Inferring Reaction Models from ODEs

François Fages, Steven Gay, and Sylvain Soliman

INRIA Paris-Rocquencourt, EPI Contraintes, France

Many models in Systems Biology are described as Ordinary Differential Equations (ODEs), which allows for numerical integration, bifurcation analyses, parameter sensitivity analyses, etc. Before fixing the kinetics and parameter values however, various analyses can be performed on the structure of the model. This approach has rapidly developed in Systems Biology in the last decade, with for instance, the analyses of structural invariants in Petri net representation [4] model reductions by subgraph epimorphims [2], qualitative attractors in logical dynamics or temporal logic properties by analogy to circuit and program verification. These complementary analysis tools do not rely on kinetic information, but on the structure of the model with reactions.

The Systems Biology Markup Language (SBML) of [3] is now a standard for sharing and publishing reaction models. However, since SBML does not enforce any coherence between the structure and the kinetics of a reaction, an ODE model can be transcribed in SBML without reflecting the real structure of the reactions, hereby invalidating all structural analyses.

In this paper we propose a general compatibility condition between the kinetic expression and the structure of a reaction, describe an algorithm for inferring a reaction model from an ODE system, and report on its use for automatically curating the writing in SBML of the models in the repository biomodels.net.

1   Theory of Well-Formed Reaction Kinetics

Let us consider a finite set $\{x_1, \ldots, x_v\}$ of molecular species. A *reaction model R* is a finite set of n reactions, written $R = \{\ e_i \ \texttt{for} \ \ r_i \ / \ m_i \ \texttt{=>} \ \ p_i \ \}_{i=1,\ldots,n}$, where e_i is a mathematical expression over species concentrations, also written x_i by abuse of notation, and symbolic parameters that are supposed positive; r_i, m_i and p_i are multisets of species which represent the reactants, the inhibitors, and the products of the reaction respectively. The species that are both reactants and products in a reaction are called catalysts. Catalysts and inhibitors are called modifiers in SBML. For a multiset r of molecular species, i.e. a function $V \to \mathbb{N}$, we denote by $r(x)$ the multiplicity of x in r, i.e. $r(x) = 0$ if x does not belong to r, and $r(x) \geq 1$ if x belongs to r, which is also written $x \in r$. The empty multiset is written ‗.

Let us call *non-decomposable* a mathematical expression that is syntactically not an addition nor a subtraction, and that cannot be reduced at top-level by the laws of distributivity of the product and division, and let us say

Definition 1 *A reaction* e `for` r $/$ m `=>` p *over molecular species* $\{x_1, \ldots, x_v\}$ *is* well-formed *if the following conditions hold:*

1. *e is a well-defined, nonnegative and partially differentiable mathematical expression for any values $x_1 \geq 0, \ldots, x_v \geq 0$;*
2. *$x \in r$ if and only if $\partial e / \partial x > 0$ for some $x_1 \geq 0, \ldots, x_v \geq 0$;*
3. *$x \in m$ if and only if $\partial e / \partial x < 0$ for some $x_1 \geq 0, \ldots, x_v \geq 0$;*

*The reaction is* non-decomposable *if its kinetic expression e is non-decomposable.*

These well-formedness conditions are met by standard kinetics, such as mass action law, Michaelis-Menten, Hill, and negative Hill kinetics. However there are ODE system which cannot be presented by a well-formed non-decomposable reaction model.

Example 1. The ODE system $\dot{x} = -k$ can be associated to the reaction `k for x => _` which is not well-formed, or to `k+l*x for x => _` and `l*x for x => 2*x` which are not non-decomposable, but cannot be associated to a well-formed non-decomposable system.

