

Identifiability of Linear Compartment Models: the impact of leaks and dividing edges

Patrick Chan, Kate Johnston, and Clare Spinner

July 23, 2020

Abstract

A mathematical model is identifiable if some subset of its unknown parameters can be recovered from data. Here, we focus on a particular class of model, linear compartment models, which are used to represent the transfer of substances in a system. We examine the effects of the removal and addition of a leak on model identifiability. We provide results on the structure of the coefficients of the input-output equation of a model with a single leak. This result is used to prove that the addition of a leak to an unidentifiable model results in an unidentifiable model and the removal of a leak from an identifiable model results in an identifiable model. We also examine the singular locus equation, introduced by Gross, Meshkat, and Shiu [2], and the effects of the removal of an edge that divides the singular locus equation on model identifiability. Our conjecture is that the removal of such an edge results in an unidentifiable model, in this paper we introduce classifications to explain this result and some interesting examples to motivate future work.

1 Introduction

A model is *identifiable* if the unknown parameters in the model can be recovered from given input/output data. In this paper, we focus on the problem of determining identifiability for a particular class of models, called *linear compartment models*. Linear compartment models are used in a variety of fields, including in biological applications, to represent the transfer of substances in a system. In the context of pharmacokinetics, the understanding of such models might allow for more precision in the introduction and measurement of drug transfers within the body.

Previous work has focused on characterizing the effects of model operations (adding or removing inputs, outputs, leaks, or edges) on the identifiability of a model. Gross et al. [1] proved that adding inputs or outputs and, under certain hypotheses, adding or removing a leak - preserves identifiability. In this paper, we work to expand upon the effects of adding or removing a leak under more general hypotheses.

In addition, we examine the singular locus equation and the information it provides on the identifiability of strongly connected submodels. More specifically, we examine the role of dividing edges in the singular locus and the identifiability of the model resulting from their removal.

The outline of this paper is as follows. In Section 2, we introduce linear compartment models and identifiability. We provide definitions and introduce notation. In Section 3, we prove our results on the effects of adding and removing a leak. We conclude with a discussion of the results in Section 4.

2 Background

2.1 Linear Compartment Models

A *linear compartment model* (alternatively, *linear compartment model*) is a tuple $(G, In, Out, Leak)$ where $G = (V, E)$ is a directed graph and $In, Out, Leak \subseteq V$. A compartment $i \in V$ indicates a node of the linear

compartment model, which can contain inputs, outputs, and leaks. Input compartments are associated to an input stream, $u_i(t)$, which fuels the system. Similarly, output compartments have an associated output measure, $y(t)$. We assume $Out \neq \emptyset$ as linear compartment models with no outputs are not identifiable.

In keeping with the literature, output compartments are indicated using this symbol: \circ . Input compartments are labeled “in” and leaks are indicated using outgoing edges. For example, the linear compartment model in Figure 1 has $In = \{1\}$, $Out = \{3\}$ and $Leak = \{1, 2\}$.

Each edge $j \rightarrow i \in E$ represents the flow or transfer from compartment j to compartment i . We assign a *flow parameter* k_{ij} to each edge, also called a *rate constant*. Leak compartments also have flow parameters, $k_{0\ell}$, which are the rates of flow exiting the system from compartment ℓ .

Figure 1 depicts a three-compartment catenary model. From a biological standpoint, this model could represent the injection and flow of a drug within the body. The input is the drug amount, and compartment 1 is the injection site. Compartments 2 and 3 represent other organs in the body where the drug travels, such as the heart and lungs, respectively. The parameters between compartments 1, 2, and 3 indicate the transfer of the drug between organs, with the leak parameters, k_{01} and k_{02} , representing the transfer of drugs from the measurable system into immeasurable parts of the body, such as the bloodstream. The output is where the final concentration of the drug is measured.

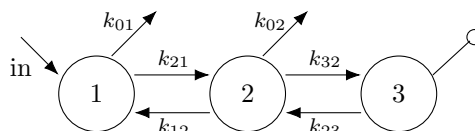


Figure 1: A linear compartment model

We now introduce some more definitions.

Definition 2.1. A directed graph is *strongly connected* if there exists a path from each vertex to every other vertex. A linear compartment model $(G, In, Out, Leak)$ is *strongly connected* if G is strongly connected.

Definition 2.2. The *compartmental matrix* of a linear compartment model $(G, In, Out, Leak)$ with n compartments is the $n \times n$ matrix A given by the following:

$$A_{ij} := \begin{cases} -k_{0i} - \sum_{p:i \rightarrow p \in E} k_{pi} & \text{if } i = j \text{ and } i \in Leak \\ -\sum_{p:i \rightarrow p \in E} k_{pi} & \text{if } i = j \text{ and } i \notin Leak \\ k_{ij} & \text{if } j \rightarrow i \text{ is an edge of } G \\ 0 & \text{otherwise} \end{cases}$$

A linear compartment model $(G, In, Out, Leak)$ defines a system of linear ODEs, with inputs $u_i(t)$ and outputs $y_i(t)$, where $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$ is the vector of concentrations in the compartments at time t :

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

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

Example 2.1. For the model in Figure 1, the ODEs are given by

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} -k_{01} - k_{21} & k_{12} & 0 \\ k_{21} & -k_{02} - k_{12} - k_{32} & k_{23} \\ 0 & k_{32} & -k_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} u_1 \\ 0 \\ 0 \end{pmatrix}$$

with output equation $y_3 = x_3$.

2.2 Input-Output Equations

The *input-output equations* of a linear compartment model are equations that hold along all solutions of the ODEs, and involve only the parameters k_{ij} , input variables u_i , output variables y_i , and their derivatives. The general form of these equations was defined by Meshkat, Sullivant, and Eisenberg [3]. A version of their result stated here is for the case of an input in compartment i and output in compartment j .

Proposition 2.1 (Meshkat, Sullivant, and Eisenberg [3]). *Consider a linear compartment model that is strongly connected, has an input in compartment i and an output in compartment j (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 $(\partial I - A)_{ij}$ be the submatrix of $(\partial I - A)$ obtained by removing row i and column j . Then the input-output equation is as follows:*

$$\det(\partial I - A)y_j = \det((\partial I - A)_{ij})u_i \quad (1)$$

Definition 2.3. From the input-output equations (1), we can derive a *coefficient map* from $\mathbb{R}^{|E|+|Leak|} \rightarrow \mathbb{R}^m$, which sends the vector of parameters (k_{ij}) to the vector of coefficients of the input-output equations. Here, m denotes the number of coefficients.

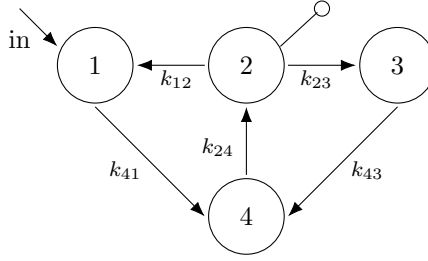


Figure 2: A linear compartment model with $In = \{1\}, Out = \{2\}$

Example 2.2. For the model in Figure 2, the compartmental matrix is:

$$A = \begin{bmatrix} -k_{41} & k_{12} & 0 & 0 \\ 0 & -k_{12} - k_{32} & 0 & k_{24} \\ 0 & k_{32} & -k_{43} & 0 \\ k_{41} & 0 & k_{43} & -k_{24} \end{bmatrix}$$

