Introducing State Constraints in Optimal Control for Health Problems

Igor Kornienko\textsuperscript{a,b}, Luís Tiago Paiva\textsuperscript{a,b}, Maria do Rosário de Pinho\textsuperscript{a,b,*}

\textsuperscript{a}Faculdade de Engenharia, Universidade do Porto, 4200-464 Porto, Portugal.
\textsuperscript{b}Instituto de Sistemas e Robótica, 4200-464 Porto, Portugal.

Abstract
An optimal control problem with state constraints based on a SEIR model to control the spreading of infectious diseases is considered. The main purpose is apply novel theoretical results to successfully validate the numerical solution, computed via direct method. The problem has simple but yet interesting features that we explore in our analysis. Of particular interest is the fact that the state constraint is of order one and that the solution is normal.

Keywords: Optimal control, Maximum Principle, State Constraints, SEIR Model.

1. Introduction

In this paper we study a particular optimal control problem with state constraints. It is a modification of an optimal control problem proposed in [9] based on a well known compartmental model called SEIR model (Susceptible, Exposed, Infectious and Recovered).

In the literature SEIR models have been extensively used to study the spreading of infectious diseases. However, to the best of our knowledge, the introduction of state constraints to such models has not been considered. This contrasts with other health related problems (see, for example, [3] and [13]).

Our problem of interest was previously proposed in [1] where numerical results are compared with other problems differing from ours with respect to the constraints. However, in [1] no theoretical discussion was made and consequently no validation of the numerical solution was discussed, a gap which we amend here.

While our initial concern was to study vaccination strategies to control the spreading of a generic infectious disease when upper bounds are imposed on state variables (in this respect, see [1]), our main concern here is on properties of the optimal control that may be of help to validate numerical solutions. The analysis we present here is of interest since it provides a systematic way of validating numerical solutions of a particular class of problems.

* Corresponding author: Tel.: +351-22-508-1617; fax: +351-22-508-1443.
E-mail address: mrpinho@fe.up.pt
As we will show, our health problem can be rewritten as:

\[
\begin{aligned}
\text{(P)} \quad & \minimize l(x(T)) + \int_0^T L(x(t), u(t)) \, dt \\
\text{subject to} \quad & \dot{x}(t) = f(x(t)) + g(x(t))u(t) \quad \text{for a.e. } t \in [0, T], \\
& h(x(t)) \leq 0 \quad \text{for all } t \in [0, T], \\
& u(t) \in U \quad \text{for a.e. } t \in [0, T], \\
& x(0) = x_0, \\
& x(T) \in \mathbb{R}^n.
\end{aligned}
\]

Here the state \( x \) takes values in \( \mathbb{R}^n \) while the control \( u \in \mathbb{R}^k \) and \( U \), the control set, is a subset of \( \mathbb{R}^k \). As for the functions we have \( I : \mathbb{R}^n \to \mathbb{R} \), \( L : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R} \), \( f : \mathbb{R}^n \to \mathbb{R}^n \), \( g : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^k \) and \( h : \mathbb{R}^n \to \mathbb{R} \).

This problem has some simple yet interesting features that we shall explore. It is solved numerically by a direct approach. As in [1] we use ICLOCS (Imperial College London Optimal Control Software), a matlab interface calling the nonlinear programming solver IPOPT (Interior Point Optimizer). A remarkable feature of IPOPT is that it provides numerical multipliers for the discrete time problem, a feature that shall be useful for our analysis.

This paper is organized in the following way. In the next section we state auxiliary results concerning (P), including necessary conditions. The state constrained optimal control problem \((P_S)\) is described in Section 3. Theoretical analysis of this problem is conducted in Section 4. In Section 5 the numerical findings and comparison between computed and analytical values is presented. The final section contains the conclusions.

We finish this section with a short description of the notation used in this paper.

