

# 1 Background

*Review of relevant other stuff.*

## *Introduction*

It is commonly taught that an organism’s genome contains the heritable material that natural selection filters and that an organism’s phenotype directly determines its evolutionary fitness. Between genotype and phenotype is an often complicated and poorly understood molecular machinery – and it is a major goal common to many different disciplines within the life sciences to elucidate its form, function, and evolution. These aims are delicately intertwined and a comprehensive understanding of a system’s evolution requires an understanding of its function and vice versa.

The molecular machinery, interacting with the environment, and bridging genotype to phenotype can be mathematically described as a dynamical system (or a system of differential equations). Previously it has been argued that any idealized study of evolution is incomplete without a mathematically sufficient description of the genotype, phenotype, and transformation from one to the other. Movement in this direction is ongoing, as researchers have begun to study the evolution of empirically inspired computational and mathematical models of gene regulatory networks (GRNs) and metabolic networks that incorporate empirical data such as sequences and expression patterns. If we allow the reasonable assumption that the genotype-phenotype map can be represented as a system of differential equations, we can immediately discuss its evolution and function in a much more mechanistic, yet general, manner.

In some fields that seek to fit parametric models to experimental data, such as control theory, chemical engineering, and statistics, it is well known that mathematical models can fundamentally be *unidentifiable* and/or *indistinguishable* – meaning that there can be uncertainty about an inferred model’s parameters or even its claims about causal structure, even with access to complete and perfect data. Models with different parameter schemes, or even different mechanics can be equally accurate, but still not *actually* agree with what is being modelled. In control theory, where electrical circuits and mechanical systems are often the focus, it is understood that there can be an infinite number of “realizations,” or ways to reverse engineer the dynamics of a black box, even if all possible input and output experiments on the black box are performed. In chemical engineering, those who study chemical reaction networks sometimes refer to the fundamental unidentifiability of these networks as “the fundamental dogma of chemical kinetics.” Although this may frustrate the occasional engineer or scientist, viewed from another angle, the concepts of unidentifiability and indistinguishability can provide a starting point for thinking about externally equivalent systems – systems that evolution can explore, so long as the parameters and structures can be realized biologically. In fact, evolutionary biologists who study homology and analogy are very familiar with such functional symmetries; macroscopically identical phenotypes in closely related species can be divergent at the molecular level.

In this paper we propose a framework to study the evolution of biological systems. To begin, we focus on the evolution of an idealized population. We consider the evolution of a perfectly adapted, large population, evolving in a static environment for an infinite number of generations. Under these ideal circumstances, we expect to observe a “conservation of phenotype,” where the population explores the manifold of phenotypically-invariant (or symmetric) genetic and developmental architectures. We would like to understand which parameters influence the distribution of

a population along the manifold of phenotypically invariant genetic systems. Further, we can show how dispersion along this manifold contributes to speciation and evolvability.

*phenomenologically or dynamically equivalent?* The set of all systems – that is all model structures and all parameter schemes – that are externally, or phenomenologically, equivalent, biologically realizable, and mutationally accessible,

What are typically referred to as “design principles” or some sort of clear structure-function relationships are commonly sought

Seeking to attribute functionality, uninformed by evolution, can lead to spurious claims (i.e. see Graur).

A linear time invariant (LTI) dynamical system is a system of linear differential equations that describes a physical process. LTI systems are used in electrical and control engineering to model a myriad of phenomena including circuits. One can use this methodology, under a set of assumptions, to reverse engineer a mechanism from impulse experiments (input/output data). The idea is that given a black box, and it’s experimental manipulation, can we describe the internal mechanisms?

