Occupation measure methods for modelling and analysis of biological hybrid systems

Alexandre Rocca [*,**] **Marcelo Forets** [*] **Victor Magron** [*]
Eric Fanchon [**] **Thao Dang** [*]

[*] *Univ. Grenoble Alpes, CNRS, Grenoble INP, VERIMAG, 700, avenue centrale,38000, Grenoble, France*
[**] *Univ. Grenoble Alpes/CNRS, TIMC-IMAG, UMR 5525, 38041, Grenoble,France* [1]

Abstract: Mechanistic models in biology often involve numerous parameters about which we do not have direct experimental information. The traditional approach is to fit these parameters using extensive numerical simulations (e.g. by the Monte-Carlo method), and eventually revising the model if the predictions do not correspond to the actual measurements. In this work we propose a methodology for hybrid system model revision, when new types of functions are needed to capture time varying parameters. To this end, we formulate a hybrid optimal control problem with intermediate points as successive infinite-dimensional linear programs (LP) on occupation measures. Then, these infinite-dimensional LPs are solved using a hierarchy of semidefinite relaxations. The whole procedure is applied on a recent model for haemoglobin production in erythrocytes.

Keywords: biological modelling, hybrid dynamical system, optimal control problem, semidefinite optimization, occupation measures.

1. INTRODUCTION AND CONTEXT

Context. Mechanistic models in biology generally involve many parameters whose values can be either measured experimentally or inferred from data which provide relationships between parameters and other biological entities. To a large extent, biological mechanisms can be modelled using ordinary differential equations (ODEs), or hybrid dynamical systems with ODEs and discrete switches, applied for example when the system is perturbed or measured during its evolution. A basic concern is to determine the numerical values for the parameters of the ODEs, or more generally a subset of the parameter space, under which the model agrees with the available data. It is common to synthesize parameters using a Monte-Carlo sampling of the parameter space, which is validated then by numerous simulations. When model simulation does not reproduce satisfactorily available experimental data, to a degree which depends on data quality, for any admissible parameter value, the model has to be revised. In this paper we develop a formal approach which does not rely on simulations, to study mechanistic biological models in their experimental context and revise parameters to produce conservative results with respect to experimental data.

Model revision. Consider the optimization problem:

$$\inf_{(\mathbf{x},\,\mathbf{u})} \quad \sum_{j=1}^{n_{exp}} \mathrm{dist}(\mathbf{m}(\mathbf{x}(T_j,\mathbf{u}(T_j))),\mathbf{z}_j) \qquad (1)$$

where $\mathbf{x}$ is a vector of biological variables, such as concentrations, whose dynamics is governed by a biological dynamical system. Time-varying parameters are represented by the input variables $\mathbf{u}$ (modelling biological parameters) such that $\forall t \in [0,T], \mathbf{u}(t) \in \mathbf{U}$. $\mathbf{X}_0$ is the set of initial conditions and the set of pairs $\{(T_j,\mathbf{z}_j)\}_j$ is the set of data points, for $1 \leq j \leq n_{exp}$, in the time frame $[0,T]$. An experimental measurement is a function of the variables $\mathbf{x}$ and is modelled via the function $\mathbf{m}(\mathbf{x})$, e.g. measurement of a kinetic parameter of a biochemical reaction in enzymology. In other words, the typoe of model revision considered here consists of finding time-varying laws of parameter evolution that minimize the error in matching experimental measurements.

Contribution. In this paper we address the model revision problem (1) for the haemoglobin production model from Bouchnita et al. (2016). The framework of our approach is a mathematical formalization of experimental protocols as hybrid dynamical systems, by formulating a particular instance of the optimal control problem with intermediate costs, when the objective function depends on the state variables and control inputs at a given set of time points. This problem is then approximated by multiple hybrid optimal control problems (HOCP) with one final cost. The solution stands on a recently developed method of Zhao et al. (2017) from the field of certified convex optimization to globally solve these HOCP. This method relies on occupation measures and a sequence of *semidefinite relaxations* to produce a sequence of polynomial controls converging to the optimal solution of a HOCP. However, the method described in Zhao et al. (2017) produces piecewise optimal control functions which either may not be biologically realistic or may be difficult to yield coherent and meaningful biological interpretations. Consequently,

[1] Authors emails follow forename.surname@univ-grenoble-alpes.fr

in order to respect realistic constraints on parameters, we use smooth approximations of the generated control input to revise the given model while maintaining good data fitting accuracy.

Related works. An important effort to formalize and validate the parameter synthesis of biological models has been made in works such as Donzé (2010); Mobilia (2015); Dreossi (2016); Beneš et al. (2016), or Rumschinski et al. (2010). Other approaches such as the ones of Cardelli et al. (2017) or Bartocci et al. (2013) design ODE models satisfying sets of temporal constraints.

The hybrid formalism has previously been used as an abstraction method to simplify complex mechanisms which are hard to analyse as seen in Noel et al. (2011); Rocca et al. (2016), or to represent "jump" evolution such as activation processes in gene regulatory networks for example using the stochastic formalism as in Li et al. (2017).