Example 2. On the other hand, the ODE with symbolic parameters $k1, k2, k3$
$$\dot{pMPF} = k2 * [MPF] * [Wee1] - k1 * [pMPF] * [Cdc25]$$
$$\dot{MPF} = k1 * [pMPF] * [Cdc25] - k2 * [MPF] * [Wee1]$$
$$\dot{Wee1} = k3/(k4 + [Clock]), \dot{Cdc25} = 0, \dot{Clock} = 0$$
is associated to the well-formed non-decomposable model
```
k1*[pMPF]*[Cdc25] for pMPF + Cdc25 => MPF + Cdc25
k2*[MPF]*[Wee1]   for MPF + Wee1 => pMPF + Wee1
k3/(k4+[Clock])   for _ / Clock => Wee1
```
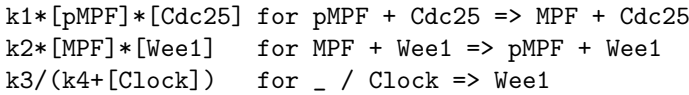
2   Reaction Model Inference Algorithm

The following algorithm for inferring reactions from ordinary differential equations is based on a syntactical normal form for ODE systems. Unlike the algorithm proposed in [5], our algorithm does not rely on uniqueness properties and always succeeds even when there are no corresponding well-formed reaction models, by inferring possibly non well-formed reactions.

Let us say that an expression is in *additive normal form* if it is of the form $\sum_{s=1}^{t} c_s * f_s$ with c_s numerical coefficients and f_s non-decomposable terms without coefficients. An ODE system is in additive normal form if each equation is in additive normal normal form as follows $\dot{x_i} = \sum_{s=1}^{t} c_{i,s} * f_s$, $1 \leq i \leq v$ where t is the number of non-decomposable terms in the system.

The idea of our reaction inference algorithm is to normalize the ODE and infer a corresponding reaction model by sorting the terms of the equations and using their coefficients as stoichiometric coefficients. To test the sign of partial derivatives, we content ourselves with an approximate test by comparing the exponents.

Algorithm 2 input: *ODE system O*

1. *$O \leftarrow$ additive-normal-form(O)*

2. $R \leftarrow \emptyset$
3. *for each non-decomposable term f of an equation in O*
   (a) *let $r \leftarrow \_$, $p \leftarrow \_$, $m \leftarrow \_$*
   (b) *for each variable x where f occurs with coefficient c in $\dot{x}$ in O*
       i. *if $c < 0$ then $r(x) \leftarrow -c$*
       ii. *if $c > 0$ then $p(x) \leftarrow c$*
   (c) *for each variable x such that $r(x) = 0$ and $\frac{\partial f}{\partial x} > 0$ for some values*
       i. *$r(x) \leftarrow 1$*
       ii. *$p(x) \leftarrow p(x) + 1$*
   (d) *for each variable x such that $\frac{\partial f}{\partial x} < 0$ for some values*
       i. *$m(x) \leftarrow 1$*
   (e) *$R \leftarrow R \cup \{f$ `for` r `/` m `=>` $p\}$*
4. **output***: reaction model R*

Proposition 3 *The reaction model inferred by Algorithm 2 from an ODE system in additive normal form $\dot{x}_i = \sum_{s=1}^{t} c_{i,s} * f_s$ for $1 \le i \le v$ is the set of non-decomposable reactions $\{f_s$ for $r_s/m_s \to p_s\}_{1 \le s \le t}$ where*

$$r_s = \sum_{c_{i,s}<0}(-c_{i,s}) * x_i + \sum_{\{i \mid c_{i,s} \ge 0, \frac{\partial f_s}{\partial x_i}>0\}} x_i \qquad p_s = \sum_{c_{i,s}>0} c_{i,s} * x_i + \sum_{\{i \mid c_{i,s} \ge 0, \frac{\partial f_s}{\partial x_i}>0\}} x_i \qquad m_s = \{x| \frac{\partial f_s}{\partial x} < 0\}$$

Algorithm 2 always computes a reaction model with an equivalent associated ODE system but this reaction model may not be well-formed. In particular, step 3b may infer reactions with reactants that do not occur in the kinetic expression. On the other hand, all variables appearing in the kinetics will now appear in the reaction as either catalysts (step 3c), inhibitors or both (step 3d). We have:

Proposition 4 *The ODE system associated to the reaction model inferred from an ODE system O is equivalent to O. The reaction model inferred from the ODEs associated to a non-decomposable well-formed reaction model is well-formed and non-decomposable.*

For space reasons, we do not describe here the preprocessor that is applied on the ODE system for detecting simplifications by mass conservation linear invariants and inferring hidden molecules, prior to the inference of reactions.

