Optimal control of membrane filtration systems
Nesrine Kalboussi, Alain Rapaport, Térence Bayen, Nihel Ben Amar, Fatma Ellouze, Jérôme Harmand

To cite this version:
Nesrine Kalboussi, Alain Rapaport, Térence Bayen, Nihel Ben Amar, Fatma Ellouze, et al.. Optimal control of membrane filtration systems. IEEE Transactions on Automatic Control, Institute of Electrical and Electronics Engineers, In press, 64 (5), pp.8704-8709. <10.1109/TAC.2018.2866638>. <hal-01854430>
Optimal control of membrane filtration systems

N. Kalboussi, A. Rapaport, T. Bayen, N. Ben Amar, F. Ellouze and J. Harmand

Abstract—This paper studies an optimal control problem related to membrane filtration processes. A generic mathematical model of membrane fouling is used to capture the dynamic behavior of the filtration process which consists in the attachment of matter onto the membrane during the filtration period and the detachment of matter during the cleaning period. We consider the maximization of the net water production of a membrane filtration system (i.e. the filtrate) over a finite time horizon, where the control variable is the sequence of filtration/backwashing cycles over the operation time of process. Based on the Pontryagin Maximum Principle, we characterize the optimal control strategy and show that it exhibits a singular arc. Moreover we prove the existence of an additional switching curve before reaching the terminal state, and also the possibility of having a dispersal curve as a locus where two different strategies are both optimal.

Index Terms—Membrane filtration process, Physical backwash strategy, Optimal Control, Pontryagin Maximum Principle, Singular Arcs.

I. INTRODUCTION

Membrane filtration systems are widely used as physical separation techniques in different industrial fields like water desalination, wastewater treatment, food, medicine and biotechnology. The membrane provides a selective barrier that separates substances when a driving force is applied across the membrane. Different fouling mechanisms are responsible of the flux decline at constant transmembrane pressure (TMP) or the increase of the TMP at a constant flux. Hence, the operation of the membrane filtration process requires to perform regularly cleaning actions like relaxation, aeration, backwashing and chemical cleaning to limit the membrane fouling and maintain a good filtrate production. Usually, sequences of filtration and membrane cleaning are fixed according to the recommendations of the membrane suppliers or chosen according to the operator’s experience. This leads to high operational cost and to performances that can be far from being optimal.

A variety of control approaches have been proposed to manage filtration processes. In practice such strategies were based on the application of a cleaning action (physical or chemical) when either the flux decline through the membrane or the TMP increase crosses predefined threshold values ([12]). Smith et al. developed a control system that monitors the TMP evolution over time and initiates a membrane backwash when the TMP exceeds a given set-point, [23]. In [18] the TMP was also used as the monitoring variable but the control action was the increase or decrease of membrane aeration. The permeate flux was used in [25] as the control variable to optimize the membrane backwashing and prevent fouling. Moreover, knowledge-based controllers found application in the control of membrane filtration process. In [22], Robles et al. proposed an advanced control system composed of a knowledge-based controller and two classical controllers (on/off and PID) to manage the aeration and backwash sequences. The permeability was used by [11] as a monitoring variable in a knowledge-based control system to control membrane aeration flow. To date, different available control systems are able to increase significantly the membrane filtration process performances. However, more enhanced optimal control strategies are needed to cope with the dynamic operation of the purifying system and to limit membrane fouling. The majority of the control strategies previously cited address energy consumption, but regulation and control have not being proved to be optimal.

In the present work, we consider the maximization of the net production (i.e. the filtrate) per area of a membrane filtration system over a given operation duration. The control variable is the direction of the flow rate: forward for filtration through the membrane and backward for backwashing attached foulants. This problem is quite generic for various fluids to be filtered. Membrane fouling is assumed to be only due to the particle deposition onto the membrane surface while pores blocking is neglected. The aim of the present work is to give a complete optimal synthesis in a quite generic way (i.e. without giving the exact expressions of the functions involved in the model) characterizing the occurrence of such singularities. The analysis of these singularities (cf. [4]) is important for practical implementation because it gives the structure of the control strategies to be applied (how many switches, where or when to switch...) and the information (i.e. which variable and when) that is needed to be measured.

In Section II, we present the model and state the optimal control problem, with preliminary results about the structure of the optimal control near to the terminal time. Section III is devoted to the analysis of singular arcs (existence and optimality). In Section IV, we show that a switching curve appears and moreover that a phenomenon of dispersion along this curve may occur. We then give a complete optimal synthesis, which is illustrated in Section V on two filtration models of the literature.

II. MODEL DESCRIPTION AND PRELIMINARY RESULTS

We consider a simple form of the model of [3] to describe the membrane filtration process. In a previous work, it was
shown that this model is very generic in the sense that it is able to capture the dynamics of a large number of models available in the literature while simple enough to be used for optimizing and control purposes, see [15]. In the present work, it is assumed that the membrane fouling is only due to the particle deposition onto the membrane surface. Let \( m \) be the mass of the cake layer formed during the water filtration \((m \geq 0)\), that is assumed to follows a dynamics \( \dot{m} = f_1(m) \) with \( f_1 : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \). We further assume that the physical cleaning of the membrane is performed by a backwashing which consists in reversing the flow. During this phase, the filtration is stopped and the mass detaches from the membrane backwash, the cake layer is decomposed and the permeate flux decreases as the extent of fouling gradually increases. Therefore, the variation of the permeate flux \( J \) can be described by a decreasing positive function of the mass of the fouling layer.

