Optimal control of the SIR epidemic model

Kazuyuki Yagasaki

Graduate School of Informatics, Kyoto University

1 Introduction

We consider the three-dimensional dynamical system

$$\dot{S} = -rSI, \quad \dot{I} = rSI - aI, \quad \dot{R} = aI, \tag{1.1}$$

where the variables S, I, R are nonnegative, and $r, a > 0$ are constants. Equation (1.1) is well-known as a basic model for spread of epidemic diseases and often called the *SIR model*. Here the state variables S, I and R represent the numbers of susceptible, infected and removed individuals, respectively, while a and r represent the infection and removal rate, respectively. We easily see that $N := S + I + R$ is constant. See, e.g., Section 10.2 of [5] for more details on the model (1.1). The dynamics of (1.1) are essentially described by the two-dimensional system

$$\dot{S} = -rSI, \quad \dot{I} = rSI - aI, \tag{1.2}$$

which is also referred to as the SIR model. We easily see that for any $c \geq 0$, $(S, I) = (c, 0)$ is an equilibrium in (1.2) and it is stable (but not asymptotically stable) or unstable, depending on whether $c < a/r$ or $c > a/r$. So we may expect that certain control techniques enable trajectories in (1.2) to converge to such an equilibrium as $c > a/r$, i.e., the number of infected individuals to decrease quickly before a critical situation occurs.

In this article, we apply optimal control [4, 6] to (1.2) as

$$\dot{S} = -(r - u_1(t))SI, \quad \dot{I} = (r - u_1(t))SI - aI - u_2(t), \tag{1.3}$$

such that $\lim_{t\to\infty} S(t) = c$ with $c > a/r$ at least and the *cost functional*

$$J = \int_0^\infty \left(k_1 u_1(t)^2 + k_2 u_2(t)^2\right) \mathrm{d}t \tag{1.4}$$

is minimized, where $k_j > 0$, $j = 1, 2$, are constants. Here the control inputs u_1, u_2 work so that the effective infection rate is decreased and infected individuals are removed. In a real issue of epidemic spread, for instance, the former can be done by decreasing the individuals' contact and the latter by finding infected individuals and isolating them immediately. The minimization of the cost functional (1.4) implies that social and economic burdens are reduced as much as possible. We choose such a special form of control inputs but our approach is generally valid for the other forms of control inputs, and probably even for different epidemic models. To obtain the optimal control inputs $u_1(t)$ and $u_2(t)$, we use the invariant manifold theory from dynamical systems theory. We demonstrate the usefulness of our approach for the optimal control in the SIR model (1.1) or (1.2) in numerical computations. Our study suggests how we can decrease the number of infected individuals quickly before a critical situation occurs while keeping social and economic burdens small. See [8] for more details on the results.

This work was partially supported by the JSPS KAKENHI Grant Number JP17H02859.

2 Optimal Control Based on Dynamical Systems Theory

We consider the general optimal control problem for systems of the form

$$\dot{x} = f(x) + g(x)u(t), \quad (x, u) \in \mathbb{R}^n \times \mathbb{R}^m, \tag{2.1}$$

where x and u, respectively, represent the state and control input vectors, $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are C^r $(r \geq 2)$ and $f(x_{\rm d}) = 0$, i.e., $x = x_{\rm d}$ is an equilibrium in (2.1) with $u = 0$. Here we take the equilibrium $x = x_{\rm d}$ as the target and determine the control input u such that the *cost functional*

$$J(x_0, u) = \int_0^\infty (q(x(t)) + \langle u(t), K(x)u(t)\rangle)\, {\rm d}t \tag{2.2}$$

is minimized when $x(0) = x_0$ and $\lim_{t\to\infty} x(t) = x_{\rm d}$, where $\langle\cdot,\cdot\rangle$ represents the inner product in $\mathbb{R}^m$ (or $\mathbb{R}^n$ below); $q : \mathbb{R} \to \mathbb{R}$ and $K : \mathbb{R}^n \to \mathbb{R}^{m\times m}$ are C^{r+1}; $q(x) \geq 0$ with $q(x_{\rm d}) = 0$; and $K(x)$ is positive definite and symmetric in $\mathbb{R}^n$. Note that ${\rm D}q(x_{\rm d}) = 0$.

