2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

Model-Based Switched Approximate Dynamic Programming for Functional Electrical Stimulation Cycling

Wanjiku A. Makumi, Max L. Greene, Kimberly J. Stubbs, Warren E. Dixon

***Abstract*—This paper applies a reinforcement learning-based approximately optimal controller to a motorized functional electrical stimulation-induced cycling system to track a desired cadence. Sufficient torque to achieve the cycling objective is achieved by switching between the quadriceps muscle and electric motor. Uniformly ultimately bounded (UUB) convergence of the actual cadence to a neighborhood of the desired cadence and of the approximate control policy to a neighborhood of the optimal control policy are proven for both motor control and muscle control via a Lyapunov-based stability analysis provided developed dwell-time conditions that determine when to switch between the motor or the muscle are satisfied. Lyapunov-based techniques are also used to derive a minimum dwell-time condition to prove UUB stability of the overall switched system.**

I. Introduction

Rehabilitation through functional electrical stimulation (FES) is a treatment for people with neurological conditions (NCs), such as stroke and spinal cord injury [1] and [2]. FES induces involuntary muscle contractions to perform a functional movement by applying an electric potential across the motor neurons of a muscle. To improve motor function and overall quality of life, multiple efforts in the rehabilitation field use FES with rehabilitation robots to facilitate human-robot therapy [3]. Stationary FES cycling is a common human-robot rehabilitative therapy for people with movement impairments resulting from NCs [4]. FES cycling has the benefits of both FES and rehabilitation robotics; it is a preferred therapy because there is minimal risk of a fall, and the repetition of coordinated limb movements improves motor skills and nervous system reorganization [5].

Optimal controllers can be established by assigning a user-defined cost to the states and control inputs, which penalizes the state and the magnitude of the control input. Through the cost function, a balance can be obtained between the accuracy of the limb motion versus the level of control effort, allowing potential tradeoffs between comfort, performance, duration of exercise, and muscle fatigue. The only results that apply optimal control methods to FES applications are [6] and [7]. These results use extremum seeking, a model-free online optimization tool, to adjust a closed-loop PID controller to minimize the cost function for upper limb electrical stimulation.

Optimal control problems can be solved via the Hamilton-Jacobi-Bellman (HJB) equation [8]. By solving the HJB equation to determine the optimal value function, an optimal control policy can be developed [8]. Generally, the HJB equation does not have a closed-form analytic solution for nonlinear systems. Motivated by the challenges of solving the HJB, especially in real-time, approximate dynamic programming (ADP) has emerged as a method to yield an approximate solution. Specficially, ADP uses a reinforcement learning (RL)-based actor-critic framework to approximate the value function in real-time [9]. Neural networks (NNs) are generally used within ADP to approximate the unknown optimal value function, but other function approximation methods could also be used [10].

In traditional adaptive control, the uncertain parameter estimates are updated using an error feedback as a performance metric; in ADP, the Bellman error (BE) is used as feedback on the level of suboptimality. Specifically, the BE is used to update the NN parameters to improve the value function approximation online. BE extrapolation yields faster policy learning over a domain by evaluating the BE over user-defined, off-trajectory regions of the state space [11]. Sufficient off-trajectory data must be selected to achieve adequate exploration. The value function approximation is updated according to the on- and off-trajectory BE.

Due to the potential benefits of using an optimal controller, it is advantageous to apply ADP to the cycling system. However, the system switches between two actuation methods: the rider's muscles and the cycle's electric motor. Therefore, FES-cycling is a switched (also called hybrid) system, which requires switched (hybrid) system analysis and design methods [12]. Until recently, switching has not been investigated in the context of ADP. The result in [13] develops a framework to estimate the optimal feedback control policy online while switching between multiple dynamic system models. When analyzing switched systems, a common problem is the growth and discontinuity of Lyapunov functions at switching instances [14]. This growth and discontinuity problem is overcome in [13] in which a dwell-time analysis is developed to determine the minimum time necessary before the system can switch to a different subsystem (i.e., a minimum dwell-time). This provides a framework to switch between the

Wanjiku A. Makumi, Max L. Greene, Kimberly J. Stubbs, and Warren E. Dixon are with the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA. Email: {makumiw, maxgreene12, kimberlyjstubbs, wdixon}@ufl.edu.

This research is supported in part by NSF Award number 1762829, Office of Naval Research Grant N00014-13-1-0151, AFOSR award number FA9550-18-1-0109, and AFOSR award number FA9550-19-1-0169. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agency.

978-1-6654-5196-3/$31.00 ©2022 AACC

two different modes of the FES-cycling controller and show stability of the overall switched system.

Motivated by our previous results in [13] and [14], this paper implements a continuous-time ADP-based tracking controller that allows for switching between mutliple cycle actuation methods to track a desired cadence. Uniformly ultimately bounded (UUB) stability of the overall switched system is proven. Moreover, the developed controller is also proven to converge to a neighborhood of the optimal controller.

Notation

For notational brevity, time-dependence is omitted while denoting trajectories of the dynamic systems. For example, the trajectory $x(t)$, where $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, is denoted as $x \in \mathbb{R}^n$ and referred to as x instead of $x(t)$. For example, the equation $f + h(y,t) = g(x)$ should be interpreted as $f(t) + h((y,t),t) = g(x(t))$. The gradient $\left[\frac{\partial f(x,y)}{\partial x_1}^T, \ldots, \frac{\partial f(x,y)}{\partial x_n}^T\right]^T$ is denoted by $\nabla_x f(x,y)$. Unless specified, let $\nabla \triangleq \nabla_\zeta$. A square diagonal matrix with elements of vector y on the main diagonal is denoted by $\text{diag}(y)$. Matrices of ones and zeros with n rows and m columns are denoted by $\mathbf{1}_{n\times m}$ and $\mathbf{0}_{n\times m}$, respectively. Both the Euclidean norm for vectors and the Frobenius norm for matrices are denoted by $\|\cdot\|$. Let $\lambda_{\min}\{\cdot\}$ denote the minimum eigenvalue of the argument. Let $r = \mod(m,p)$ denote the modulo operator where, generally, m is the dividend, p is the divisor, and r is the remainder. In this paper, the quantity or function belonging to the k^{th} mode of the switched system is denoted with the subscript k.

