Optimal control strategies in a generic class of bacterial growth models with multiple substrates (extended version) ⋆

Agustín G. Yabo [a]

[a] *MISTEA, Université Montpellier, INRAE, Institut Agro, Montpellier, France*

Abstract

Optimal control strategies are studied through the application of the Pontryagin's Maximum Principle for a class of non-linear differential systems that are commonly used to describe resource allocation during bacterial growth. The approach is inspired by the optimality of numerous regulatory mechanisms in bacterial cells. In this context, we aim to predict natural feedback loops as optimal control solutions so as to gain insight on the behavior of microorganisms from a control-theoretical perspective. The problem is posed in terms of a control function $u_0(t)$ representing the fraction of the cell dedicated to protein synthesis, and n additional controls $u_i(t)$ modeling the fraction of the cell responsible for the consumption of the available nutrient sources in the medium. By studying the necessary conditions for optimality, it is possible to prove that the solutions follow a bang-singular-bang structure, and that they are characterized by a sequential uptake pattern known as diauxic growth, which prioritizes the consumption of richer substrates over poor nutrients. Numerical simulations obtained through an optimal control solver confirm the theoretical results. Finally, we provide an application to batch cultivation of E. coli growing on glucose and lactose. For that, we propose a state feedback law that is based on the optimal control, and we calibrate the obtained closed-loop model to experimental data.

Key words: optimal control; bacterial growth models; bacterial resource allocation; diauxic growth

1 Introduction

While most of the research advancements on control and systems engineering have been focused on the development and implementation of feedback loops, control theory has also contributed substantially to our understanding of the underlying regulatory mechanisms in living organisms. For instance, at the cellular level, numerous phenomena are known to behave in a closed-loop manner, by sensing—and reacting to—changes in the environment. One of the most common ideas in cell biology is that these regulatory mechanisms are a result of the optimizing force of the natural selection, which allows living beings to survive and outgrow competing species. Under this hypothesis, optimization and optimal control theory become instrumental in elucidating the governing principles of these natural mechanisms.

Bacterial cells are constantly confronted with the problem of allocating resources to different cellular functions, such as the uptake and conversion of nutrients from the environment into building blocks (metabolism), the production of proteins from these building blocks (gene expression), and the detection of—and reaction to—environmental changes. Under the assumption that bacteria have evolved internal regulatory mechanisms to maximize growth rate [1], theoretical studies have been able to predict these natural resource allocation strategies from simple mathematical models using optimal control theory. Numerous examples can be found in the literature [2,3,4,5]. For instance, the regulatory action of the ppGpp molecule [6]—known to play a key role in growth rate control—has been compared to optimal control strategies obtained through simple bacterial growth models [2]. Another example is a cellular mechanism called CCR (Carbon Catabolite Repression), which plays a key role in how substitutable nutrients (i.e. nutrients that do not need the presence of other substrates to allow bacterial growth) are consumed from the medium. In particular, when metabolizing nutrients sequentially, the growth pattern is called *diauxic growth*.

As many empiric phenomena observed in bacteria, diauxic growth is a behavior that can potentially be predicted through optimization, provided that the adequate objective function is chosen. Previous studies

⋆ This work was partially supported by ANR project Maximic (ANR-17-CE40-0024-01) and Labex SIGNALIFE (ANR-11-LABX-0028-01).

Email address: agustin.yabo@inrae.fr (Agustín G. Yabo).

have attempted so by using simple mathematical models and numerical optimal control [7,8,9]. A more control-theoretical point of view is adopted in [10], where authors used optimal control theory to elucidate feedback control strategies of substrate uptake. However, only the case with two nutrient sources is studied, and no intermediate quantities are considered: the substrates are directly transformed into the final product, and thus the interplay between substrate uptake and protein synthesis is not captured in this simpler formulation.

The study of the feedback loops that arise in nature yields very interesting theoretical problems, that have the potential to inspire novel control strategies for non-biological fields of research and engineering. In this paper, we consider a generalized non-linear mathematical model of a population growing on n substitutable sources s_i. The model captures two natural regulation mechanisms of unicellular organisms: 1) the distribution of resources for the uptake of multiple substrates, and 2) the trade-off between metabolism and gene expression (i.e. consuming nutrients and growing). The latter bioregulation is modeled through a control function $u_0(t)$ representing the fraction of the cell dedicated to protein synthesis, as a generalization of previous bacterial growth models [2,11]. Additionally, n uptake controls $u_i(t)$ model the fractions of the cell responsible for nutrient uptake, which can produce sequential or simultaneous substrate consumption—which is an extension of the preliminary results presented in [12]. The problem of finding the optimal control functions $(u_0, u_1, \dots, u_n)$ that comply with the natural objective of maximizing biomass is written as an OCP (Optimal Control Problem), and analyzed using PMP (Pontryagin's Maximum Principle). A thorough analysis of the problem reveals that the optimal allocation strategies behind the studied regulatory mechanisms have bang-singular-bang structures, and can exhibit sequential substrate uptake patterns depending on the concentration and yield of the nutrients in the environment. Using the optimality principles obtained through PMP, we propose a control law that can be written as a function of the state of the system (i.e. feedback control) while retaining the structure of the optimal control. The approach is able to predict diauxic growth as the optimal strategy when facing multiple sources, which represents a control-theoretical argument supporting the idea that diauxie is a naturally-evolved regulation system that maximizes growth rate. Additionally, we provide numerical simulations performed with an optimal control solver verifying the obtained analytical results. Finally, a practical example is presented: the model is calibrated to account for batch processing of E. coli growing on glucose and lactose [13], and simulations with the obtained closed-loop control law show that the approach not only qualitatively predicts the diauxic growth phenomenon, but also is able to match real bacterial growth experiments.

The paper is organized as follows: in Section 2, the main biological principles and constraints are explained, and the general mathematical model is presented. In Section 3, the OCP is formulated, and the structure of the optimal control strategies is studied using PMP. In Section 4, we show the trajectories obtained with an optimal control solver to validate the theoretical results. Finally, in Section 5, we introduce the feedback control law, and we compare the closed-loop model with experimental data from [13].

2 Model definition

We introduce a model representing a bacterial population growing on n substitutable sources. Bacterial cells consume the substrates $\mathrm{s}_1, \mathrm{s}_2, \dots, \mathrm{s}_n$ from the medium and transform them into intermediate metabolites m with associated yield coefficients $Y_1, Y_2, \dots, Y_n$. Essentially, the coefficient Y_i describes the units of intermediate metabolite produced per unit of substrate s_i. The intermediate metabolites are a generalization of the compounds used to produce biomass in cells (that could represent aminoacids in bacteria, or cell quota in phytoplankton [14]). The uptake of the source s_i occurs at a rate w_i, and the synthesis rate of biomass is v_R. The mathematical model describes the time-evolution of the concentration of the i-th substrate $s_i(t)$, the concentration of intermediate metabolites $m(t)$, and the volume of the cell population $x(t)$. The states are defined as non-dimensional to simplify the computations. The dynamical system can be written as

$$\begin{cases} \dot{s}_i = -w_i(s_i)x, \qquad i = 1, 2, \dots, n \\ \dot{m} = \sum_{i=1}^{n} Y_i w_i(s_i) - w_R(m)(m+1), \\ \dot{x} = w_R(m)x, \end{cases}$$

where $w_R(m)$ corresponds to the growth rate of the bacterial population. In this formulation, the dynamics of the intermediate metabolites m are characterized by an "outflow" term $-w_R(m)$ describing the rate of protein synthesis, and a dilution term $-w_R(m)m$ resulting from the increasing bacterial volume. The yield coefficients are bounded to $Y_i \in (0, 1]$, and functions w_i and w_R are subject to the following hypotheses producing the non-linearity of the system.

Assumption 1 *Functions $w_R(x) : \mathbb{R}_+ \to \mathbb{R}_+$ and $w_i(x) : \mathbb{R}_+ \to \mathbb{R}_+$ for $i = 1, 2, \dots, n$ are*

- *Continuously differentiable,*
- *Null at the origin: $w(0) = 0$,*
- *Strictly monotonically increasing: $w'(x) > 0, \forall x \geq 0$,*
- *Strictly concave downwards: $w''(x) < 0, \forall x \geq 0$,*
- *Upper bounded: $\lim_{x\to\infty} w(x) = k$.*

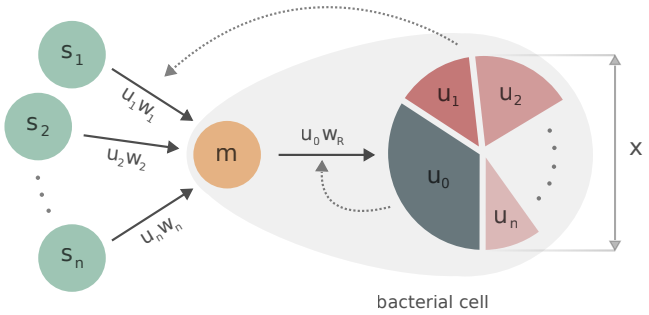

Fig. 1. Scheme of the controlled system. Each external substrate s_i is metabolized into intermediate metabolites m at rate $u_i w_i$. Then, the intermediate metabolites are used to synthesize proteins at rate $u_0 w_R$. The control u_0 represents the fraction of the cell responsible for the synthesis of biomass (lower dotted line), while the remaining $1-u_0 = u_1+\cdots+u_n$ represents the cellular resources dedicated to the consumption of sources and production of m (upper dotted line).