Optimal control theory and variation theory have been applied to biological systems in several works. Most of them address the classical problem of finding a correct input such that the system reaches a desired state. For example, one can control drug input such that a patient attains a healthy state, see Ledzewicz and Schättler (2007) or Caraguel et al. (2016). Another example is the control of some input in population studies as detailed in Bodine et al. (2008). A detailed review on the use of optimal control in systems biology can be found in Lenhart and Workman (2007). The problem of parameter estimation in presence of multiple data, also called data assimilation, is stated in (Lenhart and Workman, 2007, Chapter 26).

The optimal control problem for specific classes of hybrid systems has been investigated in several domains, such as mechanical systems in Pace and Burden (2017) and switched-mode systems in Wardi et al. (2015); Xu and Antsaklis (2004); Bengea and DeCarlo (2005). The work of Pakniyat and Caines (2014) relies on Dynamic Programming and an extension of Pontryagin's Maximum Principle. However, these approaches need a priori knowledge either on the sequence of discrete transitions, or on the number of visited subsystems. To perform optimal control on hybrid systems, we build our work on the techniques from Zhao et al. (2017), which proposes a method to obtain a global solution for hybrid systems with state-dependent transitions, without any a priori knowledge on the execution and the sequence of transitions.

Semidefinite programming (SDP) eases the resolution of hard optimization problems and yields conservative results ensured by positivity certificates. In Lasserre (2001), hierarchies of semidefinite relaxations were introduced for static polynomial optimization. The definition of an infinite-dimensional linear program (LP) over occupation measures, for optimal control problems, was first introduced in Vinter (1993). From this infinite-dimensional LP, Lasserre et al. (2008) defines hierarchies of Linear Matrix Inequalities (LMI) relaxations, to synthesize a sequence of polynomial controls converging to the solutions of the optimal control problem. In Abdalmoaty et al. (2013) the authors propose an extension to piecewise affine systems. Our underlying idea of constructing a suboptimal control with an iterative algorithm is similar to (Abdalmoaty et al., 2013, Section 4). However, we use this scheme to find input functions allowing to reproduce data not only at a final time point but also at intermediate time points.

Organization. In Section 2 we give the necessary background on hybrid systems and the optimal control problem. Section 3 presents our contribution to the resolution of the hybrid optimal control problem with intermediate points. The case study is presented and discussed in Section 4, and conclusions are drawn in Section 5.

2. PRELIMINARIES

We first give the notations and recall the definition of controlled hybrid system and hybrid optimal controlled problem that are used in the sequel.

Notation. Given $\mathbf{x} \in \mathbb{R}^n$, let x_i denote its i-th component. In general, letters in bold font denote multidimensional elements, and normal font unidimensional ones. Let $\mathbb{B} := \{\texttt{true}, \texttt{false}\}$ be the set of Booleans. Let $\mathbb{R}[\mathbf{x}]$ denote the ring of real polynomials in $\mathbf{x} \in \mathbb{R}^n$, and let $\mathbb{R}_d[\mathbf{x}]$ be the subspace of polynomials whose degree is at most d. Let $\mathcal{T}$ be the time interval $[0, T]$, where T is the final time (possibly ∞). Consider the n-dimensional ODE with inputs, $\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t))$, with $\mathbf{f} : \mathcal{T} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ a vector field which is Lipschitz continuous in $\mathbf{x}$ and piecewise continuous in $\mathbf{u}$. Let $\mathbf{X}$ and $\mathbf{U}$ be compact subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Here, $\mathbf{u} : \mathcal{T} \to \mathbf{U}$ is a feasible input function which represents time-varying parameters, or external commands. The tuple $\mathcal{F} := (\mathcal{T}, \mathbf{X}, \mathbf{U}, \mathbf{f})$ defines a continuous dynamical system.

Controlled hybrid systems. Let us recall the definition from Zhao et al. (2017). A controlled hybrid system (CHS) is defined by the tuple: $\mathcal{H} = (\mathcal{I}, \mathcal{E}, \mathbf{X}, \mathbf{U}, \mathrm{F}, \mathcal{S}, \mathcal{R})$ where:

- $\mathcal{I} \subset \mathbb{N}$ is the finite set of mode indices, and n_{modes} the number of modes.
- $\mathcal{E} \subseteq \mathcal{I} \times \mathcal{I}$ is the set of transitions $e = (i, j)$ between two modes: i is the source mode, and j the destination mode.
- $\mathbf{X} := \coprod_{i \in \mathcal{I}} \mathbf{X}_i$ is the disjoint union of domains of $\mathcal{H}$ and $\mathbf{X}_i$ the domain of the mode i. We note that $\mathbf{X}_i$ is a compact subset of $\mathbb{R}^{n_i}$ with n_i the dimension of the mode i. The disjoint union $\coprod$ can simply be considered as a labelling operation on the set of domains by $\mathcal{I}$, that is the set of mode indices.
- $\mathbf{U}$ is the set of input values of $\mathcal{H}$.
- $\mathrm{F} := \{\mathcal{F}_i\}_{i \in \mathcal{I}}$ is the set of continuous dynamical sub-systems associated to each mode. The dynamical system associated to mode i is:
$$\mathcal{F}_i := (\mathcal{T}, \mathbf{X}_i, \mathbf{U}, \mathbf{f}_i),$$
with $\mathbf{f}_i : \mathcal{T} \times \mathbf{X}_i \times \mathbf{U} \to \mathbb{R}^{n_i}$ a vector field polynomial in $\mathbf{x}$ and affine in $\mathbf{u}$.
- $\mathcal{S} := \coprod_{e \in \mathcal{E}} \mathcal{S}_e$ is the disjoint union of guards $\mathcal{S}_e \subseteq \mathbf{X}_i$ associated to each transition $e = (i, j) \in \mathcal{E}$. The guard $\mathcal{S}_{(i,j)}$ defines the switch condition from i to j: for $\mathbf{x} \in \mathbf{X}_i$, if $\mathbf{x} \in \mathcal{S}_{(i,j)}$ then the system at $\mathbf{x}$ can make the transition from mode i to mode j.
- $\mathcal{R} := \{\mathcal{R}_e\}_{e \in \mathcal{E}}$ is the set of reset maps, each reset map $\mathcal{R}_e : \mathcal{S}_e \to \mathbf{X}_j$ being associated to a transition $e := (i, j) \in \mathcal{E}$ and it defines how the continuous variables may change after the discrete transition from mode i to mode j.

Additionally, the CHS must respect a few technical assumptions, ensuring that it is deterministic, see Dang et al. (2017).

Hybrid optimal control problem. Given a CHS $\mathcal{H}$, let $\mathbf{X}_0$, and $\mathbf{X}_T$, be the initial set and target sets defined by $\mathbf{X}_0 := \coprod_{i\in\mathcal{I}} \mathbf{X}_{0,i}$ and $\mathbf{X}_T := \coprod_{i\in\mathcal{I}} \mathbf{X}_{T,i}$, where $\mathbf{X}_{0,i}$ and $\mathbf{X}_{T,i}$ are compact subsets of $\mathbf{X}_i$ for each mode $i \in \mathcal{I}$. Let i_0 and i_T be the initial mode and the final mode at time T, respectively. Then, given $(i_0, \mathbf{x}(0)) = (i_0, \mathbf{x}_0) \in \mathbf{X}_0$ and $\mathbf{u} : \mathcal{T} \to \mathbf{U}$ an input function, we say that for $T > 0$, $(\mathbf{x}(t), \mathbf{u}(t)) \in \mathcal{P}$ is an admissible pair on $\mathcal{T}$ and $\mathcal{P}$ is the set of admissible pairs, if $(i, \mathbf{x}(t)) \in \mathbf{X}$ is a trajectory of $\mathcal{H}$ (Zhao et al., 2017, Algorithm 1) and $(i_T, \mathbf{x}(T)) \in \mathbf{X}_T$. The HOCP is defined by:

$$J^*_{hocp} := \inf_{(\mathbf{x},\mathbf{u})\in\mathcal{P}} \int_0^T h_{\lambda(\mathbf{x}(t))}\left(t, \mathbf{x}_{\lambda(\mathbf{x}(t))}(t), \mathbf{u}(t)\right) dt + H_{\lambda(\mathbf{x}(T))}\left(\mathbf{x}_{\lambda(\mathbf{x}(T))}(T)\right). \quad (2)$$

where $\{h_i : [0,T] \times \mathbb{R}^{n_i} \times \mathbb{R}^m \to \mathbb{R}\}_{i\in\mathcal{I}}$ and $\{H_i : \mathbb{R}^{n_i} \to \mathbb{R}\}_{i\in\mathcal{I}}$ are measurable functions, and $\lambda(\mathbf{x}(t))$ is the function which associates to an instantaneous state $\mathbf{x}(t)$ its corresponding mode.

3. OPTIMAL CONTROL FOR MODEL REVISION

We propose in this section a method to synthesize time-varying parameters reproducing experimental results of a multiple-phase protocol modelled by a hybrid system. Indeed, it is crucial to approach model revision taking into account the biological system in the evolving environment of the complete protocol.

Problem statement. We begin by stating an optimal control problem where $\mathbf{u}(t)$ are input functions which minimize the distance of the results produced by the model and given experimental data points. Measurements modelled by a function $\mathbf{m}(\mathbf{x}(t))$, are performed at times T_j, $1 \le j \le n_{exp}$. Let $\mathbf{z}_j$ be the data point at time T_j, and n_{exp} be the number of data points. Let $\mathbf{X}_{T_j,i}$ be compact subsets of $\mathbf{X}_i$, and $\mathbf{X}_{T_j} := \coprod_{i\in\mathcal{I}} \mathbf{X}_{T_j,i}$. As in (2), let $(i_0, \mathbf{x}(0)) \in \mathbf{X}_0$, and suppose that we are given a set of time values $\{T_j\}$, with $1 \le j \le n_{exp}$, and $T_{n_{exp}} = T$. Given an input function $\mathbf{u} : \mathcal{T} \to \mathbf{U}$, we say that $(\mathbf{x}, \mathbf{u}) \in \mathcal{P}_{int}$ is an admissible pair for a problem with intermediate points and $\mathcal{P}_{int}$ the associated set of admissible pairs, if $(i(t), \mathbf{x}(t)) \in \mathbf{X}$ is a CHS trajectory, and $(i_{T_j}, \mathbf{x}(T_j)) \in \mathbf{X}_{T_j}$ for all j. Let $H(\mathbf{x}(T_j))$ be a cost at time T_j, and $h(t, \mathbf{x}(t), \mathbf{u}(t))$ a running cost for the whole $[0, T]$ interval. The model revision problem is the optimal control problem with intermediate cost for the CHS $\mathcal{H}$:

$$J^* := \inf_{(\mathbf{x},\mathbf{u})\in\mathcal{P}_{int}} \int_0^T h(t, \mathbf{x}(t), \mathbf{u}(t))dt + \sum_{j=0}^{n_{exp}} H\left(\mathbf{x}(T_j)\right) \quad (3)$$

In this paper, $H\left(\mathbf{x}(T_j)\right) = ||\mathbf{m}(\mathbf{x}(T_j)) - \mathbf{z}_j||_2^2$ and $h(t, \mathbf{x}(t), \mathbf{u}(t))$ model additional constraints on the control or temporal properties. To our best knowledge there is no method to efficiently address directly problem (3) which is a generalization of (2). Consequently, we search for an admissible solution using a *greedy approach*: we cut problem (3) into a sequence of problems (2) that can be solved using the method from Zhao et al. (2017). A good trade-off is obtained between computational cost, optimality and flexibility with this solution scheme. Moreover, this method does not constrain the form of the sought parameters which eases the modelling and biological interpretation. For each $1 \le j \le n_{exp}$ the HOCP subproblem is:

$$J^*_j := \inf_{(\mathbf{x}^{(j)}, \tilde{\mathbf{u}}^{(j)})} \int_{T_{j-1}}^{T_j} h(t, \mathbf{x}^{(j)}, \tilde{\mathbf{u}}^{(j)})dt + H(\mathbf{x}^{(j)}(T_j)), \quad (4)$$

with $(i^{(j)}(t), \mathbf{x}^{(j)}(t))$ a trajectory of a CHS $\mathcal{H}$ on the interval $\mathcal{T}_j := [T_{j-1}, T_j]$, and similarly $\tilde{\mathbf{u}}^{(j)}(t)$ the control on $\mathcal{T}_j$. We let $T_0 = 0$ and $T_{n_{exp}} = T$. We note that if a transition $i \to i'$ occurs at time T_j of the interval $[T_{j-1}, T_j]$, we retain only the left part in the mode i for the next optimization on the interval $[T_j, T_{j+1}]$. Let $\tilde{\mathbf{u}}(t)$ and $(i(t), \mathbf{x}(t))$ be respectively the control and the trajectory, for $t \in [0, T]$. They are respectively defined by the concatenation of all the controls $\tilde{\mathbf{u}}^{(j)}(t)$ and the trajectories $(i^{(j)}(t), \mathbf{x}^{(j)}(t))$ on the sub-intervals $[T_{j-1}, T_j]$. By construction, $(\mathbf{x}(t), \tilde{\mathbf{u}}(t))$ is an admissible pair for (3), as $(i_{T_j}, \mathbf{x}(T_j)) = (i^{(j)}_{T_j}, \mathbf{x}^{(j)}(T_j)) \in X_{T_j}$.

Remark 1. We emphasize that $(\mathbf{x}(t), \tilde{\mathbf{u}}(t))$ is not necessary an optimal solution for (3). Moreover, as the optimization problem (4) is obtained through a greedy scheme, we have no guarantee that its optimal cost J^*_j is inferior to a given ε. However, as we want to equally fit all the data points, searching iteratively for the control gives a satisfactory solution.