## 2 The problem

Organisms often have to respond to their environments, and the way this happens is that external input triggers a cascade of molecular signals whose eventual result is a reaction, whose appropriateness for the situation determines fitness. *A similar situation occurs in development (...?).* However, there are many ways to skin a cat: especially given a number of possible molecular intermediaries, there may be a large number of ways to construct such a molecular signalling system that has precisely the *same* input-output relationships, and hence the same function. Therefore, any mutation that changes one such system to another, equivalent version will be neutral, and may well become polymorphic in a population if there are no ill effects in heterozygotes. Certainly, mutations that move a system away from optimal functioning will also appear, but in very large populations will not rise to high frequency. *Deleterious things are quite important but somehow justify starting with the neutral ones.*

*More stuff on why we care about this: Provide neutral model; can address questions. Main point of paper.*

**Example 1 (Example: Oscillator.)** *Motivate the system:*

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

*as a special case of the below.*

**Example 2 (Example: two-component systems.)** *As a more general example, suppose that a cell produces two molecular species,  $S_1$  and  $S_2$  whose production are both stimulated by the presence of an external “input” substance. call them TF? Suppose that the two species self-regulate with strengths  $\lambda_1$  and  $\lambda_2$ , respectively, and the second species also regulates the first species with strength  $\gamma$ . Rephrase in terms of binding to promoter sequences? However, only the time course of the concentration of the first one of these species determines the fitness of the organism. (As happens for instance in XXX.) Write  $x(t) = (x_1(t), x_2(t))$  for the concentrations of the two species at time  $t$ , and write  $u(t)$  for the concentration at time  $t$  of the input substance. Then, if the rates at which*

each species are produced are linear functions of the concentrations (*look up way this is usually said in the literature.*), then the dynamics of the system are given by

$$\begin{aligned}\dot{x}_1(t) &= \lambda_1 x_1(t) + \gamma x_2(t) + u(t) \\ \dot{x}_2(t) &= \lambda_2 x_1(t) + u(t),\end{aligned}$$

where  $\dot{x}$  denotes the time derivative. The initial conditions,  $x_1(0)$  and  $x_2(0)$ , and the input  $u(t)$  then determine the concentrations through time. If we record the regulatory coefficients in the matrix:

$$A = \begin{bmatrix} \lambda_1 & \gamma \\ 0 & \lambda_2 \end{bmatrix},$$

and define the column vector  $B = [1, 1]^T$ , then in matrix notation the dynamics are

$$\dot{x}(t) = Ax(t) + Bu(t).$$

If these are all transcription factors, and we suppose their binding motifs are fixed, then the  $i^{\text{th}}$  row of  $A$  is determined by the  $i^{\text{th}}$  promoter sequence.

Since the system is linear, the state of the system at any time is the superposition of its responses to all previous inputs. This implies that if  $G(t)$  is the transient response of the system to a unit impulse,  $t$  units of time later, and  $x(0) = 0$ , then the time course of the concentration of the thing we care about,  $x_1(t)$ , can be written as

$$x_1(t) = \int_0^t G(t-s)u(s)ds. \tag{1}$$

However, it turns out that *look for simple demonstration this is true*

$$P_p = \begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix},$$

then for any  $p \neq 1$ , if we replace the regulatory matrix  $A$  with

$$A(p) = PAP^{-1},$$

then the response of  $x_1(t)$  to any particular input  $u(t)$  is identical to the original system. (However,  $x_2$  may well be different!)

For instance, setting  $p = -1$ , the system with

$$A(-1) = \begin{bmatrix} \lambda_1 + \gamma/2 & \gamma/2 \\ \lambda_2 - \lambda_1 - \gamma/2 & \lambda_2 - \gamma/2 \end{bmatrix},$$

looks very different, but gives the same input-output relationship. Although these systems are equivalent, hybrids between them may not be: *do example.*

### 3 Regulatory networks as linear, time-invariant systems

*Make sure to have strongly worded disclaimer somewhere that we know real networks aren't linear but they are locally. Plus, it works in engineering.*

We will now lay out the model in more general terms. Suppose that the *internal state* of the system is parameterized by the concentrations of a collection of  $n$  molecular species,  $S_1, \dots, S_n$ , and the vector of concentrations at time  $t$  we denote  $x(t) = (x_1(t), \dots, x_n(t))$ . There are also  $m$  “input” species, whose concentrations are determined exogenously to the system, and are denoted  $u(t) = (u_1(t), \dots, u_m(t))$ , and  $\ell$  “output” species, whose concentrations are denoted  $y(t) = (y_1(t), \dots, y_\ell(t))$ . The output is merely a linear function of the internal state:

$$y_i(t) = \sum_j C_{ij} x_j(t).$$

Since  $y$  is what natural selection acts on, we refer to it as the *phenotype*, and sometimes in contrast refer to  $x$  as the *kryptotype*, as it is “hidden” from direct selection. *Maybe*. The rate at which the  $i^{\text{th}}$  species is produced is a weighted sum of the concentrations of the other species as well as the input:

$$\dot{x}_i(t) = \sum_j A_{ij} x_j(t) + \sum_k B_{ik} u_k(t).$$

In matrix notation, this is written more concisely as

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{2}$$

$$y(t) = Cx(t). \tag{3}$$

*Everything we say I think is assuming  $x(0) = 0$ . Figure this out. “Zero-state equivalence.”*

Given an initial condition and an input, it is possible to write the solution  $x(t)$  as a convolution with an input-response kernel as in the example above (equation (1)). An alternative way of describing the input-output relationship, more common in engineering, is instead to find the *transfer function*

$$H(z) = C(zI - A)^{-1}B, \tag{4}$$

which is more mathematically readable *if you know what it means*. This can be thought of as the system’s response to input at “frequency”  $z$ . The fact that the transfer function uniquely determines the system’s input-output relationship (i.e., the mapping  $u \mapsto y$ ) follows from the fact that if  $\tilde{y}(z) = \int_0^\infty e^{-zt} y(t) dt$  is the Laplace transform of  $y$ , and  $\tilde{u}(z)$  is the Laplace transform of  $u$ , then

$$\tilde{y}(z) = H(z)\tilde{u}(z).$$

(This happens because the Laplace transform takes convolutions to products, and  $H$  is the Laplace transform of the kernel  $G$ .) Concretely,

*Should we remind what dimensions everything is?*

The state space representation is considered an internal description of the system, whereas the transfer function is an external description. “Realizations” are something.

**Definition 1 (Phenomenological equivalence of systems)** Let  $(x(t), y(t))$  and  $(\bar{x}(t), \bar{y}(t))$  be the solutions to (2) with coefficient matrices  $(A, B, C)$  and  $(\bar{A}, \bar{B}, \bar{C})$  respectively, and both  $x(0)$  and  $\bar{x}(0)$  are zero. The systems defined by  $(A, B, C)$  and  $(\bar{A}, \bar{B}, \bar{C})$  are **phenomenologically equivalent** if

$$y(t) = \bar{y}(t) \quad \text{for all } t \geq 0.$$

Equivalently, this occurs if and only if

$$H(z) = \bar{H}(z) \quad \text{for all } z \geq 0,$$

where  $H$  and  $\bar{H}$  are the transfer functions of the two systems.

One way to find other systems equivalent to a given one is by change of coordinates (“algebraic equivalence”): if  $T$  is an invertible matrix, then the systems  $(A, B, C)$  and  $(TAT^{-1}, TB, CT^{-1})$  have the same dynamics because their transfer functions are equal:

$$CT^{-1}(zI - TAT^{-1})^{-1}TB = CT^{-1}T(zI - A)^{-1}T^{-1}TB = C(zI - A)^{-1}B.$$

However, the converse is not necessarily true: systems can have identical transfer functions without being changes of coordinates of each other. In fact, systems with identical transfer functions can involve interactions between different numbers of molecular species.

The set of all systems phenomenologically equivalent to a given system  $(A, B, C)$  is elegantly described using the Kalman decomposition, which also clarifies the system dynamics? tells us a lot about how it works? *or something* To motivate this, first note that the the input  $u(t)$  only directly pushes the system in directions lying in the span of the columns of  $B$ . As a result, different combinations of input can move the system in any direction that lies in the *reachable subspace*, which we denote by  $\mathcal{R}$ , and is defined to be the closure of  $\text{span}(B)$  under applying  $A$  (or equivalently, the span of  $B, AB, A^2B, \dots, A^{n-1}B$ ). Analogously to this, we define the *observable subspace*,  $\mathcal{O}$ , to be the closure of  $\text{span}(C^T)$  under applying  $A$ . (Or:  $\bar{\mathcal{O}}$  is the largest  $A$ -invariant subspace contained in the null space of  $C$ ; and  $\mathcal{R}$  is the largest  $A$ -invariant subspace contained in the image of  $B$ .)

