DHP Method for Ramp Metering of Freeway Traffic

Dongbin Zhao, *Senior Member, IEEE*, Xuerui Bai, Fei-Yue Wang, *Fellow, IEEE*, Jing Xu, and Wensheng Yu

***Abstract*—This paper presents the design of dual heuristic programming (DHP) for the optimal coordination of ramp metering in freeway systems. Specifically, we implement the DHP method to solve both recurrent and nonrecurrent congestions with queuing consideration. A coordinated neural network controller is achieved by the DHP method with traffic models. Then, it is used for verifications with different traffic scenarios. Simulation studies performed on a hypothetical freeway indicate that the achieved neural controller maintains good control performance when compared with the classical ramp metering algorithm ALINEA. We emphasize that these neural controllers can be developed offline by using approximate traffic models. This offline mechanism avoids the risks of instability that incur during continual online training. We also discuss some real-time implementation issues.**

***Index Terms*—Congestion, dual heuristic programming (DHP), ramp metering, traffic control.**

I. Introduction

WITH THE rapid development of society, the number of vehicles and the need for mobility have increased beyond the current road capacity [2]. This has resulted in congestions, consequent excessive delays, reduced pedestrian and vehicle safety, and increased air pollution. A promising solution to these problems exploits the existing infrastructure through efficient dynamic traffic management and control. Specifically, in freeway traffic systems, many measures could be adopted to improve the service quality of freeways, such as ramp metering, route guidance, reversible lanes, speed limits, and so on [1], [2]. Among these measures, ramp metering is a well-known method extensively used in present freeway traffic systems. This method regulates the volume of traffic entering a given freeway at its entry ramps so that the freeway can operate at some desired level of service. When properly designed, ramp metering can efficiently alleviate recurrent and nonrecurrent congestions, which has been proven both by mathematically sound arguments and in practice [3].

Manuscript received December 27, 2006; revised July 7, 2008, April 12, 2009, June 22, 2010, and October 21, 2010; accepted February 21, 2011. Date of publication March 24, 2011; date of current version December 5, 2011. This work was supported in part by the National Natural Science Foundation of China under Grant 60921061, Grant 60874043, Grant 61034002, and Grant 60874010. The Associate Editor for this paper was B. De Schutter.

D. Zhao and X. Bai are with the Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: dongbin.zhao@ia.ac.cn; baixuerui05@gmail.com).

F. -Y. Wang is with the Key Laboratory of Complex Systems and Intelligence Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the Systems and Industrial Engineering Department, University of Arizona, Tucson, AZ 85721-0020 USA (e-mail: feiyue@gmail.com).

J. Xu is with the IBM China Development Lab, Xi'an 710075, China (e-mail: xjxujing@cn.ibm.com).

W. Yu is with the Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai 200062, China (e-mail: wensheng.yu@ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2011.2122257

The various existing ramp control algorithms can generally be classified into two categories: 1) fixed time metering and 2) traffic-responsive metering [34]. The latter has proved to be more effective in handling freeway congestion than the former. Typical algorithms of this kind include demand capacity, occupancy control [4], and ALINEA [5], [44]. However, these metering strategies are purely local; they are unable to harness integrated consideration over any freeway as a whole. This greatly limits their capability in combating congestions. In response to such limitations, researchers proposed several coordinated (optimal) ramp metering strategies while using various control techniques, such as Linear Quadratic Regulator [6], [7], multilayer control [8], [9], model predictive control [10], nonlinear optimal control [11], [12], reinforcement learning [13], and neural control [14], [15].

In this paper, we will further develop existing rich strategies on coordinated ramp metering control by implementing additional techniques. The technique we used is called approximate dynamic programming (ADP). The concept of ADP was introduced by Werbos in 1977 [15]–[23].

It is well known that traditional dynamic programming (DP) is limited in applications due to its high computation and storage complexity for high-order nonlinear systems; this problem has been designated the curse of dimensionality [25]. However, ADPs are able to artfully circumvent such difficulties by using a critic network for estimating the performance index or its derivatives in DP and an action network for generating optimal actions. ADPs are able to combine the concepts of backpropagation, reinforcement learning [25]–[27], and traditional DP.

The scheme of a general ADP is shown in Fig. 1. The solid lines represent signal flow, whereas the dashed lines represent pathways of backpropagation. The Action module is the controller with the system state $x(k)$ as the input, generating control action on the Plant module. The Critic module is unique in ADP, evaluating how Action works, with the system states $x(k)$ and the control action $u(k)$ as the inputs. The output of Critic $J(k)$ is defined to estimate the cost-to-go $R(k)$ in the Bellman equation. A reward is given according to the state $x(k)$ together with the control action $u(k)$. The upper right part constitutes the learning scheme of Critic. The Critic training error is defined as $r(k) + \gamma J(k) - J(k-1)$; therefore, as the error approaches zero, it could be derived that $J(k) = r(k+1) + \gamma r(k+2)^2 + \gamma^2 r(k+3)^3 + \cdots$, which is the same form

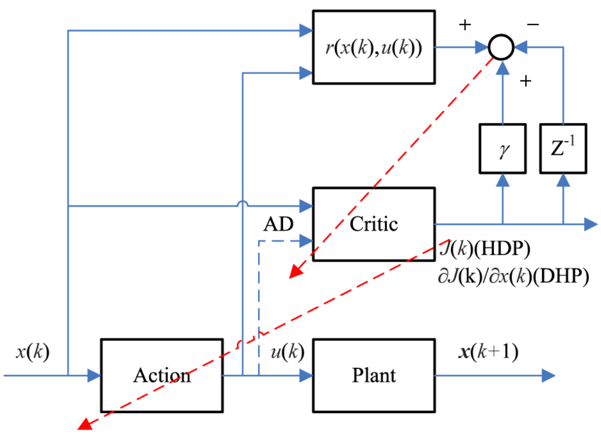


Fig. 1. Schematic of ADP methods.

as $R(k)$. That is to say, the trained Critic could guide Action toward the optimal one.

According to the objectives of the critic approximation, the existing ADP can be categorized as follows: 1) heuristic DP (HDP), which approximates the cost function; 2) dual HDP (DHP), which directly approximates derivatives of the cost function with respect to its state vector; and 3) globalized DHP (GDHP), which approximates both the cost function and its derivatives [19], [21]. Note that all critic approximations of the foregoing three ADP designs use system states as their exclusive inputs. Whereas, if the control actions were included as additional inputs to the critic network, we will get the action-dependent (AD) versions of HDP, DHP, and GDHP, which are called ADHDP, AD DHP, and AD GDHP, respectively. As indicated in Fig. 1, the critic network outputs performance index $J(k)$ in HDP and outputs the derivatives of $J(k)$ with respect to the state vector in DHP. Whether the dash dot line exists determines whether it is an AD method.

ADP-based approaches have many promising benefits, such as the optimality and feedback of DP, the numerical properties, and the real-time performance capabilities of neural networks. Another advantage is that these methods can handle systems with time-delay elements. For such systems, supervised learning may not be a valid option because it utilizes instantaneous errors between the desired output and the actual output. However, ADPs are effective under such conditions because they allow the network to learn according to the cost-to-go of the present error state. This capability of ADP makes it a preferable method for solving many traffic control problems. In [29] and [30], we have proposed an ADHDP controller for the local ramp metering problem. The cost function in ADHDP is scalar, whereas in DHP, the derivative of the cost function to the states is a vector. Thus, the derivative may be able to provide more information than the cost function in the training of critic and action networks. Therefore, in this paper, we propose the use of DHP designs to optimally coordinate metering of multiramps.