Under Hypothesis 2.1, one obtains straightforwardly the following property.

Lemma 2.1: The domain \( \{ m > 0 \} \) is positively invariant whatever is the control \( u(\cdot) \).

For convenience, we define the functions \( f_+ : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) and \( f_- : \mathbb{R}_+ \rightarrow \mathbb{R} \) defined by

\[
 f_+(m) := \frac{f_1(m) + f_2(m)}{2}, \quad f_-(m) := \frac{f_1(m) - f_2(m)}{2},
\]

thus the dynamics can be equivalently written

\[
 \dot{m} = f_-(m) + uf_+(m), \quad u \in [-1, 1].
\]

We shall use the Maximum Principle of Pontryagin (PMP) [20] to obtain necessary conditions. Defining the Hamiltonian

\[
 H(m, \lambda; u) = \lambda f_-(m) + u(\lambda f_+(m) + g(m)),
\]

the PMP states\(^2\) that if a pair \((u, m)\) is optimal, then

i. there exists an absolutely continuous adjoint variable \( \lambda \) solution of the adjoint equation \( \dot{\lambda} = -\frac{\partial H}{\partial m} \) for a.e. \( t \in [0, T] \), with the terminal condition \( \lambda(T) = 0 \).

ii. the control \( u \) satisfies the Hamiltonian condition \( u(t) \in \arg\max_{\omega \in [-1, 1]} H(x(t), \lambda(t), \omega) \) for a.e. \( t \in [0, T] \).

iii. \( t \mapsto H(m(t), \lambda(t), u(t)) \) is constant

A triplet \((x(\cdot), \lambda(\cdot), u(\cdot))\) that verifies i-ii-iii is called an extremal, and the corresponding \( x(\cdot) \) an extremal trajectory.

Here, the adjoint equation is

\[
 \dot{\lambda} = -\lambda f_-(m) + u(\lambda f_+(m) + g'(m)), \quad \text{a.e. } t \in [0, T]
\]

with

\[
 u = \begin{cases} 
 +1 & \text{when } \phi(m, \lambda) > 0, \\
 -1 & \text{when } \phi(m, \lambda) < 0, \\
 \in [-1, 1] & \text{when } \phi(m, \lambda) = 0,
\end{cases}
\]

where \( \phi \) is the switching function

\[
 \phi(m, \lambda) := \lambda f_+(m) + g(m).
\]

Proposition 2.1: An extremal satisfies \( \lambda(t) < 0 \) for any \( t \in [0, T] \). Moreover, for any initial \( m_0 \) there exists \( \bar{t} < T \) such that the control \( u(t) = 1 \) is optimal for \( t \in [\bar{t}, T] \).

Proof: \( \lambda = 0 \) implies \( \phi(m, 0) = g(m) > 0 \), \( u = 1 \) and \( \dot{\lambda} = -g'(m) > 0 \). To have \( \lambda(T) = 0 \), \( \lambda \) has to be negative for \( t < T \). As \( t \mapsto \phi(m(t), \lambda(t)) \) is positive at \( T \), it remains positive on an interval \( [\bar{t}, T] \) and \( u = 1 \) on this interval.  

\(^1\)We recall that \( f \) from \( \mathbb{R}_+ \) to \( \mathbb{R}_+ \) if a \( K \)-function if it is increasing with \( f(0) = 0 \), and a \( L \)-function if it is decreasing with \( \lim_{x \to +\infty} f(x) = +\infty \).

\(^2\)As the terminal state is free, the abnormal multiplier \( \lambda_0 \) is not considered.
III. SINGULAR ARC AND FIRST OPTIMALITY RESULTS

For convenience, we define the functions
\[
\psi(m) := g(m) \left[ f'_-(m) f_+(m) - f_-(m) f'_+(m) \right] + g'(m) f_+(m) f_-(m)
\]
\[
\gamma(m) := -\frac{g(m) f_-(m)}{f_+(m)}
\]

We now consider the following hypothesis:

**Hypothesis 3.1:** The function \( \psi \) admits an unique positive root \( \bar{m} \) and is such that \( \psi(m)(m - \bar{m}) > 0 \) for any \( m \neq \bar{m} \).

Under Hypothesis 3.1, one can characterize \( m = \bar{m} \) as the unique candidate singular arc.

**Lemma 3.1:** Consider a singular arc defined over a time interval \([t_1, t_2]\). Then the corresponding singular extremal \((m(\cdot), \lambda(\cdot), u(\cdot))\) satisfies \( m(t) = \bar{m} \) and \( u(t) = \bar{u} \), \( t \in [t_1, t_2] \), where
\[
\bar{u} := -\frac{f_-(\bar{m})}{f_+(\bar{m})}.
\]

Moreover, \( \lambda(\cdot) \) is constant equal to \( \bar{\lambda} \) where \( \bar{\lambda} \in \mathbb{R} \) is defined by
\[
\bar{\lambda} = -\frac{g(\bar{m})}{f_+(\bar{m})}.
\]

**Proof:** One can easily check that it is possible to factorize \( \phi \) and \( \psi \) in the expression of \( \phi \) as follows:
\[
\phi = \frac{\psi}{f_+} + \phi \frac{f_+ f_- f'_+}{f_+} + g,
\]
where for simplicity we wrote \( \phi \) for the derivative of \( t \mapsto \phi(m(t), \lambda(t)) \) and we dropped the \( m \) dependency of functions \( f_- \), \( f_+ \) and \( g \). The conclusions follow from (7).

