

Sensitivity Analysis of a Dynamic Fleet Management Model Using Approximate Dynamic Programming

Huseyin Topaloglu

School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853,
topaloglu@orie.cornell.edu

Warren B. Powell

Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey 08544,
powell@princeton.edu

We present tractable algorithms to assess the sensitivity of a stochastic dynamic fleet management model to fleet size and load availability. In particular, we show how to compute the change in the objective function value in response to an additional vehicle or an additional load introduced into the system. The novel aspect of our approach is that it does not require multiple simulations with different values of the model parameters, and in this respect it differs from trial-and-error-based “what-if” analyses. Numerical experiments show that the proposed methods are accurate and computationally attractive.

Subject classifications: transportation: models, network; dynamic programming/optimal control: applications.

Area of review: Optimization.

History: Received February 2004; revisions received May 2005, February 2006; accepted February 2006.

Introduction

Given a vehicle fleet and a stochastic process characterizing the load arrivals in a transportation network, the primary objective of fleet management models is to make the vehicle repositioning and vehicle-to-load assignment decisions so that some performance measure (profit, cost, deadhead miles, number of served loads, etc.) is optimized. However, besides making these vehicle repositioning and assignment decisions, an important question that is commonly overlooked by many fleet management models is how the performance measures would change in response to a change in certain model parameters. For example, freight carriers are interested in how much their profits would increase if they introduced an additional vehicle into the system or if they served an additional load on a certain traffic lane. Railroad companies want to estimate the minimum number of railcars that is necessary to cover the random shipper demands. The Airlift Mobility Command is interested in the impact of limited airbase capacities on the delayed shipments. Answering such questions, in one way or another, requires sensitivity analysis of the underlying fleet management model responsible for making the vehicle allocation decisions.

In this paper, we develop efficient sensitivity analysis methods for a stochastic fleet management model previously developed in Godfrey and Powell (2002a, b). This model formulates the problem as a dynamic program, decomposing it into time-staged subproblems, and replaces

the value functions with specially structured approximations that are obtained through an iterative improvement scheme. Two aspects of this model are crucial to our work: (1) Due to the special structure of the value function approximations, the subproblem that needs to be solved for each time period is a min-cost network flow problem. This enables us to use the well-known relationships between the sensitivity analyses of min-cost network flow problems and min-cost flow augmenting trees. In particular, we can use the fact that the change in the optimal solution of a min-cost network flow problem in response to a unit change in the supply of a node or a unit change in the upper bound of an arc is characterized by a min-cost flow augmenting path. (2) Letting \mathcal{T} be the set of time periods in the planning horizon, just as the value functions $\{V_t(\cdot) : t \in \mathcal{T}\}$ describe an optimal vehicle allocation policy through the so-called optimality equation (see Puterman 1994), a set of value function approximations $\{\tilde{V}_t^\pi(\cdot) : t \in \mathcal{T}\}$ describes a (possibly suboptimal) vehicle allocation policy π . Thus, given a trajectory of load realizations $\{d_t : t \in \mathcal{T}\}$, one can think of the trajectory of vehicle allocation decisions $\{x_t^{\pi d} : t \in \mathcal{T}\}$ induced by policy π under load realization trajectory $d = \{d_t : t \in \mathcal{T}\}$. In this paper, we exploit the aforementioned sensitivity relationships to compute the change in the decision trajectory $\{x_t^{\pi d} : t \in \mathcal{T}\}$ in response to a change in a problem parameter and to assess how changes in the current decisions affect the future time periods. In particular, we develop methods to compute how much the profits would

change if an additional vehicle or an additional load were introduced into the system.

Our work is motivated by the fact that the freight transportation industry is interested in “what-if” scenarios, and this conventionally refers to changing certain parameters of the model and re-running it. However, there are obvious advantages associated with being able to extract sensitivity information from a single model run. For example, there can be many parameters whose impact on the model performance is of interest, and making one (or more) run for each parameter might be impossible. Furthermore, the decision maker simply might not have an idea about the critical parameters, and it is important to point out where “the biggest bang for the buck” lies. Also, sensitivity information is useful in determining the optimal fleet size and mix, or making pricing decisions.

A well-known class of models, from which one can quickly obtain sensitivity information, formulates the problem over a *state-time network*, where the nodes represent the supply of vehicles in different states at different time periods and the arcs represent the vehicle movements. Examples of such models from three different industries are Dantzig and Fulkerson (1954), Hane et al. (1995), and Holmberg et al. (1998), and we refer the reader to Dejax and Crainic (1987) and Powell et al. (1995) for detailed surveys. For these types of models, sensitivity information is readily obtained by using the dual solution. However, these models are inherently deterministic and can incorporate the random future load arrivals only through the expected values.

The model we analyze in this paper falls into the category of stochastic models, which decompose the problem with respect to time periods and assess the impact of the current decisions on the future through value functions. However, because practical fleet management models involve large numbers of decision variables and possible load realizations, standard stochastic optimization methods are not feasible for computing the value functions. Therefore, most of the stochastic fleet management models revolve around the idea of approximating the value function in a tractable manner. For stochastic fleet management models other than the one in Godfrey and Powell (2002a, b), we refer the reader to Frantzeskakis and Powell (1990), Crainic et al. (1993), Carvalho and Powell (2000), and Topaloglu and Powell (2006).

Given a set of value function approximations, these models behave just like simulation models, generating different trajectories of vehicle allocation decisions for different trajectories of load realizations. Therefore, a key question for their sensitivity analysis is to be able to assess how the decision trajectories change when certain model parameters are perturbed. Our approach has similarities with infinitesimal perturbation analysis, which refers to computing the gradient of a performance measure in a discrete-event dynamic system with respect to an input parameter (see Glasserman 1991, Ho and Cao 1991). However, for the

systems we consider, the transitions between the states are more complex than those that are conventionally considered by discrete-event dynamic systems because these transitions are governed by the solutions of min-cost network flow problems.

In this paper, we make the following research contributions. We develop efficient sensitivity analysis methods for a stochastic fleet management model previously developed in Godfrey and Powell (2002a, b). In particular, we show how to compute the change in the objective value of this model in response to an additional vehicle or an additional load introduced into the system. An accurate, but tedious, sensitivity analysis method is to physically change a parameter of interest and rerun the model. We compare our methods with this “brute force” approach and show that they are quite accurate even when theoretical analysis requires certain approximations.

The rest of this paper is organized as follows. In §1, we briefly describe the fleet management model used throughout the paper. Understanding this model is important because we analyze the sensitivity of the objective value under the policy prescribed by this particular model. In §2, we consider problems with a single vehicle type and show how to compute the change in the objective value in response to an additional vehicle or an additional load introduced into the system. Section 3 extends these results to problems with multiple vehicle types. The computational experiments presented in §4 show the accuracy of the proposed sensitivity analysis methods. Section 5 contains concluding remarks.

1. Problem Formulation

We have a fleet of vehicles to serve the loads of different types that arrive over time. At every time period, a certain number of loads enter the system, and we have to decide which loads to cover and to which locations we should reposition the empty vehicles. We are interested in maximizing the total expected profit over a finite horizon, but we formulate our model to minimize cost for compatibility with the min-cost network flow literature. We assume that advance information about the future loads is not available and the loads that cannot be covered in a given time period are served by an emergency subcontractor. These enable us to assume that the uncovered loads immediately leave the system. For notational brevity, we assume that it takes one time period to move between any pair of locations. It is straightforward to extend our analysis to the case where there are multiperiod travel times by using the approach described in Topaloglu and Powell (2006). We define the following:

\mathcal{T} = set of time periods in the planning horizon, $\mathcal{T} = \{1, \dots, T\}$.

\mathcal{F} = set of locations in the transportation network.

\mathcal{K} = set of available vehicle types.

\mathcal{L} = set of movement modes, $\mathcal{L} = \{0, \dots, L\}$. (Movement modes represent different ways in which a vehicle

can move from one location to another. Movement mode 0 always corresponds to empty repositioning, whereas other modes correspond to carrying different types of loads.)

x_{ijlt}^k = number of vehicles of type k dispatched from location i to j at time period t using movement mode l .

c_{ijlt}^k = cost of dispatching one vehicle of type k from location i to j at time period t using movement mode l .

D_{ijlt} = random variable representing the number of loads that need to be carried from location i to j at time period t and correspond to movement mode l .

As will be clear shortly, the random variable D_{ijlt} serves as an upper bound on the decision variables $\{x_{ijlt}^k: k \in \mathcal{K}\}$. Because the movement mode 0 corresponds to empty repositioning and the empty repositioning movements are not bounded, we assume that $D_{ij0t} = \infty$ for all $i, j \in \mathcal{J}, t \in \mathcal{T}$. In practice, the movement modes in $\mathcal{L} \setminus \{0\}$ can correspond to different types of loads or different shippers, and we usually have $c_{ijlt}^k < 0$ when $l \in \mathcal{L} \setminus \{0\}$. The vehicle type might reflect the size of the vehicle, the skill level of the driver of the vehicle, the ability of the vehicle to satisfy certain safety or sanitary requirements, or a combination of these factors, which ultimately determine whether it is feasible to cover a certain type of load with a certain type of vehicle and what profit is obtained by doing so. If it is infeasible to cover a load of type l with a vehicle of type k , then we capture this by letting $c_{ijlt}^k = \infty$ for all $i, j \in \mathcal{J}, t \in \mathcal{T}$. Throughout this paper, we use d_{ijlt} to denote a particular realization of D_{ijlt} . By suppressing some of the indices in the variables above, we denote a vector composed of the components ranging over the suppressed indices. For example, we have $d_t = \{d_{ijlt}: i, j \in \mathcal{J}, l \in \mathcal{L}\}$ and $d = \{d_{ijlt}: i, j \in \mathcal{J}, l \in \mathcal{L}, t \in \mathcal{T}\}$.