This paper is organized as follows: In Section II, we formulate the traffic problem with macroscopic traffic models. In Section III, we present the general DHP design for the traffic problem. Then, in Section IV, we implement the DHP method to resolve recurrent and nonrecurrent traffic congestions.

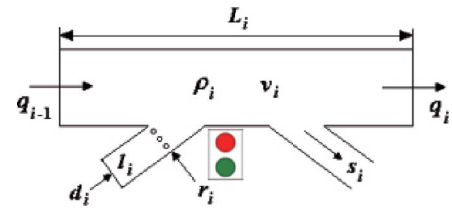


Fig. 2. Freeway section with on/off ramps.

Finally, the conclusion and topics for future research are stated in Section V.

II. Traffic Problem Formulation

The macroscopic freeway traffic model used here was originally derived by Payne [31] and modified by Papageorgiou [1] and Cremer and May [32].

Suppose a freeway lane is subdivided into N sections with a length $L_i (i = 1, \ldots, N)$, each having, at most, one on-ramp and one off-ramp (as schematically shown in Fig. 2). The evolution of freeway traffic flow can then be described by

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i}\left[q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)\right] \tag{1}$$

$$q_i(k) = \alpha\rho_i(k)v_i(k) + (1-\alpha)\rho_{i+1}(k)v_{i+1}(k) \tag{2}$$

$$\begin{aligned} v_i(k+1) = v_i(k) &+ \frac{T}{\Delta t}\left\{V_e\left[\rho_i(k)\right] - v_i(k)\right\} \\ &+ \frac{T}{L_i} v_i(k)\left[v_{i-1}(k) - v_i(k)\right] \\ &- \frac{\mu T}{\Delta t L_i}\frac{\rho_{i+1}(k) - \rho_i(k)}{\rho_i(k) + \kappa} \end{aligned} \tag{3}$$

where T is the sample time interval, $0 < \alpha < 1$ is the weighting factor, Δt, μ, and κ are positive constants, $\rho_i(k)$ is the traffic density in section i at time kT, $v_i(k)$ is the average speed, $q_i(k)$ is the traffic flow leaving section i and entering section $i+1$, $r_i(k)$ is the metering rate, $s_i(k)$ is the off-ramp volume, and $V_e[\rho_i(k)]$ is the equilibrium mean speed modeled by

$$V_e\left[\rho_i(k)\right] = v_f \exp\left(-\frac{1}{a_m}\left[\frac{\rho_i(k)}{\rho_c}\right]^{a_m}\right) \tag{4}$$

where ρ_c is the critical traffic density. v_f and a_m are constants to be identified for real traffic flow.

The off-ramp volumes $s_i(k)$ are related to the traffic volumes $q_{i-1}(k)$ through the relationship $s_i(k) = \varepsilon_i q_{i-1}(k)$ [43], where $0 < \varepsilon_i < 1$. Substituting this relationship and (2) into (1) yields

$$\begin{aligned} \rho_i(k+1) &= \rho_i(k) + \frac{T}{L_i}\left[\alpha(1-\varepsilon_i)\rho_{i-1}(k)v_{i-1}(k)\right. \\ &\qquad + (1 - 2\alpha + \varepsilon_i\alpha - \varepsilon_i)\rho_i(k)v_i(k) \\ &\qquad \left. - (1-\alpha)\rho_{i+1}(k)v_{i+1}(k) + r_i(k)\right]. \end{aligned} \tag{5}$$

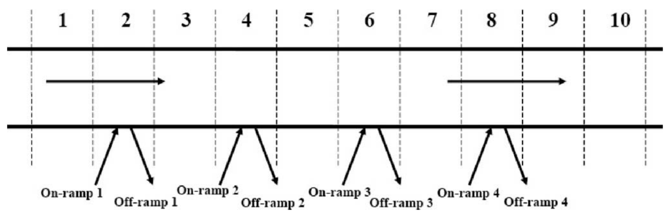


Fig. 3. Freeway stretch.

To complete the model, the dynamics of the queue l_i on the on-ramp of section i can be described as

$$l_i(k+1) = l_i(k) + T\left[d_i(k) - r_i(k)\right] \tag{6}$$

where $d_i(k)$ is the corresponding traffic demand.

The control variables $r_i(k)$ are subject to the following constraint:

$$r_{i\cdot \min \cdot k} \le r_i(k) \le r_{i\cdot \max \cdot k} \tag{7}$$

where

$$r_{i\cdot \min \cdot k} = \max\left\{0, d_i(k) - \frac{1}{T}\left[l_{i\cdot \max} - l_i(k)\right]\right\}$$

$$r_{i\cdot \max \cdot k} = \min\left\{r_{i\cdot \max}, d_i(k) + \frac{1}{T} l_i(k)\right\}$$

where $l_{i\cdot \max}$ is the maximal queue length on the on-ramp of section i, and $r_{i\cdot \max}$ is the maximum metering rates and fixed parameters that are dependent on road characteristics.

Finally, we assume the following boundary conditions [32] for the entrance and the exit:

$$\rho_0(k) = \frac{q_0(k)/v_1(k) - (1-\alpha)\rho_1(k)}{\alpha} \tag{8}$$

$$v_0(k) = v_1(k) \tag{9}$$

$$\rho_{N+1}(k) = \rho_N(k) \tag{10}$$

$$v_{N+1}(k) = v_N(k) \tag{11}$$

where $q_0(k)$ is the traffic demand on the freeway.

Based on the traffic model previously formulated, various optimization criterions can be specified and achieved by properly regulating the on-ramp metering rates. In Section IV, a hypothetical freeway consisting of ten sections with four on-ramps and four off-ramps, as shown in Fig. 3, is investigated.

III. Dual Heueristic Programming Coordinated Algorithm

Consider the following discrete-time nonlinear (time-varying) dynamical system:

$$x(k+1) = F\left[x(k), u(k), k\right], \qquad k = 0, 1, 2, \ldots \tag{12}$$

where $x \in R^n$ represents the state vector, $u \in R^m$ denotes the control action, and $F(\cdot)$ is a general nonlinear function. The performance index (or cost-to-go function) associated with this system is

$$J(k) = \sum_{i=k}^{\infty} \gamma^{i-k} U(i) = U(k) + \gamma J(k+1) \tag{13}$$

where $U(k)$ is called the primary cost function, and γ is the discount factor with $0 < \gamma \le 1$.

