Optimal feedback control strategies in a generic class of bacterial growth models with multiple substrates

Agustín Gabriel Yabo

Optimal feedback control strategies in a generic class of bacterial growth models with multiple substrates ⋆

Agustín G. Yabo [a,b]

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

[b] *Université Côte d'Azur, Inria, INRAE, CNRS, Sorbonne Université, Biocore Team, Sophia Antipolis, France*

Abstract

Optimal control strategies are studied through the application of the Pontryagin's Maximum Principle for a class of nonlinear 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 $u \in [0, 1]$ modeling the resources assigned to two main cellular functions, and n additional controls $\delta_i \geq 0$ 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—under certain conditions—they can be expressed in feedback form. The optimal solutions 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. Additionally, we provide an application to batch cultivation of E. coli growing on glucose and lactose by calibrating 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,

⋆ 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).

diauxic growth is a behavior that can potentially be predicted through optimization, provided that the adequate cost function is chosen. Previous studies 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. 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 $u(t)$ representing how resources are assigned to each cellular function, as a generalization of previous bacterial growth models [2,11]. Additionally, n uptake controls $\delta_i(t)$ decide whether the cell resources responsible for substrate uptake act upon one specific nutrient, or consume multiple substrates simultaneously—which is an extension of the preliminary results presented in [12]. The problem of finding the optimal control functions $(u, \delta_1, \dots, \delta_n)$ 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 that for "long-term" processes, they can be written as functions of the state of the system (i.e. feedback 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 theory that diauxie is a naturally-evolved control system that maximizes growth rate. 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. Finally, in Section 4, we show the trajectories obtained with an optimal control solver, as well as a comparison 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 $Y_1, Y_2, \dots, Y_n$. 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 $v_{M,i}$, 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 population, 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 $w_I(x) : \mathbb{R}_+ \to \mathbb{R}_+$ are*

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

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 $u(t) \in [0, 1]$ 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, $1 - u(t)$ accounts for the remaining fraction of the cell which is responsible for the uptake of sources s_i from the

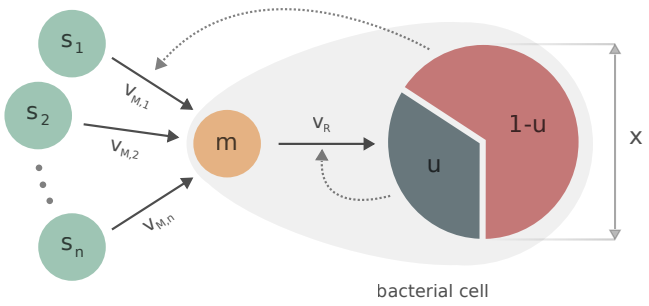

Fig. 1. Scheme of the system. Each external substrate s_i is metabolized into intermediate metabolites m at rate $w_{M,i}$. Then, the intermediate metabolites are used to synthesize proteins at rate w_R. The control $u(t)$ represents the fraction of the cell responsible for the synthesis of biomass (lower dotted line), while $1-u(t)$ represents the resources dedicated to the consumption of sources (upper dotted line).

medium, 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 rate of substrate uptake is linear in $1-u$, while the rate of protein production is linear in u. Figure 1 shows an schematic representation of the process here described. In this approach, cells are able to instantaneously stall growth ($u=0$) or substrate consumption ($u=1$), as well as to switch to balanced strategies between these two tasks ($0<u<1$). 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 remaining degrees of freedom are the distribution of resources used for substrate uptake, modeled through n time-varying functions $\delta_i \geq 0$, each responsible for regulating the rate of consumption and metabolization of the i-th substrate s_i. These control functions are subject to an inequality constraint modelling the availability limitation of the fraction of the cell responsible for substrate uptake:

$$\sum_{i=1}^{n} \delta_i \leq 1, \tag{C}$$

and they are implemented in the model by multiplying each substrate uptake rate (i.e. the terms containing $1-u$). This class of constraints can be found in previous works [17,18] to account for the limited number of enzymes in a bacterial cell. The resulting controlled model becomes

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

This way, the fraction of the cellular resources assigned to the uptake of the i-th substrate amounts to $\delta_i(1-u)$, which means that it is regulated through the combination of two controls.

2.2 Dynamics analysis

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

Lemma 1 *The set*

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

is positively invariant for the initial value problem.

Thus, we set initial conditions in Γ as

$$\begin{aligned} &s_i(0) = s_{i0} > 0, \quad i=1,2,\dots,n \\ &m(0) = m_0 > 0, \\ &x(0) = x_0 > 0. \end{aligned} \tag{IC}$$

Lemma 2 *System is upper bounded to $s_i \leq s_{i0}$, $m \leq \bar{m}$ and $x \leq \bar{x}$, with*

$$\bar{x} \doteq \sum_{i=1}^{n} Y_i s_{i0} + (m_0+1)x_0, \qquad \bar{m} \doteq \frac{\bar{x}}{x_0} - 1$$

PROOF. The lemma can be proved by observing that the quantity $\sum_{i=1}^{n} Y_i s_i + (m+1)x$ is constant and equal to $\bar{x}$ along a process starting in (IC). Then, evaluating the expression in the boundaries allows to obtain bounds on the states.

