

Remarks on structural identification, modularity, and retroactivity

Eduardo D. Sontag

Abstract—The “Modular Response Analysis” (MRA) reverse-engineering algorithm has been utilized in the systems biology literature in order to unravel the interactions between different “modules” in biochemical networks. We review the MRA technique, and discuss why its (mis-)application might lead to incorrect results due to impedance or load (“retroactivity”) effects, and present a computational criterion to discover stoichiometrically-dependent steady-state retroactivity.

I. INTRODUCTION

Reverse-engineering techniques in systems biology attempt to unravel the interactions among the components (“modules”) of biomolecular (signaling, gene regulatory, metabolic) networks, often using data gathered from steady-state perturbation experiments. Perturbations are done to particular gene or signaling components, by means of traditional genetic experiments such as gene knock-down or overexpression, RNA interference, hormones, growth factors, or pharmacological interventions such as kinase and phosphatase inhibitors. Based on steady-state changes in concentrations of active proteins, mRNA levels, transcription rates, and so forth, the objective is to map out the direct or “local” interactions among components, which capture the topology of the functional network, with the ultimate goal of elucidating the mechanisms underlying observed behavior (phenotype).

A graph is used to summarize the deduced interactions. For example, if there are two components “ A ” and “ B ”, one might perform an up-perturbation in A : if this leads to an increased value of B , a directed edge $A \rightarrow B$ labeled “activation” is introduced; if, instead, it leads to a decreased level of B , then an edge labeled “repression” is drawn; and if there is no effect on B , no edge is put into the graph.

A major problem with such steady-state (or even some time-resolved) experiments is that perturbations might propagate rapidly throughout the network, causing “global” changes which cannot be easily distinguished from direct effects. To illustrate this difficulty, consider the two graphs shown in Figure 1 (arrows indicate activation). An up-



Fig. 1. Serial vs. feedforward architectures

perturbation of the external signal u (or of the block labeled “ x ”) results in an increase in the signals represented by the block labeled “ y ” in both cases. How does one distinguish

the two architectures? The “unraveling,” or “Modular Response Analysis” (MRA) method proposed in [3] and further elaborated upon in [8], [1] and other papers, provides one approach to solving this global-to-local problem.

The MRA experimental design compares those steady states which result after performing independent perturbations to each “modular component” of a network. These perturbations might be genetic or biochemical. For example, in eukaryotes they might be achieved through the down-regulation of mRNA, and therefore protein, levels by means of RNAi. For example, the authors of [6] uncovered connectivity differences in the Raf/Mek/Erk MAPK network in rat adrenal pheochromocytoma (PC-12) cells depending on whether the cells are stimulated with epidermal growth factor (EGF) or instead with neuronal growth factor (NGF). They employed MRA in order to discover positive and negative feedback effects, using the algorithms from [8] and [1].

One possible source of errors when applying MRA is that “impedance” or “load” effects (*retroactivity* in the language of [5], see [2]) may lead to wrong conclusions when one only measures “communicating intermediates” among modules – there are subtle conditions for applicability of MRA, often not checked by those using the algorithms. When applying MRA in a modular fashion, only perturbation data on these communicating signals are collected. The connectivity strength among a pair of such intermediary signals, such as levels of activated signaling proteins, is estimated, but this apparent connectivity might not be due to a “directed” biochemical interaction. Specifically, stoichiometric constraints (conservation laws) might lead to misleading conclusions. In Section II we briefly review MRA and discuss this issue, and in Section III we present a method for detecting (exactly) retroactivity effects at steady state in a known network.

II. MODULAR RESPONSE ANALYSIS

The basic idea of MRA is easy to understand through a very simple example. Suppose that we wish to distinguish between the two possible architectures schematically shown in Figure 1. In general, components may be described by single variables, or by many variables; for instance, a gene expression component might be described at various levels of resolution: using one variable (resulting protein levels), or by a far more complicated mechanism (including binding and unbinding of transcription factors, transcription initiation and mRNA polymerase dynamics, ribosome binding and translation dynamics, etc.). For simplicity, let us take the simple model in which each component is described by a scalar linear system, so that both possible architectures are

E-mail: sontag@math.rutgers.edu. Dept. of Mathematics, Rutgers University, Piscataway, NJ

special cases of:

$$\begin{aligned}\dot{x} &= -ax + bu \\ \dot{y} &= cx - dy + pu\end{aligned}$$

where all parameters are positive but otherwise unknown ($a, d > 0$, so the model is stable). The question that we are interested in is that of deciding whether $p = 0$ or $p \neq 0$. (Obviously, it would be difficult to distinguish a small $p \neq 0$ from $p = 0$, if measurements are noisy. We assume for this introductory discussion that measurements are exact.)

