Output Feedback Speed Control for a Wankel Rotary Engine via Q-Learning

DOI:
10.1016/j.ifacol.2023.10.1014

Document Version
Final published version

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 8278–8283

Output Feedback Speed Control for a Wankel Rotary Engine via Q-Learning

Anthony Siming Chen [*,**] **Guido Herrmann** [*]
Stuart Burgess [**] **Chris Brace** [***]

[*] *Department of Electrical and Electronic Engineering, University of Manchester, Manchester, M13 9PL, UK (e-mail: siming.chen@manchester.ac.uk; guido.herrmann@manchester.ac.uk).*
[**] *Department of Mechanical Engineering, University of Bristol, Bristol, BS8 1TR, UK (e-mail: s.c.burgess@bristol.ac.uk).*
[***] *Institute for Advanced Automotive Propulsion Systems (IAAPS), University of Bath, Bath, BA2 7AY, UK (e-mail: c.j.brace@bath.ac.uk).*

Abstract: This paper develops a dynamic output feedback controller based on continuous-time Q-learning for the engine speed regulation problem. The proposed controller is able to learn the optimal control solution online in a finite time using only the measurable outputs. We first present the mean value engine model (MVEM) for a Wankel rotary engine. The regulation of engine speed can be formulated as an optimal control problem that minimises a pre-defined value function by actuating the electronic throttle. By parameterising an action-dependent Q-function, we derive a full-state adaptive optimal feedback controller using the idea of continuous-time Q-learning. The adaptive critic approximates the Q-function as a neural network and directly updates the actor, where the convergence is guaranteed by employing novel finite-time adaptation techniques. Then, we incorporate the extended Kalman filter (EKF) as an optimal reduced-order state observer, which enables the online estimation of the unknown fuel puddle dynamics, to achieve a dynamic output feedback engine speed controller. The simulation results of a benchmark 225CS engine demonstrate that the proposed controller can effectively regulate the engine speed to a set point under certain load disturbances.

Keywords: Engine control; Q-learning; Nonlinear observer; Adaptive control

1. INTRODUCTION

Recent ideas of incorporating reinforcement learning principles into feedback control have prompted extensive research on adaptive optimal control. In the computational intelligence community, reinforcement learning techniques are widely studied to solve model-free optimisation problems, where for instance an agent needs to select control actions in an environment to minimise (or maximise) its cumulative cost (or reward). The philosophy behind reinforcement learning is strongly connected to direct adaptive optimal control from a theoretical perspective (Sutton et al. (1992)). This is also referred to as approximate/adaptive dynamic programming (ADP) in the control community (Werbos (1992)). Most studies on ADP or reinforcement learning are based on the framework of Markov decision process (MDP), which is stochastic and in discrete time. The major difficulty of applying reinforcement learning techniques for continuous-time deterministic problems lies in the Hamilton-Jacobi-Bellman (HJB) equation, which requires *a priori* knowledge of the system dynamics and is usually analytically difficult to solve. Vrabie et al. (2009) proposed an adaptive optimal controller for a continuous-time linear quadratic regulation (LQR) problem and it then forms an integral reinforcement learning (IRL) approach to deal with general nonlinear affine systems (Vrabie et al. (2013)). The IRL generates a family of algorithms but most of them require complete or at least partial knowledge of the system dynamics.

Applying the idea of Q-learning and novel adaptive control techniques, our recent work (Chen and Herrmann (2019)) derived a model-free adaptive optimal controller for general unknown nonlinear affine systems. The controller performs in continuous time in a non-iterative manner and online learns the optimal solution without *a priori* knowledge of system dynamics. However, in all these results, the control design requires full-state feedback, i.e., all the state variables need to be available or directly measurable. For practical problems, it is often the case that not all the states are measurable for the controller, where the output feedback control scheme becomes necessary. Lewis and Vamvoudakis (2010) presented the output feedback ADP for discrete-time linear systems and Gao et al. (2016) developed one for continuous-time nonlinear systems using a sampled-data approach.

Speed control represents one of the most basic yet challenging automotive engine control problems, where the improvements in robustness and performance can directly result in better fuel economy, emissions, and drivability (Thornhill et al. (2000)). For example, an idle speed needs to be regulated close to the set point. In a vehicle, the load disturbance due to the events such as power steering, transmission engagement, or low-speed manoeuvring may cause idle speed excursion. Advanced techniques have been considered in the literature for idle speed control including H_∞ loop shaping (Sun et al. (2002)), adaptive control (Yildiz et al. (2007)), sliding mode control (Li and Yurkovich (2001)), model predictive control (Di Cairano et al. (2011)), etc.

10.1016/j.ifacol.2023.10.1014

In this paper, we design a novel dynamic output feedback controller using continuous-time Q-learning and extended Kalman filter (EKF) for the engine speed regulation problem. The main contributions are summarised as follows.

1) Different from the existing approaches, the proposed Q-learning-based engine speed controller is able to learn the optimal solution (throttle angle profile) online in finite time using only the measurable outputs, namely, the intake manifold pressure, temperature, and engine speed.