2.1 Controlled dynamics

The first degree of freedom of the control problem is the balance between the resources used to produce the intermediate compound m from the sources s_i, and the resources used to produce biomass x from m. In the context of cell biology, this question represents the trade-off between metabolism and gene expression. We define $u_0(t) \in [0,1]$ as the control variable representing the fraction of the cell responsible for the production of proteins (i.e. biomass x) from the compounds m. Complementary, we define n control functions $(u_1,\dots,u_n)$ satisfying $u_i(t) \in [0,1]$, each one representing the fraction of the cell assigned to the uptake and metabolization of the i-th substrate s_i, and the synthesis of intermediate metabolites m. In bacterial cells, these tasks are performed by proteins of the gene expression machinery (i.e. ribosomes) and proteins of the metabolic machinery (i.e. enzymes), respectively. Following the literature on bacterial growth laws [15,16], we assume that the consumption rate of the i-th substrate uptake is linear in $u_i(t)$, while the rate of protein production is linear in $u_0(t)$. Figure 1 shows a schematic representation of the process here described. In this approach, cells are able to instantaneously stall growth ($u_0 = 0$) or substrate consumption ($u_0 = 1$), as well as to switch to balanced strategies between these two tasks ($0 < u_0 < 1$). While this assumption is not realistic from a physical point of view, it can provide gold-standard control strategies that can be then compared to the biologically feasible ones. The control functions are subject to an inequality constraint modeling the availability limitation of cellular resources:

$$\sum_{i=0}^{n} u_i(t) \le 1. \tag{C}$$

The resulting controlled model becomes

$$\begin{cases} \dot{s}_i = -u_i w_i(s_i) x, \qquad i = 1,2,\dots,n \\ \dot{m} = \sum_{i=1}^{n} Y_i u_i w_i(s_i) - u_0 w_R(m)(m+1), \\ \dot{x} = u_0 w_R(m) x. \end{cases} \tag{S}$$

2.2 Dynamics analysis

In this section, we provide a minimal study of the asymptotic behaviour of the controlled system (S) to set initial conditions for the dynamical optimization problem.

Lemma 1 *The set*

$$\Gamma \doteq \left\{(s_1,\dots,s_n,m,x) \in \mathbb{R}^{n+2} \,:\, s_i \ge 0,\, m \ge 0,\, x \ge 0\right\}$$

is positively invariant for system (S).

PROOF. This is easily verified by evaluating each differential equation of (S) over the boundaries of Γ.

Thus, we set initial conditions for system (S) in Γ as

$$s_i(0) = s_{i0} > 0,\ m(0) = m_0 > 0,\ x(0) = x_0 > 0. \tag{IC}$$

for $i = 1,2,\dots,n$, where m_0 is chosen positive for simplicity, even though the case $m_0 = 0$ is also admissible. Additionally, we introduce the notation $s_0 \doteq (s_{10},\dots,s_{n0})$. We define the total mass of the system

$$\bar{x}(t) \doteq \sum_{i=1}^{n} Y_i s_i(t) + (m(t)+1)x(t) \tag{1}$$

for which the following result can be found.

Lemma 2 *For every solution of system* (S) *with initial conditions in* Γ, *the total mass* $\bar{x}(t)$ *is constant for all* t, *and so* $\bar{x}(t) = \bar{x}$, *and every trajectory is bounded.*

PROOF. The fact that quantity $\bar{x}(t)$ is constant can be obtained by verifying that $\dot{\bar{x}}(t) = 0$. As for the boundedness, $\dot{s}_i \le 0$ for all i, so $s_i(t) \le s_{i0}$ for all t. Then, from (1), one can see that $\bar{x} \ge (m(t)+1)x(t) \ge x(t)$ for all t. Finally, since $\dot{x}(t) \ge 0$ for all t, we have $x(t) \ge x_0$ for all t, and so $\bar{x} \ge (m(t)+1)x_0$, which implies that $m(t) \le \bar{x}/x_0 - 1$ for all t, where $\bar{x} \ge x_0$.

Finally, since the vector field of (S) is Lipschitz continuous, the Cauchy-Lipschitz theorem ensures no finite-time convergence to the equilibrium $s_i = m = 0$.

Lemma 3 *The substrate s_i and the intermediate metabolites m cannot be completely consumed in finite time (i.e. $s_i(t) > 0$ and $m(t) > 0$ for a.e. $t \geq 0$).*

3 Optimal control problem

The main assumption justifying an optimal control approach is that the feedback mechanisms regulating the distribution of resources across cellular functions have been optimized through adaptive evolution to maximize instantaneous growth rate [17]. From a mathematical perspective, the latter is equivalent to say that the control functions $u \doteq (u_0, \ldots, u_n)$ maximize the biomass after a certain period of time. Thus, the objective function to maximize is $x(t_f)$ for a fixed time interval $[0, t_f]$. It is noteworthy that, as established in [5], the case $t_f \to \infty$ for this class of systems is trivial. Indeed, any combination of controls $u_i(t)$ taking the system asymptotically to the equilibrium $(s_1, \ldots, s_n, m, x) = (0, \ldots, 0, 0, \bar{x})$ when $t \to \infty$ is optimal, as $\bar{x}$ is the maximum possible biomass. The latter is a direct consequence of the mass conservation property stated in Lemma 2 and (1). Exceptions to this case can be obtained if not all nutrients are depleted: for example, in trajectories where there exists i and ϵ such that $s_i(t) \to \epsilon$ when $t \to \infty$, produced by a control $u_i(t) = 0$ for all $t \geq 0$. Thus, in this paper we focus on the finite-time problem.

3.1 Problem statement

The OCP writes

$$\begin{cases} \textit{maximize } J(u_0, \ldots, u_n) \doteq x(t_f), \\ \textit{subject to } \textit{dynamics of } (\mathrm{S}), \\ \qquad\qquad \textit{initial conditions } (\mathrm{IC}), \\ \qquad\qquad u(\cdot) \in \mathcal{U}, \end{cases} \tag{OCP}$$

with $\mathcal{U}$ being the set of admissible controllers, which are Lebesgue measurable real-valued functions defined on the interval $[0, t_f]$ and satisfying the constraints $u_i \in [0, 1]$ and $u_0 + \cdots + u_n \leq 1$. As (OCP) has no terminal conditions, there are no controllability and reachability issues. Additionally, controls are included in a compact and convex set (as $\mathcal{U}$ is a simplex in $\mathbb{R}^{n+1}$) and, as proved in Lemma 2, every trajectory of (S) is bounded. Thus, existence of solutions of (OCP) is guaranteed by Filippov's theorem [18]. Following PMP [19], we define the adjoint state $\lambda \doteq (\lambda_{s_1}, \lambda_{s_2}, \ldots, \lambda_{s_n}, \lambda_m, \lambda_x)$, and we write the Hamiltonian as

$$H = \sum_{i=0}^{n} H_i u_i \tag{H}$$

which is constant for a.e. $t \in [0, t_f]$, as system (S) is autonomous. Functions H_i are defined as

$$\begin{aligned} H_0 &\doteq w_R(m)(x\lambda_x - (m+1)\lambda_m), \\ H_i &\doteq w_i(s_i)(Y_i\lambda_m - x\lambda_{s_i}), \qquad i = 1, 2, \ldots, n \end{aligned} \tag{2}$$

and the terminal conditions on the adjoint state are

$$\lambda(t_f) = (0, 0, \ldots, 0, 0, -\lambda_0). \tag{TC}$$

The lack of terminal conditions on the state also allows to discard abnormal extremals, and so, without loss of generality, we can fix $\lambda_0 = -1$. The dynamics of the adjoint system is given by

$$\begin{cases} \dot{\lambda}_{s_i} = -u_i \dfrac{w_i'(s_i)}{w_i(s_i)} H_i, \quad i = 1, 2, \ldots, n \\ \dot{\lambda}_m = u_0 w_R(m) \left(\lambda_m - \dfrac{w_R'(m)}{w_R^2(m)} H_0 \right), \\ \dot{\lambda}_x = \displaystyle\sum_{i=1}^{n} u_i w_i(s_i) \lambda_{s_i} - u_0 w_R(m) \lambda_x. \end{cases} \tag{AS}$$

Finally, for notation purposes, we define the function

$$\rho(s) \doteq \max(Y_1 w_1(s_1), Y_2 w_2(s_2), \ldots, Y_n w_n(s_n)) \tag{3}$$

where $s \doteq (s_1, \ldots, s_n)$ is the set of all s_i states; and the regions of the state space

$$\begin{aligned} \overline{\omega} &\doteq \left\{ (s, m) \in \mathbb{R}^{n+1} \,:\, \rho(s) > \frac{w_R^2(m)}{w_R'(m)} \right\}, \\ \underline{\omega} &\doteq \left\{ (s, m) \in \mathbb{R}^{n+1} \,:\, \rho(s) < \frac{w_R^2(m)}{w_R'(m)} \right\}, \\ \omega &\doteq \left\{ (s, m) \in \mathbb{R}^{n+1} \,:\, \rho(s) = \frac{w_R^2(m)}{w_R'(m)} \right\}, \end{aligned}$$

that we will denote as the substrate abundant case $\overline{\omega}$, the substrate deficient case $\underline{\omega}$ and the limit case ω.

3.2 Analysis of the optimal control solutions

The use of Pontryagin's Maximum Principle reduces problem (OCP) to finding the controls u_i for $i = 0, 1, 2, \ldots, n$ that maximize the Hamiltonian (H), subject to (S), (IC), (C), (TC) and (AS). Let us first establish the positivity of the Hamiltonian and the final arc.

Lemma 4 *Any optimal solution satisfies $H(t) > 0$ for*

a.e. $t \in [0, t_f]$*, and there exists* ϵ *such that* $u_0(t) = 1$ *for a.e.* $t \in [t_f - \epsilon, t_f]$.

PROOF. Evaluating the Hamiltonian at final time yields $H|_{t=t_f} = H_0(t_f)u_0(t_f)$. Using Lemma 3, we have $H_0(t_f) = w_R(m(t_f))x(t_f) > 0$. As the optimal control at final time $u_0(t_f)$ maximizes the Hamiltonian, we see that $u_0(t_f) = 1$, and thus $H > 0$ for a.e. t (as, for autonomous systems, the Hamiltonian is constant for a.e. $t \in [0, t_f]$). The existence of an ϵ such that $u_0(t) = 1$ for a.e. $t \in [t_f - \epsilon, t_f]$ is simply given by the continuity of H_0.