The available data are the steady states for both x and y , for a constant (but unknown) input u , under these three “experimental” scenarios: (1) $a = a_0, u = u_0$; (2) $a = a_0, u = u_1$; (3) $a = a_1, u = u_0$. Both u and a are unknown, except for the information that they change, one at a time, in experiments (2) and (3). (Respectively, a change in the concentration of u and a change in the degradation rate of x .) The steady state, obtained by setting $\dot{x} = \dot{y} = 0$, is, for constant u , $x(\infty) = (b/a)u$ and $y(\infty) = (cb/a + p)u/d$. We write $x_{\Delta u} = (b/a_0)u_1 - (b/a_0)u_0 = (b/a_0)\Delta u$, the difference between the measured steady state of x for experiments (2) and (1), and the corresponding quantity $y_{\Delta u} = (cb/a + p)\Delta u/d$ for y . Similarly, subtracting the data from experiments (3) and (1) provides the measured quantities $x_{\Delta a} = (1/a_1 - 1/a_0)bu_0$ and $y_{\Delta a} = (1/a_1 - 1/a_0)cbu_0/d$. Next, we compute from the data: $y_{\Delta u}/x_{\Delta u} - y_{\Delta a}/x_{\Delta a} = a_0p/(bd)$. We are done: if this number is zero, then $p = 0$ (cascade architecture), and if it is nonzero, then $p \neq 0$ (feedforward architecture). Our objective of distinguishing between the two structures has been achieved.

Moreover, we can even recover the numerical value $y_{\Delta a}/x_{\Delta a} = c/d$, and if u_0 or u_1 were also known, then we would also be able to compute b/a_0 from the steady state value of x , and hence we would also obtain the value of p/d , as $(b/a_0) \cdot (a_0p/(bd))$. In this way, the relative strengths of all the terms in the equation for \dot{y} can be computed. This is the best that one can do, because the actual values of all three constants can never be obtained from purely steady-state data, since multiplying all constants by the same number doesn’t affect the steady state location.

The MRA method generalizes the procedure just discussed. We consider systems

$$\dot{x} = f(x, p)$$

where $x = (x_1, \dots, x_n)$ is the state and $p = (p_1, \dots, p_m)$ is a vector of parameters. Parameters can be manipulated, but, once changed, they remain constant for the duration of the experiment. We will assume that $m \geq n$. In biological applications, the variables x_i might correspond to the levels of protein products corresponding to n genes in a network, and the parameters to translation rates, controlled by RNAi. Another example would be that in which the parameters represent total levels of proteins, whose half-lives are long compared to the time scale of the processes (such as phosphorylation modifications of these proteins in a signaling pathway) described by the variables x_i . Yet another

example would be one in which the parameters represent concentrations of enzymes that control the reactions, and whose turnover is slow. The goal is to obtain, for each pair of variables x_i and x_j , the relative signs and magnitudes of the partial derivatives $\frac{\partial f_i}{\partial x_j}$, which quantify the *direct* effects of each variable x_j upon each variable x_i . The entries of $\partial f_i/\partial x_j$ of the Jacobian F of f with respect to x are functions of x and p . The steady-state version of MRA estimates this Jacobian when $x = \bar{x}$ is an “unperturbed” steady state attained when the vector of parameters has an “unperturbed” value $p = \bar{p}$. The steady-state condition means that $f(\bar{x}, \bar{p}) = 0$. Ideally, one would want to find the matrix F , since this matrix completely describes the influence of each variable x_j upon the rate of change of each other variable x_i . However, since for any parameter vector p and associated steady-state x , $f(x, p) = 0$ implies that $\Lambda f(x, p) = 0$, for any diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, the best that one could hope for is for steady state data to uniquely determine each of the rows $F_i = (F_{i1}, \dots, F_{in}) = \nabla f_i$, $i = 1, \dots, n$ of F only up to a scalar multiple.

The critical assumption for MRA, and indeed the main point of [3], [4], [8], is that, while one may not know the detailed form of the vector field f , often one does know which parameters p_j directly affect which variables x_i . For example, x_i may be the level of activity of a particular protein, and p_i might be the total amount (active plus inactive) of that particular protein; in that case, we might postulate that p_i only directly affects x_i , and only indirectly affects the remaining variables. Under this assumption, the steady-state MRA experimental design consists of the following steps:

- 1) measure a steady state \bar{x} corresponding to the unperturbed vector of parameters \bar{p} ;
- 2) separately perform a perturbation to each entry of \bar{p} , and measure a new steady state.

The “perturbations” are assumed to be small, in the sense that the theoretical analysis will be based on the computation of derivatives. Under mild technical conditions, this means that a perturbed steady state can be found near \bar{x} . There are $m + 1$ experiments, and n numbers (coordinates of the corresponding steady state) are measured in each. (In practice, of course, this protocol is repeated several times, so as to average out noise and obtain error estimates.) Using these data (and assuming that a certain independence condition is satisfied), it is possible to calculate, at least in the ideal noise-free case, the Jacobian of f , evaluated at (\bar{x}, \bar{p}) , except for the unavoidable scalar multiplicative factor uncertainty on each row.