**Notation:** The inner product of two vectors \( x \) and \( y \) in \( \mathbb{R}^m \) is written as \( \langle x, y \rangle \). Also \( \| \cdot \| \) denotes the Euclidean norm or the induced matrix norm on \( \mathbb{R}^{p \times q} \). The space of continuous functions from \([0, T] \) to \( \mathbb{R} \) is \( C([0, T]; \mathbb{R}) \) and its dual, with the supremum norm, is denoted by \( C^\star([0, T]; \mathbb{R}) \). The set of elements in \( C^\star([a, b]; \mathbb{R}) \) which take nonnegative values on nonnegative valued functions in \( C([0, T]; \mathbb{R}) \) is \( C^\theta([0, T]; \mathbb{R}) \).

For (P), a pair \((x, u)\) comprising an absolutely continuous function \( x \) (state or trajectory) and a measurable function \( u \) (control), is called an **admissible process** if it satisfies all the constraints. In this paper the pair \((x^\star, u^\star)\) will always denote the solution of the optimal control problem under consideration.

We call an admissible process \((x^\star, u^\star)\) a **strong local minimum** of (P) if there exists \( \varepsilon > 0 \) such that \((x^\star, u^\star)\) minimizes the cost over all admissible processes \((x, u)\) such that
\[
|x(t) - x^\star(t)| \leq \varepsilon \quad \text{for all } t \in [0, T].
\]

2. Auxiliary Results

Next we briefly review some results on (P) that will be important in our setting.

Concerning the state constraints, an interval \([t_0^\sigma, t_1^\sigma]\) \subseteq [0, T] is called a **boundary interval** for a trajectory \( x \) if it is the maximal interval where \( h(x^\sigma(t)) = 0 \), \( \forall t \in [t_0^\sigma, t_1^\sigma] \). The points \( t_0^\sigma \) and \( t_1^\sigma \) are called **entry point** and **exit point**. An interval \([t_0^\sigma, t_1^\sigma]\) \subseteq [0, T] is an **interior interval** if \( h(x^\sigma(t)) < 0 \), \( \forall t \in [t_0^\sigma, t_1^\sigma] \). A point \( \sigma \in [0, T] \) is a **contact point** for \( x \) if it is an isolated point such that \( h(x^\sigma(\sigma)) = 0 \). Take any admissible process \((x, u)\) for (P). Set \( h^0(t, x, u) = h(x(t)) \) and \( h^1(x, u) = \frac{dh}{dt}(x(t)) \). With respect to the dynamics we have
\[
h^1(x, u) = \nabla_x h^1(x(t)) \dot{x}(t) = \left( \frac{\partial h}{\partial x}(x(t), f(x(t)) + g(x(t))u(t)) \right).
\]

If for all \( t \in [0, 1] \) we have \( \frac{\partial h^1}{\partial u}(x, u) = \left( \frac{\partial h}{\partial x}(x(t), g(x(t))) \right) \neq 0 \) then we say that the state constraint is of **order one**.

Let \((x^\star, u^\star)\) be a reference process for (P) and \( \varepsilon \) a given parameter. We impose the following condition on the data of (P).
H1. the function $L(x, \cdot)$ is continuous on $U$ for all $x \in \mathbb{R}^n$;

H2. the functions $f(\cdot), g(\cdot), L(\cdot, u)$ and $h(\cdot)$ are continuously differentiable on $x^r(t) + \epsilon B$ for all $u \in U$;

H3. the function $l$ is Lipschitz continuous on $x^r(T) + \epsilon B$;

H4. the set $U$ is compact.

We shall not dwell on the existence of solution for (P). For the sake of completeness, however, we claim that if H1–H4 are satisfied and an admissible solution exists, then Theorem 23.11 in [2] asserts that (P) has a solution.

Suppose that $(x^*, u^*)$ is a local strong minimum. Theorem 9.3.1 in [14] asserts that there exist an absolutely continuous function $p$, a scalar $\lambda$ and a measure $\mu \in \mathcal{C}(0, T]$ such that

(i) $(p, \lambda, \mu) \neq (0, 0, 0)$;

(ii) $-\dot{p}(t) = f_1^T(x^*(t))q(t) + u^*(t)g_1^T(x^*(t))q(t) - \lambda L(x^*(t), u^*(t))$;

(iii) $(g(x^*(t))u^*(t), q(t)) - \lambda L(x^*(t), u^*(t)) \geq (g(x^*(t))u, q(t)) - \lambda L(x^*(t), u), \forall u \in [0, 1]$;

(iv) $-\dot{q}(T) = 0$;

(v) $supp[\mu] \subset \{ t : h(x^*(t)) = 0 \}$,

where

$$ q(t) = p(t) + \int_{\mathcal{V}} \nabla h(x^*(s)) \mu(ds), \quad q(T) = p(T) + \int_{[0,T]} \nabla h(x^*(s)) \mu(ds). $$

The function $q$ is a bounded variation function.

