Optimal control of an artificial microbial differentiation system for protein bioproductioń

Elise Weill, Virgile Andreani, Chetan Aditya, Pierre Martinon, Jakob Ruess, Grégory Batt, Frédéric Bonnans

Optimal control of an artificial microbial differentiation system for protein bioproduction

Élise Weill[1,2], Virgile Andréani[1], Chetan Aditya[1], Pierre Martinon[2], Jakob Ruess[1] Grégory Batt[1,3], Frédéric Bonnans[2,3]

***Abstract*— The production of recombinant proteins is a problem of significant interest in bioengineering. Because of the existing trade-off between cellular growth and protein production, these two processes are separated in time in most commonly-employed strategies: a growth phase is followed by a production phase. Here, we investigate the potential of an alternative strategy using artificial cell specialization and differentiation systems in which cells either grow (”growers”) or produce proteins (”producers”) and growers can irreversibly ”differentiate” into producers. Inspired by an existing two-population system implemented in yeast, we propose a model of a ”yeast synthetic stem cell system” and define an optimal control problem to maximize bioproduction. Analytically, we first establish the well-posedness of the problem. Then, we prove the existence of an optimal control and derive non trivial optimality conditions. We finally use these results to find numerical optimal solutions. We conclude by a discussion of extensions of this work to models that capture the heterogeneity of the cell response to differentiation signals.**

I. Introduction

Microbes have a unique potential for bioproduction, with applications to the production of biofuels or to the synthesis of high added-value chemicals and therapeutic proteins [7], [3]. However, careful optimization of the production process is needed to obtain high quantities of the product of interest. Ideally, fermentation processes result in growing large quantities of cells, producing each high amounts of the target product. However, the energy and other cellular resources that are employed for bioproduction are not available to support cell growth. That is, there is a trade off at the cell level between cell population growth and production. Because this trade off results from highly complex intracellular regulations, it is hard to characterize and tame to optimize production [7].

To circumvent this problem, the strategy classically used in industry aims at dissociating in time growth from production. In a first phase, cells only grow. In a second phase, cells essentially produce the compound of interest and often grow significantly slower. Chemical induction is generally used to initiate the production phase. In this paper we investigate a different strategy in which growth and production take place simultaneously in the bioreactor but in different cells. More precisely, we envision the use of an artificial differentiation system made of two cell types. Cells of the first type produce the target product and do not grow. Cells of the second type either grow and do not produce, or switch to the first type, depending on an environmental signal. These two cell types will be referred to as *producers* and *growers*. In comparison to the classically induced cells, producers should be able to dedicate more cellular resources to the production of the molecule of interest since they do not grow.

In the following, we begin by establishing a model of the production of heterologous proteins for the differentiation and induction systems. We design optimal control problems related to the maximisation of the quantity of protein produced. We then derive from optimality conditions information on optimal controls. Using the optimal control solver Bocop, we find controls that maximise the amount of protein produced in the two systems and compare their performance.

II. Model development

In this section we introduce the models of the differentiation and induction systems we used. These systems are represented in Figure 1.

A. Model of differentiation system

Let us consider a cell culture, initially containing one type of cells, called *growers* and denoted by $\mathbf{g}$. These cells live in an environment containing nutrients (e.g. glucose), denoted by $\mathbf{s}$. We assume that the metabolic efficiency of cells depends on nutrients as follows:

$$\ell(s) = \frac{s}{K_s + s}, \tag{1}$$

where K_s is a half-velocity constant. Denoting μ the maximal growth rate and α the biomass yield, we assume the biomass growth rate to be $\mu\ell(\mathbf{s})$ and the nutrient consumption associated $\alpha\mu\ell(\mathbf{s})$. Thus, $\mu\ell(\mathbf{s})$ corresponds to a Monod growth rate. Grower cells differentiate into *producers* p at a rate $k_p\mathbf{u}$, where k_p is the maximal differentiation rate and $\mathbf{u} \in [0, 1]$ is the control. Producers synthesize proteins at a rate $k_q\ell(\mathbf{s})$, with k_q the maximal production rate. Thus, both grower and producer metabolic efficiencies *depend* on nutrient availability. However we assume that the quantity of nutrients *consumed* by producers for production is neglibible in comparison to the quantity of nutrients consumed for biomass growth. Therefore the dynamic of $\mathbf{s}$ does not depend on $\mathbf{p}$.

Two cell culture modes are considered here. In the first one, called *batch*, the environment is closed, so that there

[1]E. Weill, V. Andreani, C. Aditya, J. Ruess and G. Batt are with Inria Saclay Ile-de-France and Université Paris Saclay, and with Institut Pasteur and CNRS, C3BI - USR 3756, Paris, France

[2]E. Weill, P. Martinon, and F. Bonnans are with CMAP, Inria Saclay Ile-de-France and Ecole Polytechnique, CNRS, Université Paris Saclay, Saclay, France

[3]Equal contributions

is no nutrient addition. In the second one, called *chemostat*, the environment is renewed at a constant rate $\lambda > 0$.

