IDENTIFIABILITY OF LINEAR COMPARTMENT MODELS: THE SINGULAR LOCUS

ELIZABETH GROSS, NICOLETTE MESHKAT, AND ANNE SHIU

Abstract. This work addresses the problem of identifiability, that is, the question of whether parameters can be recovered from data, for linear compartment models. Using standard differential algebra techniques, the question of whether a given model is generically locally identifiable is equivalent to asking whether the Jacobian matrix of a certain coefficient map, arising from input-output equations, is generically full rank. We give a formula for these coefficient maps in terms of acyclic subgraphs of the model's underlying directed graph. As an application, we prove that two families of linear compartment models, cycle and mammillary (star) models with input and output in a single compartment, are identifiable, by determining the defining equation for the locus of non-identifiable parameter values. We also state a conjecture for the corresponding equation for a third family: catenary (path) models. These singular-locus equations, we show, give information on which submodels are identifiable. Finally, we introduce the *identifiability degree*, which is the number of parameter values that match generic input-output data. This degree was previously computed for mammillary and catenary models, and here we determine this degree for cycle models. Our work helps shed light on the question of which linear compartment models are identifiable.

1. Introduction

This work focuses on the identifiability problem for linear compartment models. Linear compartment models are used extensively in biological applications, such as pharmacokinetics, toxicology, cell biology, physiology, and ecology [2, 3, 7, 9, 12]. Indeed, these models are now ubiquitous in pharmacokinetics, with most kinetic parameters for drugs (half-lives, residence times, and so on) based at least in part on linear compartment model theory [13, 18].

A mathematical model is *identifiable* if its parameters can be recovered from data. Using standard differential algebra techniques, the question of whether a given linear compartment model is (generically locally) identifiable is equivalent to asking whether the Jacobian matrix of a certain *coefficient map* (arising from certain *input-output equations*) is generically full rank. Recently, Meshkat, Sullivant, and Eisenberg gave a general formula for the input-output equations of a linear compartment model [11]. Using this formula, it is easy to check whether a given model is (generically locally) identifiable. Nevertheless, we would like to bypass this formula. That is, *can we determine whether a model is identifiable by simply inspecting its underlying directed graph?* Or, as a first step, *if a model is identifiable, when can we conclude that a given submodel is too?* Our interest in submodels, obtained by removing edges, has potential applications: the operation of removing an edge in a model may correspond to a biological intervention, such as a genetic knockout or a drug that inhibits some activity. Identifiable submodels have been studied by Vajda and others [14, 16].

Date: September 28, 2017.

This work begins to answer the questions mentioned above. Our first main result is a formula for the coefficient map in terms of forests (acyclic subgraphs) in the directed graph associated to the model (Theorem 4.5). Our second result gives information on which edges of a model can be deleted while still preserving identifiability (Theorem 3.1). Our remaining results pertain to three well-known families of linear compartment models, which are depicted in Figures 1 and 2: catenary (path graph) models, mammillary (star graph) models, and cycle models [9].

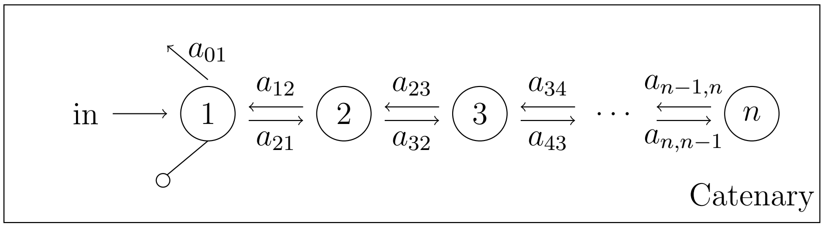


FIGURE 1. The **catenary** (path) model with n compartments, in which compartment 1 has an input, output, and leak. See Section 2.1.

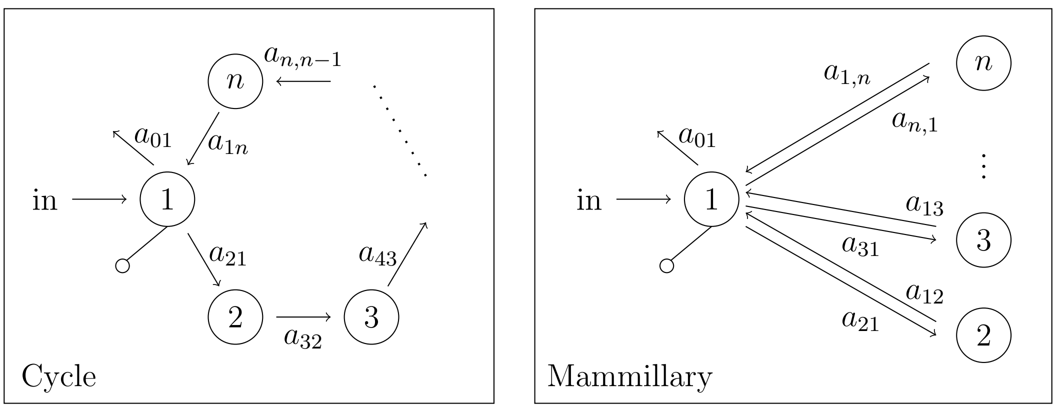


FIGURE 2. Two models with n compartments, where compartment 1 has an input, output, and leak. Left: The **cycle**. Right: The **mammillary** (star).

For these three families of models, which are (generically locally) identifiable [6, 10, 11], we obtain results and a conjecture, summarized in Table 1, on:

(1) a defining equation for the set of non-identifiable parameter values (the *singular locus*), and
(2) the *identifiability degree*: this degree is m if exactly m sets of parameter values match generic input-output data.

We are, to the best of our knowledge, the first to study this singular locus.

The outline of our work is as follows. In Section 2, we introduce linear compartment models and define the singular locus. In Section 3, we prove our result on how the singular locus gives information on identifiable submodels. In Section 4, we give a new combinatorial formula for the coefficients of the input-output equations for linear compartment models with input and output in a single compartment. We use this formula to prove, in Sections 5 and 6, the results

Model	**Equation of singular locus**	**Identifiability degree**
Catenary (path)	Conjecture: $a_{12}^{n-1}(a_{21}a_{23})^{n-2}\dots(a_{n-1,n-2}a_{n-1,n})$	1
Cycle	$a_{32}a_{43}\dots a_{n,n-1}a_{1,n}\prod_{2\leq i<j\leq n}(a_{i+1,i}-a_{j+1,j})$	$(n-1)!$
Mammillary (star)	$a_{12}a_{13}\dots a_{1,n}\prod_{2\leq i<j\leq n}(a_{1i}-a_{1j})^2$	$(n-1)!$

Table 1. Summary of theorems, conjectures, and prior results on the linear compartment models depicted in Figures 1 and 2. See Theorems 5.1, 5.2, and 6.3; Proposition 6.2; and Conjecture 5.3. Note that the $n=2$ versions of these 3 models coincide, so their equations and degrees agree. Also, note that the singular locus for all these models is a union of hyperplanes (e.g., $a_{12}=a_{13}$), including coordinate hyperplanes (e.g., $a_{12}=0$).

on the singular-locus equations and identifiability degrees mentioned above for the models in Figures 1 and 2. We conclude with a discussion in Section 7.

2. Background

In this section, we recall linear compartment models, their input-output equations, and the concept of identifiability. We also introduce the main focus our work: the locus of non-identifiable parameter values and the equation that defines it.

2.1. Linear compartment models.

A *linear compartment model* (alternatively, *linear compartmental model*) consists of a directed graph $G=(V,E)$ together with three sets $In, Out, Leak \subseteq V$. Each vertex $i\in V$ corresponds to a compartment in the model and each edge $j\to i$ corresponds to a direct flow of material from the j-th compartment to the i-th compartment. The sets $In, Out, Leak \subseteq V$ are the sets of input compartments, output compartments, and leak compartments, respectively.

Following the literature, we will indicate output compartments by this symbol:  . Input compartments are labeled by "in", and leaks are indicated by outgoing edges. For instance, each of the linear compartment models depicted in Figures 1 and 2 have $In, Out, Leak = \{1\}$.

To each edge $j\to i$ of G, we associate a parameter a_{ij}, the rate of flow from compartment j to compartment i. To each leak node $i\in Leak$, we associate a parameter a_{0i}, the rate of flow from compartment i leaving the system. Let $n=|V|$. The *compartmental matrix* of a linear compartment model $(G, In, Out, Leak)$ is the $n\times n$ matrix A with entries given by:

$$A_{ij} \ := \ \begin{cases} -a_{0i}-\sum_{k:i\to k\in E} a_{ki} & \text{if } i=j \text{ and } i\in Leak \\ -\sum_{k:i\to k\in E} a_{ki} & \text{if } i=j \text{ and } i\notin Leak \\ a_{ij} & \text{if } j\to i \text{ is an edge of } G \\ 0 & \text{otherwise.} \end{cases}$$

A linear compartment model $(G, In, Out, Leak)$ defines a system of linear ODEs (with inputs $u_i(t)$) and outputs $y_i(t)$ as follows:

$$\begin{aligned} x'(t) &= Ax(t)+u(t) \\ y_i(t) &= x_i(t) \qquad \text{for } i\in Out \ , \end{aligned} \tag{1}$$

where $u_i(t) \equiv 0$ for $i \notin In$.

