Maximizing cooperation in the prisoner's dilemma evolutionary game via optimal control

P. K. Newton[*]
Department of Aerospace & Mechanical Engineering, Mathematics, and The Ellison Institute, University of Southern California, Los Angeles, California 90089-1191, USA

Y. Ma[†]
Department of Physics & Astronomy, University of Southern California, Los Angeles, California 90089-1191, USA

(Received 15 July 2020; accepted 21 December 2020; published 13 January 2021)

The prisoner's dilemma (PD) game offers a simple paradigm of competition between two players who can either cooperate or defect. Since defection is a strict Nash equilibrium, it is an asymptotically stable state of the replicator dynamical system that uses the PD payoff matrix to define the fitness landscape of two interacting evolving populations. The dilemma arises from the fact that the average payoff of this asymptotically stable state is suboptimal. Coaxing the players to cooperate would result in a higher payoff for both. Here we develop an optimal control theory for the prisoner's dilemma evolutionary game in order to maximize cooperation (minimize the defector population) over a given cycle time T, subject to constraints. Our two time-dependent controllers are applied to the off-diagonal elements of the payoff matrix in a bang-bang sequence that dynamically changes the game being played by dynamically adjusting the payoffs, with optimal timing that depends on the initial population distributions. Over multiple cycles nT $(n > 1)$, the method is adaptive as it uses the defector population at the end of the nth cycle to calculate the optimal schedule over the $n + 1$st cycle. The control method, based on Pontryagin's maximum principle, can be viewed as determining the optimal way to dynamically alter incentives and penalties in order to maximize the probability of cooperation in settings that track dynamic changes in the frequency of strategists, with potential applications in evolutionary biology, economics, theoretical ecology, social sciences, reinforcement learning, and other fields where the replicator system is used.

DOI: 10.1103/PhysRevE.103.012304

I. INTRODUCTION

Game theory models, originally developed by von Neumann and Morgenstern in 1944 [1], are widely used as paradigms to study cooperation and conflict in fields ranging from military strategy [2], social interactions [3], economics and social sciences [1,4], computer science [5], the physics of complex systems [6], evolutionary psychology [7], evolutionary biology [8], and, more recently, cancer [9–13]. To characterize the game, a payoff matrix, A, is introduced which for a two player game is of the generic form:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \tag{1}$$

The entries of the payoff matrix determine the cost-benefit trade-off associated with each game between the two competitors who rationally compete with the goal of maximizing their own payoff [14]. We show, in Fig. 1, 12 possible games that can be played (of a total of 78 [15]) using (1) as the payoff matrix. Notice for a given value of a_{22} which we place at the origin without loss of generality, and with $a_{11} > a_{22}$ (hence the reduction in the total number of games), we can play any game if we choose the off-diagonal elements of A appropriately in the (a_{12}, a_{21}) plane. The prisoner's dilemma (PD) region of the plane occupies special territory among all other regions as it is by far the most studied and used in models that focus on the evolution of cooperation [16–18]. A well-studied framework is the iterated PD game between two players that are allowed to both view their opponent's past strategies, and adjust their own strategy for each new game based on past information. Predicting which strategy will work best is difficult, and the famous Axelrod tournaments in which competitors submitted their strategy, and the pool of strategies competed through computer simulations to see which ones worked best was the beginning of the study of such systems [3]. Recent contributions to this literature introduced a new, previously undiscovered successful strategy [19].

Evolutionary game theory, used in population dynamics models involving evolution by natural selection [8] similarly, makes use of the payoff matrix (1) but embeds it directly into a dynamical setting by assigning the payoff matrix to a dynamical system and associates payoff with reproductive prowess. A common evolutionary game theory model is the replicator dynamical system [20], in the context of two population frequencies $\vec{x} = (x_1, x_2)^T \in \mathbb{R}^2$:

$$\dot{x}_i = x_i[(A\vec{x})_i - \vec{x}^T(A\vec{x})] \quad (i = 1, 2) \tag{2}$$

[*] Corresponding author: newton@usc.edu
[†] yongqiam@usc.edu

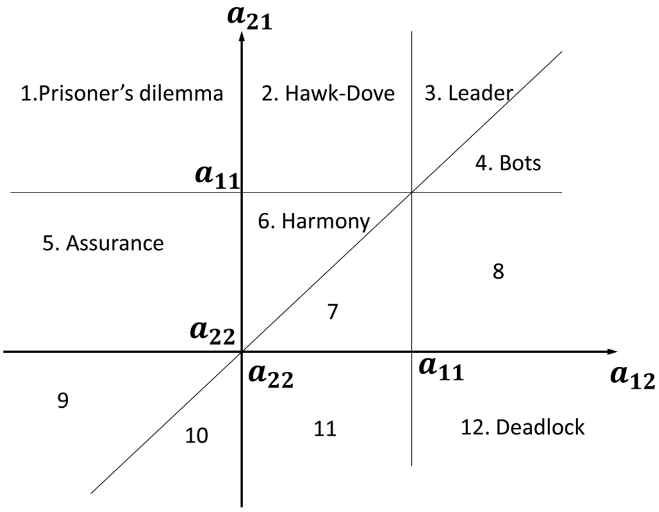

FIG. 1. Twelve regions in the (a_{12}, a_{21}) plane define which of the games is being played. There is no loss of generality in choosing a_{22} at the origin. Our control problem starts ($t = 0$) in the Prisoner's dilemma corner and asks what is the best path to travel, subject to a constraint, in order to minimize the defector population at the end of time $t = T$.