Algorithm. Algorithm 1 finds an admissible solution to (3), by solving a sequence of the HOCP (4), and it is decomposed in three steps. The first step is the procedure `HOCP`, associated to the HOCP (2) for a given pair $(T_j, \mathbf{z}_j)$. Given a starting relaxation degree r, and a relaxation order $d_r \ge r$, we solve the relaxed primal defined in (Zhao et al., 2017, Section 5.1). As described in Zhao et al. (2017), we obtain $M_{d_r}(\mathbf{y}_{\mu_i})$, the sequence moment matrices of degree d_r, associated to the occupation measure μ_i of each mode $i \in I$. We also obtain $\underline{J}_j^{(d_r)}$ an under approximation of the optimum of (4). The second step is the procedure `Synth`, which returns the admissible control $\tilde{\mathbf{u}}^{(j)}(t, \mathbf{x})$ of degree $d_u \le d_r$ using a truncated moment matrix $M_{d_u}(\mathbf{y}_{\mu_i})$ of $M_{d_r}(\mathbf{y}_{\mu_i})$ at the reduced degree d_u. The third and last step is the procedure `Simu`. It performs the validation that the synthesized control $\tilde{\mathbf{u}}^{(j)}$ yields $||\mathbf{m}(\mathbf{x}(T_j)) - \mathbf{z}_j||_2^2 \le \varepsilon$, where ε is just a stopping criterion. This last step is done with numerical simulations using an ODE solver with discrete events. If $||\mathbf{m}(\mathbf{x}(T_j)) - \mathbf{z}_j||_2^2 \le \varepsilon$, then $(i_f, \mathbf{x}^{(j)}(T_j))$ reached at $t = T_j$ give the initial conditions for the next iteration $j + 1$. Otherwise, `Ctrl_Synth` and `Simulate` are repeated while increasing the degree of the control until $d_u = d_r$. If $||\mathbf{m}(\mathbf{x}(T_j)) - \mathbf{z}_j||_2^2 \le \varepsilon$ is still not satisfied, the relaxation order d_r is increased, and the three steps are repeated. If $\varepsilon \le \underline{J}_j^{(d_r)}$ then for the considered pair (i_0, x_0), there is no control such that $||\mathbf{m}(\mathbf{x}(T_j)) - \mathbf{z}_j||_2^2 \le \varepsilon$. Consequently, we keep our previous result $\tilde{\mathbf{u}}^{(j)}$ and $(i_f, \mathbf{x}^{(j)}(T_j))$ is the initial condition for iteration $j + 1$.

4. CASE STUDY: HAEMOGLOBIN PRODUCTION

In this section, using the method developed in Section 3, we revise the model of haemoglobin production by finding a better fit for the time-varying parameter noted k_3 in

Algorithm 1.

procedure ALGORITHM 1($\mathcal{H}, \{(T_j, \mathbf{z}_j)\}_j, i_0, \mathbf{x}_0, \varepsilon, r$)
  $T_{init} = 0$
  for all experimental data $(T_j, \mathbf{z}_j)$ **do**
    $d_u = \mathbf{0}, d_r = r, \texttt{err} = +\infty$
    while $\texttt{err} \geq \varepsilon \wedge \underline{J}_j^{(d_r)} \leq \varepsilon$ **do**
      $\underline{J}_j^{(d_r)}, \mathbf{M}_{d_r}(\mathbf{y}_\mu) = \texttt{HOCP}(\mathcal{H}, i_0, ...$
        $...\mathbf{x}_0, T_{init}, T_j, \mathbf{z}_j, d_r)$
      while $\texttt{err} \geq \varepsilon$ **and** $d_u \leq d_r$ **do**
        $\tilde{\mathbf{u}}^{(j)}(\mathbf{x}(t), t) = \texttt{Synth}(\mathbf{M}_{d_r}(\mathbf{y}_\mu), d_u)$
        $(i_f, \mathbf{x}^{(j)}(t)) = \texttt{Simu}(\mathcal{H}, \tilde{\mathbf{u}}^{(j)}(\mathbf{x}(t), t), ...$
          $... \ i_0, \mathbf{x}_0, T_{init}, T_j)$
        $\texttt{err} = H(x^{(j)}(T_j), \mathbf{z}_j)$
        $\texttt{increase}\ d_u$
      end while
      $\texttt{increase}\ d_r$
    end while
    $i_0 = i_f$
    $\mathbf{x}_0 = \mathbf{x}^{(j)}(T_j)$
    $T_{init} = T_j$
  end for
end procedure

Bouchnita et al. (2016), with respect to the same error function.

The haemoglobin production model. Erythrocytes (also named red blood cells) are produced inside the bone marrow. In this place, they go through multiple differentiation stages from stem cells (also called hemocytoblasts in this context) into erythroblasts and finally erythrocyte. This differentiation process is also called erythropoiesis. During its differentiation, an erythroblast produces haemoglobin. At the final stages, the erythroblast forces out its nucleus and is released in the circulating blood. The haemoglobin stored in the erythrocyte will play the role of oxygen transport protein. Without entering into details, the haemoglobin Hb is constituted of 8 subcomponents: 4 hemes (H) and 4 globins (G). The heme contains iron Fe and its production directly depends on the iron input into the cell. The globin production is regulated by the heme. The experimental results considered in this section are the same as the ones previously fitted in Bouchnita et al. (2016), and are taken from the work of Koury and Bondurant (1988). The experiments aim to measure the rate of haemoglobin production at different steps of the erythropoiesis.

However, haemogolobin production is not directly measured, and the experiments probe it indirectly through the integration rate of radiolabelled iron ^{59}Fe in heme (free in the cell or integrated in the haemoglobin). Consequently, we only observe the quantity of radiolabelled heme $^{59}\text{H} + 4\,^{59}\text{Hb}$. Additionally, the measurements are performed on a subset of the differentiating cells and 3 hours after the injection of radiolabels. Overall, this is a multi-phase experiment that can be described with a hybrid system with modes [2] that correspond to the evolution of the control batch of cells without radiolabels and modes representing the evolution after injection of the

[2] Due to limitations, details on the transitions are given Dang et al. (2017)