We now define the concepts of *strongly connected* and *inductively strongly connected*.

Definition 2.1.

(1) A directed graph G is *strongly connected* if there exists a directed path from each vertex to every other vertex. A directed graph G is *inductively strongly connected* with respect to vertex 1 if there is an ordering of the vertices $1, \ldots, |V|$ that starts at vertex 1 such that each of the induced subgraphs $G_{\{1,\ldots,i\}}$ is strongly connected for $i = 1, \ldots, |V|$.

(2) A linear compartment model $(G, In, Out, Leak)$ is *strongly connected* (respectively, *inductively strongly connected*) if G is strongly connected (respectively, inductively strongly connected).

The two most common classes of compartmental models are *mammillary* (star) and *catenary* (path) model structures (see Figures 1 and 2). Mammillary models consist of a central compartment surrounded by and connected with peripheral (noncentral) compartments, none of which are connected to each other [7]. Catenary models have all compartments arranged in a chain, with each connected (in series) only to its nearest neighbors [7]. In a typical pharmacokinetic application, the central compartment of a mammillary model consists of plasma and highly perfused tissues in which a drug distributes rapidly. For catenary models, the drug distributes more slowly. For examples of how mammillary and catenary models are used in practice, see [7, 9, 17].

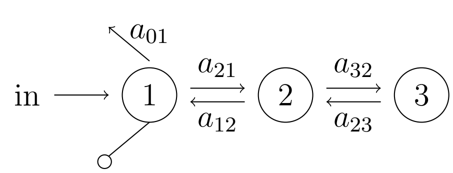
Another common class of compartmental models is formed by *cycle* models (see Figure 2). A cycle model consists of a single directed cycle. Cycle models are only strongly connected, while mammillary and catenary models are inductively strongly connected. For examples of cycle models, see [1, 8, 15].

2.2. **Input-output equations.** The *input-output equations* of a linear compartment model are equations that hold along any solution of the ODEs (1), and which involve only the parameters a_{ij}, input variables u_i, output variables y_i, and their derivatives. The general form of these equations was given by Meshkat, Sullivant, and Eisenberg [11, Corollary 1]. The version of their result we state here is for the case of a single input/output compartment:

Proposition 2.2 (Meshkat, Sullivant, and Eisenberg)**.** *Consider a linear compartment model that is strongly connected, has an input and output in compartment 1 (and no other inputs or outputs), and has at least one leak. Let A denote the compartmental matrix, let ∂ be the differential operator d/dt, and let $(\partial I - A)_{11}$ denote the submatrix of $(\partial I - A)$ obtained by removing row 1 and column 1. Then the input-output equation (of lowest degree) is the following:*

$$\det(\partial I - A) y_1 \; = \; \det\left((\partial I - A)_{11}\right) u_1 \; . \tag{2}$$

Example 2.3. Consider the following catenary model (the $n = 3$ case from Figure 1):



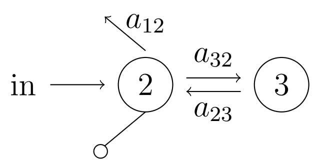
By Proposition 2.2, the input-output equation is:

$$\det \begin{pmatrix} d/dt + a_{01} + a_{21} & -a_{12} & 0 \\ -a_{21} & d/dt + a_{12} + a_{32} & -a_{23} \\ 0 & -a_{32} & d/dt + a_{23} \end{pmatrix} y_1 \ = \ \det \begin{pmatrix} d/dt + a_{12} + a_{32} & -a_{23} \\ -a_{32} & d/dt + a_{23} \end{pmatrix} u_1 \ ,$$

which, when expanded, becomes:

$$y_1^{(3)} + (a_{01} + a_{12} + a_{21} + a_{23} + a_{32})\, y_1^{(2)} + (a_{01}a_{12} + a_{01}a_{23} + a_{01}a_{32} + a_{12}a_{23} + a_{21}a_{23} + a_{21}a_{32})\, y_1' + (a_{01}a_{12}a_{23})\, y_1$$
$$= \ u_1^{(2)} + (a_{12} + a_{23} + a_{32})\, u_1' + (a_{12}a_{23})\, u_1 \ .$$

Observe, from the left-hand side of this equation, that the coefficient of $y_1^{(i)}$ corresponds to the set of forests (acyclic subgraphs) of the model that have $(3-i)$ edges and at most 1 outgoing edge per compartment. As for the right-hand side, the coefficient of $u_1^{(i)}$ corresponds to similar $(n-i-1)$-edge forests in the following model:



This combinatorial interpretation of the coefficients of the input-output equation generalizes, as we will see in Theorem 4.5.

2.3. **Identifiability.** A linear compartment model is *generically structurally identifiable* if from a generic choice of the inputs and initial conditions, the parameters of the model can be recovered from exact measurements of both the inputs and the outputs. We now define this concept precisely.

Definition 2.4. Let $\mathcal{M} = (G, In, Out, Leak)$ be a linear compartment model. The *coefficient map* is the function $c : \mathbb{R}^{|E|+|Leak|} \to \mathbb{R}^k$ that is the vector of all coefficient functions of the input-output equation (here k is the total number of coefficients). Then $\mathcal{M}$ is:

(1) *globally identifiable* if c is one-to-one, and is *generically globally identifiable* if c is one-to-one outside a set of measure zero.
(2) *locally identifiable* if around every point in $\mathbb{R}^{|E|+|Leak|}$ there is an open neighborhood U such that $c : U \to \mathbb{R}^k$ is one-to-one, and is *generically locally identifiable* if, outside a set of measure zero, every point in $\mathbb{R}^{|E|+|Leak|}$ has such an open neighborhood U.
(3) *unidentifiable* if c is infinite-to-one.

Since the coefficients in c are all polynomial functions of the parameters, the model $\mathcal{M} = (G, In, Out, Leak)$ is generically locally identifiable if and only if the image of c has dimension equal to the number of parameters, i.e., $|E| + |Leak|$. The dimension of the image of a map is equal to the rank of the Jacobian matrix at a generic point. Thus we have the following result, which is [11, Proposition 2]:

Proposition 2.5 (Meshkat, Sullivant, and Eisenberg)**.** *A linear compartment model* $(G, In, Out, Leak)$ *is generically locally identifiable if and only if the Jacobian matrix of its coefficient map* c*, when evaluated at a generic point, is equal to* $|E| + |Leak|$*.*

Remark 2.6. As an alternative to Proposition 2.5, one can test identifiability by using a Gröbner basis to solve the system of equations $c(p) = \hat{c}$, for some choice of algebraically independent $c(p)$. The model is then globally identifiable if there is a unique solution for p in terms of $\hat{c}$, locally identifiable if there are a finite number of solutions for p in terms of $\hat{c}$, and unidentifiable if there are an infinite number of solutions. In practice, Gröbner basis computations are more computationally expensive than Jacobian calculations, thus we will test local identifiability using the Jacobian test (Proposition 2.5).

We now examine when the Jacobian is *generically* full rank, but certain parameter choices lead to rank-deficiency. We call parameter values that lead to this rank-deficiency *non-identifiable*. Note that the parameters of these models are *generically* identifiable, and in the identifiability literature are called "identifiable" [7], but for our purposes, we are examining the non-generic case and thus denote the *values* of these parameters "non-identifiable".

Definition 2.7. Let $\mathcal{M} = (G, In, Out, Leak)$ be a linear compartment model that is generically locally identifiable. Let c denote the coefficient map. The *locus of non-identifiable parameter values*, or, for short, the *singular locus* is the subset of the parameter space $\mathbb{R}^{|E|+|Leak|}$ where the Jacobian matrix of c has rank strictly less than $|E| + |Leak|$.

Thus, the singular locus is the defined by the set of all $(|E| + |Leak|) \times (|E| + |Leak|)$ minors of $\mathrm{Jac}(c)$. We will focus on the cases when only a single such minor, which we give a name to below, defines the singular locus:

Definition 2.8. Let $\mathcal{M} = (G, In, Out, Leak)$ be a linear compartment model, with coefficient map $c : \mathbb{R}^{|E|+|Leak|} \to \mathbb{R}^k$. Suppose $\mathcal{M}$ is generically locally identifiable (so, $|E|+|Leak| \leq k$).

(1) If $|E| + |Leak| = k$ (the number of parameters equals the number of coefficients), then $\det(\mathrm{Jac}(c))$ is the *equation of the singular locus*.
(2) Assume $|E|+|Leak| < k$. Suppose there is a choice of $|E|+|Leak|$ coefficients from c, with $r : \mathbb{R}^{|E|+|Leak|} \to \mathbb{R}^{|E|+|Leak|}$ the resulting restricted coefficient map, such that $\det(\mathrm{Jac}(r)) = 0$ if and only if the Jacobian of c has rank strictly less than $|E|+|Leak|$. Then $\det(\mathrm{Jac}(r))$ is the *equation of the singular locus*.