We show now that for certain initial conditions the optimal solution consists in a most rapid approach path to \( \bar{m} \). This property is also known as exact turnpike in the literature [21], [24], [9]. However, Lemma 2.1 shows that an optimal solution that reaches \( \bar{m} \) has to leave it before the terminal time, at a time \( \bar{T} \) that we make explicit in the following proposition.

**Proposition 3.1:** Suppose that Hypothesis 3.1 hold true and let \( m_0 > 0 \) be an initial condition. Then, the following properties are satisfied:

i. When \( m_0 < \bar{m} \), the control \( u = 1 \) is optimal as long as the corresponding trajectory satisfies \( m(t) < \bar{m} \).

ii. When \( m_0 > \bar{m} \), either the control \( u = 1 \) is optimal until \( t = T \), or the control \( u = -1 \) is optimal until a time \( \bar{t} < T \) with \( m(\bar{t}) \geq \bar{m} \). If \( m(\bar{t}) > \bar{m} \) then \( u = 1 \) is optimal on \( [\bar{t}, T] \).

iii. Suppose that \( f_-(\bar{m}) \geq 0 \). Then, for any initial condition \( m_0 \geq \bar{m} \), an optimal control satisfies \( u = -1 \) over some time interval \([0, \bar{t}]\) with \( \bar{t} \in [0, T] \) and \( u = 1 \) over \([\bar{t}, T] \).

iv. Suppose that \( f_-(\bar{m}) < 0 \) and let \( \bar{T} \in \mathbb{R} \) be defined by
\[
\bar{T} := T - \int_{\bar{m}}^{m_T} \frac{dm}{f_1(m)} \quad \text{with} \quad m_T := g^{-1}(\gamma(\bar{m})).
\]
Then, if \( \bar{T} > 0 \), any singular trajectory is optimal until \( t = \bar{T} \).

**Proof:** From Hypothesis 3.1 and (7), we can deduce that when \( \phi(m) = 0 \) with \( m < \bar{m} \) then \( \phi < 0 \). This implies that \( \phi \) can change its sign only when decreasing. Therefore only a switching from \( u = 1 \) to \( u = -1 \) can be optimal in the domain \( \{m < \bar{m}\} \). When \( \phi(m) = 0 \) with \( m > \bar{m} \) then \( \phi > 0 \). This implies that \( \phi \) can change its sign only when increasing. Therefore, only a switching point from \( u = -1 \) to \( u = 1 \) can be optimal in the domain \( \{m > \bar{m}\} \).

Let us prove i. Take \( m_0 < \bar{m} \), and suppose that the control satisfies \( u = -1 \). It follows that the trajectory remains in the domain \( \{m < \bar{m}\} \). From Proposition 2.1, the trajectory necessarily has a switching at a time \( t_c \), (otherwise, we would have \( u = -1 \) until the terminal time \( t = T \) and a contradiction) implying \( \dot{\phi}(t_c) \geq 0 \). On the other hand, we deduce from (7) that \( \dot{\phi}(t_c) = \frac{\psi(m(t_c))}{f_+(m(t_c))} < 0 \) which is a contradiction. Hence, we must have \( u = 1 \) in the domain \( \{m < \bar{m}\} \).

The proof of ii is similar as in the domain \( \{m > \bar{m}\} \), any optimal trajectory has at most one switching from \( u = -1 \) to \( u = 1 \). It follows that only three cases may occur: either \( u = 1 \) is optimal over \([0, T]\), or the trajectory reaches \( m = \bar{m} \) at some instant \( \bar{t} < T \), or it has exactly one switching in the domain \( \{m > \bar{m}\} \) from \( u = -1 \) to \( u = 1 \).

Let us prove iii. If one has \( u = 1 \) at time 0, then the result is proved with \( \bar{t} = 0 \). Suppose now that one has \( u = -1 \) at time 0. We know that if the trajectory switches at some time \( \bar{t} \in [0, T] \) before reaching \( m = \bar{m} \), then one has \( u = 1 \) for \( t > \bar{t} \) and the result is proved. Suppose now that an optimal trajectory reaches the singular arc before \( t = T \) and that one has \( m(t) = \bar{m} \) on a time interval of non-null length.

Since the Hamiltonian is constant along any extremal, one must have \( H = \bar{\lambda} f_-(\bar{m}) \). Moreover, as the Hamiltonian at time \( T \) is given by \( H = g(m(T)) \), one should have \( \bar{\lambda} f_-(\bar{m}) = g(m(T)) > 0 \). As \( \bar{\lambda} < 0 \), we conclude that when \( f_-(\bar{m}) > 0 \), this situation cannot occur. Hence, a singular arc is not optimal.