The Hamiltonian is a convex combination of functions H_i. According to the maximization condition, any i-th control is active (i.e. $u_i > 0$) or non-active (i.e. $u_i = 0$) depending on the value of its associated function H_i with respect to the others $(H_0, \ldots, H_n)$. In order to comply with the positivity of the Hamiltonian, at least one function H_i is positive at every time instant. Let us define

$$\mathcal{I}_0(t) \doteq \Big\{ i = 0, 1, 2, \ldots, n \mid H_i = \max(H_0, \ldots, H_n) \Big\},$$
$$\mathcal{I}(t) \doteq \Big\{ i = 1, 2, \ldots, n \mid H_i = \max(H_1, \ldots, H_n) \Big\},$$

which are time-varying sets of subindexes at time t associated to the (single or multiple) maximal functions H_i. Through this reasoning, we can obtain a necessary condition for optimality of extremals.

Theorem 1 *Any optimal control solution satisfies*

$$\sum_{i=0}^{n} u_i = 1.$$

PROOF. Since $H > 0$, for a.e. t there exist i such $H_i(t) > 0$. Thus, using Lemma 4 and (C), the maximization of the Hamiltonian occurs when

$$H = \sum_{i=0}^{n} H_i u_i = \max(H_0, \ldots, H_n) \left(\sum_{i \in \mathcal{I}_0} u_i \right),$$

which implies that the optimal controls u_i satisfy $\sum_{i \in \mathcal{I}_0} u_i = 1$, and so $H = H_i$ for every $i \in \mathcal{I}_0(t)$.

The latter theorem replaces the polyhedron defined by the constraint (C) with a more strict constraint given by a simple plane in $\mathbb{R}^n$. As function H_i varies, every optimal control solution is a concatenation of bang arcs ($u = 0$ and $u = 1$) and singular arcs. Over an interval of time $[t_1, t_2] \subset [0, t_f]$, an arc is called:

$\mathcal{G}$ **(pure-growth) arc** if $\mathcal{I}_0(t) = \{0\}$ for a.e. $t \in [t_1, t_2]$, which produces $u_0(t) = 1$ for all $t \in [t_1, t_2]$.

$\mathcal{M}$ **(pure-metabolism) arc** if $0 \notin \mathcal{I}_0(t)$ for a.e. $t \in [t_1, t_2]$, and so $u_0(t) = 0$ for all $t \in [t_1, t_2]$, and thus all the resources are used for substrate consumption.

$\mathcal{S}$ **(singular) arc** if $0 \in \mathcal{I}_0(t)$ and $|\mathcal{I}_0(t)| > 1$ for a.e. $t \in [t_1, t_2]$, which describes a mixed strategy with both metabolism and growth.

Throughout this paper, and for simplicity of the notation, we refer to the *initial arc* of a solution as an arc defined in the interval $[0, \epsilon]$ for $\epsilon \in (0, t_f)$, while the *final arc* denotes an arc that occurs over $[t_f - \eta, t_f]$ for $\eta \in (0, t_f)$. Thus, Lemma 4 proves that the final arc of every optimal solution is a $\mathcal{G}$ arc.

In nature, regulatory mechanisms act by sensing the environment and cellular composition, and adjusting metabolism and gene expression accordingly. In optimal control theory, obtaining an optimal control strategy depending solely on the state is a very challenging task called *optimal synthesis*. We proceed to further investigate the optimal solutions in order to obtain explicit expressions of the controls u_i in feedback form. It is noteworthy that, so far, an active control u_i implies that $i \in \mathcal{I}(t)$. However, the inverse is not necessarily true: if $H = H_i$, then u_i can be either active or non-active. Hereunder, we analyze the dynamics of each type of arc.

3.2.1 Dynamics of the system over the arcs

$\mathcal{G}$ arcs are characterized by $H = H_0 > H_i$ for $i \in \mathcal{I}(t)$. Along this arc, every function $s_i(t) = s_i^*$ and $\lambda_{s_i}(t) = \lambda_{s_i}^*$ is constant for $i = 1, 2, \ldots, n$, where s_i^* and $\lambda_{s_i}^*$ denote those constant values over the arc; and $\dot{m} \leq 0$. In $\mathcal{M}$ arcs, $H = H_i > 0$ for all i in $\mathcal{I}(t)$. In this case, $x(t) = x^*$ and $\lambda_m(t) = \lambda_m^*$ are constant, with x^* and λ_m^* denoting said values over the arc; and $\dot{m} \geq 0$. The analysis of singular arcs is more challenging as it requires computing the successive derivatives of the function H_i, the first one being, as detailed in Appendix A,

$$\dot{H}_i = u_0 w_R(m) \left(H_i - Y_i w_i(s_i) \frac{w_R'(m)}{w_R^2(m)} H_0 \right). \tag{4}$$

The results are stated in the following lemma.

Lemma 5 *On an* $\mathcal{S}$ *arc over the interval of time* $[t_1, t_2] \subset [0, t_f]$*, the state* (s, m) *satisfies*

$$Y_i w_i(s_i) = \frac{w_R^2(m)}{w_R'(m)} \quad \textit{with } i \in \mathcal{I}(t), \tag{5}$$

which also means that

$$Y_j w_j(s_j) = Y_k w_k(s_k) \quad \textit{with } j \in \mathcal{I}(t), k \in \mathcal{I}(t). \tag{6}$$

Additionally, every optimal control is in feedback form $u_0(t) = u_{0,\text{sing}}(s,m,x) \in (0,1)$ *and* $u_i(t) = u_{i,\text{sing}}(s,m,x) \in (0, 1-u_{0,\text{sing}})$ *for* $i \in \mathcal{I}(t)$*, with*

$$u_{i,\text{sing}}(s,m,x) \doteq \frac{1-u_{0,\text{sing}}(s,m,x)}{\sum_{j\in\mathcal{I}} \frac{w_i'(s_i)}{w_j'(s_j)}}, \tag{7}$$

$$u_{0,\text{sing}}(s,m,x) \doteq \frac{x + \phi(s,m)\frac{w_R(m)}{w_R'(m)}}{x + \phi(s,m)\left(m+1+\frac{w_R(m)}{w_R'(m)}\right)}, \tag{8}$$

where $\phi(s,m) > 0$ *is defined as*

$$\phi(s,m) \doteq \left(2w_R'(m) - \frac{w_R(m)}{w_R'(m)} w_R''(m)\right)\left[\sum_{j\in\mathcal{I}} \frac{1}{w_i'(s_j)}\right].$$

Finally, the adjoint states λ_{s_i} *satisfy*

$$\frac{\lambda_{s_j}}{Y_j} = \frac{\lambda_{s_k}}{Y_k} > 0 \quad \textit{with } j \in \mathcal{I}(t), k \in \mathcal{I}(t), \tag{9}$$

PROOF. By definition of an $\mathcal{S}$ arc, we have $H = H_i = H_0$ for every $i \in \mathcal{I}_0(t)$ over $[t_1, t_2]$. This implies $\dot{H}_i = 0$ for $i \in \mathcal{I}(t)$, and so replacing in (4) yields (5). Differentiating (6) yields $w_j'(s_j)u_j = w_k'(s_k)u_k$, for $j \in \mathcal{I}(t)$, $k \in \mathcal{I}(t)$. Then, by replacing the latter in $\sum_{i=0}^{n} u_i = 1$, we can solve for $u_{i,\text{sing}}(s,u_0)$. Computing $\ddot{H}_i = 0$ yields (8) (see Appendix B). The bounds of $u_{0,\text{sing}}$ can be easily deduced from expression (8). Then, singular controls (7) can be proven to satisfy $u_{i,\text{sing}} \in (0, 1-u_0)$ for $i \in \mathcal{I}(t)$ by showing that the denominator satisfies

$$\sum_{j\in\mathcal{I}} \frac{w_i'(s_i)}{w_j'(s_j)} = 1 + \sum_{j\in\mathcal{I}\setminus\{i\}} \frac{w_i'(s_i)}{w_j'(s_j)} > 1.$$

By writing $H_j = H_k$ for $j \in \mathcal{I}(t)$, $k \in \mathcal{I}(t)$, we find $w_j(s_j)\lambda_{s_j} = w_k(s_k)\lambda_{s_k}$ for $j \in \mathcal{I}(t)$, $k \in \mathcal{I}(t)$ which, using (6), yields (9). Then, $\lambda_{s_i} > 0$ for $i \in \mathcal{I}(t)$ is a consequence of $\dot{\lambda}_{s_i} < 0$.

Expression (5) implies that, along every singular arc, the state is on the region $\omega \cup \overline{\omega}$ (as, a priori, there could be a non-active q-th control $u_q(t)$ associated to a substrate satisfying $Y_q w_q(s_q) > Y_i w_i(s_i)$). In next section, we study the structure of the optimal solutions based on the dynamics of the functions H_i.

3.3 Structure of the optimal solutions

As it is classical in optimal control theory, an extremal is composed of a concatenation of arcs—in this case, the three arcs presented before—determined by the time evolution of the functions H_i. In this section, we show that only a few of all possible combinations of arcs are admissible. To this end, we analyze the dynamics of the functions H_i on each arc.