Remark 2.9. The equation of the singular locus, when $|E| + |Leak| = k$, is defined only up to sign, as we do not specify the order of the coefficients in c. When $|E| + |Leak| < k$, there need not be a single $(|E| + |Leak|) \times (|E| + |Leak|)$ minor that defines the singular locus, and thus a singular-locus equation as defined above might not exist. We, however, have not encountered such a model, although we suspect one exists. Accordingly, we ask, *is there always a choice of coefficients or, equivalently, rows of* $Jac(c)$*, such that this square submatrix is rank-deficient if and only if the original matrix* $Jac(c)$ *is?* And, when such a choice exists, *is this choice of coefficients unique?*

Remark 2.10. In applications, we typically are only interested in the factors of the singular-locus equation: we only care whether, e.g., a_{12} divides the equation (i.e., whether a_{12} is non-identifiable) and not which higher powers a_{12}^m, for positive integers m, also divide it.

One aim of our work is to investigate the equation of the singular locus for mammillary, catenary, and cycle models with a single input, output, and leak in the first compartment. As a start, all of these families of models are (at least) generically locally identifiable:

Proposition 2.11. *The n-compartment catenary, cycle, and mammillary models in Figures 1 and 2 (with input, output, and leak in compartment 1 only) are generically locally identifiable.*

Proof. Catenary and mammillary models are inductively strongly connected with $2n-2$ edges. Thus, catenary and mammillary models with a single input and output in the first compartment and leaks from every compartment have coefficient maps with images of maximal dimension [10, Theorem 5.13]. Removing all the leaks except one from the first compartment, we can apply [11, Theorem 1] and obtain generic local identifiability.

Similarly, [10, Proposition 5.4] implies that the image of the coefficient map for cycle models with leaks from every compartment has maximal dimension. Thus removing all leaks except one results in a generically locally identifiable model, again by applying [11, Theorem 1]. □

In fact, catenary models are generically globally identifiable (Proposition 6.2). We also will investigate the identifiability degrees of the other two models, and in particular, prove that the cycle model has identifiability degree $(n-1)!$ (Theorem 6.3).

3. The singular locus and identifiable submodels

One reason a model's singular locus is of interest is because it gives us information regarding the identifiability of particular parameter values. Indeed, for generically locally identifiable models, the singular locus is the set of parameter values that cannot be recovered, even locally. A second reason for studying the singular locus, which is the main focus of this section, is that the singular-locus equation gives information about which submodels are identifiable.

Theorem 3.1 (Identifiable submodels)**.** *Let $\mathcal{M} = (G, In, Out, Leak)$ be a linear compartment model that is strongly connected and generically locally identifiable, with singular-locus equation f. Let $\widetilde{\mathcal{M}}$ be the model obtained from $\mathcal{M}$ by deleting a set of edges $\mathcal{I}$ of G. If $\widetilde{\mathcal{M}}$ is strongly connected, and f is* not *in the ideal $\langle a_{ji} \mid (i,j) \in \mathcal{I}\rangle$ (or, equivalently, after evaluating f at $a_{ji} = 0$ for all $(i,j) \in \mathcal{I}$, the resulting polynomial is nonzero), then $\widetilde{\mathcal{M}}$ is generically locally identifiable.*

Proof. Let $\mathcal{M} = (G, In, Out, Leak)$, the submodel $\widetilde{\mathcal{M}}$, the polynomial $f \in \mathbb{Q}[a_{ij} \mid (i,j) \in E(G),$ or $i = 0$ and $j \in Leak]$, and the ideal $\mathcal{I} \subseteq E(G)$ be as in the statement of the theorem. Thus, the following polynomial $\widetilde{f}$, obtained by evaluating f at $a_{ji} = 0$ for all deleted edges (i,j) in $\mathcal{I}$, is *not* the zero polynomial:

$$\widetilde{f} \ := \ f|_{a_{ji}=0 \text{ for } (i,j)\in\mathcal{I}} \ \in \ \mathbb{Q}[a_{ji} \mid (i,j) \in E(G) \setminus \mathcal{I}, \text{ or } j = 0 \text{ with } i \in Leak] \ .$$

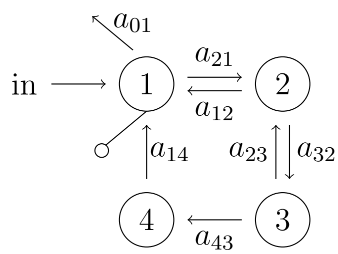
In addition, $f = \det \mathrm{Jac}(r)$, where $r : \mathbb{R}^m \to \mathbb{R}^m$ is a choice of $m := |E(G)| + |Leak|$ coefficients from $\mathcal{M}$'s input-output equation (2) in Proposition 2.2.

Let $\widetilde{m} = m - |\mathcal{I}|$. Let $J := \mathrm{Jac}(r)|_{a_{ji}=0 \text{ for } (i,j)\in\mathcal{I}}$ denote the matrix obtained from $\mathrm{Jac}(r)$ by setting $a_{ji} = 0$ for all $(i,j) \in \mathcal{I}$. The determinant of J is the nonzero polynomial $\widetilde{f}$, so J is full rank when evaluated at any parameter vector (a_{ji}) outside the measure-zero set $V(\widetilde{f}) \subseteq \mathbb{R}^{\widetilde{m}}$. (Here, $V(f)$ denotes the real vanishing set of f.) Thus, the $m \times \widetilde{m}$ matrix B obtained from J by deleting the set of columns corresponding to $\mathcal{I}$, is also full rank ($rank(B) = \widetilde{m}$) outside of $V(\widetilde{f}) \subseteq \mathbb{R}^{\widetilde{m}}$.

Choose $\widetilde{m}$ rows of B that are linearly independent outside some measure-zero set in $\mathbb{R}^{\widetilde{m}}$. (Such a choice exists, because, otherwise, B would be rank-deficient on all of $\mathbb{R}^{\widetilde{m}}$ and thus so would the generically full-rank matrix J, which is a contradiction.) These rows form an $\widetilde{m} \times \widetilde{m}$ matrix that we call $\widetilde{J}$.

Let $\widetilde{r} : \mathbb{R}^{\widetilde{m}} \to \mathbb{R}^{\widetilde{m}}$ be obtained from r by restricting to the coordinates r_i corresponding to the above choice of rows of B, and also setting $a_{ji} = 0$ for all $(i, j) \in \mathcal{I}$. By construction and by Proposition 2.2 (here we use that $\widetilde{\mathcal{M}}$ is strongly connected), $\widetilde{r}$ is a choice of $\widetilde{m}$ coefficients from the input-output equations of $\widetilde{\mathcal{M}}$, and, by construction, the Jacobian matrix of $\widetilde{r}$ is $\widetilde{J}$ (whose rows we chose to be generically full rank). Hence, $\widetilde{\mathcal{M}}$ is generically locally identifiable. □

Example 3.2. Consider the following (strongly connected) linear compartment model $\mathcal{M}$:



This model is generically locally identifiable, and the equation of the singular locus is:

$$a_{12}a_{14}a_{21}^2a_{32}(a_{12}a_{14} - a_{14}^2 - a_{12}a_{23} + a_{14}a_{23} + a_{14}a_{32} - a_{12}a_{43} + a_{14}a_{43} - a_{32}a_{43})(a_{12}a_{23} + a_{12}a_{43} + a_{32}a_{43}) .$$

This equation is *not* divisible by a_{23}, and the model $\widetilde{\mathcal{M}}$ obtained by removing that edge (labeled by a_{23}) is strongly connected. So, by Theorem 3.1, $\widetilde{\mathcal{M}}$ is generically locally identifiable.

The converse of Theorem 3.1 does not hold, as we see in the following example.

Example 3.3 (Counterexample to converse of Theorem 3.1)**.** Consider again the model $\mathcal{M}$ from Example 3.2. The submodel obtained by deleting the edges labeled by a_{12} and a_{23} is generically locally identifiable (by Theorem 5.2: the submodel is the 4-compartment cycle model). Nevertheless, the singular-locus equation of $\mathcal{M}$ is divisible by a_{12} and thus the equation is in the ideal $\langle a_{12}, a_{23} \rangle$.

Example 3.3, our counterexample to the converse of Theorem 3.1, involved deleting two edges ($|\mathcal{I}| = 2$). We do not know of a counterexample that deletes only one edge, and we end this section with the following question.

Question 3.4. In the setting of Theorem 3.1, if a parameter a_{ij} divides f, does it follow that the model $\mathcal{M}'$ obtained by deleting the edge labeled by a_{ij} is unidentifiable (assuming that $\mathcal{M}'$ is strongly connected)?

4. The coefficient map and its Jacobian matrix

Recall that for strongly connected linear compartment models with input and output in one compartment, plus at least one leak, the input-output equation was given in equation (2) (in Proposition 2.2). In this section, we give a new combinatorial formula for the coefficients of the input-output equation (Theorem 4.5).

4.1. **Preliminaries.** To state Theorem 4.5, we must define some graphs associated to a model. In what follows, we use "graph" to mean "directed graph".

Definition 4.1. Consider a linear compartment model $\mathcal{M} = (G, In, Out, Leak)$ with n compartments.