We suppose that such a minimum exists and set

$$V(x_0) = \min\bigl\{J(x_0, u) \mid u \in C((0,\infty), \mathbb{R}^m), \lim_{t\to\infty} x(t) = x_{\rm d}\bigr\},$$

which is called the *value function*. From the standard theory of optimal control [4, 6] we see that $V(x)$ satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$H(x, {\rm D}V(x)) = \langle {\rm D}V(x), f(x)\rangle - \tfrac{1}{2}\langle {\rm D}V(x), G(x){\rm D}V(x)\rangle + q(x) = 0 \tag{2.3}$$

and the optimal control input is given by

$$u = -\tfrac{1}{2}K(x)^{-1}g(x)^{\rm T}{\rm D}V(x), \tag{2.4}$$

where $G(x) = \frac{1}{2}g(x)K(x)^{-1}g(x)^{\rm T}$. In particular, $V(x_{\rm d}) = 0$ and $g(x_{\rm d}){\rm D}V(x_{\rm d}) = 0$.

We now consider the Hamiltonian system

$$\begin{aligned} \dot{x} =& f(x) - G(x)p, \\ \dot{p} =& -({\rm D}f(x))^{\rm T}p + \tfrac{1}{2}({\rm D}_x\langle p, G(x)p\rangle)^{\rm T} - {\rm D}q(x)^{\rm T} \end{aligned} \tag{2.5}$$

with the Hamiltonian

$$H(x, p) = \langle p, f(x)\rangle - \tfrac{1}{2}\langle p, G(x)p\rangle + q(x) \tag{2.6}$$

(cf. Eq. (2.3)). Using the Hamilton-Jacobi theory (e.g., Section 46D of [1]), we show that any solution $(x(t), p(t))$ to (2.5) with $\lim_{t\to\infty}(x(t), p(t)) = (x_{\rm d}, 0)$ satisfies $p(t) = {\rm D}V(x(t))$. Equation (2.5) has an equilibrium at $(x, p) = (x_{\rm d}, 0)$, at which the Jacobian matrix becomes

$$A = \begin{pmatrix} {\rm D}f(x_{\rm d}) & -G(x_{\rm d}) \\ 0 & -{\rm D}f(x_{\rm d})^{\rm T} \end{pmatrix}.$$

Noting that $H(x_{\rm d}, 0) = 0$ by $q(x_d) = 0$, we see that if a trajectory $(x(t), p(t))$ converges to $(x_{\rm d}, 0)$ as $t \to \infty$, then by (2.4) the optimal control input $u(t)$ is given by

$$u(t) = -\tfrac{1}{2}K(x)^{-1}g(x)^{\rm T}p(t). \tag{2.7}$$

If ${\rm D}f(x_{\rm d})$ has an eigenvalue with a positive real part, then A has an eigenvalue with a negative real part, so that the equilibrium $(x, p) = (x_{\rm d}, 0)$ in (2.5) has a stable manifold [3, 7]. So we compute such a trajectory $(x(t), p(t))$ as it converges to $(x_{\rm d}, 0)$ as $t \to \infty$ and obtain the optimal control input $u(t)$ from (2.7).

3 Application to the controlled SIR model (1.3)

We apply the theory of Section 2 to the controlled SIR model (1.3). The Hamiltonian (2.6) becomes

$$H(S,I,p) = -rSIp_1 + (rS-a)Ip_2 - \frac{S^2I^2}{4k_1}(p_1-p_2)^2 - \frac{p_2^2}{4k_2}. \tag{3.1}$$

So we consider the Hamiltonian system

$$\begin{aligned}
\dot{S} &= -rSI - \frac{S^2I^2}{2k_1}(p_1-p_2),\\
\dot{I} &= (rS-a)I + \frac{S^2I^2}{2k_1}p_1 - \left(\frac{S^2I^2}{2k_1} + \frac{1}{2k_2}\right)p_2,\\
\dot{p}_1 &= rIp_1 - rIp_2 + \frac{SI^2}{2k_1}(p_1-p_2)^2,\\
\dot{p}_2 &= rSp_1 - (rS-a)p_2 + \frac{S^2I}{2k_1}(p_1-p_2)^2,
\end{aligned} \tag{3.2}$$

which corresponds to (2.5).