The dynamics representing the system thus write as follows:

$$\begin{cases} \dot{\mathbf{q}} = k_q \ell(\mathbf{s})\mathbf{p} \\ \mathbf{q}(0) = 0, \end{cases} \tag{2}$$

$$\begin{cases} \dot{\mathbf{p}} = k_p \mathbf{u}\mathbf{g} - \lambda \mathbf{p} \\ \dot{\mathbf{g}} = -k_p \mathbf{u}\mathbf{g} + \mu \ell(\mathbf{s})\mathbf{g} - \lambda \mathbf{g} \\ \dot{\mathbf{s}} = -\alpha\mu \ell(\mathbf{s})\mathbf{g} + \lambda\,(s_{in} - \mathbf{s}) \\ \mathbf{p}(0) = 0,\ \mathbf{g}(0) = g_0,\ \mathbf{s}(0) = s_0. \end{cases} \tag{3}$$

We obtain the batch model when $\lambda = 0$.
The concentration of nutrients present in the media added in chemostat culture is equal to the initial amount of nutrients in the system, so $s_{in} = s_0$. Let us call T the duration of the experiment considered. Our purpose is to obtain the maximal amount of proteins $\mathbf{q}$ at time T. Since $\mathbf{q}(0) = 0$, we have:

$$\mathbf{q}(T) = \int_0^T \dot{\mathbf{q}}(t)\,\mathrm{d}t. \tag{4}$$

Therefore, maximizing $\mathbf{q}(T)$ is equivalent to minimizing the quantity $-\int_0^T \ell(s(t))p(t)\,\mathrm{d}t$. We can write the optimisation problem obtained for the differentiation model:

$$\min_{\mathbf{u}\in U} \left(-\int_0^T \ell(\mathbf{s}(t))\mathbf{p}(t)\,\mathrm{d}t \right) \quad \text{subject to (3)}, \tag{5}$$

where the space of admissible controls is

$$U := \{\mathbf{u} \in \mathbb{L}^\infty(0,T), 0 \le \mathbf{u}(t) \le 1\}.$$

B. Model of induction system

In order to evaluate the efficiency of our strategy based on artificial differentiation, we compare it to the classical strategy that relies on induction. In this system, cells grow, and upon induction, also produce the protein of interest. Control actions, $\mathbf{u} \in [0,1]$, modulate the intensity of the induction. In absence of signal of induction, that is when $\mathbf{u} = 0$, cells behave exactly as the grower cells $\mathbf{g}$ described in the differentiation model. When induction is present, that is $\mathbf{u} > 0$, protein $\mathbf{q}$ are produced by cells at a rate $k'_q \mathbf{u}\ell(\mathbf{s})$. To capture the existence of a trade-off between growth and production within our induced cells, we assume that growth is decreased by a factor $R_g\mathbf{u}$ with respect to pure growers, with $0 < R_g < 1$, and that production is less than that of pure producers (*ie,* $k'_q < k_q$). As before, we assume that only cell growth consumes nutrients.

The dynamics obtained write as follows:

$$\begin{cases} \dot{\mathbf{q}} = k'_q \mathbf{u}\ell(\mathbf{s})\mathbf{g}, \\ \mathbf{q}(0) = 0, \end{cases} \tag{6}$$

$$\begin{cases} \dot{\mathbf{g}} = (1 - R_g\mathbf{u})\,\mu\ell(\mathbf{s})\mathbf{g} - \lambda\mathbf{g} \\ \dot{\mathbf{s}} = -\alpha\mu\,(1 - R_g\mathbf{u})\,\ell(\mathbf{s})\mathbf{g} + \lambda(s_{in} - \mathbf{s}) \\ \mathbf{g}(0) = g_0,\ \mathbf{s}(0) = s_0. \end{cases} \tag{7}$$

The control problem for the induction system writes:

$$\min_{\mathbf{u}\in U} \left(-\int_0^T \mathbf{u}\ell(\mathbf{s})\mathbf{g}\,\mathrm{d}t \right) \quad \text{subject to (7)}. \tag{8}$$

All constants used in the two problems are positive.

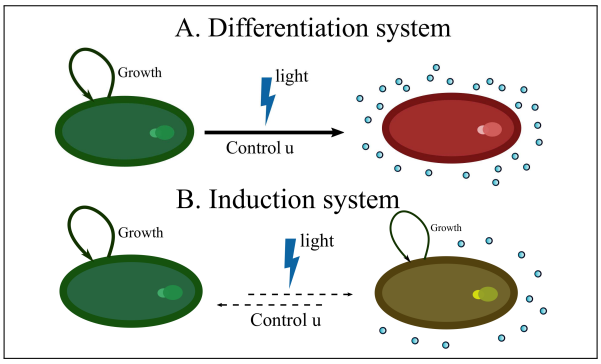

Fig. 1. Differentiation and induction systems. A. Under control (light), a fraction of growers *irreversibly* differentiates into producers, that synthesise the protein of interest. B. Uppon a gradded and *reversible* stimulation, all cells produce the protein of interest, at the cost of a lower growth rate.

III. Analytical study of the control problems

In this part, we study the optimal control problem on the system (5). Our main theoretical result is the fact that the number of switches of an optimal control is finite in the batch culture. We extend the results to the control problem related to the induction control problem (8).

A. Remark on the Monod function

We describe here some useful properties satisfied by the efficiency rate ℓ defined in (1).