To capture the state of the system at time period t , we define

r_{it}^k = number of vehicles of type k that are available at location i at time period t .

The vector $r_t = \{r_{it}^k: i \in \mathcal{J}, k \in \mathcal{K}\}$ completely defines the state of the vehicles at time period t . Given this state vector and the realization of the loads at time period t , the set of feasible decision vectors and the set of state vectors generated by these decisions at the next time period are given by

$$\mathcal{Y}(r_t, d_t) = \left\{ (x_t, r_{t+1}): \sum_{j \in \mathcal{J}} \sum_{l \in \mathcal{L}} x_{ijlt}^k = r_{it}^k \right. \\ \left. \text{for all } i \in \mathcal{J}, k \in \mathcal{K}, \quad (1) \right.$$

$$\left. \sum_{i \in \mathcal{J}} \sum_{l \in \mathcal{L}} x_{ijlt}^k - r_{j,t+1}^k = 0 \right. \\ \left. \text{for all } j \in \mathcal{J}, k \in \mathcal{K}, \quad (2) \right.$$

$$\left. \sum_{k \in \mathcal{K}} x_{ijlt}^k \leq d_{ijlt} \right. \\ \left. \text{for all } i, j \in \mathcal{J}, l \in \mathcal{L}, \quad (3) \right.$$

$$\left. \left. x_{ijlt}^k \in \mathbb{Z}_+ \right. \right. \\ \left. \left. \text{for all } i, j \in \mathcal{J}, l \in \mathcal{L}, k \in \mathcal{K} \right\}. \quad (4)$$

We are interested in Markovian deterministic policies that minimize the total expected cost over the planning horizon. A Markovian deterministic policy π can be characterized by a sequence of decision functions $\{X_t^\pi(\cdot, \cdot): t \in \mathcal{T}\}$ such that $X_t^\pi(\cdot, \cdot)$ maps the state vector r_t and the realization of the loads d_t at time period t to a decision vector x_t . One can also define the state transition functions $\{R_{t+1}^\pi(\cdot, \cdot): t \in \mathcal{T}\}$ of policy π such that $R_{t+1}^\pi(\cdot, \cdot)$ maps the state vector and the realization of the loads at time period t to a state vector for the next time period. We note that given $X_t^\pi(\cdot, \cdot)$, $R_{t+1}^\pi(\cdot, \cdot)$ can easily be defined by noting the state transition constraints in (2). Then, for a given state vector r_t and realization of future loads $\{d_t, \dots, d_T\}$ at time period t , the cumulative cost function for policy π can be written recursively as

$$F_t^\pi(r_t, d_t, d_{t+1}, \dots, d_T) \\ = c_t \cdot X_t^\pi(r_t, d_t) + F_{t+1}^\pi(R_{t+1}^\pi(r_t, d_t), d_{t+1}, d_{t+2}, \dots, d_T), \quad (5)$$

with the boundary condition $F_{T+1}^\pi(\cdot) = 0$. By repeated application of (5), we obtain

$$F_1^\pi(r_1, d_1, \dots, d_T) \\ = c_1 \cdot X_1^\pi(r_1, d_1) + c_2 \cdot X_2^\pi(R_2^\pi(r_1, d_1), d_2) \\ + \dots + c_T \cdot X_T^\pi(R_T^\pi(R_{T-1}^\pi(\dots, d_{T-2}), d_{T-1}), d_T), \quad (6)$$

which is the total cost incurred over the whole planning horizon when we use policy π ; the initial state vector is r_1 and the realization of the loads is $\{d_1, \dots, d_T\}$.

Assuming that, given r_t , D_t is independent of $\{D_{t'}: t' = 1, \dots, t-1\}$, it can be shown that the optimal policy π^* is Markovian deterministic, satisfying

$$\pi^* = \arg \min_{\pi \in \Pi} \mathbb{E}\{F_1^\pi(r_1, D_1, \dots, D_T) \mid r_1\},$$

where Π is the set of Markovian deterministic policies. This optimal policy can be found by computing the value functions through the so-called optimality equation (see Puterman 1994):

$$V_t^{\pi^*}(r_t) = \mathbb{E} \left\{ \min_{(x_t, r_{t+1}) \in \mathcal{Y}(r_t, D_t)} c_t \cdot x_t + V_{t+1}^{\pi^*}(r_{t+1}) \mid r_t \right\}. \quad (7)$$

In this case, the decision and transition functions for the optimal policy become

$$(X_t^{\pi^*}(r_t, d_t), R_{t+1}^{\pi^*}(r_t, d_t)) \\ = \arg \min_{(x_t, r_{t+1}) \in \mathcal{Y}(r_t, d_t)} c_t \cdot x_t + V_{t+1}^{\pi^*}(r_{t+1}). \quad (8)$$

Throughout this paper, to keep the presentation simple, we assume that the cost vector c_t is perturbed by small random amounts so that problem (8) has a single optimal solution. Under this assumption, the decision and state transition functions are properly defined, and our proofs become easier.

Computing the value functions $\{V_t^{\pi^*}(\cdot) : t \in \mathcal{T}\}$ through (7) is intractable for almost all problem instances of practical significance, because it requires enumerating over all possible values of r_t and taking an expectation over the multidimensional random variable D_t for all $t \in \mathcal{T}$. In this paper, we follow a class of (suboptimal) policies proposed in Godfrey and Powell (2002a), which are obtained by replacing $\{V_t^{\pi^*}(\cdot) : t \in \mathcal{T}\}$ in (8) with separable approximations $\{\tilde{V}_t^{\pi^*}(\cdot) : t \in \mathcal{T}\}$ of the form

$$\tilde{V}_t^{\pi}(r_t) = \sum_{i \in \mathcal{J}} \sum_{k \in \mathcal{K}} \tilde{V}_{it}^{\pi k}(r_{it}^k), \tag{9}$$

where each $\tilde{V}_{it}^{\pi k}(\cdot)$ is a one-dimensional piecewise-linear convex function with points of nondifferentiability being a subset of positive integers. In this case, for a policy π characterized by separable piecewise-linear convex value function approximations $\{\tilde{V}_t^{\pi}(\cdot) : t \in \mathcal{T}\}$, we can define the decision and state transition functions as

$$\begin{aligned} (X_t^{\pi}(r_t, d_t), R_{t+1}^{\pi}(r_t, d_t)) \\ = \arg \max_{(x_t, r_{t+1}) \in \mathcal{Y}(r_t, d_t)} c_t \cdot x_t + \tilde{V}_{t+1}^{\pi}(r_{t+1}). \end{aligned} \tag{10}$$

We note that although the value function approximations $\{\tilde{V}_t^{\pi}(\cdot) : t \in \mathcal{T}\}$ are separable, the cumulative cost functions $\{F_t^{\pi}(\cdot, d_t, \dots, d_T) : t \in \mathcal{T}\}$ for policy π are not necessarily separable. Furthermore, we have $V_t^{\pi^*}(r_t) = \mathbb{E}\{F_t^{\pi^*}(r_t, D_t, \dots, D_T) \mid r_t\}$ for the optimal policy π^* by the principal of optimality (see Puterman 1994), but we do not necessarily have $\tilde{V}_t^{\pi}(r_t) = \mathbb{E}\{F_t^{\pi}(r_t, D_t, \dots, D_T) \mid r_t\}$ for the policy π characterized by the value function approximations $\{\tilde{V}_t^{\pi}(\cdot) : t \in \mathcal{T}\}$.

Godfrey and Powell (2002a) give a sampling-based algorithm that can be used to obtain a “good” set of value function approximations. The question of whether these suboptimal policies yield high-quality solutions is outside the scope of this paper. We refer the reader to Godfrey and Powell (2002a, b) and Topaloglu and Powell (2006), where the experimental work indicates that this class of policies beat standard benchmarks by significant margins. Here, we assume that we already have a “good” policy π , and we are interested in computing the change in $F_1^{\pi}(r_1, d_1, \dots, d_T)$ induced by changing an element of the state vector r_1 or the load availability vector d_1 . We make this question precise in the next two sections. However, before going into the specific details, we can summarize the contents of the next two sections as follows: (1) We note that if we use a policy characterized by separable piecewise-linear convex value function approximations, then problem (10) is a min-cost network flow problem. (2) We use the well-known relationships between the sensitivity analyses of min-cost network flow problems and min-cost flow augmenting trees to find how the solution of problem (10) at time period 1 changes in response to an additional vehicle or an additional load introduced into the system. (3) We find how the state vector at time period 2 changes in response to the change in the solution of problem (10) at time period 1. (4) Finally,

we find how the solution of problem (10) at time period 2 changes in response to the change in the state vector at time period 2. We repeat the same argument in a recursive fashion for the subsequent time periods.

In §2, we start by considering problems with a single vehicle type. We generalize the ideas to multiple vehicle types in §3.