with $x_1 + x_2 = 1$, $0 \leqslant x_1 \leqslant 1$, $0 \leqslant x_2 \leqslant 1$, where each variable has the alternative interpretation as a probability. One can also think of the variables $\vec{x} = (x_1, x_2)^T$ as strategies that evolve, with the most successful strategy, say, $x_2(t) \to 1$, while the other $x_1(t) \to 0$ (as in the PD game). $(A\vec{x})_i$ is the fitness of population i, and $\vec{x}^T(A\vec{x})$ is the average fitness of both populations, so x_i in (2) drives growth if the population i is above the average and decay if it is below the average. The fitness functions in (2) are said to be population dependent (selection pressure is imposed by the mix of population frequencies) and determine growth or decay of each subpopulation.

The PD game captures, in a simple framework, the competition between two players, one of whom is labeled a cooperator (say, x_1), while the other is labeled a defector (say, x_2). As such, the prisoner's dilemma game presents a situation where if each player cooperated, they would receive a better payoff than if they both defected. The dilemma is, if they are rational players, they will both choose to defect, as this state is a Nash equilibrium of the system [21], defined as a strategy $\vec{p}^* \equiv (x_1^*, x_2^*)$ where $\vec{p}^{*T} A \vec{p}^* \geqslant \vec{p}^T A \vec{p}^*$, for all $\vec{p}$ [21]. This paradigm is well understood and has been discussed extensively in the literature [16,17].

Motivated by the use of the replicator equations and PD payoff matrix in cancer models [22,23], we develop an optimal control framework for the replicator equations (2) by allowing the off-diagonal elements of the payoff matrix, $a_{12}(t), a_{21}(t)$, to be time-dependent functions that allow us to exogenously shape the fitness landscape of two coevolving populations. As these fitness values change in time, we move to different regions of the plane shown in Fig. 1 which dynamically changes the *instantaneous* game being played during the course of the population evolution. Stated differently, if we start in the Prisoner's dilemma corner, what is the optimal path to travel in time $t = T$ that minimizes the defector population? This interpretation has been suggested recently in the context of developing adaptive therapy schedules for cancer [12,13]. The interplay between the timescales associated with the control functions $a_{12}(t)$ and $a_{21}(t)$ and the underlying dynamical timescales associated with the replicator system [governed by the eigenvalues of the linearized system (2)] makes the optimal control problem interesting. As far as we are aware, the only works we know of in which the payoff matrix is altered during the course of evolution in the context of the replicator equations is the interesting paper by Weitz *et al.* [24], who allow the payoff entries to coevolve, using feedback, along with the populations. In other contexts, such as social interactions, there is a body of work focusing on dilemma resolution by eliminating defectors, such as in the recent works of Tanimoto [25,26], as well as Refs. [27–29] and the book [4]. Our goal in this paper is, by using optimal control theory with time-dependent controllers, to show how to optimally reduce the defector population x_2 (i.e., maximize cooperation) after a fixed cycle time T in which we apply our time-dependent controller schedule, subject to fixed constraints on the control laws. We then extend the method to n (to $n = 5$) cycle times nT with an adaptive method that uses the defector population at the end of the nth cycle, $x_2(nT)$, to compute the optimal control schedule for the $n + 1$st cycle and $x_2[(n+1)T]$. We develop the mathematical framework to implement this, independent of the physical, biological, and sociological interpretations of the controllers and defector population, although we motivate our methods with an adaptive chemotherapy model. A nice, more generally focused review paper on the combined use of game theory and control is Ref. [30].

II. REPLICATOR DYNAMICS AND OPTIMAL CONTROL THEORY

A. Replicator dynamical system

We start with the uncontrolled payoff matrix of prisoner's dilemma type:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A_0 = \begin{bmatrix} 3 & 0 \\ 5 & 1 \end{bmatrix}, \tag{3}$$

where the population x_1 are the cooperators and x_2 are the defectors. The Nash equilibrium, $\vec{p}^*$, is the defector state $x_1 = 0, x_2 = 1$, which is easily shown to be a strict Nash equilibrium since $\vec{p}^{*T} A \vec{p}^* > \vec{p}^T A \vec{p}^*$ for all $\vec{p} \neq \vec{p}^*$. This implies that the defector state is also an evolutionary stable state (ESS) of the replicator system (2) as discussed in Ref. [21].