Lemma 3 *The substrate s_i and the intermediate metabolites m cannot be completely consumed in finite time.*

PROOF. Using the fact that there exists c_i and c_m such that $w_i(s_i) < c_i s_i$ for all s_i, and $w_R(m) < c_m m$ for

all m, we can bound the derivatives

$$\begin{aligned}\dot{s}_i &\geq -c_i s_i \bar{x}, \quad i = 1, 2, \ldots, n\\ \dot{m} &\geq -c_m m(\bar{m} + 1),\end{aligned}$$

showing that s_i and m decay, at worst, exponentially, and thus cannot attain 0 in finite time.

3 Optimal control problem

3.1 Problem statement

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 [19,20]. From a mathematical perspective, the latter is equivalent to say that the control functions $(u, \delta_1, \ldots, \delta_n)$ should maximize the biomass after a certain period of time. Thus, the objective function to maximize is $x(t_f)$ for a fixed time interval $[0, t_f]$. The problem writes

$$\begin{cases} \textit{maximize } x(t_f), \\ \textit{subject to dynamics of } (\mathrm{S}), \\ \qquad \textit{initial conditions } (\mathrm{IC}), \\ \qquad \textit{availability constraint } (\mathrm{C}), \\ \qquad u(\cdot) \in \mathcal{R}, \\ \qquad \delta_i(\cdot) \in \mathcal{D}, \quad i = 1, 2, \ldots, n \end{cases} \tag{OCP}$$

with $\mathcal{R}$ and $\mathcal{D}$ being the sets of admissible controllers, which are Lebesgue measurable real-valued functions defined on the interval $[0, t_f]$ and satisfying the constraints $u \in [0, 1]$ (for functions in $\mathcal{R}$) and $\delta_i \geq 0$ and (C) (for functions in $\mathcal{D}$). As (OCP) has no terminal conditions, there are no controllability and reachability issues. Additionally, controls are included in a compact and convex set and, as proved in Lemma 2, every trajectory of (S) is bounded. Thus, existence of solutions of (OCP) is guaranteed by Filippov's theorem [21]. Following PMP [22], we define the adjoint state $\lambda \doteq (\lambda_{s_1}, \lambda_{s_2}, \ldots, \lambda_{s_n}, \lambda_p, \lambda_x)$, and we write the Hamiltonian as

$$\begin{aligned} H &= \left(\sum_{i=1}^{n} H_i \delta_i\right)(1 - u) + H_u u \\ &= \left(\sum_{i=1}^{n} H_i \delta_i\right) + \left(H_u - \sum_{i=1}^{n} H_i \delta_i\right) u \end{aligned} \tag{H}$$

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

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

and the transversality conditions 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 we can fix $\lambda_0 < 0$. The dynamics of the adjoint system is given by

$$\begin{cases} \dot{\lambda}_{s_i} = -\delta_i \dfrac{w_i'(s_i)}{w_i(s_i)}(1 - u) H_i, \quad i = 1, 2, \ldots, n \\ \dot{\lambda}_p = w_R(m) u \left(\lambda_p - \dfrac{w_R'(m)}{w_R^2(m)} H_u\right), \\ \dot{\lambda}_x = \left(\displaystyle\sum_{i=1}^{n} \delta_i w_i(s_i)\right)(1 - u)\lambda_{s_i} \\ \qquad\qquad -w_R(m) u \lambda_x. \end{cases} \tag{AS}$$

Finally, for notation purposes, we define the function

$$\rho(s) \doteq \max(Y_1 w_M(s_1), Y_2 w_M(s_2), \ldots, Y_n w_M(s_n)) \tag{1}$$

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

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

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

3.2 Analysis of the optimal control solutions

The use of Pontryagin's Maximum Principle reduces problem (OCP) to finding the controls u and δ_i for $i = 1, 2, \ldots, n$ that maximize the Hamiltonian (H), subject to (S), (IC), (C), (TC) and (AS). As $u \in [0, 1]$, the Hamiltonian is a convex combination of H_u and $\sum_{i=1}^{n} H_i \delta_i$. This means the condition for maximizing the Hamiltonian depends on the sign of the difference between these two expressions, and so

$$u = \begin{cases} 0 \ \textit{if } H_u < \sum_{i=1}^{n} H_i \delta_i, \\ 1 \ \textit{if } H_u > \sum_{i=1}^{n} H_i \delta_i. \end{cases} \tag{2}$$

Analogously, the values of the metabolic controls δ_i depend on each associated function H_i, but obtaining a similar expression is not as straightforward due to constraint (C). Expression (2) defines a classical bang-bang control law, where the control takes the values of its lower and upper bounds. If $H_u = \sum_{i=1}^{n} H_i\delta_i$ over a subinterval $[t_1, t_2] \subset [0, t_f]$, an explicit expression of u cannot be directly obtained, and so further investigation of the PMP conditions is required. In optimal control problems where the Hamiltonian is linear in the control, this phenomenon results into what is typically called a *singular arc*, where the control usually takes intermediate values (i.e. not bang). In our case, the obtained Hamiltonian depends on the product between u and δ_i, which yields a more complex non-linear mathematical problem. However, throughout this paper we retain the classical terminology for "singular arcs". As evidenced in (2), every optimal control solution is a concatenation of bang arcs ($u = 0$ and $u = 1$) and singular arcs. We denote the possible arcs as:

- $\mathcal{G}$ **arc:** a pure-growth strategy given by $u = 1$, which occurs when $H_u > \sum_{i=1}^{n} H_i\delta_i$. In this case, the system does not depend on the functions δ_i, and so any combination of controls $(\delta_1, \delta_2, \ldots, \delta_n)$ is optimal.