2. Problems with a Single Vehicle Type

In this section, we assume that $|\mathcal{K}| = 1$ and drop the vehicle type superscript, in which case (9) becomes $\tilde{V}_t^{\pi}(r_t) = \sum_{i \in \mathcal{J}} \tilde{V}_{it}^{\pi}(r_{it})$. Letting R be the total number of available vehicles, the relevant domain of $\tilde{V}_{it}^{\pi}(\cdot)$ is $\{0, 1, \dots, R\}$. Therefore, assuming that $\tilde{V}_{it}^{\pi}(0) = 0$ without loss of generality, we can represent $\tilde{V}_{it}^{\pi}(\cdot)$ by a sequence of numbers $\{\tilde{v}_{it}^{\pi}(q) : q = 1, \dots, R\}$, where $\tilde{v}_{it}^{\pi}(q)$ is the slope of $\tilde{V}_{it}^{\pi}(q)$ over $(q - 1, q)$. That is, we have $\tilde{v}_{it}^{\pi}(q) = \tilde{V}_{it}^{\pi}(q) - \tilde{V}_{it}^{\pi}(q - 1)$. In this case, problem (10) can explicitly be written as

$$\min_{(x_t, r_{t+1}, z_{t+1})} \sum_{i, j \in \mathcal{J}} \sum_{l \in \mathcal{L}} c_{ijlt} x_{ijlt} + \sum_{j \in \mathcal{J}} \sum_{q=1}^R \tilde{v}_{j,t+1}^{\pi}(q) z_{j,t+1}(q) \tag{11}$$

$$\text{subject to } \sum_{j \in \mathcal{J}} \sum_{l \in \mathcal{L}} x_{ijlt} = r_{it} \quad \text{for all } i \in \mathcal{J}, \tag{12}$$

$$\sum_{i \in \mathcal{J}} \sum_{l \in \mathcal{L}} x_{ijlt} - r_{j,t+1} = 0 \quad \text{for all } j \in \mathcal{J}, \tag{13}$$

$$r_{j,t+1} - \sum_{q=1}^R z_{j,t+1}(q) = 0 \quad \text{for all } j \in \mathcal{J}, \tag{14}$$

$$x_{ijlt} \leq d_{ijlt} \quad \text{for all } i, j \in \mathcal{J}, l \in \mathcal{L}, \tag{15}$$

$$z_{j,t+1}(q) \leq 1 \quad \text{for all } j \in \mathcal{J}, q = 1, \dots, R, \tag{16}$$

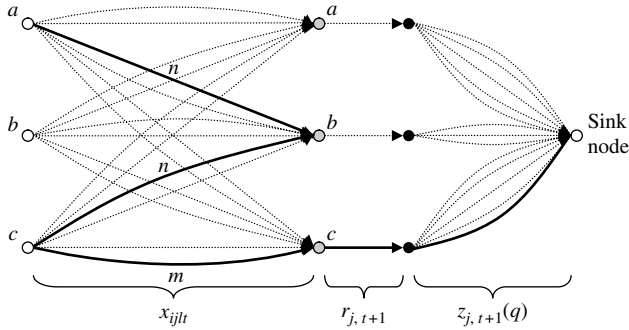
$$\begin{aligned} x_{ijlt}, r_{j,t+1}, z_{j,t+1}(q) \in \mathbb{Z}_+ \\ \text{for all } i, j \in \mathcal{J}, l \in \mathcal{L}, q = 1, \dots, R, \end{aligned} \tag{17}$$

where we use a standard technique to embed the piecewise-linear convex functions $\{\tilde{V}_{j,t+1}^{\pi}(\cdot) : j \in \mathcal{J}\}$ into the problem above through the decision variables $\{z_{j,t+1}(q) : j \in \mathcal{J}, q = 1, \dots, R\}$. In particular, due to the convexity of $\tilde{V}_{j,t+1}^{\pi}(\cdot)$, we have $\tilde{v}_{j,t+1}^{\pi}(1) \leq \tilde{v}_{j,t+1}^{\pi}(2) \leq \dots \leq \tilde{v}_{j,t+1}^{\pi}(R)$. Because the objective function is minimized, noting constraints (14), (16), and (17), we must have

$$\sum_{q=1}^R \tilde{v}_{j,t+1}^{\pi}(q) z_{j,t+1}(q) = \sum_{q=1}^{r_{j,t+1}} \tilde{v}_{j,t+1}^{\pi}(q) = \tilde{V}_{j,t+1}^{\pi}(r_{j,t+1})$$

in the optimal solution. Therefore, the second term in (11) computes $\sum_{j \in \mathcal{J}} \tilde{V}_{j,t+1}^{\pi}(r_{j,t+1})$ (see Nemhauser and Wolsey 1988). Although constraints (13) and (14) can be combined into $\sum_{i \in \mathcal{J}} \sum_{l \in \mathcal{L}} x_{ijlt} - \sum_{q=1}^R z_{j,t+1}(q) = 0$, we leave them separate to emphasize that (13) handles the state transition, whereas (14) handles the computation of the value function approximation. It is easy to see that problem (11) is the min-cost network flow problem in Figure 1. In this

Figure 1. Problem (11) is a min-cost network flow problem.



Note. The path in bold arcs represents a possible min-cost flow augmenting path from node a on the left side to the sink node. Such min-cost flow augmenting paths will be useful in §2.1.

figure, we assume that $\mathcal{I} = \{a, b, c\}$ and $\mathcal{L} = \{n, m\}$. Constraints (12), (13), and (14), respectively, correspond to the flow balance constraints for the white, gray, and black nodes. The sets of decision variables $\{x_{ijl}: i, j \in \mathcal{I}, l \in \mathcal{L}\}$, $\{r_{j,t+1}: j \in \mathcal{I}\}$, and $\{z_{j,t+1}(q): j \in \mathcal{I}, q = 1, \dots, R\}$, respectively, correspond to the arcs that leave the white, gray, and black nodes.

2.1. Policy Gradients with Respect to Vehicle Availabilities

In this section, we develop a method to compute how much the total cost under policy π would change if an additional vehicle were introduced into the system. We let $\{x_t^{\pi d}: t \in \mathcal{T}\}$ and $\{r_t^{\pi d}: t \in \mathcal{T}\}$ be the sequences of decisions and states visited by the system under policy π and load realization $d = \{d_t: t \in \mathcal{T}\}$. That is, $\{x_t^{\pi d}: t \in \mathcal{T}\}$ and $\{r_t^{\pi d}: t \in \mathcal{T}\}$ are recursively computed by

$$x_t^{\pi d} = X_t^\pi(r_t^{\pi d}, d_t), \quad r_{t+1}^{\pi d} = R_{t+1}^\pi(r_t^{\pi d}, d_t),$$

with $r_1^{\pi d} = r_1$. (18)

Then, noting (5), $F_t^\pi(r_t^{\pi d}, d_t, \dots, d_T)$ becomes the total cost incurred at time periods $\{t, \dots, T\}$ under policy π and load realization d . Letting e_i be the $|\mathcal{I}|$ -dimensional unit vector with a one in the element corresponding to $i \in \mathcal{I}$, our objective in this section is to compute

$$\Phi_t^\pi(e_i, d) = F_t^\pi(r_t^{\pi d} + e_i, d_t, \dots, d_T) - F_t^\pi(r_t^{\pi d}, d_t, \dots, d_T) \quad (19)$$

for all $i \in \mathcal{I}$, $t \in \mathcal{T}$. Then, $\mathbb{E}\{\Phi_1^\pi(e_i, D)\}$ tells us how much the total expected cost under policy π would change by introducing an additional vehicle at location i at the first time period. We note that $\Phi_t^\pi(e_i, d)$ can be computed by two simulations of policy π under load realization d , one of which starts with the state vector $r_t^{\pi d}$ and the other with $r_t^{\pi d} + e_i$. However, doing this for all $i \in \mathcal{I}$, $t \in \mathcal{T}$ and for multiple load realizations can get time consuming. Our

objective is to be able to compute $\Phi_t^\pi(e_i, d)$ for all $i \in \mathcal{I}$, $t \in \mathcal{T}$ from a single simulation.

Using (5) and (18), (19) can be written as

$$\begin{aligned} \Phi_t^\pi(e_i, d) &= c_t \cdot \{X_t^\pi(r_t^{\pi d} + e_i, d_t) - X_t^\pi(r_t^{\pi d}, d_t)\} \\ &\quad + F_{t+1}^\pi(R_{t+1}^\pi(r_t^{\pi d} + e_i, d_t), d_{t+1}, \dots, d_T) \\ &\quad - F_{t+1}^\pi(R_{t+1}^\pi(r_t^{\pi d}, d_t), d_{t+1}, \dots, d_T) \\ &= c_t \cdot \{X_t^\pi(r_t^{\pi d} + e_i, d_t) - x_t^{\pi d}\} \\ &\quad + F_{t+1}^\pi(R_{t+1}^\pi(r_t^{\pi d} + e_i, d_t), d_{t+1}, \dots, d_T) \\ &\quad - F_{t+1}^\pi(R_{t+1}^\pi(r_t^{\pi d}, d_t), d_{t+1}, \dots, d_T). \end{aligned} \quad (20)$$

As will be clear shortly, computing $X_t^\pi(r_t^{\pi d} + e_i, d_t) - x_t^{\pi d}$ and $R_{t+1}^\pi(r_t^{\pi d} + e_i, d_t) - r_{t+1}^{\pi d}$ is key to computing $\Phi_t^\pi(e_i, d)$. Because we have

$$(x_t^{\pi d}, r_{t+1}^{\pi d}) = \arg \min_{(x_t, r_{t+1}) \in \mathcal{Y}(r_t^{\pi d}, d_t)} c_t \cdot x_t + \tilde{V}_{t+1}^\pi(r_{t+1}), \quad (21)$$

$X_t^\pi(r_t^{\pi d} + e_i, d_t) - x_t^{\pi d}$ and $R_{t+1}^\pi(r_t^{\pi d} + e_i, d_t) - r_{t+1}^{\pi d}$ are related to how the solution of problem (11) changes when the right side of constraints (12) is increased from $r_t^{\pi d}$ to $r_t^{\pi d} + e_i$.