(1) The *leak-augmented graph* of $\mathcal{M}$, denoted by $\widetilde{G}$, is obtained from G by adding a new node, labeled by 0, and adding edges $j \to 0$ labeled by a_{0j}, for every leak $j \in Leak$.
(2) The graph $\widetilde{G}_i$, for some $i = 1, \ldots, n$, is obtained from $\widetilde{G}$ by completing these steps:
   - Delete compartment i, by taking the induced subgraph of $\widetilde{G}$ with vertices $\{0, 1, \ldots, n\} \setminus \{i\}$, and then:
   - For each edge $j \to i$ (with label a_{ij}) in $\widetilde{G}$,if $j \in Leak$ (i.e, $j \to 0$ with label a_{0j} is an edge in $\widetilde{G}$), then label the leak $j \to 0$ in $\widetilde{G}_i$ by $(a_{0j} + a_{ij})$; if, on the other hand, $j \notin Leak$, then add to $\widetilde{G}_i$ the edge $j \to 0$ with label a_{ij}.

Example 4.2. Figure 3 displays a model $\mathcal{M}$, its leak-augmented graph $\widetilde{G}$, and the graphs $\widetilde{G}_1$ and $\widetilde{G}_2$. The compartmental matrix of $\mathcal{M}$ is:

$$A = \begin{pmatrix} -a_{01} - a_{21} & a_{12} \\ a_{21} & -a_{12} \end{pmatrix} .$$

The compartmental matrix that corresponds to $\widetilde{G}_1$ is obtained from A by removing row 1 and column 1. Similarly, for $\widetilde{G}_2$, the corresponding compartmental matrix comes from deleting row 2 and column 2 from A. This observation generalizes (see Lemma 4.3).

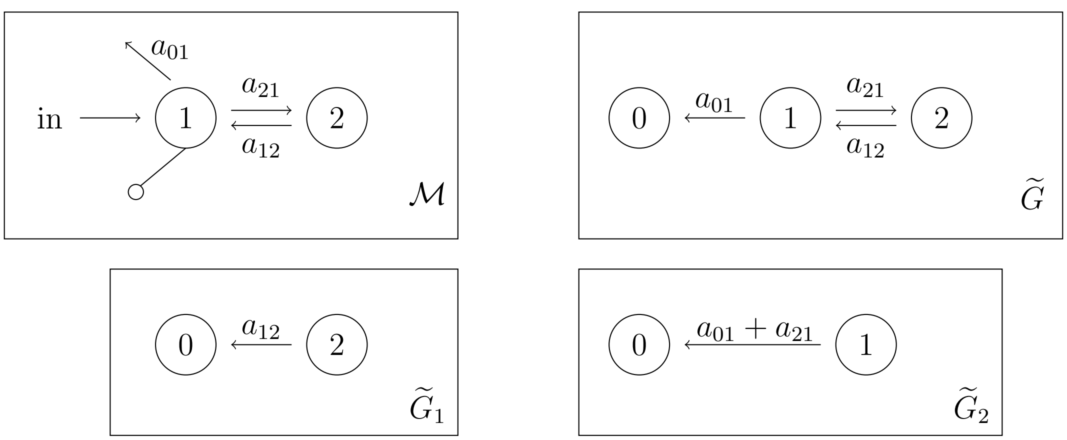


FIGURE 3. A model $\mathcal{M}$, its leak-augmented graph $\widetilde{G}$, and the graphs $\widetilde{G}_1$ and $\widetilde{G}_2$.

Lemma 4.3. *Consider a linear compartment model $\mathcal{M}$ with compartmental matrix A and n compartments. Let $\widetilde{G}_i$, for some $i = 1, \ldots, n$, be as in Definition 4.1. Then for any model $\mathcal{M}'$ whose leak-augmented graph is $\widetilde{G}_i$, the compartmental matrix of $\mathcal{M}'$ is the matrix obtained from A by removing row i and column i.*

Proof. Let $\mathcal{M}$, A, and $\mathcal{M}'$ be as in the statement of the lemma. Let A_{ii} denote the matrix obtained from A by removing row i and column i. We must show that the compartmental matrix of $\mathcal{M}'$ equals A_{ii}. The graph $\widetilde{G}_i$ is obtained by taking the induced subgraph of $\widetilde{G}$

formed by all vertices except i – which ensures that the off-diagonal entries of the compartmental matrix of $\mathcal{M}'$ equal those of A_{ii} – and then replacing edges directed toward i with leak edges (and combining them as necessary with existing leak edges) – which ensures that the diagonal entries of the compartmental matrix also equal those of A_{ii}. Thus, A_{ii} is the compartmental matrix of $\mathcal{M}'$. □

The following terminology matches that of Buslov [4]:

Definition 4.4. Let G be a (directed) graph.

(1) A *spanning subgraph* of G is a subgraph of G with the same set of vertices as G.
(2) An *incoming forest* is a directed graph such that (a) the underlying undirected graph has no cycles and (b) each node has at most one outgoing edge.
(3) For an incoming forest F, let π_F denote the product of the labels of all edges in the forest, that is, $\pi_F = \prod_{(i,j)\in E(F)} a_{ji}$, where a_{ji} labels the edge $i \to j$.
(4) Let $\mathcal{F}_k(G)$ denote the set of all k-edge, spanning, incoming forests of G.

4.2. **A formula for the coefficient map.** Our formula for the coefficient map expresses each coefficient as a sum, over certain spanning forests, of the product of the edge labels in the forest (Theorem 4.5). The formula is an "expanded out" version of a result of Meshkat and Sullivant [10, Theorem 3.2] that showed the coefficient map factors through the cycles in the leak-augmented graph. The difference is due to the fact that Meshkat and Sullivant treated diagonal entries of A as separate variables (e.g., a_{ii}), while our diagonal entries are negative sums of a leak and/or rates (e.g., $-a_{0i} - a_{2i} - a_{3i}$).

Theorem 4.5 (Coefficients of input-output equations)**.** *Consider a linear compartment model $\mathcal{M} = (G, In, Out, Leak)$ that is strongly connected and has an input and output in compartment 1 (and no other inputs or outputs) and at least one leak. Let n denote the number of compartments, and A the compartmental matrix. Write the input-output equation* (2) *as:*

$$y_1^{(n)} + c_{n-1}y_1^{(n-1)} + \cdots + c_1y_1' + c_0y_1 \;=\; u_1^{(n-1)} + d_{n-2}u_1^{(n-2)} + \cdots + d_1u_1' + d_0u_1 \ . \tag{3}$$

Then the coefficients of this input-output equation are as follows:

$$c_i \;=\; \sum_{F\in\mathcal{F}_{n-i}(\widetilde{G})} \pi_F \qquad \textit{for } i = 0, 1, \ldots, n-1 \ , \quad \textit{and}$$

$$d_i \;=\; \sum_{F\in\mathcal{F}_{n-i-1}(\widetilde{G}_1)} \pi_F \qquad \textit{for } i = 0, 1, \ldots, n-2 \ .$$

The proof of Theorem 4.5 requires the following result, which interprets the coefficients of the characteristic polynomial of a compartmental matrix (with no assumptions on the model's connectedness, number of leaks, etc.):

Proposition 4.6. *Let A be the compartmental matrix of a linear compartment model with n compartments and leak-augmented graph $\widetilde{G}$. Write the characteristic polynomial of A as:*

$$\det(\lambda I - A) \;=\; \lambda^n + e_{n-1}\lambda^{n-1} + \cdots + e_0 \ .$$

Then e_i (for $i = 0, 1, \dots, n-1$) is the sum over $(n-i)$-edge, spanning, incoming forests of $\widetilde{G}$, where each summand is the product of the edge labels in the forest:

$$e_i \;=\; \sum_{F \in \mathcal{F}_{n-i}(\widetilde{G})} \pi_F \;.$$

In the Appendix, we prove Proposition 4.6 and explain how it is related to similar results.

Proof of Theorem 4.5. By Proposition 2.2, the coefficient c_i of $y^{(i)}$ in the input-output equation (3) is the coefficient of λ^i in the characteristic polynomial $\det(\lambda I - A)$ of the compartmental matrix A. Hence, the desired result follows immediately from Proposition 4.6.

Now consider the right-hand side of the input-output equation (3). Let A_{11} denote the matrix obtained from A by removing row 1 and column 1. By Lemma 4.3, A_{11} is the compartmental matrix for any model with leak-augmented graph $\widetilde{G}_1$. So, by Proposition 4.6, the sum $\sum_{F \in \mathcal{F}_{n-i-1}(\widetilde{G}_1)} \pi_F$ equals the coefficient of λ^i in the characteristic polynomial $\det(\lambda I - A_{11}) = \det(\lambda I - A)_{11}$ (where the first identity matrix I has size n and the second has size $n-1$). This coefficient, by Proposition 2.2, equals d_i, and this completes the proof. ☐

Remark 4.7 (Jacobian matrix of the coefficient map)**.** In the setting of Theorem 4.5, each coefficient c_k of the input-output equation is the sum of products of edge labels of a forest, and thus is multilinear in the parameters a_{ij}. Therefore, in the row of the Jacobian matrix corresponding to c_k, the entry in the column corresponding to some a_{lm} is obtained from c_k by setting a_{lm}=1 in those terms divisible by a_{lm} and then setting all other terms to 0.

5. The singular locus: mammillary, catenary, and cycle models

In this section, we establish the singular-locus equations for the mammilary (star) and cycle models, which were displayed in Table 1 (Theorems 5.1 and 5.2). We also state our conjecture for the singular-locus equation for the catenary (path) model (Conjecture 5.3). We also pose a related conjecture for models that are formed by bidirectional trees, which include the catenary model (Conjecture 5.6).