2) Q-learning is applied to the engine speed regulation problem in a continuous-time framework (in contrast to common iterative ADP algorithms).

3) The state feedback Q-learning combined with the EKF yields an observer-based dynamic output feedback Q-learning controller.

2. REVIEW OF WANKEL ENGINE MODEL

This section presents a physics-based mean value engine model (MVEM) for the engine speed control design. The model was initially developed by Hendricks and Luther (2001) for reciprocating engines and then enhanced in our work (Chen et al. (2018) and Chen et al. (2020)) for a Wankel rotary engine. The research on Wankel engine modelling and control is motivated by recent interest in using it as a range extender for hybrid electric vehicles (HEV). The engine operates with the same Otto cycle, i.e., a single-rotor Wankel engine is equivalent to a two-cylinder four-stroke reciprocating engine.

2.1 Throttle Body Model

Assuming the one-dimensional, steady, isentropic compressible flow of an ideal gas, the air mass flow rate $\dot{m}_{at}$ passing the throttle can be described as a linear function of the throttle angle α as

$$\dot{m}_{at} = K_\alpha \alpha \tag{1}$$

with the linearised flow rate sensitivity K_α (Cook and Powell (1988)). The throttle angle α is one of the major control inputs in the engine speed control system (while older engines often use an air bypass valve).

2.2 Intake Manifold Model

The air flow passing the throttle will enter the intake manifold before entering the housing. An adiabatic model (which was found to be more accurate than an isothermal one) of the air-filling dynamics (Hendricks (2001)) in the intake manifold can be given in terms of the pressure p_m and the temperature T_m as

$$\dot{p}_m = \frac{\kappa R}{V_m}(\dot{m}_{at} T_a - \dot{m}_a T_m) \tag{2}$$

$$\dot{T}_m = \frac{R T_m}{p_m V_m}[\dot{m}_{at}(T_a \kappa - T_m) - \dot{m}_a (T_m \kappa - T_m)] \tag{3}$$

where p_a and T_a are the ambient pressure and temperature, respectively; κ is the ratio of the specific heats; R is the ideal gas constant; V_m is the manifold volume; the port air flow rate $\dot{m}_a$ can be given as a nonlinear function of the intake manifold pressure p_m and engine speed N such that

$$\dot{m}_a(p_m, N) = \frac{V_d}{120 R T_m}\eta_{vol}(p_m, N) p_m N \tag{4}$$

where V_d is the engine displacement and η_{vol} is the volumetric efficiency that can be determined from a static map with respect to p_m and N.

2.3 Fuel Puddle Model

The fuelling of the engine is controlled by a fuel injector fitted near the port on the intake manifold. Due to the port fuel injection (PFI) configuration, a fraction of the injected fuel is deposited on the manifold walls and becomes fuel puddles, which is referred to as the "wall-wetting" phenomenon. The final fuel flow rate $\dot{m}_f$ entering the housing is the sum of the fuel puddle flow rate $\dot{m}_{fpe}$ and the fuel vapour flow rate $\dot{m}_{fve}$ entering the combustion chamber

$$\dot{m}_f = \dot{m}_{fpe} + \dot{m}_{fve} = m_{fp}/\tau_p + m_{fv}/\tau_m \tag{5}$$

where τ_p and τ_m are the characteristic manifold time constants for the puddle m_{fp} and vapour m_{fv} fuel mass, respectively. Their dynamics can be taken as a set of two first-order processes with a time constant τ_f as

$$\begin{cases} \dot{m}_{fp} = \chi \dot{m}_{fi} - (1/\tau) m_{fp} - \dot{m}_{fpe} \\ \dot{m}_{fv} = (1-\chi)\dot{m}_{fi} + (1/\tau) m_{fp} - m_{fv}/\tau_m \end{cases} \tag{6}$$

where $\dot{m}_{fi}$ is the injected fuel flow rate (i.e., the control input in the air-fuel ratio (AFR) control system) and χ ($0 \le \chi < 1$) is the fraction of injected fuel that deposits on the manifold walls as fuel puddles (see Arsie et al. (2003) for more details).

2.4 Combustion Model

For the sake of simplicity, we assume an ideal fuelling of the engine and there is no crevice or leakage between the chambers. Hence, the AFR $\lambda = \dot{m}_a/\dot{m}_f$ is regulated around the stoichiometric value, i.e., $\lambda_d = 14.67$ for petrol. The indicated engine torque τ_{ind} from combustion can be determined as

$$\tau_{ind}(N, p_m, \dot{m}_f) = H_u \frac{\eta_{th}(N, p_m, \theta_{SA}, \lambda)\dot{m}_f}{N} \tag{7}$$

where H_u is the fuel energy constant and η_{th} is a nonlinear function of the engine speed N, the manifold pressure p_m, the spark advance angle θ_{SA}, and the AFR λ (Hendricks and Luther (2001)). For the engine speed control problem, the engine operates within a certain region of conditions and the spark angle θ_{SA}, as well as λ, remains constant.