- $\mathcal{S}$ **arc:** the singular arc, that occurs when $H_u = \sum_{i=1}^{n} H_i\delta_i$ along an interval of time $[t_1, t_2] \subset [0, t_f]$.

- $\mathcal{M}$ **arc:** given by $u = 0$, and thus all the resources are used for substrate consumption. This arc is optimal when $H_u < \sum_{i=1}^{n} H_i\delta_i$.

From a control theory point of view, expression (2) is an open-loop control law, as the values of u are not expressed exclusively in terms of the state $(s_1, \ldots, s_n, p, x)$ (but also in terms of the adjoint state). 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 control problem in order to obtain explicit expressions of the controls u and $(\delta_1, \ldots, \delta_n)$ in a feedback form. Let us first establish the positivity of the Hamiltonian and the final arc.

Lemma 4 *An optimal process should finish with a $\mathcal{G}$ arc, and should satisfy $H > 0$ for every t.*

PROOF. Evaluating the Hamiltonian at final time yields $H|_{t=t_f} = w_R(p(t_f))x(t_f)u(t_f)$. Using Proposition 3, we have $w_R(p(t_f))x(t_f) > 0$. As the optimal control at final time $u(t_f)$ should maximize the Hamiltonian, we see that $u(t_f) = 1$, and thus $H > 0$ for all t. Then, the final arc should be $u(t_f) = 1$.

Let us analyze the conditions for optimality along $\mathcal{S}$ and $\mathcal{M}$ arcs (i.e. the arcs enabling substrate uptake), where $u \in [0, 1)$. In this case, $H = \sum_{i=1}^{n} H_i\delta_i$, and so the i-th control is active (i.e. $\delta_i > 0$) or non-active (i.e. $\delta_i = 0$) depending on the value of its associated function H_i with respect to the others $(H_1, \ldots, H_n)$. We can see that, in order to comply with the positivity of the Hamiltonian, at least one function H_i should be positive. Let us denote $\mathcal{I}(t) \doteq \{j, \ldots, k\}$ with $\mathcal{I} \subset \{1, 2, \ldots, n\}$ the time-varying set of subindexes at time t that satisfy

$$\max(H_1, H_2, \ldots, H_n) = H_i,$$

which can contain a single or multiple elements. Analogously, we define $\mathcal{I}^C(t)$ as the complement of $\mathcal{I}$, such that $\mathcal{I}^C = \{1, \ldots, n\} \backslash \mathcal{I}$. Then, using Lemma 4 and (C), the maximization of the Hamiltonian occurs when

$$H = \sum_{i=1}^{n} H_i\delta_i = \max(H_1, H_2, \ldots, H_n) \left(\sum_{i \in \mathcal{I}} \delta_i \right).$$

From the latter expression, the controls δ_i should satisfy $\sum_{i \in \mathcal{I}} \delta_i = 1$, which then implies $H = H_i$ for every $i \in \mathcal{I}$. We formalize this results in the following theorem.

Theorem 1 *An optimal control solution should satisfy*

$$\sum_{i=1}^{n} \delta_i = 1$$

along any arc enabling substrate uptake (i.e. $u \neq 1$).

The latter theorem replaces the polyhedron defined by the constraint (C) with a more strict constraint given by a simple plane in $\mathbb{R}^n$. It is noteworthy that, so far, an active control δ_i implies that its associated function satisfies $H = H_i$. However, the inverse is not necessarily true: if $H = H_i$, then δ_i can be either active or non-active. Hereunder, we analyze the dynamics of each type of arc.

3.2.1 Dynamics of the arcs

$\mathcal{G}$ arcs are characterized by $H_u > \max(H_1, H_2, \ldots, H_n)$, which yields $H = H_u$. Along this arc, $s_i = s_i^*$ and $\lambda_{s_i} = \lambda_{s_i}^*$ are constant, and $\dot{m} \leq 0$. In $\mathcal{M}$ arcs, and as previously shown, there is at least one $H_i > 0$ solution of $\max(H_1, H_2, \ldots, H_n)$, and so the Hamiltonian becomes simply $H = H_i$ for all i in $\mathcal{I}$. In this case, $x = x^*$ and $\lambda_p = \lambda_p^*$, are constant, and $\dot{m} \geq 0$.