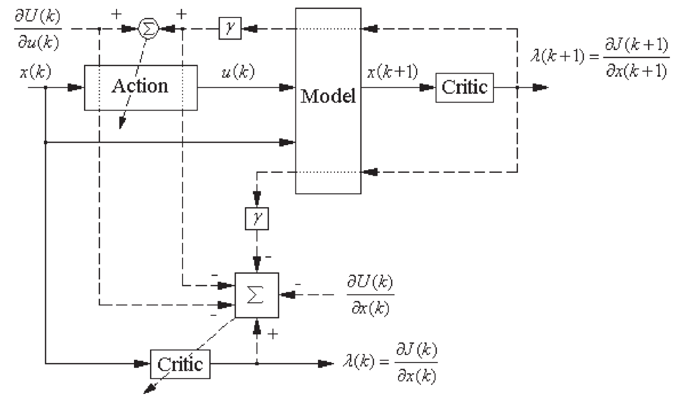


Fig. 4. Adaptation in DHP. The solid lines represent signal flow, whereas the dashed lines represent pathways of backpropagation. Components of the vector $\lambda(k+1)$ are propagated back from outputs $x(k+1)$ of the model to its inputs $x(k)$ and $u(k)$, yielding the third terms of (16) and the term of (21), respectively. The latter is propagated back from outputs $u(k)$ of the action network to its inputs $x(k)$, thus getting the fourth term of (16). Backpropagation of the vector $\partial U(k)/\partial u(k)$ through the action network yields a vector with components consisting of the second term of (16). Following (15) and (20), the summators produce error vectors $E_c(k)$ and $E_a(k)$ that are used to adapt the critic and action networks, respectively.

The curse of dimensionality often makes it untenable to run true DP. Fortunately, ADPs provide us with feasible ways for approximate optimal solutions by including a critic network that approximates the cost function or its derivatives. HDP is the simplest design of ADP. The critic network in HDP estimates the function J in the Bellman equation of DP. However, it has been criticized for its limitations when handling complex problems [19]. DHP is one of the more complex forms of ADP. The critic network in DHP estimates the derivatives of J with respect to the state vector. Derivatives of the cost function are more effective for the given sizes of the network and training data. The output of DHP is a vector instead of a scalar, as in the case of HDP. The derivatives provide additional information indicating which action to change [19].

The schematic for DHP is described in Fig. 4. Clearly, there exist three modules: 1) the model; 2) the action network; and 3) the critic network (two critic networks are the same when shown in two consecutive moments in time). The model module functions as a differentiable form of the controlled plant. It can directly represent the analytical formulation of the plant if such a formulation exists, or it can indirectly represent a differentiable neural network trained to approximate the plant if such a formulation does not exist. The model network can be trained beforehand offline or trained parallel to the critic and action networks. This module connects with two other modules, expediting both the forward signal path and the backward error signal path.

A. Critic Network

The critic network is expected to output derivatives of the cost function with respect to system states. For this purpose, we try to minimize the following errors over time:

$$\|E_c\| = \sum_k E_c^T(k) E_c(k) \tag{14}$$

where

$$E_c(k) \triangleq col\{\lambda_s(k) - \lambda_s^o(k), \quad s = 1, \ldots, n\} \tag{15}$$

where $\lambda_s(k)$ is the sth output of the critic network, which approximates the derivative of the cost function with respect to the state $x_s(k)$ (also called the co-state), and $\lambda_s^o(k)$ is its desired value. Applying (13), we get

$$\begin{aligned}\lambda_s^o(k) &= \frac{\partial J(k)}{\partial x_s(k)} \\ &= \frac{\partial}{\partial x_s(k)}[U(k) + \gamma J(k+1)] \\ &= \frac{\partial U(k)}{\partial x_s(k)} + \sum_{j=1}^{m} \frac{\partial U(k)}{\partial u_j(k)} \frac{\partial u_j(k)}{\partial x_s(k)} \\ &\quad + \gamma \sum_{i=1}^{n} \lambda_i(k+1) \frac{\partial x_i(k+1)}{\partial x_s(k)} \\ &\quad + \gamma \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_i(k+1) \frac{\partial x_i(k+1)}{\partial u_j(k)} \frac{\partial u_j(k)}{\partial x_s(k)} \end{aligned} \tag{16}$$

where $\partial U(k)/\partial x_s(k)$ and $\partial U(k)/\partial u_j(k)$ are directly calculated according to the specified primary cost function $U(k)$. $\partial x_i(k+1)/\partial x_s(k)$ and $\partial x_i(k+1)/\partial u_j(k)$ are determined from the system model, and the partial derivative $\partial u_j(k)/\partial x_s(k)$ is calculated by backpropagation through the action network. The word "dual" is used to describe a situation where the desired derivatives are calculated using dual subroutines (states and co-states) to backpropagate the estimated derivatives through the model and the action network, as shown in Fig. 4, e.g.,

$$\mathbf{w}_c(k+1) = \mathbf{w}_c(k) + \Delta \mathbf{w}_c(k) \tag{17}$$

$$\Delta \mathbf{w}_c(k) = l_c \left[-\frac{\partial E_c(k)}{\partial \mathbf{w}_c(k)} \right] \tag{18}$$

where l_c is the learning rate with critic network. For the weights updating in the critic network, the least mean square (LMS) training algorithm can be applied.

B. Action Network

The action network is adapted in Fig. 4 by propagating $\lambda(k+1)$ back through the model down to the action, and it tries to minimize the following errors over time:

$$\|E_a\| = \sum_k E_a^T(k) E_a(k) \tag{19}$$

where

$$E_a(k) \triangleq col\left\{ \frac{\partial U(k)}{\partial u_j(k)} + \gamma \frac{\partial J(k+1)}{\partial u_j(k)}, \quad j = 1, \ldots, m \right\} \tag{20}$$

where

$$\frac{\partial J(k+1)}{\partial u_j(k)} = \sum_{i=1}^{n} \lambda_i(k+1) \frac{\partial x_i(k+1)}{\partial u_j(k)} \tag{21}$$

$$\mathbf{w}_a(k+1) = \mathbf{w}_a(k) + \Delta \mathbf{w}_a(k) \tag{22}$$

$$\Delta \mathbf{w}_a(k) = l_a \left[-\frac{\partial E_a(k)}{\partial \mathbf{w}_a(k)} \right] \tag{23}$$

where l_a is the learning rate with the action network. For the weights updating in the action network, the LMS training algorithm can be applied.

There are many methods suggested for the training of critic and action networks [19], [20]. One way is to split the process into two separate training cycles: one cycle for the action and one cycle for the critic. For example, we can train the critic network while keeping the action network fixed, and then, we train the action network while keeping the critic network fixed. Such an alternating training process is repeated until an acceptable level of system performance is achieved. Another way is to conduct the training of the two networks concurrently at each control step. We adopt the latter in this paper, and this training procedure is described as follows:

Concurrent training of the action and critic networks:

1) $k = 1$. Randomly initialize traffic densities, average speeds, and traffic demands according to the model previously described.
2) Shift and scale $x(k)$, and apply it to the action network, obtaining $u(k)$.
3) Scale $u(k)$, and apply it to the plant, obtaining $x(k+1)$.
4) If the traffic densities $\rho_i(k+1)$ lie outside of the specified range, e.g., [10, 180], or the queue lengths $l_i(k+1)$ are beyond the specified range of $[0, l_{i\cdot\max}]$, or the metering rates $r_i(k)$ are beyond the specified range defined by (7), then go to 1 to start a new training epoch.
5) Shift and scale $x(k+1)$, and apply it to the critic network, obtaining $\lambda(k+1)$.
6) Shift and scale $x(k)$, and apply it to the critic network, obtaining $\lambda(k)$.
7) Calculate and execute weights updating for the action network.
8) Calculate and execute weights updating for the critic network.
9) If $k < epoch$, then increment k and go to 2).
10) Output the trained parameters.

Finally, we note that the critic and action networks do not require exclusively neural network implementations, and any differentiable structure (such as fuzzy model) suffices as the building block.