2.5 Eccentric Shaft Model

The rotating dynamics of the eccentric shaft can be expressed using Newton's second law as

$$J\dot{N} = \tau_{ind} - \tau_{fric} - \tau_{load} \tag{8}$$

where J is the scaled engine moment of inertia, τ_{fric} and τ_{load} refer to the friction and the load torque, respectively.

3. STATE FEEDBACK VIA Q-LEARNING

In this section, we propose an adaptive optimal state feedback controller for the engine speed regulation problem. The adaptive optimal control algorithm is derived based on the authors' previous work (Chen and Herrmann (2019)) of Q-learning for unknown continuous-time nonlinear systems. The controller can solve the nonlinear optimal control problems online and is model-free.

3.1 Optimal Control Problem Formulation

The state variables can be chosen from the MVEM as

$$\mathcal{X}(t) = \begin{bmatrix} p_m(t) \\ T_m(t) \\ m_{fp}(t) \\ m_{fv}(t) \\ N(t) \end{bmatrix} \begin{matrix} \text{intake manifold pressure} \\ \text{intake manifold temperature} \\ \text{fuel puddle mass} \\ \text{fuel vapour mass} \\ \text{engine speed} \end{matrix} \tag{9}$$

For the engine speed control problem, the objective is to regulate the engine speed $N(t)$ around a certain low set point, namely, $N^0 = 3000$ RPM for the Wankel rotary engine. We can shift the coordinate of the equilibrium point to zero by translating the engine speed as $\mathcal{N}(t) = N(t) - N^0$. Similarly, by translating the other states to a nominal operating point, a new state vector x is defined such that

$$x(t) = \mathcal{X}(t) - \mathcal{X}^0 \tag{10}$$

where $\mathcal{X}^0$ is the nominal operating point around which the engine operates as in Table 1. By inspection of (1) and the MVEM, the throttle angle α is an affine input of the whole system. We can represent the MVEM into a continuous-time nonlinear time-invariant system in state space as

$$\dot{x}(t) = f(x(t)) + g(x(t))\alpha(t), \quad x(0) = x_0 \tag{11}$$

where the state vector $x(t)$ is defined as (9), the throttle angle $\alpha(t)$ is the control policy for the electronic throttle, and $f(x(t))$, $g(x(t))$ are the system drift and the input gain functions, respectively. It is reasonable to assume that the engine dynamics $f(x) + g(x)\alpha$ is Lipschitz continuous on a compact set $\Omega \in \mathbb{R}^5$ that contains the origin.

Table 1. Nominal engine operating point

Engine state	Symbol	Value	Units
Intake manifold pressure	p_m^0	0.8	bar
Intake manifold temperature	T_m^0	25	^{o}C
Fuel puddle mass	m_{fp}^0	0.11	g
Fuel vapour mass	m_{fv}^0	0.25	g
Engine speed	N^0	3000	RPM

We define the infinite-horizon integral cost

$$V(x(t)) := \int_t^\infty r(x(\tau), \alpha(\tau))\mathrm{d}\tau \tag{12}$$

with the utility $r(t) = S(x(t)) + R\alpha^2(t)$. The utility $r(t)$ is positive definite, i.e., $S(x(t)) > 0$ and $R > 0$. One can simply choose a quadratic utility term $S(x(t)) = x^\mathsf{T} S^0 x$ with a positive definite matrix $S^0 > 0$ for the engine speed regulation problem.

The optimal control problem is to minimise the value function (12) by choosing the optimal stabilising (admissible) control policy $\alpha^*(t)$. The optimal value function $V^*(x)$ can be determined as

$$V^*(x(t)) := \min_\alpha \int_t^\infty r(x(\tau), \alpha(\tau))\mathrm{d}\tau \tag{13}$$

A general solution to the nonlinear optimal control problem can be formulated as a partial differential equation for the optimal value function $V^*(x)$. We define the Hamiltonian of the problem as

$$\mathcal{H}(x, \alpha, \nabla V_x) := r(x, \alpha) + (\nabla V_x)^\mathsf{T}(f(x) + g(x)\alpha) \tag{14}$$

with the gradient vector $\nabla V_x = \partial V/\partial x$. The optimal value function $V^*(x)$ in (13) satisfies the *Hamilton-Jacobi-Bellman* (HJB) equation

$$0 = \min_\alpha \ \mathcal{H}(x, \alpha, \nabla V_x^*) \tag{15}$$

It is noted that the throttle angle α should be in the range of $[0, 90^o]$. However, the engine speed control problem, e.g. idle speed control, usually provides a local span of the operation condition near the low speed where the throttle angle is small, e.g., $\alpha \approx 30^o$ when all the states are around the nominal operating point in Table 1. The control action α is far below the upper limit and locally unconstrained. Therefore, the optimal control α^* can be found by setting $\partial\mathcal{H}(x, u, \nabla V_x^*)/\partial\alpha = 0$ so that

$$\alpha^* = -\frac{1}{2}R^{-1}g(x)^\mathsf{T}\nabla V_x^* \tag{16}$$

Inserting the optimal control (16) into (15) gives the HJB equation in terms of ∇V_x^* as

$$0 = S(x) + (\nabla V_x^*)^\mathsf{T} f(x) - \frac{1}{4}(\nabla V_x^*)^\mathsf{T} g(x) R^{-1} g(x)^\mathsf{T} \nabla V_x^* \tag{17}$$

In general, the HJB equation (15) is difficult to solve due to its nonlinearity and the requisite for *a priori* knowing the system drift dynamics $f(x)$ and input gain dynamics $g(x)$.