Proposition 3.1: The following holds when $s \ge 0$: -

- $0 \le \ell(s) \le 1$.
- $\ell'(s) = K_s/(K_s + s)^2$, and $\ell(\mathbf{s})/\ell'(s) = s\,(K_s + s)\,/K_s$.
- $\ell''(s) = -2K_s/(K_s + s)^3) < 0$.

B. Energy of the system

We here introduce the quantity $\mathbf{y} = \alpha(\mathbf{g} + \mathbf{p}) + \mathbf{s}$, which can be interpreted as an abstract notion of energy, in nutrient equivalent. This energy corresponds to the total amount in the environment of unconsumed nutrient and of nutrients that has been consumed by cells for biomass growth. It will be used later on in the investigations about singular arcs in chemostat mode

Combining equations in (3), let us remark that $\mathbf{y}$ follows the following dynamics:

$$\dot{\mathbf{y}} = \lambda(s_{in} - \mathbf{y}). \tag{9}$$

Since $\mathbf{y}(0) = \alpha g_0 + s_0 := y_0$, we get:

$$\mathbf{y}(t) = s_{in} + (y_0 - s_{in})\exp(-\lambda t), \tag{10}$$

The energy $\mathbf{y}$ has a stable limit point, $\mathbf{y}_\infty = s_{in}$. This means that for long time, the energy present in the reactor is constant.

C. Positivity constraints

Since the variables of the systems studied represent quantities and concentrations, any solution of (3) should stay positive. In the sequel, we verify that this condition is automatically satisfied for any control $\mathbf{u} \in U$.

Theorem 3.2: Given $\mathbf{u} \in U$, the system defined in (3) admits only positive Lipschitz solutions defined on $(0, T)$.

Proof: We give here an overview of the key points of the proof. First, the dynamics (3) can be written :

$$\dot{\mathbf{Y}} = f(\mathbf{Y}, \mathbf{u}), \tag{11}$$

with f locally Lipschitz. Therefore, there exists τ so that there exist a maximal solution to (11). It can be proved that any solution $\mathbf{Y}$ remains positive over $(0, \tau)$. This implies that f has linear growth, so that $\dot{\mathbf{Y}}$ is bounded, and $\tau = T$. ∎

D. Existence of an optimal control

Theorem 3.3: There exists an optimal control $\mathbf{u} \in U$ to the problem (5).

Proof: This derives from the classical arguments, extracting from a minimizing sequence a weakly-* converging subsequence. The key points are the boundedness of the set of admissible controls U, and that solutions of the dynamics of (3) are bounded in $\mathbb{L}^\infty(0, T)$ by a constant depending only on the initial conditions. ∎

E. Optimality conditions

In this section, we apply Pontryagin Maximum Principle (PMP) in order to obtain optimality conditions. These conditions will be used in the subsequent sections.

Let us first write the Hamiltonian related to the differentiation model (5) :

$$\begin{aligned} H(p, g, s, p_p, p_g, p_s, u) \quad &= -\ell(s)p + p_p\left(k_p u g - \lambda p\right) \\ &+ p_g\left(-k_p u g + \mu\ell(s)g - \lambda g\right) \\ &+ p\left(-\alpha\mu\ell(s)g + \lambda(s_0 - s)\right) \end{aligned} \tag{12}$$

We derive the adjoint equations. In order to do so, let us introduce $\Delta^1_{\mathbf{p}} := \mathbf{p}_p - \mathbf{p}_g$ and $\Delta^2_{\mathbf{p}} := \mathbf{p}_g - \alpha\mathbf{p}_s$. The dynamics write:

$$\begin{cases} -\dot{\mathbf{p}}_p &= -\ell(\mathbf{s}) - \lambda\mathbf{p}_p \\ -\dot{\mathbf{p}}_g &= k_p\mathbf{u}\Delta^1_{\mathbf{p}} + \mu\ell(\mathbf{s})\Delta^2_{\mathbf{p}} - \lambda\mathbf{p}_g \\ -\dot{\mathbf{p}}_s &= -\ell'(\mathbf{s})\mathbf{p} + \mu\ell'(\mathbf{s})\mathbf{g}\Delta^2_{\mathbf{p}} - \lambda\mathbf{p}_s, \end{cases} \tag{13}$$

with final conditions

$$\mathbf{p}_p(T) = \mathbf{p}_g(T) = \mathbf{p}_s(T) = 0. \tag{14}$$

We also derive the dynamics of $\Delta^1_{\mathbf{p}}$ and $\Delta^2_{\mathbf{p}}$:

$$\begin{cases} \dot{\Delta}^1_{\mathbf{p}} &= \left(k_p\mathbf{u} + \lambda\right)\Delta^1_{\mathbf{p}} + \ell(\mathbf{s})\left(1 + \mu\Delta^2_{\mathbf{p}}\right), \\ \dot{\Delta}^2_{\mathbf{p}} &= -\alpha\ell'(\mathbf{s})\mathbf{p} + \left(\alpha\mu\ell'(\mathbf{s})\mathbf{g} - \mu\ell(\mathbf{s}) + \lambda\right)\Delta^2_{\mathbf{p}} \\ &\quad - k_p\mathbf{u}\Delta^1_{\mathbf{p}}, \end{cases} \tag{15}$$

and we introduce

$$\mathbf{\Delta} := 1 + \mu(\Delta^1_{\mathbf{p}} + \Delta^2_{\mathbf{p}}). \tag{16}$$

