1 What are LTI systems?

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?

The inpulse data is summarized by the transfer function $G(s) = C(sI - A)^{-1}B + D$, a Laplace transform of the state space equation of the system. While the transfer function description of a system can be unique, the state space representation is not always unique.

The state space representation is as follows:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{1}$$

Where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^q$, $u \in \mathbb{R}^p$, $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times p}, C \in \mathbb{R}^{q \times n}, D \in \mathbb{R}^{q \times p}$.

The state space representation is considered an internal description of the system, whereas the transfer function is an external description. $\vec{x}$ is generally considered to be a state of the system, $\vec{u}$ to be the input, and $\vec{y}$ to be the observed output.

Multiple systems can interact in serial,

$$\begin{aligned} \dot{x}_1 &= A_1 x_1 + B_1 u \\ \dot{x}_2 &= A_2 x_2 + B_2 C_1 x_1 \\ y &= C x_2 \end{aligned} \tag{2}$$

or in parallel as well:

$$\begin{aligned} \dot{x}_1 &= A_1 x_1 + B_1 u \\ \dot{x}_2 &= A_2 x_2 + B_2 u \\ y &= C_1 x_1 + C_2 x_2 \end{aligned} \tag{3}$$

Let $P \in \mathbb{R}^{n \times n}$ be a non-singular matrix, and let $\dot{\overline{x}} = Px$, then

$$\begin{aligned} \dot{\overline{x}} &= \overline{A}\overline{x} + \overline{B}u \\ y &= \overline{C}\overline{x} + \overline{D}u \end{aligned} \tag{4}$$

be an algebraically equivalent (AE) "realization" of the system, where $\overline{A} = PAP^{-1}, \overline{B} = PB, \overline{C} = CP^{-1}$, and $\overline{D} = D$.

If multiple realizations of a system are algebraically equivalent, then they must also have identical transfer functions $G(s) = \overline{G}(s)$. However, the opposite relation is not necessarily true. Multiple systems can have identical transfer functions and not have algebraic equivlaence. In fact, two systems with identical transfer functions can even have different dimensions.

Two LTI systems $\{A, B, C, D\}$ and $\{\overline{A}, \overline{B}, \overline{C}, \overline{D}\}$ have "Zero-State Equivalence (ZSE)" if they have the same transfer function. (AE $\Rightarrow$ ZSE).

2   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 absense of genetic drift or adaptation? Is rewiring typically compensatory and ususally follow the fixation of a deleterious allele, or can networks neutrally reorganize? How does the size, average degree, function, peiotropy, 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 topoligical 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?

3   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 concered, 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 possbily 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' bean $(n + m \times n + m)$ 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.

4   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. 152192, 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 Necomb, 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 effecrively equivlaent?

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 identifiabile if and only if it is also controllable and observable.