Consider problem (21) and its network representation in Figure 1. Set the flows on the arcs in this network such that these flows correspond to the optimal solution $(x_t^{\pi d}, r_{t+1}^{\pi d})$. Let $\mathcal{P}_t^\pi(e_i, d)$ be the min-cost flow augmenting path from node $i \in \mathcal{I}$ on the left side to the sink node in this figure. One possible flow augmenting path when $i = a$ is shown in bold arcs. We define the vector $\xi_t^\pi(e_i, d) = \{\xi_{ijl}^\pi(e_i, d): i, j \in \mathcal{I}, l \in \mathcal{L}\}$ as

$$\xi_{ijl}^\pi(e_i, d) = \begin{cases} +1 & \text{if the arc corresponding to variable } x_{ijl} \text{ is a forward arc in } \mathcal{P}_t^\pi(e_i, d), \\ -1 & \text{if the arc corresponding to variable } x_{ijl} \text{ is a backward arc in } \mathcal{P}_t^\pi(e_i, d), \\ 0 & \text{if the arc corresponding to variable } x_{ijl} \text{ is not in } \mathcal{P}_t^\pi(e_i, d). \end{cases}$$

Similarly, we define the vector $\delta_{t+1}^\pi(e_i, d) = \{\delta_{j,t+1}^\pi(e_i, d): j \in \mathcal{I}\}$ as

$$\delta_{j,t+1}^\pi(e_i, d) = \begin{cases} +1 & \text{if the arc corresponding to variable } r_{j,t+1} \text{ is a forward arc in } \mathcal{P}_t^\pi(e_i, d), \\ -1 & \text{if the arc corresponding to variable } r_{j,t+1} \text{ is a backward arc in } \mathcal{P}_t^\pi(e_i, d), \\ 0 & \text{if the arc corresponding to variable } r_{j,t+1} \text{ is not in } \mathcal{P}_t^\pi(e_i, d). \end{cases} \quad (22)$$

For example, for the flow augmenting path in Figure 1, we have $\xi_{abnt}^\pi(e_a, d) = +1$, $\xi_{cbnt}^\pi(e_a, d) = -1$, $\xi_{cmt}^\pi(e_a, d) = +1$, and $\delta_{c,t+1}^\pi(e_a, d) = +1$. The following result characterizes how the solution of problem (21) changes when the number of vehicles available at location i is increased by 1.

LEMMA 1. *The following results hold:*

(1) *We have $X_i^\pi(r_i^{\pi d} + e_i, d_t) = x_i^{\pi d} + \xi_i^\pi(e_i, d)$ and $R_{t+1}^\pi(r_i^{\pi d} + e_i, d_t) = r_{t+1}^{\pi d} + \delta_{t+1}^\pi(e_i, d)$.*

(2) *One element of the vector $\delta_{t+1}^\pi(e_i, d)$ is equal to +1 and the other elements are equal to 0.*

PROOF. The first part is a direct result of Theorem 1 in Powell (1989). The second part holds because any acyclic path from node $i \in \mathcal{J}$ on the left side of Figure 1 to the sink node traverses exactly one of the arcs corresponding to one of the variables $\{r_{j,t+1} : j \in \mathcal{J}\}$. \square

The second part of the lemma implies that if an additional vehicle is introduced at a certain location at time period t , then exactly one element of the state vector at time period $t + 1$ will increase by 1. The following proposition gives an efficient method to compute $\Phi_t^\pi(e_i, d)$.

PROPOSITION 2. *We have $\Phi_t^\pi(e_i, d) = c_t \cdot \xi_t^\pi(e_i, d) + \Phi_{t+1}^\pi(\delta_{t+1}^\pi(e_i, d), d)$ for all $i \in \mathcal{J}$, $t \in \mathcal{T}$ with the boundary condition $\Phi_{T+1}^\pi(\cdot, d) = 0$.*

PROOF. Using Lemma 1, (20) can be written as

$$\begin{aligned} \Phi_t^\pi(e_i, d) &= c_t \cdot \xi_t^\pi(e_i, d) \\ &\quad + F_{t+1}^\pi(r_{t+1}^{\pi d} + \delta_{t+1}^\pi(e_i, d), d_{t+1}, \dots, d_T) \\ &\quad - F_{t+1}^\pi(r_{t+1}^{\pi d}, d_{t+1}, \dots, d_T) \\ &= c_t \cdot \xi_t^\pi(e_i, d) + \Phi_{t+1}^\pi(\delta_{t+1}^\pi(e_i, d), d). \quad \square \end{aligned}$$

The first term in $c_t \cdot \xi_t^\pi(e_i, d) + \Phi_{t+1}^\pi(\delta_{t+1}^\pi(e_i, d), d)$ captures how much the cost incurred at time period t changes in response to an additional vehicle at location i at time period t , whereas the second term captures how much the cost incurred at time periods $\{t + 1, \dots, T\}$ changes in response to an additional vehicle at location i at time period t .

Thus, the idea is to start with the last time period T and let $\Phi_T^\pi(e_i, d) = c_T \cdot \xi_T^\pi(e_i, d)$ for all $i \in \mathcal{J}$. Then, we move to time period $T - 1$. Because $\delta_T^\pi(e_i, d)$ is always a positive integer unit vector, $\Phi_{T-1}^\pi(e_i, d)$ can easily be computed as $c_{T-1} \cdot \xi_{T-1}^\pi(e_i, d) + \Phi_T^\pi(\delta_T^\pi(e_i, d), d)$. We continue in a similar fashion until we reach the first time period.

We note that to evaluate the expected cost impact of an additional vehicle, we need to compute $\mathbb{E}\{\Phi_t^\pi(e_i, D)\}$ as opposed to $\Phi_t^\pi(e_i, d)$ for a particular load realization d . In this case, because computing this expectation is usually intractable, we can sample N load realizations, say d^1, \dots, d^N , use the method described in this section to compute $\Phi_t^\pi(e_i, d^n)$ for all $n = 1, \dots, N$, and use the standard confidence interval methodology to estimate $\mathbb{E}\{\Phi_t^\pi(e_i, D)\}$. By carrying out a “pilot run” that uses a small number of load realizations, we can assess the number of load realizations that are needed to estimate $\mathbb{E}\{\Phi_t^\pi(e_i, D)\}$ with a certain precision (see Law and Kelton 2000).

Finally, we note that a similar method to compute

$$\begin{aligned} \Phi_t^\pi(-e_i, d) &= F_t^\pi(r_i^{\pi d} - e_i, d_t, \dots, d_T) - F_t^\pi(r_i^{\pi d}, d_t, \dots, d_T) \quad (23) \end{aligned}$$

can be developed by using min-cost flow decreasing trees. This will be useful in the next section.

2.2. Policy Gradients with Respect to Load Availabilities

Freight carriers continuously face the problem of evaluating newly arriving loads to decide whether to accept or reject them. In this section, we develop a method to compute how much the total cost under policy π would change if an additional load were introduced into the system. This information can, in turn, be used for load evaluation decisions. The class of policies we consider assume that there is no advance information about future load realizations. For this reason, we assume that if t is the current time period, then the additional load that is introduced into the system is a load that needs to be served at time period t .

Letting e_{ijl} and e_{ijl} be the $|\mathcal{J}|^2|\mathcal{L}||\mathcal{T}|$ and $|\mathcal{J}|^2|\mathcal{L}|$ -dimensional unit vectors with a one in the element corresponding to $i, j \in \mathcal{J}$, $l \in \mathcal{L}$, $t \in \mathcal{T}$ and $i, j \in \mathcal{J}$, $l \in \mathcal{L}$, we want to compute

$$\begin{aligned} \Psi_t^\pi(e_{ijl}, d) &= F_t^\pi(r_i^{\pi d + e_{ijl}}, d_t + e_{ijl}, d_{t+1}, \dots, d_T) \\ &\quad - F_t^\pi(r_i^{\pi d}, d_t, d_{t+1}, \dots, d_T), \quad (24) \end{aligned}$$

which is the change in the total cost of policy π under load realization d in response to an additional load of type l on lane (i, j) at time period t . Using an argument similar to the one in (20) and noting that $r_i^{\pi d + e_{ijl}}$ is a function of the load realizations up to (but not including) time period t , (24) can be written as

$$\begin{aligned} \Psi_t^\pi(e_{ijl}, d) &= F_t^\pi(r_i^{\pi d}, d_t + e_{ijl}, d_{t+1}, \dots, d_T) \\ &\quad - F_t^\pi(r_i^{\pi d}, d_t, d_{t+1}, \dots, d_T) \\ &= c_t \cdot \{X_t^\pi(r_i^{\pi d}, d_t + e_{ijl}) - x_i^{\pi d}\} \\ &\quad + F_{t+1}^\pi(R_{t+1}^\pi(r_i^{\pi d}, d_t + e_{ijl}), d_{t+1}, \dots, d_T) \\ &\quad - F_{t+1}^\pi(r_{t+1}^{\pi d}, d_{t+1}, \dots, d_T). \quad (25) \end{aligned}$$

To compute $\Psi_t^\pi(e_{ijl}, d)$, we now need to characterize $X_t^\pi(r_i^{\pi d}, d_t + e_{ijl}) - x_i^{\pi d}$ and $R_{t+1}^\pi(r_i^{\pi d}, d_t + e_{ijl}) - r_{t+1}^{\pi d}$. These quantities are related to how the solution of the min-cost network flow problem (11) changes when the upper bound on the decision variables x_i is increased from d_t to $d_t + e_{ijl}$.