Therefore the input-output equation is:

$$\det(\partial I - A)y_2 = \det((\partial I - A)_{12})u_1$$

$$\det \left(\begin{bmatrix} \frac{d}{dt} + k_{41} & k_{12} & 0 & 0 \\ 0 & \frac{d}{dt} + k_{12} + k_{32} & 0 & k_{24} \\ 0 & k_{32} & \frac{d}{dt} + k_{43} & 0 \\ k_{41} & 0 & k_{43} & \frac{d}{dt} + k_{24} \end{bmatrix} \right) y_2 = \det \left(\begin{bmatrix} 0 & 0 & -k_{24} \\ 0 & \frac{d}{dt} + k_{43} & 0 \\ -k_{41} & -k_{43} & \frac{d}{dt} + k_{24} \end{bmatrix} \right) u_1$$

which simplifies to the following:

$$y_2^{(4)} + y_2^{(3)}(k_{12} + k_{24} + k_{32} + k_{41} + k_{43}) + y_2^{(2)}(k_{12}k_{24} + k_{12}k_{41} + k_{12}k_{43} + k_{24}k_{32} + k_{24}k_{41} + k_{32}k_{41} + k_{32}k_{43} + k_{41}k_{43}) + y_2^{(1)}(k_{12}k_{24}k_{43} + k_{12}k_{41}k_{43} + k_{24}k_{32}k_{41} + k_{24}k_{41}k_{43} + k_{32}k_{41}k_{43}) = u_1^{(1)}(-k_{24}k_{41}) - k_{24}k_{41}k_{43}$$

Below is the coefficient map derived from the input-output equation:

The coefficient map $c : \mathbb{R}^5 \rightarrow \mathbb{R}^5$ is given by:

$$(k_{12}, k_{24}, k_{41}, k_{23}, k_{43}) \mapsto (k_{12} + k_{24} + k_{32} + k_{41} + k_{43}, k_{12}k_{24} + k_{12}k_{41} + k_{12}k_{43} + k_{24}k_{32} + k_{24}k_{41} + k_{32}k_{41} + k_{32}k_{43} + k_{41}k_{43}, k_{12}k_{24}k_{43} + k_{12}k_{41}k_{43} + k_{24}k_{32}k_{41} + k_{24}k_{41}k_{43} + k_{32}k_{41}k_{43}, -k_{24}k_{41}, k_{24}k_{41}k_{43})$$

2.3 Identifiability

A linear compartment model is *structurally identifiable* if all $k_{i,j}$ parameters can be recovered from the data. More broadly, given a certain linear compartment model, representing some physical system, can we recover all of the flow parameters between compartments.

There are different kinds of identifiability, but in this paper we focus on *generic local identifiability*. The following definition defines the concept more precisely.

Definition 2.4. Let $\mathcal{M} = (G, In, Out, Leak)$ be a linear compartment model that is strongly connected with at least one input. Let $c : \mathbb{R}^{|E|+|Leak|} \rightarrow \mathbb{R}^m$ be the coefficient map derived from the input-output equations. 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 $c : U \rightarrow \mathbb{R}^k$ 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 in U .
3. *unidentifiable* if c is infinite-to-one.

Now, we state a prior result of Meshkat, Sullivant, and Eisenberg [3] which is a criterion for identifiability.

Proposition 2.2 (Meshkat, Sullivant, and Eisenberg [3]). *A linear compartment model $(G, In, Out, Leak)$, with $G = (V, E)$, is generically locally identifiable if and only if the rank of the Jacobian matrix of its coefficient map c , when evaluated at a generic point, is equal to $|E| + |Leak|$.*

Now that we have established a basic understanding of identifiability and its relation to linear compartment models, we introduce some prior results on the effects of removing/adding leaks on the identifiability of a model.

Proposition 2.3 (Theorem 4.3 [1]). *Let M be a linear compartment model that is strongly connected and has at least one input and no leaks. Let M' be a model obtained from M by adding one leak. If M is generically locally identifiable from the coefficient map, then so is M' .*

Proposition 2.4 (Proposition 4.6 [1]). *Let M be a linear compartment model that is strongly connected and has an input, output, and leak in a single compartment (and has no other inputs, outputs, or leaks). If M is generically locally identifiable from the coefficient map, then so is the model obtained from M by removing the leak.*

2.4 Singular Locus

Definition 2.5. Let $\mathcal{M} = (G, In, Out, Leak)$ be a generically locally identifiable linear compartment model. Let c be the coefficient map derived from the input-output equation. 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 $|E| + |Leak|$.

Therefore, for identifiable linear compartment models, the *singular locus* is defined by the set of all $(|E| + |Leak|) \times (|E| + |Leak|)$ minors of the Jacobian matrix of the coefficient map. In certain cases, the model results in a square Jacobian matrix where there is then only a single minor.

A model's singular locus is of interest because it provides information regarding the identifiability of particular parameter values. Previously Gross, Meshkat, and Shiu [2] have investigated models with a single singular locus equation and the edges that divide the singular locus, which we call *dividing edges* and showed that removing edges that are not dividing edges from an identifiable model maintains identifiability.

Proposition 2.5 (Theorem 3.1 [2]). *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 $\langle a_{ji} \mid (i, j) \in \mathcal{I} \rangle$, then $\widetilde{\mathcal{M}}$ is generically locally identifiable.*

Their work motivated the following question that we address in this paper.

Question 2.1 ([2]). *If a parameter k_{ij} divides the singular locus equation, does it follow that the model \mathcal{M}' obtained by deleting the edge k_{ij} is unidentifiable?*

In addressing this question, we make use of the following conjecture on the number of coefficients of a linear compartment model with n compartments.

Conjecture 2.1. *Consider a strongly connected linear compartment model $\mathcal{M} = (G, In, Out, Leak)$ with $|In| = |Out| = 1$. Let n be the number of compartments, and let L be the length of the shortest (directed) path in G from the input compartment to the output compartment. Then, in the input-output equation 1, the number of non-monic, nonzero coefficients on the left- and right-hand sides are as follows:*

$$\# \text{ coefficients on LHS} = \begin{cases} n & \text{if Leak} \neq \emptyset \\ n-1 & \text{if Leak} = \emptyset \end{cases} \quad \# \text{ coefficients on RHS} = \begin{cases} n-1 & \text{if In} = \text{Out} \\ n-L & \text{if In} \neq \text{Out} \end{cases}$$

All models we have observed thus far have been consistent with Conjecture 2.1. The paper (in preparation) "Identifiability of linear compartment models: The effect of moving inputs and outputs" by Cashous Bortner, Elizabeth Gross, Nicolette Meshkeat, Anne Shiu and Seth Sullivant will resolve at least some cases of the conjecture.

3 Main Results

In this section, we will outline our results on the effects of identifiability when adding or removing a leak. More specifically, we address the following questions: If a model is unidentifiable with no leaks and a single leak is added, is the resulting model unidentifiable? If a model containing a single leak is identifiable and the leak is removed, is the resulting model identifiable? In addition, we address the question of removing edges that divide the singular locus and the effect on the identifiability of the model (Question 2.1).

3.1 Preliminary Results

First we introduce some preliminary results regarding the structure of the coefficients of the input-output equation which will be used in the proofs of our main results.

Lemma 3.1. *Let \mathcal{M} be a strongly connected linear compartment model with $|In| = |Out| = 1$. Denote the coefficients of the input-output equation by c_i . Each coefficient of the model $\tilde{\mathcal{M}}$ obtained by adding a single leak at compartment ℓ has the form: $\tilde{c}_i = c_i + k_{0\ell}(g_i)$, where $k_{0\ell}$ is the leak parameter and g_i is a polynomial in the parameters k_{ij} 's where $i \geq 1$.*

Proof. The left-hand coefficients (c_{y_i}) of the input-output equations of the models \mathcal{M} and $\tilde{\mathcal{M}}$ are the result of the expansion of $\det(\partial I - A)$ and $\det(\partial I - \tilde{A})$, respectively. It is clear from the structure of the compartmental matrix A that $A = \tilde{A} \Big|_{k_{0\ell}=0}$. In addition, $\det(\partial I - \tilde{A}) \Big|_{k_{0\ell}=0} = \det(\partial I - \tilde{A} \Big|_{k_{0\ell}=0})$. It is straightforward to verify: $\det(\partial I - \tilde{A}) \Big|_{k_{0\ell}=0} = \det(\partial I - A)$. So, comparing coefficients of $y_3^{(p)}$ in the resulting determinants, we note the following relationship: $\tilde{c}_{y_i} \Big|_{k_{0\ell}=0} = c_{y_i}$, for all i and therefore, $\tilde{c}_i = c_i + k_{0\ell}(g_i)$ where g_i is a polynomial in the parameters k_{ij} 's where $i \geq 1$.