II. Problem Formulation

Following the development in [15], the dynamics of the combined one-legged cycle and rider system are

$$\tau = M(q)\ddot{q} + V_c(q,\dot{q})\dot{q} + G(q) + P(q,\dot{q}) + b_c\dot{q}, \quad (1)$$

where q, $\dot{q}$, and $\ddot{q} \in \mathbb{R}$ denote the angle, angular velocity, and angular acceleration, of the crank arm respectively. $M : \mathbb{R} \to \mathbb{R}_{>0}$ denotes the inertia matrix, $V_c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ denotes the centripetal-Coriolis matrix, $G : \mathbb{R} \to \mathbb{R}$ denotes the gravitational effects, $P : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ denotes the passive viscoelastic tissue forces, $b_c \in \mathbb{R}_{>0}$ denotes the cycle's viscous damping effect, and τ denotes the torque applied by the quadriceps muscle and the cycle motor, which is subsequently defined.

The torque is applied by two different actuators, corresponding to either the torque due to the FES-induced muscle contractions or the torque due to the electric motor. Given the need to use the different actuators at different times, we define two sets: $\mathcal{Q}$, when the crank angle is in the kinematically effective quadricep region, and $\mathcal{Q}^c$, when the crank angle is in the region of poor kinematic efficiency [16]. Let $\mathcal{Q} \subset [0^\circ, 360^\circ)$ denote where electrical stimulation is active and $\mathcal{Q}^c$ denote the complement of $\mathcal{Q}$, where the electric motor is active.

The torque $\tau : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ in (1) is defined as

$$\tau(q,\dot{q}) \triangleq \begin{cases} b_1(q,\dot{q})u_1 & \mod(q,360) \in \mathcal{Q} \\ b_2 u_2 & \mod(q,360) \in \mathcal{Q}^c \end{cases}, \quad (2)$$

where $b_1 : \mathbb{R}\times\mathbb{R} \to \mathbb{R}_{>0}$ is the assumed known muscle control effectiveness, $u_1 \in \mathbb{R}$ is the muscle control input, $b_2 \in \mathbb{R}$ is the known motor control constant, and $u_2 \in \mathbb{R}$ is the motor control input. From (2) the dynamics for each mode are [15]

$$b_k u_k = M(q)\ddot{q} + V_c(q,\dot{q})\dot{q} + G(q) + P(q,\dot{q}) + b_c\dot{q}, \quad (3)$$

where k represents the active switched subsystem. Let $k \in \mathbb{S}$, where $\mathbb{S} \triangleq \{1,2\}$ is the switching index set.

A. Background Information

Following the development in [17], the dynamics in (3) can be rewritten in the control-affine form[1]

$$\dot{x} = f(x) + g_k(x)u_k, \quad (4)$$

where $x \triangleq [q,\dot{q}]^T$, and a subsequently defined control input $u_k \in \mathbb{R}$ represents the control input for the k^{th} system. The drift dynamics $f : \mathbb{R}^2 \to \mathbb{R}^2$ are defined as $f(x) \triangleq \begin{bmatrix} \dot{q} \\ M(q)^{-1}(-V(q,\dot{q})\dot{q} - G(q) - P(q,\dot{q}) - b_c\dot{q}) \end{bmatrix}$, and the control effectiveness $g_k : \mathbb{R}^2 \to \mathbb{R}^2$ is defined as

$$g_k(x) \triangleq \begin{cases} \left[0, M(q)^{-1}b_1(q,\dot{q})\right]^T & \mod(q,360) \in \mathcal{Q} \\ \left[0, M(q)^{-1}b_2\right]^T & \mod(q,360) \in \mathcal{Q}^c \end{cases}.$$

The control objective is to track a time-varying continuously differentiable signal $x_d \in \mathbb{R}^2$. To quantify the tracking objective, the tracking error $e \in \mathbb{R}^2$ is defined as $e \triangleq x - x_d$. Using the technique in [17], the control affine dynamics in (4) can be expressed as

$$\dot{\zeta} = F_k(\zeta) + G_k(\zeta)\mu_k, \quad (5)$$

where $\zeta \in \mathbb{R}^4$ is the concatenated state $\zeta \triangleq \left[e^T, x_d^T\right]^T$, $\mu_k \triangleq u_k - u_{d,k}(x_d)$ is the transient component of the controller, $u_{d,k} : \mathbb{R}^2 \to \mathbb{R}$ is the subsequently-defined trajectory tracking component of the controller, $F_k : \mathbb{R}^4 \to \mathbb{R}^4$ is the concatenated drift dynamics defined as $F_k(\zeta) \triangleq \begin{bmatrix} f(e - x_d)^T - h_d(x_d)^T + u_{d,k}(x_d)g_k(e - x_d)^T \\ h_d(x_d)^T \end{bmatrix}$, and $G_k : \mathbb{R}^4 \to \mathbb{R}^4$ is the concatenated control effectiveness defined as $G_k(\zeta) \triangleq \left[g_k(e - x_d)^T, \mathbf{0}_{1\times 2}\right]^T$. Furthermore, $h_d : \mathbb{R}^2 \to \mathbb{R}^2$ is a locally Lipschitz function such that $h_d(x_d) \triangleq \dot{x}_d$. The following properties and assumptions facilitate the development of the desired approximate optimal tracking controller.

Property 1. The drift dynamics f are continuously differentiable [15], which, using [18, Lemma 3.2], means that f is a locally Lipschitz function and $f(0) = 0$.

[1]The cycle-rider dynamics do not differ between modes. The only difference between the switching modes is the actuation methods.

Property 2. The control effectiveness matrix g_k is continuously differentiable [15] and therefore a locally Lipschitz function [18, Lemma 3.2]. The matrix g_k is bounded such that $0 < \|g_k(x)\| \leq \overline{g}_k \; \forall x \in \mathbb{R}^n$, where $\overline{g}_k \in \mathbb{R}_{>0}$ is the supremum over all x of the maximum singular value of $g_k(x)$, respectively, for all k. It follows that $\|G_k(\zeta)\| \leq \overline{G}$ [19].