3.2 Parameterisation of Nonlinear Q-Function

The idea of continuous-time Q-learning is to create an action-dependent version of value function $Q(x, \alpha)$: such that $Q^*(x, \alpha^*) = V^*(x)$. For the continuous-time nonlinear affine system (11), the Q-function can be explicitly defined by adding the Hamiltonian (14) onto the optimal value (13) as

$$\begin{aligned} Q(x, \alpha) :&= V^*(x) + \mathcal{H}(x, \alpha, \nabla V_x^*) \\ &= \underbrace{V^*(x) + S(x) + (\nabla V_x^*)^\mathsf{T} f(x)}_{F_{xx}(x)} + \\ &\quad \underbrace{(\nabla V_x^*)^\mathsf{T} g(x)\alpha}_{F_{x\alpha}(x,\alpha)} + \underbrace{R\alpha^2}_{F_{\alpha\alpha}(\alpha)} \end{aligned} \tag{18}$$

where $F_{xx}(x)$, $F_{x\alpha}(x, \alpha)$, and $F_{\alpha\alpha}(\alpha)$ are the lumped terms that can be approximated respectively via neural networks.

Lemma 1. The Q-function defined in (18) is positive definite with the optimisation $Q^*(x, \alpha^*) = \min_\alpha Q(x, \alpha)$. The optimal Q-function $Q^*(x, \alpha^*)$ has the same optimal value $V^*(x)$ (13) as for the value function $V^\alpha(x)$ (12), i.e. $Q^*(x, \alpha^*) = V^*(x)$ when applying the optimal control α^*.

Proof. The proof follows from *Lemma 3* of Chen and Herrmann (2019). ■

3.3 Adaptive Critic Design and Q-Learning

We approximate the Q-function (18) using a critic neural network by

$$Q(x, \alpha) = W^\mathsf{T}\Phi(x, \alpha) + \varepsilon_Q(x, \alpha) \tag{19}$$

where $\Phi(x, \alpha) \in \mathbb{R}^n$ denotes the activation function vector with the number n of neurons in the hidden layer; $W \in \mathbb{R}^n$ is the weight vector; $\varepsilon_Q(x, \alpha)$ is the neural network approximation error; and $W^\mathsf{T}\Phi(x, \alpha)$ can be explicitly expressed according to the three components $F_{xx}(x)$, $F_{x\alpha}(x, \alpha)$, and $F_{\alpha\alpha}(\alpha)$ in (18) as

$$W^\mathsf{T}\Phi(x, \alpha) = \begin{bmatrix} W_{xx}^\mathsf{T} & W_{x\alpha}^\mathsf{T} & W_{\alpha\alpha}^\mathsf{T} \end{bmatrix} \begin{bmatrix} \Phi_{xx}(x) \\ \Phi_{x\alpha}(x)\alpha \\ \Phi_{\alpha\alpha}(\alpha) \end{bmatrix} \tag{20}$$

where $\Phi_{xx} \in \mathbb{R}^{n_{xx}}$, $\Phi_{x\alpha} \in \mathbb{R}^{n_{x\alpha}}$ and $\Phi_{\alpha\alpha} = \alpha^2$. The regressor $\Phi(x, \alpha)$ is selected to provide a complete independent basis such that $Q(x, \alpha)$ is uniformly bounded with $n = n_{xx} + n_{x\alpha} + 1$). Recall the Weierstrass higher-order approximation theorem (Abu-Khalaf and Lewis (2005)), the approximation error $\varepsilon_Q(x, \alpha)$ is bounded for a fixed n within the compact set Ω and as the number of neurons $n_{xx} \to \infty$ and $n_{x\alpha} \to \infty$, i.e., $n \to \infty$, we have $\varepsilon_Q(x, \alpha) \to 0$.

One needs to derive the Bellman equation in terms of the Q-function to update the critic. By Bellman's principle of optimality (Lewis et al. (2012)), we have the following optimality equation

$$V^*(x(t-T)) = \int_{t-T}^{t} r(x(\tau), \alpha^*(\tau))\mathrm{d}\tau + V^*(x(t)) \quad (21)$$