Finally, let us prove iv and suppose that \( f_-(\bar{m}) < 0 \). Accordingly to Propositions 2.1 and 3.1, an optimal control satisfies \( u = 1 \) in a left neighborhood of \( t = T \). Let us compute the last instant \( \bar{T} < T \) (if it exists) until a singular arc is possible. From the previous analysis, we necessarily have \( u = 1 \) on \([\bar{T}, T]\). This imposes (using that the Hamiltonian is constant) the final state to be \( \bar{m}_T = m(T) \) as a solution of
\[
g(\bar{m}_T) = \bar{\lambda} f_-(\bar{m}) = \frac{-\bar{g}(\bar{m}) f_-(\bar{m})}{f_+(\bar{m})} = g(\bar{m}),
\]
which is uniquely defined as \( g \) is decreasing, \( \lim_{m \to +\infty} g(m) = 0 \) and \( \frac{-f_-(m)}{f_+(m)} \in [0, 1] \). This also imposes that the switching time \( \bar{T} \) can be determined integrating backward the Cauchy problem \( m' = f_1(m) \), \( m(\bar{T}) = \bar{m}_T \) until \( m(\bar{T}) = \bar{m} \), which amounts to have the expression (8).

Finally we show that an extremal trajectory leaving the singular arc \( m = \bar{m} \) at a time \( t < \bar{T} \) cannot be optimal. To do so, consider a trajectory \( m(\cdot) \) leaving the singular arc at a time \( t < \bar{T} \) (necessarily with \( u = 1 \) until the terminal time \( T \)). In particular, we have \( m(T) > \bar{m}_T \). Since the dynamics is \( m = f_1(m) \) with \( u = 1 \), the corresponding cost from time \( t \) can be written as follows:
\[
J_1(t) := \int_{m(t)}^{m(T)} g(m) f_1(m) \, dm = \int_{\bar{m}_T}^{m(T)} g(m) f_1(m) + \int_{\bar{m}_T}^{m(T)} g(m) f_1(m)
\]
be compared with the cost $J_u(t)$ of the singular arc strategy from time $t$ (i.e. $u = \bar{u}$ over $[t, \bar{T}]$ and then $u = 1$ over $[\bar{T}, T]$), which is equal to

$$J_u(t) := -\int_{\bar{m}}^{m(T)} \frac{g(m)}{f_1(m)} \, dm + \int_{\bar{m}}^{\bar{m}} \frac{g(m)}{f_1(m)} \, dm.$$ 

Thanks to (8) and $T - t = \int_{\bar{m}}^{m(T)} \frac{dm}{f_1(m)}$, we get

$$\bar{T} - t = (T - t) - \int_{\bar{m}}^{m(T)} \frac{dm}{f_1(m)} = \int_{\bar{m}}^{m(T)} \frac{dm}{f_1(m)} - \int_{\bar{m}}^{m(T)} \frac{dm}{f_1(m)} = \int_{\bar{m}}^{m(T)} \frac{dm}{f_1(m)}.$$ 

The difference of costs $d(m(T)) := J_1(t) - J_u(t)$ can be then written as:

$$d(m(T)) = \int_{\bar{m}}^{m(T)} \left( \frac{g(m)}{f_1(m)} + \frac{g(m) - g(\bar{m})}{f_1(m)} \right) \, dm,$$

Let us now study the behavior of $d$ as a function of $m(T)$. For convenience, we write $m$ in place of $m(T)$ and recall that $m \geq \bar{m}$ since $m(T) \geq \bar{m}$. By a direct computation, one has:

$$\delta'(m) = \frac{g(m) + \bar{\alpha}}{f_1(m)},$$

$$\delta''(m) = \frac{g(m)f_1(m) - (g(m) + \bar{\alpha})f_1'(m)}{f_1^2(m)},$$

where $\bar{\alpha} := \frac{g(\bar{m})}{f_1(\bar{m})}$. From this last expression, since $g' < 0$, one has at each $m > 0$:

$$\delta'(m) = 0 \implies \delta''(m) < 0.$$ 

Now, it is to be observed that $\delta(\bar{m}) = 0$ and that $\delta'(\bar{m}) = 0$ (from (9)). The previous analysis then shows that $\delta' < 0$ on $(\bar{m}, +\infty)$. It follows that $\delta$ is decreasing over $(\bar{m}, +\infty)$. Hence, we obtain that $d(m) < 0$ for any $m > \bar{m}$. To conclude, we have proved that $J_1(t) < J_u(t)$ at any $t \in [0, \bar{T})$, thus any singular trajectory is such that it is optimal to stay on the singular arc until $\bar{T}$ (and then use $u = 1$ up to $T$).

**Remark 3.1:** One can easily check that Hypothesis 3.1 implies that $(\bar{m}, \bar{u})$ is the unique steady state that maximizes the integrand $g(m^*)u^*$ among steady states $(m^*, u^*)$, and that the dynamics is strictly dissipative for the supply rate $w(m,u) = g(m)\bar{u} - g(m)u$ (with the storage function $S(m) = \int_{\bar{m}}^{m} \frac{-g(\bar{u})}{f_1(\bar{u})} \, \xi(t)$). The turnpike property of the optimal solution can be deduced from [10, Th. 2], but here we characterize explicitly the last switching time $\bar{T}$.

Considering the following disjoint sub-domains:

$$D_{-} := \{ (t,m) \in [0,T] \times [\bar{m}, \bar{m}] \},$$

$$D_{+} := \{ (t,m) \in [0,T] \times (\bar{m}, +\infty) \},$$

one deduces straightforwardly the following result.