Assumption 1. The desired trajectory is upper-bounded by a known positive constant $\overline{x_d} \in \mathbb{R}$ such that $\sup_{t \in \mathbb{R}_{\geq 0}} \|x_d\| \leq \overline{x_d}$ [17].

Assumption 2. $g_k^+ : \mathbb{R}^2 \to \mathbb{R}^{1\times 2}$ is the left pseudoinverse, defined as $g_k^+(x) \triangleq \left(g_k^T(x) g_k(x)\right)^{-1} g_k^T(x)$, where $g_k(x_d) g_k^+(x_d)(h_d(x_d) - f(x_d)) = h_d(x_d) - f(x_d)$, $\forall t \in \mathbb{R}_{\geq 0}$, $\forall k \in \mathbb{S}$ [17].

Based on the above assumptions, the trajectory tracking component of the controller $u_{d,k}(x_d)$ is defined as $u_{d,k}(x_d) \triangleq g_k^+(x_d)(h_d(x_d) - f(x_d))$.

B. Control Objective

The control objective is to solve the infinite-horizon optimal tracking problem i.e. to find a control policy μ_k that minimizes the cost function

$$J_k(\zeta, \mu_k) \triangleq \int_{t_0}^{\infty} \zeta^T \overline{Q}_k \zeta + \mu_k^T R_k \mu_k \, d\tau, \tag{6}$$

where $\overline{Q}_k \in \mathbb{R}^{4\times 4}$ is a user-defined positive semidefinite (PSD) symmetric cost matrix, and $R_k \in \mathbb{R}_{>0}$ is a positive constant. Let $\overline{Q}_k \triangleq \operatorname{diag}\{Q_k, \mathbf{0}_{2\times 2}\}$, where $Q_k \in \mathbb{R}^{2\times 2}$ is a positive definite (PD) cost matrix. Note that $\overline{Q}_k$ is PSD and Q_k is PD so that the cost in (6) does not depend on the desired trajectory.

Property 3. The state cost matrix Q_k satisfies $\underline{q}_k \leq Q_k \leq \overline{q}_k$ where $\underline{q}_k, \overline{q}_k \in \mathbb{R}_{>0}$ are the minimum and maximum eigenvalues of Q_k, respectively.

The infinite horizon value function (i.e. the cost-to-go) for the k^{th} mode $V_k^* : \mathbb{R}^4 \to \mathbb{R}_{\geq 0}$ is defined as

$$V_k^*(\zeta) \triangleq \min_{\mu_k \in U} \int_t^{\infty} \zeta^T \overline{Q}_k \zeta + \mu_k^T R_k \mu_k \, d\tau, \tag{7}$$

where $U \subset \mathbb{R}$ is the action space for μ_k.

Assumption 3. The optimal value function V_k^* is continuously differentiable for all $k \in \mathbb{S}$ [17].

The optimal transient control policy $\mu_k^* : \mathbb{R}^4 \to \mathbb{R}$ is defined as

$$\mu_k^*(\zeta) = -\frac{1}{2} R_k^{-1} G_k(\zeta)^T \left(\nabla V_k^*(\zeta)\right)^T. \tag{8}$$

Each k^{th} optimal value function and optimal control policy satisfy the HJB equation

$$0 = \nabla V_k^*(\zeta)\left(F_k(\zeta) + G_k(\zeta)\mu_k^*\right) + \zeta^T \overline{Q}_k \zeta + \mu_k^{*T} R_k \mu_k^*, \tag{9}$$

which has the boundary condition $V_k^*(0) = 0$.

C. Value Function Approximation

The optimal value function V_k^* is unknown for general nonlinear systems. Let $\Omega \subset \mathbb{R}^4$ be a compact set such that $\zeta \in \Omega$. The value function can be approximated with a NN in Ω by invoking the Stone-Weierstrass Theorem to obtain

$$V_k^*(\zeta) = W_k^T \phi(\zeta) + \epsilon_k(\zeta), \tag{10}$$

where $W_k \in \mathbb{R}^L$ is a vector of unknown weights, $\phi : \mathbb{R}^4 \to \mathbb{R}^L$ is a user-defined vector of basis functions, and $\epsilon_k : \mathbb{R}^4 \to \mathbb{R}$ is the bounded function reconstruction error.[2] Substituting (10) into (8) yields a NN representation of the optimal control policy

$$\mu_k^*(\zeta) = -\frac{1}{2} R_k^{-1} G_k(\zeta)^T \left(\nabla \phi(\zeta)^T W_k + \nabla \epsilon_k(\zeta)^T\right). \tag{11}$$

Assumption 4. There exists a set of known positive constants $\overline{W}, \overline{\phi}, \overline{\nabla\phi}, \overline{\epsilon}, \overline{\nabla\epsilon} \in \mathbb{R}_{>0}$ such that $\sup_{k\in\mathbb{S}} \|W_k\| \leq \overline{W}$, $\sup_{\zeta\in\Omega,\, k\in\mathbb{S}} \|\phi(\zeta)\| \leq \overline{\phi}$, $\sup_{\zeta\in\Omega,\, k\in\mathbb{S}} \|\nabla\phi(\zeta)\| \leq \overline{\nabla\phi}$, $\sup_{\zeta\in\Omega,\, k\in\mathbb{S}} \|\epsilon_k(\zeta)\| \leq \overline{\epsilon}$, and $\sup_{\zeta\in\Omega,\, k\in\mathbb{S}} \|\nabla\epsilon_k(\zeta)\| \leq \overline{\nabla\epsilon}$ [20, Assumptions 9.1.c-e].