The analysis of singular arcs is more challenging. Again, we have $H = H_u = H_i$ for every $i \in \mathcal{I}$ over a subinterval $[t_1, t_2] \subset [0, t_f]$. This implies $\dot{H}_i = 0$, which yields

$$Y_i w_i(s_i) = \frac{w_R^2(m)}{w_R'(m)}, \quad i \in \mathcal{I}, \tag{3}$$

meaning that, along every singular arc, the state is on the region $\overline{\omega} \cup \overline{\omega}$ (as, a priori, there could be a non-active q-th control associated to a substrate satisfying $Y_q w_q(s_q) > Y_i w_i(s_i)$). Expression (3) also means

$$Y_j w_j(s_j) = \cdots = Y_k w_k(s_k), \tag{4}$$

where the subindexes $\{j, \ldots, k\}$ correspond to the elements of $\mathcal{I}$. Differentiating (4) yields $w_j'(s_j)\delta_j = \cdots = w_k'(s_k)\delta_k$. Then, by using Theorem 1, we can see that every control δ_i with $i \in \mathcal{I}$ should be active, and equal to

$$\delta_{i,\text{sing}}(s) = \frac{1}{\sum_{j\in\mathcal{I}} \frac{w_i'(s_i)}{w_j'(s_j)}}, \tag{5}$$

which is a feedback singular control depending only on substrate availability in the medium. Computing $\ddot{H}_i = 0$ yields

$$u_{\text{sing}}(s,p,x) = \frac{x + \phi(s,m)\frac{w_R(m)}{w_R'(m)}}{x + \phi(s,m)\left(m + 1 + \frac{w_R(m)}{w_R'(m)}\right)}, \tag{6}$$

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

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

and thus u_{sing} also depends only on the state. By writing $H_j = \cdots = H_k$, we find $w_j(s_j)\lambda_{s_j} = \cdots = w_k(s_k)\lambda_{s_k}$ or, using (4),

$$\frac{\lambda_{s_j}}{Y_j} = \cdots = \frac{\lambda_{s_k}}{Y_k}, \tag{7}$$

which, since $\dot{\lambda}_{s_i} < 0$, implies that $\lambda_{s_i} > 0$ for $i \in \mathcal{I}$.

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 and H_u. 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 and H_u on each arc. Across a $\mathcal{G}$ arc, we have

$$H_u > \max(H_1, H_2, \ldots, H_n),$$
$$\dot{H}_i = w_R(m)\left(H_i - Y_i w_i(s_i^*)\frac{w_R'(m)}{w_R^2(m)} H_u\right),$$

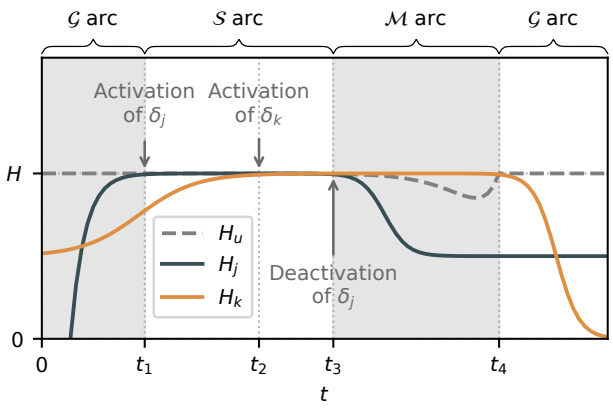

Fig. 2. Example of the evolution of functions H_i and H_u along a trajectory of the system. In $[0, t_1]$, every $H_i < H_u$, and so the system is on a $\mathcal{G}$ arc until H_j reaches $H = H_u$. At that point, a switch is produced to a $\mathcal{S}$ arc, and δ_j becomes active. At time t_2, δ_k also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

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 $\dot{H}_i < 0$. Analogously, across a $\mathcal{M}$ arc, we have

$$H_u < \max(H_1, H_2, \ldots, H_n),$$
$$\dot{H}_u = w_R(m)\left(H_u \frac{w_R'(m)}{w_R^2(m)} \sum_{i\in\mathcal{I}} Y_i \delta_i w_i(s_i) - H_i\right), \tag{8}$$

for $i \in \mathcal{I}$, and so, if $(s,p) \in \underline{\omega}$, then $\dot{H}_u < 0$. Additionally, there could exist a function H_q for $q \in \mathcal{I}^C$ (i.e. associated to a non-active control), and thus satisfying $H_q < H$. By computing $\dot{H}_q$, we can verify that, if such function exist, it is constant. In other words, every non-active control remains non-active along a $\mathcal{M}$ arc. However, along a singular arc $\mathcal{S}$, the function H_q for $q \in \mathcal{I}^C$ has dynamics

$$\dot{H}_q = w_R(m) u_{\text{sing}}(s,p,x)\left(H_q - \frac{Y_q w_q(s_q^*)}{Y_i w_i(s_i)} H_i\right), \tag{9}$$

and, as the q-th substrate is not being consumed, both $s_q = s_q^*$ and $\lambda_{s_q} = \lambda_{s_q}^*$ are constant. 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_u can present along an extremal. Below, we exploit the dynamics of these functions, and the transversality conditions (TC), to further investigate the possible structures of the optimal control solutions.

Proposition 1 *If every active control satisfies*

$$Y_i w_i(s_i) \leq \frac{w_R^2(m)}{w_R'(m)}, \quad i \in \mathcal{I}, \tag{10}$$

then a $\mathcal{M}$ arc is not admissible.

PROOF. Along a $\mathcal{M}$ arc, one has $\dot{m} > 0$, and every active control δ_i satisfies $\dot{s}_i \leq 0$. Then, condition (10) becomes $Y_i w_i(s_i) < w_R^2(m)/w_R'(m)$ along the arc for every $i \in \mathcal{I}$. Additionally, we can see that

$$\sum_{i \in \mathcal{I}} Y_i \delta_i w_i(s_i) \leq Y_k w_k(s_k) < \frac{w_R^2(m)}{w_R'(m)}$$

for any k-th substrate satisfying $Y_k w_k(s_k) \geq Y_i w_i(s_i)$ for $i \in \mathcal{I}$. Using (8), this means $\dot{H}_u < 0$, and thus the arc is optimal until $t_2 = t_f$. But this contradicts Lemma 4, as the process cannot end with a $\mathcal{M}$ arc, so the arc is not admissible.