Consider problem (21) and its network representation in Figure 2. Set the flows on the arcs in this network such that these flows correspond to the optimal solution $(x_i^{\pi d}, r_{t+1}^{\pi d})$. Let $\mathcal{Q}_t^\pi(-e_i, e_j, d)$ be the min-cost flow augmenting path from node $j \in \mathcal{J}$ in the middle section to node $i \in \mathcal{J}$ on the left side in Figure 2. Denote the cost of this min-cost flow augmenting path by $\mathcal{C}_t^\pi(-e_i, e_j, d)$. One possible flow augmenting path when $i = a$, $j = c$ is shown in dashed arcs. We define

an additional load of type l on lane (i, j) at time period t does change the state vector at time period $t + 1$. Given the fact that $\delta_{t+1}^{\pi+}(-e_i, e_j, d)$ and $\delta_{t+1}^{\pi-}(-e_i, e_j, d)$ are positive integer unit vectors, (26) can be easily computed once we know $\Phi_{t+1}^{\pi}(\mp e_i, d)$ for all $i \in \mathcal{J}$.

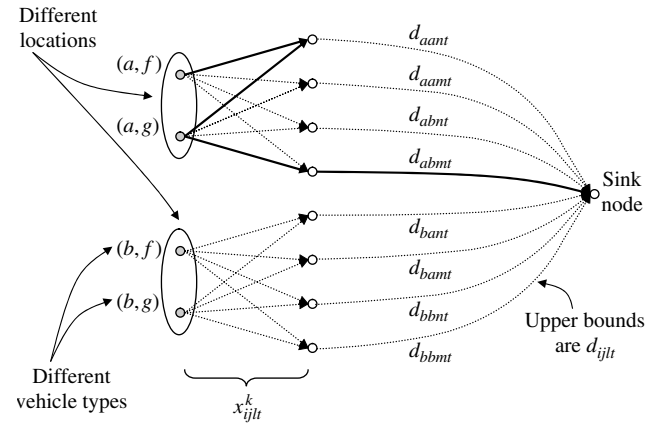
As noted in §1, $F_{t+1}^{\pi}(\cdot, d_{t+1}, \dots, d_T)$ is not necessarily a separable function. However, we propose using (26) as an approximation to $\Psi_t^{\pi}(e_{ijl}, d)$ even when $F_{t+1}^{\pi}(\cdot, d_{t+1}, \dots, d_T)$ is not separable. Our computational experiments show that this approximation yields accurate results. We believe that the accuracy of this approximation is due to the following. The expression in (26) captures the change in the total cost of policy π under load realization d in response to an additional load of type l on lane (i, j) at time period t . Among the three terms on the right side of (26), the first term accurately captures the change in the cost incurred at time period t , whereas the sum of the second and third terms approximately captures the change in the cost incurred at time periods $\{t + 1, \dots, T\}$. Therefore, accurately capturing the change in the cost incurred at the current time period and approximately capturing the change in the cost incurred at the future time periods appear to be adequate to obtain a good approximation.

Assuming that $\mathcal{C}_t^{\pi}(-e_a, e_c, d) + c_{acmt} < 0$ and noting that $\delta_{a,t+1}^{\pi}(-e_a, e_c, d) = -1$, $\delta_{b,t+1}^{\pi}(-e_a, e_c, d) = +1$ for the min-cost flow augmenting path in Figure 2, the approximation in (26) can be interpreted as follows. The first term gives the change in the immediate cost due to the change in the decisions at time period t . The second term gives the change in the future cost due to having an additional vehicle at location b at time period $t + 1$. The third term gives the change in the future cost due to having one less vehicle at location a at time period $t + 1$.

3. Problems with Multiple Vehicle Types

In this section, we extend the ideas in §2 to the case where there are multiple vehicle types. Topaloglu and Powell (2006) note that if there are multiple vehicles types and policy π is characterized by a set of separable piecewise-linear convex value function approximations, then problem (10) becomes a min-cost integer multicommodity network flow problem, and this inhibits exploiting properties of min-cost flow augmenting and decreasing trees as we did in the previous section. Nonetheless, they also show that if each $\tilde{V}_t^{\pi}(\cdot)$ is a linear function of the form $\tilde{V}_t^{\pi}(r_t) = \sum_{i \in \mathcal{J}} \sum_{k \in \mathcal{K}} \tilde{v}_{it}^{\pi k} r_{it}^k$, then problem (10) is a min-cost network flow problem. Furthermore, over a limited domain, say $[0, 1]$ or $[0, 2]$, linear functions approximate piecewise-linear functions quite well. Because the sum of the elements of the $|\mathcal{J}||\mathcal{K}|$ -dimensional vector r_t is always equal to the number of available vehicles, say R , if $|\mathcal{J}||\mathcal{K}| \gg R$, then we expect the elements of the vector r_t to be mostly 0s, 1s, or 2s. In this case, using piecewise-linear approximations does not bring too much advantage over linear approximations. For these reasons, when working on problems with

Figure 3. Problem (27) is a min-cost network flow problem.



multiple vehicle types, we use policies characterized by linear value function approximations.

We now extend the results of §2.1 to the case of multiple vehicle types. Noting (2), the objective function of problem (10) under a policy defined by linear value function approximations is

$$c_t \cdot x_t + \tilde{V}_{t+1}^{\pi}(r_{t+1}) = \sum_{i, j \in \mathcal{J}} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} c_{ijlt}^k x_{ijlt}^k + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \tilde{v}_{j,t+1}^{\pi k} \left(\sum_{i \in \mathcal{J}} \sum_{l \in \mathcal{L}} x_{ijlt}^k \right).$$

Then, the decision function for policy π can be written as

$$X_t^{\pi}(r_t, d_t) = \arg \min_{x_t} \sum_{i, j \in \mathcal{J}} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} (c_{ijlt}^k + \tilde{v}_{j,t+1}^{\pi k}) x_{ijlt}^k \quad (27)$$

subject to (1), (3), (4),

which is the min-cost network flow problem in Figure 3. In this figure, we assume that $\mathcal{J} = \{a, b\}$, $\mathcal{L} = \{n, m\}$, and $\mathcal{K} = \{f, g\}$. Constraints (1) represent the flow balance constraints for the gray nodes. Defining the additional decision variables $\{w_{ijlt}; i, j \in \mathcal{J}, l \in \mathcal{L}\}$ and splitting constraints (3) into two sets of constraints $\sum_{k \in \mathcal{K}} x_{ijlt}^k - w_{ijlt} = 0$ and $w_{ijlt} \leq d_{ijlt}$ for all $i, j \in \mathcal{J}, l \in \mathcal{L}$, the first set represents the flow balance constraints for the white nodes in the middle section.

We let π be a policy characterized by the linear value function approximations $\{\tilde{V}_t^{\pi}(\cdot); t \in \mathcal{T}\}$ with the decision, state transition, and cumulative cost functions $X_t^{\pi}(\cdot, \cdot)$, $R_{t+1}^{\pi}(\cdot, \cdot)$, $F_t^{\pi}(\cdot, \cdot, \dots, \cdot)$. We also define $\{x_t^{\pi d}; t \in \mathcal{T}\}$ and $\{r_t^{\pi d}; t \in \mathcal{T}\}$ similar to their counterparts in §2.1. Letting e_i^k be the $|\mathcal{J}||\mathcal{K}|$ -dimensional unit vector with a one in the element corresponding to $i \in \mathcal{J}, k \in \mathcal{K}$, we want to compute $\Phi_t^{\pi}(e_i^k, d) = F_t^{\pi}(r_t^{\pi d} + e_i^k, d_t, \dots, d_T) - F_t^{\pi}(r_t^{\pi d}, d_t, \dots, d_T)$ for all $i \in \mathcal{J}, k \in \mathcal{K}, t \in \mathcal{T}$.

Consider problem (27) and its network representation in Figure 3. Set the flows on the arcs in this network such that these flows correspond to the optimal solution $x_t^{\pi d}$.

Let $\mathcal{P}_t^\pi(e_i^k, d)$ be the min-cost flow augmenting path from node $(i, k) \in \mathcal{F} \times \mathcal{K}$ on the left side to the sink node in this figure. One possible flow augmenting path when $(i, k) = (a, f)$ is shown in bold arcs. We define the vector $\xi_t^{\pi\kappa}(e_i^k, d) = \{\xi_{ij\ell t}^{\pi\kappa}(e_i^k, d) : i, j \in \mathcal{F}, \ell \in \mathcal{L}, \kappa \in \mathcal{K}\}$ as

$$\xi_{ij\ell t}^{\pi\kappa}(e_i^k, d) = \begin{cases} +1 & \text{if the arc corresponding to variable} \\ & x_{ij\ell t}^\kappa \text{ is a forward arc in } \mathcal{P}_t^\pi(e_i^k, d), \\ -1 & \text{if the arc corresponding to variable} \\ & x_{ij\ell t}^\kappa \text{ is a backward arc in } \mathcal{P}_t^\pi(e_i^k, d), \\ 0 & \text{if the arc corresponding to} \\ & \text{variable } x_{ij\ell t}^\kappa \text{ is not in } \mathcal{P}_t^\pi(e_i^k, d). \end{cases}$$

We also define the vector $\delta_{j,t+1}^{\pi\kappa}(e_i^k, d) = \{\delta_{j,t+1}^{\pi\kappa}(e_i^k, d) : j \in \mathcal{F}, \kappa \in \mathcal{K}\}$ as

$$\delta_{j,t+1}^{\pi\kappa}(e_i^k, d) = \sum_{i \in \mathcal{F}} \sum_{\ell \in \mathcal{L}} \xi_{ij\ell t}^{\pi\kappa}(e_i^k, d). \quad (28)$$

For example, for the flow augmenting path in Figure 3, we have

$$\begin{aligned} \xi_{aant}^{\pi f}(e_a^f, d) &= +1, & \xi_{aant}^{\pi g}(e_a^f, d) &= -1, & \xi_{abmt}^{\pi g}(e_a^f, d) &= +1, \\ \delta_{a,t+1}^{\pi f}(e_a^f, d) &= +1, & \delta_{a,t+1}^{\pi g}(e_a^f, d) &= -1, \end{aligned}$$

and

$$\delta_{b,t+1}^{\pi g}(e_a^f, d) = +1.$$

The following result characterizes how the solution of problem (27) changes when the number of vehicles of type k available at location i is increased by one.