We assume given a parameter vector \bar{p} and state \bar{x} such that $f(\bar{x}, \bar{p}) = 0$ and so that the following generic condition holds for the Jacobian of f : $\det F(\bar{x}, \bar{p}) = \det \frac{\partial f}{\partial x}(\bar{x}, \bar{p}) \neq 0$. Therefore, we may apply the implicit function theorem and conclude the existence of a mapping φ , defined on a neighborhood of \bar{p} , with the property that, for each row i ,

$$f_i(\varphi(p), p) = 0 \quad \text{for all } p \approx \bar{p}, \quad (1)$$

and $\varphi(\bar{p}) = \bar{x}$ (and, in fact, $x = \varphi(p)$ is the unique state x near \bar{x} such that $f(x, p) = 0$).

We next discuss how one reconstructs the gradient $\nabla f_i(\bar{x}, \bar{p})$, up to a constant multiple. (The index i is fixed from now on, and the procedure must be repeated for each row f_i .) We do this under the assumption that it is possible to apply $n - 1$ independent parameter perturbations: there are $n - 1$ indices j_1, j_2, \dots, j_{n-1} with the following two properties:

- (a) f_i does not depend directly on any p_j : $\partial f_i / \partial p_j \equiv 0$, for $j \in \{j_1, j_2, \dots, j_{n-1}\}$, and
- (b) the vectors $v_j = (\partial \varphi / \partial p_j)(\bar{p})$, for these j 's, are linearly independent.

We then have, taking total derivatives in (1): $\nabla f_i(\bar{x}, \bar{p})v_j = 0$, $j \in \{j_1, j_2, \dots, j_{n-1}\}$. Thus, the vector $F_i = \nabla f_i(\bar{x}, \bar{p})$ which we wish to estimate is known to be orthogonal to the $n - 1$ dimensional subspace spanned by $\{v_1, \dots, v_{n-1}\}$. Therefore, it is uniquely determined, up to multiplication by a positive scalar. The row vector F_i satisfies $F_i \Sigma = 0$ where Σ is defined as the $n \times (n - 1)$ matrix whose columns are the v_i 's. Generically, we assume that the rank of Σ is $n - 1$. Thus, F_i can be computed by Gaussian elimination, as a vector orthogonal to the span of the columns of Σ . Of course, the sensitivities represented by the vectors v_i (entries of the matrix Σ) cannot be directly obtained from typical experimental data. However, approximating the vectors v_j by finite differences, one has that $\nabla f_i(\bar{x}, \bar{p})$ is approximately orthogonal to these differences as well.

Noise: We refer the reader to [1] for modifications to allow for noisy measurements, using total least-squares.

Modular Approach: Let us suppose that the entire network consists of an interconnection of n subsystems or "modules", each of which is described by a set of differential equations such as:

$$\begin{aligned} \dot{x}_j &= g_j(\mathbf{y}_j, x_1, \dots, x_n, p_1, \dots, p_m), \quad j = 1, \dots, n \\ \dot{\mathbf{y}}_j &= G_j(\mathbf{y}_j, x_1, \dots, x_n, p_1, \dots, p_m), \quad j = 1, \dots, n, \end{aligned}$$

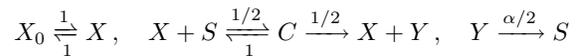
where the variables x_j represent "communicating" or "connecting" intermediaries of module j that transmit information to other modules, whereas the *vector variables* \mathbf{y}_j represent chemical species that interact within module j . Each vector \mathbf{y}_j has dimension ℓ_j . The integers ℓ_j , $j = 1, \dots, n$ are in general different for each of the n modules, and they represent one less than the number of chemical species in the j th module respectively. Observe that, for each j , the rate of change of the communicating variable depends only on the remaining communicating variables x_i , $i \neq j$, and on the variables \mathbf{y}_j in its own block, but does not directly depend on the internal variables of other blocks. In that sense, we think of the variables \mathbf{y}_j as "hidden" (except from the communicating variable in the same block).

We assume that each Jacobian of G_j with respect to the vector variable \mathbf{y}_j , evaluated at the steady state corresponding to \bar{p} (assumed to exist, as before) is nonsingular. By the Implicit Mapping Theorem, in a neighborhood of this steady state, we may solve $G_j(\mathbf{y}_j, x, p) = 0$ (x denotes the vector x_1, \dots, x_n , and similarly for p) for the vector variable \mathbf{y}_j , as a function of x, p , locally by a function $\mathbf{y}_j = \mathbf{M}_j(x, p)$.