3. SEIR Constrained Problem

Our problem of interest is an optimal control problem with a scalar state constraint proposed in [1]. It is based on the so called “SEIR” model (for more information on SEIR models, we refer the reader to [9] and the references within).

The SEIR model is a compartmental model dividing the total population $N$ into four different compartments relevant to the epidemic. Those compartments are susceptible ($S$), exposed ($E$), infectious ($I$), and recovered (or immunized by vaccination) ($R$). Those in the $S$ compartment are susceptible to catching the disease. A person who is infected with the disease but is currently in latency is in the $E$ compartment. Infectious individuals are in the $I$ compartment and immune ones are in the $R$ compartment. Any newborn is considered susceptible, a susceptible individual becomes exposed when in contact with infectious individuals by what is known as horizontal transmission (see [9] for a more complete description). The exposed ones may die of natural causes or become infectious. The infectious ones can either die or recover completely. Finally, all individuals who recovered (those susceptible who were vaccinated or those who recovered from the disease) are considered immune. Let $S(t), E(t), I(t)$, and $R(t)$ denote the number of individuals in all those compartment at time $t$. The total population is $N(t) = S(t) + E(t) + I(t) + R(t)$. Only susceptible individuals are vaccinated and the rate of vaccination is denoted by $u(t)$, taking values in $[0, 1]$ (we have $u(t) = 1$ if at a given instant all those susceptible are vaccinated). We assume that every vaccinated individual becomes immune, that is, an individual in the compartment $S$, treated with vaccine, proceeds to the $R$ compartment.

We shall look at the evolution of the disease over a certain period of time $T$. The parameters describing the population and the disease transmission, assumed constant over the period of time of interest, are the following. The birth rate of the population is $b$ while $d$ denotes the natural death rate. The rate at which the exposed individuals become infectious is $e$, $g$ is the rate at which infectious individuals recover and $a$ denotes the death rate due to the disease. The rate of transmission is described by the number of contacts between susceptible and infectious individuals. If $c$ is the incidence coefficient of horizontal transmission, such rate is $cs(t)I(t)$.

Taking all the above considerations into account we are led to the following dynamical system:

$$ S(t) = bN(t) - dS(t) - cS(t)I(t) - u(t)S(t) \tag{2} $$

$$ E(t) = cS(t)I(t) - (e + d)E(t) \tag{3} $$

$$ I(t) = eE(t) - (g + a + d)I(t) \tag{4} $$

$$ R(t) = gI(t) - dR(t) + u(t)S(t) \tag{5} $$

$$ \dot{N}(t) = (b - d)N(t) - aI(t) \tag{6} $$
with the initial conditions \( S(0) = S_0, \ E(0) = E_0, \ I(0) = I_0, \ R(0) = R_0 \) and \( N(0) = N_0 \).

We retain the cost functional which was introduced by Neill and Lenhart in [9] (and also in [1]):

\[
J(u) = \int_0^T \left( AI(t) + u^2(t) \right) dt.
\]

The differential equation for the recovered compartment \( R \) can be removed since the state variable \( R \) only appears in the corresponding differential equation and the number of recovered individual at each instant \( t \) is obtained from \( R(t) = N(t) - S(t) + E(t) + I(t) \). However, to count the number of vaccinated individuals we introduce an extra variable \( W \) and the differential equation \( W(t) = u(t)S(t) \) with the initial condition \( W(0) = 0 \). As far as the optimal control problem is concerned, this new differential equation is redundant.

