Transportation Science

On the Concept of Opportunity Cost in Integrated Demand Management and Vehicle Routing

David Fleckenstein, Robert Klein, Vienna Klein, Claudius Steinhardt

Please scroll down for article—it is on subsequent pages

On the Concept of Opportunity Cost in Integrated Demand Management and Vehicle Routing

David Fleckenstein,[a] Robert Klein,[a] Vienna Klein,[a] Claudius Steinhardt[b,*]

[a] Analytics & Optimization, University of Augsburg, 86159 Augsburg, Germany; [b] Business Analytics & Management Science, University of the Bundeswehr Munich, 85577 Neubiberg, Germany
*Corresponding author
Contact: david.fleckenstein@wiwi.uni-augsburg.de, https://orcid.org/0000-0003-2732-9239 (DF); robert.klein@wiwi.uni-augsburg.de, https://orcid.org/0000-0001-9843-2984 (RK); vienna.klein@uni-a.de; https://orcid.org/0000-0001-8224-1868 (VK); claudius.steinhardt@unibw.de, https://orcid.org/0000-0003-4263-6608 (CS)

Abstract. Integrated demand management and vehicle routing problems are characterized by a stream of customers arriving dynamically over a booking horizon and requesting logistical services, fulfilled by a given fleet of vehicles during a service horizon. Prominent examples are attended home delivery and same-day delivery problems, where customers commonly have heterogeneous preferences regarding service fulfillment and requests differ in profitability. Thus, demand management methods are applied to steer the booking process to maximize total profit considering the cost of the routing decisions for the resulting orders. To measure the requests' profitability for any demand management method, it is common to estimate their opportunity cost. In the context of integrated demand management and vehicle routing problems, this estimation differs substantially from the estimation in the well-examined demand management problems of traditional revenue management applications as, for example, found in the airline or car rental industry. This is because of the unique interrelation of demand control decisions and vehicle routing decisions as it inhibits a clear quantification and attribution of cost, and of displaced revenue, to certain customer requests. In this paper, we extend the theoretical foundation of opportunity cost in integrated demand management and vehicle routing problems. By defining and analyzing a generic Markov decision process model, we formally derive a definition of opportunity cost and prove opportunity cost properties on a general level. Hence, our findings are valid for a wide range of specific problems. Further, based on these theoretical findings, we propose approximation approaches that have not yet been applied in the existing literature, and evaluate their potential in a computational study. Thereby, we provide evidence that the theoretical results can be practically exploited in the development of solution algorithms.

1. Introduction

The widespread adoption of digital distribution channels both enables and forces more and more logistical service providers to manage booking processes actively in order to maintain competitiveness. As a result, their operational planning is no longer limited to solving vehicle routing problems (VRPs). Instead, providers integrate demand management to steer the booking process and either make established business models more profitable or operate novel ones profitably in the first place. These demand management approaches can comprise *demand control decisions* on *prices* of fulfillment

options, the *availability* of fulfillment options, or the *acceptance/rejection* of requests.

Generally, the resulting integrated demand management and vehicle routing problems (i-DMVRPs) share a common structure (Fleckenstein, Klein, and Steinhardt 2023, Waßmuth et al. 2023): A service provider offers *logistical services* characterized by origin and destination in combination with other parameters like service fees, time commitments, or vehicle types. These services are sold throughout a *booking horizon*, with customer requests arriving dynamically. The provider specifies a set of *fulfillment options* to offer in response to an

incoming customer request, consisting of only a single option or multiple options with fixed or varying fees. Subsequently, the customer makes a purchase choice, that is, places an order, based on their individual preferences and the offered options. Fulfillment of all customer orders takes place throughout the *service horizon* by means of a fixed number of vehicles. Capacities of other resources, like driver working hours, may also be limited. The booking and service horizons can be disjoint, which is typical for attended home delivery problems, or overlapping, which is common for same-day delivery and mobility-on-demand problems. Given the capacity restrictions as well as other operational constraints, such as potentially guaranteed service levels, the provider's typical objective is to control demand and routing in a profit-maximizing way, that is, to maximize the difference between revenue and cost.

Because i-DMVRPs are stochastic and dynamic, they can be modeled as Markov decision processes (MDPs), and theoretically, decisions can be optimized by solving the well-known Bellman equation. However, as in demand management problems from traditional revenue management applications, like the airline or car rental industry (Klein et al. 2020), solving the Bellman equation is intractable for industry-sized instances. Therefore, it is common to approach demand management problems by decomposing them into two subproblems (Gallego and Topaloglu 2019, p. 25), with the aim of eliminating the Bellman equation's recursiveness (Klein et al. 2018). The respective subproblems are (1) approximating the opportunity cost for every potential fulfillment option to measure its profitability considering the remaining booking process and (2) solving the actual demand control problem based on the opportunity cost approximation. For this general decomposition-based solution concept, the overall performance largely depends on the quality of the underlying opportunity cost approximation (Klein et al. 2018), which is typically calculated as the difference of the state value approximation for the two resulting postdecision states (selling the fulfillment option versus not selling it). Hence, in revenue management, the analysis of opportunity cost and its properties has already become a standard tool (Talluri and Van Ryzin 2004, p. 92; Adelman 2007; Gallego and Topaloglu 2019, p. 10). However, the corresponding results cannot be transferred directly to i-DMVRPs, due to the mutual interdependencies between demand control decisions and vehicle routing decisions across the entire planning horizon (Agatz et al. 2013).