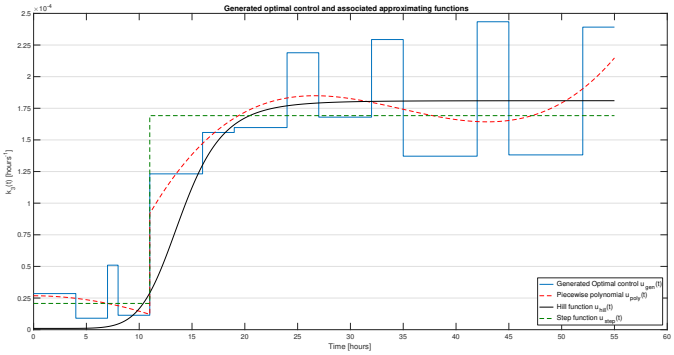

Fig. 1. Synthesized optimal control (blue) and various approximations that yield a realistic interpretation.

radiolabels. Here, we search a better fit of the time-varying parameter $k_3(t)$ modelling the production mechanism of heme. Consequently, this parameter is directly linked to the observations: the iron integration rate in heme and haemoglobin.

Scaling and balancing. For numerical reasons, it is necessary to scale the parameters and state variables, making it easier for the solver to succeed in solving the relaxed problem. Similarly, to facilitate the numerical optimization we rewrite the control variable $u(t) \in \mathbf{U} = [0, 1]$ as $u(t) = \zeta k_3(t)$, with $\zeta \ll 1$ and $k_3(t) \in [0, 1/\zeta]$. While the scale factor ζ may take different values depending on the numerical optimization details, the objective control $u(t)$ always evolves in $[0, 1]$. We solve the optimal control problem with intermediate time points defined in (3), using the method from Section 4. The experimental measurement is $m(\mathbf{x}) := {}^{59}\text{H} + 4\,{}^{59}\text{Hb}$. Thus, we set $H(\mathbf{x}(T_j)) := ({}^{59}\text{H}(T_j) + 4\,{}^{59}\text{Hb}(T_j) - z_j)^2$, as we search to minimize the total residual error term:

$$\varepsilon_{total} = \sum_{1 \leq j \leq n_{exp}} \frac{\sqrt{H(\mathbf{x}(T_j))}}{\sum_{1 \leq j \leq n_{exp}} z_j}, \tag{5}$$

where $\mathbf{x}$ is the vector of biological concentrations. The original experimental data points (T_j, z_j) are given in Table 1. Here, the input control $k_3(t)$ models some hidden

Time (h)	7	11	19	27	35	45	55
Measure ($\frac{cpm}{1e^{-7}\text{L}\cdot h^{-1}}$)	16	85	348	391	399	481	395

Table 1. Experimental data points (T_j, z_j) used as references.

mechanism which evolves with the differentiation of the cells. It should be the same function of time for both the control and the radioactive cell batch.

However, as the control generated by Algorithm 1 is piecewise, with 1 continuous piece for each mode, and the fact that our data are on the radioactive species only, the solution of the optimization problem with only a final cost $H(\mathbf{x}(T_i))$ is not *balanced*, having a much stronger control in the modes where the radioactive species are evolving. A workaround for the balancing problem is the following. We add a small penalization cost $c_i^1(t) = (0.01u(t))^2$ to equilibrate the control when i corresponds to a mode with radioactive species, otherwise $c_i^1(t) = 0$. In a similar vein, we add another penalization cost $c_i^2(t) = (u(T_j) - u(t))^2$ to avoid when the control strongly varies between

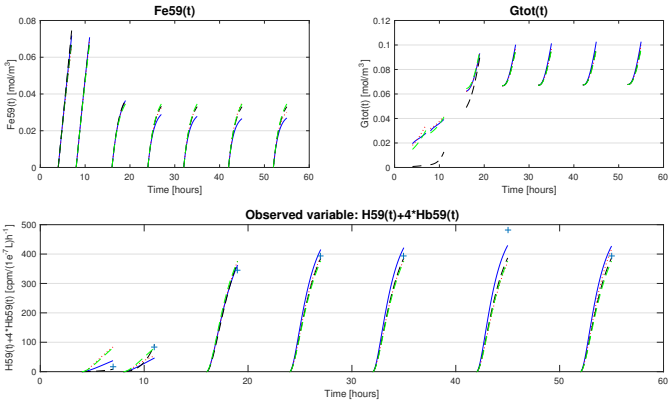

Fig. 2. Top: Radioactive quantities 59Fe, Gtot. Bottom: comparison of the measurement function results (^{59}H + 4 ^{59}Hb)to the data of Table 1 (crosses).

two iterations j on the interval $[T_{j-1}, T_j]$ and $j+1$ on $[T_j, T_{j+1}]$ (with the exception of the first iteration). This leads to $h_i(t, \mathbf{x}(t), u(t)) = c_i^1(t) + c_i^2(t)$. Let us note that, even if these additional costs can eventually degrade the accuracy of the data fitting, we gain in terms of biological interpretation of the resulting traces.