We expand the model described hereto to accommodate state constraints. A first thought would be to keep a pointwise upper bound on the number of infectious individuals. However, this would be a state constraint of order one since the number of susceptible individuals will certainly increase given that any newborn is considered susceptible and infectious individuals will be driven down because of the upper bound on the number of susceptible individuals. Note that the number of susceptible individuals will certainly increase given that any newborn is considered susceptible but, after vaccination a susceptible individual becomes immune. The translation of the upper bound on the number of susceptible individuals into mathematical terms is the state constraint \( S(t) \leq S_{\max} \). Putting all together, our problem is then

\[
(P_3)
\begin{align*}
\text{Minimize} & \quad \int_0^T \left( AI(t) + u^2(t) \right) dt \\
\text{subject to} & \\
\dot{S}(t) & = bN(t) - dS(t) - cS(t)I(t) - u(t)S(t), \\
\dot{E}(t) & = cS(t)I(t) - (e + d)E(t), \\
\dot{I}(t) & = eE(t) - (g + a + d)I(t), \\
\dot{N}(t) & = (b - d)N(t) - aI(t), \\
W(t) & = u(t)S(t), \\
S(t) & \leq S_{\max}, \\
u(t) & \in [0, 1] \quad \text{for a.e. } t \in [0, T], \\
S(0) & = S_0, \ E(0) = E_0, \ I(0) = I_0, \ N(0) = N_0, \ W(0) = W_0.
\end{align*}
\]

This problem is in the form of \((P)\) as it can be seen by setting

\[
x(t) = (S(t), E(t), I(t), N(t)), \quad \tilde{A} = (0, 0, A, 0), \quad C = (1, 0, 0, 0),
\]

\[
A_1 = \begin{bmatrix}
-d & 0 & 0 & b \\
0 & -(e + d) & 0 & 0 \\
0 & e & -(g + a + d) & 0 \\
0 & 0 & -a & b - d
\end{bmatrix}, \quad B = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]

and defining \( f(x_0, x_1) = 0, \ L(x, u) = \langle \tilde{A}, x \rangle + u^2 \) and \( f(x, u) = f_1(x) + g(x)u \) where \( f_1(x) = A_1x + c(-S, S, 0, 0)^T, \ g(x) = Bx \) and \( h(x) = \langle C, x \rangle - S_{\max} = S - S_{\max} \) for some fixed \( S_{\max} > S(0) \).

Note that \((P_3)\) has free end states, a quadratic cost with respect to \( u \) and that the differential equation \( \dot{x}(t) = f_1(x(t)) + g(x(t))u(t) \) is affine in the control and nonlinear in the state \( x \) due to the term \( f_1 \).

4. Necessary Conditions for \((P_3)\)

We may omit the variable \( W \) throughout the rest of the paper because it is redundant in the forthcoming analysis. It reappears in the next section only to keep track of the number of vaccines used in the period of interest.

The initial conditions we shall work with are those given in Table 1 below. For those values it is simple to see that there are constants \( U_S, L_S, U_N, L_N, U_E, U_I \) such that \( \forall t \in [0, T] 0 < L_S \leq S(t) \leq U_S, 0 < L_N \leq N(t) \leq U_N, 0 \leq I(t) \leq U_I \) and \( 0 \leq E(t) \leq U_E \).
Now we concentrate on the necessary conditions given by (i)-(v) above. Consider \( q = (q_s, q_e, q_l, q_u) \) and analogously \( p = (p_s, p_e, p_l, p_u) \). It can be shown that Theorem 4.1 in [11] holds meaning that (i)-(v) hold with \( \lambda = 1 \). Those we have normality of the optimal solution.

Apply (i)-(v) to (P). If \( u^*(t) \in [0, 1] \), the Weierstrass Condition (iii) yields \( \langle g(x^*(t))u^*, q(t) \rangle - u^* u^2 \geq \langle g(x^*(t))u, q(t) \rangle - u^2 \) for all \( u \in [0, 1] \). Taking into account that \( g(x^*(t)) = (S^*(t), 0, 0, 0) \), we deduce that

\[
\frac{\dot{q}_s(t)}{S^*(t)} = -\frac{q_s(t)S^*(t)}{2}.
\]