LEMMA 5. *The following results hold:*

(1) We have $X_t^\pi(r_t^{\pi d} + e_i^k, d_t) = x_t^{\pi d} + \xi_t^\pi(e_i^k, d)$ and $R_{t+1}^\pi(r_t^{\pi d} + e_i^k, d_t) = r_{t+1}^{\pi d} + \delta_{t+1}^\pi(e_i^k, d)$.

(2) There exist two disjoint subsets of $\mathcal{F} \times \mathcal{K}$, say $\Delta_{t+1}^{\pi+}(e_i^k, d)$ and $\Delta_{t+1}^{\pi-}(e_i^k, d)$, such that $\delta_{t+1}^\pi(e_i^k, d)$ can be written as

$$\delta_{t+1}^\pi(e_i^k, d) = \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi+}(e_i^k, d)} e_j^\kappa - \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi-}(e_i^k, d)} e_j^\kappa.$$

PROOF. In the first part, the first equality is a direct result of Theorem 1 in Powell (1989) and the second equality follows from the definition of $\delta_{t+1}^\pi(e_i^k, d)$ in (28) and the state transition constraints (2). We show the second part in the appendix. \square

The second part of the lemma shows that an additional vehicle of type k at location i at time period t could change the state vector at time period $t + 1$ in a complicated manner, but each component of the state vector at time period $t + 1$ changes by at most one. The following proposition gives an efficient method to compute $\Phi_t^\pi(e_i^k, d)$.

PROPOSITION 6. *If $F_{t+1}^\pi(\cdot, d_{t+1}, \dots, d_T)$ is a separable function, then we have*

$$\begin{aligned} \Phi_t^\pi(e_i^k, d) &= c_t \cdot \xi_t^\pi(e_i^k, d) + \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi+}(e_i^k, d)} \Phi_{t+1}^\pi(e_j^\kappa, d) \\ &\quad + \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi-}(e_i^k, d)} \Phi_{t+1}^\pi(-e_j^\kappa, d) \end{aligned} \quad (29)$$

for all $i \in \mathcal{F}$, $k \in \mathcal{K}$, $t \in \mathcal{T}$ with the boundary condition $\Phi_{T+1}^\pi(\cdot, d) = 0$.

PROOF. Using Lemma 5, we have

$$\begin{aligned} \Phi_t^\pi(e_i^k, d) &= c_t \cdot \xi_t^\pi(e_i^k, d) \\ &\quad + F_{t+1}^\pi \left(r_{t+1}^{\pi d} + \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi+}(e_i^k, d)} e_j^\kappa - \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi-}(e_i^k, d)} e_j^\kappa, d_{t+1}, \dots, d_T \right) \\ &\quad - F_{t+1}^\pi(r_t^{\pi d}, d_{t+1}, \dots, d_T) \\ &= c_t \cdot \xi_t^\pi(e_i^k, d) \\ &\quad + \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi+}(e_i^k, d)} \{ F_{t+1}^\pi(r_{t+1}^{\pi d} + e_j^\kappa, d_{t+1}, \dots, d_T) \\ &\quad \quad - F_{t+1}^\pi(r_{t+1}^{\pi d}, d_{t+1}, \dots, d_T) \} \\ &\quad + \sum_{(j, \kappa) \in \Delta_{t+1}^{\pi-}(e_i^k, d)} \{ F_{t+1}^\pi(r_{t+1}^{\pi d} - e_j^\kappa, d_{t+1}, \dots, d_T) \\ &\quad \quad - F_{t+1}^\pi(r_{t+1}^{\pi d}, d_{t+1}, \dots, d_T) \}, \end{aligned}$$

where the second equality uses the separability assumption and Lemma 8 in the appendix. \square

On the right side of (29), the first term captures how much the cost incurred at time period t changes in response to an additional vehicle of type k at location i at time period t , whereas the second and third terms capture how much the cost incurred at time periods $\{t+1, \dots, T\}$ changes in response to an additional vehicle of type k at location i at time period t .

$F_{t+1}^\pi(\cdot, d_{t+1}, \dots, d_T)$ is not necessarily a separable function. However, we propose using (29) as an approximation to $\Phi_t^\pi(e_i^k, d)$ even when $F_{t+1}^\pi(\cdot, d_{t+1}, \dots, d_T)$ is not separable.

4. Computational Experiments

This section focuses on the results of §§2.2 and 3 and numerically establishes the accuracy of the methods proposed to compute $\Psi_t^\pi(e_{ijl}, d)$ and $\Phi_t^\pi(e_i^k, d)$. In particular, we use a variety of test problems to show that (26) and (29) can approximate $\Psi_t^\pi(e_{ijl}, d)$ and $\Phi_t^\pi(e_i^k, d)$ accurately even when $F_{t+1}^\pi(\cdot, d_{t+1}, \dots, d_T)$ is not a separable function. The method proposed to compute $\Phi_t^\pi(e_i, d)$ in §2.1 is exact and does not require numerical validation.

Our test problems involve 40 locations and 41 movement modes (40 load types and one movement mode

Table 1. Accuracy of the policy gradients with respect to load availabilities.

Problem	Corr. coeff.	Avg. % dev.	Histogram				Time (sec.)	Coeff. of variation		
			2.5%	5%	10%	25%		Avg.	20%	80%
(10,1000,1,100,2)	0.99	6.27	45	56	71	100	0.09	0.55	0.20	0.82
(10,1000,1,100,4)	0.98	8.91	58	58	61	86	0.08	0.99	0.39	1.56
(10,1000,1,100,8)	0.99	6.95	65	68	70	89	0.08	1.01	0.46	1.47
(10,1000,1,200,2)	0.99	6.31	62	65	72	92	0.08	0.27	0.14	0.36
(10,1000,1,200,4)	0.98	5.73	67	71	74	93	0.08	0.35	0.16	0.45
(10,1000,1,200,8)	0.98	2.85	87	88	89	94	0.08	0.37	0.15	0.52
(30,3000,1,100,2)	0.98	9.48	54	55	62	81	0.31	0.68	0.19	0.99
(30,3000,1,100,4)	0.99	5.71	72	72	75	90	0.28	1.07	0.45	1.47
(30,3000,1,100,8)	0.98	5.88	76	76	76	88	0.27	1.52	0.64	2.26
(30,3000,1,200,2)	0.97	9.64	43	48	59	87	0.31	0.41	0.21	0.57
(30,3000,1,200,4)	0.96	8.87	49	52	61	90	0.28	0.55	0.28	0.76
(30,3000,1,200,8)	0.98	4.52	77	77	81	92	0.28	0.81	0.36	1.11

Note. Percent deviation is $100|\Psi_1^\pi(e_{ijl}, d^n) - \tilde{\Psi}_1^\pi(e_{ijl}, d^n)|/|\Psi_1^\pi(e_{ijl}, d^n)|$, and we ignore the data points with $\Psi_1^\pi(e_{ijl}, d^n)=0$.

for empty repositioning). We label our test problems by (T, D, K, R, e) , where T is the length of the planning horizon, D is the expected number of loads over the planning horizon, K is the number of vehicle types, R is the number of available vehicles, and e is the empty repositioning cost applied on a “per-mile” basis.

Accuracy of the Policy Gradients with Respect to Load Availabilities. We start by testing the accuracy of the method proposed to compute $\Psi_1^\pi(e_{ijl}, d)$. Our experimental setup is as follows. For each test problem, we first obtain a “good” vehicle allocation policy π by using the sampling-based method of Godfrey and Powell (2002a). Having obtained a policy π , we sample N load realizations, say d^1, \dots, d^N . For each load realization d^n , we approximate $\Psi_1^\pi(e_{ijl}, d^n)$ for all $i, j \in \mathcal{J}$, $l \in \mathcal{L}$ by using (26). We let $\{\Psi_1^\pi(e_{ijl}, d^n): i, j \in \mathcal{J}, l \in \mathcal{L}, n = 1, \dots, N\}$ be these approximations. Because the method given in §2.2 requires simulating the behavior of policy π under load realization d^n , at this point we can also compute $F_1^\pi(r_1, d_1^n, d_2^n, \dots, d_T^n)$ as $\sum_{l \in \mathcal{J}} c_l \cdot x_l^{\pi d^n}$. We then physically increase the number of loads of type l on lane (i, j) at time period 1 by 1 and compute $F_1^\pi(r_1, d_1^n + e_{ijl}, d_2^n, \dots, d_T^n)$ by simulating the behavior of policy π under load realization $d^n + e_{ijl1}$. In this way, we can accurately compute $\Psi_1^\pi(e_{ijl}, d^n)$ in a “brute force” fashion as $F_1^\pi(r_1, d_1^n + e_{ijl}, d_2^n, \dots, d_T^n) - F_1^\pi(r_1, d_1^n, d_2^n, \dots, d_T^n)$. Our aim is to compare the approximation $\Psi_1^\pi(e_{ijl}, d^n)$ that is computed through (26) with $\Psi_1^\pi(e_{ijl}, d^n)$ that is computed in a “brute force” fashion.