This observation is the motivation for our work, which aims to inform and accelerate the development of more accurate opportunity cost approximation approaches. To this end, we consider opportunity cost at a formal level, and hence, draw on MDP models of i-DMVRPs as our primary object of study. Investigating these models, we derive mathematical properties and prove that they are valid for the entire family of i-DMVRPs. In a further step, we present three opportunity cost approximation approaches, which exploit the central property, namely the decomposability of opportunity cost. In a computational study, we analyze the potential of these approaches for a variety of problem settings. Primarily, this study is intended as a proof of concept for how the theoretical knowledge about the concept of opportunity cost can drive the development of solution approaches. In addition, the performance evaluation of the presented approaches can serve as a starting point for future research because it hints which approaches have the greatest potential in a certain setting.

Overall, our work has the following contributions:

1. We deepen the theoretical foundation of opportunity cost in the context of i-DMVRPs. We do so by elaborating on decisive differences between opportunity cost in traditional revenue management applications and in i-DMVRPs and by introducing a formal opportunity cost definition that is specifically tailored to i-DMVRPs.

2. We contribute to the existing literature on modeling i-DMVRPs and strengthen the connection between models and solution approaches by introducing a generic MDP model for i-DMVRPs. Further, to the best of our knowledge, we are the first to show formally how to separate demand control decisions from vehicle routing decisions at each decision epoch. This allows investigating the profit impact of demand control decisions in isolation. Additionally, we present a valid model transformation to restore properties in case a certain i-DMVRP application does not naturally inherit them.

3. We introduce and prove four central properties of opportunity cost for our i-DMVRP model that can be exploited within opportunity cost approximation approaches. Those properties are decomposability into two components, potential component-wise negativity, overall nonnegativity, and state value monotonicity.

4. Based on our theoretical findings and focusing on the decomposability as the central property, we present three types of approximation approaches that exploit this property and have yet to be applied to i-DMVRPs: single component approximation, a rather naive hybrid reward approximation, and a more sophisticated hybrid reward approximation. Thereby, we illustrate how the theoretical results lead to direct, practical advances in algorithm development.

The remainder of this paper is structured as follows: In Section 2, we review the literature on opportunity cost in the context of demand management problems in general and on i-DMVRPs in particular. In Section 3, we first model the generic i-DMVRP, for which we show how its opportunity cost differs from the traditional interpretation in revenue management. Then, we present a formal definition of opportunity cost for

i-DMVRPs. In Section 4, we elaborate and prove four central opportunity cost properties, which hold for the generic i-DMVRP. In Section 5, we present the approximation approaches and discuss the computational results. In Section 6, we summarize our work and outline opportunities for future research.

2. Literature Review

In this section, we give an overview of the related literature. We divide this overview into two parts. First, we briefly sketch the evolution of the integration of demand management and vehicle routing as a distinct research area (Section 2.1). Second, we review the scientific contributions to the analysis of opportunity cost for both traditional revenue management problems and i-DMVRPs and position our own work relative to these publications (Section 2.2).

2.1. Integrating Demand Management and Vehicle Routing

Specific similarities between traditional revenue management applications and selling logistical services, such as fixed resources and heterogeneous demand, have prompted the establishment of vehicle routing as a new application for demand management (Agatz et al. 2013). The differentiation of i-DMVRPs from "pure" stochastic and dynamic vehicle routing problems (for a recent review, see Soeffker, Ulmer, and Mattfeld (2022)) is nontrivial. In the remainder of this work, we follow the definition in Fleckenstein, Klein, and Steinhardt (2023). According to that, the control of demand with respect to profitability instead of only feasibility is the distinguishing feature of i-DMVRPs.

Thus, although there is some earlier work in the field of stochastic and dynamic vehicle routing, the works by Campbell and Savelsbergh (2005) and Campbell and Savelsbergh (2006) can be viewed as the first contributions to integrating active demand management and vehicle routing. These publications initiate the literature stream on attended home delivery problems (Yang et al. 2016, Koch and Klein 2020, Vinsensius et al. 2020), for which Snoeck, Merchán, and Winkenbach (2020) and Waßmuth et al. (2023) provide in-depth reviews. The corresponding problems feature disjoint booking and service horizons, meaning that bookings for a specific service horizon, usually a working day, are only possible until a certain cutoff time, such that there is no temporal overlap between booking and fulfillment processes.