We then introduce an augmented payoff matrix A of the form:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A_0 + A_1(t), \tag{4}$$

$$= \begin{bmatrix} 3 & 0 \\ 5 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 4u_2(t) \\ -6u_1(t) & 0 \end{bmatrix}, \tag{5}$$

where $A_1(t)$ represents our control. The time-dependent functions $\vec{u}(t) = (u_1(t), u_2(t)) \in \mathbb{R}^2$ are bounded above and below, $0 \leqslant u_1(t) \leqslant 1$, $0 \leqslant u_2(t) \leqslant 1$ and have a range that allows use to traverse the plane depicted in Fig. 1 to play each of the 12 possible games at any given snapshot in time.

We will use the two control functions to *shape the fitness landscape* of the system appropriately by dynamically changing the game being played between the two populations as the system evolves (only the difference in the row values of A_0 determine the dynamics which makes two controllers sufficient).

To motivate our problem further and understand why the prisoner's dilemma game has been used as a paradigm for tumor growth [10,22,23,31], think of a competing population of healthy cells, x_1, and cancer cells, x_2, where the healthy cells play the role of cooperators and the cancer cells play the role of the defectors in a PD evolutionary game [9]. Using (2), starting with any tumor cell population $0 < x_2 \leqslant 1$, it is straightforward to show (i) $x_2 \to 1$, as $t \to \infty$; (ii) the average fitness of the healthy cell state $x_1 = 1, x_2 = 0$ is greater than the average fitness of the cancer cell state $x_1 = 0, x_2 = 1$ since $\vec{x}^T A_0 \vec{x} \equiv 3x_1^2 + 5x_1x_2 + x_2^2$ is the average fitness and $3 > 1$; (iii) the average fitness of the total population decreases as the cancer cell population saturates; and (iv) the tumor growth curve $x_2(t)$ vs. t yields an S-shaped curve very typical of tumor growth curves [32].

In this context, the controller $\vec{u}(t)$ can be thought of as a chemotherapy schedule which will alter the fitness balance of the uncontrolled healthy cell–tumor cell populations. The goal of a simple chemotherapy schedule might be to shape the fitness landscape so that the defector (tumor) cells $x_2(t)$ do not reach the saturated state $x_2(t) = 1$. If we denote the dose $\vec{D} \in \mathbb{R}^2$:

$$\vec{D}(t) = (D_1(t), D_2(t)) = \int_0^t \vec{u}(t)dt \tag{6}$$

as the total dose delivered in time t, then:

$$\dot{\vec{D}}(t) = \vec{u}(t) \tag{7}$$

and:

$$\vec{D}(0) = 0, \tag{8}$$

$$\vec{D}(T) = \int_0^T \vec{u}(t)dt = \vec{D}_T, \tag{9}$$

where T denotes a final time in which we implement the control. Thus, $\vec{D}_T$ denotes the total dose delivered in time T which we will use that as a constraint on the optimization problem. Then, our control goal is to maximally reduce the tumor cell (defector) population $x_2(t)$ at the end of one chemotherapy cycle $0 \leqslant t \leqslant T$, given constraints on the total dose delivered. We are particularly interested in the optimal schedule $\vec{u}(t)$ that produces the optimal response and we will compare it with the responses produced by other schedules (that have chemotherapeutic interpretations). Stated simply, we are interested in obtaining the time-dependent schedule $(u_1(t), u_2(t))$ in (5) that maximizes cooperation $x_1(t)$ for the system (2) at the end of time $t = T$, subject to the constraint $\vec{D}_T =$ const. We then extend our time-horizon to n cycles $t = nT$ (up to $n = 5$) by solving the optimal control problem in each cycle. This general framework extends to most other applications in which evolutionary game theory is used.

B. Pontryagin maximum principle

We utilize the standard form for implementing the maximum (minimum) principle with boundary value constraints:

$$\vec{X} = [\vec{x}(t), \vec{D}(t)]^T, \quad \vec{X} \in \mathbb{R}^4, \tag{10}$$

$$\dot{\vec{X}} = \vec{F}(\vec{X}) = [\dot{\vec{x}}, \dot{\vec{D}}(t)]^T, \quad \vec{F} : \mathbb{R}^4 \to \mathbb{R}^4, \tag{11}$$

with the goal of minimizing a general cost function:

$$\int_0^T L(\vec{x}(t), \vec{u}(t), t)dt + \varphi(\vec{x}(T)). \tag{12}$$

Since the method is standard, we will just briefly describe the basic framework and refer readers to Ref. [33–36] for more details on how to implement the approach. Following Ref. [36] in particular (see Theorem 4.2.1), we construct the control theory Hamiltonian:

$$H(\vec{x}(t), \vec{D}(t), \vec{\lambda}, \vec{u}(t)) = \vec{\lambda}^T \vec{F}(\vec{x}) + L(\vec{x}, \vec{u}(t), t), \tag{13}$$