The ideal weights W_k are unknown a priori; hence, an approximation of W_k is desired. The critic weight estimate vector $\hat{W}_{c,k} \in \mathbb{R}^L$ is substituted into (10) to obtain the optimal value function estimate $\hat{V}_k : \mathbb{R}^4 \times \mathbb{R}^L \to \mathbb{R}$, defined as

$$\hat{V}_k\left(\zeta, \hat{W}_{c,k}\right) \triangleq \hat{W}_{c,k}^T \phi(\zeta). \tag{12}$$

The actor weight estimate vector $\hat{W}_{a,k} \in \mathbb{R}^L$ is substituted into (11) to obtain the optimal transient control policy estimate $\hat{\mu}_k : \mathbb{R}^4 \times \mathbb{R}^L \to \mathbb{R}$, defined as

$$\hat{\mu}_k\left(\zeta, \hat{W}_{a,k}\right) \triangleq -\frac{1}{2} R_k^{-1} G_k(\zeta)^T \left(\nabla \phi(\zeta)^T \hat{W}_{a,k}\right). \tag{13}$$

The overall controller $u_k \in \mathbb{R}$ is defined as $u_k \triangleq \hat{\mu}_k + u_{d,k}(x_d)$.

III. Bellman Error

To calculate the BE $\delta_k : \mathbb{R}^4 \times \mathbb{R}^L \times \mathbb{R}^L \to \mathbb{R}$, the optimal value function $V_k^*(\zeta)$ and the optimal control policy $\mu_k^*(\zeta)$ in (9) are replaced by the approximate optimal value function $\hat{V}_k\left(\zeta, \hat{W}_{c,k}\right)$ and the approximate optimal control policy $\hat{\mu}_k\left(\zeta, \hat{W}_{a,k}\right)$, respectively, where

$$\begin{aligned}\delta_k\left(\zeta, \hat{W}_{c,k}, \hat{W}_{a,k}\right) &= \zeta^T \overline{Q}_k \zeta \\ &+ \hat{\mu}_k\left(\zeta, \hat{W}_{a,k}\right)^T R_k \hat{\mu}_k\left(\zeta, \hat{W}_{a,k}\right) \\ &+ \nabla \hat{V}_k\left(\zeta, \hat{W}_{c,k}\right)\left(F_k + G_k(\zeta)\hat{\mu}_k\left(\zeta, \hat{W}_{a,k}\right)\right).\end{aligned} \tag{14}$$

The value of the BE indicates how close the actor and critic weight estimates are to their respective ideal weight values. By subtracting (9) from (14), substituting (10)-(13), and denoting the difference between the actual and ideal weight

[2] To focus the scope of this manuscript, each switched system will use the same dimension vector of basis functions $\phi(\zeta)$ i.e., $L_1 = L_2 = \ldots = L_k$

values by $\tilde{W}_{c,k} \triangleq W_k - \hat{W}_{c,k}$ and $\tilde{W}_{a,k} \triangleq W_k - \hat{W}_{a,k}$ the analytical form of the BE in (14) is

$$\delta_k\left(\zeta, \hat{W}_{c,k}, \hat{W}_{a,k}\right) = \frac{1}{4}\tilde{W}_{a,k}^T G_{\phi,k}(\zeta)\tilde{W}_{a,k} - \omega_k^T \tilde{W}_{c,k} + \Theta_k(\zeta), \quad (15)$$

where $\omega_k : \mathbb{R}^4 \times \mathbb{R}^L \to \mathbb{R}^4$ is $\omega_k\left(\zeta, \hat{W}_{a,k}\right) \triangleq \nabla\phi(\zeta)\left(F_k(\zeta) + \hat{\mu}\left(\zeta, \hat{W}_{a,k}\right) G_k(\zeta)\right)$ and $\Theta_k(\zeta) \triangleq \frac{1}{2}W_k^T \nabla\phi(\zeta) G_{R,k} \nabla\epsilon^T + \frac{1}{4}G_{\epsilon,k} - \nabla\epsilon_k F_k$.[3]

Remark 1. Although they are equivalent, (14) is used in implementation and (15) is used in the stability analysis.

Bellman Error Extrapolation

Using the control policy given in (13), the current system state, the critic weight estimate, and the actor weight estimate, the estimated BE in (14) can be evaluated to calculate the instantaneous BE denoted by $\delta_k\left(\zeta, \hat{W}_{c,k}, \hat{W}_{a,k}\right)$ at each time instance $t \in \mathbb{R}_{\geq 0}$. The exploration versus exploitation problem is well-known for learning-based control methods. In results such as [21], an exploration signal is required to successfully explore the operating domain. Results such as [11] use BE extrapolation, which simultaneously evaluates the BE along the system trajectory and at user-defined points in the state space. The BE extrapolation technique eliminates the need for the exploration signal by providing simulation of experience, thus yielding a better value function approximation [11].

The BE is extrapolated from the user-defined off-trajectory points $\{\zeta_i : \zeta_i \in \Omega\}_{i=1}^{N_k}$ set by the user, where $N_k \in \mathbb{N}$ denotes a user-specified number of overall extrapolation trajectories in the compact set Ω. The tuple $(\Sigma_{c,k}, \Sigma_{a,k}, \Sigma_{\Gamma,k})$ represents the data stacks defined as $\Sigma_{c,k} \triangleq \frac{1}{N_k}\sum_{i=1}^{N_k}\frac{\omega_{i,k}}{\rho_{i,k}}\delta_{i,k}$, $\Sigma_{a,k} \triangleq \frac{1}{N_k}\sum_{i=1}^{N_k}\frac{G_{\sigma i,k}^T \hat{W}_{a,k}\omega_{i,k}^T}{4\rho_{i,k}}$, $\Sigma_{\Gamma,k} \triangleq \frac{1}{N_k}\sum_{i=1}^{N_k}\frac{\omega_{i,k}\omega_{i,k}^T}{\rho_{i,k}^2}$, where $\delta_{i,k} \triangleq \delta_k\left(\zeta_i, \hat{W}_{c,k}, \hat{W}_{a,k}\right)$, $\omega_{i,k} \triangleq \omega_k\left(\zeta_i, \hat{W}_{a,k}\right)$, and $\rho_{i,k} = 1 + \nu_k \omega_{i,k}^T \Gamma_k \omega_{i,k}$. $\nu_k \in \mathbb{R}_{>0}$ is a user-defined gain, and $\Gamma_k : \mathbb{R}^{L\times L}$ is a time-varying least-squares gain matrix. Each subsystem has its own distinct set of data, gain values, and update laws.

