EFFICIENT LINEAR APPROXIMATIONS TO STOCHASTIC VEHICULAR COLLISION-AVOIDANCE PROBLEMS

Dmitri Dolgov
Toyota Technical Center, USA, Inc.
2350 Green Rd., Suite 101, Ann Arbor, MI 48105, USA

Ken Laberteaux
Toyota Technical Center, USA, Inc.
2350 Green Rd., Suite 101, Ann Arbor, MI 48105, USA

Keywords: Decision support systems, Vehicle control applications, Optimization algorithms.

Abstract: The key components of an intelligent vehicular collision-avoidance system are sensing, evaluation, and decision making. We focus on the latter task of finding (approximately) optimal collision-avoidance control policies, a problem naturally modeled as a Markov decision process. However, standard MDP models scale exponentially with the number of state features, rendering them inept for large-scale domains. To address this, factored MDP representations and approximation methods have been proposed. We approximate collision-avoidance factored MDP using a composite approximate linear programming approach that symmetrically approximates objective functions and feasible regions of the LP. We show empirically that, combined with a novel basis-selection method, this produces high-quality approximations at very low computational cost.

1 INTRODUCTION

Vehicular collisions are a leading cause of death and injury in many countries around the world: in the United States alone, on an average day, auto accidents kill 116 and injure over 7900, with an annual economic impact of around $200 billion (NHTSA, 2003); the situation in the European Union is similar with over 100 deaths and 4600 injuries daily, and the annual cost of €160 billion (CARE, 2004). Governments and automotive companies are responding by making the reduction of vehicular fatalities a top priority (e.g., (ITS, 2003; Toyota, 2004)).

Key to reducing auto collisions is improving drivers' recognition and response behaviors, technology often described as an Intelligent Driver Assistant (e.g., (Batavia, 1999)). This driver assistant would direct a driver's attention to a safety risk and potentially advise the driver of appropriate counter-measures. Such a system would require new sensing, evaluation, and decision-making technologies.

This work focuses on the latter task of constructing approximately optimal collision-avoidance policies. We represent the stochastic collision-avoidance problem as a Markov decision process (MDP) – a well-studied, simple, and elegant model of stochastic sequential decision-making problems (Puterman, 1994). Unfortunately, classical MDP models scale very poorly, as the size of the *flat* state space increases exponentially with the number of environment features (e.g., number of vehicles) – the effect commonly referred to as the *curse of dimensionality*.

Fortunately, many problems are well-structured and admit compact, *factored* MDP representations (Boutilier et al., 1995), leading to drastic reductions in problem size. However, a challenge in solving factored MDPs is that well-structured problems do not always lead to well-structured solutions (Koller and Parr, 1999), making approximations a necessity. One such technique that has recently proved successful in many domains is *Approximate Linear Programming* (ALP) (Schweitzer and Seidmann, 1985; de Farias and Roy, 2003; Guestrin et al., 2003).

We show that vehicular collision-avoidance domains can be modeled compactly as factored MDPs, and further, that ALP techniques can be successfully applied to such problems, yielding very high-quality results at low computational cost (attaining exponential speedup over flat MDP models). We use the *composite ALP* formulation of (Dolgov and Durfee, 2005), which approximates both the primal and the dual variables of LP formulations of MDPs, thus symmetrically approximating their objective functions and feasible regions. ALP methods are extremely sensitive to the selection of *basis functions* and the specifics of the approximation of the feasible region, with only greedy and domain-dependent basis-

Dolgov D. and Laberteaux K. (2005).
EFFICIENT LINEAR APPROXIMATIONS TO STOCHASTIC VEHICULAR COLLISION-AVOIDANCE PROBLEMS.
In *Proceedings of the Second International Conference on Informatics in Control, Automation and Robotics*, pages 275-278

selection methods currently available (Patrascu et al., 2002; Poupart et al., 2002). The second contribution of this work is a method for automatically constructing basis functions, which, as demonstrated by our empirical evaluation, works very well for collision-avoidance problems (the idea also extends naturally to other domains that are similarly well-structured).

2 FACTORED MDPS AND ALP

We model the collision-avoidance problem as a stationary, discrete-time, fully-observable, discounted MDP (Puterman, 1994), which can be defined as $\langle \mathcal{S}, \mathcal{A}, P, R, \gamma \rangle$, where: $\mathcal{S} = \{i\}$ and $\mathcal{A} = \{a\}$ are finite sets of states and actions, $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \mapsto [0, 1]$ is the transition function (P_{iaj} is the probability of transitioning to state j upon executing action a in state i), and $R : \mathcal{S} \times \mathcal{A} \mapsto [R_{\min}, R_{\max}]$ defines the bounded reward function (R_{ia} is the reward for executing action a in state i), and γ is the discount factor.