The right-hand coefficients (c_{u_i}) of the input-output equations of the models \mathcal{M} and $\tilde{\mathcal{M}}$ are the result of the expansion of $\det((\partial I - A)_{in/out})$ and $\det((\partial I - \tilde{A})_{in/out})$, respectively. The proof follows that above to show $\tilde{c}_{u_i} = c_{u_i} + k_{0\ell}(x)$, where x is some polynomial combination of k_{ij} with $i \geq 1$. \square

Lemma 3.2. *Let \mathcal{M} be a strongly connected linear compartment model with $|In| = |Out| = 1$, $Leak = \emptyset$ and r coefficients in the input-output equation. The model $\tilde{\mathcal{M}}$ obtained by adding a single leak $k_{0\ell}$ has exactly $r + 1$ coefficients.*

Proof. It is clear, by Lemma 3.1, that there are at least as many coefficients in the model $\widetilde{\mathcal{M}}$ as in \mathcal{M} . We show that there is exactly one additional coefficient by showing that the number of y -coefficients (c_{y_i}) on the left-hand side of the input-output equation increases by 1 and the number of u -coefficients (c_{u_i}) on the right-hand side of the input-output equation does not change.

First, we note that the powers of $\frac{d}{dt}$ in the expansion of $\det(\partial I - A)$ and $\det((\partial I - A)_{in/out})$ are not affected by the addition of a leak. Any new $\frac{d}{dt}$ terms created by the addition of the leak will include the leak term and will therefore not cancel out any terms that existed in the model with no leak. Therefore, the only coefficient that is affected by the leak is the constant coefficient which corresponds to $\det(A)$.

The structure of A ensures that given a strongly connected model with no leaks the sum of the entries of each column is 0. Therefore, the rows are linearly dependent and thus, $\det(A) = 0$. So \mathcal{M} does contain the constant term y -coefficient. However, in the model $\widetilde{\mathcal{M}}$ the rows of the matrix \widetilde{A} are linearly independent. We prove this by supposing there is a linear combination of rows $b_1 \cdot \vec{a}_1 + b_2 \cdot \vec{a}_2 + \dots + b_n \cdot \vec{a}_n = \vec{0}$, where $b_i \in \mathbb{R}$ and \vec{a}_i is the vector of the entries of the i th row of \widetilde{A} and showing the only solution is the trivial one.

We know \vec{a}_ℓ contains $k_{0\ell}$ and no other vector does, therefore $b_\ell = 0$. Because the model is strongly connected, there exists some edge $k_{\ell j}$. Each edge $k_{i j}$ where $i \geq 1$ appears in \widetilde{A} two times. Therefore, with $b_\ell = 0$, there is now only one instance of $k_{\ell j}$ which appears in \vec{a}_j . So, $b_j = 0$. We continue by induction to show that $b_i = 0$ for all $1 \leq i \leq n$. Thus, the only solution is the trivial one and the rows of \widetilde{A} are linearly independent. Therefore, $\det(\widetilde{A}) \neq 0$. This means that the constant y -coefficient exists in the model $\widetilde{\mathcal{M}}$ and the number of y -coefficients has increased by 1.

Now we show that the constant u -coefficient existed in \mathcal{M} and therefore the number of u -coefficients does not change following the addition of the leak. The u -coefficients appear in the expansion of $\det((\partial I - A)_{in/out})$. Again we consider the linear combination of rows such that the sum is 0 and show that the only solution is the trivial one. Suppose, $b_1 \cdot \vec{a}_1 + b_2 \cdot \vec{a}_2 + \dots + b_{n-1} \cdot \vec{a}_{n-1} = \vec{0}$ where $b_i \in \mathbb{R}$ and \vec{a}_i contains the entries of the i th row of $A_{in/out}$. We follow the same process as with the y -coefficients. Instead of starting with the leak, we consider the edge $k_j in$. The removal of the row in means that this edge appears only once in $A_{in/out}$ in row j . So, $b_j = 0$. Since the model is strongly connected, there is an edge $k_{s j}$. When $b_j = 0$, there is only one occurrence of the edge $k_{s j}$ which appears in \vec{a}_s , thus $b_s = 0$. We continue by induction to conclude that $b_i = 0$ for all $1 \leq i \leq n - 1$. So the rows are linearly independent and $\det(\widetilde{A}) \neq 0$. Therefore the model \mathcal{M} contains the constant coefficient. This constant coefficient remains in the model $\widetilde{\mathcal{M}}$, as any new constant terms will contain the leak edge and will not cancel with terms from the model with no leak. Therefore the number of u -coefficients is unchanged.

So, we have showed that, following the addition of a single leak, the number of y -coefficients increases by 1 and the number of u -coefficients is unchanged. Therefore the total number of coefficients increases by 1 to $r + 1$.

To conclude the proof, we state explicitly what is implied - that the addition leaks beyond the first leak does not change the number of coefficients in the model. As stated previously, a leak affects only the constant term coefficient in the expansion of the determinant. It is clear that there can be at most one constant term coefficient. Furthermore, the addition of a leak will not cancel out the constant term coefficient since, as stated previously, additional terms that appear in the model with the leak will contain the leak term and will therefore not cancel any terms that existed in the model without that leak. So, for a model with no leaks and r coefficients, the addition of any number of leaks results in a model with exactly $r + 1$ coefficients. \square

In addition to the previous results, our observations of several classes of models lead us to believe the following conjecture is true regarding the edges that divide the singular locus equation.

Conjecture 3.1. *Let \mathcal{M} be a strongly connected linear compartment model such that $|In| = |Out| = 1$ and $Leak \neq \emptyset$. Then, $k_{0\ell} \nmid \det(J_{\mathcal{M}})$ for all $\ell \in Leak$.*

3.2 Adding a Leak

In this section we address the following question:

Question 3.1 (Question 5.2 [1]). *Let \mathcal{M} be an unidentifiable linear compartment model, if one leak is added, is the resulting model $\widetilde{\mathcal{M}}$ always unidentifiable?*

We address this question by considering the two subcases that arise naturally through Proposition 2.2 and classify an unidentifiable model by the way in which the Jacobian matrix of the coefficient map is not full rank. In the first subcase, a model is unidentifiable because the Jacobian matrix has a determinant of 0. In the second subcase, a model is unidentifiable because the Jacobian matrix has more columns than rows (model has more parameters in the model than coefficients of the input-output equation). The following Theorem 3.1 addresses the second subcase and Conjecture 3.2 concerns the first.

Theorem 3.1. *Let \mathcal{M} be a linear compartment model with $|In| = |Out| = 1$. Consider the model $\widetilde{\mathcal{M}}$ formed by adding a single leak to \mathcal{M} . If \mathcal{M} has more parameters in the model than coefficients in the input-output equation, then $\widetilde{\mathcal{M}}$ also has more parameters in the model than coefficients in the input-output equation.*

We give two proofs, one using the Conjecture 2.1 and the other using Lemma 3.2.

Proof using Conjecture 2.1:

Proof. Consider a strongly connected linear compartment model \mathcal{M} with n compartments. Let K be the number of parameters of \mathcal{M} and C be the total number of coefficients of the input-output equation and assume that $K - C > 0$.

We start with the case where $Leak = \emptyset$ in \mathcal{M} . Then, by Conjecture 2.1, the number of y-coefficients on the left-hand side of the input-output equation is $n - 1$. Let x be the number of u-coefficients on the right-hand side of the equation. Thus, $C = n - 1 + x$ so $K - n + 1 - x > 0$.

Now consider the model $\widetilde{\mathcal{M}}$ obtained by adding a single leak. We show that $\widetilde{K} - \widetilde{C} > 0$. With the additional leak, the number of parameters is $\widetilde{K} = K + 1$. By Conjecture 2.1, the number of y-coefficients is n . The number of u-coefficients remains x . Thus, $\widetilde{C} = n + x$. Therefore, $\widetilde{K} - \widetilde{C} = K + 1 - n - x > 0$. Thus, there are more parameters than coefficients in the model $\widetilde{\mathcal{M}}$ and, by Proposition 2.2, $\widetilde{\mathcal{M}}$ is unidentifiable.