Since \( u^*(t) \) may be 0 or 1 or in \( ]0, 1[ \), we conclude that

\[
u(t) = \max\left\{0, \min\left\{1, -\frac{q_s(t)S^*(t)}{2}\right\}\right\}.
\]

Suppose now that \( \left[t_0^*, t_1^*\right] \) is a boundary interval. Then for \( t \) on this interval we have \( S^*(t) = S_{\max} \) and consequently

\[
\dot{S}^*(t) = bN^*(t) - dS^*(t) - cS^*(t)I'(t) - u^*(t)S^*(t) = 0.
\]

It then follows that for \( t \in \left[t_0^*, t_1^*\right] \) we get

\[
u(t) = \frac{bN^*(t)}{S^*(t)} - d - cI^*(t).
\]

Recall now that \( q(t) = p(t) + \int_{[0,t]} \nabla h(x^*(t)) \mu(ds), \nabla h(x^*(t)) = (1, 0, 0, 0) \) and

\[
\int_{[0,t]} (1, 0, 0, 0) \mu(ds) = \left(\int_{[0,t]} \mu(ds), 0, 0, 0\right).
\]

Thus we have \( q_S(t) = p_S(t) + \int_{[0,t]} \mu(ds) \). Now we explore regularity properties of the multipliers. It is a simple matter to see that the main results in [12] apply and so we deduce that \( \mu \) is absolutely continuous w.r.t. Lebesgue measure. This means that there exists an integrable function \( v \) such that \( \int_0^t v(s) ds = \int_{[0,t]} \mu(ds) \). Consequently, \( q \) is absolutely continuous on \( [0, T] \) and \( \dot{q}_s(t) = \dot{p}_s(t) + v(t) \). It is now a simple matter to see that

\[
\dot{q}_s(t) = (d + cI^*(t) + u^*(t))q_s(t) - v(t) + cI^*(t)q_u(t)
\]

Let us now concentrate on a boundary interval. Taking the above expression of \( \dot{q}_s \), we deduce with the help of (8) and (9) that

\[
v(t) = -(d + cI^*(t) + u^*(t))q_s(t) - cI^*(t)q_u(t) + 2c \frac{\dot{I}^*(t)}{S_{\max}} - \frac{2bN^*(t)}{S_{\max}^2}.
\]

The function \( v \) is indeed defined in the whole interval but it is \( v(t) = 0 \) in any interior interval.

In the next section we will call \( v \) our analytical multipliers considering that it is defined as in (10) for all \( t \). Such \( v \) will then be compared with its computed counterpart. Finally, it is worth mentioning that \( q_s \) may have a jump when \( t = T \) and that \( p_r(t) = q_r(t), p_l(t) = q_l(t), p_u(t) = q_u(t) \) and \( p_n(T) = p_n(T) = p_n(T) = 0 \).

5. Numerical Results

In Table 1 we present the values of the parameters and constants used in all our simulations. Such values are exactly as in [9]. The values of \( S_0, E_0, J_0, N_0 \) and \( W_0 \) appear in the last five lines of Table 1.

Recall that we allow the control to take values in the interval \([0, 1]\) while in [9] the control is constrained to \([0, 0.9]\).

To do our simulations we use the Imperial College London Optimal Control Software – ICLOCS – version 0.1b ([4]). ICLOCS is an optimal control interface, implemented in Matlab, for solving optimal control problems. It calls IPOPT – Interior Point OPTimizer – an open-source software package for large-scale nonlinear optimization [15]. For extra explanations see [10].
Table 1. Parameters with their clinically approved values and constants as in [9].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>natural birth rate</td>
<td>0.525</td>
</tr>
<tr>
<td>$d$</td>
<td>natural death rate</td>
<td>0.5</td>
</tr>
<tr>
<td>$c$</td>
<td>incidence coefficient</td>
<td>0.001</td>
</tr>
<tr>
<td>$e$</td>
<td>exposed to infectious rate</td>
<td>0.5</td>
</tr>
<tr>
<td>$g$</td>
<td>recovery rate</td>
<td>0.1</td>
</tr>
<tr>
<td>$a$</td>
<td>disease induced death rate</td>
<td>0.2</td>
</tr>
<tr>
<td>$A$</td>
<td>weight parameter</td>
<td>0.1</td>
</tr>
<tr>
<td>$T$</td>
<td>number of years</td>
<td>20</td>
</tr>
<tr>
<td>$S_0$</td>
<td>initial susceptible population</td>
<td>1000</td>
</tr>
<tr>
<td>$E_0$</td>
<td>initial exposed population</td>
<td>100</td>
</tr>
<tr>
<td>$I_0$</td>
<td>initial infected population</td>
<td>50</td>
</tr>
<tr>
<td>$R_0$</td>
<td>initial recovered population</td>
<td>15</td>
</tr>
<tr>
<td>$N_0$</td>
<td>initial population</td>
<td>1165</td>
</tr>
<tr>
<td>$W_0$</td>
<td>initial vaccinated population</td>
<td>0</td>
</tr>
</tbody>
</table>