If we define

1. The columns of  $P_{r\bar{o}}$  are an orthonormal basis for  $\mathcal{R} \cap \bar{\mathcal{O}}$ .
2. The columns of  $P_{ro}$  are an orthonormal basis of the complement of  $\mathcal{R} \cap \bar{\mathcal{O}}$  in  $\mathcal{R}$ .
3. The columns of  $P_{\bar{r}o}$  are an orthonormal basis of the complement of  $\mathcal{R} \cap \bar{\mathcal{O}}$  in  $\bar{\mathcal{O}}$ .
4. The columns of  $P_{\bar{r}\bar{o}}$  are an orthonormal basis of the the remainder of  $\mathbb{R}^n$ .

If we then define

$$P = \left[ \begin{array}{c|c|c|c} P_{r\bar{o}} & P_{ro} & P_{\bar{r}o} & P_{\bar{r}\bar{o}} \end{array} \right],$$

then

$$P^T P = \left[ \begin{array}{c|c|c|c} I & 0 & 0 & 0 \\ \hline 0 & I & U & 0 \\ \hline 0 & V & I & 0 \\ \hline 0 & 0 & 0 & I \end{array} \right].$$

*Check this. Can we get  $U = V = 0$ ?*

The following theorem can be found in SOME REFERENCE.

**Theorem 1 (Kalman decomposition)** For any system  $(A, B, C)$  with corresponding Kalman basis matrix  $P$ , the transformed system  $(PAP^{-1}, PB, CP^{-1})$  has the following form:

$$\hat{A} = PAP^{-1} = \begin{bmatrix} A_{r\bar{o}} & A_{r\bar{o},ro} & A_{r\bar{o},\bar{r}\bar{o}} & A_{r\bar{o},\bar{r}o} \\ 0 & A_{ro} & 0 & A_{ro,\bar{r}o} \\ 0 & 0 & A_{\bar{r}\bar{o}} & A_{\bar{r}\bar{o},\bar{r}o} \\ 0 & 0 & 0 & A_{\bar{r}o} \end{bmatrix},$$

and

$$\hat{B} = PB = \begin{bmatrix} B_{r\bar{o}} \\ B_{ro} \\ 0 \\ 0 \end{bmatrix},$$

and

$$\hat{C} = CP^{-1} = \begin{bmatrix} 0 & C_{ro} & C_{\bar{r}\bar{o}} & 0 \end{bmatrix}.$$

The transfer function of both systems is given by

$$H(z) = C_{ro}(zI - A_{ro})^{-1}B_{ro}.$$

In the latter case, we say that the system is *minimal* – there is no equivalent system with a smaller number of species. Note that this says that any two equivalent minimal systems are changes of basis of each other.

Since any system can be put into this form, and once in this form, its transfer function is determined only by  $C_{ro}$ ,  $A_{ro}$ , and  $B_{ro}$ , therefore, the set of all equivalent systems are parameterized by the dimension  $n$ , the choice of basis ( $P$ ), the remaining submatrices in  $\hat{A}$ ,  $\hat{B}$ , and  $\hat{C}$  (which are unconstrained), and an invertible transformation of  $\text{span}(P_{ro})$ , which we call  $T_{ro}$ .

**Theorem 2 (Parameterization of equivalent systems)** Let  $(A, B, C)$  be a minimal system.

- (a) Every equivalent system is of the form given in Theorem 1, i.e., can be specified by choosing a dimension,  $n$ ; submatrices in  $\hat{A}$ ,  $\hat{B}$ , and  $\hat{C}$  except for  $A_{ro} = A$ ,  $B_{ro} = B$ , and  $C_{ro} = C$ ; and choosing an invertible matrix  $P$ .
- (b) *conjecture*: The parameterization is unique if  $P$  is furthermore chosen so that each  $P_x$  other than  $P_{ro}$  is a projection matrix, and that

$$0 = P_x^T P_y$$

for all  $(x, y)$  except  $(ro, \bar{r}\bar{o})$ .