For the case where $Leak \neq \emptyset$ in \mathcal{M} , by Conjecture 2.1, the number of y-coefficients is n . Let x be the number of u-coefficients. Then, $C = n + x$ so $K - n - x > 0$.

Consider the model $\widetilde{\mathcal{M}}$ obtained by adding a single leak. The addition of this leak increases the number of parameters to $\widetilde{K} = K + 1$. The number of y-coefficients and number of u-coefficients do not change, so $\widetilde{C} = n + x$. So, $\widetilde{K} - \widetilde{C} = K + 1 - n - x > 0$. Therefore, by Proposition 2.2, $\widetilde{\mathcal{M}}$ is unidentifiable. \square

Proof using the Lemma 3.2:

Proof. Let \mathcal{M} be a strongly connected linear compartment model with $|In| = |Out| = 1$. Assume the number of parameters in the model K is greater than the number of coefficients in the input-output equation C . So $K - C > 0$. We consider two cases : the case when $Leak = \emptyset$ and the case when $Leak \neq \emptyset$. Let $\widetilde{\mathcal{M}}$ be the model obtained by adding a single leak.

Suppose $Leak = \emptyset$ in \mathcal{M} . The addition of a single leak increases the number of parameters to $\widetilde{K} = K + 1$. By Lemma 3.2, following the addition of a single leak, the number of u-coefficients on the right-hand side of the input-output equation does not change and the number of y-coefficients on the left-hand side of the equation increases by exactly 1. So the total number of coefficients increases to $\widetilde{C} = C + 1$. Then, $\widetilde{K} - \widetilde{C} = (K + 1) - (C + 1) = K - C > 0$. Therefore $\widetilde{K} > \widetilde{C}$ and by Proposition 2.2 $\widetilde{\mathcal{M}}$ is unidentifiable.

Suppose $Leak \neq \emptyset$ in \mathcal{M} . The addition of a single leak increases the number of parameters to $\widetilde{K} = K + 1$. By Lemma 3.2, the total number of coefficients does not change. So $\widetilde{C} = C$.

Therefore, $\widetilde{K} - \widetilde{C} = K + 1 - C > 0$. Therefore $\widetilde{K} > \widetilde{C}$, and by Proposition 2.2, $\widetilde{\mathcal{M}}$ is unidentifiable. \square

In addition to this result, our observations have led us to the following conjecture concerning the subcase of Question 3.1 where the determinant of the Jacobian matrix is 0.

Conjecture 3.2. *Let \mathcal{M} be a linear compartment model with $|In| = |Out| = 1$ and $Leak = \emptyset$. Consider the model $\widetilde{\mathcal{M}}$ formed by adding some number of leaks to \mathcal{M} . If the determinant of the Jacobian matrix of the coefficient map of \mathcal{M} is 0, or, in the case of a non-square Jacobian matrix, all minors of the Jacobian are 0, then the model \mathcal{M} is also unidentifiable.*

We end this section with a discussion of the addition of leaks to models that are identifiable. The following result follows directly from Lemma 3.2.

Theorem 3.2. *Let \mathcal{M} be an identifiable strongly connected linear compartment model with K parameters and C coefficients such that $C - K \geq 0$ and $Leak = \emptyset$. If $n > C - K + 1$ leaks are added, the resulting model is unidentifiable.*

Proof. The proof follows easily from Lemma 3.2. By Lemma 3.2, the addition of any number of leaks increases the number of coefficients in the input-output equation by exactly 1. The number of parameters increase by 1 for each additional leak. So, suppose $C - K + 2$ leaks are added to a model. Then the number of coefficients increases to $C' = C + 1$ and the number of parameters increases to $K' = K + (C - K + 2) = C + 2$. Therefore, there are more parameters than coefficients and, by Proposition 3.2, the model is unidentifiable. \square

Remark 3.1. In the case that less than $n \leq C - K + 1$ leaks are added to the model. The model may or may not be identifiable. If the model is not identifiable, it has a Jacobian matrix with a determinant of zero. A fuller understanding of the cases when the determinant of the Jacobian is zero might clarify this question.

3.3 Removing a Leak

In this section we address the following conjecture, originally posed by Gross, Harrington, Meshkat and Shiu [1, Conjecture 4.5].

Conjecture 3.3 (Removing a leak). *Let $\widetilde{\mathcal{M}}$ be a linear compartment model that is strongly connected and has at least one input and exactly one leak. If $\widetilde{\mathcal{M}}$ is generically locally identifiable, then so is the model \mathcal{M} obtained from $\widetilde{\mathcal{M}}$ by removing the leak.*

Previously Gross et al. [1] proved the conjecture holds for the case where $In = Out = Leak$. (Proposition 2.4). We have expanded upon this result to prove the conjecture whenever Conjecture 3.1 holds.

Theorem 3.3. *Conjecture 3.3 is equivalent to Conjecture 3.1.*

Proof. Lemmas 3.1, 3.2, and 3.3, provide us with a very specific structure of the coefficients c_i and \tilde{c}_i of the models \mathcal{M} and $\widetilde{\mathcal{M}}$, respectively. Using this structure, we can compare the structure of the Jacobian matrices of the coefficient map for each model to show that $\det(J_{\widetilde{\mathcal{M}}}) \neq 0 \Rightarrow \det(J_{\mathcal{M}}) \neq 0$ and therefore \mathcal{M} is identifiable.

It is clear, $\widetilde{\mathcal{M}}$ has one additional parameter, $k_{0\ell}$. By Lemma 3.3, if $\widetilde{\mathcal{M}}$ has r y-coefficients on the left-hand side of the input-output equation, then \mathcal{M} has $(r - 1)$ y-coefficients. By Lemma 3.1, the additional coefficient has the form: $\tilde{c}_{y_r} = k_{0\ell}(g_\ell)$ where g_ℓ is some polynomial combination of edges k_{ij} with $i \geq 1$. Thus, the row of $J_{\widetilde{\mathcal{M}}}$ corresponding to \tilde{c}_{y_r} has entries of $\frac{\partial c_{y_r}}{\partial k_{ij}} = 0$ if k_{ij} is not in g_ℓ , and entries of $k_{0\ell}(g_i)$, where g_i is some polynomial combination of edges, if k_{ij} is in g_ℓ . The entry $\frac{\partial c_{y_r}}{\partial k_{0\ell}} = g_\ell \neq 0$.

Consider the submatrix, B , consisting of the rows, $\tilde{c}_{y_1}, \dots, \tilde{c}_{y_{r-1}}, \tilde{c}_{u_1}, \dots, \tilde{c}_{u_m}$ and all columns except the column corresponding to $k_{0\ell}$. By Lemma 3.1, the entries of this matrix have the form: $\frac{\partial \tilde{c}_p}{\partial k_{ij}} = \frac{\partial c_p}{\partial k_{ij}} + k_{0\ell}(b_{p,i,j})$ where $b_{p,i,j}$ is a polynomial in the parameters. Thus, $B|_{k_{0\ell}=0} = J_{\mathcal{M}}$.

Now, consider the matrix, $J_{\widetilde{\mathcal{M}}}|_{k_{0\ell}=0}$. Take the determinant by expanding along the row corresponding to c_{y_r} , which now has all entries 0 except the entry g_ℓ . The calculation is as follows:

$$\begin{aligned} \det(J_{\widetilde{\mathcal{M}}})|_{k_{0\ell}=0} &= 0 + 0 + \dots + 0 \pm g_\ell \cdot (\det(B|_{k_{0\ell}=0})) \\ &= \pm a \cdot (\det(J_{\mathcal{M}})) \end{aligned}$$

If Conjecture 3.1 holds for $\widetilde{\mathcal{M}}$, then $k_{0\ell}$ does not divide $J_{\widetilde{\mathcal{M}}}$ and therefore $\det(J_{\widetilde{\mathcal{M}}})|_{k_{0\ell}=0} \neq 0$. Then, since $g_\ell \neq 0$, we can conclude that $\det(J_{\mathcal{M}}) \neq 0$ and therefore, \mathcal{M} is identifiable. So we can conclude that Conjecture 3.3 is true if and only if Conjecture 3.1 is true.