Table 1 summarizes our findings. The first set of columns give the average percent deviation and the coefficient of correlation between $\{\Psi_1^\pi(e_{ijl}, d^n): i, j \in \mathcal{J}, l \in \mathcal{L}, n = 1, \dots, N\}$ and $\{\tilde{\Psi}_1^\pi(e_{ijl}, d^n): i, j \in \mathcal{J}, l \in \mathcal{L}, n = 1, \dots, N\}$. The second set of columns give a histogram for the percent deviations that shows what fraction of the percent deviations is less than 2.5%, 5%, 10%, and 25%. The next column gives the average time to compute $\{\tilde{\Psi}_1^\pi(e_{ijl}, d^n): i, j \in \mathcal{J}, l \in \mathcal{L}\}$ for a particular load realization d^n . This time

includes the time spent simulating the behavior of policy π under load realization d^n . The last set of columns give summary statistics for the coefficients of variation of $\{\Psi_1^\pi(e_{ijl}, D): i, j \in \mathcal{J}, l \in \mathcal{L}\}$. We estimate the coefficient of variation of $\Psi_1^\pi(e_{ijl}, D)$ as μ_{ijl}/σ_{ijl} , where μ_{ijl} and σ_{ijl}^2 are the sample mean and sample variance of $\{\Psi_1^\pi(e_{ijl}, d^n): n = 1, \dots, N\}$. Using this estimate of coefficient of variation, one can have an idea of how many load realizations are needed to estimate $\mathbb{E}\{\Psi_1^\pi(e_{ijl}, D)\}$ with a certain precision (see Law and Kelton 2000). Because μ_{ijl}/σ_{ijl} depends on $i, j \in \mathcal{J}, l \in \mathcal{L}$, we give the mean, and the 20th and 80th percentiles of $\{\mu_{ijl}/\sigma_{ijl}: i, j \in \mathcal{J}, l \in \mathcal{L}\}$. We note that the coefficient of variation estimates are highly problem-specific and one should not draw general conclusions from them.

The high coefficient of correlation and the low average percent deviation figures in Table 1 show that $\{\Psi_1^\pi(e_{ijl}, d^n): i, j \in \mathcal{J}, l \in \mathcal{L}, n = 1, \dots, N\}$ and $\{\tilde{\Psi}_1^\pi(e_{ijl}, d^n): i, j \in \mathcal{J}, l \in \mathcal{L}, n = 1, \dots, N\}$ are in close agreement. The histograms show that, about 90% of the time, our approximations are within 25% of the true value. If we were working on problems with deterministic load arrivals, then $\Psi_1^\pi(e_{ijl}, d)$ could also be approximated by using the optimal value of the dual variable associated with the load availability constraint $x_{ijl1} \leq d_{ijl1}$ in the “state-time network” formulation of the problem. Powell (1989) reports that, 10% of the time, approximating $\Psi_1^\pi(e_{ijl}, d)$ by using the dual solution brings an error of 50% or more. Therefore, our method can approximate $\Psi_1^\pi(e_{ijl}, d)$ noticeably better than the dual solution of the “state-time network” formulation.

Accuracy of the Policy Gradients with Respect to Vehicle Availabilities. We now compare the approximations obtained through (29), say $\{\tilde{\Phi}_1^\pi(e_i^k, d^n): i \in \mathcal{J}, k \in \mathcal{K}, n = 1, \dots, N\}$, with the values of $\{\Phi_1^\pi(e_i^k, d^n): i \in \mathcal{J}, k \in \mathcal{K}, n = 1, \dots, N\}$ obtained in a “brute force” fashion by increasing the number of vehicles of type k at location i by one and simulating the behavior of policy π under load realization d^n . The results in Table 2 indicate that (29) yields accurate results.

Table 2. Accuracy of the policy gradients with respect to vehicle availabilities.

Problem	Corr. coeff.	Avg. % dev.	Histogram				Time (sec.)	Coeff. of variation		
			2.5%	5%	10%	25%		Avg.	20%	80%
(10, 1000, 20, 100, 2)	0.97	3.28	88	89	91	92	0.41	0.91	0.56	1.16
(10, 1000, 20, 100, 4)	0.99	2.37	88	90	93	96	0.41	0.93	0.56	1.19
(10, 1000, 20, 100, 8)	0.99	3.46	87	89	89	93	0.41	0.93	0.57	1.23
(10, 1000, 20, 200, 2)	0.99	1.63	91	92	94	98	0.31	1.35	0.87	1.78
(10, 1000, 20, 200, 4)	0.97	3.28	88	89	91	92	0.41	1.29	0.86	1.71
(10, 1000, 20, 200, 8)	0.99	2.30	88	89	91	97	0.31	1.26	0.77	1.74
(10, 1000, 40, 100, 2)	1.00	0.77	94	95	97	99	0.59	1.14	0.67	1.55
(10, 1000, 40, 100, 4)	1.00	1.25	95	95	95	98	0.56	0.97	0.62	1.28
(10, 1000, 40, 100, 8)	1.00	0.77	94	95	97	99	0.58	1.14	0.67	1.55
(10, 1000, 40, 200, 2)	0.99	1.38	93	93	94	99	0.61	1.29	0.80	1.72
(10, 1000, 40, 200, 4)	1.00	1.11	93	94	96	99	0.60	1.26	0.75	1.68
(10, 1000, 40, 200, 8)	1.00	0.94	96	96	97	98	0.59	1.22	0.68	1.79
(30, 3000, 20, 100, 2)	0.98	6.78	55	61	67	97	0.86	1.42	0.92	1.89
(30, 3000, 20, 100, 4)	0.99	2.43	77	83	92	99	0.86	1.38	0.91	1.84
(30, 3000, 20, 100, 8)	0.99	3.74	68	70	82	98	0.85	1.56	1.01	2.06
(30, 3000, 20, 200, 2)	0.99	3.26	78	83	84	99	0.90	1.83	1.30	2.41
(30, 3000, 20, 200, 4)	0.99	4.34	64	73	82	97	0.90	1.86	1.20	2.55
(30, 3000, 20, 200, 8)	0.98	2.73	80	82	90	98	0.89	1.75	1.19	2.37
(30, 3000, 40, 100, 2)	0.98	3.90	79	79	83	94	1.66	1.06	0.70	1.32
(30, 3000, 40, 100, 4)	0.99	2.49	82	84	89	98	1.66	0.95	0.63	1.23
(30, 3000, 40, 100, 8)	0.99	3.22	77	78	85	98	1.65	0.87	0.61	1.04
(30, 3000, 40, 200, 2)	0.99	2.38	82	88	89	99	1.70	1.25	0.84	1.60
(30, 3000, 40, 200, 4)	0.99	2.40	82	84	89	99	1.71	1.37	0.92	1.84
(30, 3000, 40, 200, 8)	0.99	2.02	85	88	90	100	1.71	1.26	0.83	1.63

5. Conclusions

We presented efficient methods to assess the sensitivity of a stochastic dynamic fleet management model to fleet size and load availability. Numerical experiments indicated that these methods are accurate and computationally tractable. Information about the cost impact of an additional vehicle or an additional load can, in turn, be used when making fleet sizing, load evaluation, and pricing decisions. Using the method described in §2.2 for load pricing is the topic of another paper (see Topaloglu and Powell 2005).

Appendix

This section presents the omitted proofs. The following result is useful when proving Lemma 3.

LEMMA 7. *If $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl} \geq 0$, then we have $X_i^\pi(r_i^{\pi d}, d_i + e_{ijl}) = x_i^{\pi d}$.*

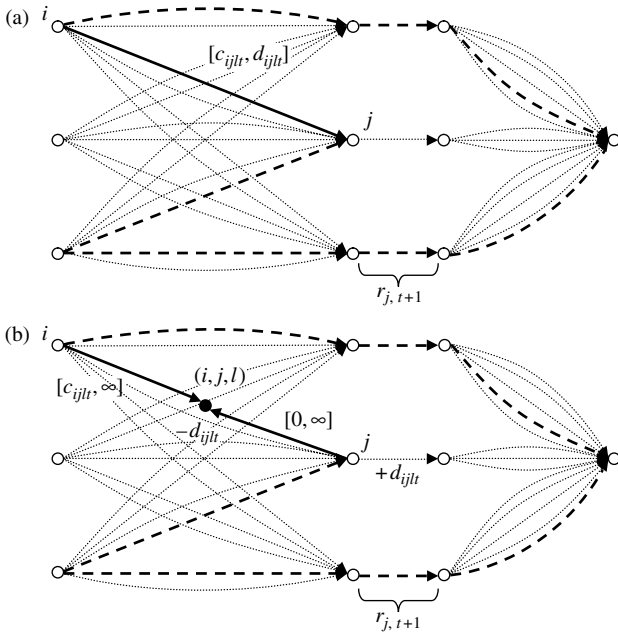
PROOF OF LEMMA 7. Consider problem (21) and its network representation in Figure 2. In this problem, d_{ijl} acts as an upper bound on the decision variable x_{ijl} , and a min-cost network flow problem with upper bounds can be converted to an equivalent problem without upper bounds by the transformation shown in Figure A.1 (see Vanderbei 1997). Therefore, if d_{ijl} is increased by one, then the change in the optimal solution of problem (21) is given by a min-cost flow augmenting path from node j to node (i, j, l) in Figure A.1(b). We denote this min-cost flow augmenting path by \mathcal{Q}' . Because node (i, j, l) has exactly two inbound arcs, there are two possible cases to consider

for \mathcal{Q}' : (1) Either \mathcal{Q}' includes only the bold arc that connects node j to node (i, j, l) —in this case, the cost of \mathcal{Q}' is 0; or (2) \mathcal{Q}' connects node j to node i , and then node i to node (i, j, l) . We let $\mathcal{C}_i^\pi(-e_i, e_j, d)$ be the cost of the min-cost flow augmenting path from node j to node i in Figure A.1(a). Then, the cost of the min-cost flow augmenting path from node j to node i in Figure A.1(b) is also $\mathcal{C}_i^\pi(-e_i, e_j, d)$. Hence, for the second case, the cost of \mathcal{Q}' is $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl}$.