IV. Simulation

Using the foregoing traffic model, we will demonstrate the use of DHP for the coordinated control of ramp metering in recurrent and nonrecurrent congestions. Recurrent congestion reflects the day-to-day buildup of traffic on urban freeway and arterials, which are, notably, during the morning and afternoon commuter peak periods. Nonrecurrent congestion reflects the delays caused by accidents, truck spills, inclement weather, etc. [46].

The simulation parameters of the traffic model are chosen from the literature [45] for comparison. The segment length is 500 m. We assume that the free flow speed will be smaller than 120 km/h. The simulation step T is equal to 10 s. Then, for traffic models, relative parameters are specified as follows: $T = 10$ s, $L_i = 0.5$ km, $\alpha = 0.9$, $\varepsilon_i = 0.15$, $\Delta t = 18$ s, $\mu = 21.6$ km^2/h, $\kappa = 40$ veh/km, $v_f = 110$ km/h, $\rho_{jam} = 180$ veh/km, $\rho_c = 35$ veh/km, and $a_m = 1.636$. For the sake of convenience, each ramp metering rate is confined within the range of [0, 1000] veh/h. The capacity of the simulated freeway stretch is 8000 veh/h for four lanes.

A. ALINEA

ALINEA [5], [44] is a linearized feedback control algorithm that adjusts the metering rate to keep the occupancy downstream of on-ramps at a desired level, which is adopted here for comparison. ALINEA maintains the desired level of occupancy by using feedback regulation. A popular measure to solve the queuing problem is to place a detector at the on-ramp and release ramp metering when the occupancy exceeds the maximal allowed queue length. Therefore, the ALINEA closed-loop ramp metering strategy considering queue length is

$$\begin{cases} r(k) = r(k-1) - Y\left[\rho(k) - \rho_d\right], & \text{if } l \le l_{\max} \\ r(k) = d(k), & \text{if } l > l_{\max} \end{cases} \tag{24}$$

where Y is a positive parameter (specified as 50 km/h), and ρ_d is the desired traffic density (specified as 34 veh/km, which is a little lower than the critical density).

B. DHP for Nonrecurrent Congestion

1) DHP Training: We first intend to resolve the nonrecurrent traffic congestion with queuing consideration with DHP controller. It has been argued that the control method, such as ALINEA and the proposed control algorithm, may result in long queues on metered ramps because these control methods only consider the mainstream traffic on the freeway. One way to resolve this problem is to increase the control value with a higher metering rate until ramp queues are reduced below certain thresholds. A better way is to consider queuing with more explicit terms in control algorithms. Therefore, we define the primary cost function in system performance (13) as

$$U(k) = c_1 \sum_{i=1}^{10} T\rho_i(k)L + c_2 \sum_{j=1}^{4} l_j^2(k) \tag{25}$$

where c_1 and c_2 are positive weighting parameters set as $c_1/c_2 = 36\,000$. The first term in (25) is used to minimize the total time spent (TTS) on freeways, and the second term is used to diminish queues down to comparable length. With the queuing consideration, the queue at an on-ramp cannot grow larger than the physically available space, which is added as a hard constraint to the optimization by limiting the queue lengths at the on-ramps to $l_{i \cdot \max} = 200$ vehicles or less.

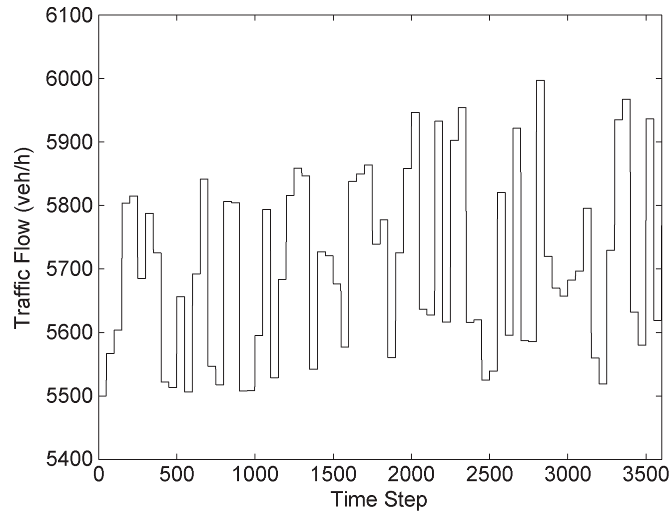


Fig. 5. Typical scheme of freeway demand for the training.

According to the primary cost function (25), the system state vector is chosen as $x(k) \triangleq col\{\rho_i(k), l_j(k), i = 1, \ldots, 10, j = 1, 2, 3, 4\}$. Corresponding to each system state vector, the critic network is structured as ten input neurons, 15 hidden neurons (determined experimentally) with $logsig(\cdot)$ activation functions, and ten output neurons with linear activation functions. As for the action network, its inputs are the same as those of the critic network, and its outputs (properly scaled) end up as ramp metering rates. Thus, the action network consists of ten input neurons, 15 hidden neurons (determined experimentally) with $logsig(\cdot)$ activation functions, and four output neurons with $logsig(\cdot)$ activation functions. The critic and action networks are trained with learning rates $l_c = 0.1$ and $l_a = 0.2$, respectively.

We consider one epoch to be 3600 control (or training) steps. During each training epoch, we set the freeway traffic demand $q_0(k)$ in uniform random from the interval [5500, 6000] veh/h, and each value lasts for 50 time steps. A typical scheme of traffic demand for training epochs is shown in Fig. 5, which seems as a superimposed standard deviation noise. Such a scheme is motivated by the desire to have excitations across a range of possible system states. Moreover, the stochastic variation of traffic demands appears to have an annealing effect, reducing the likelihood of getting stuck in some local optimum. Of course, these benefits mean that more training effort is needed. We uniformly initialize the queue length using values within the range of [20, 60] veh and using traffic densities within the range of [20, 30] veh/km. For traffic demands on on-ramps presented in sections 2, 4, 6, and 8, we use constant values of 500 veh/h during training for each on-ramp, respectively. The total flows of the mainstream, on-ramps, and off-ramps are below the capacity. We use the aforementioned concurrent strategy to train critic and action networks. We expect that the learned feedback mechanism of our proposed neural controller is capable of handling changes in traffic demand.

2) DHP Testing: After training 3600 epochs, we obtain the neural controller for testing. Then, we start to test the capability of the DHP controller in combating congestions on the freeway.

TABLE I
INITIAL VALUES USED FOR THE SECOND TEST

Section No.	1	2	3	4	5	6	7	8	9	10
$\rho_i(0)$	30	30	30	30	30	30	30	85	85	30
$v_i(0)$	66	66	66	66	66	66	66	5	5	66
$l_i(0)$		30		30		30		30		

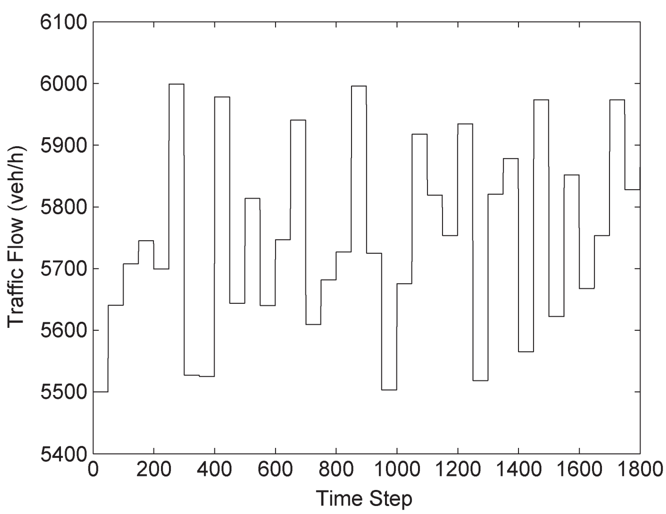


Fig. 6. Freeway demand for the testing.