We easily see that for any $c \geq 0$

$$(S,I,p_1,p_2) = (c,0,0,0) \tag{3.3}$$

is an equilibrium in (3.2). The Jacobian matrix of the right-hand-side of (3.2) at the equilibrium (3.3) is given by

$$A = \begin{pmatrix} 0 & -rc & 0 & 0\\ 0 & rc-a & 0 & -1/2k_2\\ 0 & 0 & 0 & 0\\ 0 & 0 & rc & -(rc-a) \end{pmatrix}, \tag{3.4}$$

for which $\lambda = \pm(rc-a)$ are simple eigenvalues and $\lambda = 0$ is a double eigenvalue. Hence, the equilibrium (3.3) has a two-dimensional center manifold W^{c} and one-dimensional stable and unstable manifolds, W^{s} and W^{u}, if $c \neq a/r$.

These equilibria construct a one-dimensional invariant manifold

$$\{(S,I,p_1,p_2) = (c,0,0,0) \mid c > a/r\} \tag{3.5}$$

in the four-dimensional phase space. Especially, the invariant manifold has two-dimensional stable, unstable and center manifolds. Using the computer software `AUTO` [2], we compute the stable manifold to apply the approach of Section 2. From (2.7) we obtain the optimal control input as

$$u_1(t) = \frac{S(t)I(t)}{2k_1}(p_2(t)-p_1(t)), \quad u_2(t) = \frac{p_2(t)}{2k_2}, \tag{3.6}$$

where $(S(t),I(t),p_1(t),p_2(t))$ is a trajectory on the computed two-dimensional stable manifold of the invariant manifold (3.5). For a wide range of an initial condition (S_0,I_0) of (S,I), the control input (3.6) can make the corresponding trajectory in (1.3) converge to an equilibrium $(S,I) = (c,0)$ for some $c > a/r$ although the value of c depends on the initial condition.

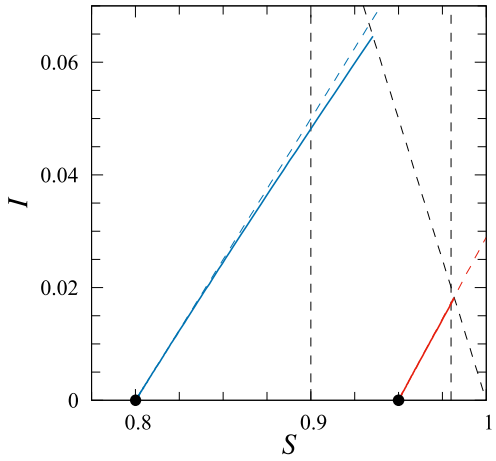

Figure 1. Projection of the stable manifold W^{s} onto the (S, I)-plane in case (i) for $(a, r) = (0.4, 1)$. The stable manifold is plotted as the blue and red lines for $c = 0.8$ and 0.95, respectively, and the corresponding stable subspace is plotted as the dashed lines with the same colors. The bullets '•' represent the loci of equilibria, and the black dashed lines represent the relations $S + I = 1$, $S = 0.9$ and $S = 0.98$.