Assumption 5. On the compact set, Ω, a finite set of off-trajectory points $\{\zeta_i : \zeta_i \in \Omega\}_{i=1}^{N_k}$ exists such that $0 < \underline{c}_k \triangleq \inf_{t\in\mathbb{R}_{\geq 0}} \lambda_{\min}\{\Sigma_{\Gamma,k}\}$ for all $k \in \mathbb{S}$, where $\underline{c}_k$ is a constant scalar lower bound of the value of each input-output data pair's minimum eigenvalues for the k^{th} subsystem [11].

IV. Update Laws for Actor and Critic Weights

The critic and actor weights are updated according to the subsequent laws while each mode is active. In the weight update laws, $\eta_{c1,k}$, $\eta_{c2,k}$, $\eta_{a1,k}$, $\eta_{a2,k}$, $\lambda_k \in \mathbb{R}$ are positive constant learning gains, and $\underline{\Gamma}_k$, $\overline{\Gamma}_k \in \mathbb{R}_{>0}$ are the upper and lower projection operator bounds for Γ_k. The critic update law for the k^{th} mode $\dot{\hat{W}}_{c,k} \in \mathbb{R}^L$ is defined as

$$\dot{\hat{W}}_{c,k} \triangleq \text{proj}\{\Phi_{c,k}\}, \quad (16)$$

where $\Phi_{c,k} \triangleq -\eta_{c1,k}\Gamma_k \frac{\omega_k}{\rho_k}\delta_k - \eta_{c2,k}\Sigma_{c,k}$. The actor update law for the k^{th} mode $\dot{\hat{W}}_{a,k} \in \mathbb{R}^L$ is defined as

$$\dot{\hat{W}}_{a,k} \triangleq \text{proj}\{\Phi_{a,k}\}, \quad (17)$$

where $\Phi_{a,k} \triangleq -\eta_{a1,k}\left(\hat{W}_{a,k} - \hat{W}_{c,k}\right) - \eta_{a2,k}\hat{W}_{a,k} + \frac{\eta_{c1,k}G_{\phi,k}^T \hat{W}_{a,k}\omega_k^T}{4\rho_k}\hat{W}_{c,k} + \eta_{c2,k}\Sigma_{a,k}\hat{W}_{c,k}$. The operator $\text{proj}\{\cdot\}$ denotes the smooth projection operator defined in [22, Appendix E, Eq. E.4] and is designed such that $\left\|\hat{W}_{c,k}\right\| \in \left[\underline{\hat{W}}_{c,k}, \overline{\hat{W}}_{c,k}\right]$ and $\left\|\hat{W}_{a,k}\right\| \in \left[\underline{\hat{W}}_{a,k}, \overline{\hat{W}}_{a,k}\right]$ under the assumption $\|W_k\| \in \left[\underline{\hat{W}}_k, \overline{\hat{W}}_k\right]$. The least-squares gain matrix update law of the k^{th} mode $\dot{\Gamma}_k \in \mathbb{R}^{L\times L}$ is

$$\dot{\Gamma}_k \triangleq \left(\lambda_k\Gamma_k - \eta_{c1,k}\frac{\Gamma_k\omega_k\omega_k^T\Gamma_k}{\rho_k^2} - \eta_{c2,k}\Gamma_k\Sigma_{\Gamma,k}\Gamma_k\right) \cdot \mathbf{1}_{\left\{\underline{\Gamma}_k \leq \|\Gamma_k\| \leq \overline{\Gamma}_k\right\}}, \quad (18)$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.[4] While the k^{th} mode is inactive $\dot{\hat{W}}_{c,k} = \mathbf{0}_{L\times 1}$, $\dot{\Gamma}_k = \mathbf{0}_{L\times L}$, and $\dot{\hat{W}}_{a,k} = \mathbf{0}_{L\times 1}$.[5]

Remark 2. Under Assumptions 1-3, the PD solution of the HJB equation is the optimal value function for each system. The approximation of the PD solution to the HJB equation is guaranteed by the appropriate selection of Lyapunov-based update laws and initial weight estimates [23].

V. Stability Analysis

It is possible for a switched system to become unstable, even if the individual subsystems of a switched system are stable [12, Ch. 3]. Hence, the stability of each subsystem must be investigated along with the switching between the systems. In the subsequent development, k subsystems, each with a class of dynamics in (4), are analyzed with the control policy in (13) and update laws outlined in (16)-(18).

A. Subsystem Stability Analysis

Since the state penalty matrix $\overline{Q}_k$ is PSD, the optimal value function V_k^* is PSD and is not a valid Lyapunov function. However, a nonautonomous form of the optimal value function denoted as $V_{t,k}^* : \mathbb{R}^2 \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is defined such that $V_{t,k}^*(e,t) = V_k^*(\zeta)$, and is PD and decrescent [17].

To facilitate the stability analysis, let $z_k \in \mathbb{R}^{2+2L}$ be a concatenated state vector defined as $z_k \triangleq \begin{bmatrix} e^T & \tilde{W}_{c,k}^T & \tilde{W}_{a,k}^T \end{bmatrix}^T$, and let $V_{L,k} : \mathbb{R}^{2+2L} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a candidate Lyapunov function defined as

$$V_{L,k}(z_k, t) \triangleq V_{t,k}^*(e,t) + \frac{1}{2}\tilde{W}_{c,k}^T\Gamma_k^{-1}\tilde{W}_{c,k} + \frac{1}{2}\tilde{W}_{a,k}^T\tilde{W}_{a,k}. \quad (19)$$

[3] $G_{R,k}$, $G_{\phi,k}$, and $G_{\epsilon,k}$ are defined as $G_{R,k}(\zeta) \triangleq G_k(\zeta)R_k^{-1}G_k(\zeta)^T$, $G_{\phi,k}(\zeta) \triangleq \nabla_\zeta\phi(\zeta)G_{R,k}(\zeta)\nabla_\zeta\phi(\zeta)^T$, and $G_{\epsilon,k}(\zeta) \triangleq \nabla_\zeta\epsilon_k(\zeta)G_{R,k}(\zeta)\nabla_\zeta\epsilon_k(\zeta)^T$ respectively.

[4] Using (18) ensures that each $\underline{\Gamma}_k \leq \|\Gamma_k\| \leq \overline{\Gamma}_k$ for all $t \in \mathbb{R}_{>0}$.

