Stochastic Dynamic Programming Control Policies for Fuel Efficient Vehicle Following

Kevin McDonough, Ilya Kolmanovsky, Dimitar Filev, Diana Yanakiev, Steve Szwabowski, and John Michelini

Abstract—Stochastic control policies are developed for fuel-efficient following of a lead vehicle moving at a constant speed. Assuming that a lead vehicle is travelling at a specified speed and that the road grade is modeled stochastically, we solve numerically a Stochastic Dynamic Programming (SDP) problem to determine follower vehicle speed control policies subject to constraints on vehicle following distance. The cost function to which SDP is applied reflects fuel consumption, travel time and includes a penalty for distance constraint violation. A virtual testing environment based on CarSim is used for simulations and fuel economy assessment. We demonstrate the potential for fuel consumption improvements (over 5%) and examine properties of the speed control policy of the follower vehicle, including the emergent ‘pulse and glide’ behavior. The policies do not require preview information of the route and of the road grade thereby providing a potentially simpler alternative to preview-based systems. The results complement our previous results on fuel efficient in traffic driving.

I. INTRODUCTION

With increasing gasoline prices, fuel efficient driving has been receiving increasing attention. Fuel economy maximizing techniques are actively discussed in the hypermiling/eco-driving community [11]. Fuel economy gains can be realized, to a varying extent, by driving at “efficient speeds,” “accelerating quickly, but smoothly,” “coasting” and “burn and coast” (also known as “pulse and glide”). Even though there is much on-going discussion, the best driving strategies in real traffic on real roads are insufficiently understood.

If driving on a constant grade, the fuel economy dependence on the vehicle speed shows the trends apparent in Figure 1. The discontinuous changes are due to gear shifts and the optimal fuel economy is achieved between 30 and 50 mph depending on the grade.

Periodic optimal control [2] and saw-tooth policies [10] were examined through the application of optimal control techniques in the literature. The main conclusion, under ideal assumptions, appears that fuel economy can be improved versus steady speed driving by varying vehicle speed periodically. This conclusion appears to be counterintuitive as the general belief is that driving smoothly is better for fuel economy. Periodic control policies for battery power modulation have also been examined for hybrid electric vehicle applications [8].

The fuel efficient driving strategies when the road grade is time-varying (as is the case on real roads) or when driving in-traffic are even more complicated. If following a vehicle with known vehicle speed versus time on a grade known in advance, the vehicle speed trajectory can be optimized. For an aggressive driving cycle, such as US06, by allowing the follower vehicle speed to deviate in +/-1.5 mph (+/- 1 sec) band of the lead vehicle, subject to 2% deviation on driving distance, over 10% fuel economy improvement can be demonstrated [6]. Driving optimally within 5 m distance band of the lead vehicle during last 100 sec of US06 cycle translates into 36% fuel economy improvement [6]. The notion of a distance hybrid vehicle has been introduced in [6] to refer to exploiting this distance storage mechanism for fuel economy improvement. Related recent work uses deterministic dynamic programming and model predictive control (see e.g., [3], [4], [5], [9] and references therein) to improve fuel economy.

![Figure 1: Fuel economy of constant speed driving versus vehicle speed at different grades.](image-url)

When the driving cycle (road grade, vehicle speed, wind speed and direction, etc.) cannot be predicted in advance, the fuel efficient policies can be obtained by modeling the driving cycle statistical properties in a given geographical region, and using the stochastic dynamic programming techniques [6], [10], [11]. The resulting policy for adjusting the vehicle speed relative to the traffic speed is time-independent and achieves on average best performance in terms of fuel economy and travel time. Stochastic model
predictive control [1] is another approach that has been proposed for these applications.

In this paper, we address a lead vehicle following scenario where the lead vehicle is driving at a constant speed. The grade may vary in-time, and we develop Transition Probability Model (TPM) to represent the grade statistics. Through the use of the penalty term in the cost function we constrain the distance variation to stay bounded in a specified range, and we use the stochastic dynamic programming to generate control policies for the follower vehicle speed which minimize the expected fuel consumption. The resulting policies are time-invariant and specify the offset in the follower vehicle speed relative to the lead vehicle speed based on the relative distance and the follower vehicle speed. Our objective is to determine if these time-invariant policies improve fuel economy and if they may result in the periodic oscillations of the follower vehicle speed that one may expect from the periodic optimal control considerations [2]. We use a high fidelity simulation modeling setup in CarSim1 and Simulink2 to evaluate the control policies. The difference with our previous work [10] is that in this paper we in focus on the constant lead vehicle speed case and we incorporate the treatment of relative distance constraints in the optimization.