We next state the switch function and its first two derivatives:

$$H_{\mathbf{u}} = k_p\mathbf{g}\Delta^1_{\mathbf{p}}, \tag{17}$$

$$\dot{H}_{\mathbf{u}} \quad = k_p\ell(\mathbf{s})\mathbf{g}\mathbf{\Delta}, \tag{18}$$

$$\ddot{H}_{\mathbf{u}} \quad = k_p\left(\ell'(\mathbf{s})\mathbf{g}\mathbf{\Delta}\dot{\mathbf{s}} + \ell(\mathbf{s})\mathbf{\Delta}\dot{\mathbf{g}} + \ell(\mathbf{s})\mathbf{g}\dot{\mathbf{\Delta}}\right). \tag{19}$$

As expected (see e.g. [1]), $\dot{H}_u$ does not depend explicitly on u since the dynamics are affine with respect to the control.

When $H_{\mathbf{u}} = 0$ on an time interval (t_1, t_2), we have: By (17):

$$\Delta^1_{\mathbf{p}}(t) = \dot{\Delta}^1_{\mathbf{p}}(t) = 0. \tag{20}$$

In view of the first line of (15), we also deduce that:

$$1 + \mu\Delta^2_{\mathbf{p}} = 0 \text{ over } (t_1, t_2), \text{ thus } \dot{\Delta}^2_{\mathbf{p}} = 0. \tag{21}$$

By applying PMP, we can state that the value of any control $\mathbf{u}$ satisfying the necessary conditions is determined by the sign of $\Delta^1_{\mathbf{p}}$: whenever $\Delta^1_{\mathbf{p}}(t) > 0$, $\mathbf{u}(t) = 0$, and when $\Delta^1_{\mathbf{p}}(t) < 0$, $\mathbf{u}(t) = 1$.

In the sequel, we will prove that in batch cultures, that is when $\lambda = 0$, an optimal control is Bang-Bang with a finite number of switches. We will also show in the general case that the control must be maximized at final time T.

F. Non existence of singular arcs in batch mode

In this part, we suppose that $\lambda = 0$. We prove here the non existence of singular arcs.

Theorem 3.4: There exist no singular arc associated to the control problem (5) in batch culture ($\lambda = 0$).

Proof: Suppose that $H_{\mathbf{u}}(t) = 0$ on some interval (t_1, t_2). Then, for every $t \in (t_1, t_2)$, (20) and (21) hold. Combining with the second line of (15), we obtain

$$\mathbf{p} + \mathbf{g} = \frac{\ell(\mathbf{s})}{\alpha\ell'(\mathbf{s})} = \frac{\mathbf{s}}{\alpha K_s}(K_s + \mathbf{s}). \tag{22}$$

By (5), we have:

$$\dot{\mathbf{p}} + \dot{\mathbf{g}} = \mu\ell(\mathbf{s})\mathbf{g} > 0;\ \dot{\mathbf{s}} = -\alpha\mu\ell(\mathbf{s})\mathbf{g} < 0. \tag{23}$$

Therefore $\mathbf{p}$+$\mathbf{g}$ is increasing and $\mathbf{s}$ decreases, so equality (22) cannot hold on the entire interval (t_1, t_2). The conclusion follows. ∎

G. Singular arcs in chemostat mode

1) Sufficient condition of non existence of singular arcs in chemostat mode: In this part, $\lambda > 0$. We give here a condition on parameter values so that no singular arc exist in chemostat mode.

Theorem 3.5: If $\mu \leq \lambda$, then there exists no singular arc.

Proof: We suppose $H_{\mathbf{u}} = 0$ on an interval (t_1, t_2), then (20) and (21) hold. Replacing it in the equation of (15), we obtain:

$$-\alpha(\mathbf{p} + \mathbf{g}) + \frac{\ell(\mathbf{s})}{\ell'(\mathbf{s})} - \frac{\lambda}{\mu\ell'(\mathbf{s}))} = 0 \text{ on } (t_1, t_2). \tag{24}$$

Using the expression of ℓ and ℓ' in (1), and recalling that $\mathbf{y} := \alpha(\mathbf{p}+\mathbf{g})+\mathbf{s}$, we obtain:

$$\left(1-\frac{\lambda}{\mu}\right)\mathbf{s}^2 + 2K_s\left(1-\frac{\lambda}{\mu}\right)\mathbf{s} - K_s\mathbf{y} - \frac{K_s^2\lambda}{\mu} = 0, \quad (25)$$

Therefore, on a singular arc, $\mathbf{s}$ must satisfy a second order equation. The discriminant of this second degree equation writes:

$$\Delta_d \quad = 4K_s\left(1-\frac{\lambda}{\mu}\right)(K_s+\mathbf{y}). \quad (26)$$

Existence of a solution implies that the discriminant is nonnegative. So we must have $\mu \geq \lambda$. When $\lambda = \mu$, (25) writes:

$$\mathbf{y} = -K_s, \quad (27)$$

which is impossible since $\mathbf{y} > 0$. ∎

2) Value of the control on singular arcs: In this part, we give a formal formula of the optimal control on singular arcs. We suppose here that $\frac{\lambda}{\mu} < 1$.