5.1. **Mammillary (star) models.**

Theorem 5.1 (Mammillary)**.** *Assume $n \geq 2$. The n-compartment mammillary (star) model in Figure 2 is generically locally identifiable, and the equation of the singular locus is:*

$$(a_{12}a_{13}\dots a_{1,n}) \prod_{2 \leq i < j \leq n} (a_{1i} - a_{1j})^2 \;. \tag{4}$$

Proof. The compartmental matrix for this model is

$$A \;=\; \begin{pmatrix} -a_{01} - (a_{21} + \cdots + a_{n1}) & a_{12} & a_{13} & \dots & a_{n1} \\ a_{21} & -a_{21} & 0 & \dots & 0 \\ a_{31} & 0 & -a_{13} & & 0 \\ \vdots & \vdots & & \ddots & \\ a_{n1} & 0 & 0 & & -a_{1n} \end{pmatrix} .$$

Let $E_j(x_1, \dots, x_m)$ denote the j-th elementary symmetric polynomial on $x_1, x_2, \dots, x_m$; and let $E_j(\hat{a}_{1k})$ denote the j-th elementary symmetric polynomial on $a_{12}, \dots, a_{1,k-1}, a_{1,k+1}, \dots, a_n$.

Then, the coefficients on the left-hand side of the input-output equation (2) are, by Theorem 4.5, the following:

$$\begin{aligned} c_i &= a_{01}E_{n-i-1}(a_{12},\ldots,a_{1n}) + (a_{21}E_{n-i}(\hat{a}_{12}) + a_{31}E_{n-i}(\hat{a}_{13}) + \cdots + a_{n1}E_{n-i}(\hat{a}_{1n})) \\ &\quad + E_{n-i}(a_{12},\ldots,a_{1n}) \end{aligned}$$

for $i = 0, 1, \ldots, n-1$. As for the coefficients of the right-hand side of the input-output equation, they are as follows, by Proposition 2.2:

$$d_i = E_{n-i-1}(a_{12},\ldots,a_{1n}) \qquad \text{for } i = 0, 1, \ldots, n-2 \ .$$

Consider the coefficient map $(c_{n-1}, c_{n-2}, \ldots, c_0,\ d_{n-2}, d_{n-3}, \ldots, d_0)$. Its Jacobian matrix, where the order of variables is $(a_{21}, a_{31}, \ldots, a_{n1},\ a_{01},\ a_{12}, a_{13}, \ldots, a_{1n})$, has the following form:

$$\begin{pmatrix} [\ M\ ] & \star & \star \\ \mathbf{0} & [\ a_{12}a_{13}\cdots a_{1n}\ ] & \star \\ \mathbf{0} & \mathbf{0} & [\ M\ ] \end{pmatrix} ,$$

where M is the following $(n-1)\times(n-1)$ matrix:

$$M = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ E_1(\hat{a}_{12}) & E_1(\hat{a}_{13}) & \ldots & E_1(\hat{a}_{1n}) \\ E_2(\hat{a}_{12}) & E_2(\hat{a}_{13}) & \ldots & E_2(\hat{a}_{1n}) \\ \vdots & \vdots & & \vdots \\ E_{n-2}(\hat{a}_{12}) & E_{n-2}(\hat{a}_{13}) & \ldots & E_{n-2}(\hat{a}_{1n}) \end{pmatrix} .$$

Thus, to prove the desired formula (4), we need only show that the determinant of M equals, up to sign, the *Vandermonde polynomial* on $(a_{12}, \ldots, a_{1n})$:

$$\det M = \pm \prod_{2\le i<j\le n} (a_{1i} - a_{1j}) \ . \tag{5}$$

To see this, note first that both polynomials have the same multidegree: the degree with respect to the a_{1j}'s of $\det M$ is $0 + 1 + \cdots + (n-2)$ (because the entries in row-i of M have degree $i-1$), which equals $\binom{n-1}{2}$, and this is the degree of the Vandermonde polynomial on the right-hand side of equation (5). Also, note that both polynomials are, up to sign, monic.

So, to prove the claimed equality (5), it suffices to show that when $2 \le i < j \le n$, the term $(a_{1i} - a_{1j})$ divides $\det M$. Indeed, when $a_{1i} = a_{1j}$, then the columns of M that correspond to a_{i1} and a_{1j} (namely, the $(i-1)$-st and $(i-1)$-st columns) coincide, and thus $\det M = 0$. Hence, $(a_{1i} - a_{1j}) | \det M$ (by the Nullstellensatz). □

5.2. **Cycle models.**

Theorem 5.2 (Cycle)**.** *Assume $n \ge 3$. The n-compartment cycle model in Figure 2 is generically locally identifiable, and the equation of the singular locus is:*

$$a_{32}a_{43}\ldots a_{n,n-1}a_{1,n} \prod_{2\le i<j\le n} (a_{i+1,i} - a_{j+1,j}) \ .$$

Proof. The compartmental matrix for this model is

$$A = \begin{pmatrix} -a_{01}-a_{21} & 0 & 0 & \dots & 0 & a_{1n} \\ a_{21} & -a_{32} & 0 & \dots & 0 & 0 \\ 0 & a_{32} & -a_{43} & & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & & a_{n-1,n-2} & -a_{n,n-1} & 0 \\ 0 & 0 & \dots & 0 & a_{n,n-1} & -a_{1n} \end{pmatrix} .$$

Let $E_j := E(a_{32}, a_{43}, \dots, a_{1,n})$ denote the j-th elementary symmetric polynomial on $a_{32}, a_{43}, \dots, a_{1,n}$; and let $E_j(\hat{a}_{k+1,k})$ denote the j-th elementary symmetric polynomial on $a_{32}, a_{43}, \dots, a_{k,k-1}, a_{k+2,k+1}, \dots, a_{1,n}$, where $E_j(\hat{a}_{n+1,n}) := E_j(\hat{a}_{1,n})$. Then, the coefficients on the left-hand side of the input-output equation (3) are, by Theorem 4.5, the following:

$$c_0 = a_{01} E_{n-1} , \quad \text{and} \tag{6}$$

$$c_i = (a_{01} + a_{21})\, E_{n-i-1} + E_{n-i-2} \qquad (\text{for } i = 1, 2, \dots, n-1) . \tag{7}$$

As for the coefficients of the right-hand side of the input-output equation, they are as follows, by Proposition 2.2:

$$d_i = E_{n-i-1} \qquad (\text{for } i = 0, 1, \dots, n-2). \tag{8}$$

Consider the coefficient map $(c_0, c_1, \dots, c_{n-1},\ d_0, d_1, \dots, d_{n-2})$. Its Jacobian matrix, where the order of variables is $(a_{01}, a_{21}, a_{32}, \dots, a_{1n})$, has the following form:

$$J = \begin{pmatrix} \begin{bmatrix} E_{n-1} & 0 \\ E_{n-2} & E_{n-2} \\ \vdots & \vdots \\ E_1 & E_1 \\ 1 & 1 \end{bmatrix} & \star \\ \mathbf{0} & \begin{bmatrix} M \end{bmatrix} \end{pmatrix} , \tag{9}$$

where M is the following $(n-1) \times (n-1)$ matrix:

$$M = \begin{pmatrix} E_{n-2}(\hat{a}_{32}) & E_{n-2}(\hat{a}_{43}) & \dots & E_{n-2}(\hat{a}_{1n}) \\ E_{n-3}(\hat{a}_{32}) & E_{n-3}(\hat{a}_{43}) & \dots & E_{n-3}(\hat{a}_{1n}) \\ \vdots & \vdots & & \vdots \\ E_1(\hat{a}_{32}) & E_1(\hat{a}_{43}) & \dots & E_1(\hat{a}_{1n}) \\ 1 & 1 & \dots & 1 \end{pmatrix} .$$

In the upper-left $(n \times 2)$-block of the matrix J in equation (9), rows 2 through $(n-1)$ are scalar multiples of the bottom row of 1's. Thus, if we let $\widetilde{J}$ denote the square matrix (of size $n+1$) obtained by removing from J rows 2 through $(n-1)$, then the singular-locus equation of the model is $\det \widetilde{J}$. Indeed, all nonzero $(n+1) \times (n+1)$ minors of J are scalar multiples of $\det \widetilde{J}$, and thus the singular locus is defined by the single equation $\det \widetilde{J} = 0$.

From equality (5) in the proof of Theorem 5.1, we know $\det M = \prod_{2 \le i < j \le n} (a_{i+1,i} - a_{j+1,j})$. Thus, the equation of the singular locus is

$$\det \widetilde{J} = (E_{n-1})(\det M) = a_{32} a_{43} \dots a_{n,n-1} a_{1,n} \prod_{2 \le i < j \le n} (a_{i+1,i} - a_{j+1,j}) .$$

□

5.3. Catenary (path) models.

Conjecture 5.3. *Assume $n \geq 2$. For the n-compartment catenary (path) model in Figure 1, the equation of the singular locus is:*

$$a_{12}^{n-1}(a_{21}a_{23})^{n-2}(a_{32}a_{34})^{n-3}\dots(a_{n-1,n-2}a_{n-1,n}) \ . \tag{10}$$

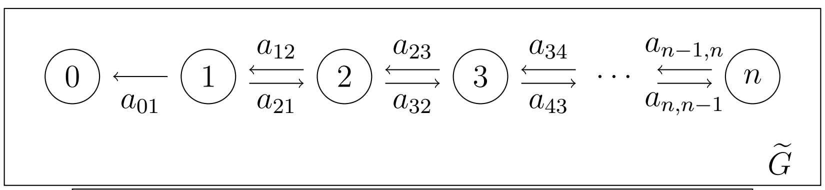
Remark 5.4. The structure of the conjectured equation (10) suggests a proof by induction on n, but we currently do not know how to complete, for such a proof, the inductive step.