Below is a diagram illustrating the structure of $J_{\widetilde{\mathcal{M}}}$ and $J_{\mathcal{M}}$ described above.

$$\begin{array}{c}
 J_{\widetilde{\mathcal{M}}} = \left[\begin{array}{c} \tilde{c}_{y_1} \\ \vdots \\ \tilde{c}_{y_{r-1}} \\ \tilde{c}_{y_r} \\ \tilde{c}_{u_1} \\ \vdots \\ \tilde{c}_{u_m} \end{array} \right] \begin{array}{c} \boxed{\frac{\partial c_{y_i}}{\partial k_{ij}} + k_{0\ell}(b)} \\ \text{either } 0 \text{ or } k_{0\ell}(\dots) \\ \boxed{\frac{\partial c_{u_i}}{\partial k_{ij}} + k_{0\ell}(b)} \end{array} \begin{array}{c} B \\ x \\ k_{0\ell} \end{array}
 \end{array}
 \qquad
 \begin{array}{c}
 J_{\mathcal{M}} = \left[\begin{array}{c} c_{y_1} \\ \vdots \\ c_{y_{r-1}} \\ c_{u_1} \\ \vdots \\ c_{u_m} \end{array} \right] \begin{array}{c} \boxed{\frac{\partial c_{y_i}}{\partial k_{ij}}} \\ \boxed{\frac{\partial c_{u_i}}{\partial k_{ij}}} \end{array}
 \end{array}
 \quad \square$$

Remark 3.2. The proof above handles the case of a model with a square Jacobian matrix of the coefficient map. If an identifiable model has more coefficients in the input-output equation than parameters in the model, then the Jacobian matrix is non-square. The model being identifiable means that at least one minor of the Jacobian matrix is non-zero. If such a minor corresponds to a submatrix that contains the row corresponding to the y -coefficient with the form $k_{0\ell}(g_\ell)$, where $k_{0\ell}$ is the leak and g_ℓ is a polynomial in the parameters k_{ij} , then the proof above applies. However, if the model is such that the non-zero minors are associated to submatrices that do not contain the row corresponding to the y -coefficient $k_{0\ell}(g_\ell)$, then the proof above does not apply. We have not observed such a case ever occurring, however a full proof to address scenario this has not yet been attempted.

Remark 3.3. Theorem 3.2 describes the the equivalence of Conjecture 3.1 and Conjecture 3.3. Proposition 2.4 [2] resolves Conjecture 3.3 for models where $|In| = |Out| = |Leak| = 1$ and $Leak \in In \cup Out$. Therefore, Conjecture 3.1 holds for this class of models.

3.4 The Singular Locus and Dividing Edges

In this section we address Question 2.1 and limit our discussion to models with a square Jacobian matrix and therefore a single singular locus equation. Our observations led us to the following conjecture.

Conjecture 3.4. *Let \mathcal{M} be a strongly connected, generically locally identifiable linear compartment model. If parameter k_{ij} divides the singular locus equation of model \mathcal{M} , then the model \mathcal{M}' obtained by deleting the edge k_{ij} is unidentifiable.*

Furthermore, we have observed that many dividing edges of the singular locus equation belong to one or more of the following classifications based on the effects of their removal:

1. Removal of the dividing edge breaks the strong connectedness of the model.
2. Removal of the dividing edge increases the length of the shortest path from input to output.
3. Removal of the dividing edge creates “pseudo-leaks” or “pseudo-inputs.”

3.4.1 Classification 1: Breaking Strong Connectedness

It is our observation that in many cases the dividing edges of a model are such that the removal of that edge would break the strong connected characteristic of the model. The effect of breaking the strong connectedness of the model is unsurprising as we conjecture that the model obtained by removing a dividing edge is unidentifiable. The lack of the strong connected characteristic means that information is not flowing completely through the model from input to output resulting in a situation that would lend itself to an unidentifiable result.

Remark 3.4. It is important to note that many of the methods and known results pertaining to identifiability rely on the strong connected nature of a linear compartment model. Therefore, when examining models that are no longer strongly connected when a dividing edge has been removed, we must keep in mind that the method we have outlined in the Background that uses the input-output equation may not be accurate. (Remark 3.11 in [1] one of the potential issues that may arise)

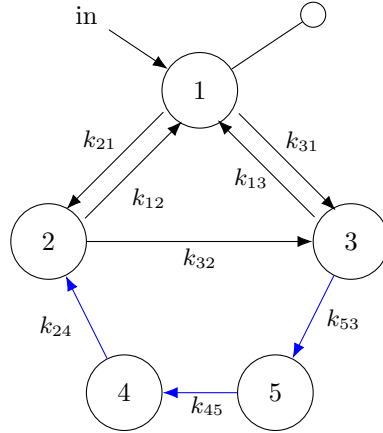


Figure 3: Model with $In = Out = 1$ and dividing edges k_{24}, k_{45}, k_{53}

Figure 3 depicts a strongly connected linearly compartment model with dividing edges $k_{24}, k_{45},$ and k_{53} . The removal of any one of these edges results in an unidentifiable linear compartment model. The removal also results in a model that is not strongly connected. All other edges in the model do not break strong connectedness in the model when removed and do not appear as dividing edges in the singular locus equation.

It is important to note that while in many instances each dividing edge is such that removing it results in the model no longer being strongly connected, this is not always the case.

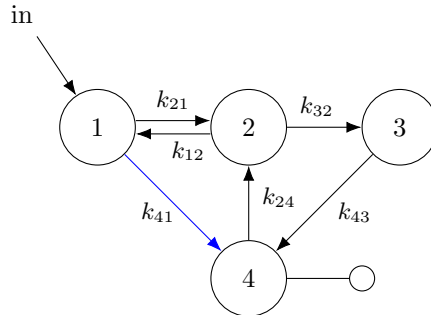


Figure 4: Model with $In = \{1\}, Out = \{4\}$ and dividing edges k_{41}

Figure 4 depicts a linear compartment model where k_{41} is the only dividing edge. The model obtained by removing k_{41} is unidentifiable. In addition, removing k_{41} does not break the strong connectedness of the

model. It does, however, affect the length of the shortest path from input to output which brings us to our second classification.

3.4.2 Classification 2: Affecting Length of Path from Input to Output

We have observed that, at times, the dividing edge of the singular locus equation is an edge in the shortest path from the input compartment to the output compartment and its removal increases the length of the shortest path from input to output.

Dividing edges that fall into this classification may also be a part of Classification 1 in that their removal results in a model that is not strongly connected. Some, however, do not fall into Classification 1, meaning that their removal does not break the strongly connected nature of the resulting model. In this latter case, we can apply Conjecture 2.1 to resolve Conjecture 3.4 for this subcase.