[5] The update laws will not update a subsystem k's weight estimates or least-squares matrix unless subsystem k is active.

According to [17] and [18, Lemma 4.3], (19) can generally be bounded as $\underline{v}_{l,k}(\|z_k\|) \leq V_{L,k}(z_k) \leq \overline{v}_{l,k}(\|z_k\|)$ using class $\mathcal{K}$ functions $\underline{v}_{l,k}, \overline{v}_{l,k} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. To facilitate the subsequent dwell-time analysis, the following more restrictive assumption is required.

Assumption 6. The optimal value function $V^*_{t,k}(e,t)$ can be bounded by the square of the norm of its argument times a positive constant, i.e.,[6]

$$\beta_{1,k}\|e\|^2 \leq V^*_{t,k}(e,t) \leq \beta_{2,k}\|e\|^2,$$

for all $k \in \mathbb{S}$, where $\beta_{1,k}$, $\beta_{2,k} \in \mathbb{R}_{\geq 0}$.

Using Assumption 6, (19) can be bounded as

$$\alpha_{1,k}\|z_k\|^2 \leq V_{L,k}(z_k,t) \leq \alpha_{2,k}\|z_k\|^2, \tag{20}$$

where $\alpha_{1,k}, \alpha_{2,k} \in \mathbb{R}_{>0}$.

The normalized regressors $\frac{\omega_k}{\rho_k}$ and $\frac{\omega_{i,k}}{\rho_{i,k}}$ are bounded as $\sup_{t\in\mathbb{R}_{\geq 0}}\left\|\frac{\omega_k}{\rho_k}\right\| \leq \frac{1}{2\sqrt{\nu_k \underline{\Gamma}_k}}$ and $\sup_{t\in\mathbb{R}_{\geq 0}}\left\|\frac{\omega_{i,k}}{\rho_{i,k}}\right\| \leq \frac{1}{2\sqrt{\nu_k \underline{\Gamma}_k}}$ for all $\zeta \in \Omega$ and $\zeta_i \in \Omega$, respectively. $G_{R,k}$ is bounded as $\sup_{\zeta\in\Omega}\|G_{R,k}\| \leq \overline{G}^2 \lambda_{\max}\{R^{-1}\}$, and $G_{\phi,k}$ is bounded as $\sup_{\zeta\in\Omega}\|G_{\phi,k}\| \leq \left(\overline{\nabla\phi G}\right)^2 \lambda_{\max}\{R^{-1}\}$.

Remark 3. Using the projection operator from the critic update law in (16) and [22, Lemma E.1], $-\tilde{W}^T_{c,k}\Gamma_k^{-1}\dot{\hat{W}}_{c,k}$ is bounded from above as

$$\begin{aligned}-\tilde{W}^T_{c,k}\Gamma_k^{-1}\dot{\hat{W}}_{c,k} &= -\tilde{W}^T_{c,k}\Gamma_k^{-1}\text{proj}\{\Phi_{c,k}\}\\ &\leq -\tilde{W}^T_{c,k}\Gamma_k^{-1}\Phi_{c,k}.\end{aligned}$$

Using the projection operator from the actor update law in (17) and [22, Lemma E.1], $-\tilde{W}^T_{a,k}\dot{\hat{W}}_{a,k}$ is bounded from above as

$$\begin{aligned}-\tilde{W}^T_{a,k}\dot{\hat{W}}_{a,k} &= -\tilde{W}^T_{a,k}\text{proj}\{\Phi_{a,k}\}\\ &\leq -\tilde{W}^T_{a,k}\Phi_{a,k}.\end{aligned}$$

To facilitate the subsequent analysis, let $\mathcal{R} \in \mathbb{R}_{>0}$ be the radius of a compact ball $\mathcal{B}_\mathcal{R} \subset \mathbb{R}^2 \times \mathbb{R}^L \times \mathbb{R}^L$ centered at the origin.

Theorem 1. *While each subsystem is active, if Assumptions 1-6 hold, the control policy in (13) and the weight update laws in (16)-(18) are implemented, and the conditions*

$$\eta_{a1,k} + \eta_{a2,k} \geq \frac{1}{\sqrt{\nu_k \underline{\Gamma}_k}}(\eta_{c1,k} + \eta_{c2,k})\overline{WG_\phi}, \tag{21}$$

$$\underline{c}_k \geq \frac{3\eta_{a1,k}}{\eta_{c2,k}} + \frac{3(\eta_{c1,k}+\eta_{c2,k})^2 \overline{W}^2 \overline{G_\phi}^2}{16\eta_{c2,k}\nu_k\underline{\Gamma}_k(\eta_{a1,k}+\eta_{a2,k})}, \tag{22}$$

$$\sqrt{\frac{2\alpha_{2,k}l_k}{\alpha_{1,k}\Lambda_k}} < \mathcal{R}, \tag{23}$$

are satisfied for each individual subsystem, then the tracking error e_k, the critic weight estimate error $\tilde{W}_{c,k}$, and the actor weight estimate error $\tilde{W}_{a,k}$ are uniformly ultimately bounded (UUB) . Therefore, the transient tracking control policy $\hat{\mu}_k$ converges to a neighborhood of the optimal control policy μ^*_k.

Proof: Using (5) and the fact that $V^*_{t,k}(e,t) = V^*_k(\zeta)$, $\forall e \in \mathbb{R}^2$, $t \in \mathbb{R}_{\geq 0}$ and taking the time derivative of the candidate Lyapunov function in (19) yields

$$\begin{aligned}\dot{V}_{L,k}(z_k) = \nabla V^*_k\dot{\zeta} - \tilde{W}^T_{c,k}\Gamma_k^{-1}\dot{\hat{W}}_{c,k} - \tilde{W}^T_{a,k}\dot{\hat{W}}_{a,k}\\ - \frac{1}{2}\tilde{W}^T_{c,k}\Gamma_k^{-1}\dot{\Gamma}_k\Gamma_k^{-1}\tilde{W}_{c,k},\end{aligned} \tag{24}$$

where the fact that $\frac{d}{dt}\Gamma_k^{-1} = \Gamma_k^{-1}\dot{\Gamma}_k\Gamma_k^{-1}$ is used. Under the sufficient gain conditions in (21) and (22), and using (9), (15), and the update laws in (16)-(18), the expression in (24) can be bounded as