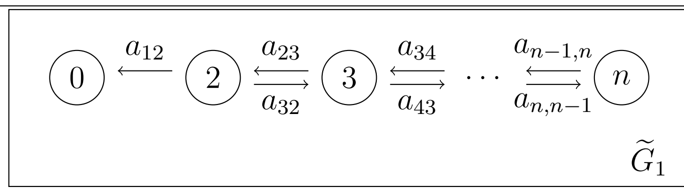
We can prove the following weaker version of Conjecture 5.3:

Proposition 5.5. *For the n-compartment catenary (path) model in Figure 1, the following parameters divide the equation of the singular locus:*

$$a_{21} \qquad \text{and} \qquad a_{12},\ a_{23},\ \dots\ ,\ a_{n-1,n}\ .$$

Proof. By Theorem 4.5, the coefficients of the input-output equation of the catenary model arise from spanning forests of the following graphs:





More specifically, some of the coefficients are as follows:

$$
\begin{aligned}
c_0 &= (a_{01})a_{12}a_{23}\dots a_{n-1,n} && \text{(corresponds to Row 1)}\\
c_1 &= \sum_{F\in\mathcal{F}_{n-1}(\widetilde{G})} \pi_F && \text{(Row 2)}\\
c_{n-1} &= (a_{01}+a_{21})+a_{12}+a_{23}+a_{32}+\dots+a_{n-1,n}+a_{n,n-1} && \text{(Row } n\text{)}\\
d_0 &= a_{12}a_{23}\dots a_{n-1,n} && \text{(Row } n+1\text{)}\\
d_1 &= \sum_{F\in\mathcal{F}_{n-1}(\widetilde{G}_1)} \pi_F = \sum_{F\in\mathcal{F}_{n-1}(\widetilde{G})\ :\ F \text{ does not involve } a_{01} \text{ or } a_{21}} \pi_F && \text{(Row } n+2\text{)}\\
d_{n-2} &= a_{12}+a_{23}+a_{32}+\dots+a_{n-1,n}+a_{n,n-1} && \text{(Row } 2n-1\text{)}
\end{aligned} \tag{11}
$$

For each coefficient in (11), we indicated the corresponding row of the $(2n-1)\times(2n-1)$ Jacobian matrix for the coefficient map $(c_0,\ c_1,\ \dots,\ c_{n-1},\ d_0,\ d_1,\ \dots,\ d_{n-2})$.

Perform the following elementary row operations (which do not affect the determinant) on the Jacobian matrix:

(1) Row 1 := Row 1 - a_{01}Row $(n+1)$
(2) Row 2 := Row 2 - a_{01}Row $(n+2)$ - Row $(n+1)$

(3) Row n := Row n - Row $(2n-1)$.

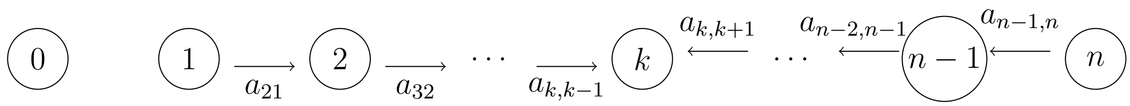
Next, we reorder the columns so that the first four columns are indexed by $a_{01},\ a_{n,n-1},\ a_{21},\ a_{12}$. We claim that in the resulting matrix, the submatrix formed by Rows 1, 2, and n has the following form:

$$\left(\begin{array}{cccc|ccc} a_{12}a_{23}a_{34}...a_{n-1,n} & 0 & 0 & 0 & 0 & \dots & 0 \\ * & a_{21}a_{32}...a_{n-1,n-2} & * & 0 & \bigstar & \dots & \bigstar \\ 1 & 0 & 1 & 0 & 0 & \dots & 0 \end{array}\right) \tag{12}$$

The forms of the first and third rows follow from (11) and the row operations (1) and (3). As for the second row of the matrix (12), consider an entry that is *not* labeled by $*$, i.e., an entry in a column indexed by some a_{ij} with $a_{ij} \neq a_{01}, a_{21}$. This entry, via a straightforward argument using (11) and the row operation (3), is the following sum over the $(n-1)$-edge incoming forests of $\widetilde{G}$ that involve both edges a_{21} and a_{ij}:

$$\sum_{H \sqcup \{a_{21}, a_{ij}\} \in \mathcal{F}_{n-i}(\widetilde{G})} a_{21}\pi_H \ .$$

(Here $\sqcup$ denotes disjoint union.) Each forest in such a sum has the following form, for some $k = 2, 3, \dots, n$:



The only such forest involving the edge $a_{n,n-1}$ is the $k = n$ case, so the $(2,2)$-entry in matrix (12) is indeed $a_{21}a_{32}...a_{n-1,n-2}$. Also, note that a_{21} divides all entries labeled by $\bigstar$.

Next, it is straightforward to row-reduce the matrix (12), without affecting the determinant or the values of the entries labeled by $\bigstar$, to obtain:

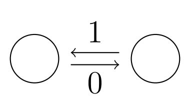
$$\left(\begin{array}{cccc|ccc} a_{12}a_{23}a_{34}...a_{n-1,n} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & a_{21}a_{32}...a_{n-1,n-2} & 0 & 0 & \bigstar & \dots & \bigstar \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 \end{array}\right)$$

Thus, from examining Rows 1 and 2, and the fact that all $\bigstar$-labeled entries are multiples of a_{21}, we conclude that a_{21} and $a_{12}a_{23}a_{34}...a_{n-1,n}$ both divide the determinant of the full Jacobian matrix. This determinant is the singular-locus equation, so we are done. □

5.4. **Tree conjecture.** In this subsection, we generalize Conjecture 5.3, which pertained to catenary (path) models, to "tree models" (Conjecture 5.6). To motivate the new conjecture, we begin by revisiting the 4-compartment catenary and mammillary models. We depict these models in Figure 4, where instead of labeling the edge (i,j) with a_{ji}, we label the edge with the *multiplicity* of a_{ji} in the equation of the singular locus, i.e., the largest p such that a_{ji}^p divides the equation (recall Conjecture 5.3 and Theorem 5.1).

Now consider the model in Figure 5, which also has edges labeled by multiplicities.

Notice that all leaf-edges in Figures 4 and 5 have the following labels:



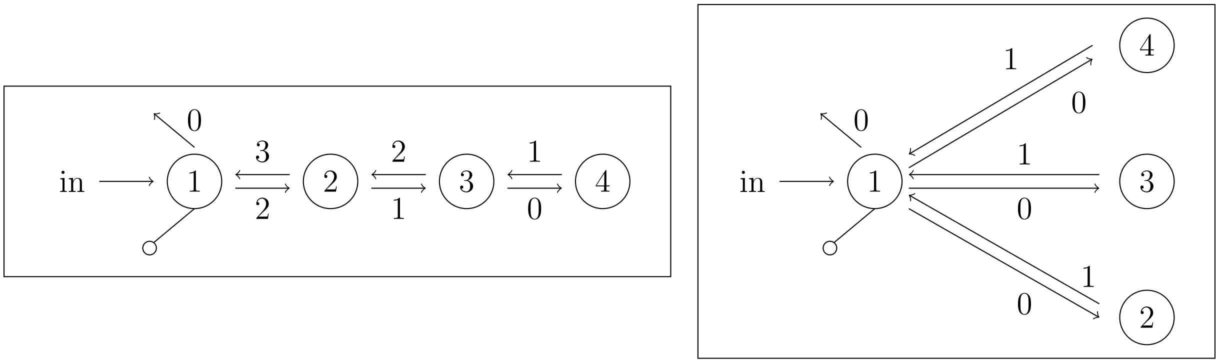


FIGURE 4. A catenary and a mammillary model, with edges (i, j) labeled by the multiplicity of a_{ji} in the corresponding singular-locus equation.

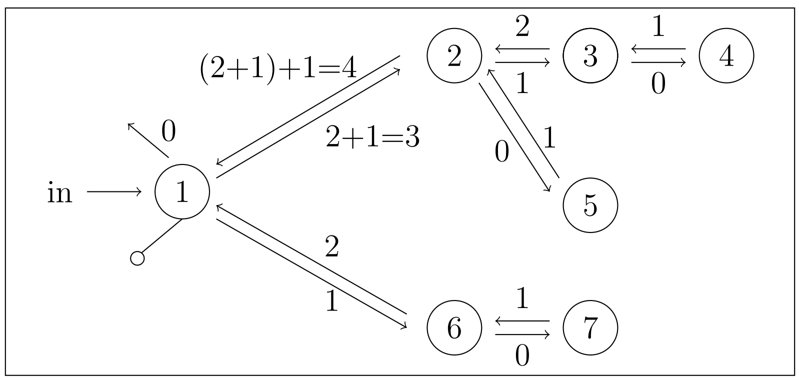


FIGURE 5. A linear compartment model, with edges (i, j) labeled by the multiplicity of a_{ji} in the corresponding singular-locus equation.