Condition (10) is less strict than $(s, p) \in \underline{\omega}$, 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 2 *A $\mathcal{M}$ arc is only admissible at the beginning of the process.*

PROOF. In order to prove the proposition, let us show that a $\mathcal{M}$ arc cannot occur after any other arc:

- If the optimal control solution is at a $\mathcal{S}$ arc, with active controls δ_i with $i \in \mathcal{I}$ satisfying (3), the solution cannot switch to a $\mathcal{M}$ arc, as this contradicts Proposition 1.
- If the optimal control solution is at a $\mathcal{G}$ arc, every function H_i satisfies $H_i < H_u$. Additionally, $\dot{s}_i = 0$ for every $i = 1, 2, \ldots, n$ and $\dot{m} < 0$. Suppose that a function H_j becomes equal to H_u. This means that, before the switch time, $\dot{H}_j > 0$, which, according to (8), only happens if $Y_j w_j(s_j) \leq w_R^2(m)/w_R'(m)$. Again, this contradicts Proposition 1.
- If the optimal control solution is at a $\mathcal{M}$ arc, and H_u becomes equal to H, it means $\dot{H}_u > 0$ before the switch, and thus there should be at least one active substrate satisfying $Y_i w_i(s_i) \geq w_R^2(m)/w_R'(m)$. As, after the switch, $\dot{H}_u < 0$ in order to avoid entering a $\mathcal{S}$ arc, the substrate satisfies $Y_i w_i(s_i) = w_R^2(m)/w_R'(m)$ at the switch instant, and thus, according to Proposition 1, the $\mathcal{M}$ arc is not admissible.

Lemma 5 *Along a singular arc, the active controls are those associated to substrates solution of* (1)*, and so the state is in $\underline{\underline{\omega}}$.*

PROOF. Along a singular arc, the substrates associated to active controls satisfy (3). By way of contradiction, suppose at some time instant that the q-th substrate with $q \in \mathcal{I}^C$ satisfies $Y_q w_q(s_q^*) \geq Y_i w_i(s_i)$ for every $i \in \mathcal{I}$. Using (9), we have that $\dot{H}_q < 0$, and thus the q-th control remains non-active along the whole arc, which also means that $Y_q w_q(s_q^*) > w_R^2(m)/w_R'(m)$ for the remaining of the arc (as $w_R^2(m)/w_R'(m)$ is decreasing). Additionally, $H_q < H_i$, which means

$$w_q(s_q^*)\lambda_{s_q}^* \geq w_i(s_i)\lambda_{s_i} > 0 \tag{11}$$

along the arc, with $\lambda_{s_i} > 0$, as shown in (7). Using Proposition 2, it is easy to see that the q-th control remains non-active until the end of the process: the function $w_R^2(m)/w_R'(m)$ can only increase along a $\mathcal{M}$ arc, which cannot occur after any other arc. This means $Y_q w_q(s_q) = w_R^2(m)/w_R'(m)$ cannot be reached, and thus $\delta_q = 0$ until $t = t_f$. But, according to (TC), λ_{s_q} is null at t_f, and so it should be null along the whole arc, which contradicts (11).

Lemma 5 establishes a first clear relation between the consumption of the i-th substrate and its associated uptake control δ_i. In order to maximize its growth, the system should prioritize the metabolization of substrates solution of $\rho(s)$, which implies favoring the substrates that, for a given distribution of resources u, 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 [10], as well as with experimental observations of bacterial growth laws [13]. Another important consequence of Proposition 5 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 $\overline{\omega}$, which contradicts Lemma 5. 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 should be followed by the final $\mathcal{G}$ arc. The proof is based on Lemma 5:

when exiting a singular arc, the state is in $\overline{\underline{\omega}}$, which means that, as $\dot{m} < 0$, it enters $\overline{\omega}$. As $\overline{\underline{\omega}}$ cannot be reached along the $\mathcal{G}$ arc, the process finishes with the $\mathcal{G}$ arc (and, consequently, no other singular arc is allowed). Let us analyze the remaining admissible structures, taking into account Proposition 2.

- If the initial conditions are in $\overline{\omega}$:
  - and the initial arc is a $\mathcal{G}$ arc: the control is a single $\mathcal{G}$ arc and the state stays in $\overline{\omega}$.
  - and the initial arc is a $\mathcal{M}$ arc, the system should at some point reach $H_u = \max(H_1, H_2, \ldots, H_n)$ (according to Lemma 4). At that point, the solution can:
    switch to a $\mathcal{S}$ arc if the state is in $\overline{\underline{\omega}}$, and then to the final $\mathcal{G}$ arc.
    switch to a $\mathcal{G}$ arc if the state is in $\overline{\omega}$ (as $\dot{H}_i$ should be negative).
- If the initial conditions are in $\underline{\omega}$:
  - and the initial arc is a $\mathcal{M}$ arc: the control is a single $\mathcal{M}$ arc, which contradicts Lemma 4.
  - and the initial arc is a $\mathcal{G}$ arc, the system can
    stay in $\underline{\omega}$ or switch to $\overline{\omega}$, without reaching $H_u = \max(H_1, H_2, \ldots, H_n)$, and thus the control is a single $\mathcal{G}$ arc.
    reach $H_u = \max(H_1, H_2, \ldots, H_n)$, point at which the solution can:
      switch to a $\mathcal{S}$ arc if the state is in $\overline{\underline{\omega}}$, and then to the final $\mathcal{G}$ arc.
      switch to a $\mathcal{G}$ arc. Then, at the switch point, $\dot{H}_i = 0$, as it is positive before and negative after (otherwise, the arc becomes a $\mathcal{S}$ arc). Thus, at switch time, the system is in $\overline{\underline{\omega}}$, and so the state enters $\overline{\omega}$.