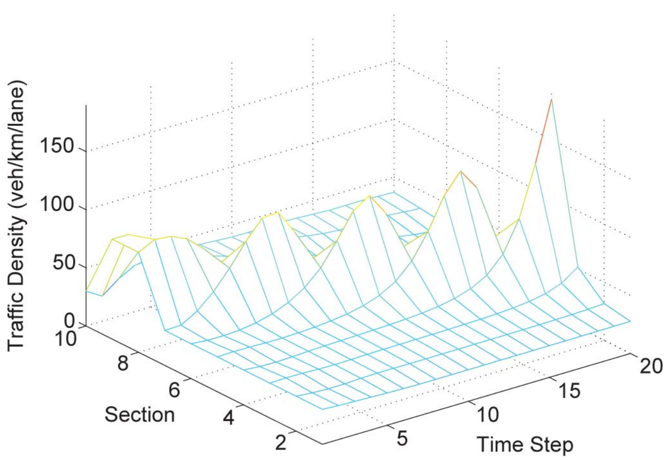


Fig. 7. Traffic density of ALINEA in nonrecurrent congestion.

Given a jam situation as in sections 8 and 9. The initial congestion is specified as shown in Table I, and the initial queues are 30 vehicles in length. The traffic demands for each on-ramp are 850, 650, 350, and 550 veh/h, respectively. The traffic demand in the mainstream is shown in Fig. 6.

Due to the high traffic density on the freeway and the high on-ramp demands, the traffic density grows rapidly. It can clearly be seen from Fig. 7 that the ALINEA method leads to the jammed traffic. Because the densities are far more than the critical density, only the first 20 steps are shown in Fig. 7. The traffic breakdown is due to the limitation of the ALINEA ramp metering algorithm in regulating the traffic flow and the severe nonrecurrent congestion.

However, since the use of a more effective learning mechanism and a reasonable performance index function, the DHP controller handles this congestion quite efficiently by imple-

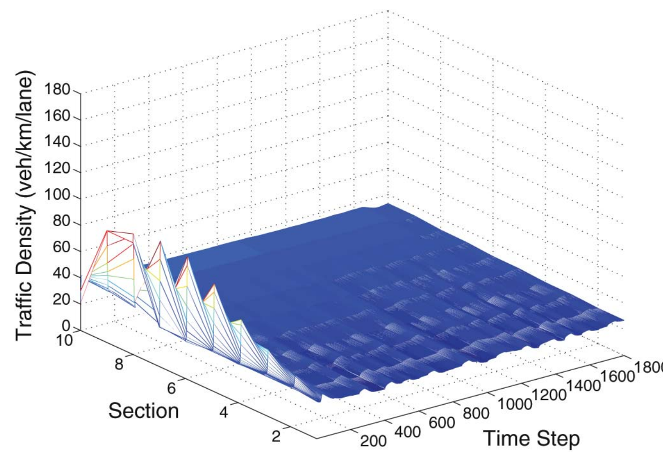


Fig. 8. Traffic density of DHP in nonrecurrent congestion.

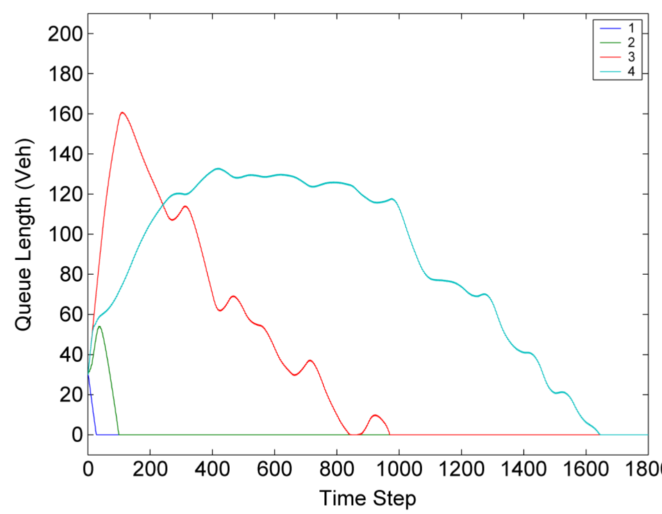


Fig. 9. Queue length under DHP control in nonrecurrent congestion.

menting coordinated ramp metering, as shown in Fig. 8. At the beginning, the congestion starts to spread to the upper sections in this freeway, and the queue length starts to increase. By diverting the traffic flow in the congested section and limiting the inflow of vehicles onto the freeway, after a couple of time steps, the density gradually decreases to desired density, and the queue length gradually approaches to zero. The congestion that occurs in sections 8 and 9 under ALINEA does not appear again under the DHP controller. Fig. 9 depicts the queue evolution profile. It is clear that the congestion is successfully resolved and the queues length are efficiently regulated under DHP control.

C. *DHP for Recurrent Congestion*

Next, we intend to test the performance of resolving the recurrent traffic congestion with the previously obtained DHP controller.

A simulation is conducted with the traffic model in the morning rush hour. The rush hour ranges from 5 A.M. to 10 A.M. As a result, there are $N_{\text{sim}} = 1800$ simulation steps. The initial states are specified as shown in Table II. The evolution of traffic flow at the entrance of the freeway is presented in Fig. 10. The

TABLE II
INITIAL VALUES IN THE MORNING RUSH-HOUR RECURRENT CONGESTION

Section No.	1	2	3	4	5	6	7	8	9	10
$\rho_i(0)$	10	10	10	10	10	10	10	10	10	10
$v_i(0)$	97	97	97	97	97	97	97	97	97	97
$l_i(0)$		30		30		30		30		

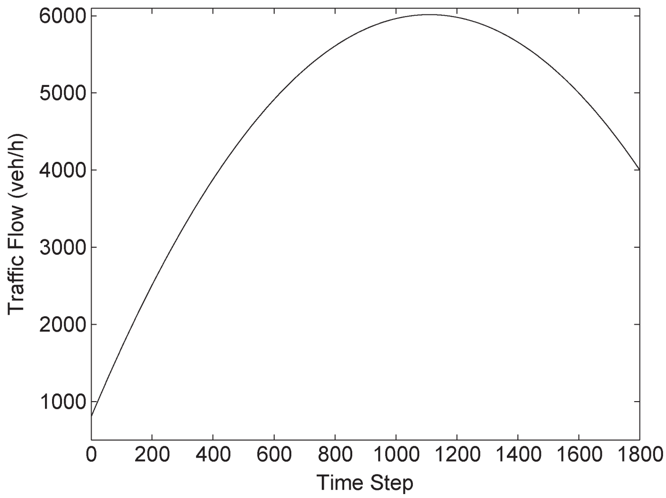


Fig. 10. Traffic flow during the morning rush-hour recurrent congestion.

traffic demands are constant at 850, 650, 350, and 550 veh/h for each on-ramp, respectively.