Implementation. Finally, by partitioning the computation in the time domain, we can greatly reduce the computational cost at each iteration. More technically, since the transitions of the hybrid system $\mathcal{H}$ are fully determined by the time t, we can pre-compute the function $\lambda(\mathbf{x}(t))$ from Section 2. Thus, each iteration j of Algorithm 1 can be restrained to the sub-hybrid system $\mathcal{H}^j$ of $\mathcal{H}$, constituted by the modes visited in the interval $[T_{j-1}, T_j]$. For numerical implementation, the problem on measures is formulated in SPOTLESS [3], and then we extract the primal solution provided by a primal-dual SDP solver. To do so, we use the implementation from Zhao et al. (2017) to generate the dual problem. We used the SDP solver MOSEK v.7.1 (Andersen and Andersen, 2000), in MATLAB v.9.0 (R2016a). Results are obtained with a processor i7-5600U CPU (2.60 GHz) with 16Gb of RAM on Debian 8. We only solve the problem for a relaxation order $r = 4$, as any higher order would be too memory expensive. Here, we did not impose any constraint on ε. Using this configuration, the total time taken by Algorithm 1 is 2107s, with 1700s spent in the `HOCP` procedure, and 390s in the `Synth` procedure. On Figure 1, the control generated by Algorithm 1 is shown in blue. This control is piecewise, and clearly divided in two phases: before and after t equals 11 hours. However, the control synthesized is still difficult to interpret as a biological phenomenon. Consequently, we propose and analyze three fits of this control by using functions closer to biological knowledge. In Figure 2, we show a graphical representation of how closely each function can control the model to reach the desired data points.

Results without fitting. In a simulation-based approach, we have to propose a template function to fit the data, e.g. a polynomial of given degree, for the desired time-varying parameter. If we want to fit a polynomial of higher degree, the simulations have to be run again multiple times. On the contrary, the proposed approach returns a control signal, and since the fit to data points is performed a posteriori, there is no additional computation cost in refining the model. In this case study, from the form of the experimental data points, a usual hypothesis is that $k_3(t)$ should be similar to a jump function, with a low value for the two first points, and a higher one for the following ones. However, even with such information a good fit is not easily achieved with simulations. The control generated with Algorithm 1 returns the expected "jump" behaviour for $k_3(t)$, and the total residual error is 9.59% which is much lower than the 22.8% from Bouchnita et al. (2016).

Results with fitting. We first fit a step function to the generated control, with a change at $t = 11$. The associated error of 12.24% is still lower than Bouchnita et al. (2016), yet being higher than the generated control mainly due to the second-to-last point. We also fit a piecewise polynomial function in two pieces. The first piece, for $t \in [0, 11]$, is a polynomial of degree 2 while the degree of the second, for $t \in [11, 55]$, is 4. This proposed input control allows to reproduce more accurately, than the step function, the third data point. However, the total error is 13%, being overall the worst of the proposed fits. Lastly, we fit a Hill function, a function used to model the kinetics of a class of biochemical reactions and which is a very common way to represent biological activation processes. The associated total error is 7.5%, which is the lowest, taking advantages from both the step function and the piecewise polynomial function. In this case, the inaccuracy also mainly comes from the second-to-last point, which is quite separated from trend of the other experimental points, and may be due to some experimental problems (no standard deviation results were available). Without taking this point into consideration for the error computation the error falls to 3% for the Hill function fit.

Discussion. On this particular example, the generated control is accurate, and computed in a reasonable time (∼35min), even for a large hybrid system of 14 modes with at most 9 continuous variables. Using some fitting functions afterwards, it is even possible to refine the results and also obtain a biologically meaningful interpretation for the desired time-varying parameters.
The presence of a Hill function (jump behaviour) in the evolution of the k_3 parameter suggests that the consumption of iron for heme synthesis is regulated, and triggered at a given point in the differentiation process. The evolution law of k_3 is thus a way to account for a regulation process not explicitly described in this model.

5. CONCLUSION

In this work, we have addressed an important problem arising in biological modelling: model revision. We propose a method for revising an experiment modelled by a hybrid system, given a set of experimental data points. The method scales even on large hybrid systems such as the haemoglobin production model, while providing an accurate result, and a meaningful interpretation, as an activation process, for the mechanism underlying the revised parameter. The CHS formalism is motivated by the development of an automatic, and formal modelling of

[3] https://github.com/spot-toolbox/spotless

multiple-step experimental protocols, and to develop new methods for their analysis. Such formal representations had already been used as alternative, non-ambiguous languages, in contrast with the natural language, for the description of experiments Soldatova et al. (2008). However, those works do not consider an underlying mechanistic model in the form of ODEs. In future work we plan to investigate, using the CHS formalism, two other relevant problems in biological systems modelling: finding valid subsets of the parameters space fitting multiple data points (as an extension Shia et al. (2014)), and the validation of biological experiments. The focus of this work has been on ODEs, but multicellular systems and transport processes are described by partial differential equation (PDE) models. The extension of semidefinite programming techniques to PDEs (Mevissen et al., 2011) and their application to biological models would require further theoretical and numerical developments.

ACKNOWLEDGEMENTS

This work is partially supported by the ANR CADMIDIA project (grant ANR-13-CESA-0008-03) and the ANR MALTHY project (grant ANR-12-INSE-003)

REFERENCES

Abdalmoaty, M.R., Henrion, D., and Rodrigues, L. (2013). Measures and lmis for optimal control of piecewise-affine systems. In *ECC*.

