Theoretical Computer Science 641 (2016) 11–24

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

# Complete characterization for the fit-preserving data refinement of mass-action reaction networks

Cristian Gratie *, Ion Petre

*Computational Biomodeling Laboratory, Turku Centre for Computer Science and Department of Computer Science, Åbo Akademi University, Joukahaisenkatu 3–5 A, 20520 Turku, Finland*

ARTICLE INFO

*Article history:*
Received 14 June 2015
Received in revised form 29 February 2016
Accepted 26 March 2016
Available online 30 March 2016
Communicated by S. Miyano

*Keywords:*
Biomodeling
Model fit
Quantitative model refinement
Fit-preserving refinement
Canonical refinement
Brusselator

ABSTRACT

The data refinement of reaction-based models consists in substituting species from the original model with several subspecies in the refined one. Fit-preserving refinement, where the goal is to capture the same species dynamics as the original model, helps reduce the computational cost of model fitting by reusing previously fit rate constants. In this paper we give a complete characterization of fit-preserving refinement, as necessary and sufficient linear constraints on the reaction rate constants. Our result is applicable for mass-action reaction networks with uniquely identifiable rate constants. We demonstrate our result on the well-known Brusselator model.



## 1. Introduction

The top-down development of large biological models relies on iteratively adding details to an initial abstraction of the biological phenomena. At each step, the numerical setup of the mathematical model that describes the dynamic behavior of the system needs to be fit against existing experimental data. Model fitting is often a computationally expensive process, thus it would be beneficial to reuse previously fit values to initialize the parameter estimation routines for the refined model.

We focus here on reaction-based models that rely on mass-action kinetics [1]. Data refinement, introduced in [2], refers to the replacement of one (or more) species with several variants, or *subspecies*, in the refined model. The approach presented in [2] for the numerical setup of the refined model aims to preserve the fit of the original model and consists in assigning parameter values in such a way that the ODEs describing the original model can be recovered as a sum of ODEs from the refined model. We formalized this idea in [3] as *fit-preserving refinement* and provided a sufficient condition for preserving the numerical fit. The condition links the refined parameters to those of the original model without the need for inspecting the ODEs.

The approach from [3] provides a partial answer to the open problem, formulated in [4], of finding values for the unknown parameters of a partially specified refined model so that it preserves the numerical fit of the original model: as long as the known values do not already lead to a violation of the fit-preservation constraints, the problem has at least one solution, which can be effectively computed.

* Corresponding author.
*E-mail addresses:* cgratie@abo.fi (C. Gratie), ipetre@abo.fi (I. Petre).

http://dx.doi.org/10.1016/j.tcs.2016.03.027
0304-3975/

Since the proposed condition is only sufficient, for some models it is possible to build fit-preserving refinements for which the parameter values do not satisfy the condition. We address this issue here and show that all such models are in fact "pathological" cases that do not have uniquely identifiable rate constants, a critical requirement for being able to fit the model with experimental data, see [5].

This paper is an extended version of [3]. In addition to [3], we include here the full proof of its main result, as well as all mathematical considerations leading to it. Moreover, we prove that the sufficient condition given in [3] is in fact a necessary and sufficient condition, i.e. a complete characterization of fit-preserving refinement, provided that the original model has uniquely identifiable rate constants. Finally, we illustrate our approach on a new case study, that of the Brusselator [6], to show that fit-preserving refinement alone can give rise to very different refined models.

While this paper focuses on the forward direction, note that the idea of refinement can also be played backwards, i.e. one can start with a detailed model, which may be difficult to study and understand due to its complex dynamics, group together species that behave similarly and obtain a smaller, more general model, for which the dynamic behavior can be explored and used to help understand the initial model.

The paper is structured as follows. Section 2 provides an introduction to chemical reaction networks and the main formal results related to uniquely identifiable rate constants and solutions of ordinary differential equations. In Section 3 we formally discuss fit-preserving refinement and prove the main result of this paper. We apply our approach to the study of a simple example, the Brusselator [6], in Section 4, where we analyze four different refinements of the model. We discuss the implications of our result in Section 5.

## 2. Preliminaries

We first fix some notations used throughout the paper. We denote by $\mathbb{N}$ the set of non-negative integers and by $\mathbb{R}_{\geq 0}$ the set of non-negative real numbers. For two sets $X$, $Y$ we denote by $X^Y$ the set of mappings $f : Y \to X$; for a finite set $Y$, $X^Y$ can also be seen as the set of vectors of dimension $|Y|$ with elements from $X$. Throughout this paper we will always denote vectors with a lower-case bold-faced letter. We will use bold-faced upper-case letters to denote matrices or functions that have multiple inputs and outputs.

### 2.1. Solutions of autonomous ordinary differential equations

Let $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. We focus on the solutions of the following autonomous ordinary differential equation (ODE):

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) \tag{1}$$

with the initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x_0}$. Following common practice in the literature, we will also call this an initial value problem (IVP). We can also understand equation (1) as a system of differential equations if we consider the components of $\boldsymbol{x}$ and $\boldsymbol{F}(\boldsymbol{x})$ separately. A solution of such an equation is a function $\boldsymbol{x} : I \to \mathbb{R}^n$, where $I \subseteq \mathbb{R}$ is an interval containing $t_0$, such that $\boldsymbol{x}$ is differentiable and, moreover, $\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(\boldsymbol{x}(t))$, for all $t \in I$. It is well known that such IVPs have a unique solution in the neighborhood of $t_0$, as long as $\boldsymbol{F}$ satisfies some reasonable assumptions, for example see [7]. Since our work involves the careful manipulation of these ideas, we provide (without proof) formal results for both the existence and uniqueness in what follows.

First, let us see that if we are given a solution $\boldsymbol{x}$ of equation (1) with the given initial condition, then the function $\boldsymbol{x}(t - t_0)$ satisfies the same differential equation, with the initial condition specified at 0, i.e. $\boldsymbol{x}(0) = \boldsymbol{x_0}$. Thus, it suffices to only consider initial conditions specified for $t_0 = 0$.

**Theorem 1.** *Let $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Then there exists a closed interval $[-a, a]$ and a unique function $\boldsymbol{x} : [-a, a] \to \mathbb{R}^n$ such that $\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(\boldsymbol{x}(t))$ for all $t \in [-a, a]$ and $\boldsymbol{x}(0) = \boldsymbol{x_0}$.*

For a complete proof of Theorem 1, the reader may see [7]. The uniqueness applies also for extensions of the solution to larger intervals. In particular, we can consider the maximal time domain for which a solution exists and refer to the corresponding solution as *the solution* of the equation for the given initial condition.

**Definition 1.** Let $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. For every $\boldsymbol{\alpha} \in \mathbb{R}^n$, we will use $\boldsymbol{a}[\boldsymbol{\alpha}]$ to refer to the unique, maximal (with respect to its time domain) solution of the differential equation $\dot{\boldsymbol{a}} = \boldsymbol{F}(\boldsymbol{a})$, with $\boldsymbol{a}(0) = \boldsymbol{\alpha}$. We will use $dom(\boldsymbol{a}[\boldsymbol{\alpha}])$ to denote the domain of $\boldsymbol{a}[\boldsymbol{\alpha}]$.

Note that the time domain of a solution $\boldsymbol{a}[\boldsymbol{\alpha}]$ may also depend on the actual value of $\boldsymbol{\alpha}$.

### 2.2. Reaction networks

We consider in this paper that all reactions are irreversible; any reversible reaction is replaced by its "left-to-right" and "right-to-left" irreversible reactions. We formalize in the following the notion of a reaction, using both a rewriting rule style, and a vectorial style.