Theorem 3.4 (Subcase of Conjecture 3.4). *Let \mathcal{M} be a linear compartment model with $|In| = |Out|$, $Leak = \emptyset$, and a square Jacobian matrix of the coefficient map, for which Conjecture 2.1 holds. If there exists a dividing edge k_{ij} in \mathcal{M} such that the model \mathcal{M}' obtained by removing k_{ij} is strongly connected and the length of the shortest path from input to output has increased by at least 2, then \mathcal{M}' is unidentifiable.*

Proof. Let L be the length of the shortest path from input to output in \mathcal{M} . Let K be the number of parameters in \mathcal{M} and C be the number of coefficients of the input-output equation. The Jacobian matrix is square, so $K = C$. Furthermore, if the removal of the edge k_{ij} increases the length of the path from input to output, then $In \neq Out$. Thus, by Conjecture 2.1, the number of coefficients is $C = n - 1 + n - L = 2n - L - 1$. (Here we assume the model has no leaks, the case for a model with leaks follows a similar proof). When the dividing edge is removed and the length of the path from input to output increases to $L' \geq L + 2$. Thus, the number of coefficients decreases to $C' = n - 1 + n - L'$, which is at most, $C' = 2n - L - 3$. The number of parameters decreases to $K' = K - 1 = 2n - L - 2$. Thus, the Jacobian matrix of the model \mathcal{M}' has dimensions $(2n - L - 3) \times (2n - L - 2)$. Therefore, by Proposition 2.2 \mathcal{M}' is not identifiable. \square

Theorem 3.3 is demonstrated in the model depicted in Figure 4. In the original model, the shortest path from input to output has length 1 (the edge k_{41}). The model is identifiable with 6 parameters and $2n - 2 = 2(4) - 2 = 6$ coefficients, resulting in a 6×6 Jacobian matrix of the coefficient map. When k_{41} is removed, the resulting model is again strongly connected, however the length of the shortest path from input to output has increased to 3 (the edges k_{21}, k_{32}, k_{43}). There are now 5 parameters and $2n - 4 = 2(4) - 4 = 4$ coefficients resulting in a 4×5 Jacobian matrix. By Proposition 2.2, this model is unidentifiable.

A second example of this subcase of Conjecture 3.4 is shown below.

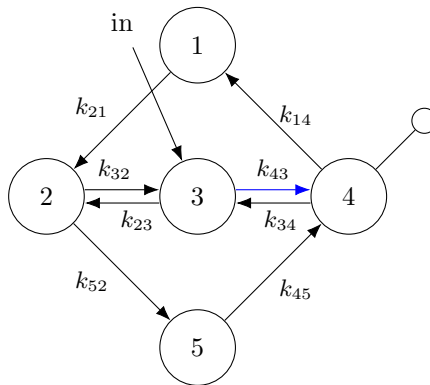


Figure 5: Model with $In = \{3\}$, $Out = \{4\}$ and dividing edge k_{43}

Figure 5 depicts an identifiable, strongly connected linear compartment model with dividing edge k_{43} . The model has 8 parameters and the shortest path from input to output has length 1 (the edge k_{43}) so the number of coefficients is $2n - 2 = 2(5) - 2 = 8$. This results in an 8×8 Jacobian matrix. When the edge

k_{43} is removed, the length of the shortest path from input to output increases to 3 (edges k_{23} , k_{52} , k_{45}). The resulting model has 7 parameters and $2n - 4 = 2(5) - 4 = 6$ coefficients, resulting in a 6×7 Jacobian matrix. By Proposition 2.2, the model is unidentifiable.

3.4.3 Classification 3: Creation of Pseudo-leaks and Pseudo-inputs

A final interesting classification that our observations have led us to develop is the dividing edge that when removed creates what we call “pseudo-leaks” and “pseudo-inputs”. All models that belong to this classification must also belong to Classification 1

Definition 3.1. A *pseudo-leak* is an edge from one compartment to a compartment that has no outgoing edges. A *pseudo-input* an edge to a compartment from a compartment that has no incoming edges.

The goal of defining this vocabulary is to emphasize the way in which we suspect the behavior of pseudo-inputs and pseudo-leaks mimics the behavior of traditional inputs and leaks. For the case of a pseudo-input, the concept is not fully understood in the context of a physical application, as it is unclear what - if anything - leaves the compartment through the pseudo-input when no substance is flowing into the compartment.

However, for the case of a pseudo-leak the physical application supports the conjecture that these pseudo-leaks behave in the same manner as a traditional leak. Furthermore, we have observed models with pseudo-leaks and compared them to the corresponding model with the traditional leak and our observations indicate that these types of edges do in fact behave in the same manner. This means that results on the effect of leaks on a strongly connected model could be applied to cases of models with pseudo-leaks.

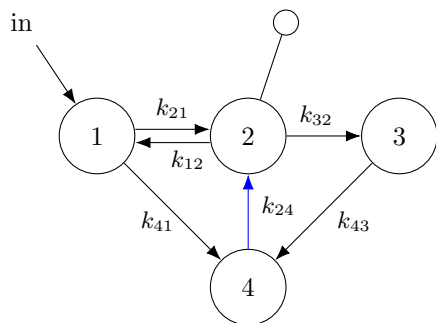


Figure 6: Model with $In = \{1\}$, $Out = \{2\}$ and dividing edge k_{24}

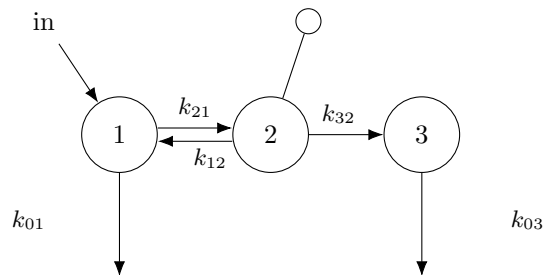


Figure 7: Model with $In = \{1\}$, $Out = \{2\}$

Figure 6 depicts a strongly connected linear compartment model with dividing edge k_{24} . The removal of k_{24} creates two pseudo-leaks, k_{41} and k_{43} . We compare the coefficients of the model in Figure 6 to the coefficients of the model depicted in Figure 7 where we have removed compartment 4 and replaced the pseudo-leaks by traditional leaks k_{01} and k_{03} . The coefficients of the input-output equation of each model are as follows:

$$\begin{aligned}
 c_{y_1} &= k_{12} + k_{21} + k_{32} + k_{41} + k_{43} & c_{y_1} &= k_{12} + k_{21} + k_{32} + k_{01} + k_{03} \\
 c_{y_2} &= k_{12}k_{41} + k_{21}k_{32} + k_{12}k_{43} + k_{32}k_{41} + k_{41}k_{43} + k_{21}k_{43} & c_{y_2} &= k_{12}k_{01} + k_{21}k_{32} + k_{12}k_{03} + k_{32}k_{01} + k_{01}k_{03} + k_{21}k_{03} \\
 c_{y_3} &= k_{12}k_{41}k_{43} + k_{21}k_{43}k_{32} + k_{32}k_{41}k_{43} & c_{y_3} &= k_{12}k_{01}k_{03} + k_{21}k_{03}k_{32} + k_{32}k_{01}k_{03} \\
 c_{u_1} &= -k_{21} & c_{u_1} &= -k_{21} \\
 c_{u_2} &= -k_{21}k_{43} & c_{u_2} &= -k_{21}k_{03}
 \end{aligned}$$

The direct replacement of k_{41} and k_{43} by k_{01} and k_{03} in the coefficients of the input-output equation indicates that the pseudo-leaks and leaks behave the same way in the analysis of the model.

Remark 3.5. The implication of pseudo-leaks behaving as traditional leaks is that one can apply previous results on the behavior of leaks in a model to the model created by the removal of dividing edges. However, it is important to note that this relies on the assumption that the all compartments, except the compartment

to which the pseudo-leaks are directed, induce a graph that is strongly connected. We have not yet observed an interesting case in which this is true. Therefore, previous results on the effects of leaks may not apply due to the lack of strong connectedness.

3.4.4 Interesting Examples

We now finish our discussion of the dividing edges of the singular locus with some particularly interesting examples of models that we have observed in our research.

Example 1:

The first example is a model with a dividing edge that does not fall into any of the Classifications described above.