Andersen, E.D. and Andersen, K.D. (2000). The Mosek Interior Point Optimizer for Linear Programming: An Implementation of the Homogeneous Algorithm.

Bartocci, E., Bortolussi, L., and Nenzi, L. (2013). A temporal logic approach to modular design of synthetic biological circuits. In *CMSB*.

Beneš, N., Brim, L., Demko, M., Pastva, S., and Šafránek, D. (2016). Parallel smt-based parameter synthesis with application to piecewise multi-affine systems. In *ATVA*.

Bengea, S.C. and DeCarlo, R.A. (2005). Optimal control of switching systems. *automatica*.

Bodine, E.N., Gross, L.J., and Lenhart, S. (2008). Optimal control applied to a model for species augmentation. *Mathematical biosciences and engineering*.

Bouchnita, A., Rocca, A., Fanchon, E., Koury, M., Moulis, J., and Volpert, V. (2016). Multi-scale modelling of erythropoiesis and hemoglobin production. *JIOPM*.

Caraguel, F., Lesart, A.C., Estève, F., van der Sanden, B., and Stéphanou, A. (2016). Towards the design of a patient-specific virtual tumour. *Computational and mathematical methods in medicine*.

Cardelli, L., Češka, M., Fränzle, M., Kwiatkowska, M., Laurenti, L., Paoletti, N., and Whitby, M. (2017). Syntax-guided optimal synthesis for chemical reaction networks. In *CAV*.

Dang, T., Fanchon, E., Forets, M., Magron, V., and Rocca, A. (2017). Occupation measure methods for modelling and analysis of biological hybrid automata. *arXiv preprint arXiv:1710.03158*.

Donzé, A. (2010). Breach, a toolbox for verification and parameter synthesis of hybrid systems. In *CAV*.

Dreossi, T. (2016). Sapo: Reachability computation and parameter synthesis of polynomial dynamical systems. *arXiv*.

Koury, M.J. and Bondurant, M.C. (1988). Maintenance by erythropoietin of viability and maturation of murine erythroid precursor cells. *Journal of cellular physiology*.

Lasserre, J.B. (2001). Global optimization with polynomials and the problem of moments. *SIOPT*.

Lasserre, J.B., Henrion, D., Prieur, C., and Trélat, E. (2008). Nonlinear optimal control via occupation measures and lmi-relaxations. *SICON*.

Ledzewicz, U. and Schättler, H. (2007). Optimal controls for a model with pharmacokinetics maximizing bone marrow in cancer chemotherapy. *Mathematical biosciences*.

Lenhart, S. and Workman, J.T. (2007). *Optimal control applied to biological models*. CRC Press.

Li, X., Omotere, O., Qian, L., and Dougherty, E.R. (2017). Review of stochastic hybrid systems with applications in biological systems modeling and analysis. *EURASIP Journal on Bioinformatics and Systems Biology*.

Mevissen, M., Lasserre, J.B., and Henrion, D. (2011). Moment and SDP relaxation techniques for smooth approximations of problems involving nonlinear differential equations. *IFAC*.

Mobilia, N. (2015). *Méthodologie semi-formelle pour l'étude de systèmes biologiques: application à l'homéostasie du fer*. Ph.D. thesis, UGA.

Noel, V., Vakulenko, S., and Radulescu, O. (2011). Algorithm for identification of piecewise smooth hybrid systems: application to eukaryotic cell cycle regulation. In *WABI*.

Pace, A.M. and Burden, S.A. (2017). Piecewise-differentiable trajectory outcomes in mechanical systems subject to unilateral constraints. In *HSCC*.

Pakniyat, A. and Caines, P.E. (2014). On the relation between the minimum principle and dynamic programming for hybrid systems. In *CDC*.

Rocca, A., Dang, T., Fanchon, E., and Moulis, J.M. (2016). Application of the reachability analysis for the iron homeostasis study. In *HSB*.

Rumschinski, P., Borchers, S., Bosio, S., Weismantel, R., and Findeisen, R. (2010). Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC systems biology*.

Shia, V., Vasudevan, R., Bajcsy, R., and Tedrake, R. (2014). Convex computation of the reachable set for controlled polynomial hybrid systems. In *CDC*.

Soldatova, L.N., Aubrey, W., King, R.D., and Clare, A. (2008). The exact description of biomedical protocols. *Bioinformatics*.

Vinter, R. (1993). Convex duality and nonlinear optimal control. *SICON*.

Wardi, Y., Egerstedt, M., and Hale, M. (2015). Switched-mode systems: gradient-descent algorithms with armijo step sizes. *Discrete Event Dynamic Systems*.

Xu, X. and Antsaklis, P.J. (2004). Optimal control of switched systems based on parameterization of the switching instants. *IEEE trans. on automatic control*.

Zhao, P., Mohan, S., and Vasudevan, R. (2017). Optimal control for nonlinear hybrid systems via convex relaxations. *arXiv preprint arXiv:1702.04310*.