Those steady states that are obtained by small perturbations of \bar{p} are the same as the steady states of the "virtual" system $\dot{x}_j = h_j(x_1, \dots, x_n, p_1, \dots, p_m) = g_j(\mathbf{M}_j(x, p), x, p)$, $j = 1, \dots, n$. Now the analysis continues as before, using the h_j 's instead of the f_j 's.

Using quasi-steady state data: In certain problems, steady-state data do not provide enough information, for example when the system "adapts" to perturbations. A variant of MRA, which allows for the use of general non-steady-state, time-series data was developed in [8]. However, that method requires one to compute second-order time derivatives, and hence is hard to apply when time measurements are spaced far apart and/or are noisy. An intermediate possibility is to use *quasi-steady state data*, meaning that one employs data collected at those times when a variable has been observed to attain a local maximum (peak of activity) or a local minimum. (This is the approach taken in [6], which, for EGF stimulation, measured network responses at the time of peak Erk activity, approximately 5 minutes, and not at steady state.) The paper [7] provided an extension to quasi-steady state MRA.

Effect of stoichiometric constraints on MRA: We will use a very simple example in order to show how stoichiometric constraints may lead to misleading results when using the MRA algorithm. Suppose that we want to study a system in which we postulate that there are two "modules" involving enzymes X and Y , the "active forms" of which are the "communicating variables." The active form X is reversibly produced from an inactive form X_0 , and Y is formed when X reversibly binds to a substrate S thus producing a complex C , which may dissociate into X and S or into X and Y (this is a standard Michaelis-Menten type of reaction). We also assume that Y can revert to S in one step; a more complicated model could be used as well, by modeling the phosphatase action in a Michaelis-Menten form, or by modeling mechanistically its binding and unbinding to Y , but the principle is the same. The network of reactions is as follows:



and we consider experiments in which $X(0)=3$, $C(0)=Y(0)=0$, and $S(0)=\beta$, and either α or β is to be perturbed experimentally. We think of X_0 and X as constituting the first "module" and S, C, Y as the second one. The two parameters α, β are usually viewed as affecting only the second module. The unique positive steady state (X, S, C, Y) is then obtained by solving: $2X + C = 3, C = XS, C = \alpha Y, S + C + Y = \beta$ (and $X_0 = X$). We will consider perturbations around $\alpha=1$ and $\beta=3$. For these nominal parameter values, $(X, S, C, Y) = (1, 1, 1, 1)$. Taking implicit derivatives with respect to α , evaluating at $X=S=C=Y=1$, $\alpha=1$, $\beta=3$, and denoting $x = \partial X / \partial \alpha$, $u = \partial S / \partial \alpha$, $v = \partial C / \partial \alpha$, $y = \partial Y / \partial \alpha$, we have that: $2x + v = 0, v = x + u, v = 1 + y, u + v + y = 0$ which solves to: $x = -1/7, u = -3x, v = -2x, y = 5x$ and thus "dX/dY" computed as $\frac{\partial X / \partial \alpha}{\partial Y / \partial \alpha}$ equals $1/5 > 0$. The MRA

method, or any other sensitivity-based approach, applied to the phenomenological model in which only X and Y are “communicating intermediaries” will lead us to include an edge $Y \rightarrow X$ labeled “activating”. But *such an edge does not represent a true feedback effect*: for example, it is not possible to delete this edge with a “mutation” in the system that does not affect the forward edge. The edge merely reflects a “loading” or impedance effect. In fact, the situation is even more confusing. Arguing in the same way with derivatives with respect to β (evaluated at the same values), one obtains $x = -1/7, u = -3x, v = -2x, y = -2x$ and thus “ dX/dY ” computed as $\frac{\partial X/\partial \beta}{\partial Y/\partial \beta}$ is $-1/2 < 0$. Now the (false) effect is inhibitory. (The intuition is that when we increase α , the substrate for X increases, sequestering more of X , and also D is smaller. If instead we over-express B , then both X is sequestered more and D is larger. But intuition is not enough: for some parameters, $dX/dY < 0$ for both experiments.) Experimentally, it is often the case that one measures $X + C$ and Y , instead of X and Y , so that one would be interested in the relative variations of $\hat{x} = x + v$ and y . Since $2x + v = 0$, it follows that $\hat{x} = -x$. Thus, $d(X + C)/dY = -dX/dY$, so the signs are reversed, but are, again, ambiguous.

Of course, there is a simple explanation for the problem: the parameter α affects the differential equation for X , and the variables S and C in fact enter that differential equation. Thus, the conditions for applicability of MRA have been violated. The point, however, is that a naive application of sensitivity analysis (as usually done in practice) that does not account for these subtle dependencies is wrong. One way to avoid this potential pitfall is to insure that the postulated mechanism (without additional feedback loops) does not exhibit such “load” effects. The next section presents an algorithm to detect such effects (at steady state).