*Another way of saying it: pick the  $\mathcal{R}$  and  $\bar{\mathcal{O}}$  subspaces, that must intersect in something of the minimal dimension; then let  $P$  be the appropriate basis?*

In some situations we may be interested in only “network rewiring”, where  $A$  changes while  $B$  and  $C$  do not. For instance, if all non-regulatory functions of each molecule are strongly constrained, then  $C$  cannot change. Likewise, if responses of each molecule to the external inputs are not changed by evolution, then  $B$  does not change.

**Lemma 1** *Let  $(A, B, C)$  be a system. Every matrix  $\hat{A}$  such that  $(\hat{A}, B, C)$  is equivalent to  $(A, B, C)$  can be specified in the form given in Theorem 1, with *them having the same row bit* and satisfying*

$$\begin{aligned}\hat{B} &= PB \\ \hat{C}P &= C\end{aligned}$$

It is remarkable to note that even with the relationships between environment, cryptotype, and phenotype constant, and in the minimal dimension, there are still almost always degrees of freedom. These correspond to distinct genetic networks that perform indistinguishable functions. For example, the equivalent systems of Example 2 above are minimal, and share common  $B$  and  $C$  matrices.

*Maybe say that stability to noise can be determined by  $A_{\bar{r}_0}$  (and others?), while complexity of the internal, unobserved dynamics are somehow given by the difference between the dimension of  $y$  and the dimension of  $A_{r_0}$ .*

*General argument why Brownian motion on set of equivalent things is maybe ok.*

**Example 3** *The system in Example 2 is minimal, so  $A = A_{r_0}$ , and the only free parameter is  $P$ . Suppose also that the effects of the input and output are strongly constrained, so that as in that example,  $B$  and  $C$  cannot change, and neutral evolution moves through the set of  $P$  such that  $PB = B$  and  $CP = C$ . *Clarify we mean with no gene duplication and only neutral drift.* How will this system change over evolutionary time in the manner described *above*? An unbiased Brownian motion on this set can be written in terms of a one-dimensional Brownian motion  $\beta_\tau$  with  $\beta_0 = 0$  as*

$$P(\tau) = \begin{bmatrix} 1 & 0 \\ e^{\beta_\tau} - 1 & e^{\beta_\tau} \end{bmatrix}.$$

*Therefore, the state of the system after (evolutionary) time  $\tau$  is  $(P(\tau)AP(\tau)^{-1}, B, C)$ .*

## 4 What biological question are we asking?

How do gene regulatory networks (GRNs) change through evolutionary time? Specifically, under constant selection and environmental pressures, how does a maximally fit population's GRN change and what molecular and population parameters are significant to the process? How frequent does gene network rewiring occur in the absence of genetic drift or adaptation? Is rewiring typically compensatory and usually follow the fixation of a deleterious allele, or can networks neutrally reorganize? How does the size, average degree, function, pleiotropy, etc of a network influence its likelihood to drift? If networks can drift, how much topological heterogeneity can a well-mixed population tolerate, if any? Does topological heterogeneity and/or the size of the neutral genotype-set, if consequences of network organization, constrain or entail exaptation and evolvability? Are some network organizations (maybe higher dimensional realizations?) able to more rapidly evolve to construct novel phenotypes and meet the demands of changes in selection pressures?

## 5 How can we mathematically model this question?

In some respect an organism can be thought of as a black box that responds to some set of inputs (it's environment and other initial conditions) and outputs a phenotype. The phenotype is then

evaluated by natural selection. The black box metaphor holds, because much of the details of what happens to construct a phenotype are unimportant as far as selection is concerned, so long as the phenotype reliably performs its function. If multiple internal mechanisms can map the same set of inputs to the same set of outputs, and these mechanisms are mutationally nearby (in genotype space), then evolution may drift from mechanism to mechanism.

In this simple model we can say that  $B\vec{u}$  is a list of initial GRN protein concentrations determined by the more general initial environmental conditions  $\vec{u}$ .  $A$  represents all the genetic interactions of a particular GRN, and  $C\vec{x}$  is an organism's phenotype. Fitness could be calculated by comparing a subset of the phenotype to an optimum. Maybe,  $f(\cdot) = e^{-\int_a^b \|G(s) - G^*(s)\| ds}$ . Alternatively, fitness scores could simply assess whether or not a phenotype is within some range or breaks some threshold by some time point. Or possibly the simplest: any phenotype not exactly equal to  $G(s)$  is lethal, and  $G(s)$  is perfectly fit.