**Corollary 3.1:** Under Hypothesis 3.1, are verified:

i. When $f_-(\bar{m}) \geq 0$, $u = 1$ is optimal at any $(t,x) \in D_{-}$. 

ii. When $f_-(\bar{m}) < 0$ with $T \leq 0$, where $T$ is defined in (8), $u = 1$ is optimal at any $(t,x) \in D_{-}$. 

iii. When $f_-(\bar{m}) < 0$ and $\bar{T} \in (0,T)$ with $\bar{T}$ defined in (8), 

$$u = \begin{cases} 
1 & \text{if } m < \bar{m} \text{ or } t \geq \bar{T}, \\
\bar{u} & \text{if } m = \bar{m} \text{ and } t < \bar{T}, 
\end{cases}$$

is optimal at any $(t,x) \in D_{-}$. 

iv. $D_{+}$ is optimally invariant (i.e. for any initial condition in $D_{+}$, an optimal trajectory stays in $D_{+}$ at any time).

**IV. SWITHCING LOCUS AND FULL SYNTHESIS**

**A. Study of the switching locus in $D_{+}$**

Accordingly to Proposition 3.1, a switching outside the singular arc can occur only in $D_{+}$ and from $u = -1$ to $u = 1$. When $f_-(\bar{m}) < 0$, consider the (possibly empty) curve in $D_{+}$:

$$C := \{ (\bar{T}(\bar{m}), \bar{m}) \mid \bar{m} \geq \bar{m} \text{ and } \bar{T}(\bar{m}) > 0 \},$$

where $\bar{T}: [\bar{m}, +\infty) \to \mathbb{R}$ is the function defined by

$$\bar{T}(\bar{m}) := T - \int_{\bar{m}}^{g^{-1}(\gamma(\bar{m}))} \frac{dm}{f_1(m)}, \quad \bar{m} \geq \bar{m}.$$ 

**Proposition 4.1:** Assume that Hypothesis 3.1 is fulfilled.

i. When $f_-(\bar{m}) \geq 0$, $u = 1$ is optimal for $(t,m) \in D_{+}$.

ii. When $f_-(\bar{m}) < 0$, one as:

a. If $C$ is empty, $u = 1$ is optimal for $(t,m) \in D_{+}$.

b. When $C$ is non-empty,

$$u = \begin{cases} 
-1 & \text{if } (t,m) \in W, \\
\bar{u} & \text{if } m = \bar{m} \text{ and } t < \bar{T}, \\
1 & \text{otherwise.}
\end{cases}$$

is optimal, where the domain $W$ is defined as

$$W := \{ (t,m) \in [0,T] \times [\bar{m}, +\infty) \mid t < \bar{T}(m) \}. \quad (13)$$

Furthermore, $C$ is tangent to the trajectory that leaves the singular arc at $(T, \bar{m})$ with the control $u = 1$.

**Proof:** Suppose that $f_-(\bar{m}) \geq 0$ and let us prove i. We only have to show that any optimal control satisfies $u = 1$ in $D_{+}$. In this case, we know that no singular arc occurs, therefore it is enough to exclude switching from $u = -1$ to $u = 1$ in $D_{+}$. Also, since one has $u = 1$ in a neighborhood of $t = T$, it is enough to consider terminal states $m_T \geq \bar{m}$.

By integrating backward the dynamics with $u = 1$, one has

$$\phi(m, \lambda) = g(m) + \lambda f_+(m)$$

$$= g(m) + (g(m_T) - g(m)) \frac{f_+(m)}{f_1(m)},$$

$$= \frac{f_+(m)}{f_1(m)} (g(m_T) - \gamma(m)), \quad (14)$$

remains positive. As $f_-(\bar{m}) \geq 0$, one has $\gamma(\bar{m}) \leq 0$. Notice also that for $m \geq 0$, one has

$$\gamma'(m) = -\frac{\psi(m)}{f_+(m)^{\frac{1}{2}}}, \quad (15)$$

so that $\gamma$ is increasing over $[0, \bar{m}]$ and decreasing over $[\bar{m}, +\infty)$. We deduce that $\gamma(m(t)) \leq 0$ for any $t \in [0,T]$. Consequently, $\phi$ cannot change its sign, and $u = 1$ is optimal at any time.