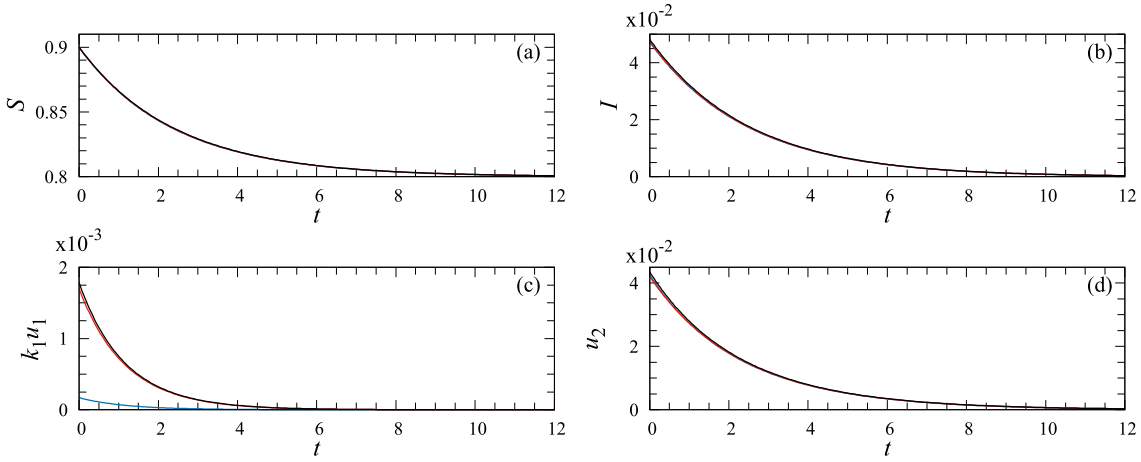

Figure 2. Controlled trajectory with $S_0 = 0.9$ converging to $(S, I) = (0.8, 0)$ in (1.3) for $(a, r) = (0.4, 1)$: (a) S; (b) I; (c) k_1u_1; (d) u_2. These components are plotted as black, red and blue lines for cases (i), (ii) and (iii), respectively. The value of I_0 is approximately 0.0482 but slightly different in the three cases.

4 Numerical Examples

We set $N = 1$, $a = 0.4$ and $r = 1$. Especially, $S + I \leq 1$. Without control, the equilibrium point $(S, I) = (c, 0)$ is stable if $c < 0.4$ and it is unstable if $c > 0.4$ in (1.2). Moreover, the variable I increases at least till the variable S decreases to 0.4. We consider the following three cases for the parameters k_1, k_2 in the cost functional (1.4):

(i) $(k_1, k_2) = (1, 1)$; (ii) $(0.1, 1)$; (iii) $(1, 0.1)$.

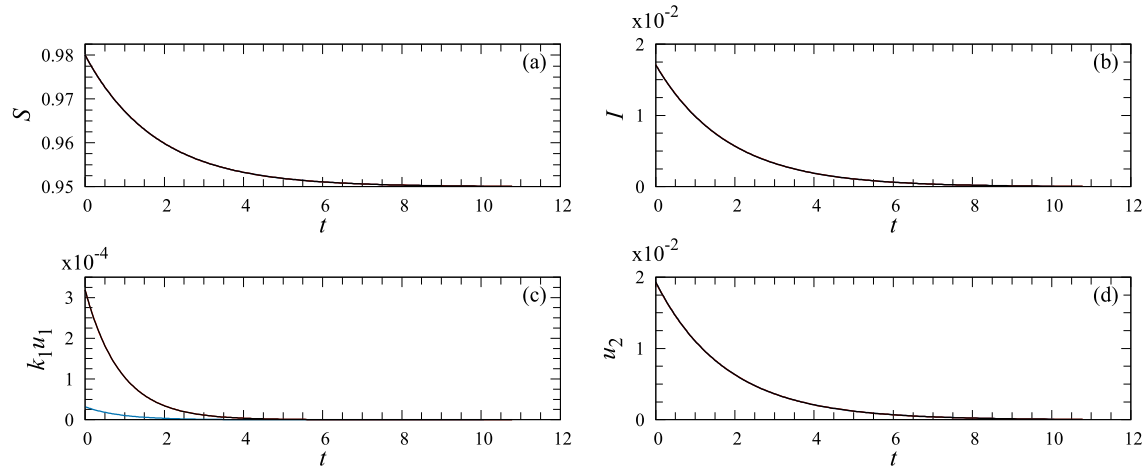

Figure 3. Controlled trajectory with $S_0 = 0.98$ converging to $(S, I) = (0.95, 0)$ in (1.3) for $(a, r) = (0.4, 1)$: (a) S; (b) I; (c) k_1u_1; (d) u_2. These components are plotted as black, red and blue lines for cases (i), (ii) and (iii), respectively. The value of I_0 is approximately 0.0171 but slightly different in the three cases.