where $\vec{\lambda} = [\lambda_1, \lambda_2, \mu_1, \mu_2]^T$ are the costate functions (i.e., momenta) associated with $\vec{x}$ and $\vec{D}$ respectively. Assuming that $\vec{u}^*(t)$ is the optimal control for this problem, with corresponding trajectory $\vec{x}^*(t), \vec{D}^*(t)$, the canonical equations satisfy:

$$\dot{x_i}^*(t) = \frac{\partial H}{\partial \lambda_i^*}, \tag{14}$$

$$\dot{D_i}^*(t) = \frac{\partial H}{\partial \mu_i^*}, \tag{15}$$

$$\dot{\lambda_i}^*(t) = -\frac{\partial H}{\partial x_i^*}, \tag{16}$$

$$\dot{\mu_i}^*(t) = -\frac{\partial H}{\partial D_i^*}, \tag{17}$$

where $i = (1, 2)$. The corresponding boundary conditions are as follows:

$$\vec{x}^*(0) = \vec{x}_0, \tag{18}$$

$$\vec{D}^*(0) = 0, \vec{D}^*(T) = \vec{D}_T^*, \tag{19}$$

$$\lambda_i^*(T) = \frac{\partial \varphi(\vec{x}(T))}{\partial x_i^*(T)}. \tag{20}$$

Then, at any point in time, the optimal control $\vec{u}^*(t)$ will minimize the control theory Hamiltonian:

$$\vec{u}^*(t) = \underset{\vec{u}(t)}{\arg\min}\, H(\vec{x}^*(t), \vec{D}^*(t), \vec{\lambda}^*(t), \vec{u}(t)). \tag{21}$$

The optimization problem becomes a two-point boundary value problem [using (18)–(20)] with unknowns $(\lambda_2^*(0), x_2^*(T))$ whose solution gives rise to the optimal trajectory $\vec{x}^*(t)$ (from (14)) and the corresponding control $\vec{u}^*(t)$ that produces it [33–36]. For the optimization, we take $\vec{D}(T) = (0.5, 0.5)$ and to minimize the defector frequency at the end of one cycle $t = T$, we choose our cost function (12):

$$L = 0; \varphi(\vec{x}(T)) = x_2(T). \tag{22}$$

We solve this problem by standard shooting type methods [37].

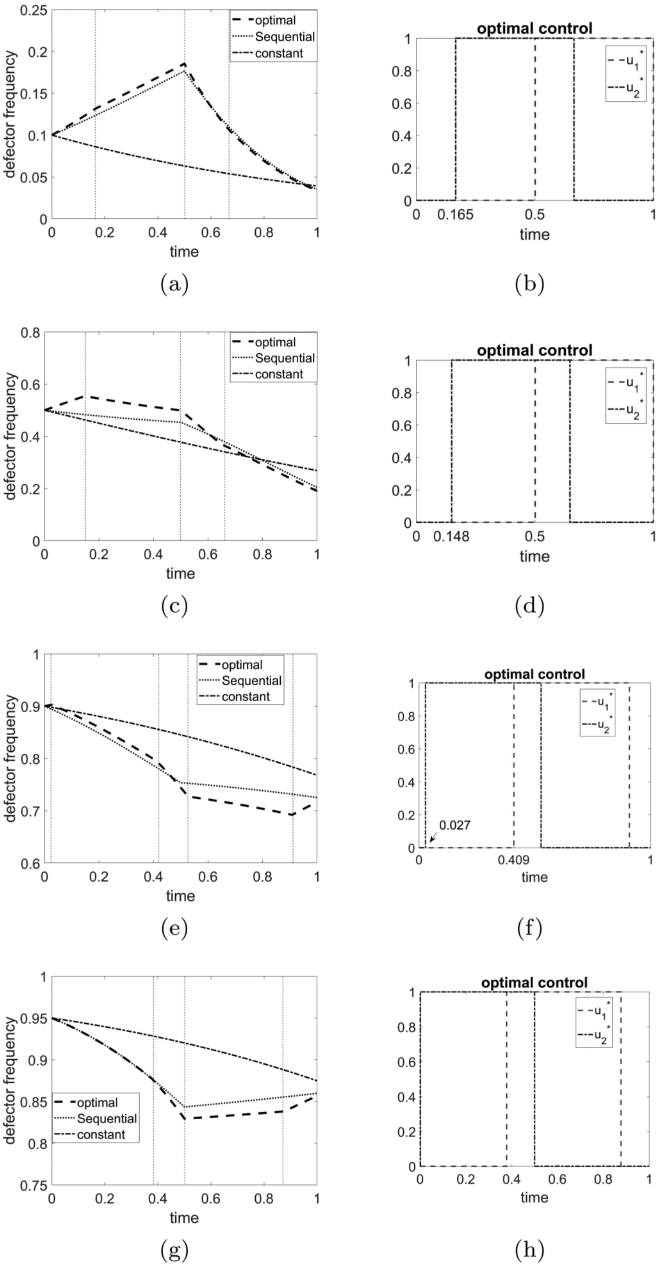

FIG. 2. Optimal trajectories and optimal control sequences for small, medium, and large initial defector populations, with $\vec{D}_T = (0.5, 0.5)$. (a) Defector frequency for initial state $x_2(0) = 0.1$; (b) optimal control sequence for initial state $x_2(0) = 0.1$; (c) defector frequency for initial state $x_2(0) = 0.5$; (d) optimal control sequence for initial state $x_2(0) = 0.5$; (e) defector frequency for initial state $x_2(0) = 0.9$; (f) optimal control sequence for initial state $x_2(0) = 0.9$; (g) defector frequency for initial state $x_2(0) = 0.95$; and (h) optimal control sequence for initial state $x_2(0) = 0.95$.