Theorem 3.6: The following results hold on every singular arcs: i)

$$\mathbf{s} = -K_s + \frac{K_s}{\sqrt{1-\frac{\lambda}{\mu}}}\sqrt{1+\frac{\mathbf{y}}{K_s}}, \quad (28)$$

$$k_p\mathbf{u} = \frac{\ddot{\mathbf{s}}+\lambda\dot{\mathbf{s}}}{\lambda s_{in}-\lambda\mathbf{s}-\dot{\mathbf{s}}} + \frac{\ell'(\mathbf{s})}{\ell(\mathbf{s})}\dot{\mathbf{s}} + \mu\ell(\mathbf{s}) - \lambda. \quad (29)$$

ii) When the energy $\mathbf{y}$ has reached its equilibrium value, that is $\mathbf{y} = s_{in}$, on singular arcs, the states variables are constant, and $\mathbf{u}$ is equal to the following constant:

$$\mathbf{u} = \frac{\mu}{k_p}\left(1-\sqrt{\frac{1-\frac{\lambda}{\mu}}{1+\frac{s_{in}}{K_s}}}\right) - \frac{\lambda}{k_p}. \quad (30)$$

Proof: On any singular arc, the state $\mathbf{s}$ must satisfy (25). We thus obtain (28).

We notice that, by (3), $\lambda s_{in} - \lambda\mathbf{s} - \dot{\mathbf{s}} \neq 0$, otherwise we would have $\mathbf{g} = 0$. Using (3), we express $\mathbf{g}$ as a function of $\mathbf{s}, \dot{\mathbf{s}}$,differentiate and obtain (29).

To prove ii), we assume $\mathbf{y} = s_{in}$. By (28), we deduce that $\mathbf{s}$ must be constant, so $\dot{\mathbf{s}} = \ddot{\mathbf{s}} = 0$. We obtain the following expression of $\mathbf{u}$.

$$\begin{aligned} k_p\mathbf{u} \quad &= \mu\ell(\mathbf{s}) - \lambda \\ &= \mu\frac{\frac{1}{\sqrt{1-\frac{\lambda}{\mu}}}\sqrt{1+\frac{\mathbf{y}}{K_s}} - 1}{\frac{1}{\sqrt{1-\frac{\lambda}{\mu}}}\sqrt{1+\frac{\mathbf{y}}{K_s}}} - \lambda \\ &= \mu\left(1-\sqrt{\frac{1-\frac{\lambda}{\mu}}{1+\frac{s_{in}}{K_s}}}\right) - \lambda. \end{aligned} \quad (31)$$

∎

H. Finiteness of the set of switching points

1) A priori regularities and useful property: We recall that any function belonging in $\mathbb{L}^\infty(0,T)$ is well-defined for almost every $t \in (0,T)$. Therefore, before considering such functions at a precise time t, we have to analyse their regularities at this point.

In the following part, we will need to gain regularity results. In this scope, we recall the following property:

Lemma 3.7: Let $f \in \mathbb{L}^\infty(0,T)$ and $g \in \mathcal{C}^0(0,T)$. Let $t_0 \in (0,T)$.
i) If $g(t_0) = 0$, then for any bounded representative of f, $fg(t_0)$ is well-defined and continuous on t_0, and $fg(t_0) = 0$.
ii) If moreover g is differentiable at t_0 and $g'(t_0) = 0$, then $(fg)'(t_0)$ has a zero Frechet-derivative at t_0.

Proof: Continuity is directly obtained when h goes to zero in the following inequalities:

$$-C|g(t_0+h)| \leq fg(t_0+h) \leq C|g(t_0+h)| \quad (32)$$

If moreover g is differentiable at t_0, $g'(t_0) = 0$, by considering the following Taylor expansion and passing to the limit when h goes to zero, we obtain the expected result.

$$\begin{aligned} (fg)(t_0+h) - (fg)(t_0) &= f(t_0+h)g(t_0+h) \\ &= f(t_0+h)\,(o(h)) \end{aligned} \quad (33)$$

∎

2) Finiteness of switching points: In the sequel we show that any switching point is isolated. Since the time interval considered is bounded, the number of switches is finite.

Definition 3.1: A point t_0 is said to be critical if $H_{\mathbf{u}}(t_0) = 0$. It is said to be bicritical if moreover $\dot{H}_{\mathbf{u}}(t_0) = 0$.

Proposition 3.8: Any optimal control is Bang-bang (a.e. equal to one of its bounds), with finitely many switches.

Proof: In this proof we establish the following results: If t_0 is a bicritical point, then $\ddot{H}_{\mathbf{u}}$ is continuous at t_0, thus well-defined. If moreover $\ddot{H}_{\mathbf{u}}(t_0) = 0$, then $H^{(3)}_{\mathbf{u}}(t_0)$ is well-defined and negative. This implies the finiteness of switching points. Let t_0 be a bicritical point. By (17) and (18), this implies:

$$0 = \Delta^1_{\mathbf{p}}(t_0) = \dot{\Delta}^1_{\mathbf{p}}(t_0) = 1 + \mu\Delta^2_{\mathbf{p}}(t_0), \quad (34)$$

