A. Lane

A gene regulatory network can be described as a system of equations in which any gene , where is a set of all genes in a network, arbitrarily depends on gene such that for non-self-regulating genes. As such, a matrix can be used to describe the set of interdependencies between the sets of genes. This such matrix will be denoted as , with dimensions and is defined as such:

We will denote as the gene expression level of gene, where is a function of all. Furthermore, a steady state of a regulatory network corresponding to can be described as a set of functions. The steady state of said network, when expressed as a system of linear equations, can be described as matrix multiplication. We’ll denote matrix with same dimensions as, so that (where), and gene expression vector such that , where is a nonzero vector. For clarity, a nonzero value in indicates that the expression level for gene depends on gene by a factor of and implies. The following proof can be induced to find the steady state. Assume is an identity matrix.

However, for many purposes the regulatory network is unknown and thus must be found. One such way of finding is through a series of perturbations unto. Doing so will yield new steady states which can be observed to estimate. For the case of linear systems of equations, can also be found (or estimated, if noise is present). will be denoted as the steady state after perturbing gene by amount .

It will next be shown that genes will be perturbed given the steady state .

We’ll denote: , ,

This concludes the proof of all perturbations in a linear regulatory network.

Once the perturbations on all genes have been performed, the differences between their steady states from the original can be measured. The deviations can then be used to recover. We’ll denote these deviations as :

For clarity, is an matrix, where the column is equal to: .

The above equation can also be expressed as a dot product between two vectors. Since due to the genes being non self-regulating, we can omit a term in the dot product. This means any can be expressed as a sum of terms.

Vectors are not invertible. However row of (omitting element ) can be recovered by instead using a matrix by defining as a set of dot products, all of which use the same row in .

The matrix is invertible, and taking its inverse will recover row of .

For nonlinear networks however, linear algebra cannot be used find the steady-state of . Instead, the network is defined as a set of nonlinear functions.

Where denotes a set of indices of all nonzero values of row of matrix , denotes the power set of , denotes the relative activation of and is within the domain , and , where K and N are matrices who’s nonzero elements correspond to the nonzero elements of matrix . The values of and are the dissociation constants and Hill coefficients, respectively. Here, a steady state can be found by converging an arbitrary vector toward by using the property: such that:

For the nonlinear case, can also be estimated by inducing perturbations on all genes as with the linear case. What differs, however, is the method of induction. We’ll again denote as the steady state of the expression levels of the regulatory network when perturbing gene by . The perturbations can be simulated by asserting:

We can take advantage of the nonlinear convergence property to find:

such that: for all iterations .

As with the linear case, the deviations are: . They can be used as before to recover a linear matrix to estimate the set of dependencies defined by :

This however, is a naïve approach. The perturbations imply: . The contrapositive statement: Can be used to estimate , but not with high precision especially in cases where there are cycles in the regulatory network. We’ll denote a cycle as a set of genes: . In those such cases, a perturbation may cause a recovery error due to it perturbing such that an indirect dependency exists: , but the recovered matrix implies that there exists a direct dependency: .

We’ll next take into account a level of molecular noise for both linear and nonlinear cases by introducing a Gaussian random variable of a normal distribution with mean 0 and standard deviation .

This will cause fluctuations in and affect the recovery process. Precision will be lost due to the noise in the linear case, and in the nonlinear case the noise will cause further errors in the recovery of addition to the errors already present.

In both the linear and nonlinear cases, the number of errors in the recovered matrix is proportional to the factor of noise. As the factor of noise increases linearly, the average number of errors also increases linearly. The number of errors can be minimized by maximizing the perturbations for the linear case. This is due to the genes having dependencies of linear magnitude, and so when the perturbation of gene is increased by a large factor, the deviations will also increase by a linear magnitude, and so the noise can be dwarfed in comparison (i.e. ). This is not necessarily true for the nonlinear case due to the fact that genes may not share a linear dependence.