Also, as edges move one step closer to compartment 1, the corresponding edge labels increase by 1, except at compartment 2 in Figure 5. Incident to that compartment are the edges (2,1) and (1,2), whose labels are written as sums, $(2+1)+1 = 4$ and $2+1 = 3$, respectively. These observations suggest the following procedure to predict multiplicities:

(CONJECTURED) PROCEDURE TO OBTAIN EXPONENTS IN SINGULAR-LOCUS EQUATION

Input: A linear compartment model $\mathcal{M} = (G, In, Out, Leak)$ with input, output, and leak in compartment 1 only ($In = Out = Leak = \{1\}$), and such that G is a bidirectional tree.

Output: One integer associated to each edge (i, j) of G (which is the purported multiplicity of a_{ji} in the singular-locus equation of $\mathcal{M}$).

Steps:

- *Part 1: outgoing edges* (directed away from compartment 1)
  (1) Label each outgoing leaf-edge with 0.
  (2) As long as there are unlabeled outgoing edges, consider an outgoing edge (i, j) such that all outgoing edges of the form $j \to \star$ have already been labeled. Add 1 to each of these labels, and then compute their sum S. Label edge (i, j) with S.
- *Part 2: incoming edges* (directed toward compartment 1)

(1) Label each incoming leaf-edge with 1.
(2) As long as there are unlabeled incoming edges, consider an incoming edge (j, i) such that all incoming edges of the form $\star \to j$ have already been labeled. Label the edge (j, i) with 1 plus the sum of the labels of all edges incoming to j.

The above procedure and the following conjecture are due to Molly Hoch, Mark Sweeney, and Hwai-Ray Tung (personal communication).

Conjecture 5.6 (Tree conjecture). *The procedure above yields the multiplicities of parameter variables a_{ji} in the equation of the singular locus.*

Hoch, Sweeney, and Tung verified that the conjecture holds for trees on up to 4 nodes.

6. Identifiability degree: mammillary, catenary, and cycle models

In this section, we discuss the identifiability degrees of mammillary, catenary, and cycle models (Proposition 6.2 and Theorem 6.3). The identifiability degree, a term we introduce here, is m if exactly m sets of parameter values match generic input-output data:

Definition 6.1. The *identifiability degree* of a (generically locally identifiable) model is m if the coefficient map is generically m-to-1.

In other words, the identifiability degree is the number of elements in the fiber of the coefficient map over a generic point.

Cobelli, Lepschy, and Romanin Jacur [6] showed that the identifiability degrees of mammillary and catenary models are, respectively, $(n-1)!$ and 1:

Proposition 6.2 (Mammillary and catenary [6]). *Assume $n \geq 2$.*

(1) The identifiability degree of the mammillary (star) model in Figure 2 is $(n-1)!$, where n is the number of compartments.
(2) The identifiability degree of the catenary (path) model in Figure 1 is 1. That is, the model is generically globally identifiable.

Cobelli, Lepschy, and Romanin Jacur also computed the identifiability degrees for the versions of the mammillary and catenary models in which the input/output compartment need not be, respectively, the central compartment or an "end" compartment of the path [6].

Here we prove that the identifiability degree of a cycle model is $(n-1)!$.

Theorem 6.3 (Cycle). *Assume $n \geq 3$. The identifiability degree of the cycle model in Figure 2 is $(n-1)!$, where n is the number of compartments.*

Proof. Let $E_j := E(a_{32}, a_{43}, \ldots, a_{1,n})$ be the j-th elementary symmetric polynomial on the parameters $a_{32}, a_{43}, \ldots, a_{1,n}$.

Recall from the proof of Theorem 5.2, specifically, equations (6)–(8), that the coefficients on the left-hand side of the input-output equation are $c_0 = a_{01}E_{n-1}$ and $c_i = (a_{01} + a_{21})\, E_{n-i-1} + E_{n-i-2}$ (for $i = 1, 2, \ldots, n-1$), and those on the right-hand side are $d_i = E_{n-i-1}$ (for $i = 0, 1, \ldots, n-2$). These coefficients c_i and d_i are invariant under permutations of the $a_{32}, a_{43}, \ldots, a_{1,n}$, so the identifiability degree is at least $(n-1)!$.

We also see that the coefficients are related by the equation $c_0 = a_{01} d_0$ and as follows:

$$\begin{aligned} c_1 &= (a_{01} + a_{21}) d_1 + d_0 \\ c_2 &= (a_{01} + a_{21}) d_2 + d_1 \\ &\vdots \\ c_{n-1} &= (a_{01} + a_{21}) d_{n-1} + d_{n-2} \ . \end{aligned}$$

Thus, $a_{01} = c_0/d_0$ and $a_{21} = (c_{n-1} - d_{n-2})/d_{n-1} - c_0/d_0$, so both a_{01} and a_{21} can be uniquely recovered (when the parameters a_{ij} are generic).

Now consider the remaining coefficients $a_{32}, a_{43}, \ldots, a_{1,n}$. We may assume, by genericity, that these a_{ij}'s are distinct. Having proven that the identifiability degree is *at least* $(n-1)!$, we need only show that the set $\mathcal{A} := \{a_{32}, a_{43}, \ldots, a_{1,n}\}$ can be recovered from the coefficients c_i and d_i (because this would imply that the identifiability degree is *at most* $(n-1)!$). To see this, first recall that the d_i's comprise all the elementary symmetric polynomials, from E_1 to E_n, on $a_{32}, a_{43}, \ldots, a_{1,n}$, and these E_i's are, up to sign, the coefficients of the following (monic) univariate polynomial:

$$\prod_{i=2}^{n} (x - a_{i+1,i}) \ .$$

In turn, a monic polynomial in $\mathbb{R}[x]$ is uniquely determined by its set of roots (in $\mathbb{C}$), so the set $\mathcal{A}$ is uniquely determined by the d_i's. This completes the proof. □

The proof of Theorem 6.3 showed that for the cycle model, under generic conditions, the parameters a_{01} and a_{21} can be uniquely recovered from input-output data, but only the *set* of the remaining parameters $\{a_{32}, a_{43}, \ldots, a_{1,n}\}$ can be identified. Also, this set does *not* reflect some underlying symmetry in the cycle model in Figure 2 (the symmetry of the cycle is broken when one compartment is chosen for the input/output/leak). In contrast, the identifiability degree of the mammillary model, which is $(n-1)!$ (Proposition 6.2), *does* reflect the symmetry of its graph: the $n-1$ non-central compartments can be permuted (see Figure 2).

7. Discussion

In this work, we investigated, for linear compartment models, the set of parameter values that are non-identifiable. Specifically, we focused on the equation that defines this set. We showed first that this equation gives information about which submodels are identifiable, and then computed this equation for certain cycle and mammillary (star) models. These equations revealed that when the parameters are known to be positive, then these parameters can be recovered from input-output data, as long as certain pairs of parameters are not equal. We also stated a conjecture for the singular-locus equation for tree models.

Another topic we examined is the *identifiability degree*, the number of parameter sets that match generic input-output data. We computed this degree for the cycle models, and noted that the degree was already proven for catenary (path) and mammillary models [6].

A natural future problem is to investigate how our results change when the input, output, and/or leak are moved to other compartments. As mentioned earlier, results in this direction were obtained by Cobelli, Lepschy, and Romanin Jacur for catenary and mammillary models [6]. We also are interested in the effect of adding more inputs, outputs, or leaks.

Finally, let us revisit the question that we began with, namely, which models can we conclude are identifiable simply from inspecting the underlying graph? Even when we restrict our attention to models in which a single compartment has an input, output, and leak (and no other compartments has an input, output, or leak), the only such models currently are:

(1) models in which the underlying graph is inductively strongly connected [11, Theorem 1], such as bidirectional trees (e.g., catenary and mammillary models) [6], and
(2) cycle models (Theorem 5.2).

We hope in the future to add to this list by harnessing our new combinatorial interpretation of the input-output coefficients (Theorem 4.5).

Acknowledgements. This project began at a SQuaRE (Structured Quartet Research Ensemble) at AIM, and the authors thank AIM for providing financial support and an excellent working environment. The authors thank Luis García Puente and Heather Harrington for their insights and feedback throughout the course of this project. EG was supported by the NSF (DMS-1620109). NM was partially supported by the Clare Boothe Luce Program from the Luce Foundation. AS was supported by the NSF (DMS-1312473/1513364). Molly Hoch, Mark Sweeney, and Hwai-Ray Tung posed Conjecture 5.6 as a result of research conducted in the NSF-funded REU in the Department of Mathematics at Texas A&M University (DMS-1460766), in which AS served as mentor.