The result from *Lemma 1* showed that $Q^*(x, \alpha^*) = V^*(x)$, which means we can employ *Lemma 10.2-1* of Lewis et al. (2012) and rewrite (21) in terms of $Q^*(x, \alpha^*)$ as

$$\begin{aligned} &\overbrace{-\int_{t-T}^{t} r(x,\alpha)\mathrm{d}\tau}^{-\rho(x,\,\alpha)} = Q(x(t), \alpha(t)) \\ &\quad - Q(x(t-T), \alpha(t-T)) + \rho_r \\ &= \underbrace{W^\mathsf{T}\Phi(x(t), \alpha(t)) - W^\mathsf{T}\Phi(x(t-T), \alpha(t-T))}_{W^\mathsf{T}\Delta\Phi(x,\,\alpha)} \\ &\quad + \varepsilon_{BQ} + \rho_r \end{aligned} \quad (22)$$

with the integral reinforcement $\rho(x, \alpha)$, the difference $\Delta\Phi(t) = \Phi(x(t), \alpha(t)) - \Phi(x(t-T), \alpha(t-T))$, the Bellman equation residual errors $\varepsilon_{BQ} = \varepsilon_Q(x(t), \alpha(t)) - \varepsilon_Q(x(t-T), \alpha(t-T))$ being bounded for bounded ε_Q, and $\rho_r = \rho_{r1} + \rho_{r2} + \rho_{r3}$ is the Bellman residual error with

$$\rho_{r1} = -\int_{t-T}^{t} R(\alpha - \alpha^*)^2 d\tau \quad (23)$$

$$\rho_{r2} = -R(\alpha(t) - \alpha^*(t))^2 \quad (24)$$

$$\rho_{r3} = R(\alpha(t-T) - \alpha^*(t-T))^2 \quad (25)$$

Define two auxiliary variables $\mathcal{P} \in \mathbb{R}^{n\times n}$ and $\mathcal{Q} \in \mathbb{R}^n$ by low-pass filtering the variables in (22) as

$$\begin{cases} \dot{\mathcal{P}} = -\ell\mathcal{P} + \Delta\Phi(t)\Delta\Phi(t)^\mathsf{T}, & \mathcal{P}(0) = 0 \\ \dot{\mathcal{Q}} = -\ell\mathcal{Q} + \Delta\Phi(t)\rho(x,\alpha), & \mathcal{Q}(0) = 0 \end{cases} \quad (26)$$

with a filter parameter $\ell > 0$.

The adaptive critic neural network can be written as

$$\hat{Q}(x, \alpha) = \hat{W}^\mathsf{T}\Phi(x, \alpha) \quad (27)$$

where $\hat{W}$ and $\hat{Q}(x, \alpha)$ denote the current estimate of W and $Q(x, \alpha)$, respectively.

Now we design the adaptation law using the sliding mode technique to update $\hat{W}$ such that

$$\dot{\hat{W}} = -\Gamma\mathcal{P}\frac{M}{||M||} \quad (28)$$

where $M \in \mathbb{R}^n$ is defined as $M = \mathcal{P}\hat{W} + \mathcal{Q}$ and $\Gamma \succ 0$ is a diagonal adaptive learning gain to be tuned.

We reconstruct the optimal control α^* from (16) based on the parameterisation of $Q(x, \alpha)$ (18) such that

$$\alpha^* = -\frac{1}{2}W_{\alpha\alpha}^{-1}W_{x\alpha}^\mathsf{T}\Phi_{x\alpha}(x) + \varepsilon_\alpha \quad (29)$$

where ε_α is a bounded approximation error due to ε_Q, $W_{x\alpha}^\mathsf{T}\Phi_{x\alpha}(x)$ accounts for the term $g(x)^\mathsf{T}\nabla V_x^*$, and $W_{\alpha\alpha}$ is essentially predefined R (see (18)). Therefore, one can determine the optimal control directly using the adaptive critic (27) if the weight $\hat{W}$ converges to the actual weight W. The control law (actor) will be

$$\alpha = -\frac{1}{2}R^{-1}\hat{W}_{x\alpha}^\mathsf{T}\Phi_{x\alpha}(x) \quad (30)$$

At this stage, we have provided all the elements of the adaptive optimal state feedback controller: an adaptive critic (27), an adaptation law (28), and a control law (30). We summarise the result for this Q-learning-based state feedback control in the following remark.

Remark. Given the engine system (11) with the value function (12) and Q-function (18), the adaptive critic neural network (27) with the adaptation law (28) and the actor (30) form an adaptive optimal control so that the adaptive critic weight estimation error $\tilde{W}$ will converge to a compact set and the throttle angle (the actor) α will converge to a small bounded set around its optimal control solution α^* in finite time. The proof of the finite-time convergence of the weight estimation error $\tilde{W}$ and the control α follows from *Lemma 2* of Chen and Herrmann (2019). Its detailed proof will be provided in *Theorem 4* in Section 4.2. The proposed engine speed controller is a *model-free* algorithm that can approximately solve the optimal control problem *online* without the *a priori* knowledge of the system drift $f(x)$ and input gain $g(x)$.

It is important to note that the control algorithm is data-driven and requires complete knowledge of all the state variables $x(t)$.

4. OUTPUT FEEDBACK SPEED CONTROL

This section describes the design and analysis of the output feedback control for the engine speed regulation problem, which synthesises a reduced-order optimal observer with the previous Q-learning-based state feedback control.