The paper is organized as follows. Section II describes the identification of transition probabilities from road grade data. The fuel consumption model based on neural networks is discussed in Section III. In Section IV, we discuss the optimization problem formulation and the solution method. In Section V we illustrate the control policies. We describe the simulation results in Section VI. Finally, concluding remarks are made in Section VII.

II. CONSTRUCTION OF TRANSITION PROBABILITY MATRICES

For the road grade, the TPMs were constructed using measured road grade along a 17-mile stretch of M-39 near Dearborn, Michigan. The spatial interval size was selected to be 15 m and the values of the road grade were quantized as \([-6, -5, ..., 0, ..., 5, 6]\)% . The spatial road grade data from the M-39 route were then used to generate TPMs.

Transition frequencies between the states corresponding to the intervals around quantized grade values within the spatially mapped experimental data were counted so that TPMs were defined based on

\[
T(\theta^+|\theta) = \frac{N_{\theta,\theta^+}}{M_\theta}
\]

where \(T(\theta^+|\theta)\) is the transition probability between states \(\theta\) and \(\theta^+\), \(N_{\theta,\theta^+}\) is the number of transitions between states \(\theta\) and \(\theta^+\), and \(M_\theta\) is the total number of transitions initiated in state \(\theta\).

Figure 2 shows an example of the developed TPM. Note the closeness of non-zero probabilities to the main diagonal implying that the transitions tend to happen between discrete states that are relatively close to each other.

![Figure 2: Road grade transition probabilities from current grade, \(\theta\), to the next grade, \(\theta^+\).](image)

III. FUEL CONSUMPTION MODELING

As in [10], we model the fuel consumption as a function of current and next states, i.e.,

\[
W_f = f(v, v^+, \theta, \theta^+),
\]

where \(W_f\) is the fuel consumption over a 15 m segment, \(v\) is the current vehicle speed, \(v^+\) is the next vehicle speed, \(\theta\) is the current grade, and \(\theta^+\) is the next grade. Such a model for fuel consumption can be developed based on inferring wheel power and using engine map and transmission/driveline efficiencies to determine fuel flow. In practice, the development of a first principle based model is complicated by the need to account for gear shifts and friction brake activation. Consequently, an empirical, data-driven model is employed as (2). This model has a form of a 2-layer, 5-input, single output multilayer perceptron (MLP) feed forward neural network of the form,

\[
W_f = \sigma_2(w_2\sigma_1(w_1u + b_1) + b_2),
\]

where \(\sigma_2\) and \(\sigma_1\) are linear and hyperbolic activation functions, respectively, \(w_2\) and \(w_1\), \(b_2\) and \(b_1\) are the corresponding vectors of weights and biases of the linear and hyperbolic activation functions, and \(u = (v, v^+, v^+ - v, \theta, \theta^+)\) is the vector of model inputs. The Bayesian regularization back propagation training function \texttt{trainbr} available from the Neural Network tool box in MATLAB was applied for training the neural network model. The fuel flow data used for training were based on CarSim simulated drives. The fuel consumption model is quite accurate, see Figure 3. This fuel consumption model is only used to generate the control policy, for the actual evaluation and comparison of the fuel consumption we employ CarSim-based model.

---

1 CarSim is a registered trademark of Mechanical Simulation Corporation.

2 Simulink is a registered trademark of The Mathworks, Inc.
The Stochastic Dynamic Programming (SDP) is used for the control policy generation. The states are the road grade, \( \theta \), the offset of the speed of the follower vehicle relative to the speed of the lead vehicle, \( u \), and the distance, \( \rho \) (over a nominal safety margin) between the lead vehicle and the follower vehicle. Note that the follower vehicle velocity is given by

\[
v = v_l + u,
\]

where \( v_l \) is the speed of the lead vehicle. The relative distance satisfies the following relation,

\[
\rho^* = \rho + u \frac{\Delta s}{v_l}
\]

where \( \rho^* \) is the relative distance at the end of the distance segment \( \Delta s \).

The fuel consumption model is averaged with respect to the (next one step ahead) value of road grade:

\[
\bar{W}_f(\theta, u, u^*) = E_{\theta + f}(v_l + u, v_l + u^*, \theta, \theta^*),
\]

where \( u^* \) (next follower vehicle speed offset relative to the given constant lead vehicle speed, \( v_l \)) is the decision (control) variable and \( f \) is defined in (2). The approximation (6) is justified given that the transition probability matrix is close to being diagonal.