An optimal solution to such an MDP is a stationary, deterministic policy $\pi : \mathcal{S} \times \mathcal{A} \mapsto [0, 1]$, and the key to obtaining it is to compute the optimal value function $v : \mathcal{S} \mapsto \mathbb{R}$, which specifies, for every state, the expected total reward of starting in that state and acting optimally thereafter. The value function can be computed, for example, using the following *primal* linear program of an MDP (Puterman, 1994):

$$\min \sum_i \alpha_i v_i \;\Big|\; v_i \geq R_{ia} + \gamma \sum_j P_{iaj} v_j, \quad (1)$$

where $\alpha_i > 0$ are arbitrary constants. It is often useful to consider the equivalent *dual LP*:

$$\max \sum_{i,a} R_{ia} x_{ia} \Big| \sum_a x_{ja} - \gamma \sum_{i,a} x_{ia} P_{iaj} = \alpha_j \quad (2)$$

where $x \geq 0$ are called the *occupation measure* (x_{ia} is the expected discounted number executions of action a in state i). Thus, the constraints in (eq. 2) ensure the conservation of flow through each state.

A weakness of such MDPs is that they require an explicit enumeration of all system states. To address this issue, factored MDPs have been proposed (Boutilier et al., 1995; Koller and Parr, 1999) that define the transition and reward functions on *state features* $z \in \mathcal{Z}$. The transition function is specified as a dynamic Bayesian network, with the current state features viewed as the parents of the next step features:

$$P_{iaj} = P(\mathbf{z}(j)|\mathbf{z}(i), a) = \prod_{n=1}^{N} p_n(z_n(j)|a, \mathbf{z}_{p_n}(i)),$$

The reward function for a factored MDP is compactly defined as $R_{ia} = \sum_{m=1}^{M} r_m(\mathbf{z}_{r_m}(i), a)$.

Approximate linear programming (Schweitzer and Seidmann, 1985; de Farias and Roy, 2003) lowers the dimensionality of the objective function of the primal LP (eq. 1) by restricting the space of value functions to a linear combination of predefined basis functions:

$$v_i = v(\mathbf{z}(i)) = \sum_{k=1}^{K} h_k(\mathbf{z}_{h_k}(i)) w_k, \quad (3)$$

where $h_k(\mathbf{z}_{h_k})$ is the k^{th} basis function defined on a small subset of the state features $\mathcal{Z}_{h_k} \subset \mathcal{Z}$, and w are the new optimization variables. Such a reduction is only beneficial if each basis function h_k depends on a small number of state features. Using the above substitution, LP (eq. 1) can be approximated as follows:

$$\min \alpha^T H w \;\Big|\; AHw \geq r, \quad (4)$$

where we introduce $A_{ia,j} = \delta_{ij} - \gamma P_{iaj}$ (δ_{ij} is the Kronecker delta $\delta_{ij} = 1 \Leftrightarrow i = j$).

This approximation reduces the number of optimization variables from $|\mathcal{S}|$ to $|w|$, but the number of constraints remains exponential at $|\mathcal{S}||\mathcal{A}|$. There are several ways of addressing this. One way is to sample the constraint set (de Farias and Roy, 2004), which works, intuitively, because once the number of optimization variables is reduced as in (eq. 4), only a small number of constraints remain active. This method is very sensitive to the distribution over which the constraint set is sampled, i.e., a poor choice of a subset of the constraints could significantly impair the effectiveness of this method. Another method (which does not further approximate the solution, beyond (eq. 3)) is to restructure the constraints using the principles of non-serial dynamic programming (Guestrin et al., 2003). Unfortunately, the worst-case complexity of this approach still grows exponentially with the number of state features.

We use another approach, proposed in (Dolgov and Durfee, 2005), which applies linear approximations to both the primal (v) and the dual (x) coordinates, effectively approximating both the objective function and the feasible region of the LP. Let us consider the dual of the ALP (eq. 4) and apply a linear approximation of the dual coordinates: $x = Qy$ to the result, yielding the following *composite* ALP:

$$\max r^T Q y \Big| H^T A^T Q y = H^T \alpha, \quad Qy \geq 0. \quad (5)$$

The constraint $Qy \geq 0$ in this composite ALP still has exponentially many ($|\mathcal{S}||\mathcal{A}|$) rows, but this can be resolved in several ways. For example, $Qy \geq 0$ can be reformulated as a more compact set using the ideas of (Guestrin et al., 2003), but the resulting constraint set still scales exponentially in the worst case. Another way of handling this is to restrict Q to be non-negative and replace the constraints with a stricter condition $y \geq 0$ (introducing yet another source of approximation error), leading to the following LP:

$$\max r^T Q y \Big| H^T A^T Q y = H^T \alpha, \quad y \geq 0, \quad (6)$$

or, equivalently, its dual:

$$\min \alpha^T H w \Big| Q^T A H w \geq Q^T r. \quad (7)$$

Recall that the quality of the primal ALP (eq. 4) is very sensitive to the choice of the primal basis H. Similarly, the quality of policies produced by the composite ALPs ((eq. 6) and (eq. 7)) greatly depends on the choice of both H and Q. However, as we empirically show below, the approach lends itself to an intuitive algorithm for constructing small and compact basis sets H and Q that yield high quality solutions for the collision-avoidance domain.

Finally, let us also note that, while feasibility of the primal ALP (eq. 4) can be ensured by simply adding a constant $h_0 = 1$ to the basis H (de Farias and Roy, 2003), it is slightly more difficult to ensure the feasibility of the composite ALP (eq. 6) (or the boundedness of (eq. 7)). Let us note that in practice, for any primal basis H, boundedness and feasibility of the composite ALPs can be ensured by constructing a sufficiently large dual basis Q.

3 COLLISION-AVOIDANCE MDP MODEL

We conducted experiments on several two-dimensional collision-avoidance scenarios, and the high-level results were consistent across the domains. To ground the discussion, we report our findings for a simplified model of the task of driving on a two-way street. We model the problem as a discrete-state MDP, by using a grid-world representation for the road, with the x-y positions of all cars as the state features (the flat state space is given by their cross-product).

In this domain, we are controlling one of the cars, and the goal is to find a policy that minimizes the aggregate probability of collisions with other cars. Each uncontrolled vehicle is modeled to strictly adhere to the right-hand-side driving convention. Within these bounds, the vehicles stochastically change lanes while drifting with varying speed in the direction of traffic.

This model can be naturally represented as a factored MDP. Indeed, the reward function lends itself to a factored representation, because we only penalize collisions with other cars, so the total reward can be represented as a sum of local reward functions, each one a function of the relative positions of the controlled car and one of the uncontrolled cars.[1] The transition function of the MDP also factors well, because each car moves mostly independently, so the factored transition function can be represented as a Bayesian network with each node depending on a small number of world features.

[1]We also experimented with other more interesting domains and reward functions (e.g., roads with shoulders where moving on a shoulder gave a small penalty); the high-level results were consistent across such modifications.

4 BASIS SELECTION AND EVALUATION

As mentioned earlier, ALP is very sensitive to the choice of basis functions H and Q. Therefore, our main goal is to design procedures for constructing primal (H) and dual (Q) basis sets that are compact, but at the same time yield high-quality control policies.

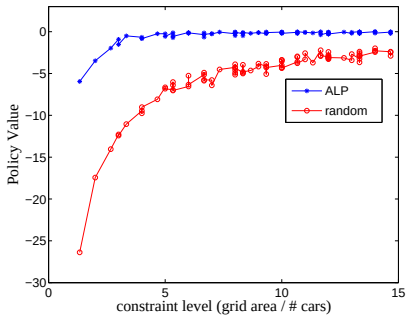
The basic domain-independent idea behind our algorithm is to use solutions to smaller MDPs as basis functions for larger problems. For our collision-avoidance domains, we implemented this idea as follows. For every pair of objects, we constructed an MDP with the original topology but without any other objects, and then used these optimal value functions as the primal basis H and the optimal occupation measures as the dual basis Q for the original MDP.

We empirically evaluated this method on the car domain from Section 3.[2] In our experiments, we varied the geometry of the grid and the number of cars, and for each configuration, we solved the corresponding factored MDP using the ALP method described above, and evaluated the resulting policies using a Monte Carlo simulation (an exact evaluation is infeasible, due to the curse of dimensionality).

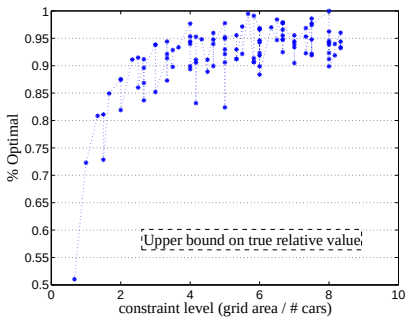
Figure 1a shows the value of the approximate policies computed in this manner, as a function of how highly constrained the problem is (the ratio of the grid area to the number of cars), with the average values of random policies shown for comparison. The important question is, of course, how close our solution is to the optimum. Unfortunately, for all but the most trivial domains, computing the optimal solution is infeasible, so we cannot directly answer that question. However, for our collision-avoidance domains, where only negative rewards are obtained in collision states, we can upper-bound the value of any policy by zero. Using this upper bound on the quality of the optimal solution, we can compute a lower bound on the relative quality of our approximation, which is shown in Figure 1b. Notice that, for highly constrained problems (where optimal solutions have large negative values), this lower bound can greatly underestimate the quality of our solution, which explains low numbers in the left part of the graph. However, even given this pessimistic view, our ALP method produced policies that were, on the average, no worse than 92% of the optimum (relative to the optimal-random gap).