4.1 Optimal State Observer

It is uncommon to directly use state feedback for a realistic system since not all states are measurable in practice. By inspection of the engine states (9), the intake manifold pressure p_m, temperature T_m, and the engine speed N are commonly measurable through the manifold absolute pressure (MAP) sensor, the intake air temperature (IAT) sensor, and the tachometer, respectively. The other two states: the fuel puddle mass m_{fp} and the fuel vapour mass m_{fv} are not directly measurable in practice. If the system is observable, one can design a reduced-order observer to online estimate only the unknown states. In our work (Chen et al. (2018)), an extended Kalman filter (EKF) is proposed to estimate the fuel puddle model using the mass air flow (MAF) sensor and the Lambda sensor. This is integrated as part of a nonlinear observer-based AFR controller, where the EKF estimates the fuel puddle fraction for the control law. Since only the fuel puddle variables m_{fp} and m_{fv} need to be estimated, we can re-use the EKF as a reduced-order optimal observer for our Q-learning-based state feedback engine speed controller. We briefly present the result as follows (see Chen et al. (2018) for more details).

For the nonlinear fuel puddle model (5)(6) with the unknown parameters τ_f and χ, the parameters can be taken as extra states and the term $\dot{m}_{fpe}$ is shown to be negligible (Arsie et al. (2003). We write the fuel puddle process (5)(6) as a set of stochastic first-order equations

$$\begin{cases} \dot{\tau_f} = w_1 \\ \dot{\chi} = w_2 \\ \dot{m}_{fp} = \chi\dot{m}_{fi} - (1/\tau_f)m_{fp} + w_3 \\ \dot{m}_{fv} = (1-\chi)\dot{m}_{fi} + (1/\tau_f)m_{fp} - m_{fv}/\tau_m + w_4 \\ \dot{m}_f = m_{fv}/\tau_m + v \end{cases} \quad (31)$$

or in the form of state equations as

$$\begin{cases} \dot{y} = \xi(y, \dot{m}_{fi}) + w \\ \dot{m}_f = h(y) + v \end{cases} \tag{32}$$

where $y = [\tau_f \ \chi \ m_{fp} \ m_{fv}]^T$ is the state vector, $f(y, \dot{m}_{fi})$ and $h(y)$ denote the (non-)linear functions in (31), $w = [w_1 \ w_2 \ w_3 \ w_4]^T$ and v are the zero mean multivariate Gaussian noises that account for the model inaccuracy and measurement noise with pre-defined covariance $\mathcal{W}$ and $\mathcal{Y}$. It is assumed that the noises are bounded, i.e., $||w|| \leq \varpi$ and $|v| \leq \mu$ with the constants $\varpi > 0$ and $\mu > 0$. In practice, the measurement of $\dot{m}_f$ can be obtained by dividing the reading of the MAF sensor by the reading of Lambda sensor since $\dot{m}_f = \dot{m}_a/\lambda$.

For the stochastic system (32), an extended Kalman filter can be designed accordingly with the Kalman gain vector L as

$$\dot{\hat{y}} = \xi(\hat{y}, \dot{m}_{fi}) + L(\dot{m}_f - h(\hat{y})) \tag{33}$$

where $\hat{y}$ is the estimate of the state vector y the optimal Kalman gain is chosen as $L = PH^{\mathsf{T}}\mathcal{Y}^{-1}$; Ξ and H are the Jacobian matrix of $\xi(y, \dot{m}_{fi})$ and $h(y)$ with respect to y as

$$\Xi = \frac{\partial \xi}{\partial y} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ m_{fp}/\tau_f^2 & \dot{m}_{fi} & -1/\tau_f & 0 \\ -m_{fp}/\tau_f^2 & -\dot{m}_{fi} & 1/\tau_f & -1/\tau_m \end{bmatrix} \tag{34}$$

$$H = \frac{\partial h}{\partial y} = [0 \ 0 \ 0 \ 1/\tau_m] \tag{35}$$

P is the covariance prediction matrix which can be solved via the algebraic Riccati equation

$$\dot{P} = \Xi P + P\Xi^{\mathsf{T}} - PH^{\mathsf{T}}\mathcal{Y}^{-1}HP + \mathcal{W} \tag{36}$$

Lemma 2. For the system (32) with the EKF (33), the estimation error $\tilde{y} = y - \hat{y}$ will exponentially converge towards a compact set around zero.

Proof. See *Proposition 1* in Chen et al. (2018) for the detailed proof. ■

The approximation error of the higher order terms due to the linearisation is denoted as $o(||\tilde{y}||)$, i.e.,

$$\dot{y} = \Xi y + o(||\tilde{y}||), \quad ||o(||\tilde{y}||)|| < \delta \tag{37}$$

which has an upper bound $\delta > 0$. It is clear from *Theorem 4* that the EKF, as a reduced-order optimal observer, can online estimate the two unknown states m_{fp} and m_{fv}.

4.2 Output Feedback Synthesis

The proposed Q-learning-based state feedback engine speed controller requires the complete knowledge of the system states. A reduced-order optimal state observer is designed to estimate the unknown states. The synthesis of the two naturally leads to an observer-based dynamic output feedback Q-learning engine speed controller. We summarise the main result of this paper in the following theorem.

Theorem 3. Given the engine system (11) with the stochastic fuel puddle process (32) and the prescribed value function (12) and Q-function (18), the adaptive critic neural network (27) with the adaptation law (28) and the actor (30) and the EKF (33) form a dynamic output feedback control so that the throttle angle α will converge to a small bounded set near its optimal control solution α^* in finite time, i.e., the engine speed will be regulated to the set point subject to the prescribed value function (12).