Suppose now $f_-(\bar{m}) < 0$ and let us prove ii. Again, we consider terminal states $m_T \geq \bar{m}$. Note that when $m_T = \bar{m}$, one has $g(m_T) = \gamma(\bar{m})$ by conservation of the Hamiltonian.
Consider now an initial state \( m_T > \bar{m}_T \) and the system backward in time with \( u = 1 \). If an optimal control is such that \( u = 1 \) until reaching the singular arc, we deduce (by (14))
\[
g(m_T) - \gamma(m_T) < g(\bar{m}_T) - \gamma(\bar{m}) = 0,
\]
(since \( g \) is decreasing). Thus, the switching function is negative when \( \bar{m} \) is reached backward in time. By the mean value Theorem, we conclude that there necessarily exists a switching at some value \( \bar{m} > m \) such that \( \gamma(m) = g(m_T) \), and accordingly to Proposition 3.1 this switching (from \( u = 1 \) to \( u = 1 \)) is unique. From the monotonicity of \( \gamma \) on \([\bar{m}, +\infty)\), for each \( m > \bar{m}, \bar{m} \) is uniquely defined by \( \bar{m} = \gamma^{-1}(g(m_T)) \), or reciprocally, for any \( m > \bar{m}, m_T \) is uniquely defined as a function of \( \bar{m}; m_T(\bar{m}) = g^{-1}(\gamma(m)) \) (as \( g \) is also a decreasing invertible function), with
\[
m_T'(\bar{m}) = \frac{\gamma'(\bar{m})}{g'(m_T(\bar{m}))} \geq 0. \tag{16}
\]
Then, the corresponding switching time \( \bar{T}(\bar{m}) \) satisfies
\[
T - \bar{T}(\bar{m}) = \int_{\bar{m}}^{m_T(\bar{m})} \frac{d}{f_1(m)}. \tag{17}
\]
If \( \bar{T}(\bar{m}) \leq 0 \) then no switch occurs at \( \bar{m} \), i.e. the constant control \( u = 1 \) is optimal from 0 to \( T \). It follows that if \( C \) is empty, then \( u = 1 \) is optimal in \( D_+ \), as was to be proved.

When a switching occurs, that is when \( C \) is non-empty, the previous analysis shows that it occurs on the curve of \( D_+ \) given by (10) and the corresponding switching times are given by (11) as was to be proved. The optimality of the feedback control (12) follows by noting that in \( D_+ \), optimal trajectories have at most one switching point \( u = 1 \) to \( u = 1 \) or from \( u = 1 \) to \( \bar{u} \).

Finally, the derivative of \( \bar{T} \) with respect to \( \bar{m} \) can be determined from expressions (17) and (16) as
\[
\bar{T}'(\bar{m}) = \frac{1}{f_1(\bar{m})} - \frac{m_T'(\bar{m})}{f_1(m_T(\bar{m}))} \frac{\gamma'(\bar{m})}{g'(m_T(\bar{m}))} = \frac{1}{f_1(\bar{m})} - \frac{1}{g'(m_T(\bar{m}))} \frac{\gamma'(\bar{m})}{\gamma(\bar{m})}.
\]

Remark 4.1: When \( f_-(\bar{m}) < 0 \) and \( \bar{T} > 0 \), \( (\bar{T}, \bar{m}) \) belongs to \( C \) which is then non-empty. \( C \) could be a set of disjoint curves (for instance if the function \( \bar{T} \) has several changes of sign). However, in the examples, we met, it is always a single curve (bounded or not), see Section V. Notice also that the map \( \bar{m} \mapsto \bar{T}(\bar{m}) \) has no a priori reason to be monotonic, as it is shown in examples in Section V.

**B. Dispersal locus**

When the set \( C \) is non-empty (under the condition \( f_-(\bar{m}) < 0 \)), consider the partition: \( C = C_s \cup C_d \) with
\[
C_s := \{(t, m) \in C; 1 + \bar{T}(m)f_2(m) > 0\},
\]
\[
C_d := \{(t, m) \in C; 1 + \bar{T}(m)f_2(m) \leq 0\}.
\]

**Corollary 4.1:** Assume that Hypothesis 3.1 is fulfilled with \( f_-(\bar{m}) < 0 \) and \( C \) defined in Proposition 3.1 non empty.

- \( C_s \) is not reduced to \( \{(\bar{T}, \bar{m})\} \) and is a switching locus.
- \( C_d \) (when it is non-empty) is a dispersion locus i.e. from any state in \( C_d \) the two trajectories
  1. with \( u = 1 \) up to the terminal time,
  2. with \( u = -1 \) up to reaching \( m = \bar{m} \) or \( C_s \),

are both optimal.

**Proof:** The domain \( W \) (when it is not empty) is exactly the set of points \( (t, m) \in D_+ \) for which the optimal control satisfies \( u = 1 \) (see Proposition 3.1). From such a state, the optimal trajectory has to leave the domain \( W \) (as \( \bar{m} \) is bounded from above by \( -f_2(\bar{m}) < 0 \) in this set) reaching either the singular arc or the set \( C \). At some point \( (t, m) \in \bar{C} \), an outward normal \( n \) to \( W \) is then given by
\[
n(t, m) = \left[\begin{array}{c} 1 \\ -\bar{T}'(m) \end{array}\right],
\]
and the velocity vectors \( v_-, v_1 \) for the control \( u = -1 \) and \( u = 1 \) respectively are
\[
v_-(t, m) = \left[\begin{array}{c} 1 \\ -f_2(m) \end{array}\right], \quad v_1(t, m) = \left[\begin{array}{c} 1 \\ f_1(m) \end{array}\right].
\]

Notice that by construction of the set \( C \), the velocity vector \( v_1 \) points outward of \( W \) at any point \((t, m) \in C \). Hence, the velocity vector \( v_- \) points outward when the scalar product \( n \cdot v_- \) is positive, that is when \((t, m) \in C_s \).

We consider now optimal trajectories that reach \( C \) from \( W \) and distinguish two cases.

1. At states in \( C_s \), the velocity vectors \( v_- \), \( v_1 \) both point outward of the set \( W \). Therefore an optimal trajectory reaching \( C_s \) with \( u = -1 \) leaves it with \( u = 1 \). Then, accordingly to Proposition 2.1, the optimal control stays equal to \( 1 \) up to the terminal time.