Proposition 1 *If, over an interval of time* $[t_1, t_2] \subset [0, t_f]$*, the solution corresponds to a:*

- $\mathcal{G}$ ***arc****, we have* $H_0 > H_i$ *for* $i = 1, 2, \ldots, n$ *and*

$$\dot{H}_i = w_R(m)\left(H_i - Y_i w_i(s_i^*)\frac{w_R'(m)}{w_R^2(m)} H_0\right), \tag{10}$$

for $i = 1, 2, \ldots, n$*, and thus, every substrate* s_i *satisfying* $Y_i w_i(s_i) > w_R^2(m)/w_R'(m)$ *also satisfies* $\dot{H}_i < 0$*.*
- $\mathcal{M}$ ***arc****, we have* $H_i > H_0$ *for* $i \in \mathcal{I}(t)$ *and*

$$\dot{H}_0 = w_R(m)\left(H_0 \frac{w_R'(m)}{w_R^2(m)} \sum_{i\in\mathcal{I}} Y_i u_i w_i(s_i) - H_i\right), \tag{11}$$

for $i \in \mathcal{I}(t)$*, and* $\dot{H}_q = 0$ *for* $q \notin \mathcal{I}(t)$*; and so* $\dot{H}_0 < 0$ *if* $(s,m) \in \underline{\omega}$*.*
- $\mathcal{S}$ ***arc****, every function* H_q *for* $q \notin \mathcal{I}(t)$ *has dynamics*

$$\dot{H}_q = w_R(m) u_{0,\text{sing}}(s,m,x)\left(H_q - \frac{Y_q w_q(s_q^*)}{Y_i w_i(s_i)} H_i\right), \tag{12}$$

and both $s_q(t) = s_q^*$ *and* $\lambda_{s_q}(t) = \lambda_{s_q}^*$ *are constant, where* s_q^* *and* $\lambda_{s_q}^*$ *refer to the constant values of such functions over the arc.*

PROOF. Expressions (10) and (12) can be simply obtained by replacing the dynamics of each arc in (4), while (11) is the derivative of the function H_0 defined in (2) (see Appendix C).

In other words, every non-active control remains non-active along an $\mathcal{M}$ arc. As the derivative of H_q can be positive, it could happen that a non-active control becomes active. Figure 2 shows an example of the interplay between the analyzed functions, to illustrate the class of trajectories that functions H_i and H_0 could present along an extremal. Below, we exploit the dynamics of these functions, and the terminal conditions on the adjoint state (TC), to further investigate the possible structures of the optimal control solutions.

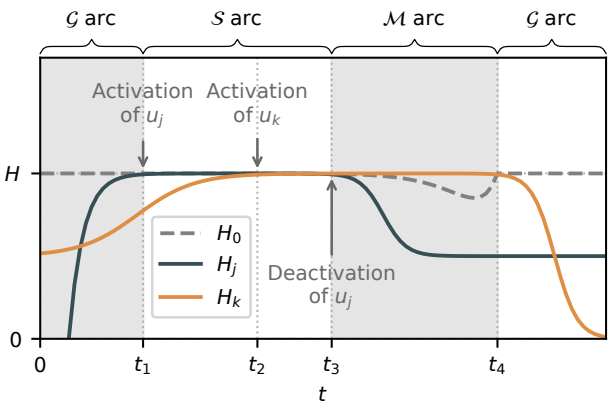

Fig. 2. Example of the evolution of functions H_i along a trajectory of the system. In $[0, t_1]$, every $H_i < H_0$, and so the solution is a $\mathcal{G}$ arc until H_j reaches $H = H_0$, and $\mathcal{I}(t_1) = \{j\}$. At that point, a switch is produced to an $\mathcal{S}$ arc, and $u_j = 1$. At time t_2, u_k also becomes active until $t = t_3$ where both H_j and H_0 decrease, producing the deactivation of u_j and a switch to an $\mathcal{M}$ arc. The latter arc is optimal until H_0 again reaches H at $t = t_4$, point at which the maximal H_k decreases, producing a switch to a $\mathcal{G}$ arc.

Proposition 2 *If the sources satisfy*

$$Y_i w_i(s_i(t)) \leq \frac{w_R^2(m(t))}{w_R'(m(t))}, \quad \forall i \in \mathcal{I}(t), \tag{13}$$

for a.e. $t \in [t_1, t_2] \subset [0, t_f]$, then an $\mathcal{M}$ arc is not admissible on that interval.

PROOF. By way of contradiction, suppose that for a.e. $t \in [t_1, t_2] \subset [0, t_f]$, condition (13) holds and the optimal solution is an $\mathcal{M}$ arc. Then, according to the dynamics of the system over $\mathcal{M}$ arcs, condition (13) becomes $Y_i w_i(s_i(t)) < w_R^2(m(t))/w_R'(m(t))$ for a.e. $t \in (t_1, t_2]$ and for every $i \in \mathcal{I}(t)$ (since the right-hand side of the inequality is increasing w.r.t. m, and $\dot{m}(t) > 0$ for a.e. $t \in [t_1, t_2]$). Then,

$$\sum_{i \in \mathcal{I}} Y_i u_i(t) w_i(s_i(t)) < \frac{w_R^2(m(t))}{w_R'(m(t))} \sum_{i \in \mathcal{I}} u_i(t) < \frac{w_R^2(m(t))}{w_R'(m(t))}$$

for a.e. $t \in (t_1, t_2]$. Using (11), this means

$$\dot{H}_0(t) < w_R(m(t)) \left(H_0(t) - H_i(t)\right) < 0$$

for a.e. $t \in (t_1, t_2]$, and thus the arc is optimal until $t_2 = t_f$, as leaving the $\mathcal{M}$ arc requires $\dot{H}_0 > 0$ (so that eventually $H_0 = H$). But this contradicts Lemma 4, as the process cannot end with an $\mathcal{M}$ arc, so the arc is not admissible over the interval $[t_1, t_2]$.

Condition (13) is less strict than $(s, m) \in \underline{\omega}$, as it only applies to those substrates that are active. However, it is useful to discard $\mathcal{M}$ arcs as intermediate arcs, as we show in the following proposition.

Proposition 3 *An $\mathcal{M}$ arc is only admissible over an interval $[0, t_1] \subset [0, t_f]$ (i.e. at the beginning of the process).*

PROOF. In order to prove the proposition, let us show that an $\mathcal{M}$ arc cannot occur after an $\mathcal{S}$ arc or $\mathcal{G}$ arc. Suppose that, over an interval $[t_1, t_2] \subset [0, t_f]$, the optimal solution is:

- an $\mathcal{S}$ arc. Then, every active control $u_i(t)$ with $i \in \mathcal{I}(t)$ satisfies (5) for a.e. $t \in [t_1, t_2]$, and so the solution cannot enter an $\mathcal{M}$ arc at time t_2, as stated in Proposition 2.
- a $\mathcal{G}$ arc. Then $H_i(t) < H_0(t)$ for every $i = 1, 2, \ldots, n$ and for a.e. $t \in [t_1, t_2]$. Suppose that an $\mathcal{M}$ arc is optimal for $t \geq t_2$. This implies that there exists j such that $H_j(t_2) = H_0(t_2)$, and so $j \in \mathcal{I}(t_2)$. Then, there exists $\epsilon > 0$ such that $\dot{H}_j(t) > 0$ over an interval $[t_2 - \epsilon, t_2) \subset [0, t_f]$ (i.e. before the switch time). According to (10), this can only happen if

$$Y_j w_j(s_j(t)) < w_R^2(m(t))/w_R'(m(t)) \tag{14}$$

for a.e. $t \in [t_2 - \epsilon, t_2)$. The admissibility of the $\mathcal{M}$ arc at $t = t_2$ also implies that $Y_j w_j(s_j(t_2)) > w_R^2(m(t_2))/w_R'(m(t_2))$ (otherwise, it contradicts Proposition 2). Over the $\mathcal{G}$ arc (i.e. for a.e. $t \in [t_1, t_2]$) we have $\dot{s}_i(t) = 0$ for every $i = 1, 2, \ldots, n$ and $\dot{m}(t) < 0$. Thus, by continuity of the states, this implies the existence of $\eta > 0$ such that $Y_j w_j(s_j(t)) \geq w_R^2(m(t))/w_R'(m(t))$ for a.e. $t \in [t_2 - \eta, t_2] \subset [0, t_f]$. But this contradicts (14), and so the $\mathcal{M}$ arc is not admissible.

Lemma 6 *Over a $\mathcal{S}$ arc over $t \in [t_1, t_2] \subset [0, t_f]$, $\max(Y_1 w_1(s_1(t)), \ldots, Y_n w_n(s_n(t))) = Y_i w_i(s_i(t))$ for $i \in \mathcal{I}(t)$, and for a.e. $t \in [t_1, t_2]$, and so $(s, m) \in \omega$.*

PROOF. Over a $\mathcal{S}$ arc occurring over $[t_1, t_2]$, the conditions (5) holds for every $i \in \mathcal{I}(t)$ and for a.e. $t \in [t_1, t_2]$. By way of contradiction, suppose that at time $t^* \in [t_1, t_2]$ there exists an $s_q(t)$ substrate with $q \notin \mathcal{I}(t)$ (such that $s_q(t) = s_q^*$) satisfying

$$Y_q w_q(s_q^*) \geq Y_i w_i(s_i(t^*)) \quad \textit{for } i \in \mathcal{I}(t).$$

Using (12), we have that $\dot{H}_q(t) < 0$ for a.e. $t \in [t^*, t_2]$, and thus $q \notin \mathcal{I}(t)$ for a.e. $t \in [t^*, t_2]$, which also means that $Y_q w_q(s_q^*) > w_R^2(m(t))/w_R'(m(t))$ for a.e. $t \in [t^*, t_2]$ (as $w_R^2(m)/w_R'(m)$ is decreasing w.r.t. m). Additionally, $H_q(t) < H_i(t)$ for a.e. $t \in [t^*, t_2]$, which means

$$w_q(s_q^*) \lambda_{s_q}^* \geq w_i(s_i(t)) \lambda_{s_i}(t) > 0 \tag{15}$$

for a.e. $t \in [t^*, t_2]$, with $\lambda_{s_i} > 0$, as shown in (9). It is easy to see that $q \notin \mathcal{I}(t)$ for a.e. $t \geq t^*$, that is, until the end of the process: the function $w_R^2(m)/w_R'(m)$ can only increase along an $\mathcal{M}$ arc, which cannot occur for $t \geq t^*$ (as established in Proposition 3). This means $Y_q w_q(s_q) = w_R^2(m)/w_R'(m)$ cannot be reached, and thus $u_q(t) = 0$ for a.e. $t > t^*$. But, according to (TC), $\lambda_{s_q}(t_f) = 0$, and so $\lambda_{s_q}(t) = 0$ for a.e. $t \geq t^*$ (due to the fact that $\dot{\lambda}_{s_q}$ is linear on $u_q(= 0)$, which contradicts (15)).