Proof. The proof follows from the Lyapunov theorem and is omitted due to limited space. ■

5. SIMULATIONS

An MVEM of a Wankel rotary engine is created in Matlab/Simulink, where the model parameters are calibrated based on the experimental data sets (Chen et al. (2018), Chen et al. (2020)). The engine is a 225CS rotary engine produced by Advanced Innovative Engineering (AIE) UK Ltd, which is being tested as a range extender for a hybrid automotive powertrain due to its high specific power output. For the engine speed control problem, we choose the value function as (12) with $S^0 = diag[1 \ \ 1 \ \ 1 \ \ 1 \ \ 4]$ and $R = 1$. The activation function $\Phi(x, \alpha)$ of the adaptive critic neural network (20) is selected as $\Phi(x, \alpha) = [p_m^2 \ \ p_m T_m \ \ T_m^2 \ \ m_{fp} \ \ m_{fv} \ \ p_m N \ \ N^2 \ \ p_m\alpha \ \ T_m\alpha \ \ m_{fp}\alpha \ \ m_{fv}\alpha \ \ N\alpha \ \ \alpha^2]^{\mathsf{T}}$ with the number of neurons $n = 13$. We initialise the state $x(0) = [1 \ 1 \ 1 \ 1 \ 1]^{\mathsf{T}}$ and the weight $\hat{W}(0) = [0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 0.5 \ \ 1]^{\mathsf{T}}$. The tuning parameters are chosen as such: the sample period $T = 2$s, the filter parameter $\ell = 1$, and the adaptive learning gain $\Gamma = 7$. We first inject the exploration noise onto the throttle angle for the learning period to ensure the persistent excitation of the signals. The engine load is set to be constant when learning. Fig. 1 presents the engine trajectories with the exploration noise for a period of 30 s. The state variables are normalised into the value range of [0,1].

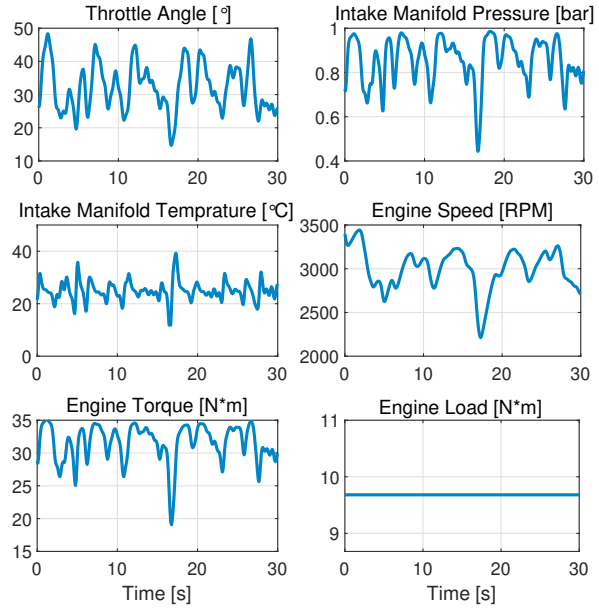

Fig. 1. Engine trajectories with the exploration noise

In order to validate the performance of the resulting controller after learning, we simulate a load disturbance (caused for instance by power steering or transmission engagement) at 900s and 1000s. The results are presented in Fig. 2, where the engine speed response under the resulting controller is plotted against that when there is no control action. The controller can effectively reject the disturbance in either case of an increase or decrease in the engine load. However, we have not explored rigorously the question of robustness to wide operating point variations; a simulation around 3000 RPM was promising but not conclusive. In any case, we would anticipate that a practical implementation would provide adaptation without requiring extra sensors.