First, the model is simulated without any control to have a reference value of the performance measure, which can subsequently be compared with the performance values resulting from simulations with ALINEA and DHP control. Fig. 11 shows the simulation results of the traffic density evolution under no control. The traffic density in the first section varies with the input traffic demand. It indicates the occurrence of congestion in the first section of the freeway. For example, in section 1, the traffic density grows larger than the critical density during rush hour. Then, the recurrent congestion sets in. We apply ALINEA and DHP methods to resolve such problems. The simulation results in Figs. 12 and 13 show that both ALINEA and DHP can efficiently alleviate the congestion. Note that their vertical axis limits are different from the no control case. However, Figs. 14 and 15 show that ALINEA-based control results in large queue length. The coordinated ramp metering method developed with DHP can handle both the congestion and queue length well.

The TTS by all the vehicles is also defined for explicit comparison as

$$\mathrm{TTS} = \sum_{k=0}^{N_{\mathrm{sim}}-1} \left[\sum_{i=1}^{10} \rho_i(k) L + \sum_{j=1}^{4} l_j(k) \right] T \qquad (26)$$

where N_{sim} is the number of simulation steps in the simulated period. TTS consists of the TTS by all the vehicles on the freeway sections [the first term in (26)] plus the TTS by the vehicles in the queues at the on-ramps [the second term in (26)].

The TTS of each ramp metering method is summarized in Table III. The lower the TTS during the simulated 5-h

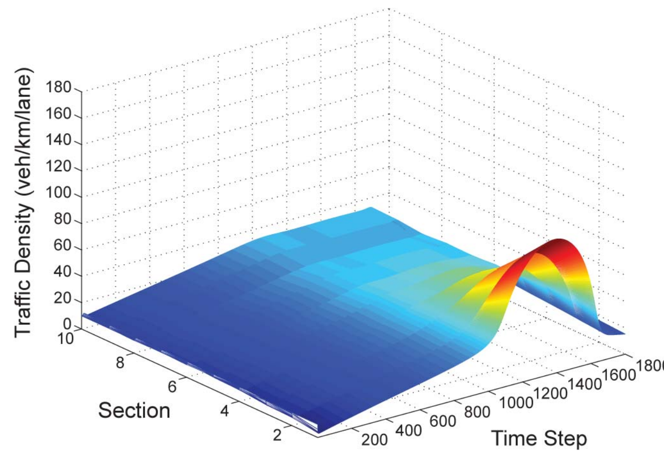


Fig. 11. Traffic density without control during the morning rush-hour recurrent congestion.

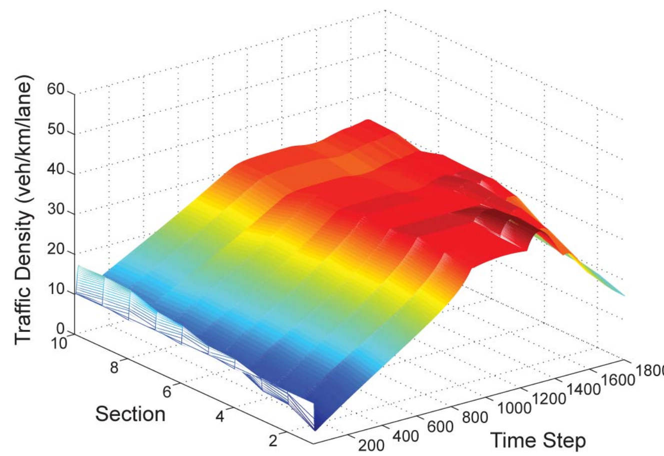


Fig. 12. Traffic density under ALINEA control during the morning rush-hour recurrent congestion.

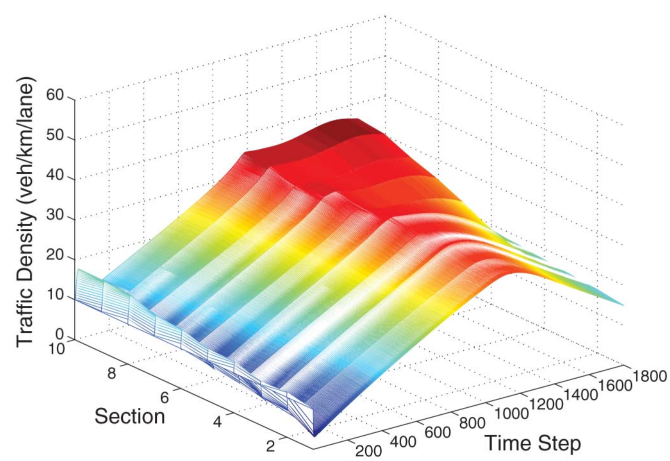


Fig. 13. Traffic density profile under DHP control during the morning rush-hour recurrent congestion.

period, the higher the performance of the freeway system. The performance improves by 7.4% and 12.3% with ALINEA and DHP method, respectively, compared with the no-control case. It can be seen that the DHP method manages to comply with the queue constraints imposed and at the same time regulate the TTS on the freeway. The coordination of the metering

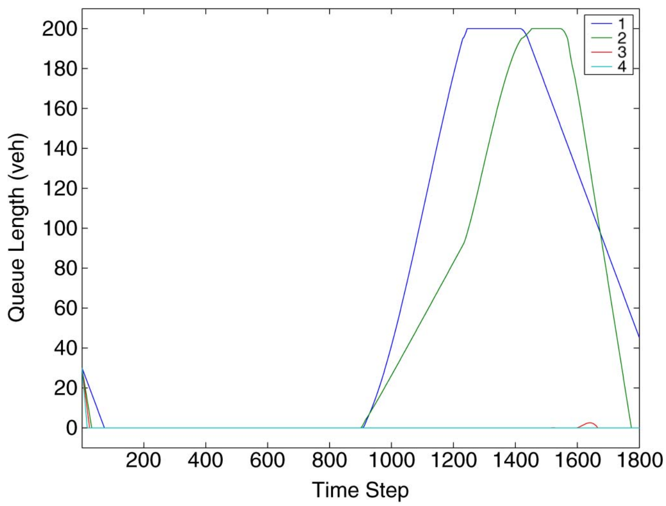


Fig. 14. Queue length under ALINEA control during the morning rush-hour recurrent congestion.

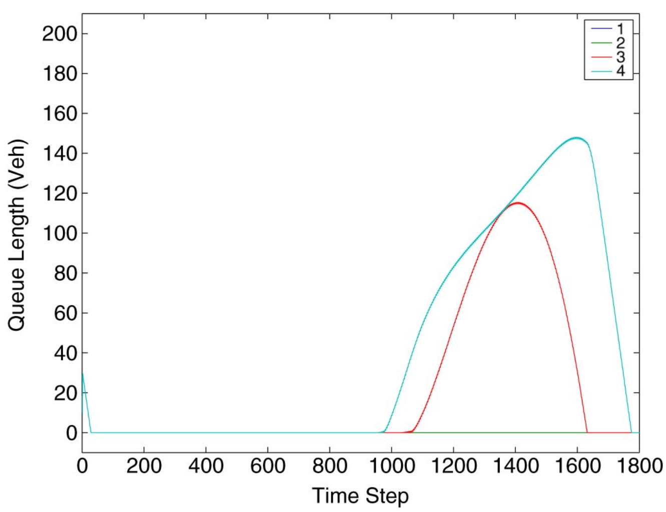


Fig. 15. Queue length profile under DHP control during the morning rush-hour recurrent congestion.