Lemma 6 establishes a first clear relation between the consumption of the i-th substrate and its associated uptake control u_i. In order to maximize its growth, the system should prioritize the metabolization of substrates that are solutions of $\rho(s)$, which implies favoring the substrates that allow maximal synthesis of intermediate metabolites from the nutrients in the environment. This result is in accordance with those obtained in simpler mathematical models of bacterial growth with no intermediate metabolites [7,10], as well as with experimental observations of bacterial growth laws [13]. In other words, over an $\mathcal{S}$ arc, the criterion for substrate consumption $\mathcal{I}(t) = \mathcal{I}(s)$, depending solely on environmental conditions (and not on t). In practice, under this control law, the i-th substrate corresponding to the maximal function $Y_i w_i(s_i)$ is consumed first, and so s_i decreases until it reaches the second maximal function $Y_j w_j(s_j)$ associated to the j-th substrate. At that point, both substrates s_i and s_j start being consumed simultaneously, and the cycle is repeated until all the sources are (asymptotically) depleted, or until the end of the singular arc. Another important consequence of Lemma 6 is that no active control can become inactive. Indeed, if a substrate stops being consumed along a singular arc, it would satisfy $Y_q w_q(s_q) > w_R^2(m)/w_R'(m)$, and so the state would enter the region $\overline{\omega}$, which contradicts Lemma 6. Finally, we proceed to enumerate the possible structures of the optimal solutions.

Theorem 2 *The optimal control solutions can be:*

- *A single $\mathcal{G}$ arc*
- *$\mathcal{M} - \mathcal{G}$ for initial conditions in $\overline{\omega}$*
- *$\mathcal{M} - \mathcal{S} - \mathcal{G}$ for initial conditions in $\overline{\omega}$*
- *$\mathcal{G} - \mathcal{S} - \mathcal{G}$ for initial conditions in $\underline{\omega}$*

PROOF. First, we state the fact that an optimal solution can admit at most one $\mathcal{S}$ arc, which is followed by the final $\mathcal{G}$ arc. The proof is based on Lemma 6: suppose that an $\mathcal{S}$ arc occurs over the interval $[t_1, t_2] \subset [0, t_f]$. Then, $(s(t_2), m(t_2)) \in \omega$, and the solution enters a $\mathcal{G}$ arc (as, according to Lemma 4 and Proposition 3, $t_2 \neq t_f$, and there cannot be an $\mathcal{M}$ arc at $t = t_2$). This means that there exists $\epsilon > 0$ such that $\dot{m}(t) < 0$ for a.e. $t \in (t_2, t_2 + \epsilon]$, and so $(s(t), m(t)) \in \overline{\omega}$ for a.e. $t \in (t_2, t_2 + \epsilon]$. As ω cannot be reached along the $\mathcal{G}$ arc, the process finishes with the $\mathcal{G}$ arc (and, consequently, no other $\mathcal{S}$ arc is allowed). Let us analyze the remaining admissible structures, taking into account Proposition 3:

- If $(s_0, m_0) \in \overline{\omega}$ and the initial arc is a $\mathcal{G}$ arc: the control is a single $\mathcal{G}$ arc and $(s(t), m(t)) \in \overline{\omega}$ for a.e. $t \in [0, t_f]$.
- If $(s_0, m_0) \in \overline{\omega}$ and the initial arc is an $\mathcal{M}$ arc, there exists $t_1 \in (0, t_f)$ such that $0 \in \mathcal{I}_0(t_1)$ (according to Lemma 4). At that point, the solution can: 1) enter a $\mathcal{S}$ arc if $(s(t_1), m(t_1)) \in \omega$, and then the final $\mathcal{G}$ arc; or 2) enter a $\mathcal{G}$ arc if $(s(t_1), m(t_1)) \in \overline{\omega}$ (as $\dot{H}_i < 0$).
- If $(s_0, m_0) \in \underline{\omega}$ and the initial arc is an $\mathcal{M}$ arc: the control is a single $\mathcal{M}$ arc, which contradicts Lemma 4.
- If $(s_0, m_0) \in \underline{\omega}$ and the initial arc is a $\mathcal{G}$ arc, either: 1) $0 \notin \mathcal{I}_0(t)$ for a.e. $t \in [0, t_f]$, and thus the control is a single $\mathcal{G}$ arc; or 2) there exists $t_1 \in (0, t_f)$ such that $0 \in \mathcal{I}_0(t_1)$, and so the solution can:
  - enter an $\mathcal{S}$ arc if $(s(t_1), m(t_1)) \in \omega$, and then the final $\mathcal{G}$ arc.
  - continue over the $\mathcal{G}$ arc. Then, $\dot{H}_i(t_1) = 0$, as $\dot{H}_i(t) > 0$ for a.e. $t \in [0, t_1)$ and there exists ϵ such that $\dot{H}_i(t) < 0$ for a.e. $t \in (t_1, t_1 + \epsilon]$ (otherwise, the arc would enter an $\mathcal{S}$ arc at t_1). Thus, $(s(t_1), m(t_1)) \in \omega$, and so $(s(t), m(t)) \in \overline{\omega}$ for a.e. $t > t_1$, and the control is a single $\mathcal{G}$ arc.

From a practical point of view, the control strategies that do not admit a singular arc ($\mathcal{G}$ and $\mathcal{M} - \mathcal{G}$) are optimal when the state ω is not reachable. For example, an optimal solution with $(s_0, m_0) \in \underline{\omega}$ that does not reach $(s, m) \in \omega$ results in a pure $\mathcal{G}$ control strategy, either because t_f is too small with respect to the reaction rates, or because $\rho(s) \ll w_R^2(m)/w_R'(m)$. However, for large values of t_f, singular arcs become admissible. We will then focus on the cases where t_f is sufficiently large so as to allow singular arcs, which reduces the analysis to $\mathcal{G} - \mathcal{S} - \mathcal{G}$ and $\mathcal{M} - \mathcal{S} - \mathcal{G}$ solutions. In order to conclude the analysis, it suffices to study the behavior of the optimal trajectories along the initial arc. The initial arc can be either $\mathcal{G}$ or $\mathcal{M}$ depending on the initial conditions. If the initial conditions $(s_0, m_0) \in \overline{\omega}$, the initial arc is an $\mathcal{M}$ arc. Along an initial $\mathcal{M}$ arc over the interval $[0, t_1]$, every function H_i for $i = 1, 2, \ldots, n$ is constant, and thus $\mathcal{I}(t)$ is constant for a.e. $t \in [0, t_1]$. However, along an $\mathcal{M}$ arc, it is possible to have $u_i = 0$ for $i \in \mathcal{I}(t)$. As there is no biomass production (and x is constant), the pattern and order in which the substrates are consumed does not affect the state at the junction between arcs t_1 as long as the conditions for optimality are met. More precisely, any combination of controls u_i satisfying Theorem 1 and taking the state from the initial condition (IC) to ω is optimal. In the interest of homogenizing the control criteria, one could choose the same strategy used along the singular arc, by adopting the control law obtained in (7). In next section, we validate the analytical results with numerical simulations.

4 Numerical simulations

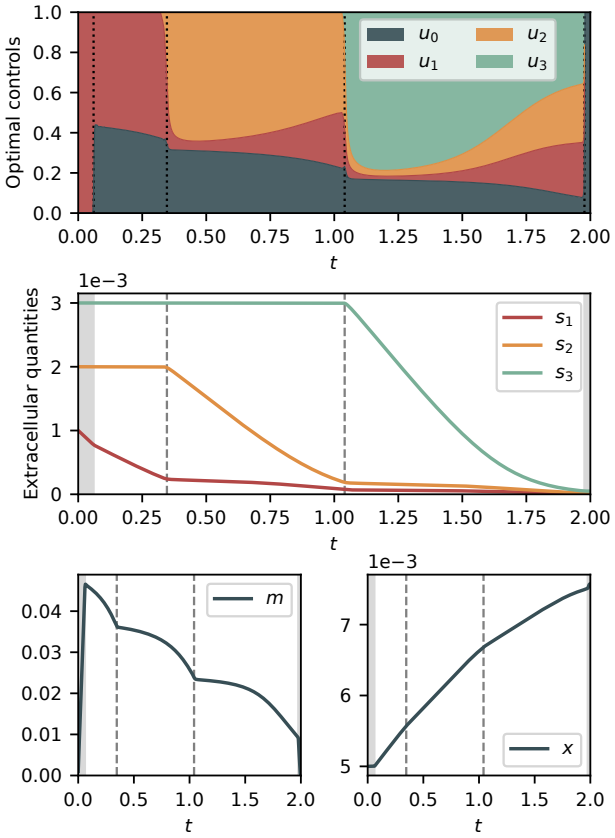

Fig. 3. Optimal trajectory of (OCP) obtained using Bocop. Parameter values are $k = 10\ \mathrm{h}^{-1}$, $K_R = 1$, $K_i = 0.1$, $Y_1 = 1$, $Y_2 = 0.5$ and $Y_3 = 0.2$ and initial conditions are set to $s_{10} = 0.001$, $s_{20} = 0.002$, $s_{30} = 0.003$, $m_0 = 0.001$ and $x_0 = 0.005$. The shaded regions at the beginning and end of each plot denote the time intervals where the optimal solution is either an $\mathcal{M}$ or a $\mathcal{G}$, while the non-shaded area in the middle denotes the singular arc. Two dashed vertical lines indicate the time instants where s_2 and s_3 start being consumed.