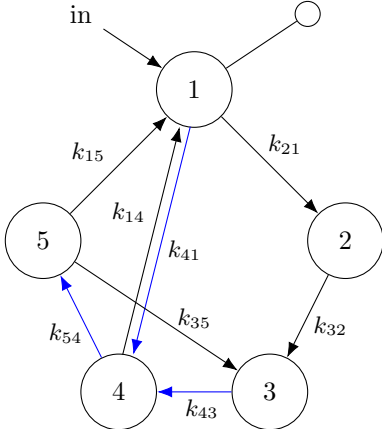


Figure 8: Model with $In = Out = \{1\}$

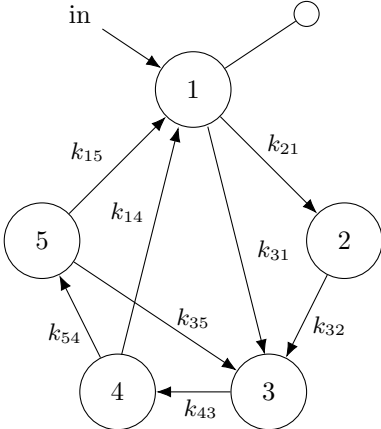


Figure 9: Model with $In = Out = \{1\}$

Figure 8 depicts a strongly connected, linear compartment model with $In = Out = \{1\}$ and dividing edges k_{54}, k_{43} , and k_{41} . The edge k_{41} is an interesting case of a dividing edge in that the removal of k_{41} results in a model that is still strongly connected (so no pseudo-leaks or pseudo-inputs were created) and the length of the shortest path from input to output has not increased. The model obtained by removing k_{41} has an 8×7 Jacobian matrix for which all minors are 0. Thus, the model obtained by removing k_{41} is unidentifiable, as expected by Conjecture 3.4, however this is not explained by any of the classifications previously described.

Also interesting is the model depicted in Figure 9 where the edge k_{41} (in Figure 8) is replaced by the edge k_{31} . What is interesting is that while it appears that k_{31} would act in a similar manner to k_{41} , in fact the model in Figure 9 is unidentifiable.

Example 2:

The next interesting examples we consider involved edges that appear as dividing edges following small changes to the model structure.

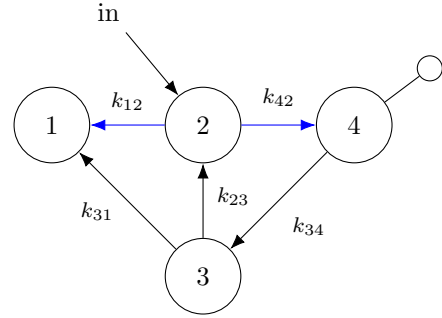
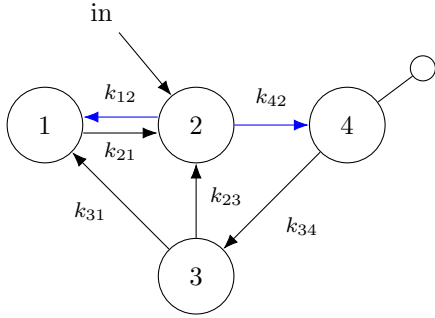


Figure 10: Model 1 with $In = \{2\}$ and $Out = \{4\}$ **Figure 11:** Model 2 with $In = \{2\}$ and $Out = \{4\}$

Figure 10 displays a strongly connected linear compartment model, Model 1, that is identifiable with a 6×6 Jacobian matrix. The dividing edges are k_{12} and k_{42} . Figure 11 displays the not strongly connected linear compartment model, Model 2, obtained by removing the edge k_{21} from the model in Figure 10. This model is also identifiable with a 6×5 Jacobian matrix for which the dividing edge k_{42} divides all minors. We want to draw attention to the fact that when k_{21} is removed the edge k_{12} does not appear as a dividing edge in the singular locus of Model 2.

Next we consider removing the dividing edges from Model 1. When the edge k_{12} is removed from Model 1, the resulting model is unidentifiable with a 6×5 Jacobian matrix for which all minors are 0. When the edge k_{42} is removed from Model 1, the resulting model is unidentifiable with a 3×5 Jacobian matrix.

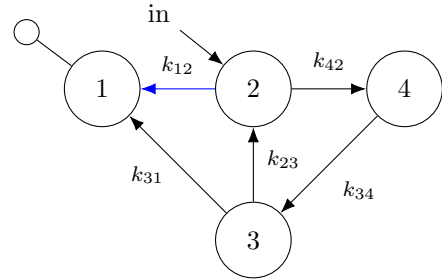
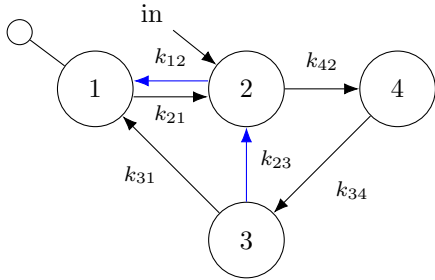


Figure 12: Model 3 with $In = \{2\}$ and $Out = \{1\}$ **Figure 13:** Model 4 with $In = \{2\}$ and $Out = \{1\}$

Figure 12 displays a strongly connected, linear compartment model, Model 3, that is identifiable with a 6×6 Jacobian matrix. The dividing edges are k_{12} and k_{23} . Figure 13 displays a model, Model 4, that is not strongly connected. This model is identifiable with a 6×5 Jacobian matrix for which the dividing edge k_{12} which divides all minors. Again, we want to draw attention to the fact that the removal of k_{21} results in a model in which k_{23} no longer appears as a dividing edge in the singular locus.

When dividing edge k_{12} is removed from Model 3, the resulting model is unidentifiable with 3×5 Jacobian matrix. When dividing edge k_{23} is removed from Model 3, the resulting model is unidentifiable with a 5×5 Jacobian matrix that has a determinant of 0.

It is our observation that there are similarities between the dividing edge of Model 1 that stopped dividing the singular locus of Model 2, and the dividing edge in Model 3 that stopped dividing in Model 4. In both

cases, when this edge was removed from Model 1 and Model 3 to obtain the resulting models were unidentifiable, with a square or tall rectangular Jacobian matrix. That is, for a given number of coefficients, C , and number of parameters, K , the dividing edges from Model 1 and Model 3 which stopped dividing Model 2 and Model 4, respectively, when removed resulted in models with $K \leq C$. In contrast, the dividing edges which remained from Model 1 to Model 2 and Model 3 to Model 4 when removed resulted in models $C \leq K$.

Example 3:

We conclude our discussion of the singular locus with a comment on the case where a model has a non-square Jacobian and therefore the singular locus is defined by the set of all maximal minors of the Jacobian matrix. In this case we specify that a dividing edge is such that it divides every such minor. Conjecture 3.4 states that the removal of such an edge results in an unidentifiable model. However, it is interesting to consider the case where an edge divides one or multiple minors of the Jacobian but not all. The following example depicts such a case.

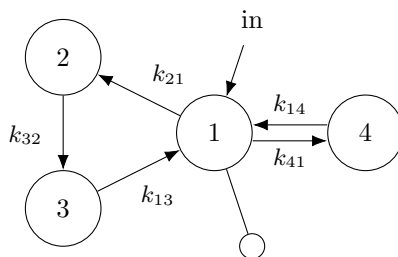


Figure 14: Model with $In = Out\{1\}$

The strongly connected model in Figure 14 has 5 parameters and 6 coefficients resulting in a 6×5 Jacobian matrix of the coefficient map. Therefore, the singular locus consists of the 6 minors of the Jacobian matrix. There is no dividing edge k_{ij} that divides all 6 such minors. However, the edge k_{14} divides one of the 6 minors. Furthermore, when the edge k_{14} is removed, the resulting model is *identifiable*.

This example is interesting because it provides a counterexample to the possible conjecture that if an edge k_{ij} divides a minor of the Jacobian matrix, then the model obtained by removing k_{ij} is unidentifiable. Rather, this example indicates that the edge must divide all minors of the Jacobian for its removal to result in an unidentifiable model. More work remains to be done in understanding the differences between each of the minors of the Jacobian matrix and the information each holds.

4 Discussion

In this work, we examined the effect of two types of model operation on the identifiability of a linear compartment model: the first being the addition and removal of leaks, and the second being the removal of a dividing edge of the singular locus equation. In examining these questions we obtained results on the structure of coefficients in models with and without leaks and resolved part of Conjecture 2.1 for the case of the coefficients on the left-hand side of the equation. Using these results, we concluded that the addition of a leak to a model that is unidentifiable with more parameters than coefficients results in a model that is unidentifiable with more parameters than coefficients. Future work would look to resolve the case where a leak is added to a model that is unidentifiable and has a Jacobian matrix with a determinant of zero.