Therefore $\mathbf{\Delta}(t_0) = 0$. Since $\mathbf{\Delta}$ is continuous at t_0 applying lemma 3.7 (i) to $\dot{\mathbf{g}}\mathbf{\Delta}$ in (19), we obtain that $\ddot{H}_{\mathbf{u}}$ is well-defined at t_0. Suppose $\ddot{H}_{\mathbf{u}}(t_0) = 0$, then we obtain in (19):

$$0 = \dot{\mathbf{\Delta}}(t_0) = \dot{\Delta}^1_{\mathbf{p}} + \dot{\Delta}^2_{\mathbf{p}}. \quad (35)$$

Using equality (34) in (15), we obtain $\dot{\Delta}^2_{\mathbf{p}}(t_0) = 0$, so that $\dot{\mathbf{\Delta}}(t_0) = 0$. Applying lemma 3.7 (i) to $(\mathbf{u}\Delta^1_{\mathbf{p}})$ in (15), we obtain that $\Delta^1_{\mathbf{p}}$ and $\Delta^2_{\mathbf{p}}$ are twice differentiable at t_0. Therefore $\mathbf{\Delta}$ is twice differentiable at t_0. Moreover, since $\mathbf{\Delta}(t_0) = \dot{\mathbf{\Delta}}(t_0) = 0$, applying lemma 3.7 (ii) to $\mathbf{\Delta}\dot{\mathbf{g}}$ in (19), we obtain that $\ddot{H}_{\mathbf{u}}$ is differentiable at time t_0, and

$$H^{(3)}_{\mathbf{u}}(t_0) = k_p\mu\ell(\mathbf{s}(t_0))\mathbf{g}(t_0)\ddot{\mathbf{\Delta}}(t_0). \quad (36)$$

Differentiating the dynamics(15) and using (34), we obtain $\ddot{\Delta}^1_{\mathbf{p}}(t_0) = 0$, thus, $\ddot{\mathbf{\Delta}}(t_0) = \ddot{\Delta}^2_{\mathbf{p}}(t_0)$, and at time t_0:

$$\begin{aligned} \ddot{\Delta}^2_{\mathbf{p}} \quad &= -\alpha\ell'(\mathbf{s})\dot{\mathbf{p}} - \alpha\ell''(\mathbf{s})\mathbf{p}\dot{\mathbf{s}} \\ &\quad + (\alpha\ell''(\mathbf{s})\mathbf{g}\dot{\mathbf{s}} + \alpha\ell'(\mathbf{s})\dot{\mathbf{g}} - \mu\ell'(\mathbf{s}))\Delta^2_{\mathbf{p}} \\ &\quad + (\alpha\ell'(\mathbf{s})\mathbf{g} - \mu\ell(\mathbf{s}))\dot{\Delta}^2_{\mathbf{p}} - k_p\dot{\overbrace{\mathbf{u}\Delta^1_{\mathbf{p}}}} \\ &= -\alpha\ell'(\mathbf{s})k_p\mathbf{u}\mathbf{g} - \alpha\ell''(\mathbf{s})(-\alpha\mu\ell(\mathbf{s})\mathbf{g})\mathbf{p} \\ &\quad - \alpha\ell''(\mathbf{s})\mathbf{g}(-\alpha\mu\ell(\mathbf{s})\mathbf{g}) - \alpha\ell'(\mathbf{s})(-k_p\mathbf{u}\mathbf{g} + \mu\ell(\mathbf{s})\mathbf{g}) \\ &\quad + \ell'(\mathbf{s})(-\alpha\mu\ell(\mathbf{s})\mathbf{g}) \\ &= \alpha^2\mu\ell(\mathbf{s})\ell''(\mathbf{s})\mathbf{g}(\mathbf{g}+\mathbf{p}). \end{aligned} \quad (37)$$

By 3.1, $\ell''(\mathbf{s}) < 0$, $\ddot{\mathbf{\Delta}}(t_0) < 0$, and thus:

$$H^{(3)}_{\mathbf{u}}(t_0) < 0. \tag{38}$$

The conclusion follows.

■

I. Behaviour near the final time

We now come back to the general case, $\lambda \geq 0$. In this section, we prove the following result:

Theorem 3.9: Near the final time T, any optimal control takes the value 1

Proof: The sign of an optimal control $\mathbf{u}$ is determined by the sign of $\Delta^1_{\mathbf{p}}$ near the final time T. By the final conditions, $\Delta^1_{\mathbf{p}}(T) = 0 = \Delta^2_{\mathbf{p}}(T)$.
The dynamic of $\Delta^1_{\mathbf{p}}$ at time T thus writes:

$$\dot{\Delta}^1_{\mathbf{p}}(T) = \ell(\mathbf{s}(T)). \tag{39}$$

Since $\ell(\mathbf{s}(t)) > 0$ for any time, near final time T, the sign of the dynamics of $\Delta^1_{\mathbf{p}}$ is given by the sign of $\ell(\mathbf{s}(t)$, and more specifically it should be increasing.
Thus $\Delta^1_{\mathbf{p}}$ is negative, and $\mathbf{u}$ must be equal to 1 near final time T. ■

J. Induction model

The study of the induction model is quite similar as the one we did for differentiation model. Therefore, we should only outline the study.