Because \mathcal{Q}' is the min-cost flow augmenting path, if $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl} > 0$, then the first case must hold for \mathcal{Q}' . Thus, we have $X_i^\pi(r_i^{\pi d}, d_i + e_{ijl}) = X_i^\pi(r_i^{\pi d}, d_i)$. We conclude by noting that the possibility of having $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl} = 0$ is ruled out by the random perturbation of the costs so that problem (21) does not have alternative optima. \square

PROOF OF LEMMA 3. To show the first part, we consider two cases depending on the sign of $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl}$: (1) If $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl} \geq 0$, then by Lemma 7, an additional load of type l on lane (i, j) does not change the solution of problem (21). Therefore, $X_i^\pi(r_i^{\pi d}, d_i + e_{ijl}) = x_i^{\pi d}$ and $R_{i+1}^\pi(r_i^{\pi d}, d_i + e_{ijl}) = r_{i+1}^{\pi d}$ hold. (2) Following an argument similar to the proof of Lemma 7, if we introduce an additional load of type l on lane (i, j) , then the change in the solution of problem (21) is given by the min-cost flow augmenting path from node j to (i, j, l) in Figure A.1(b). Let this flow augmenting path be \mathcal{Q}' . If $\mathcal{C}_i^\pi(-e_i, e_j, d) + c_{ijl} < 0$, then the second case in the proof of Lemma 7 holds. Therefore, \mathcal{Q}' first connects node j to node i , and

Figure A.1. (a) The cost and the upper bound for the bold arc that connects node i to node j are c_{ijlt} and d_{ijlt} . (b) The min-cost network flow problem in Figure A.1(a) can be converted to one without the aforementioned upper bound by a simple transformation. We introduce an extra node (i, j, l) with supply $-d_{ijlt}$ and set the supply of node j to $+d_{ijlt}$.



then, node i to node (i, j, l) . This means that \mathcal{Q} is equal to the min-cost flow augmenting path $\mathcal{Q}_i^\pi(-e_i, e_j, d)$ appended by the arc that connects node i to node (i, j, l) . Then, the result follows.

The second part holds because any acyclic path from node $j \in \mathcal{F}$ in the middle section of Figure A.1(a) (or Figure A.1(b)) to node $i \in \mathcal{F}$ on the left side traverses either zero or two of the arcs corresponding to the variables $\{r_{j,t+1}: j \in \mathcal{F}\}$. If a path traverses two of these arcs, then one of these arcs is a forward arc and the other is a backward arc in the path. \square

PROOF OF LEMMA 5. We now show the second part. We first note that any acyclic path from node $(i, k) \in \mathcal{F} \times \mathcal{K}$ on the left side of Figure 3 to the sink node can only visit the nodes $\{(i, k'): k' \in \mathcal{K}\}$.

Assume that the result does not hold. This means that $\delta_{j',t+1}^\pi(e_i^k, d)$ cannot be written as a vector whose elements are $+1, -1$, or 0 , and hence, there exist $j' \in \mathcal{F}, k' \in \mathcal{K}$ such that $|\delta_{j',t+1}^\pi(e_i^k, d)| \geq 2$. Assume that $\delta_{j',t+1}^\pi(e_i^k, d) \geq 2$. Because we have

$$\delta_{j',t+1}^\pi(e_i^k, d) = \sum_{i \in \mathcal{F}} \sum_{\ell \in \mathcal{L}} \xi_{ij'\ell t}^{\pi k'}(e_i^k, d),$$

there exist $i', i'' \in \mathcal{F}$ and $\ell', \ell'' \in \mathcal{L}$ such that $\xi_{ij'\ell' t}^{\pi k'}(e_i^k, d) = +1$ and $\xi_{ij'\ell'' t}^{\pi k'}(e_i^k, d) = +1$. Because of our initial observa-

tion, we must have $i' = i'' = i$. But, having $\xi_{ij'\ell' t}^{\pi k'}(e_i^k, d) = +1$ and $\xi_{ij'\ell'' t}^{\pi k'}(e_i^k, d) = +1$ implies that, on the min-cost flow augmenting path from node (i, k) to the sink node, there are two forward arcs that leave node (i, k') . This contradicts the fact that the min-cost flow augmenting path is acyclic. One can also reach a contradiction by assuming that $\delta_{j',t+1}^\pi(e_i^k, d) \leq -2$. \square

Proofs of Propositions 4 and 6 use the following result.

LEMMA 8. For a separable function $G(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}$, we have

$$G\left(x + \sum_{i=1}^n \alpha_i e_i\right) - G(x) = \sum_{i=1}^n \{G(x + \alpha_i e_i) - G(x)\},$$

where $\alpha_i \in \mathbb{R}$ for all $i = 1, \dots, n$, and e_i is the n -dimensional unit vector with a one in the i th element.

PROOF OF LEMMA 8. Letting $G(x) = \sum_{i=1}^n g_i(x_i)$, where $g_i(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ for all $i = 1, \dots, n$, we have

$$\begin{aligned} G\left(x + \sum_{i=1}^n \alpha_i e_i\right) - G(x) &= \sum_{i=1}^n \{g_i(x_i + \alpha_i) - g_i(x_i)\} \\ &= \sum_{i=1}^n \left\{ \left[\sum_{j=1}^{i-1} g_j(x_j) + g_i(x_i + \alpha_i) + \sum_{j=i+1}^n g_j(x_j) \right] - \sum_{j=1}^n g_j(x_j) \right\}. \quad \square \end{aligned}$$

Acknowledgments

The authors gratefully acknowledge the comments of two anonymous referees that tightened the presentation significantly. The work of the first author was supported in part by National Science Foundation grant DMI-0422133. The work of the second author was supported in part by Air Force Office of Scientific Research grant AFOSR-FA9550-05-1-0121 and National Science Foundation grant CMS-0324380.

References

Carvalho, T. A., W. B. Powell. 2000. A multiplier adjustment method for dynamic resource allocation problems. *Transportation Sci.* **34** 150–164.

Crainic, T., M. Gendreau, P. Dejax. 1993. Dynamic and stochastic models for the allocation of empty containers. *Oper. Res.* **41** 102–126.

Dantzig, G., D. Fulkerson. 1954. Minimizing the number of tankers to meet a fixed schedule. *Naval Res. Logist. Quart.* **1** 217–222.

Dejax, P., T. Crainic. 1987. A review of empty flows and fleet management models in freight transportation. *Transportation Sci.* **21** 227–247.

Frantzeskakis, L., W. B. Powell. 1990. A successive linear approximation procedure for stochastic dynamic vehicle allocation problems. *Transportation Sci.* **24**(1) 40–57.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*. Kluwer Academic Publishers, Norwell, MA.

- Godfrey, G. A., W. B. Powell. 2002a. An adaptive, dynamic programming algorithm for stochastic resource allocation problems I: Single period travel times. *Transportation Sci.* **36**(1) 21–39.
- Godfrey, G. A., W. B. Powell. 2002b. An adaptive, dynamic programming algorithm for stochastic resource allocation problems II: Multi-period travel times. *Transportation Sci.* **36**(1) 40–54.
- Hane, C., C. Barnhart, E. Johnson, R. Marsten, G. Nemhauser, G. Sigismondi. 1995. The fleet assignment problem: Solving a large-scale integer program. *Math. Programming* **70** 211–232.
- Ho, Y.-C., X.-R. Cao. 1991. *Perturbation Analysis of Discrete Event Dynamic Systems*. Kluwer Academic Publishers, Norwell, MA.
- Holmberg, K., M. Joborn, J. T. Lundgren. 1998. Improved empty freight car distribution. *Transportation Sci.* **32** 163–173.
- Law, A. L., W. D. Kelton. 2000. *Simulation Modeling and Analysis*. McGraw-Hill, Boston, MA.
- Nemhauser, G., L. Wolsey. 1988. *Integer and Combinatorial Optimization*. John Wiley & Sons, Chichester, UK.
- Powell, W. B. 1989. A review of sensitivity results for linear networks and a new approximation to reduce the effects of degeneracy. *Transportation Sci.* **23**(4) 231–243.
- Powell, W. B., P. Jaillet, A. Odoni. 1995. Stochastic and dynamic networks and routing. C. Monma, T. Magnanti, M. Ball, eds. *Networks. Handbook in Operations Research and Management Science*. North-Holland, Amsterdam, The Netherlands, 141–295.
- Puterman, M. L. 1994. *Markov Decision Processes*. John Wiley & Sons, New York.
- Topaloglu, H., W. B. Powell. 2005. Incorporating pricing decisions into the stochastic dynamic fleet management problem. Technical report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.
- Topaloglu, H., W. B. Powell. 2006. Dynamic programming approximations for stochastic, time-staged integer multicommodity flow problems. *INFORMS J. Comput.* **18**(1) 31–42.
- Vanderbei, R. 1997. *Linear Programming: Foundations and Extensions*. Kluwer's International Series, Norwell, MA.