2. At states in \( C_d \), \( v_- \) points inward of \( W \) while \( v_1 \) points outward. Therefore an optimal trajectory cannot reach a point located on \( C_d \). From states in \( C_d \), there exist two extremal trajectories: one with \( u = 1 \) up to the terminal time, and another one with \( u = -1 \) up to the singular arc or to the curve \( C_s \) (accordingly to Propositions 2.1 and 3.1) and then \( u = 1 \) up to the terminal time. At these points, the Hamiltonian dynamics of \((m, \lambda)\) is an upper semi-continuous multi-valued map with convex compact values, and therefore its solutions are continuous with respect to the initial condition (see for instance [13]). Moreover, the value function of a Lagrange problem with smooth data is everywhere Lipschitz continuous (see for instance [1]). For initial condition on each side of \( C_d \), we have shown that the trajectory with \( u = -1 \) or \( u = 1 \), depending on the side, are optimal. Therefore, by continuity w.r.t. to the initial condition, the two extremal trajectories leaving the set \( C_d \) have the same (optimal) cost (and are thus optimal).

Finally, let us show that \( C_s \) is not reduced to a singleton. The state \((\bar{m}, \bar{T})\) belongs to \( C \) (as it is indeed a point where the switching function vanishes) but it also belongs to the singular arc \( m = \bar{m} \). So, there exists a trajectory with \( u = -1 \) that crosses \( C \) transversely at this point. By continuity of the
solutions with $u = -1$ w.r.t. the initial condition, we deduce that there exist locally other trajectories that cross the non-empty curve $\mathcal{C}$ transversely with the control $u = -1$.

Figure 1 illustrates the two kinds of points in $\mathcal{C}$.

![Figure 1. Switching point (left) versus dispersion point (right) on the set $\mathcal{C}$.](image)

**Remark 4.2:** For a problem with free terminal state and smooth data, the value function $V$ is Lipschitz continuous (see for instance [1]) but not necessarily differentiable. When there is no dispersion locus i.e. $C_d = \emptyset$, the extremals, or equivalently the characteristics of the HJB equation, do not intersect in the $(t, m)$ plane. Then the relation between the adjoint variable and the sub- and super-differentials of $V$ (see [26]) allows to show that $V$ is indeed differentiable. However, in presence of dispersion locus, the lack of uniqueness of optimal trajectories leads to the non differentiability of $V$ [14].

### C. Feedback synthesis

We reformulate Corollary 3.1 and Proposition 4.1 in terms of feedback $u: (t, m) \in [0, T] \times [0, +\infty) \mapsto u(t, m)$.

**Theorem 4.1:** Assume that Hypothesis 3.1 is fulfilled.

i. If $f_{-}(\bar{m}) \geq 0$ or $f_{+}(\bar{m}) < 0$ and $\mathcal{C}$ empty, then $u(t, m) = 1$ is optimal for any $(t, m) \in [0, T] \times (0, +\infty)$.

ii. If $f_{-}(\bar{m}) < 0$ and $\mathcal{C}$ non-empty, then

$$u(t, m) = \begin{cases} -1 & \text{if } (t, m) \in \mathcal{W}, \\ \bar{u} & \text{if } m = \bar{m} \text{ and } t < \bar{T}, \\ 1 & \text{otherwise} \end{cases}$$

is an optimal feedback, where $\mathcal{W}$ is defined in (13).

### V. TWO NUMERICAL CASE STUDIES

#### A. The Benyahia et al model

The next functions have been experimentally validated [3]:

$$f_1(m) = \frac{b}{e + m}, \quad f_2(m) = am, \quad g(m) = \frac{1}{e + m},$$

where $a$, $b$ and $e$ are positive numbers. One can check that Hypothesis 2.1 is fulfilled. A straightforward computation of the function $\psi$ gives the following expression

$$\psi(m) = \frac{a^2 e^2 m^2 + 2a^2 e^2 m^2 + a^2 e^2 m^2 + 2a^2 e^2 m^2 + 4a e m^3 - 4 a b e m + 4 a b m^2 + 2b^2}{4(e + m)^2},$$

and a further computation of its derivative

$$\psi'(m) = \frac{a^2 e^2 m + a e m}{2(e + m)^3},$$

which allows to conclude that $\psi$ is increasing on $\mathbb{R}_+$. As one has $\psi(0) = -(2a b e + b^2) / (4e^4) < 0$ and $\lim_{m \to +\infty} \psi(m) = +\infty$, we deduce that Hypothesis 3.1 is fulfilled. When $\psi$ is null for $m = \bar{m}$, one has

$$d(\bar{m}) = f_{-}'(\bar{m}) f_{+}(\bar{m}) - f_{-}(\bar{m}) f_{+}'(\bar{m}) = -\frac{g'(\bar{m}) f_{+}(\bar{m})}{g(\bar{m})} f_{-}(\bar{m}).$$

Therefore $f_{-}(\bar{m})$ and $d(\bar{m})$ have the same sign. A straightforward computation gives

$$d(m) = -\frac{ab(e + 2m)}{2(e + m)^2} < 0,$$

and thus one has $d(\bar{m}) < 0$. Therefore, from Proposition 3.1 and Corollary 4.1, there exists a singular arc when $\bar{T} > 0$ and a switching locus when $\bar{T}(\bar{m}) > 0$. Figure 2 gives the optimal synthesis, where $\mathcal{C}$ is entirely a switching locus (i.e. $\mathcal{C} = \mathcal{C}_s$).