Considering a time interval of 20 years ($T = 20$), a time-grid with 10000 nodes was created, i.e., for $t \in [0, 20]$ we get $\Delta t = 0.002$. Since our problem is solved by a direct method we impose an acceptable convergence tolerance at each step of $\varepsilon_{\text{rel}} = 10^{-9}$. In this respect we refer the reader to [4].

We now show the numerical simulations of (P\text{state}). We consider the state constraint $S(t) \leq S_{\text{max}}$ with $S_{\text{max}} = 1100$.

Fig. 1. The optimal trajectories and optimal vaccination rate for (P\text{state}).

About 6 345 individuals were vaccinated during the whole period. Figure 1 shows that the computed optimal control is 1 in the beginning dropping to approximately 0.2 and increasing from then on to keep the number of susceptible individuals equal or below 1 100. Observe that the state constraint has a boundary interval and that the state constraint is active at the end point (i.e., $S(20) = 1100$). The multiplier associated with the $S$ variable, $p_s$, is not 0 when $T = 20$, as shown in Figure 2. This behaviour can be explained since the measure $\mu$ has an atom at $t = T$ although it is absolutely continuous with respect to the Lebesgue measure on $[0, T]$. 


To validate the numerical solution, we first use (7). As shown in the top left graph of Figure 3, the computed optimal control satisfies (8). In the top right graph of Figure 3, we also show that the computed optimal control matches the control defined by (9) when the state constraint is active. We go a step further and compare the multipliers $\nu$ computed by ICLOCS with the analytical (10). This comparison is shown in the bottom right graph of Figure 3. Indeed, we have a match except for $t = T$ where the numerical multiplier $\nu$ has a jump at $t = T$, as seen in Figure 2. For the sake of completeness, we show the graph of the multipliers $q_s$ and $p_s$ in the bottom left graph of Figure 3. Recall that (iv) asserts $q_s(T) = 0$. It is clear that $q_s$ has a jump at $T$ due to the atom of the measure $\mu$.

Fig. 3. The adjoint multipliers for (P_S).
6. Conclusion

We considered an optimal control problem with state constraints to obtain optimal vaccination schedules and control strategies for an SEIR epidemic model of human infectious diseases. First order necessary conditions were applied to extract analytical information on the solution and multipliers of such problem that was then confronted with the computed counterparts. Taking into account the special structure of our problem the analysis allows us to review noble results on the normality of the Maximum Principle and on the regularity of multipliers and solution. Our approach is of interest since it illustrates a systematic way of validating numerical solutions for a class of optimal control problems. Many questions however remain unanswered. Indeed, it would be interesting to find an analytical way of determining the point where the state constraints touches the boundary. Also, stability of the solution with respect to the parameters should be studied. This last issue calls for second order conditions sufficient conditions as in, for example, [6], [7] and [8], and thus it is out of scope of this paper.

Acknowledgements

We would like to thank Prof. Helmut Maurer for his valuable suggestions. The financial support of FCT projects PTDC/EEA-CRO/116014/2009, PTDC/EEA-ELC/122203/2010 and PTDC/EEI-AUT/1450/2012, and European Union FP7 (FP7-PEOPLE-2010-ITN, Grant Agreement no. 264735-SADCO) are gratefully acknowledged.

References