In this section, we simulate a system with three sources s_1, s_2 and s_3, with associated yields satisfying $Y_1 > Y_2 > Y_3$ and initial conditions chosen to produce sequential substrate uptake, such that $Y_1 w_1(s_{10}) > Y_2 w_2(s_{20}) > Y_3 w_3(s_{30})$. We resort to the case where the reaction rates are defined as Michaelis-Menten kinetics in terms of the availability of the quantities as

$$w_R(m) = k_R \frac{m}{K_R + m}, \qquad w_i(s_i) = k_i \frac{s_i}{K_i + s_i}$$

where K_R and K_i are the half-saturation constants of the synthesis rates, and k_R and k_i are the maximal reaction rates, that are fixed to $k_R = k_i = k$ in the simulations (as the difference in the maximal rates can be also set through the parameters Y_i). The numerical optimal trajectories are computed with Bocop [20], an open-source toolbox that computes optimal solutions using direct methods, by discretizing the time variable and then solving a finite-dimensional optimization problem that approximates the OCP. The discretization algorithm used is Gauss II (implicit, 2-stage, order 4) with 10000 time steps. Parameter values are chosen to emphasize the structure of the optimal trajectory (rather than the biological meaning). Figure 3 shows an optimal trajectory with $(s_0, m_0) \in \overline{\omega}$ where s_1, s_2 and s_3 are consumed sequentially. In this example, t_f is relatively small compared to biologically relevant values, and yet, the solution admits a singular arc that represents more than 95% of the time interval. This results in an optimal solution following a structure $\mathcal{M} - \mathcal{S} - \mathcal{G}$. This phenomenon is strongly related to the so-called *turnpike* properties [21], where solutions of OCPs for sufficiently large final times are described by a main arc spending most of the time near a steady-state, enclosed by transient arcs at the beginning and end of the process. We can also see that the sequential activation of controls u_i occurs when the previous substrate s_{i-1} attains a very low concentration level, and is followed by a phase where u_i reaches a maximum value—taking up most of the resources dedicated to substrate uptake—and decreases progressively until the next activation or the end of the process. Figure 4 confirms the main theoretical results. At $t = 0$, the solution of (3) is $Y_1 w_1(s_1)$ and $\mathcal{I}(0) = \{1\}$. Thus, u_1 is the only active control, and it remains the only active control until the end of the initial $\mathcal{M}$ arc. As expected, H_0 and m increase until $(s, m) \in \omega$, and H_0 becomes equal to H_1. At that point, the solution enters a singular arc, and continues to adopt the uptake control strategy $u_1 = 1$ until $Y_1 w_1(s_1) = Y_2 w_2(s_2)$, which occurs when H_2 reaches H, activating the second control u_2. The same sequence occurs for the 3rd substrate, point at which all three substrates are being consumed simultaneously until the final $\mathcal{G}$ arc. Once in the final arc, (most of) the remaining intermediate metabolites are converted into biomass and, simultaneously, all three H_i functions converge to 0, while H_0 remains constant. Figure 5 shows another case with three substrates that have the exact same initial concentration. The initial metabolite concentration m_0 is set to a high value so that $(s_0, m_0) \in \underline{\omega}$, which produces an initial $\mathcal{G}$ arc. This also increases the duration of the final $\mathcal{G}$ arc, as depleting the pool of intermediate metabolites requires additional time.

5 Application to experimental data

Previous studies showed that the bacterium *Escherichia coli* exhibits sequential substrate uptake patterns in environments where the only available carbon sources are glucose and lactose. The latter has been explained by the fact that biomass yield on glucose is higher than on lactose, a criterion that is successfully captured in the singular control stated in Lemma 6. The objective of this section is to use the theoretical results to predict the natural behavior aforementioned. For that, we propose

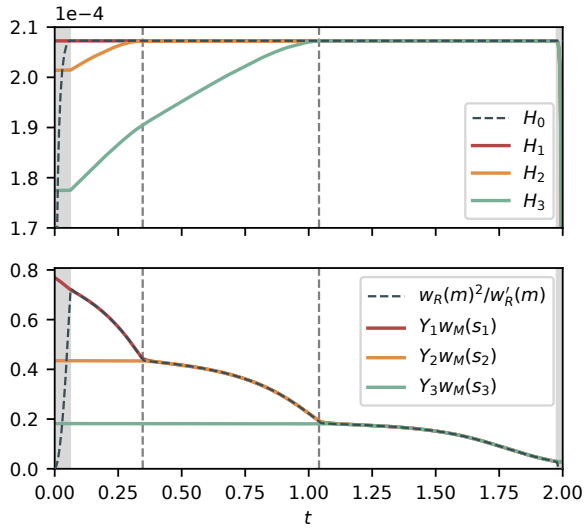

Fig. 4. Functions $w_R^2(m)/w'_R(m)$ and $Y_i w_i(s_i)$ (lower plot) and H_0 and H_i (upper plot) related to the trajectories plotted in Figure 3. Shaded regions and dashed vertical lines indicate arcs and uptake control activation, respectively. Functions H_i associated to non-active controls increase until they reach H, while functions $Y_i w_i(s_i)$ decrease.

a feedback control law (i.e. a control function expressed in terms of the state) that is based on the optimality principles obtained through PMP. Then, we build a two-substrate model representing E. coli growing on glucose and lactose, and we match the results to experimental data.

5.1 Feedback control law

The pool of intermediate metabolites m acts as a buffer compartment regulated according to the phase of the process. Its presence along the resource pathway results in the trade-off between gene expression and metabolism—otherwise, consuming substrates and producing proteins become the same task, as in the classical Monod model. While the pool is regulated to optimal levels throughout the singular arc, substrate uptake becomes unnecessary towards the end of an optimal bioprocess, as there is no time remaining to produce biomass from the available intermediate metabolites. This is the role of the final $\mathcal{G}$ arc, which engages all the available cell resources to the task of emptying the pool of intermediate metabolites while arresting substrate uptake. As t_f increases, the duration of the final arc becomes negligible in comparison to the duration of the singular arc, a classical behavior in solutions exhibiting the turnpike phenomenon. Thus, in the interest of studying the long-term perspective of the biological phenomenon, we define a suboptimal control law that can be expressed in feedback form by neglecting the

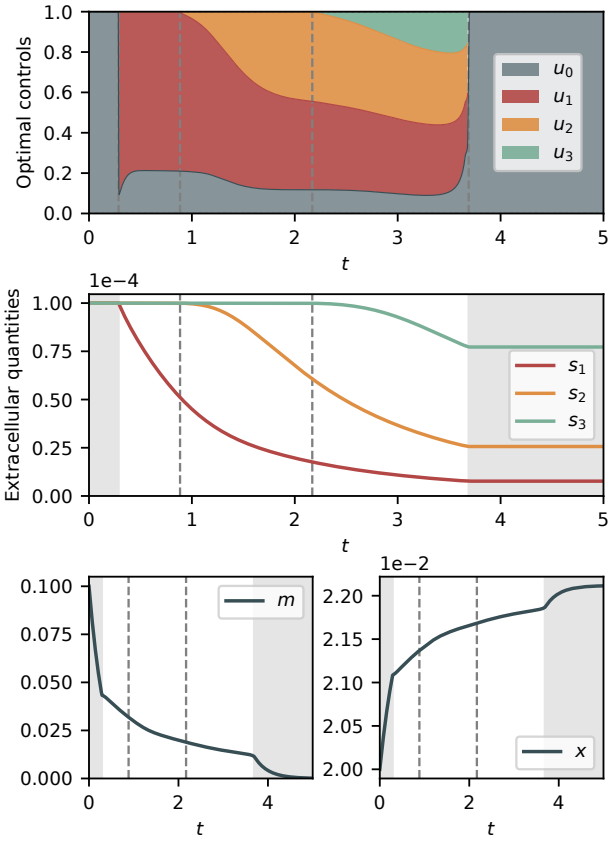

Fig. 5. Optimal trajectory of (OCP) obtained using Bocop. Parameter values are $k = 10\ \mathrm{h}^{-1}$, $K_R = 1$, $K_i = 0.05$, $Y_1 = 1$, $Y_2 = 0.3$ and $Y_3 = 0.1$ and initial conditions are set to $s_{10} = s_{20} = s_{30} = 0.0001$, $m_0 = 0.1$ and $x_0 = 0.02$. The shaded regions at the beginning and end of each plot denote the time intervals where the optimal solution is a $\mathcal{G}$ arc, and the non-shaded area in the middle shows the singular arc. The two dashed vertical lines indicate the time instants where s_2 and s_3 start being consumed.

final arc of the optimal control structure:

$$u_0(s,m,x) = \begin{cases} 0 & \text{if } (s,m) \in \overline{\omega}, \\ 1 & \text{if } (s,m) \in \underline{\omega}, \\ u_{0,\mathrm{sing}}(s,m,x) & \text{if } (s,m) \in \omega, \end{cases}$$

$$u_i(s,m,x) = \begin{cases} \dfrac{1-u_0(s,m,x)}{\sum_{j\in\mathcal{I}} \dfrac{w'_i(s_i)}{w'_j(s_j)}} & \text{if } i \in \mathcal{I}(s), \\ 0 & \text{otherwise}, \end{cases}$$

$$\mathcal{I}(s) = \{\, i = 1,2,\ldots,n \mid \rho(s) = Y_i w_i(s_i) \,\}. \tag{16}$$

for $i = 1,2,\ldots,n$. When applying the control law to system (S), it produces trajectories $\mathcal{G}-\mathcal{S}$ and $\mathcal{M}-\mathcal{S}$,

depending on whether the initial conditions are in $\underline{\omega}$ or $\overline{\omega}$, respectively. Indeed, the closed-loop structure consists of a single bang arc that takes the state to ω, and "slides" over the singular surface until the end of the process. Similar behaviors can be also found in the theory of sliding mode control, where a discontinuous control function forces the system to reach and stay in a particular region of the state space. Hereunder, we study the case of a 2-substrate culture under the proposed state feedback control and compare the results to experimental data of a batch process of E. coli.