![Figure 2. Optimal synthesis for model V-A with $a = b = e = 1$ and $T = 10$ hours. The set $\mathcal{W}$ is depicted in blue and the switching locus in yellow.](image)

#### B. The Cogan-Chellam model

The following functions have been proposed [7], [8]

$$f_1(m) = \frac{b}{e + m}, \quad f_2(m) = \frac{a m}{e + m}, \quad g(m) = \frac{1}{e + m},$$

where $a$, $b$ and $e$ are positive numbers. Clearly Hypothesis 2.1 is fulfilled. Moreover, one has

$$\psi(m) = \frac{(a b - b^2) - 2 a b e - 2 b^2}{4(e + m)^4},$$

Therefore, the function $\psi$ can have two changes of sign at

$$\bar{m}_- = \frac{b - \sqrt{2b^2 + 2ab}}{a}, \quad \bar{m}_+ = \frac{b + \sqrt{2b^2 + 2ab}}{a},$$

where $\bar{m}_-$ and $\bar{m}_+$ are respectively negative and positive numbers. One has also

$$\psi'(m) = \frac{a^2 e m + a e m + a b m}{2(e + m)^3},$$

which is positive. Therefore $\psi$ is an increasing function and Hypothesis 3.1 is fulfilled with $\bar{m} = \bar{m}_+$. Moreover one can write

$$f_{-}(\bar{m}) = -\frac{\sqrt{b^2 + 2ab}}{e + \bar{m}} < 0.$$
Then, as for the previous model, Proposition 3.1 and Corollary 4.1 allow to conclude that there exists a singular arc when \( T > 0 \) and a switching locus when \( T(\bar{m}) > 0 \). Figure 3 gives the optimal feedback synthesis, where \( C \) is spitted into two non-empty subsets \( C_s \) and \( C_d \).

![Figure 3. Optimal synthesis for model V-B with \( a = b = e = 1 \) and \( T = 40 \) hours. The set \( W \) is depicted in blue, the switching locus in yellow and the dispersion locus in gray.](image)

C. Discussion

Although the two models are very close and have similar optimal syntheses, a main difference occurs on the size and on the shape of the domain \( W \) where backwash has to be applied (see Figures 2, 3). In particular, its boundary \( C \) is entirely a switching curve in one case while most of it is a dispersal curve in the second case. This should give valuable information to the practitioners about when and how long backwashing (i.e. \( u = -1 \)) has to be applied out of the singular arc. Notice that the duration \( T - \bar{T} \) in (8) is independent of \( T \). For arbitrary large horizon \( T \), the simpler strategy which consists in reaching as fast as possible the singular arc and stay on it for ever gives a good approximation of the optimal average cost, which moreover converges to to the optimal value in infinite horizon. For the practical implementation of the optimal control (where only the values \( u = -1 \) and \( u = 1 \) can be applied) it is not possible to stay exactly on the singular arc \( m = \bar{m} \), but an approximation by a sequence of filtration/backwashing can be applied to stay on the vicinity of the singular arc. This sequence can be chosen so that the average value of \( m \) is \( \bar{m} \), which provides a good approximation of the optimal cost as it has been tested in [16], [17]. One may argue that the problem could be reformulated in discrete time where the time step is the smallest period of switching between filtration and backwashing that could be applied in practice. We believe that this approach gives less geometric insights of the nature of the optimal control than the continuous formulation. Moreover, computing the optimal cost in continuous time gives an upper bound of what could be intrinsically expected from the process, independently of its practical implementations.

VI. Conclusion

The Pontryagin Maximum Principle has been applied to a membrane filtration model and shows interesting results for maximizing the net water production (per filtrate). The optimal synthesis exhibits bang-bang controls with a most rapid approach to a singular arc and a switching curve before reaching the final time. We have shown that a dispersal locus may occur, leading to the non-uniqueness of optimal trajectories. Practically, the determination of the singular arc allows to compute a sequence of filtration/backwashing to stay about the singular arc, and the determination of the curve \( C \) provides the information about the domain where backwashing has to be applied. The synthesis also reveals that if one wants to implement a feedback controller, which is more robust than an open-loop one, the on-line measurement of the mass deposit \( m \) or of any invertible function of \( m \), such as the water flow rate, is crucial. The main advantage of the present analysis is to describe an optimal synthesis for a very large class of models relying on simple qualitative properties of the functions \( f_1 \), \( f_2 \) and \( g \).

Perspectives of this work are first to implement the optimal synthesis with real process constraints, and compare the water production (per filtrate) of the membrane filtration process with the classical operating strategies that are proposed in the literature and currently used. Extensions to other fluids or non constant TMP and consideration of multiple objectives (production and energy consumption) could be also the matter of future works, as well as possibilities of multiple singular arcs.

Acknowledgments. The authors thank the TREASURE research network\(^1\) for its financial support. The third author would like to thank the MIA Division of INRA and the MISTEA lab, Montpellier for providing him a half year delegation during the academic year 2017-2018.

REFERENCES


\(^1\)http://www.inra.fr/treasure