TABLE III
OVERVIEW OF THE TTS ON THE FREEWAY FOR DIFFERENT RAMP METERING METHODS IN THE MORNING RUSH-HOUR RECURRENT CONGESTION

Method	No Control	ALINEA	DHP
TTS (*veh.h*)	2861	2649	2499
TTS Reduction (%)	0	7.4	12.3

rates of the different on-ramps assures that the control actions taken at different locations in the network reinforce rather than counteract each other.

D. Real-Time Implementation Issues

The simulation computer is with Pentium-4 2.4-GHz CPU and 1-GB memory. The average learning time is 50 s. We use offline training in this paper. If the freeway size increases, then the input to the neural network increases, which may result in larger computational burden. However, it is still acceptable for offline training process.

For control systems with shorter time constants, the computation time could be problematic for a neural network with many weight requirements for continual online adaptation. In addition, frequent online training can also lead to instability. Therefore, we adopt the offline training mechanism in this paper. The convergence guarantee of the critic and the action networks during offline training is described in [35], [36], and [40]–[42].

During real-time implementation, our approach can be deployed according to the local simple and remote complex (LSRC) design principle [37]. For example, the offline training is originally conducted in the remote traffic control center. Once trained, the neural controller is sent back into the local field for metering control with fixed weight parameters. With the information collected during online control, the neural controller may then be retrained in the control center. Using this mechanism, we can avoid the risk of instability and guarantee real-time demands. We note that the use of approximate models during DHP design makes the LSRC design principle more practical.

V. CONCLUSION AND FUTURE RESEARCH

In this paper, we have proposed a design of DHP to solve the traffic congestion problems. Using a hypothetical freeway stretch, we resolve the nonrecurrent and recurrent traffic congestions with queuing consideration. Simulation results indicate that the DHP controllers are effective against traffic congestions on the freeway.

Although DHP has proved effective in combating congestion problems, there are still many other problems to be solved, which may be taken as future research directions, as follows:

1) further evaluation of the DHP metering controllers using microscopic traffic simulators (such as TSIS, Paramics, etc.);
2) further investigation of the effectiveness of DHP by including other traffic control measures (such as speed limits, route guidance, reversible lanes, dynamic lane assignment, etc.) for wider traffic networks;
3) real-time implementation of the DHP controller on a real freeway stretch according to the LSRC principle.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers' comments and suggestions on improving the quality of the paper.

REFERENCES

[1] M. Papageorgiou, *Application of Automatic Control Concepts to Traffic Flow Modeling and Control*. New York: Springer-Verlag, 1983.

[2] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 271–281, Dec. 2002.

[3] J. Robinson and M. Doctor, "Ramp metering status in North America," Office Traffic Oper., Fed. Highway Admin., U.S. Dept. Transp., Washington, DC, 1995.

[4] R. Wilshire, R. Black, R. Grochoske, and J. Higonbotham, *ITE Traffic Control Systems Handbook*, Washington, DC, Dec. 1985.

[5] M. Papageorgiou, H. Hadj-Salem, and J. M. Blosseville, "ALINEA: A local feedback control law for on-ramp metering," *Transp. Res. Rec.*, no. 1320, pp. 58–64, 1991.

[6] N. B. Golstein and K. S. P. Kumar, "A decentralized control strategy for freeway regulation," *Transp. Res. Part B, Methodological*, vol. 16, no. 4, pp. 279–290, Aug. 1982.

[7] M. Papageorgiou, J. M. Blosseville, and H. S. Habib, "Modeling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: Coordinated on-ramp metering," *Transp. Res. Part A, General*, vol. 24, no. 5, pp. 361–370, Sep. 1990.

[8] M. Papageorgiou, "Multilayer control system design applied to freeway traffic," *IEEE Trans. Autom. Control*, vol. AC-29, no. 6, pp. 482–490, Jun. 1984.

[9] P. Y. Li, R. Horowitz, L. Alvarez, J. Frankel, and A. M. Robertson, "Traffic flow stabilization," in *Proc. Amer. Control Conf.*, Seattle, WA, 1995, pp. 144–149.

[10] A. Hegyi, B. D. Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transp. Res. Part C, Emerging Technol.*, vol. 13, no. 3, pp. 185–209, Jun. 2005.

[11] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Salem, "Coordinated and integrated control of motorway networks via non-linear optimal control," *Transp. Res. Part C, Emerging Technol.*, vol. 10, no. 1, pp. 65–84, Feb. 2002.

[12] Y. H. Chiang and J. C. Juang, "Control of freeway traffic flow in unstable phase by H_∞ theory," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 2, pp. 193–208, Jun. 2008.

[13] C. Jacob and B. Abdulhai, "Automated adaptive traffic corridor control using reinforcement learning: Approach and case studies," in *Proc. 85th Annu. Meeting Transp. Res. Board*, 2006, pp. 1–8.

[14] A. D. Febbraro, T. Parisini, S. Sacone, and R. Zoppoli, "Neural approximations for feedback optimal control of freeway systems," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 302–313, Jan. 2001.

[15] H. M. Zhang, S. G. Ritchie, and R. Jayakrishnan, "Coordinated traffic-responsive ramp control via nonlinear state feedback," *Transp. Res. Part C, Emerging Technol.*, vol. 9, no. 5, pp. 337–352, Oct. 2001.

[16] P. J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *Gen. Syst. Yearbook*, vol. 22, pp. 25–38, 1977.

[17] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York: Van Nostrand, 1992, ch. 13.

[18] R. A. Santiago and P. J. Werbos, "New progress towards truly brain-like intelligent control," in *Proc. World Congr. Neural Netw.*, San Diego, CA, 1994, pp. 27–33.

[19] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, Sep. 1997.

[20] G. G. Lendaris and C. Paintz, "Training strategies for critic and action neural networks in dual heuristic programming method," in *Proc. Int. Joint Conf. Neural Netw.*, Houston, TX, 1997, vol. 2, pp. 712–717.

[21] J. Si and Y. T. Wang, "On-line learning control by association and reinforcement," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 264–276, Mar. 2001.

[22] D. Liu, X. Xiong, and Y. Zhang, "Action-dependent adaptive critic designs," in *Proc. Int. Joint Conf. Neural Netw.*, Washington, DC, 2001, vol. 2, pp. 990–995.

[23] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 2, pp. 140–153, May 2002.

[24] L. Yang, R. Enns, Y. T. Wang, and J. Si, "Direct neural dynamic programming," in *Stability and Control of Dynamical Systems With Applications*, D. Liu and P. J. Antsaklis, Eds. Boston, MA: Birkhaüser, 2003, ch. 10.

[25] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

[26] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, Aug. 1988.

[27] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3, pp. 279–292, 1992.

[28] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[29] X. R. Bai, D. B. Zhao, and J. Q. Yi, "Ramp metering based on adaptive dynamic programming," in *Proc. Int. Joint Conf. Neural Netw.*, Hong Kong, SAR, 2008, pp. 1847–1852.

[30] X. R. Bai, D. B. Zhao, and J. Q. Yi, "Coordinated control of multiple ramps metering based on ADHDP (λ) controller," *Int. J. Innov. Comput., Inf. Control*, vol. 5, no. 10(B), pp. 3471–3481, 2009.

