 0 \\ 0 & \Lambda \end{bmatrix} P^{-1} : PB = B \text{ and}$$

$$CP = C, \text{ and } \Lambda = \begin{bmatrix} \lambda_{n+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n+m} \end{bmatrix}$$

Where some elements in P and all eigenvalues in Λ are free to take any value. (B and C can be augmented with 0s).

DMIs

Define reproduction as

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} A + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \bar{A}$$

Let phenotypic error be $\epsilon_0 := \big(\epsilon_1 + \epsilon_2 + \epsilon_k\big)$, where $\epsilon_{1,2} := |\lambda_{1,2} - \lambda^H_{1,2}|$ and $\epsilon_k := |k - k^H|$.
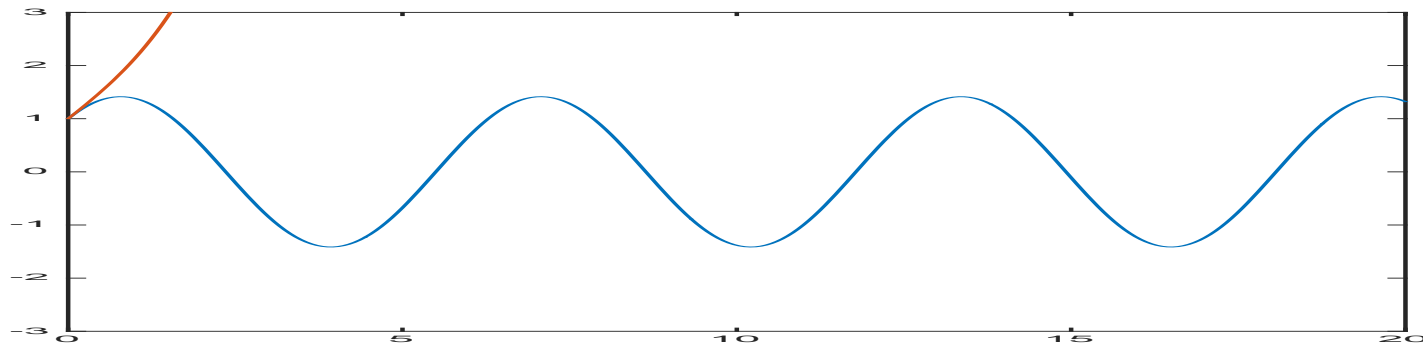
$$\lambda^H_{1,2} = \tfrac{1}{2}\left(\lambda_1 + \lambda_2 + \zeta - \bar{\zeta} \pm \Big((\lambda_1 - \lambda_2)^2 - (\zeta - \bar{\zeta})\big((\zeta + 3\bar{\zeta}) + 2(\lambda_1 + \right.$$

$$\left. \lambda_2) + \tfrac{4(\lambda_1 - \bar{\zeta})(\lambda_2 - \bar{\zeta})}{k + \bar{\zeta}}\big)\Big)^{\frac{1}{2}} \right)$$

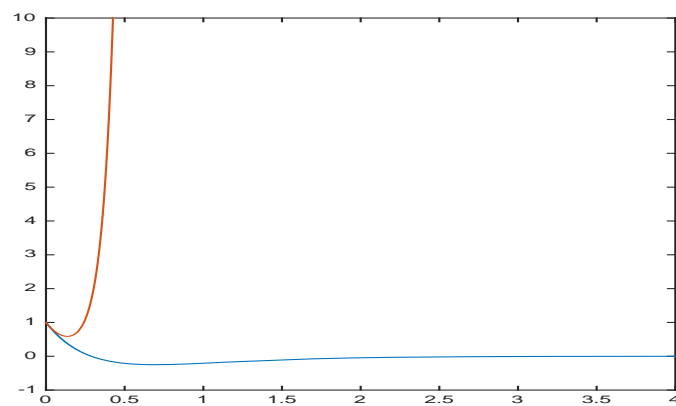$$k^H = \zeta - \bar{\zeta}.$$

Let GRNs that survive have $\epsilon_0 < \delta$.

Hybrid expression dynamics deviate significantly

Both $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ and $\begin{bmatrix} -1 & 2 \\ -1 & 1 \end{bmatrix}$ have identical phenotypes, however their hybrids do not.

$$\begin{bmatrix} 0 & -6 \\ 1 & -5 \end{bmatrix} \text{ and } \begin{bmatrix} -12 & 6 \\ -15 & 7 \end{bmatrix}$$

Evolution as Brownian Motion

If $\mathcal{B}$ is an η-dimensional (where η consists of the free parameters in P and eigenvalues λ_{n+1} through λ_{n+m}) Brownian Motion with covariance matrix $\sigma^2 I$ then $\mathbb{E}\left[\|\mathcal{B}\|^2\right] = \eta\sigma^2 t$, however, $A = F(\mathcal{B})$.

How to model gene addition and deletion.

- Gene addition: add row and column to the GRN matrix; augment B and C with zeros.

- Gene deletion: more complicated. What is the probability that,
  - a set of dimensions (genes) we don't care about.
  - a set of dimensions only affects the other genes in the network "a little" bit (the GRN is mostly lower-triangular).

$$Z = \begin{bmatrix} \bigstar & |\epsilon \\ \cdots & \cdots \end{bmatrix} \qquad \|G(s) - \bar{G}(s)\| \leq \delta$$

  - if the phenotypic effect of these extra dimensions diminishes as gene networks evolve to be larger, pherhaps there will be an equillibrium level GRN complexity. (note: empirical network connectedness and robustness).

Thank you!

Peter Ralph

Sergey Nuzhdin

Erik Lundgren