$$\dot{V}_{L,k}(z_k) \leq -\frac{\Lambda_k}{2\alpha_{2,k}}V_{L,k}(z_k)\,\forall\sqrt{\frac{2l_k}{\Lambda_k}} \leq \|z_k\| \leq \mathcal{R}, \tag{25}$$

for all $k \in \mathbb{S}$ and $t \in \mathbb{R}_{\geq 0}$, where $\Lambda_k \triangleq \min\left[\ \lambda_{\min}(Q_k),\ \ \frac{1}{6}\eta_{c2,k}\underline{c}_k,\ \ \frac{1}{8}(\eta_{a1,k}+\eta_{a2,k})\ \right]$, and l_k is a positive constant that depends on the control gains and NN bounding constants in Assumption 4. Using the bounds in (20), the time derivative in (24), Λ_k, and (23), [18, Thm. 4.18] can be invoked to prove that z_k is UUB such that $\limsup_{t\to\infty}\|z_k\| \leq \sqrt{\frac{2\alpha_{2,k}l_k}{\alpha_{1,k}\Lambda_k}}$, and the transient tracking control policy $\hat{\mu}_k$ converges to a neighborhood of the optimal control policy μ^*_k. Since $z_k \in \mathcal{L}_\infty$, it follows that $e, \tilde{W}_{c,k}, \tilde{W}_{a,k} \in \mathcal{L}_\infty$, and since $\hat{\mu}_k \in \mathcal{L}_\infty$ and $x_d \leq \overline{x}_d$, it follows that $u_k \in \mathcal{L}_\infty$. Furthermore, every trajectory z_k that is initialized in the ball $\mathcal{B}_\mathcal{R}$ is bounded such that $z_k \in \mathcal{B}_\mathcal{R}$, $\forall t \in \mathbb{R}_{\geq 0}$, $\forall k \in \mathbb{S}$. Since $z_k \in \mathcal{B}_\mathcal{R}$, it follows that the individual elements of z_k lie in a compact set, i.e. $e, \tilde{W}_{c,k}$, and $\tilde{W}_{a,k}$ lie in a compact set. Additionally, since $x_d \leq \overline{x}_d$, then the concatenated state $\zeta \in \Omega$, $\forall t \in \mathbb{R}_{\geq 0}$, $\forall k \in \mathbb{S}$, which facilitates value function approximation. ■

Remark 4. See [11] for insight into satisfying the gain conditions in (21) and (22).

B. Switched Subsystems

Let $t^{ON}_k \in [0,t]$ denote a time instant when the k^{th} subsystem of the switching sequence is activated. Let $t^{OFF}_k \in [0,t]$ denote a time instant when the k^{th} subsystem in the switching sequence is deactivated. The dwell-time in any active mode of a subsystem denoted by $\tau_k \in \mathbb{R}_{>0}$ is defined as $\tau_k = t^{OFF}_k - t^{ON}_k$ and represents the amount of time a subsystem must be active before switching to the next. The minimum dwell time for any active mode of a system is denoted by $\tau^* \in \mathbb{R}_{>0}$. There are a finite number of switches, and $N_\sigma \in \mathbb{N}_{<\infty}$ denotes the number of switching events. The sequence of time instants at which a switching event occurs is defined as $\{t^{ON}_{N_\sigma}\}$, such that $0 = t^{ON}_1 < t^{ON}_2 < \cdots < t^{ON}_{N_\sigma} < t^{ON}_{N_\sigma+1}$.

C. Dwell-Time Analysis

The stability analysis proves that each subsystem is UUB while active. However, the Lyapunov function for the overall switching system may instantaneously increase due to the change in the optimal value function and set of new weight parameters. The value function corresponding to mode $k+1$,

[6] Assumption 6 is a stricter version of [10, Lemma 3.14].

$V^*_{t,k+1}(\zeta)$, may be larger than the value function corresponding to mode k, $V^*_{t,k}(\zeta)$. Similarly, the magnitude of the actor and critic weight errors could be larger in mode $k+1$ than in mode k. Therefore, a dwell time condition must be designed to account for switching between the subsystems which ensures that the switched system is stable [12, Ch. 3].

Theorem 2. *The system consisting of a family of subsystems with the dynamics in (4) with a properly designed minimum dwell-time, $\tau^* \in \mathbb{R}_{>0}$ ensures that the tracking error, critic estimate errors, and actor estimate errors will converge to a neighborhood of the origin in the sense that $\|z_k\| \leq \max_{k\in\mathbb{S}} \sqrt{\frac{2\alpha_{2,k} l_k}{\alpha_{1,k}\Lambda_k}}$ for all $t \geq T$, where $\max_{k\in\mathbb{S}} \sqrt{\frac{2\alpha_{2,k} l_k}{\alpha_{1,k}\Lambda_k}} \in \mathbb{R}_{>0}$ is the maximum ultimate bound for all subsystems, and $T \in \mathbb{R}_{\geq 0}$ is the time required to reach the ultimate bound.*

The proof follows that of [13, Theorem 2] and is available upon request.

Remark 5. From Section II, the system switches modes based on if the crank angle q belongs to $\mathcal{Q}$ or $\mathcal{Q}^c$, i.e., the switching is state-based; furthermore the switches occur more frequently at higher desired cadence values. The user cannot directly control the time of the switching instances. For the system to be stable using the previous analysis, the dwell-time must be significantly smaller than the time required to travel through the regions $\mathcal{Q}$ and $\mathcal{Q}^c$. The dwell-time τ^* is composed of many user-selected parameters. Notably, τ^* can be decreased by increasing the decay rate γ_0, i.e. stronger convergence parameters result in a shorter dwell-time. The dwell time τ^* is inversely proportional to the decay rate γ_0, and γ_0 is proportional to Λ. Hence, maximizing Λ will decrease the dwell time. This is achieved by maximizing the term $\min\left[\lambda_{\min}(Q_k), \frac{1}{6}\eta_{c2,k}\underline{c}_k, \frac{1}{8}(\eta_{a1,k}+\eta_{a2,k})\right]$. While this maximization decreases the dwell-time so that it is significantly smaller than the time dictated by the desired trajectory, there are some practical drawbacks. A larger state cost matrix $\overline{Q}_k$ will increase the penalty on the error; paired with larger actor and critic learning gains $\eta_{a1,k}$, $\eta_{a2,k}$, and $\eta_{c2,k}$, this may lead to a more aggressive controller, which may cause rider discomfort. Furthermore, increasing $\underline{c}_k$ relies on using more BE extrapolation data pairs, which may become computationally intensive. Motivated by these practical considerations, additional analysis methods that can potentially eliminate the need for a minimum dwell-time are motivated.