Appendix

Here we prove Proposition 4.6, which, for convenience, we restate here:

Proposition 7.1 (Proposition 4.6)**.** *For a linear compartment model with n compartments, let A be the compartmental matrix. Write the characteristic polynomial of A as:*

$$\det(\lambda I - A) \ = \ \lambda^n + e_{n-1}\lambda^{n-1} + \cdots + e_0 \ .$$

Then e_i (for $i = 0, 1, \ldots, n-1$) is the sum – over $(n-i)$-edge, spanning, incoming forests of the model's leak-augmented graph $\widetilde{G}$ – of the product of the edge labels in the forest:

$$e_i \ = \sum_{F \in \mathcal{F}_{n-i}(\widetilde{G})} \pi_F \ . \tag{13}$$

To prove Proposition 4.6, we need a closely related result, Proposition 7.2 below, which is due to Buslov [4, Theorem 2]. For that result, recall that the *Laplacian matrix* of a graph G with n vertices and edges $i \to j$ labeled by b_{ji} is the $(n \times n)$-matrix L with entries as follows:

$$L_{ij} \ := \ \begin{cases} -b_{ji} & \text{if } i \neq j \\ \sum_{k \neq i} b_{ki} & \text{if } i = j \ . \end{cases}$$

For a model $\mathcal{M} = (G, In, Out, Leak)$ with no leaks ($Leak = \emptyset$), the Laplacian matrix of G is the transpose of the negative of the compartmental matrix of $\mathcal{M}$.

Proposition 7.2 (Buslov)**.** *Let L be the Laplacian matrix of a directed graph G with n edges. Write the characteristic polynomial of L as:*

$$\det(\lambda I - L) \ = \ \lambda^n + e_{n-1}\lambda^{n-1} + \cdots + e_0 \ .$$

Then

$$e_i \;=\; (-1)^{n-i} \sum_{F \in \mathcal{F}_{n-i}(G)} \pi_F \qquad \textit{for } i = 0, 1, \ldots, n-1.$$

In particular, $e_0 = 0$.

Remark 7.3. Buslov's statement of Proposition 7.2 (namely, [4, Theorem 2]) differs slightly from ours: it refers to forests with i connected components rather than $n-i$ edges. That version is equivalent to ours, because a spanning, incoming forest of a graph G with n vertices has i connected components if and only if it has $n-i$ edges.

Remark 7.4. Propositions 4.6 and 7.2 are closely related to the all-minors matrix-tree theorem [5]. The all-minors matrix-tree theorem is a formula for the minors of the Laplacian matrix of a (weighted, directed) graph G, and it is a sum over certain forests in G.

We obtain the following consequence of Proposition 7.2:

Proposition 7.5. *Let L be the Laplacian matrix, or its transpose, of a directed graph G with n edges. Write the characteristic polynomial of $-L$ as:*

$$\det(\lambda I + L) \;=\; \lambda^n + e_{n-1}\lambda^{n-1} + \cdots + e_0 \;.$$

Then

$$e_i \;=\; \sum_{F \in \mathcal{F}_{n-i}(G)} \pi_F \qquad \textit{for } i = 0, 1, \ldots, n-1.$$

Proof. Let $p(\lambda) := \det(\lambda I - L)$. The result follows directly from Proposition 7.2, the equality $\det(\lambda I + L) = (-1)^n \cdot p(-\lambda)$, and (for the case of the transpose) the invariance of the determinant under taking transposes. □

We can now prove Proposition 4.6.

Proof of Proposition 4.6. Let A be the compartmental matrix of a model $\mathcal{M}$ with n compartments and k leaks (so, $1 \leq k \leq n$). The case of no leaks ($k = 0$ and any n) is known, by Proposition 7.5: A is the transpose of the negative of the Laplacian of the graph G of $\mathcal{M}$. Also, the case of $k = 1$ leak and $n = 1$ compartment is straightforward: the compartmental matrix is $A = [a_{01}]$, so $\det(\lambda I - A) = \lambda + a_{01}$, which verifies equation (13).

The above cases form the base cases for proving equation (13) by strong induction on (n, k). Next we prove the inductive step (i.e., the $(n, k+1)$ case, assuming all smaller cases):

Claim: Equation (13) holds for all linear compartment models with n compartments and $k+1$ leaks (where $n \geq 2$ and $1 \leq k \leq n-1$), assuming that the equation holds for all linear compartment models with $(\widetilde{n}, \widetilde{k})$ for which either $\widetilde{n} = n$ and $\widetilde{k} \leq k$, or $\widetilde{n} \leq n-1$ and $\widetilde{k} \leq \widetilde{n}$.

Relabel the compartments so that compartment 1 has a leak (labeled by a_{01}), and, also, there is a (directed) edge from compartment 1 to compartments $2, 3, \ldots, m$ and *not* to compartments $m+1, m+2, \ldots, n$ (for some $1 \leq m \leq n$). This operation permutes the rows and columns of A in the same way, which does not affect the characteristic polynomial $\det(\lambda I - A)$.

We compute as follows:

$$\det(\lambda I - A) = \det \begin{bmatrix} [\ \lambda + (a_{21} + \cdots + a_{m1}) + a_{01}\ ] & \star \\ \star & [\ \lambda I - M\ ] \end{bmatrix}$$

$$= \det \begin{bmatrix} [\ \lambda + (a_{21} + \cdots + a_{m1})\ ] & \star \\ \star & [\ \lambda I - M\ ] \end{bmatrix} + \det \begin{bmatrix} [\ a_{01}\ ] & \star \\ \star & [\ \lambda I - M\ ] \end{bmatrix}$$

$$= \det(\lambda I - N) + a_{01} \det(\lambda I - M)\ , \tag{14}$$

where N is the compartmental matrix for the model obtained from $\mathcal{M}$ by removing the leak from compartment 1 (so, this model has k leaks and n compartments), and (by Lemma 4.3) M is the compartmental matrix for an $(n-1)$-compartment model $\mathcal{M}_1$ for which the leak-augmented graph is $\widetilde{G}_1$.

By the inductive hypothesis, the coefficient of λ^i in the first summand in (14) is:

$$\sum_{F \text{ an } (n-i)\text{-edge, spanning, incoming forest of } \widetilde{G} \text{ } \textit{not} \text{ involving } a_{01}} \pi_F\ . \tag{15}$$

Also by our inductive hypothesis, the coefficient of λ^i in the second summand in (14) is a sum over $(n-i-1)$-edge forests in $\widetilde{G}_1$. Specifically, that coefficient is:

$$a_{01} \sum_{F \in \mathcal{F}_{n-i-1}(\widetilde{G}_1)} \pi_{F'} = \sum_{F \text{ an } (n-i)\text{-edge, spanning, incoming forest of } \widetilde{G} \text{ that involves } a_{01}} \pi_F\ . \tag{16}$$

Taken together, equations (15) and (16) prove the Claim. □

References

[1] S. Audoly and L. D'Angio. On the identifiability of linear compartmental systems: A revisited transfer function approach based on topological properties. *Math. Biosci.*, 66:201–228, 1983.

[2] M Berman and R Schoenfeld. Invariants in experimental data on linear kinetics and the formulation of models. *J. Appl. Phys.*, 27:1361–1370, 1956.

[3] M Berman, E Shahn, and MF Weiss. Some formal approaches to the analysis of kinetic data in terms of linear compartmental systems. *Biophys. J.*, 2:289–316, 1962.

[4] V. A. Buslov. On coefficients of the characteristic polynomial of the Laplace matrix of a weighted digraph and the all minors theorem. *J. Math. Sci.*, 212(6):643–653, 2016.

[5] Seth Chaiken. A combinatorial proof of the all minors matrix tree theorem. *SIAM J. Algebra Discr.*, 3(3):319–329, 1982.

[6] C. Cobelli, A. Lepschy, and G. Romanin Jacur. Identifiability results on some constrained compartmental systems. *Math. Biosci.*, 47(3):173–195, 1979.

[7] Joseph Distefano III. *Dynamic systems biology modeling and simulation.* Academic Press, 2015.

[8] E. Egri, J. Toth, C. Brochot, and F. Y. Bois. Symbolic lumping of some catenary, mamillary and circular compartmental systems. *arXiv*, 0802.2806.

[9] Keith Godfrey. *Compartmental models and their application.* Academic Press Inc., 1983.

[10] Nicolette Meshkat and Seth Sullivant. Identifiable reparametrizations of linear compartment models. *J. Symb. Comput.*, 63:46–67, 2014.

[11] Nicolette Meshkat, Seth Sullivant, and Marisa Eisenberg. Identifiability results for several classes of linear compartment models. *Bull. Math. Biol.*, 77(8):1620–1651, 2015.

[12] RJ Mulholland and MS Keener. Analysis of linear compartment models for ecosystems. *J. Theoret. Biol.*, 44:105–116, 1974.

[13] TN Tozer. Concepts basic to pharmacokinetics. *Pharmacol. Therapeut.*, 12:109–131, 1981.

[14] S. Vajda. Analysis of unique structural identifiability via submodels. *Math. Biosci.*, 71:125–146, 1984.

[15] S. Vajda. Structural equivalence and exhaustive compartmental modeling. *Math. Biosci.*, 69:57–75, 1984.

[16] S. Vajda, J. J. DiStefano, K. R. Godfrey, and J. Fagarasan. Parameter space boundaries for unidentifiable compartmental models. *Math. Biosci.*, 97:27–60, 1989.
[17] Paolo Vicini, Hsiao-Te Su, and Joseph J DiStefano III. Identifiability and interval identifiability of mammillary and catenary compartmental models with some known rate constants. *Math. Biosci.*, 167(2):145–161, 2000.
[18] JG Wagner. History of pharmacokinetics. *Pharmacol. Therapeut.*, 12:537–562, 1981.

San José State University

Santa Clara University

Texas A&M University