We made significant progress toward resolving Conjecture 3.3 in showing that this conjecture is equivalent to Conjecture 3.1 and therefore allowing conclusions to be drawn for classes of models where one of the conjectures holds. Future work would seek to resolve Conjecture 3.1.

On the topic of the singular locus equation and dividing edges, we have generated several interesting examples to provide insight on the behavior of those dividing edges and the effect of removing one such edge. In addition

we have proposed three classifications corresponding to characteristics of dividing edges that we observed as occurring often. Future work would seek to better understand these characteristics and their potential connection to resolving Conjecture 3.4. Finally, we propose approach to examining this topic. Can we identify which edges in a linear compartment model divide the singular locus? An answer to this question would provide insight on the effects of their removal.

5 Acknowledgements

We would like to thank Dr. Anne Shiu and Aleksandra Sobieska for their assistance and mentorship throughout this research project. This paper would not have been possible without their trust and dedication.

This research was completed in the 2020 REU in the Department of Mathematics at Texas A&M University, supported by NSF grant DMS-1757872.

6 Appendix A: Matlab code for calculation of the singular locus of a linear compartment model

Here we include the Matlab code for the program used to determine the identifiability of linear compartment models, calculate their singular locus and determine dividing edges for all examples included in this paper.

Program : Calculate the Singular Locus of Linear Compartment Model

Date : 06/09/2020

Author : Kate Johnston

Purpose : Calculates the singular locus of a linear compartment model as well as the dividing edges.

Input: Input compartment, output compartment, leak compartments and an adjacency matrix, X, describing the edges where $X_{ij} = 1$ if the edge from compartment i to compartment j exists.

Output: 'identifiable' or 'unidentifiable', the size of the Jacobian matrix, and - if the model is identifiable - the singular locus and dividing edges.

```
%user input
in = ; %input node
out = ; %output node
n = ; %total number of nodes
leak = []; %contains each node with a leak, empty when there is no leak

%edges
%Xij = 1 - if edge ij exists (from node j to node i), else 0
X = [];
k1 = []; %array to hold created edges - (leave empty - no input required)

%create matrix A
A = sym(zeros(n,n));
for(i = 1:n)
    for(j = 1:n)
        if(i == j)
            lk = false;
            for(k = 1:length(leak)) %check to see if node has a leak
                if(i == leak(k)) %if node has a leak
                    p = 0; %variable to hold sum of edges
                    %create edge for leak k0i and add to array of edges k1
                    a = str2sym(strcat('k', num2str(0), num2str(i)));
```



```

k1 = [k1 a];
p = -a;
%sum of edges
for(t = 1:size(X,1))
    %check ith column (corresponds to edges leaving node i of form kji)
    if(X(t,i) == 1) %if there is an edge i to t
        %create a sym k representing the edge
        a = str2sym(strcat('k', num2str(t), num2str(i)));
        %check if edge already exists
        exists = false;
        if(not(isempty(k1))) %check there are edges in k1
            for(s = 1:length(k1)) %check all entries of k1 for a match
                if(isequal(k1(s),a)) %if there is a match set 'exists' = true
                    exists = true;
                    break
                end
            end
        end
        %if edge is not in array then add edge
        if(exists == false)
            k1 = [k1 a];
        end
        %add edge to sum
        p = p - a;
    end
end
%assign entry of A to be sum of edges
A(i,j) = p;
lk = true; %there is a leak
end
end
%if i is not a leak
if(lk == false)
    p = 0;%variable to hold computed sum
    %iterate over the rows of matrix X
    for(t = 1:size(X,1))
        %check ith column (corresponds to edges leaving node i)
        if(X(t,i) == 1) %if there is an edge i to t
            %create a sym k representing the edge
            a = str2sym(strcat('k', num2str(t), num2str(i)));
            %check if edge already exists
            exists = false;
            if(not(isempty(k1))) %check there are edges in k1
                for(s = 1:length(k1)) %check all entries of k1 for a match
                    if(isequal(k1(s),a) == 1) %there is a match
                        exists = true;
                        break
                    end
                end
            end
            %if edge is not in array then add edge
            if(exists == false)
                k1 = [k1 a];
            end
        end
    end
end

```

```

        end
        %add edge to sum
        p = p + a;
    end
    end
    %entry of matrix (is negative according to formula)
    A(i,j) = -p;
end
else %else  $i \neq j$ 
    %check if there is an edge j to i
    if(X(i,j) == 1)
        %create corresponding k edge
        a = str2sym(strcat('k', num2str(i), num2str(j)));
        %check if edge already exists
        exists = false;
        if(not(isempty(k1))) %check there are edges in k1
            for(s = 1:length(k1)) %check all entries of k1 for a match
                if(isequal(k1(s),a) == 1)
                    exists = true;
                    break
                end
            end
        end
        %if edge is not in array then add edge
        if(exists == false)
            k1 = [k1 a];
        end
        %add edge to matrix A
        A(i,j) = a;
    else %else -; there is no edge j to i
        A(i,j) = 0;
    end
end
end
end
end

%create dI -; identity matrix
syms ddt;
dI = ddt.*eye(n);

%compute det(dI-A)
M = dI - A;
collect(det(M),ddt); %will print the determinant with terms collected by d/dt power
c_y = coeffs(det(M),ddt); %removes coefficients of d/dt terms
c_y = c_y(c_y = 1); %remove the 1 coefficient
c_y = flip(c_y); %flip so in proper order (descending powers of ddt)

%compute det((dI-A)in/out)
B = M;
B(in,:) = []; %remove "in" row of M
B(:,out) = []; %remove "out" column of M
%create vector holding u coefficients
collect(det(B),ddt); %will print determinant with terms collected by d/dt power
c_u = coeffs(det(B), ddt); %removes coefficients of d/dt terms

```

```

%check for coefficient of 1 and remove if present
c_u = c_u(c_u == 1);
%flip vector to match order of ascending powers
c_u = flip(c_u);

%make a vector of all coefficients by combining c_u and c_y
c = cat(2, c_y, c_u);
%make jacobian matrix - rows correspond to coeffs, columns to edges (kij)
J = jacobian(c, k1);

%find determinants of all submatrices (if matrix is rectangular - else det of matrix)
if(size(J,2) > size(J,1)) %check there are more coeffs than parameters
    if(size(J,2) == size(J,1)) %if matrix is square
        if det(J) == 0
            fprintf('identifiable');
            size(J)
            det(J)
            %dividing edges
            factor(det(J))
        else
            fprintf('unidentifiable')
            size(J)
            det(J)
        end
    else %matrix is not square -> so multiple dets
        dets = []; %array holds all determinants of submatrices
        subrows = num2cell(nchoosek(1:size(J,1), size(J,2)), 2);
        subm = cellfun(@(rows) J(rows, :), subrows, 'UniformOutput', false);
        for i = 1:length(subm)
            dets = [dets det(cell2sym(subm(i)))]; %populate array with dets
        end
        size(J)
        dets
        %determine the dividing edges of singular locus by finding common
        %dividing edges between determinants
        dets2 = nonzeros(dets); %remove determinants = 0
        k = intersect(factor(dets2(1)), factor(dets2(2)));
        %check for common divisor between first and second det
        for i = 2:length(dets2)
            = intersect(k, factor(dets2(i)));
        end
        %print list of dividing edges
        k
    end
else %there are more parameters than coeffs so matrix is not full rank and unidentifiable
    size(J)
    fprintf('matrix is unidentifiable')
end

```

References

- [1] Elizabeth Gross, Heather A. Harrington, Nicolette Meshkat, and Anne Shiu. Linear compartmental models: input-output equations and operations that preserve identifiability. *SIAM J. Appl. Math.*, 79(4):1423–1447, 2019.
- [2] Elizabeth Gross, Nicolette Meshkat, and Anne Shiu. Identifiability of linear compartment models: the singular locus. *Preprint*, [arXiv:1709.10013](https://arxiv.org/abs/1709.10013), 2017.
- [3] Nicolette Meshkat, Seth Sullivant, and Marisa Eisenberg. Identifiability results for several classes of linear compartment models. *B. Math. Biol.*, 77(8):1620–1651, 2015.