5.2 Diauxic growth on glucose/lactose

The experimental data used in this section corresponds to a batch process of a wild-type strain of E. coli growing on glucose and lactose [13], where the data points are 15 measurements of the concentration of each substrate and of the bacterial biomass. In order to model the process, we define a system (S) with $n = 2$, where the states s_1 and s_2 account for the concentrations of glucose and lactose respectively, and thus $Y_1 > Y_2$. The model is simulated in its closed-loop form, with the feedback control law defined in (16). The closed-loop model was calibrated by adjusting the parameters $\theta \doteq (k, K_1, K_2, Y_1, Y_2)$ through a least-squares algorithm minimizing the difference between model simulation and experimental measurements, given by the cost function

$$f(\theta) \doteq \sum_{i=1}^{15} \left[(s_1(t_i) - \beta \hat{s}_1^i)^2 + (s_2(t_i) - \beta \hat{s}_2^i)^2 + (x(t_i) - \beta \hat{x}^i)^2 \right],$$

where $\hat{s}_1^i$, $\hat{s}_2^i$ and $\hat{x}^i$ correspond to the i-th measurement of glucose, lactose and biomass respectively, and t_i is the time instant at which the i-th measurement has been obtained. For calibration and simulation, the non-dimensional quantities s_1, s_2 and x are divided by a parameter $\beta = 0.003$ L/g representing the inverse of the cell density in E. coli [2], which yields concentrations in grams per liter. The resulting calibrated parameters are $k = 1.77$ h^{-1}, $K_1 = 0.037$ g/L, $K_2 = 0.01$ g/L, $Y_1 = 0.77$ and $Y_2 = 0.25$, which are consistent with previous studies [10,2]. For the numerical simulation, the initial conditions are set to $s_1(t_1) = 6.522 \times 10^{-4}$, $s_2(t_1) = 3.42 \times 10^{-3}$, $m(t_1) = 0.01$ and $x(t_1) = 9.477 \times 10^{-5}$, in accordance with the experimental data. Comparison between the experiments and the model behavior are shown in Figure 6. The model successfully predicts the diauxic growth behavior as an optimal response to the difference between carbon sources (in yield and concentration). The largest deviation between model output and data occurs when u_2 becomes active: experiments indicate that bacteria would fully exhaust the glucose in the medium before switching to lactose, a strategy that slightly differs from the optimal control approach. First,

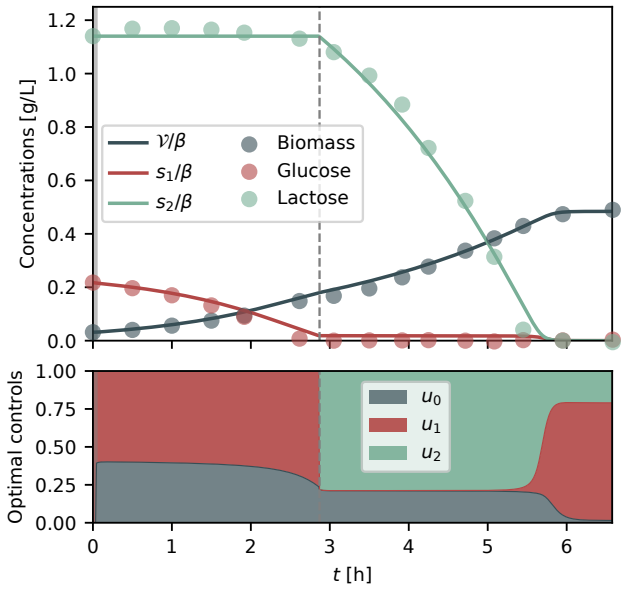

Fig. 6. Comparison with experimental measurements (of glucose, lactose and biomass) from [13]. In the top plot, the data from the experiment is represented in circles, while the simulation of model (S) is plotted in lines. The time-instant at which $Y_1 w_1(s_1)$ becomes equal to $Y_2 w_2(s_2)$ is marked with a vertical dashed line. The shaded region at the beginning of each plot indicates the initial $\mathcal{M}$ arc.

and as shown in Lemma 3, the model presented in this paper does not allow for full consumption of the substrates in finite time, as the depletion occurs—at worse—at exponential rate. Additionally, when a certain control becomes active, it remains active until the end of the process. This means that, even though $u_1 \approx 0$ after the switch, it cannot vanish exactly. Therefore, this behavior is not captured by our approach. It is also noteworthy that the length of the initial $\mathcal{M}$ arc is marginal with respect to the duration of the process ($\approx 0.5\%$ of the final time). This is consistent with the fact that, in a real biological process, bacteria alter their cellular composition progressively—constrained by the maximal synthesis rates of proteins and resource availability—and not instantaneously as modelled in this paper. However, the approach remains applicable along the singular surface, where cellular composition is governed by the expression defined in (8). When governed by this expression, the resources allocated to protein synthesis u_0 are gradually decreased as $(s, m) \to 0$. Rather than a biological mechanism, the latter appears to be a mathematical artifact consequence of the assumptions and simplifications of the approach. In nature, additional cellular functions come into play in long-term survival, e.g. related to the feast/famine cycle of cells.

6 Conclusion

This paper addressed the optimal control of a generic class of growth models, inspired by the naturally-evolved

regulatory mechanisms employed by bacterial cells. From a biological point of view, the model aims to predict how resources are divided across cellular functions in order to maximize biomass. The problem is posed in the context of optimal control, and necessary conditions for optimality are derived by applying PMP. The dynamics of the system and optimal control along each type of arc are investigated, and the admissible structures of the optimal solutions are proven to be a simple concatenation of the studied arcs. Based on the optimality principles obtained with PMP, a simple feedback control law is proposed, that retains the main features of the optimal control. The resulting closed-loop model that can be calibrated to represent batch processing of E. coli on glucose and lactose. In spite of the simplicity of the approach, the model is capable of reproducing the substrate uptake pattern known as diauxic growth, and to successfully fit experimental results.

While the approach aims to better understand the self-adaptive capabilities of living organisms, and their—sometimes hidden—naturally-evolved feedback loops, its formulation draws upon a general non-linear dynamical model which can be easily extrapolated to other systems. As pointed out in previous works, biologically-inspired resource distribution can be often related to results from different disciplines, such as the law of equi-marginal utility in economic theory [7]. In this line of work, numerous natural phenomena remain unstudied, that have the potential to provide guidance in designing artificial control loops.

Acknowledgements

I am very grateful to Jean-Luc Gouzé, Jean-Baptiste Caillau and Alain Rapaport for the enriching discussions and valuable suggestions they provided during the production of this paper.

References

[1] Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592, 2005.

[2] Nils Giordano, Francis Mairet, Jean-Luc Gouzé, Johannes Geiselmann, and Hidde De Jong. Dynamical allocation of cellular resources as an optimal control problem: novel insights into microbial growth strategies. *PLoS Computational Biology*, 12(3):e1004802, 2016.

[3] Ivan Yegorov, Francis Mairet, Hidde De Jong, and Jean-Luc Gouzé. Optimal control of bacterial growth for the maximization of metabolite production. *Journal of Mathematical Biology*, pages 1–48, 2018.

[4] Agustín G. Yabo, Jean-Baptiste Caillau, and Jean-Luc Gouzé. Optimal bacterial resource allocation: metabolite production in continuous bioreactors. *Mathematical Biosciences and Engineering*, 17(6):7074–7100, 2020.

[5] Agustín G. Yabo, Jean-Baptiste Caillau, and Jean-Luc Gouzé. Optimal bacterial resource allocation strategies in batch processing. Preprint, June 2022.

[6] Katarzyna Potrykus, Helen Murphy, Nadège Philippe, and Michael Cashel. ppGpp is the major source of growth rate control in E. coli. *Environmental Microbiology*, 13(3):563–575, 2011.

[7] P Dhurjati, D Ramkrishna, MC Flickinger, and GT Tsao. A cybernetic view of microbial growth: modeling of cells as optimal strategists. *Biotechnology and Bioengineering*, 27(1):1–9, 1985.

[8] Andreas Kremling, Johannes Geiselmann, Delphine Ropers, and Hidde de Jong. An ensemble of mathematical models showing diauxic growth behaviour. *BMC Systems Biology*, 12(1):1–16, 2018.

[9] Pierre Salvy and Vassily Hatzimanikatis. Emergence of diauxie as an optimal growth strategy under resource allocation constraints in cellular metabolism. *Proceedings of the National Academy of Sciences*, 118(8):e2013836118, 2021.

[10] Aravinda R Mandli and Jayant M Modak. Optimal control analysis of the dynamic growth behavior of microorganisms. *Mathematical Biosciences*, 258:57–67, 2014.

[11] Agustín G. Yabo, Jean-Baptiste Caillau, Jean-Luc Gouzé, Hidde de Jong, and Francis Mairet. Dynamical analysis and optimization of a generalized resource allocation model of microbial growth. *SIAM Journal on Applied Dynamical Systems*, 21(1):137–165, 2022.

[12] Agustín G. Yabo. Predicting microbial cell composition and diauxic growth as optimal control strategies. *IFAC-PapersOnLine*, 56(2):6217–6222, 2023.

[13] Katja Bettenbrock, Sophia Fischer, Andreas Kremling, Knut Jahreis, Thomas Sauter, and Ernst-Dieter Gilles. A quantitative approach to catabolite repression in Escherichia coli. *Journal of Biological Chemistry*, 281(5):2578–2584, 2006.

[14] Michaël R Droop. Vitamin B12 and marine ecology. IV. the kinetics of uptake, growth and inhibition in Monochrysis lutheri. *Journal of the Marine Biological Association of the United Kingdom*, 48(3):689–733, 1968.