We also evaluated our approximate solution by its relative gain in efficiency. In our experiments, the sizes of the primal and dual basis sets grow quadratically with the number of cars, while the size of the exact LP (eq. 1) grows exponentially. Table 1 illustrates the complexity reduction achieved by using the composite ALP approach. In fact, the difference in

[2]Other collision-avoidance domains had similar results.



(a)



(b)

Figure 1: Absolute value (a) and lower bound on relative value (b) of ALP solutions. The lower-bound estimate of the ALP quality of ALP policies is, on average, 92% of optimum (relative to random policies).

Table 1: Problem size of exact LP (eq. 1) and composite ALP (eq. 7); former scales exponentially and the latter quadratically with the number of cars.

# cars	3	4	5	6	7	8	9
ALP	3660	4960	6300	7680	9100	10560	12060
exact	4e+09	6e+12	1e+16	1e+19	2e+22	4e+25	6e+28

complexity between the flat LP and the ALP is so significant that, the bottleneck was not the ALP itself, but the smaller 2-car MDPs which were solved for the exact solution to obtain the basis functions. Thus, an interesting direction of future work is to experiment with approximate solution techniques for the small MDPs in the basis-generation phase.

5 CONCLUSIONS

We have analyzed the sequential decision-making problem of vehicular collision avoidance in a stochastic environment, modeled as a Markov decision process. Although classical MDP representations and solution techniques are not feasible for realistic domains, we have empirically demonstrated that collision-avoidance problems can be represented compactly as factored MDPs and, moreover, that they admit high-quality ALP solutions.

The core of our algorithm, the composite ALP (eq. 7), relies on two basis sets – the primal basis H and the dual basis Q. We have presented a simple procedure for constructing these basis sets, where optimal solutions to scaled-down problems are used as basis functions. This method attains an exponential reduction in problem complexity (Table 1), while producing policies that were very close to optimal (above 90% of the random-optimal gap, according to the pessimistic estimate of Figure 1b). Moreover, we believe that this general basis-selection methodology is more widely applicable and can be fruitfully used in other domains that are similarly well-structured. An analysis of this methodology for other problems is a direction of our ongoing and future work.

REFERENCES

Batavia, P. (1999). *Driver Adaptive Lane Departure Warning Systems*. PhD thesis, CMU.

Boutilier, C., Dearden, R., and Goldszmidt, M. (1995). Exploiting structure in policy construction. In *IJCAI-95*.

CARE (2004). Community road accident database. http://europa.eu.int/comm/transport/care/.

de Farias, D. and Roy, B. V. (2004). On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. of OR*, 29(3):462–478.

de Farias, D. P. and Roy, B. V. (2003). The linear programming approach to approximate dynamic programming. *OR*, 51(6).

Dolgov, D. A. and Durfee, E. H. (2005). Towards exploiting duality in approximate linear programming for MDPs. In *AAAI-05*. To appear.

Guestrin, C., Koller, D., Parr, R., and Venkataraman, S. (2003). Efficient solution algorithms for factored MDPs. *JAIR*, 19:399–468.

ITS (2003). US Department of Transportation press release. http://www.its.dot.gov/press/fhw2003.htm.

Koller, D. and Parr, R. (1999). Computing factored value functions for policies in structured MDPs. In *IJCAI*.

NHTSA (2003). Traffic safety facts. Report DOT HS 809 767, http://www-nrd.nhtsa.dot.gov.

Patrascu, R., Poupart, P., Schuurmans, D., Boutilier, C., and Guestrin, C. (2002). Greedy linear value-approximation for factored Markov decision processes. In *AAAI-02*, pages 285–291.

Poupart, P., Boutilier, C., Patrascu, R., and Schuurmans, D. (2002). Piecewise linear value function approximation for factored MDPs. In *AAAI-02*, pages 292–299.

Puterman, M. L. (1994). *Markov Decision Processes*. John Wiley & Sons, New York.

Schweitzer, P. and Seidmann, A. (1985). Generalized polynomial approximations in Markovian decision processes. *J. of Math. Analysis and App.*, 110:568 582.

Toyota (2004). Toyota safety: Toward realizing zero fatalities and accidents. http://www.toyota.co.jp/en/safety_presen/index.html.