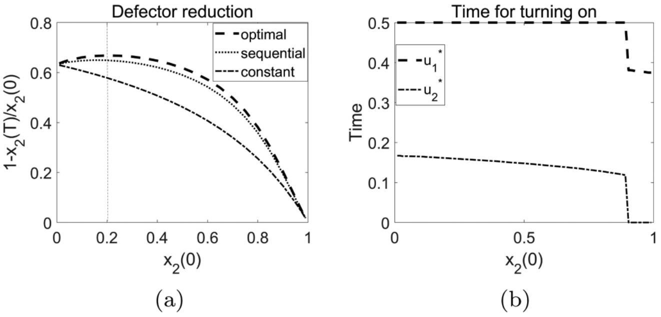

FIG. 3. (a) Relative reduction as a function of initial defector proportion. Maximum reduction (shown as dashed vertical line) occurs for initial defector proportion of around 22%. In the limits $x_2(0) \to 0, 1$, the exact schedule $\vec{u}(t)$ does not matter, only the total dose $\vec{D}(T)$. The controls schedule is much more efficient at reducing the proportion of defectors for small populations than for large; (b) turn-on time for $u_1(t)$ and $u_2(t)$ as a function of initial defector population $x_2(0)$. In the narrow (approximate) range $0.89 \leqslant x_2(0) \leqslant 0.91$, the times change abruptly and the control sequence has five segments as shown in Figs. 2(e) and 2(f).

III. RESULTS

Because the Hamiltonian (13) is linear in the controllers [with $L = 0$ in (21)], it is straightforward to prove that the optimal control $\vec{u}^*(t)$ is bang-bang. We compare the optimal schedule and trajectories with two other (nonoptimal) ones in Fig. 2, for a small value of $x_2(0) = 0.1$ [Figs. 2(a) and 2(b)], intermediate value of $x_2(0) = 0.5$ [Figs. 2(c) and 2(d)], large value of $x_2(0) = 0.9$ Figs. 2(e) and 2(f)], and a value near saturation $x_2(0) = 0.95$ [Figs. 2(g) and 2(h)]. There are several interesting points to make. First, for initial values $x_2(0) < 0.95$, the optimal control schedule allows the defector proportion to increase for a short time before u_2 turns on, which is perhaps counter-intuitive. For initial values sufficiently small, this initial growth phase is compensated by keeping u_1 turned on until the end of the cycle time T [Figs. 2(b) and 2(d)]. The larger the initial defector proportion, the earlier the control u_2 turns on [Figs. 2(f) and 2(h)], until for values above a threshold of $x_2(0) \sim 0.91$, the initial time abruptly goes to zero [Fig. 3(b)]. It is more beneficial to control the growth of $x_2(t)$ at the beginning of the cycle than at the end [Fig. 2(g) and 2(h)]. Also, notice that for the optimal control sequences shown in Figs. 2(b), 2(d) 2(f), and 2(h), the controllers u_1 and u_2 partially overlap in the time they are turned on in the middle of the cycle at the expense of leaving both off either at the beginning or end. We compare the optimal trajectories [Figs. 2(a), 2(c) and 2(e)] with the case where there is no time overlap (i.e., sequential: first u_2 followed by u_1), and the case where u_1 and u_2 are held constant throughout the cycle time T, all with the same value of $\vec{D}_T$. For small initial values [Fig. 2(a) and 2(b)] there is very little difference between the optimal trajectory and the one produced by a sequential controller, particularly for small initial conditions.

In Fig. 3(a) we show the relative reduction of the defector frequency $[1 - x_2(T)/x_2(0)]$ as a function of the initial frequency $x_2(0)$. The maximum reduction (dashed vertical line) occurs for an initial defector proportion of around 22%. The curve also shows that the optimal schedule is much more efficient at reducing the proportion of defectors for small populations than for large ones, with reduction approaching zero as the initial defector population approaches one. Also, in the limits $x_2(0) \to 0, 1$, all three schedules converge to the same value, showing that the exact schedule $\vec{u}(t)$ matters less than the total dose $\vec{D}_T$ which is the same for each. This is because in those regimes the system is essentially linear, and the exact