[15] Sheng Hui, Josh M Silverman, Stephen S Chen, David W Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11(2):784, 2015.

[16] David W Erickson, Severin J Schink, Vadim Patsalo, James R Williamson, Ulrich Gerland, and Terence Hwa. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. *Nature*, 551(7678):119–123, 2017.

[17] Nathan E Lewis, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, Gunnar Schramm, Samuel O Purvine, Daniel Lopez-Ferrer, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6(1):390, 2010.

[18] MI Zelikin, Andrei A Agrachev, Yuri Sachkov, and Yuri L Sachkov. *Control theory from the geometric viewpoint*, volume 2. Springer Science & Business Media, 2004.

[19] Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. Routledge, 2018.

[20] Inria Saclay Team Commands. Bocop: an open source toolbox for optimal control. http://bocop.org, 2017.

[21] Emmanuel Trélat and Enrique Zuazua. The turnpike property in finite-dimensional nonlinear optimal control. *Journal of Differential Equations*, 258(1):81–114, 2015.

A Computation of $\dot{H}_i$

Differentiating H_i for any i yields

$$\begin{aligned}
\dot{H}_i =& w_i'(s_i)(Y_i\lambda_m - x\lambda_{s_i})\dot{s}_i + w_i(s_i)(Y_i\dot{\lambda}_m - \dot{x}\lambda_{s_i} - x\dot{\lambda}_{s_i}) \\
=& - w_i'(s_i)(Y_i\lambda_m - x\lambda_{s_i})u_i w_i(s_i)x \\
&+ w_i(s_i)Y_i u_0 w_R(m)\left(\lambda_m - \frac{w_R'(m)}{w_R^2(m)}H_0\right) \\
&- w_i(s_i)u_0 w_R(m)x\lambda_{s_i} \\
&+ w_i(s_i)xu_i\frac{w_i'(s_i)}{w_i(s_i)}H_i \\
=& -w_i'(s_i)\underbrace{(Y_i\lambda_m - x\lambda_{s_i})w_i(s_i)}_{H_i}\cancel{u_i x} \\
&+ u_0 w_R(m)\underbrace{w_i(s_i)(Y_i\lambda_m - x\lambda_{s_i})}_{H_i} \\
&- w_i(s_i)Y_i u_0 w_R(m)\frac{w_R'(m)}{w_R^2(m)}H_0 \\
&\cancel{+xu_i w_i'(s_i)H_i} \\
=& u_0 w_R(m)\left(H_i - Y_i w_i(s_i)\frac{w_R'(m)}{w_R^2(m)}H_0\right).
\end{aligned}$$

Along a $\mathcal{S}$ arc, if $i \in \mathcal{I}(t)$, we have $\dot{H}_i = 0$ and $H_i = H_0$, which implies that

$$Y_i w_i(s_i) = \frac{w_R^2(m)}{w_R'(m)}. \tag{A.1}$$

Along a $\mathcal{G}$ arc, $u_0 = 1$, and so

$$\dot{H}_i = w_R(m)\left(H_i - Y_i w_i(s_i)\frac{w_R'(m)}{w_R^2(m)}H_0\right).$$

B Computation of singular controls from $\ddot{H}_i$

Along a $\mathcal{S}$ arc, using Theorem 1 and the fact that $u_j = u_k w_k'(s_k)/w_j'(s_j)$, we have

$$\begin{aligned}
u_0 + u_1 + \cdots + u_i + \cdots + u_n &= 1 \\
u_0 + u_i\frac{w_i'(s_i)}{w_1'(s_1)} + \cdots + u_i + \cdots + u_i\frac{w_i'(s_i)}{w_n'(s_n)} &= 1 \\
u_0 + u_i\left(\frac{w_i'(s_i)}{w_1'(s_1)} + \cdots + 1 + \cdots + \frac{w_i'(s_i)}{w_n'(s_n)}\right) &= 1 \\
u_0 + u_i\sum_{j\in\mathcal{I}}\frac{w_i'(s_i)}{w_j'(s_j)} &= 1
\end{aligned}$$

and so

$$u_i(u_0, s) = \frac{1 - u_0}{\sum_{j\in\mathcal{I}}\frac{w_i'(s_i)}{w_j'(s_j)}}.$$

Then, by computing the derivative of (A.1), we obtain

$$\begin{aligned}
0 =& - Y_i w_i'(s_i)\dot{s}_i + w_R(m)\left(2 - \frac{w_R(m)}{w_R'(m)^2}w_R''(m)\right)\dot{m} \\
0 =& Y_i w_i'(s_i)u_i w_i(s_i)x \\
&+ \Phi(m)\left(\sum_{i=1}^n Y_i u_i w_i(s_i) - u_0 w_R(m)(m+1)\right),
\end{aligned}$$

with

$$\Phi(m) \doteq w_R(m)\left(2 - \frac{w_R(m)}{w_R'(m)^2}w_R''(m)\right).$$

Solving for u_0 yields

$$\begin{aligned}
0 =& \underbrace{Y_i w_i(s_i)}_{w_R^2(m)/w_R'(m)} w_i'(s_i)\frac{1 - u_0}{\sum_{j\in\mathcal{I}}\frac{w_i'(s_i)}{w_j'(s_j)}}x \\
&+ \Phi(m)\left(\underbrace{Y_i w_i(s_i)}_{w_R^2(m)/w_R'(m)}\underbrace{\sum_{i=1}^n u_i}_{1-u_0} - u_0 w_R(m)(m+1)\right) \\
0 =& \frac{w_R^2(m)}{w_R'(m)}\frac{1 - u_0}{\sum_{j\in\mathcal{I}}\frac{1}{w_j'(s_j)}}x \\
&+ \Phi(m)\left(\frac{w_R^2(m)}{w_R'(m)}(1 - u_0) - u_0 w_R(m)(m+1)\right) \\
0 =& \frac{w_R^2(m)}{w_R'(m)}\frac{1}{\sum_{j\in\mathcal{I}}\frac{1}{w_j'(s_j)}}x + \Phi(m)\frac{w_R^2(m)}{w_R'(m)} \\
&- u_0\left(\frac{w_R^2(m)}{w_R'(m)}\frac{1}{\sum_{j\in\mathcal{I}}\frac{1}{w_j'(s_j)}}x\right. \\
&\left.+\Phi(m)\frac{w_R^2(m)}{w_R'(m)} + \Phi(m)w_R(m)(m+1)\right),
\end{aligned}$$

and so

$$u_0(s, m, x) = \frac{x + \phi(s, m)\frac{w_R(m)}{w_R'(m)}}{x + \phi(s, m)\left(m + 1 + \frac{w_R(m)}{w_R'(m)}\right)},$$

with

$$\phi(s, m) = \left(2w_R'(m) - \frac{w_R(m)}{w_R'(m)}w_R''(m)\right)\left[\sum_{j\in\mathcal{I}}\frac{1}{w_j'(s_j)}\right].$$

C Computation of $\dot{H}_0$ along a $\mathcal{M}$ arc

Differentiating H_0 over an $\mathcal{M}$ arc gives

$$
\begin{aligned}
\dot{H}_0 =& \Big(w_R'(m)(x\lambda_x - (m+1)\lambda_m) - w_R(m)\lambda_m\Big)\dot{m} \\
&+ w_R(m)\dot{x}\lambda_x + w_R(m)x\dot{\lambda}_x - w_R(m)(m+1)\dot{\lambda}_m \\
=& \Big(w_R'(m)(x\lambda_x - (m+1)\lambda_m) - w_R(m)\lambda_m\Big) \\
&\times \left(\sum_{i=1}^{n} Y_i u_i w_i(s_i) - u_0 w_R(m)(m+1)\right) \\
&+ \cancel{w_R(m)u_0 w_R(m)x\lambda_x} \\
&+ w_R(m)x\left(\sum_{i=1}^{n} u_i w_i(s_i)\lambda_{s_i} - \cancel{u_0 w_R(m)\lambda_x}\right) \\
&- w_R(m)(m+1)u_0 w_R(m)\left(\lambda_m - \frac{w_R'(m)}{w_R^2(m)}H_0\right) \\
=& \Big(w_R'(m)\underbrace{(x\lambda_x - (m+1)\lambda_m)}_{H_0/w_R(m)} - w_R(m)\lambda_m\Big)\times \\
&\sum_{i=1}^{n} Y_i u_i w_i(s_i) + w_R(m)x\sum_{i=1}^{n} u_i w_i(s_i)\lambda_{s_i} \\
&- \Big(w_R'(m)\underbrace{\cancel{(x\lambda_x - (m+1)\lambda_m)}}_{H_0/w_R(m)}{}^{(2)} - \cancel{w_R(m)\lambda_m}^{(1)}\Big) \\
&\times u_0 w_R(m)(m+1) \\
&- w_R(m)(m+1)u_0 w_R(m)\left(\cancel{\lambda_m}^{(1)} - \frac{\cancel{w_R'(m)}}{\cancel{w_R^2(m)}}{}^{(2)}H_0\right) \\
=& w_R(m)\left(\frac{w_R'(m)}{w_R^2(m)}H_0 - \lambda_m\right)\sum_{i=1}^{n} Y_i u_i w_i(s_i) \\
&+ w_R(m)x\sum_{i=1}^{n} u_i w_i(s_i)\lambda_{s_i} \\
=& w_R(m)\frac{w_R'(m)}{w_R^2(m)}H_0\sum_{i=1}^{n} Y_i u_i w_i(s_i) \\
&- w_R(m)\underbrace{\sum_{i=1}^{n} u_i \underbrace{w_i(s_i)(Y_i\lambda_m + x\lambda_{s_i})}_{H_i}}_{H_i} \\
=& w_R(m)\left(H_0\frac{w_R'(m)}{w_R^2(m)}\sum_{i\in\mathcal{I}} Y_i u_i w_i(s_i) - H_i\right).
\end{aligned}
$$