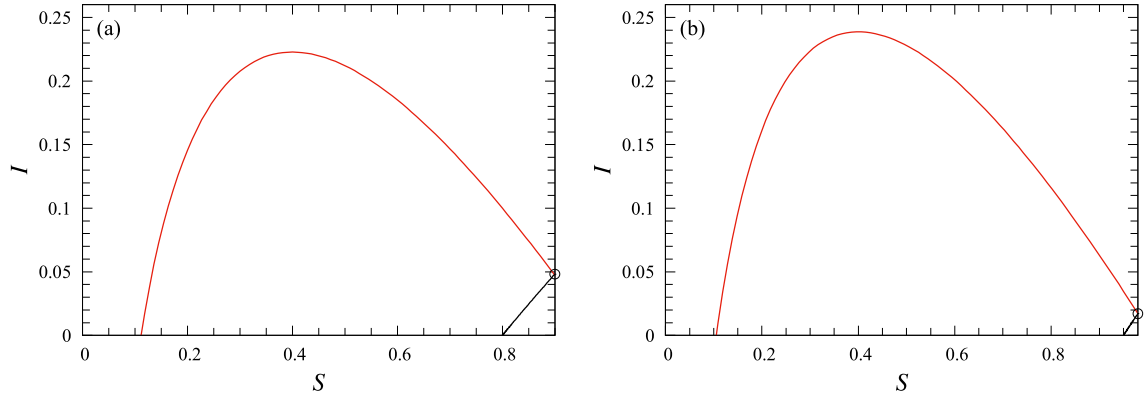

Figure 4. Uncontrolled and controlled trajectories with the same initial point (S_0, I_0) in (1.2) and (1.3) for $(a, r) = (0.4, 1)$ in case (i): (a) $(S_0, I_0) = (0.9, 0.04720\ldots)$; (b) $(S_0, I_0) = (0.98, 0.017268\ldots)$. Uncontrolled and controlled ones are plotted as red and black lines, respectively. The circle '$\circ$' represents the initial point.

When we consider an issue of suppressing epidemic spread, an attempt to find infected individuals and isolate them immediately (resp. to decrease contacts between infected ans susceptible individuals) is made more actively in case (iii) (resp. in case (ii)). Case (i) is moderate.

Figure 1 shows a projection of the stable manifold $W^{\rm s}$ onto the (S, I)-plane for $c = 0.8$ and 0.95 in case (i). We see that it almost coincides with the corresponding stable subspace but find a small difference between them for $c = 0.8$. For the other cases similar results were obtained.

Controlled trajectories converging to $(S, I) = (0.8, 0)$ and $(0.95, 0)$ in (1.3) along with the control inputs u_1, u_2 are displayed in Figs. 2 and 3, respectively, for the three cases.