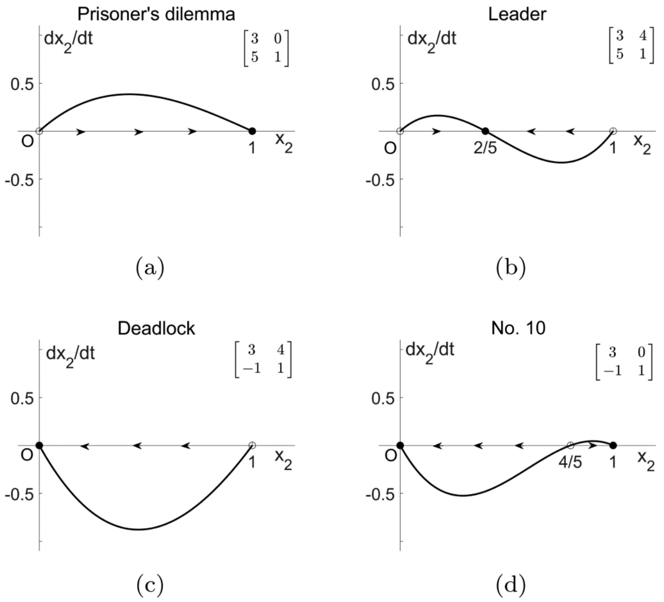

FIG. 4. The sequence of four games that produce the optimal trajectory if the switching times are chosen as shown in Fig. 2. (a) The first game is a prisoner's dilemma game, where $x_2 = 1$ is an asymptotically stable fixed point; (b) the second game is a Leader game with interior fixed point $x_2 = 2/5$ being an asymptotically stable fixed point; (c) the third game is a Deadlock game with $x_2 = 0$ being the asymptotically stable fixed point; (d) the fourth game is game No. 10 in Fig. 1 with $x_2 = 0, 1$ being asymptotically stable fixed points.

solution depends only on $\int_0^t u(t)dt$. Figure 3(b) shows the time at which each of the controllers is turned on, with an abrupt change for large enough initial defector populations $x_2(0) \sim 0.91$ as mentioned.

Figure 4 shows the sequence of distinct games that we cycle through to produce the optimal trajectory:

(1) Prisoner's dilemma (asymptotically stable state $x_2 = 1$);

(2) Leader (asymptotically stable state $x_2 = 2/5$);

(3) Deadlock (asymptotically stable state $x_2 = 0$);

(4) Game No. 10 in Fig. 1 (asymptotically stable states $x_2 = 0, 1$).

This is easiest to understand by decoupling the replicator system (2) and writing the cubic nonlinear ordinary differential equation for x_2:

$$\begin{aligned}\dot{x}_2 &= x_2(1-x_2)[(A\vec{x})_2 - (A\vec{x})_1],\\ &= x_2(1-x_2)[(2-6u_1)-(1-6u_1+4u_2)x_2].\end{aligned}\tag{23}$$

The sequence of games is obtained using:

(i) $u_1 = 0, u_2 = 0$

(ii) $u_1 = 0, u_2 = 1$

(iii) $u_1 = 1, u_2 = 1$

(iv) $u_1 = 1, u_2 = 0$.

To produce the optimal trajectory, the switching times must be chosen as shown in Figs. 2(b), 2(d) 2(f), and 2(h) [note in Figs. 2(f) and 2(h) the game switches back to PD just before T], and these times depend on the initial state $x_2(0)$. The flow associated with the sequence of four games depicted in Fig. 4 makes it clear why the switching times are tied to the initial condition.