III. RETROACTIVITY AT STEADY STATES

We now turn to characterizations of steady-state retroactivity (we do not define “retroactivity” but instead use the term only informally; the results to be given provide a precise content to the term under slightly different contexts). The main issue is to understand what is the relation between the steady states of individual systems (described by chemical reactions) and the steady states of their interconnection. One expects that retroactivity at steady state arises only when there are more conservation laws imposed by an interconnection, in addition to those that hold for each of the interconnected systems separately. Making this intuition precise is not completely trivial. In fact, unless certain properties are imposed on interconnections, it is not even correct. Our main results are Theorems 1 and 2. They give sufficient conditions for retroactivity to exist or not, respectively. Neither is necessary. However, we will define a “consistency” property for interconnections, under which Theorem 1 and Theorem 2 constitute a dichotomy.

Chemical reaction networks: We assume that the entries of $S = \text{col}(S_1, \dots, S_N)$ are the (concentrations of the) reacting species, and each $S_i = S_i(t)$ is a non-negative function

of time, evolving according to $\dot{S} = \Gamma R(S)$. The matrix $\Gamma \in \mathbb{R}^{N \times r}$ is the *stoichiometry matrix*, and $R(S) \in \mathbb{R}^r$ is the vector of *reactions*: $R(S(t))$ indicates the values of the reaction rates when the species concentrations are $S(t)$. A technical assumption is that solutions that start non-negative remain so. This property is automatically satisfied for the usual chemical reaction rate forms. We will also assume that each $S_0 \in \mathbb{R}_{\geq 0}^N$ the solution $\varphi(t, S_0)$ of $\dot{S} = \Gamma R(S)$ with $S(0) = S_0$ is defined for all $t \geq 0$.

For any chemical reaction system $\dot{S} = \Gamma R(S)$, and any state S_0 , the *stoichiometry equivalence class* of S_0 , denoted here as $\Delta(S_0)$, is the intersection of the affine manifold $S_0 + \Delta$ with $\mathbb{R}_{\geq 0}^N$, where Δ is the span of the columns of Γ . Thus, two states S_0 and S_1 are in the same stoichiometry class if and only if $S_0 - S_1 \in \Delta$, or equivalently if $\Delta(S_0) = \Delta(S_1)$. Observe that $\varphi(t, S_0) \in \Delta(S_0)$ for all $t \geq 0$. Moreover, since $\Delta(S_0)$ is a closed set, any S in the closure of the forward orbit $\mathcal{O}^+(S_0) = \{\varphi(t, S_0), t \geq 0\}$ is also in $\Delta(S_0)$.

We also introduce the vector space of “conservation laws”. This is the set of all vectors perpendicular to the stoichiometry space, written as rows: $\Pi := \Delta^\perp = \{\pi \in \mathbb{R}^{1 \times N} \mid \pi \Gamma = 0\}$. Observe that a state S_1 is in the stoichiometry class of a state S_0 (that is, $S_1 - S_0 \in \Delta$) iff $\pi(S_1 - S_0) = 0$ for all $\pi \in \Pi$.

For any chemical reaction system $\dot{S} = \Gamma R(S)$, and any steady state \bar{S} (that is, $\Gamma R(\bar{S}) = 0$) we say that \bar{S} is a *local attractor relative to its stoichiometry class* if there is some neighborhood \mathcal{U} of \bar{S} in $\mathbb{R}_{\geq 0}^N$ such that, for each $S_0 \in \mathcal{U} \cap \Delta(\bar{S})$, $\varphi(t, S_0) \rightarrow \bar{S}$. A *positive* state S is one for which all components are strictly positive, that is, $S \in \mathbb{R}_{> 0}^N$. Under certain hypotheses on the structure of the chemical reaction network, one may insure that in each stoichiometry class there is at least one positive steady state that is a local attractor relative to the class. Moreover, this steady state is often unique and is a global attractor relative to the class (deficiency theory).

Modular systems: We will assume that the vector S of species has $N = n + m$ components, which we partition into two vectors $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$: $S = (x', z')'$ (we use primes to indicate transpose). Corresponding to these coordinates, the reaction vector is partitioned into two vectors $R_1(x)$ and $R_2(x, z)$ of dimensions r_1 and $r_2 = r - r_1$ respectively: $R(x, z) = (R_1(x)', R_2(x, z)')$. We also assume that, in terms of this partition, the stoichiometry matrix looks as follows:

$$\Gamma = \begin{pmatrix} P \\ Q \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where $P \in \mathbb{R}^{n \times r}$, $Q \in \mathbb{R}^{m \times r}$, $A \in \mathbb{R}^{n \times r_1}$, $B \in \mathbb{R}^{n \times r_2}$, $C \in \mathbb{R}^{m \times r_2}$. Sometimes it is more convenient to use the “ (P, Q) ” form or the “ (A, B, C) ” form. The equations for the system take the following partitioned form:

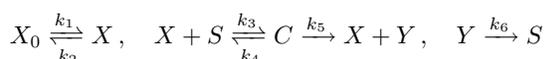
$$\begin{aligned} \dot{x} &= A R_1(x) + B R_2(x, z) \\ \dot{z} &= C R_2(x, z) \end{aligned}$$

If there are no reactions involving x alone, we write $A = 0$, thought of as an $n \times 1$ matrix. Of course, the actual reactions entering x and z need not be the same, since B and C may

multiply different elements of the vector $R_2(x, z)$ by zero coefficients. We think of the overall system as an interconnection of the “upstream” subsystem described by the x -variables, that feeds a signal to the “downstream” subsystem described by the z variables. The “ x ” appearing in $CR_2(x, z)$ is seen, in that sense, as an input signal to the second system. The role of $BR_2(x, z)$ is different. This term represents the “retroactivity to the output,” denoted by the letter “ r ” in [2], and is interpreted as a “load” effect that arises due to the physical interconnection. Of course, these interpretations are subjective, and partitioning a system into an interconnection can be done in non-unique ways. However, the questions to be posed depend on one such partition. The system $\dot{x} = AR_1(x)$ is by definition the *isolated system*, and the full system $\dot{S} = \Gamma R(S)$ the *interconnected system*.

We use the notation $\Delta_1(x)$ for the stoichiometry class of a state x of the isolated system: $\Delta_1(x) = (x + \Delta_1) \cap \mathbb{R}_{\geq 0}^n$, where Δ_1 is the span of the columns of A .

Example 3.1: Our first example is this network:



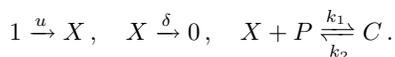
where the upstream system is described by the vector $x = (X_0, X)'$ which drives the downstream system described by the vector $z = (S, C, Y)'$. Thus, with $n = 2$, $m = 3$, $r_1 = 2$, and $r_2 = 3$, we take, using mass-action kinetics, $R_1(x) = (k_1 X_0, k_2 X)'$ and $R_2(x, z) = (k_3 X S, k_4 C, k_5 C, k_6 Y)$. Note that

$$A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 \end{pmatrix}$$

$$C = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

There are conservation laws in this system which tie together the isolated (x) system to the downstream (z) system, and one may expect that retroactivity effects appear. Indeed, this system will satisfy the sufficient condition for retroactivity given in Theorem 1 below. \square

Example 3.2: Consider these reactions:

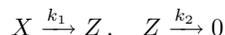


Then, with $x = X$ and $z = (P, C)'$, and listing reactions in the obvious order:

$$A = \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Because of the production and/or decay of X , there are no conservation laws tying together the X and the P, C systems, and there is no retroactivity effect. Indeed this system will satisfy the sufficient condition in Theorem 2 for non-retroactivity. \square

Example 3.3: Consider the following reaction:



with $x = X$ and $z = Z$. Here $B = (-1 \ 0)$ and $C = (1 \ -1)$, and $A = 0$. This example is one in which there are no

conservation laws whatsoever, yet retroactivity holds. Neither Theorem 1 nor Theorem 2 applies to this example, showing the gap between the conditions. However, this example is somewhat pathological, as it represents an “inconsistent” interconnection in the sense defined below. \square

Main results: Consider the following property:

$$\text{rank} \begin{pmatrix} P \\ Q \end{pmatrix} = \text{rank } P + \text{rank } Q \quad (*)$$

Remark 3.4: Since the weak inequality “ \leq ” is always true, the negation of (*) is equivalent to:

$$\text{rank} \begin{pmatrix} P \\ Q \end{pmatrix} < \text{rank } P + \text{rank } Q \quad (*')$$

or, equivalently, the requirement that the row spaces of P and Q have a nonzero intersection. \square

If property (*) holds, then there is retroactivity at steady state. The precise statement is as follows:

Lemma 3.5: Suppose that Property (*) does not hold. Then, for each positive state $\bar{S} = (\bar{x}', \bar{z}')'$ of the interconnected system, there exists a state $S_0 = (x'_0, z'_0)'$ such that $\Delta(S_0) = \Delta(\bar{S})$ but $\Delta_1(x_0) \neq \Delta_1(\bar{x})$. Moreover, S_0 can be picked arbitrarily close to \bar{S} .