[31] H. J. Payne, "Models of freeway traffic and control," in *Proc. Math. Models Public Syst., Simul. Council*, 1971, pp. 51–61.

[32] M. Cremer and A. D. May, "An Extended Traffic Model for Freeway Control," Inst. Transp. Stud., Univ. California, Berkeley, Res. Rep. UCB-ITS-RR-85-7, 1985.

[33] T. T. Shannon and G. G. Lendaris, "Qualitative models for adaptive critic neurocontrol," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 1999, vol. 1, pp. 455–460.

[34] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proc. IEEE*, vol. 91, no. 12, pp. 2043–2067, Dec. 2003.

[35] D. Prokhorov and D. C. Wunsch, "Convergence of critic-based training," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 1997, vol. 4, pp. 3057–3060.

[36] D. Prokhorov and L. A. Feldkamp, "Analyzing for Lyapunov stability with adaptive critics," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 1998, vol. 2, pp. 1658–1661.

[37] F. -Y. Wang, "OSGi and agent-based technology for control and management of networked applications," in *Proc. Chin. Conf. Sci. Tech. Household Electron.*, Hangzhou, China, 2001, pp. 81–93.

[38] F. -Y. Wang and G. N. Saridis, "Suboptimal control for nonlinear stochastic systems," in *Proc. 31st IEEE Conf. Decision Control*, Tucson, AZ, 1992, pp. 1856–1861.

[39] F.-Y. Wang and G. N. Saridis, "On successive approximation of optimal control of stochastic dynamic systems," in *Modeling Uncertainty An Examination of Stochastic Theory, Methods, and Applications*, M. Dror, P. L. Ecuyer, and F. Szidarovszky, Eds. Boston, MA: Kluwer, 2002.

[40] K. Doya, "Reinforcement learning in continuous time and space," *Neural Comput.*, vol. 12, no. 1, pp. 219–245, Jan. 2000.

[41] L. C. Baird, "Reinforcement learning with function approximation," in *Proc. 12th Int. Conf. Mach. Learn.*, 1995, pp. 30–37.

[42] C. J. Lin and C. T. Lin, "Reinforcement Learning for an ART-based fuzzy adaptive learning control network," *IEEE Trans. Neural Netw.*, vol. 7, no. 3, pp. 709–731, May 1996.

[43] T. Bellemans, B. D. Schutter, and B. D. Moor, "Model predictive control for ramp metering of motorway traffic: A case study," *Control Eng. Pract.*, vol. 14, no. 7, pp. 757–767, 2006.

[44] M. Papageorgiou, H. Hadj-Salem, and F. Middelham, "ALINEA local ramp metering—Summary of field results," *Transp. Res. Rec.*, no. 1603, pp. 90–98, 1998.

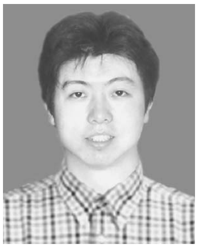
[45] A. Kotsialos, M. Papageorgiou, and A. Messmer, "Optimal coordinated and integrated control motorway network traffic control," in *Proc. 14th Int. Symp. Transp. Traffic Theory*, Jerusalem, Israel, 1999, pp. 621–644.

[46] iTRANS Consulting Inc., "Costs of non-recurrent congestion in Canada," Ottawa, ON, Canada, TP 14664E, 2006.

Dongbin Zhao (M'06–SM'10) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in August 1994, August 1996, and April 2000, respectively, all in materials processing engineering.

He was a Postdoctoral Fellow of humanoid robotics with the Department of Mechanical Engineering, Tsinghua University, Beijing, China, from May 2000 to January 2002. He is currently an Associate Professor with the Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing. He has published one book and over 30 international journal papers. His current research interest lies in the area of computational intelligence, adaptive dynamic programming, robotics, intelligent transportation systems, and process simulation.

Dr. Zhao has been a Senior Member of the Chinese Mechanical Engineering Society since 2006. He served as a member of many international program committees and is the Financial Chair of the 2011 International Symposium on Neural Networks. He received the Second Award for Scientific Progress of National Defense from the Commission of Science Technology and Industry from the National Defense of China in 1999, the First Award for Scientific Progress of Chinese Universities from the Ministry of Education of China in 2001, and the Third Award for Scientific and Technology Progress from the China Petroleum and Chemical Industry Association in 2009.



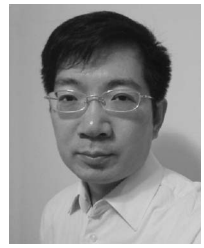
Xuerui Bai received the Ph.D. degree in control theory and control engineering from the Chinese Academy of Sciences (CAS), Beijing, China, in 2010.

Since 2010, he has been an Assistant Researcher with the Key Laboratory of Complex Systems and Intelligence Science, CAS. His research interests include computational intelligence and intelligent transportation.

Fei-Yue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1990.

He joined the University of Arizona, Tucson, in 1990 and became a Professor and the Director of the Robotics Laboratory and the Program in Advanced Research for Complex Systems. In 1999, he founded the Intelligent Control and Systems Engineering Center, Chinese Academy of Sciences (CAS), Beijing, China, with the support of the Outstanding Oversea Chinese Talents Program. Since 2002, he has been the Director of the Key Laboratory on Complex Systems and Intelligence Science, CAS. From 1995 to 2000, he was the Editor-in-Chief of the *International Journal of Intelligent Control and Systems* and the World Scientific Series on *Intelligent Control and Intelligent Automation*. His research interests include social computing, web science, and intelligent control.

Dr. Wang is a member of Sigma Xi; an elected Fellow of the International Council on Systems Engineering, the International Federation of Automatic Control, the American Society of Mechanical Engineers, and the American Association for the Advancement of Science; and Vice President and Secretary-General of the Chinese Association of Automation. He has served as Chair of more than 20 IEEE, Association for Computing Machinery (ACM), and Institute for Operations Research and Management Sciences conferences. He was the President of the IEEE Intelligent Transportation Systems Society from 2005 to 2007, the Chinese Association for Science and Technology, USA, in 2005, and the American Zhu Kezhen Education Foundation from 2007 to 2008. He is currently the Editor-in-Chief of IEEE INTELLIGENT SYSTEMS and the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He was the recipient of the National Prize in Natural Sciences of China in 2007 and the Outstanding Scientist Award from the ACM for his work in intelligent control and social computing.



Jing Xu received the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007.

Since 2007, he has been a Staff Researcher with the IBM China Development Lab, Xi'an, China. He has broad research interests, including control theory and control engineering, adaptive dynamic programming, neural optimization, intelligent transportation systems, statistical analysis, data mining, predictive analytics, social network analysis, etc.

Wensheng Yu received the Ph.D. degree in dynamics and control from Peking University, Beijing, China, in 1998.

He was with the Institute of Automation, Chinese Academy of Sciences, Beijing, where he was promoted from Assistant Researcher to Associate Professor in 1999 and to Professor in 2004. He was a Visiting Research Fellow with the University of Melbourne, Melbourne, Australia, from 2001 to 2002. Since 2009, he has been a Professor with Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China. He is the author or coauthor of numerous articles in journals and conference proceedings, which span the fundamental fields of control theory and control engineering and applied mathematics. His current research interests include robust and optimal control, adaptive filter theory and linear estimation, control theory and control engineering, robotics and control, complex systems and control, signal processing and communications, algorithm research for robust control, mechanization for control theory, etc.