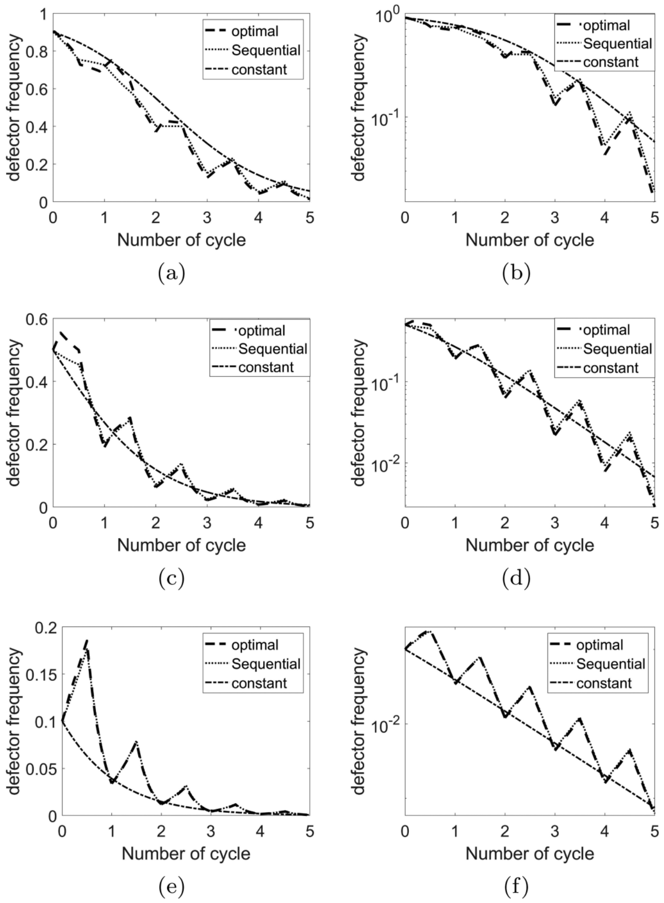

FIG. 5. Defector frequency reduction using optimal control over five consecutive cycles. (a) Defector optimal frequency with $x_2(0) = 0.9$ over five cycles as compared with frequency associated with sequential control and constant control; (b) same as (a) plotted on log-linear scale; (c) defector optimal frequency with $x_2(0) = 0.5$ over five cycles as compared with frequency associated with sequential control and constant control; (d) same as (c) plotted on log-linear scale; (e) defector optimal frequency with $x_2(0) = 0.1$ over five cycles as compared with frequency associated with sequential control and constant control; and (f) same as (e) plotted on log-linear scale.

Figure 5 shows the defector frequencies $x_2^*(t)$ over a sequence of five cycles $0 \leqslant t \leqslant 5T$ for the optimal control sequence, as compared with the sequential and constant controllers. In the case of optimization over multiple cycles, the method is adaptive as it must use the frequency obtained at the end of the nth cycle, $x_2^*(nT)$, to calculate the optimal schedule associated with the $n+1$st cycle. For large initial values $x_2(0) = 0.9$ [shown in Figs. 5(a) and 5(b)], the frequency reduction curves start decreasing (nearly) linearly, but deviates from linear over time. For small initial values [shown in Figs. 5(e) and 5(f) for $x_2(0) = 0.1$], the reduction is very close to linear.

IV. DISCUSSION

We have developed a constrained optimization procedure for maximizing cooperation in the prisoner's dilemma evolutionary game by dynamically altering the payoffs during the course of evolution. The interpretation of the schedules and payoffs depends on the application, but we focus on a tumor growth model with adaptive chemotherapy schedules as the controllers [10,22,23]. The optimal control schedule is bang-bang for each of the two controllers, with switching times that depend on the initial proportion of cooperators to defectors. One interpretation of the method is that we cycle through a sequence of four distinct payoff matrix types (PD → Leader → Deadlock → No. 10), switching the game at precisely the right times to maximize group cooperation. Interestingly, a recent paper interprets a strategy used by non-small-cell lung cancer cocultures as one of switching from playing a Leader game to playing a Deadlock game (which is one transition of our four game sequence) in order to develop resistance [13].

The interpretation of the control functions depends on the specific application with many potential interpretations in the context of controlling the balance of evolving populations. Aside from optimizing chemotherapy schedules to control tumors [22,23,38–40], one might think of the application of antibiotic schedules to control microbial populations [41], or the strategic application of toxins to control infestations using integrated pest management protocols [42]. In economics, one might think of time-dependent payoffs as dynamic policy actions that seek to optimize certain societal economic goals [43], in reinforcement learning applications [44], controlling the payoff entries alters the learning dynamics of multi-agent systems, or in traffic flow applications, one might seek dynamic incentives for minimizing transit times by dynamic payoff schedules [45]. We see no particular technical roadblock to generalizing the optimal control method to $N \times N$ evolutionary games, although the numerical challenges of solving the necessary two-point boundary value problems becomes more severe, particularly searching for appropriate initial guesses to ensure convergence.

As a final note, we mention that a system related to (2) is the adjusted replicator equation [31] with growth and decay terms normalized by the average fitness. While the normalization term has no effect on ratios $\dot{x}_i(t)/\dot{x}_j(t)$, it does effect the time scaling of each variable, hence changes the optimal control problem. Most importantly, the control Hamiltonian is no longer linear in $\vec{u}(t)$, so the optimal control may not necessarily be bang-bang, and we have preliminary results that show this. Since the adjusted replicator system is known to be the deterministic limit of the finite-population stochastic Moran process in the limit of infinite populations [46,47], this deterministic optimal control problem should be related to the stochastic optimal control problem associated with the Moran process via a Markov chain approach.

ACKNOWLEDGMENT

We gratefully acknowledge support from the Army Research Office MURI Award No. W911NF1910269 (2019-2024).

[1] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Press, Princeton, NJ, 1944).

[2] D. Snidal, World Politics **38**, 25 (1985).

[3] R. Axelrod, *The Evolution of Cooperation* (Basic Books, New York, 1984).

[4] J. Tanimoto, *Fundamentals of Evolutionary Game Theory and Its Applications* (Springer-Verlag, Berlin, 2015).

[5] T. Roughgarden, Commun. ACM **53** 7 (2010).

[6] M. Bertotti and M. Delitala, Math. Models Methods Appl. Sci. **14**, 1061 (2004).

[7] K. McCabe, J. Rassenti, and V. Smith, Proc. Natl. Acad. Sci. USA **93**, 13421 (1996).