Example 3. The model of Example 2 has one invariant: $pMPF + MPF$ is a constant c. Replacing $pMPF$ by $c - MPF$ yields to the system
$$\dot{MPF} = k1 * (c - [MPF]) * [Cdc25] - k2 * [MPF] * [Wee1]$$
$$\dot{Wee1} = k3/(k4 + [Clock]) \qquad\qquad \dot{Cdc25} = 0 \qquad\qquad \dot{Clock} = 0$$
From this system, Algorithm 2 would infer the reactions:

```
c*k1*[Cdc25]        for Cdc25 => Cdc25 + MPF
  k1*[Cdc25]*[MPF]  for MPF + Cdc25 => Cdc25
  k2*[MPF]*[Wee1]   for MPF + Wee1 => Wee1
  k3/(k4+[Clock])   for _ / Clock => Wee1
```

However, the preprocessor recognizes the linear invariant and introduces a molecule y for the expression $[y] = c - [MPF]$ which yields in this case to the same reactions as of Example 2 with $y = pMPF$.

3   Evaluation Results on Biomodels.net

Out of the 409 models from the curated branch of the latest version (21) of `biomodels.net`, 340 models have proper kinetic laws. We compare the number of non well formed reaction models before and after the automatic curation obtained by exporting SBML to ODE format and applying our reaction inference algorithm to the ODE. The following table summarizes the improvement by counting the number of models with BIOCHAM warnings: *well-formed*:

- "K not R" denotes the number of models in which the concentration of some compound appears in a kinetic law but it is neither a reactant nor a modifier;
- "R not K" denotes the number of models in which some compound is marked as reactant or modifier but does not appear in its kinetic law.
- "Negative" denotes the number of models where a minus sign appears in the kinetic expression at some place that is not inside an exponent expression.

225 models, i.e. 66% of the 340 reaction models of the original "curated" part of biomodels.net are non well formed and produce some warning. Our algorithm is able to automatically reduce the number of non well-formed models by 58%, from 66% to 28%:

Biomodels.net	K not R	R not K	Neg.	Any warning
Originally Curated	165	120	148	225 (66.17%)
Automatically Curated	0	81	39	97 (28.52%)

The algorithm completely removes the "K not R" warnings. For the two other warnings, since the algorithm focuses on *non-decomposable* kinetics, it results in curated models quite close to the original ones, but does not tackle thoroughly the case of reactions with rates independent of some reactant, as in Example 1. For these reasons, 97 over 340 models remain with a non well-formedness warning after automatic curation [1].

References

1. Fages, F., Gay, S., Soliman, S.: Automatic curation of SBML models based on their ODE semantics. Technical Report 8014, Inria Paris-Rocquencourt (July 2012)
2. Gay, S., Soliman, S., Fages, F.: A graphical method for reducing and relating models in systems biology. Bioinformatics 26(18), i575–i581 (2010), special issue ECCB 2010
3. Hucka, M., et al.: The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. Bioinformatics 19(4), 524–531 (2003), http://sbml.org/
4. Reddy, V.N., Mavrovouniotis, M.L., Liebman, M.N.: Petri net representations in metabolic pathways. In: Hunter, L., Searls, D.B., Shavlik, J.W. (eds.) Proceedings of the 1st International Conference on Intelligent Systems for Molecular Biology (ISMB), pp. 328–336. AAAI Press (1993)
5. Soliman, S., Heiner, M.: A unique transformation from ordinary differential equations to reaction networks. PLoS One 5(12), e14284 (2010)