From a practical point of view, the control strategies that do not admit a singular arc are optimal when the state $\overline{\underline{\omega}}$ is not reachable. For example, an optimal solution with initial conditions in $\underline{\omega}$ that does not reach $\overline{\underline{\omega}}$ 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.

Assumption 2 *t_f and $(s_{10}, \ldots, s_{n0}, p_0)$ are chosen such that every optimal solution admits a singular arc.*

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 state is in $\underline{\omega}$, the initial arc is a $\mathcal{G}$ arc, and thus the δ_i controls do not intervene. If the initial conditions $(s_0, p_0) \in \overline{\omega}$, the initial arc is a $\mathcal{M}$ arc. Along an initial $\mathcal{M}$ arc, every function H_i for $i = 1, 2, \ldots, n$ is constant, and thus, the functions H_i for $i \in \mathcal{I}$ that are a solution of $\max(H_1, H_2, \ldots, H_n)$ in the initial arc, are also active in the following singular arc. However, along a $\mathcal{M}$ arc, an active control does not necessarily mean $\delta_i > 0$. 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 as long as the conditions for optimality are met. More precisely, any combination of controls δ_i satisfying Theorem 1 and taking the state from the initial condition (IC) to $\overline{\underline{\omega}}$ 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 (5).

3.4 Towards feedback control

The pool of intermediate metabolites 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. The final arc is optimal when the process has a fixed horizon, and its duration becomes smaller as $t_f \to \infty$ (and $m \to 0$ at the end of the singular arc). Thus, the final $\mathcal{G}$ arc can be neglected when using infinite time approaches (see [2] for a discussion of infinite horizons and overtaking optimality) aiming to study the "long-term" perspective of bacterial growth. In that case, the optimal control u depends only on the state—and not on the length of the bioprocess—and can be written in a feedback form. We formalize the latter in the following theorem.

Theorem 3 *The optimal long-term feedback control law for resource distribution is*

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

and the optimal i-th substrate uptake control is

$$\delta_i(s) = \frac{1}{\sum_{j\in\mathcal{I}} \dfrac{w_i'(s_i)}{w_j'(s_j)}},$$

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

It is noteworthy that a similar expression of the closed-loop control $\delta_i(s)$ can be found in simpler bacterial growth models [7], as well as in results from microeconomics theory [23]. 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. In next section, we validate the analytical results for fixed final time processes; and we study the particular case of a 2-substrate culture and compare the results to experimental data of a batch process of E. coli.

4 Numerical simulations

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). In the first part, we confirm the analytical findings by computing numerical optimal trajectories with Bocop [24]. The latter is an open-source toolbox that computes optimal trajectories using direct methods, by discretizing the time variable and then solving a finite-dimensional optimization problem that approximates the OCP. Then, we build a two-substrate model representing E. coli growing on glucose and lactose, and we match the results to experimental data. In this case, the optimal control is replaced by a closed-loop control as defined in Theorem 3.

4.1 General trajectories

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 $Y_1 w_1(s_{10}) > Y_2 w_2(s_{20}) > Y_3 w_3(s_{30})$. For the simulations performed with Bocop, 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 initial conditions 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