[8] J. Maynard-Smith, *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, 1982).

[9] R. Axelrod, D. E. Axelrod, and K. J. Pienta, Proc. Natl. Acad. Sci. USA **103**, 13474 (2006).

[10] J. West, Z. Hasnain, J. Mason, and P. Newton, Converg. Sci. Phys. Oncol. **2**, 035002 (2016).

[11] D. Basanta, M. Simon, H. Hatzikirou, and A. Deutsch, Cell Prolif. **41**, 980 (2008).

[12] M. Gluzman, J. Scott, and A. Vladimirsky, Proc. Roy. Soc. B **287**, 20192454 (2020).

[13] A. Kaznatcheev, J. Peacock, D. Basanta, A. Marusyk, and J. Scott, Nature Ecol. Evol. **3**, 450 (2019).

[14] A. Rapoport, *Two-Person Game Theory* (Dover, London, 1966).

[15] A. Rapoport, *Two-Person Game Theory*, (Dover Publications Inc. Mineola, New York, 1966).

[16] R. Axelrod and W. Hamilton, Science **211**, 1390 (1981).

[17] R. Axelrod and D. Dion, Science **242**, 1385 (1988).

[18] M. Nowak, Science **314**, 1560 (2006).

[19] W. Press and F. Dyson, Proc. Natl. Acad. Sci. USA **109**, 10409 (2012).

[20] P. Taylor and L. Jonker, Math. Biosci. **40**, 145 (1978).

[21] J. Hofbauer *et al.*, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).

[22] P. K. Newton and Y. Ma, Phys. Rev. E **99**, 022404 (2019).

[23] P. Newton and Y. Ma, bioRxiv 2020.05.13.094375 [Physics Rev. E (to be published)].

[24] J. Weitz, C. Eksin, K. Paarporn, S. Brown, and W. Ratcliff, Proc. Natl. Acad. Sci. **113**, E7518 (2016).

[25] H. Ito and J. Tanimoto, R. Soc. Open Sci. **5**, 181085 (2018).

[26] H. Ito and J. Tanimoto, R. Soc. Open Sci. **7**, 200891 (2020).

[27] F. F. Ferreira and P. R. A. Campos, Phys. Rev. E **88**, 014101 (2013).

[28] R. J. Requejo and J. Camacho, Phys. Rev. Lett. **108**, 038701 (2012).

[29] W. Huang, P. de Araujo Campos, V. Moraes de Oliveira, and F. Fagundes Ferreira, Peer J **4**, e2329 (2016).

[30] J. Marden and J. Shamma, Annu. Rev. Control Robot. Auton. Syst. **1**, 105 (2018).

[31] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, MA, 2006).

[32] A. Laird, Br. J. Cancer **19**, 278 (1965).

[33] L. Pontryagin, *Mathematical Theory of Optimal Processes* (CRC Press, Boca Raton, FL, 1987).

[34] E. B. Lee and L. Markus, Foundations of Optimal Control Theory, Technical Report (University of Minnesota Center For Control Sciences, Minneapolis, MN 1967).

[35] I. Ross, *A Primer on Pontryagin's Principle in Optimal Control*, 2nd ed. (Collegiate Press, Boston, MA, 2015).

[36] K. L. Teo, C. Goh, and K. Wong, A unified computational approach to optimal control problems, *Pitman Monographs and Surveys in Pure and Applied Mathematics* (Longman Scientific & Technical, New York, 1991).

[37] A. Rao, Adv. Astro. Sci. **135**(1), 497 (2009).

[38] T. Browder, C. Butterfield, B. Krälling, B. Shi, B. Marshall, M. O'Rielly, and J. Folkman, Cancer Res. **60**, 1878 (2000).

[39] M. Engelhart, D. Lebiedz, and S. Sager, Math. Biosci. **229**, 123 (2011).

[40] H. Schättler and U. Ledzewicz, *Optimal Control for Mathematical Models of Chemotherapy* (Springer-Verlag, Berlin, 2015).

[41] M. Baym, T. Lieberman, E. Kelsic, R. Chait, R. Gross, I. Yelin, and R. Kishony, Science **353**, 1147 (2016).

[42] W. Lewis, J. van Lenteren, S. Phatak, and J. Tumlinson, Proc. Natl. Acad. Sci. USA **94**, 12243 (1997).

[43] T. Basar, *Dynamic Games and Applications in Economics* (Springer-Verlag, Berlin, 1986).

[44] D. Bloembergen, K. Tuyls, D. Hennes, and M. Kaisers, J. Artif. Intell. Res. **53**, 659 (2015).

[45] W. Sandholm, *Population Games and Evolutionary Dynamics* (MIT Press, Cambridge, MA, 2010).

[46] A. Traulsen, J. C. Claussen, and C. Hauert, Phys. Rev. Lett. **95**, 238701 (2005).

[47] A. Traulsen, J. C. Claussen, and C. Hauert, Phys. Rev. E **74**, 011901 (2006).