VI. Conclusion

This paper develops an ADP-based controller for switched cycle dynamics while achieving a time-varying tracking objective. The stability of each individual subsystem is proven by a Lyapunov-based analysis, and the stability of the overall switched system is proven via a dwell-time analysis. The entire switched system is proven to be UUB such that the control policy is proven to converge to a neighborhood of the optimal policy and to track the cadence within a neighborhood of its desired value. Future work will investigate the application of the developed controller to an FES-cycling testbed and the development of analysis methods free of minimum dwell-time requirements.

References

[1] S. Ferrante, A. Pedrocchi, G. Ferrigno, and F. Molteni, "Cycling induced by functional electrical stimulation improves the muscular strength and the motor control of individuals with post-acute stroke," *Eur. J. Phys. Rehabil. Med.*, vol. 44, no. 2, pp. 159–167, 2008.

[2] S. P. Hooker, S. F. Figoni, M. M. Rodgers, R. M. Glaser, T. Mathews, A. G. Suryaprasad, and S. C. Gupta, "Physiologic effects of electrical stimulation leg cycle exercise training in spinal cord injured persons," *Arch. Phys. Med. Rehabil.*, vol. 73, no. 5, pp. 470–476, 1992.

[3] F. Anaya, P. Thangavel, and H. Yu, "Hybrid FES–robotic gait rehabilitation technologies: a review on mechanical design, actuation, and control strategies," *Int. J. Intell. Robot. Appl.*, pp. 1–28, 2018.

[4] D. Kuhn, V. Leichtfried, and W. Schobersberger, "Four weeks of functional electrical stimulated cycling after spinal cord injury: a clinical study," *Int. J. Rehab. Res.*, vol. 37, pp. 243–250, March 2014.

[5] V. R. Edgerton, R. D. de Leon, S. J. Harkema, J. A. Hodgson, N. London, D. J. Reinkensmeyer, R. R. Roy, R. J. Talmadge, N. J. Tillakaratne, W. Timoszyk, and A. Tobin, "Retraining the injured spinal cord," *J. Physiol.*, vol. 533, no. 1, pp. 15–22, 2001.

[6] T. R. Oliveira, L. R. Costa, and A. V. Pino, "Extremum seeking applied to neuromuscular electrical stimulation," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 188–193, 2016.

[7] P. Paz, T. R. Oliveira, A. V. Pino, and A. P. Fontana, "Model-free neuromuscular electrical stimulation by stochastic extremum seeking," *IEEE Trans. Control Sys. Tech.*, vol. 28, no. 1, pp. 238–253, 2020.

[8] D. Kirk, *Optimal Control Theory: An Introduction*. Mineola, NY: Dover, 2004.

[9] F. L. Lewis and D. Liu, *Reinforcement learning and approximate dynamic programming for feedback control*. John Wiley & Sons, 2013, vol. 17.

[10] R. Kamalapurkar, P. S. Walters, J. A. Rosenfeld, and W. E. Dixon, *Reinforcement learning for optimal feedback control: A Lyapunov-based approach*. Springer, 2018.

[11] R. Kamalapurkar, P. Walters, and W. E. Dixon, "Model-based reinforcement learning for approximate optimal regulation," *Automatica*, vol. 64, pp. 94–104, 2016.

[12] D. Liberzon, *Switching in Systems and Control*. Birkhauser, 2003.

[13] M. Greene, M. Abudia, R. Kamalapurkar, and W. E. Dixon, "Model-based reinforcement learning for optimal feedback control of switched systems," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 162–167.

[14] A. Parikh, T.-H. Cheng, R. Licitra, and W. E. Dixon, "A switched systems approach to image-based localization of targets that temporarily leave the camera field of view," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 6, pp. 2149–2156, 2018.

[15] M. Bellman, "Control of cycling induced by functional electrical stimulation: A switched systems theory approach," Ph.D. dissertation, University of Florida, 2015.

[16] M. J. Bellman, R. J. Downey, A. Parikh, and W. E. Dixon, "Automatic control of cycling induced by functional electrical stimulation with electric motor assistance," *IEEE Trans. Autom. Science Eng.*, vol. 14, no. 2, pp. 1225–1234, April 2017.

[17] R. Kamalapurkar, H. Dinh, S. Bhasin, and W. E. Dixon, "Approximate optimal trajectory tracking for continuous-time nonlinear systems," *Automatica*, vol. 51, pp. 40–48, Jan. 2015.

[18] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[19] C. A. Cousin, C. A. Rouse, V. H. Duenas, and W. E. Dixon, "Controlling the cadence and admittance of a functional electrical stimulation cycle," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1181–1192, June 2019.

[20] D. Vrabie, K. G. Vamvoudakis, and F. L. Lewis, *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. The Institution of Engineering and Technology, 2013.

[21] K. G. Vamvoudakis, D. Vrabie, and F. L. Lewis, "Online adaptive algorithm for optimal control with integral reinforcement learning," *Int. J. of Robust and Nonlinear Control*, vol. 24, no. 17, pp. 2686–2710, 2014.

[22] M. Krstic, I. Kanellakopoulos, and P. V. Kokotovic, *Nonlinear and Adaptive Control Design*. New York, NY, USA: John Wiley & Sons, 1995.

[23] P. Deptula, Z. Bell, E. Doucette, W. J. Curtis, and W. E. Dixon, "Data-based reinforcement learning approximate optimal control for an uncertain nonlinear system with control effectiveness faults," *Automatica*, vol. 116, pp. 1–10, June 2020.