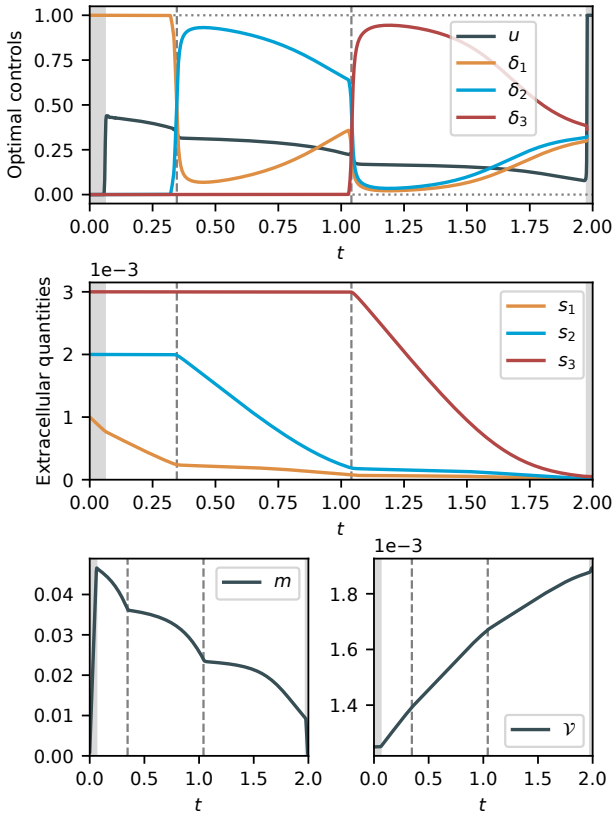

Fig. 3. Optimal trajectory of (OCP) obtained using Bocop. Parameter values are $k = 10\ \mathrm{h}^{-1}$, $K_R = 1$, $K_i = 0.1$, $Y_1 = 1$, $Y_2 = 0.5$ and $Y_3 = 0.2$ and initial conditions are set to $s_{10} = 0.001$, $s_{20} = 0.002$, $s_{30} = 0.003$, $m_0 = 0.001$ and $x_0 = 0.005$. The shaded regions at the beginning and end of each plot denote the time intervals where the optimal solution is either a $\mathcal{M}$ 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. Controls δ_i are not plotted along the final $\mathcal{G}$, as any combination of $(\delta_1, \delta_2, \delta_3)$ is optimal in that interval.

singular arc that represents more than 95% of the time interval. This results in an optimal solution following a structure $\mathcal{M} - \mathcal{S} - \mathcal{G}$. We can also see that the sequential activation of controls δ_i occurs when the previous substrate s_{i-1} attain very low concentration level, and is followed by a period of time where δ_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 of this paper. At $t = 0$, the solution of (1) is $Y_1 w_1(s_1)$ and H_1 is the solution of $\max(H_1, H_2, \ldots, H_n)$. Thus, δ_1 is the only active control, as predicted in Theorem 3, and it remains the only active control until the end of the initial $\mathcal{M}$ arc. As expected, H_u and m increase until $\overline{\omega}$ is reached, and H_u becomes equal to H_1. At that point, the solution switches to a singular arc, and continues to adopt the uptake control strategy $\delta_1 = 1$ until $Y_1 w_1(s_1) = Y_2 w_2(s_2)$, which

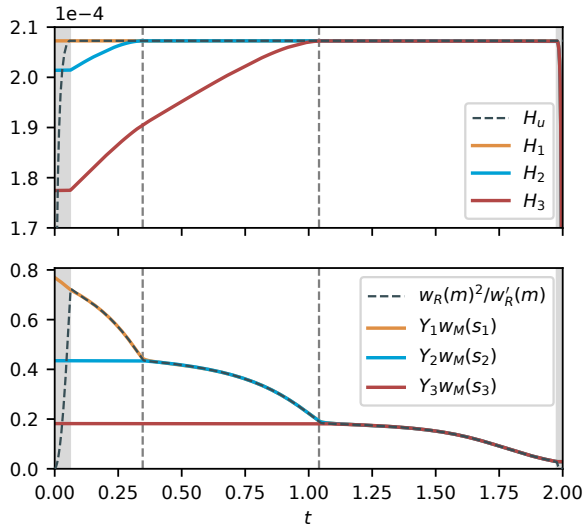

Fig. 4. Functions $w_R^2(m)/w'_R(m)$ and $Y_i w_i(s_i)$ (left-hand side plot) and H_u and H_i (right-hand side plot) related to the trajectories plotted in Figure 3. Shaded regions and dashed vertical lines indicate arcs and uptake control activation, respectively. Functions H_i associated to non-active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

occurs when H_2 reaches H, activating the second control δ_2. The same sequence occurs for the 3rd substrate, point at which all three substrates are being consumed simultaneously until the final switch to the $\mathcal{G}$ 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_u 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 as to start in $\underline{\omega}$, and thus to produce 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.

4.2 Case study: diauxic growth on glucose/lactose

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 optimal feedback control stated in Theorem 3. The objective of this section is to predict the natural behavior aforementioned using the optimality criteria obtained through PMP. The experimental data used to this end 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 $n = 2$,

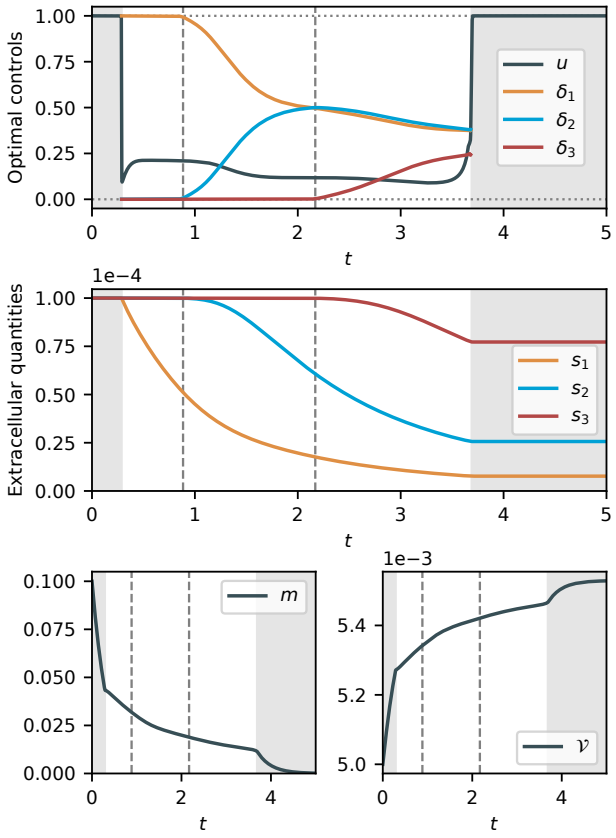

Fig. 5. Optimal trajectory of (OCP) obtained using Bocop. Parameter values are $k = 10\ \mathrm{h}^{-1}$, $K_R = 1$, $K_i = 0.05$, $Y_1 = 1$, $Y_2 = 0.3$ and $Y_3 = 0.1$ and initial conditions are set to $s_{10} = s_{20} = s_{30} = 0.0001$, $m_0 = 0.1$ and $x_0 = 0.02$. The shaded regions at the beginning 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. Controls δ_i are not plotted along the initial and final $\mathcal{G}$ arcs.