The incremental cost function is of the form

\[
R(\rho, \theta, u, u^*, \lambda) = \bar{W}_f(\theta, u, u^*) + \lambda T_r(u, u^*) + \psi(\rho),
\]

where \( \bar{W}_f \) is the expected fuel cost over a 15 m segment described by (6), \( \lambda \) is a trade-off parameter between fuel consumption and travel time, \( T_r \) is the expected segment travel time estimated to be

\[
T_r(u, u^*) = \frac{2v_l + u + u^*}{2\Delta s},
\]

and the function \( \psi(\rho) \) is a penalty for the distance constraint violation defined below.

For the incremental cost function (7) we formulate and solve the SDP problem numerically by using the value iteration approach to find a value function that approximately satisfies the Bellman equation

\[
V(\rho, \theta, u, \lambda) = \min_{u^*} Q(\rho, \theta, u, u^*, \lambda),
\]

where

\[
Q(\rho, \theta, u, u^*, \lambda) = R(\rho, \theta, u, \lambda) + \sum_{\theta^*} qV(\rho^*, \theta^*, u^*, \lambda)|P(\theta^* | \theta),
\]

\( q, 0 < q < 1 \), is a discount factor and \( P(\theta^* | \theta) \) is the grade transition probability. The optimal control policy is a minimizer of (10), that is

\[
U^*(\rho, \theta, u, \lambda) \in \arg\min_{u^*} Q(\rho, \theta, u, u^*, \lambda). \tag{11}
\]

The value function iterations applied to solve (9)-(10) numerically have the following form,

\[
V_{n+1}(\rho, \theta, u, \lambda) = \min_{u^*} Q_n(\rho, \theta, u, u^*, \lambda),
\quad Q_n(\rho, \theta, u, u^*, \lambda) = R + \sum_{\theta^*} qV(\rho^*, \theta^*, u^*, \lambda)|P(\theta^* | \theta),
\quad V_0(\rho, \theta, u, \lambda) = 0, \tag{12}
\]

with \( O(\theta, u) \) denoting the set of feasible actions, i.e., the next speed offsets.

Under standard assumptions, the iterations (12) converge. Computationally, the iterations are performed until a termination criterion

\[
\|V_{n+1}(\cdot) - V_n(\cdot)\| \leq CT, \tag{13}
\]

is satisfied where \( CT \) is a defined convergence threshold.

While beyond the scope of this conference paper, the driver comfort and traffic flow disruption constraints may be incorporated either by additional terms in the cost function or by restricting \( O(\theta, u) \) to a subset of appropriate control actions.

\[\text{V. CONTROL POLICIES}\]

For the control policy generation, we restrict the search to

\[
u^* \in O(\theta, u) = \{-3, -2, -1, 0, 1, 2, 3\} \left\lfloor \frac{m}{\text{grid}} \right\rfloor,
\]

as the grid. The penalty function \( \psi(\rho) \) in (7) has the form

\[
\psi(\rho) = \kappa l_{\text{dist}}(\rho),
\]
where $\kappa$ is the cost to violate the distance constraint and $I_{\text{dist}}(\rho)$ is the indicator function of the distance constraint violation,

$$ I_{\text{dist}}(\rho) = \begin{cases} 
1, & \text{if } \rho < 3 \text{ or if } \rho > 10 \\
0, & \text{otherwise}
\end{cases} $$

Figures 4 and 5 show the cross cut surface of the value function and control policy. According to this control policy, the next offset appears to have a strong dependency on the current relative distance between the lead vehicle and the follower vehicle. If the distance becomes too great, the follower vehicle will increase its speed and if the distance becomes too small, the follower vehicle will decrease its speed.

In addition to the policy for the time-varying grade case, another policy which applies to zero grade case has been generated for comparison.

In order to test the control policies and ensure repeatability, a virtual testing environment for in-traffic driving was implemented using CarSim and Simulink software. See [10] for additional details. Fig. 6 illustrates the fuel consumption savings as a percent of the lead vehicle fuel consumption, that is

$$ FC_{\%\text{save}} = \frac{FC_{\text{lead}} - FC_{\text{follow}}}{FC_{\text{lead}}} \times 100, $$

and the dependence of the fuel consumption savings on the lead vehicle velocity.