Proof: Suppose that (*) holds, and pick any positive state \bar{S} . By (*'), there is some nonzero row vector θ which is in the row spaces of P and Q , that is to say, there are two row vectors μ_0 and ν_0 such that $\theta = \mu_0 P = \nu_0 Q \neq 0$. Replacing ν_0 by $-\nu_0$, we will assume that $\mu_0 P = -\nu_0 Q \neq 0$. Let r be any vector such that $\mu_0 P r \neq 0$ (for example, one may pick $r = P' \mu'_0$), and let $u := P r$ and $v := Q r$. Let, as earlier, $\Pi := \Delta^\perp$. Note that, for each $\pi = (\mu, \nu) \in \Pi$, $\mu P + \nu Q = 0$, by definition of Π , and therefore also $\mu u + \nu v = (\mu P + \nu Q) r = 0$. In particular, $(\mu_0, \nu_0) \in \Pi$ satisfies that $\mu_0 u = \mu_0 P r \neq 0$ and also $\nu_0 v = \nu_0 Q r = -\mu_0 P r \neq 0$. Since $P = (A \ B)$ and $Q = (0 \ C)$, every element $(\mu, \nu) \in \Pi$ has the property that, in particular, $\mu A = 0$.

Notice that one could pick u and v as close to zero as wanted (multiplying, if necessary, u and v by a common small positive factor). So, without loss of generality, we assume that both $x_0 := \bar{x} + u$ and $z_0 := \bar{z} + v$ are non-negative, and write $S_0 := (x'_0, z'_0)$. We claim that S_0 and \bar{S} are in the same stoichiometry class. Indeed, for any $(\mu, \nu) \in \Pi$: $\mu \bar{x} + \nu \bar{z} = \mu \bar{x} + \nu \bar{z} + 0 = \mu \bar{x} + \nu \bar{z} + \mu u + \nu v = \mu(\bar{x} + u) + \nu(\bar{z} + v) = \mu x_0 + \nu z_0$.

Finally, we claim that x_0 and \bar{x} are not in the same stoichiometry class for the isolated system. Since $\mu_0 A = 0$, μ_0 is a conservation law for the isolated system. So it will be enough to show that $\mu_0 \bar{x} \neq \mu_0 x_0$. Indeed, $\mu_0 \bar{x} = \mu_0(x_0 - u) = \mu_0 x_0 - \mu_0 u$, and $\mu_0 u \neq 0$. \blacksquare

Lemma 3.5 implies a steady-state retroactivity effect, in the following sense. Suppose that \bar{S} is an attractor for points near it and in $\Delta(\bar{S})$. If x_0 is taken as the initial state of a trajectory $x(t)$ for the isolated system, then every limit point ξ of this trajectory is in $\Delta_1(x_0)$. On the other hand, if the composite system is initialized at this same state x_0 for the x -subsystem, and at z_0 for the z -subsystem, then the ensuing trajectory converges to the steady state \bar{S} , with x -component

\bar{x} . But $\xi \neq \bar{x}$, because $\bar{x} \notin \Delta_1(x_0)$. The following result formalizes this fact.

Theorem 1: Suppose that there is some positive steady state $\bar{S} = (\bar{x}', \bar{z}')'$ of the interconnected system which is a local attractor relative to its stoichiometry class. If Property (*) is false, then there exist x_0 and z_0 such that, with the initial condition $S_0 = (x'_0, z'_0)'$:

- 1) $\varphi(t, S_0) \rightarrow \bar{S}$ as $t \rightarrow +\infty$, but
- 2) for the solution $x(t)$ of the isolated system $\dot{x} = AR_1(x)$ with $x(0) = x_0$, $\bar{x} \notin \text{clos} \{x(t), t \geq 0\}$.

Proof: We use Lemma 3.5. Let S_0 be as there. Since S_0 can be picked arbitrarily close to \bar{S} and in $\Delta(S)$, we may assume that S_0 belongs to the domain of attraction of the steady state \bar{S} . Property (1) in the Theorem statement is therefore satisfied. Finally, we consider the solution $x(t)$ of the isolated system $\dot{x} = AR_1(x)$ with initial condition $x(0) = x_0$, and pick any state $\xi \in \text{clos} \{x(t), t \geq 0\}$. As $\Delta_0(\xi) = \Delta_0(x_0) \neq \Delta_0(\bar{x})$, it follows that $\xi \neq \bar{x}$. ■

Next, consider the following property:

$$\text{rank}(AB) = \text{rank}A \quad (**)$$

i.e., the column space of B is included in that of A . Note that if this condition holds, then (*) holds too.

Lemma 3.6: Suppose that (**) holds. Pick any two states $\bar{S} = (\bar{x}', \bar{z}')'$ and $S_0 = (x'_0, z'_0)'$ of the interconnected system. Then

$$\Delta(\bar{S}) = \Delta(S_0) \Rightarrow \Delta_1(\bar{x}) = \Delta_1(x_0). \quad (2)$$

Proof: As $\bar{S} - S_0$ belongs to the column space Δ of Γ , in particular, $\bar{x} - x_0$ is in the column space of (AB) . Since the latter equals the column space of A , it follows that $\bar{x} - x_0$ is in the column space of A , which means that x_0 and \bar{x} are in the same stoichiometry class in the isolated system. ■

Lemma 3.5 implies a steady-state retroactivity effect, in the following sense. Suppose that there is a unique steady state in each stoichiometry class in the isolated system, and that this steady state is a global attractor relative to its class. Then, every omega-limit point of the composite system has the property that its x -component equals this same steady state of the isolated system. The following result formalizes this discussion..