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 defined in Theorem 3, which accounts for long-term biomass maximization. The optimal substrate uptake in its two-dimensional form is given by

$$\delta_1(s) = \begin{cases} 1 & \text{if } Y_1 w_1(s_1) > Y_2 w_2(s_2), \\ 0 & \text{if } Y_1 w_1(s_1) < Y_2 w_2(s_2), \\ \dfrac{w_2'(s_2)}{w_1'(s_1) + w_2'(s_2)} & \text{if } Y_1 w_1(s_1) = Y_2 w_2(s_2). \end{cases}$$

and $\delta_2(s) = 1-\delta_1$, which yields the closed-loop bacterial growth model

$$\begin{cases} \dot{s}_1 = -\delta_1(s) w_1(s_1)(1-u(s,p,x))x, \\ \dot{s}_2 = -\delta_2(s) w_2(s_2)(1-u(s,p,x))x, \\ \dot{m} = \left[Y_1\delta_1(s)w_1(s_1) + Y_2\delta_2(s)w_2(s_2)\right](1-u(s,p,x)) \\ \qquad -w_R(m)u(s,p,x)(m+1), \\ \dot{x} = w_R(m)u(s,p,x)x. \end{cases} \tag{R}$$

The model was calibrated by adjusting the parameters $\theta = (k, K_1, K_2, Y_1, Y_2)$ 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 $s_1(t_1) = 6.522 \times 10^{-4}$, $s_2(t_1) = 3.42 \times 10^{-3}$, $p(t_1) = 0.01$ 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 δ_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, 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 $\delta_1 \approx 0$ after the switch, it cannot vanish exactly. Therefore, this behavior is not captured by our approach.

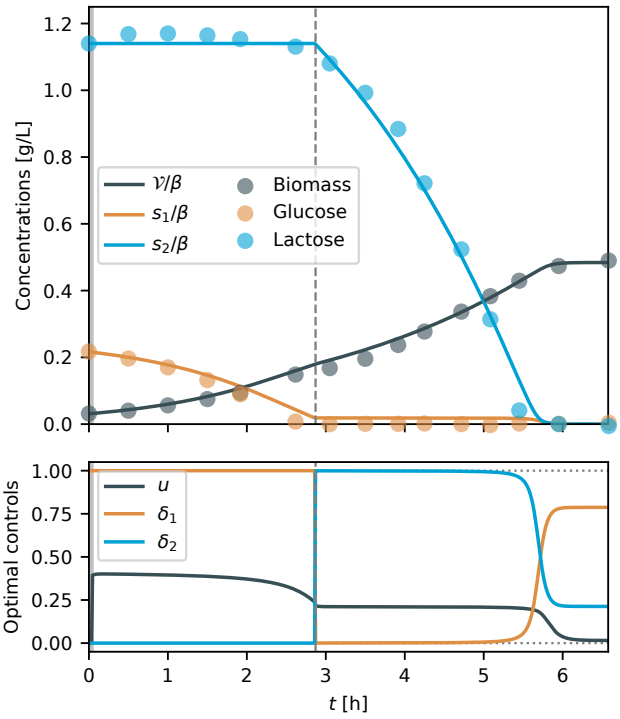

Fig. 6. Comparison with experimental measurements (of glucose, lactose and biomass) from [13]. In the top plot, the data from the experiment is represented in circles, while the simulation of model (R) is plotted in lines. The time-instant at which $Y_1 w_1(s_1)$ becomes equal to $Y_2 w_2(s_2)$ is marked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

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 arc, where cellular composition is governed by the expression defined in (6). Overall, the resources allocated to substrate uptake (represented by $1-u$) are gradually increased throughout the process, until the control u becomes almost null at the final time.

5 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 long-term 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. Finally, the model is calibrated to represent batch processing of E. coli on glucose and lactose. Despite 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 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 Gabriel 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 Gabriel 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 Gabriel Yabo. Predicting microbial cell composition and diauxic growth as optimal control strategies. 2022. Preprint.

[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 b 12 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] Guy C Brown. Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *Journal of theoretical biology*, 153(2):195–203, 1991.

[18] Edda Klipp, Reinhart Heinrich, and Hermann-Georg Holzhütter. Prediction of temporal gene expression: Metabolic optimization by re-distribution of enzyme activities. *European journal of biochemistry*, 269(22):5406–5413, 2002.

[19] Richard E Lenski and Michael Travisano. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, 91(15):6808–6814, 1994.

[20] 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.

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

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

[23] Richard J Herrnstein. Formal properties of the matching law 1. *Journal of the experimental analysis of behavior*, 21(1):159–164, 1974.

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