The two scenarios shown in Fig. 6 concern the evaluation of the control policy developed under the time-varying grade assumption and the evaluation of the control policy developed under the zero grade assumption. In the first scenario the vehicle is simulated on a time-varying grade, in the second scenario the vehicle is simulated on the zero grade. For the time-varying grade, the control policy uses as inputs the current grade, the current distance and the current follower vehicle speed offset. For the zero grade, the control policy uses as inputs the current distance and the current follower vehicle speed offset. Since the control policies are generated on a grid, the nearest point interpolation is employed to determine the control action for state values in the simulation not at the grid points. In all cases, the SDP policy improves the fuel consumption of the follower vehicle relative to the fuel consumption of the lead vehicle.

Fig. 7 illustrates the behavior of the relative distance over the simulated drive when the vehicle speed is 55 mph. Note that the distance is maintained within a restricted range due to the penalty term in the cost function, however, the controller is not able to strictly enforce the distance range constraint, which is effectively treated as soft. At the time instants when the relative distance exceeds the range for which the policy has been generated, an appropriate control action is generated based on policy extension through nearest point extrapolation. Fig. 8 shows that the follower vehicle speed is not constant. It varies periodically and exhibits characteristics of ‘pulse and glide’ type. Fig. 10 illustrates that the fuel flow exhibits significantly larger variations with the optimal control policy of the follower vehicle than for the lead vehicle. The zoom-in of the fuel flow time histories (Fig. 9) shows that more fuel is consumed during accelerations of the follower vehicle and less during coasting-like phases as compared to the lead vehicle.

Figs. 12-14 illustrate the performance of the zero grade policy on the zero grade simulated route when the lead vehicle speed is 55 mph. As these figures confirm, the periodic, ‘pulse and glide’ character of the follower vehicle speed is not due to the grade variation with time, and also occurs for zero grade policies and driving scenarios.

**VII. CONCLUDING REMARKS**

In this paper, the stochastic dynamic programming is applied to construct control policies that prescribe follower vehicle speed relative to the lead vehicle which is driven at a
constant speed on a time-varying grade. The policies are time-independent and use current relative distance, current road grade, and current follower vehicle speed offset as inputs. The policies maintain the relative distance between the vehicles in a prescribed range, and allow the relative distance to vary within this range to improve fuel economy. The policies do not require preview information of the route and of the road grade. The fuel economy improvements of up to 13% (at 45 mph) have been demonstrated in detailed CarSim simulations. Since the policies prescribe the offset relative to the lead vehicle speed, these policies can be applied when lead vehicle speeds that vary slightly relative to the lead vehicle speed value for which these policies have been generated.

The ability of the control policies to yield periodic vehicle speed trajectories, resembling ‘pulse and glide’, has been confirmed through simulations.

**Figure 6:** Percent fuel savings while following a lead vehicle moving at a constant speed on (i) a road with non-zero grade using a policy that accounts for grade, and (ii) on a road with zero grade using a policy that assumes zero grade.

**Figure 7:** Relative distance (m) changing on grade using a policy that accounts for grade variation. Vehicle following of 55 mph lead vehicle.

**Figure 8:** Follower vehicle speed (mph) versus the lead vehicle speed (mph) for the simulation with a time-varying grade, grade-dependent policy and lead vehicle speed of 55 mph.

**Figure 9:** A zoomed in version of Fig. 8: Follower vehicle speed (mph) versus the lead vehicle speed (mph) for the simulation with a time-varying grade, grade-dependent policy and lead vehicle speed of 55 mph.

**Figure 10:** Fuel flow for the follower vehicle (kg/sec) versus the fuel flow for the lead vehicle speed (kg/sec) for the simulation with a time-varying grade, grade-dependent policy and lead vehicle speed of 55 mph.
Figure 11: Fuel flow for the follower vehicle (kg/sec) versus the fuel flow for the lead vehicle speed (kg/sec) for the simulation with a time-varying grade, grade-dependent policy and lead vehicle speed of 55 mph.

Figure 12: Relative distance (m) changing on zero grade when using a zero grade policy for vehicle following a 55 mph lead vehicle.

Figure 13: Follower vehicle speed (mph) versus the lead vehicle speed (mph) for the simulation with a zero grade, grade-independent policy and lead vehicle speed of 55 mph.

Figure 14: Zoomed-in time history of vehicle speed in Fig. 13 with sharp acceleration followed by extended deceleration.

REFERENCES