Theorem 2: Suppose that (**) holds. For any initial condition $S_0 = (x'_0, z'_0)'$, if a state $\bar{S} = (\bar{x}', \bar{z}')'$ of the interconnected system is in the omega-limit set of S_0 , then x_0 and \bar{x} are in the same stoichiometry class relative to the isolated system.

Proof: If $\bar{S} = (\bar{x}', \bar{z}')'$ is in the omega-limit set of S_0 then $\Delta(S_0) = \Delta(\bar{S})$. The conclusion thus follows from Lemma 3.6. ■

There is a gap between the negation of Property (*) in Theorem 1 and Property (**) in Theorem 2. In order to bridge this gap, we introduce the following property:

$$\ker C \subseteq \ker B \quad (C)$$

which we call *consistency*.

An interpretation of property (C) is as follows. Suppose that $S = (\bar{x}', \bar{z}')'$ is a steady state of the interconnected

system. That is to say, $AR_1(\bar{x}) + BR_2(\bar{x}, \bar{z}) = 0$ and $CR_2(\bar{x}, \bar{z}) = 0$. Since then $R_2(\bar{x}, \bar{z}) \in \ker C \subseteq \ker B$, this means that also $BR_2(\bar{x}, \bar{z}) = 0$, and therefore we can conclude that $AR_1(\bar{x}) = 0$. In summary, *under consistency, the x -component of every steady state of the interconnected system is a steady state of the isolated system*. Moreover, the “retroactivity” signal $BR_2(x, z)$ also vanishes at steady state. This property is satisfied in most interesting interconnections.

Property (C) is equivalent to the requirement that the row space of B be a subspace of the row space of C . Under this property, $\text{rank} \Gamma = \text{rank} A + \text{rank} C$, and therefore Property (*), i.e. $\text{rank} \Gamma = \text{rank}(AB) + \text{rank} C$ is equivalent to Property (**). In other words, for consistent interconnections, the two Theorems provide a dichotomy. Summarizing this discussion and consequences of the two technical lemmas:

Corollary 3.7: Suppose that Property (C) holds. Then, the following statements are equivalent:

- (a) Property (*) holds.
- (b) Property (**) holds.
- (c) Property (2) holds for any two states.

Example 3.1 fails Property (*): the ranks of P and Q are 2 and 3 respectively, but the composite matrix has rank $4 < 5$. Thus this example exhibits retroactivity, by Theorem 1. This example is consistent.

Example 3.2 does not exhibit any retroactivity effects, as is easy to see directly, or appealing to Theorem 2, since Property (**) is satisfied. This example is consistent.

Example 3.3 satisfies Property (*), but nonetheless exhibits a retroactivity effect, in the sense that every state of the isolated system is a steady state, but for the interconnected system $\dot{x} = -x$, $\dot{y} = x - y$ every solution converges to $x = y = 0$. However, Property (**) cannot be used to show retroactivity, since this property also fails. Intuitively, this is a system that has no conservation laws, yet retroactivity fails. However, this system is “inconsistent” in the sense that property (C) does not hold.

REFERENCES

- [1] M. Andrec, B.N. Kholodenko, R.M. Levy, and E.D. Sontag. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *J. Theoret. Biol.*, 232(3):427–441, 2005.
- [2] D. Del Vecchio, A.J. Ninfa, and E.D. Sontag. Modular cell biology: Retroactivity and insulation. *Molecular Systems Biology*, 4:161, 2008.
- [3] B.N. Kholodenko, A. Kiyatkin, F. Bruggeman, E.D. Sontag, H. Westerhoff, and J. Hoek. Untangling the wires: a novel strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences USA*, 99:12841–12846, 2002.
- [4] B.N. Kholodenko and E.D. Sontag. Determination of functional network structure from local parameter dependence data. Technical report, 2002. arXiv physics/0205003, May 2002.
- [5] J. Saez-Rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, and E. D. Gilles. Modular analysis of signal transduction networks. *IEEE Control Systems Magazine*, pages 35–52, 2004.
- [6] S.D.M. Santos, P.J. Verveer, and P.I.H. Bastiaens. Growth factor induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nature Cell Biology*, 9:324–330, 2007.
- [7] E.D. Sontag. Network reconstruction based on steady-state data. *Essays in Biochemistry*, 45:161–176, 2008.
- [8] E.D. Sontag, A. Kiyatkin, and B.N. Kholodenko. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics*, 20(12):1877–1886, 2004.