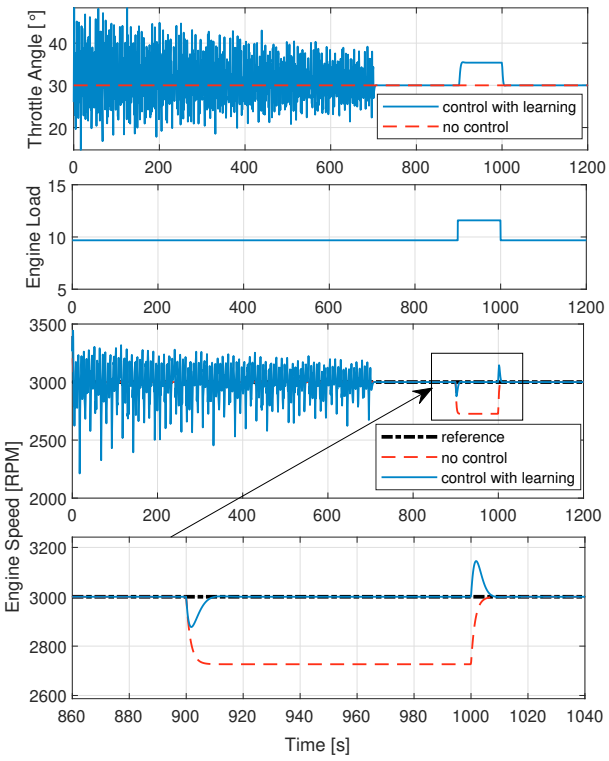

Fig. 2. Simulation results of the learning-based controller

6. CONCLUSIONS

We proposed a dynamic output feedback controller for the engine speed regulation problem using reinforcement learning principles, namely, Q-learning. The Q-learning-based controller is data-driven in continuous time. Via an EKF, a dynamic output feedback control is achieved using only the measurable outputs (intake manifold pressure and temperature, engine speed), i.e. no extra sensor is needed. The simulation on a Wankel engine model shows the proposed controller, after learning, can effectively reject load disturbance and regulate the engine speed around a desired point. Future work will focus on the practical validation of the proposed controller with engine tests.

ACKNOWLEDGEMENTS

This work is funded jointly by the University of Bristol and China Scholarship Council (CSC) and the Engineering and Physical Sciences Research Council (EPSRC) as part of the RAIN+ Research Hub (EP/W001128/1). The authors would like to thank Dr Giovanni Vorraro, Mr Matthew Turner, Dr Reza Islam, and Prof Jamie Turner at the Institute for Advanced Automotive Propulsion Systems (IAAPS), the University of Bath, for their continued advice and support on engine modelling and control.

REFERENCES

Abu-Khalaf, M. and Lewis, F.L. (2005). Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network hjb approach. *Automatica*, 41(5), 779–791.

Arsie, I., Pianese, C., Rizzo, G., and Cioffi, V. (2003). An adaptive estimator of fuel film dynamics in the intake port of a spark ignition engine. *Control Engineering Practice*, 11(3), 303–309.

Chen, A.S. and Herrmann, G. (2019). Adaptive optimal control via continuous-time Q-learning for unknown nonlinear affine systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 1007–1012. IEEE.

Chen, A.S., Herrmann, G., Na, J., Turner, M., Vorraro, G., and Brace, C. (2018). Nonlinear observer-based air-fuel ratio control for port fuel injected Wankel engines. In *2018 UKACC 12th International Conference on Control (CONTROL)*, 224–229. IEEE.

Chen, A.S., Vorraro, G., Turner, M., Islam, R., Herrmann, G., Burgess, S., Brace, C., Turner, J., and Bailey, N. (2020). Control-oriented modelling of a Wankel rotary engine: A synthesis approach of state space and neural networks. SAE Technical Paper, No. 2020-01-0253.

Cook, J.A. and Powell, B.K. (1988). Modeling of an internal combustion engine for control analysis. *IEEE Control Systems Magazine*, 8(4), 20–26.

Di Cairano, S., Yanakiev, D., Bemporad, A., Kolmanovsky, I.V., and Hrovat, D. (2011). Model predictive idle speed control: Design, analysis, and experimental evaluation. *IEEE Transactions on Control Systems Technology*, 20(1), 84–97.

Gao, W., Jiang, Y., Jiang, Z.P., and Chai, T. (2016). Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. *Automatica*, 72, 37–45.

Hendricks, E. (2001). Isothermal vs. adiabatic mean value si engine models. *IFAC Proceedings Volumes*, 34(1), 363–368.

Hendricks, E. and Luther, J.B. (2001). Model and observer based control of internal combustion engines. In *Proceedings of International Workshop on Modeling, Emissions and Control in Automotive Engines*. Citeseer.

Lewis, F.L. and Vamvoudakis, K.G. (2010). Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(1), 14–25.

Lewis, F.L., Vrabie, D., and Syrmos, V.L. (2012). *Optimal control*. John Wiley & Sons.

Li, X. and Yurkovich, S. (2001). Sliding mode control of delayed systems with application to engine idle speed control. *IEEE Transactions on Control Systems Technology*, 9(6), 802–810.

Sun, X.D., Scotson, P.G., and Balfour, G. (2002). A further application of loop shaping h-infinity control to diesel engine control-driven-idle speed control. Technical report, SAE Technical Paper.

Sutton, R.S., Barto, A.G., and Williams, R.J. (1992). Reinforcement learning is direct adaptive optimal control. *IEEE Control Systems*, 12(2), 19–22.

Thornhill, M., Thompson, S., and Sindano, H. (2000). A comparison of idle speed control schemes. *Control Engineering Practice*, 8(5), 519–530.

Vrabie, D., Pastravanu, O., Abu-Khalaf, M., and Lewis, F.L. (2009). Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2), 477–484.

Vrabie, D., Vamvoudakis, K.G., and Lewis, F.L. (2013). *Optimal adaptive control and differential games by reinforcement learning principles*, volume 2. IET.

Werbos, P. (1992). Approximate dynamic programming for realtime control and neural modelling. *Handbook of intelligent control: neural, fuzzy and adaptive approaches*, 493–525.

Yildiz, Y., Annaswamy, A., Yanakiev, D., and Kolmanovsky, I. (2007). Adaptive idle speed control for internal combustion engines. In *2007 American Control Conference*, 3700–3705. IEEE.