Next, we will have to define mutation and recombination. Mutation needs to add and remove genetic interactions and recombination needs to shuffle genes during sexual reproduction. We could adapt methods from Lynch 2007 and Lynch and Hagner 2015 to model the probability of a TFBS appearing or disappearing due to mutational pressure. Every time a new binding site is gained or lost, the values within the  $A$  matrix are either modified or rows and columns are added removed following a specific set of rules. I am not sure if it would be easier to maintain a matrix size (say  $m \times m$ ) where  $m$  is arbitrarily large, with most entries being zero, or if a matrix should only have as many dimensions as the number of active TFs. Recombination should then shuffle rows of the  $A$  matrix randomly, as each row represents the regulatory region of a given gene (assuming only cis-regulation, and that the regulatory element is small enough to not break up during recombination).

To study some of the above posed questions, we can simulate (or analytically determine) the rates of GRN change, and how frequently these changes lead to significant rewiring or speciation. We would want to know the dynamics of a brownian motion over the set of all realizations. Another interesting question would be to determine the size of the neutral genotype space as well as the number of connections from the neutral genotype space to the non-neutral genotype space. Maybe higher dimensional realizations of a GRN will have significantly more connections to non-neutral genotypes and therefore be more "evolvable" or exapted to novel environments?

Let  $M \in \mathbb{R}^{n \times n}$  be a least dimensional realization of  $A$ , and let  $M'$  be an  $(n + m \times n + m)$  matrix where  $M'_{(i,j)} = M_{(i,j)}$  for  $i, j \leq n$ , and 0 otherwise. Let  $P$  be any  $(n + m \times n + m)$  matrix with rank  $n$ . Then for any matrix  $B$  such that  $BP = PM'$ ,  $B$  is a realization of  $A$ .

## 6 List of papers with some notes.

R. E. Kalman, Mathematical description of linear dynamical systems, Journal of the Society for Industrial and Applied Mathematics, Series A: Control, vol. 1, no. 2, pp. 152-192, 1963. This is a good overview written by Kalman. Probably the most important paper on this list.

Minimal state-space realization in linear system theory: an overview. B. De Schutter. 2000. Another, more recent, overview that I found helpful to understand controllability and observability.

The Inverse Problem of Stationary Covariance Generation. BDO Anderson. 1969.

Equivalence of Linear Time-Invariant Dynamical Systems. BDO Anderson, BW Nocomb, RE Kalman. 1966.

B. Ho and R. E. Kalman, Effective construction of linear state-variable models from input/output functions, Proceedings of the 3rd Annual Allerton Conference on Circuit and System Theory, vol.



1, no. 2, p. 449459, 1965.

Dynamic Structure Functions for the Reverse Engineering of LTI Networks. J Goncalves. 2007. I have not read this paper yet, but apparently it outlines a method for recovering the state space equation from specific sets of measurements.

New Developments in Systems Theory Relevant to Biology. RE Kalman. 1968. Not technical but kind of interesting.

Old and New Directions of Research in Systems Theory. RE Kalman. 2010?

Minimality and Observability of Group Systems. HA Loeliger, GD Forney Jr., T Mittelholzer, MD Trott. 1994. Another paper I just found. It looks like it might be useful.

Partial Realization of Descriptor Systems. P Benner and VI Sokolov. 2006. I think this is about recovering a finite realization from infinite systems. Maybe it would be interesting to think about the similarity between transfer functions though? If selection only cares about a specific interval of the transfer function, say  $G(10)$ , then maybe two partial transfer functions will be effectively equivalent?

There are also the previous papers I sent on identifiability and distinguishability. In the original Bellman and Astrom paper where they introduce “structural identifiability” they show that a linear dynamical system is identifiable if and only if it is also controllable and observable.

## 7 Notes on what to do when allowing deleterious mutations

Epistasis, coming from mutations deleterious in some networks but not in others (come from incompatibilities).

Describe heterosis.