The Hamiltonian related to the induction model writes:

$$\begin{aligned} H(g, s, p_g, p_s, u) \quad &= -u\ell(s)g \\ &\quad + p_g((1 - R_g u)\mu\ell(s)g - \lambda g) \\ &\quad + p_s(-\alpha\mu(1 - R_g u)\ell(s)g + \lambda(s_{in} - s)) \end{aligned} \tag{40}$$

The adjoint equations associated write:

$$\begin{cases} -\dot{\mathbf{p}}_g = -\mathbf{u}\ell(\mathbf{s}) + \mu\ell(\mathbf{s})(1 - R_g\mathbf{u})(\mathbf{p}_g - \alpha\mathbf{p}_s) - \lambda\mathbf{p}_g \\ -\dot{\mathbf{p}_s} = -\mathbf{u}\ell'(\mathbf{s})\mathbf{g} + \mu\ell'(\mathbf{s})\mathbf{g}(1 - R_g\mathbf{u})(\mathbf{p}_g - \alpha\mathbf{p}_s) - \lambda\mathbf{p}_s, \end{cases} \tag{41}$$

with final values $\mathbf{p}_g(T) = \mathbf{p}_s(T) = 0$. We define the variable $\Delta_{\mathbf{p}} := 1 + R_g\mu(\mathbf{p}_g - \alpha\mathbf{p}_s)$. The switching function is :

$$H_{\mathbf{u}} = -\ell(\mathbf{s})\mathbf{g}\Delta_{\mathbf{p}}. \tag{42}$$

The value of an optimal control at final time is here easily deduced. We here do not study more this control problem. We leave further analysis of this control problem for future work.

IV. Numerical analysis

A. Parameterization of differentiation and induction models

Cell densities in liquid cultures are generally expressed in units of optical density (OD600). In yeast cultures, 1 OD typically corresponds to $\sim 10^7$ cells/mL [6]. Here the inital amount of (grower) cells is set to 0.1 OD.
We assume that glucose is the major source of energy for growth and we set its initial concentration to a typical value of $s_0 = 10$ m.mL^{-1}. For the continuous culture mode, we set $s_{in} = s_0$. Experimental values found in the literature for K_s vary significantly, depending on the specific experimental conditions. We fixed it to $s_0/10$, so that growth in batch mode will be relatively fast until at least 90% of the nutriments are consumed.

Parameters	Values	Units
g_0	0.1	OD
s_0	10	mg.mL^{-1}
s_{in}	10	mg.mL^{-1}
K_s	$\frac{s_0}{10}$	mg.mL^{-1}
α	0.3	mg.mL^{-1}.OD^{-1}
μ	$\frac{\log(2)}{90}$	min^{-1}
k_p	2μ	min^{-1}
k_q	1	min^{-1}.OD^{-1}
k'_q	$k_q/2$	min^{-1}.OD^{-1}
R_g	0.1	-
λ	0.1	min^{-1}

TABLE I
PARAMETER VALUES FOR THE DIFFERENTIATION AND INDUCTION MODELS

The biomass yield parameter α corresponds to the amount of nutrients needed to make the cell biomass present in one mL at 1 OD. A maximum yield for yeast cells would correspond to approximately $\alpha = 0.15$ mg.mL^{-1}.OD^{-1} of glucose. We assume here that the yield is 50%, hence $\alpha = 0.3\, mg/mL/OD$. The maximal growth rate of cells μ is set such that the cell minimal generation time is 90 minutes, a value typical for yeast. To have a relatively efficient control of the differentiation, we set the maximal differentiation rate k_p to twice the maximal growth rate.
Lastly the specific value of k_q, in arbitrary protein concentration per time units and per cell, does not influence the optimization problem and we set it arbitrarily to 1.
The impact of exogenous protein production on cell growth, and conversely, is still relatively poorly characterized. Based on the results of Kafri and colleagues [5], we fixed the maximal growth decrease R_g to 10%. We also assumed that the capacity of production of an induced cell is half the capacity of a pure producer, that is, that $k'_q = k_q/2$.
In experiments in chemostat mode, the constant of renewal λ should be set carefully. It should neither be to small -otherwise it has the same behaviour as in batch-, nor too high -otherwise, yeasts cannot grow fast enough without being washed away.

B. Graphs of the optimal solutions for the different alternatives and comparison.

We now describe the numerical results obtained after optimisation by the solver *Bocop* [2]. Discrete equations were written with a midpoint scheme with 100 time steps, with an exception for the differentiation method in long time chemostat mode, where we used 10000 time steps.

In batch mode, for a 20 hours long experiment, the final quantity of proteins obtained with an optimal control for differentiation method is more than two times higher than what we obtain with induction (see Figure 2). We interpret this as follows. On the one side, induction provides a very limited control on growth and cannot limit the consumption of nutrients. On the other side, differentiation allows to tightly control the number of growers, and through that,

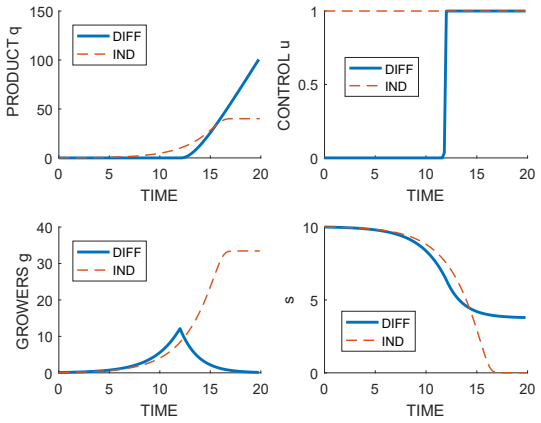

Fig. 2. **Batch mode,** 20 **hours**. Optimal solutions for induction and differentiation systems.

the quantity of nutrients that is consumed. Instead of fully exhausting nutrients, nutrients in the differentiation system are preserved, leading to a significantly longer period of efficient protein production.