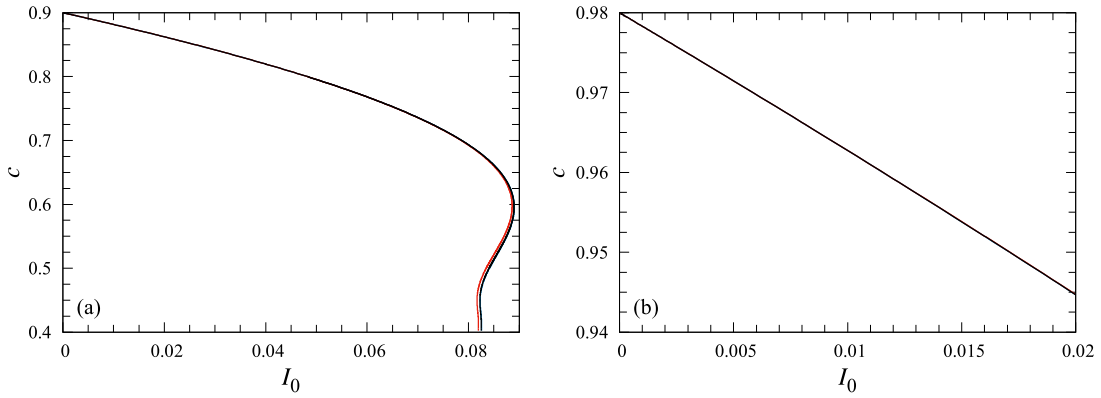

Figure 5. Target equilibrium $(c, 0)$ to which the controlled trajectory starting at (S_0, I_0) converges in (1.3) for $a = 0.4$ and $r = 1$: (a) $S_0 = 0.9$; (b) $S_0 = 0.98$. It is plotted as black, red and blue lines for cases (i), (ii) and (iii), respectively.

The results for case (i) in Figs. 2 and 3, respectively, correspond to the points $(S_0, I_0) = (0.9, 0.0482\ldots)$ and $(0.98, 0.0171\ldots)$ on $W^{\rm s}$ in Fig. 1. We see that the S, I and u_2 vary along almost the same curves in the three cases except that the control input u_1 is very different: It is approximately a hundred times bigger in case (ii) with $k_1 = 0.1$ than in case (iii) with $k_1 = 1$ for the displayed range, where it should be noticed that "k_1u_1" is used as the ordinates in Figs. 2(c) and 3(c). This means that finding infected individuals and isolating them immediately is more effective than decreasing their contacts from a viewpoint of suppressing epidemic spreading. One of the reason for this observation is that the dynamics of (3.2) are well approximated by its linearized system, which is independent of k_1, since $(S(t), I(t))$ is very close to the target.

Figure 4 shows a comparison between the uncontrolled and controlled trajectories for the same initial point (S_0, I_0) in (1.2) and (1.3) for case (i). We see that the optimal control succeeds in decreasing the final total number of infected individuals drastically and preventing a critical situation.

As stated at the end of Section 3, the target equilibrium $x_{\rm d} = (c, 0)$ to which the controlled trajectory $(S(t), I(t))$ converges depends on its initial point (S_0, I_0) in (1.3). Figure 5 shows how the target equilibrium $(c, 0)$ depends on I_0 for $S_0 = 0.9$ and 0.98 in the three cases. We find almost no difference between the three cases in Fig. 5. In Fig. 5(a) we see that two or more (resp. no) values of c correspond to one value of $I_0 \in (0.082, 0.089)$ (resp. of $I_0 > 0.089$) approximately. Thus, for $S_0 = 0.9$, the controlled trajectory converges to one of two or more equilibria with $c > a/r = 0.4$, depending on which optimal control inputs are chosen, for $I_0 \in (0.082, 0.089)$ approximately, and only converge to equilibria with $c < a/r = 0.4$ for $I_0 > 0.089$ approximately.

References

[1] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed., Springer, New York, 1989.

[2] E. Doedel and B.E. Oldeman, *AUTO-07P: Continuation and Bifurcation Software for Ordinary Differential Equations*, 2012, available online from `http://cmvl.cs.concordia.ca/auto`.

[3] J. Guckenheimer and P.J. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer, New York, 1983.

[4] F.L. Lewis, D. Vrabie and V.L. Syrmos, *Optimal Control*, 3rd ed. John Wiley and Sons, Hoboken, NJ, 2012.

[5] J.D. Murray, *Mathematical Biology I: An Introduction*, 3rd ed., Springer, New York, 2002.

[6] A.J. van der Schaft, *L_2-Gain and Passivity Techniques in Nonlinear Control*, 2nd ed., Springer, London, 2000.

[7] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, 2nd ed., Springer, New York, 2003.

[8] K. Yagasaki, Optimal control of the SIR epidemic model based on dynamical systems theory, submitted for publication.