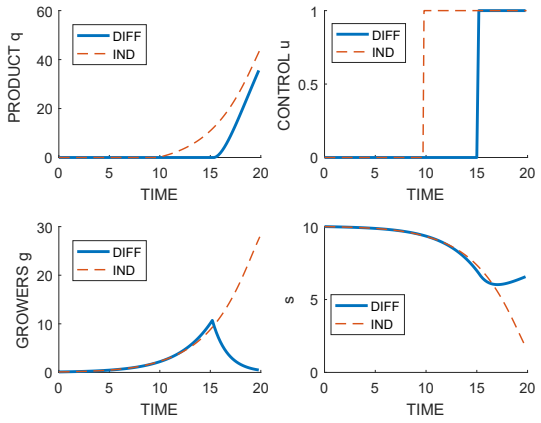

Fig. 3. **Chemostat mode,** 20 **hours**. Optimal solutions for induction and differentiation systems.

In chemostat mode, for $T = 20$ hours, induction method out competes the differentiation method (see Figure 3). In comparison to the previous setting, the lack of nutrients is tackled by nutrients renewal. Note also that no steady state is attained in 20 hours.

For long time experiments (here $T = 40$ hours), differentiation method appears to be more efficient than induction (see Figure 4). We also observe for differentiation system the existence of a singular arc. The value taken by the control on this singular arc is close to the value it should take in (30), when the energy of the system $\mathbf{y}$ is stationary. We think that this value corresponds to the best stationary trade-off between growth and production, that maximises the product of the efficiency rate by the number of producers. Therefore, with our parameter setting, differentiation appears to be more performing in batch for shorter times, and in chemostat for long times.

V. Conclusions

Through the help of control theory we obtained information on the differentiation control problem we proposed

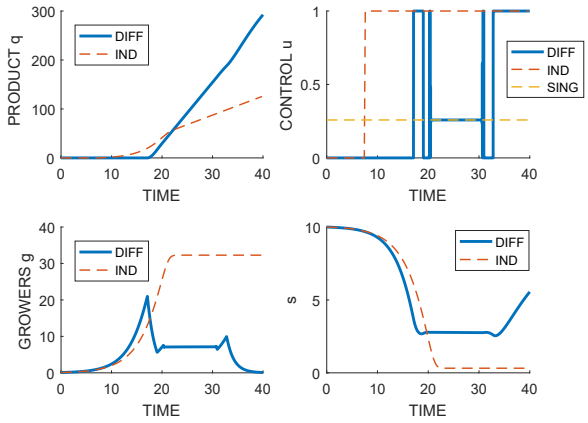

Fig. 4. **Chemostat mode,** 40 **hours**. Optimal solutions for induction and differentiation systems. The value of u on a singular arc in steady state is superposed to the graph. Since the control value cannot be deduced from the expression of $\dot{H}_u$ over a singular arc, we may expect a chattering phenomenon of the control at junctions with singular arcs of Fuller type [4]. And this is indeed what we observe.

to study. The numerical results we obtained are encouraging: in batch mode, differentiation seems to have better performances than induction. This also holds for long time experiments in chemostat mode. Since chemostat mode is suited for longer experiments, where media need to be renewed, these simulations speak in favour of differentiation. For future investigations, we plan to elaborate more complex models for differentiation system, taking in account the cellular heterogeneity. Indeed, in reality cells do not respond the same way to the differentiation signal. This will lead to an optimal control problem subject to constraints in the form of a mixed system of ordinary differential equations and partial differential equations.

Acknowledgments

This work has received funding from Inria (IPL Cosy), H2020 FET Open (COSY-BIO 766840) and Agence Nationale de la Recherche (ANR-16-CE12-0025).

References

[1] F. Bonnans, P. Rouchon, Commande et optimisation de systèmes dynamiques,2005, p.209-210
[2] Team Commands, Inria Saclay, BOCOP: an open source toolbox for optimal control, http://bocop.org, 2017.
[3] J.M. Carothers, J.A. Goler and J.D. Keasling Chemical synthesis using synthetic biology, Current Opinion in Biotechnology, 2009, 20(4):498:503
[4] Fuller, A. T., Study of an optimum non-linear control system, J. Electronics Control (1), vol 15, 1963, p63-71.
[5] Kafri, Moshe et al. "The cost of protein production." Cell reports 14.1(2016), p.22-31
[6] F. Margesin, Effect of temperature on growth parameters of psychrophilic bacteria and yeasts. Extremophiles. 2009 Mar13(2):257-62 p.261 fig.2 right bottom graphPubMed ID19057843
[7] D. Mattanovich, M. Sauer and B. Gasser, Yeast biotechnology: teaching the old dog new tricks, Microbial cell factories, 2014, 13(1):34
[8] F. Sherman, Getting started with yeast. Methods Enzymol. 2002 350: 3-41. p.15 table VPubMed ID12073320
[9] H. Youk, W.A. Lim, Science 343, 1242782 (2014). DOI:10.1126 / science.1242782