On the contrary, Azi, Gendreau, and Potvin (2012) present the first work on steering booking processes in parallel to fulfillment operations, that is, problems with overlapping booking and service horizons. Therewith, they start the literature stream on same-day delivery

(Ulmer 2020, Klein and Steinhardt 2023), which is reviewed thoroughly by Li, Archetti, and Ljubic (2024).

Another stream of literature on i-DMVRPs is initiated by Atasoy et al. (2015) and Hosni, Naoum-Sawaya, and Artail (2014) and considers (shared) passenger transportation problems summed up under the term mobility-on-demand (Bertsimas, Jaillet, and Martin 2019, Al-Kanj, Nascimento, and Powell 2020, Arian, Bai, and Chen 2022, Kullman et al. 2022). The corresponding problems commonly feature overlapping booking and service horizons as well. For a more extensive, cross-application review of the literature on i-DMVRPs, we refer the interested reader to the recent survey by Fleckenstein, Klein, and Steinhardt (2023).

2.2. Theoretical Analysis of Opportunity Cost

As explained in Section 1, instead of approaching demand management problems holistically, solution concepts usually rely on decomposition. The idea is to separate the opportunity cost approximation from optimizing the demand control decision. Both subproblems can be tackled with separate approaches. In traditional revenue management applications, properties of the state values and opportunity cost, such as monotonicity or nonnegativity, have been successfully exploited to improve both the performance of opportunity cost approximation approaches and approaches for optimizing the subsequent demand control decisions. For the first task, that is, the approximation of opportunity cost, approaches based on linear programming (Adelman 2007) as well as on statistical learning (Koch 2017) are known to perform better if constraints are imposed to ensure that the resulting approximation also exhibits existing properties of opportunity cost. For the second task, that is, solving the demand control problem, certain properties simplify the computation of the optimal demand control decisions (shown in Talluri and Van Ryzin (2004, p. 38) for single-resource capacity control). When designing solution approaches for specific applications of demand management, it is necessary to prove whether such structural properties hold or do not hold for the specific demand management problem, that is, under the assumptions associated with the underlying application (see Maddah et al. (2010) for an example from cruise ship revenue management or Quante, Fleischmann, and Meyr (2009) for a manufacturing problem).

Because of the unique problem structure of i-DMVRPs, we cannot directly transfer findings from analyses of demand management problems in traditional revenue management or from other application areas of demand management. In the academic literature on i-DMVRPs, the majority of authors follow the general decomposition-based solution approach described in Section 1 (Fleckenstein, Klein, and Steinhardt 2023). However, the development of solution approaches is mainly driven

by structural reasoning based on characteristics of specific i-DMVRPs paired with the validation of the respective approaches in a computational study. First, there are works that design opportunity cost approximations with the aim of capturing displacement effects regarding potential future orders (Prokhorchuk, Dauwels, and Jaillet 2019, Ulmer 2020, Avraham and Raviv 2021, Lang, Cleophas, and Ehmke 2021). Second, some authors suggest that there is another component of opportunity cost in i-DMVRPs to be considered besides displacement cost: Arian, Bai, and Chen (2022) define opportunity cost as a difference in future profit, which includes fulfillment cost. According to Klein et al. (2018), opportunity cost quantifies the "[…]' consequences" concerning potential future requests and the resulting routing cost […]" (p. 971). Koch and Klein (2020), Yang et al. (2016), and Campbell and Savelsbergh (2005) state that the lost revenue of potential future orders as well as final fulfillment cost have to be anticipated when approximating opportunity cost. Vinsensius et al. (2020) introduce the term "marginal fulfillment cost" of a potential order, and Akkerman, Mes, and Lalla-Ruiz (2022) aim at approximating changes in transportation cost. Abdollahi et al. (2023), Strauss, Gülpinar, and Zheng (2021), Mackert (2019), and Yang and Strauss (2017) provide the most extensive discussion of both revenue-side and cost-side future effects of a demand control decision.

Despite the substantial progress regarding approximation approaches, which the aforementioned publications have contributed to, there is hardly any work on formalizing and generalizing the underlying considerations aside from the following three publications (Waßmuth et al. 2023): Lebedev, Goulart, and Margellos (2021) and Asdemir, Jacob, and Krishnan (2009) conduct a structural analysis of a specific i-DMVRP, namely an attended home delivery dynamic pricing problem with disjoint horizons. Based on MDP formulations that only implicitly model vehicle routing, they derive properties of the value function and the optimal pricing policy. In contrast to Asdemir, Jacob, and Krishnan (2009), Lebedev, Goulart, and Margellos (2021) also model the cost-side via a fulfillment cost approximation. In comparison, our study is more general in that it considers a generic i-DMVRP and explicit vehicle routing decisions. It also focuses on the concept of opportunity cost independent from a specific demand management approach such as dynamic pricing. Although the former two works take up a demand management-oriented view, Ulmer et al. (2020) focus on dynamic vehicle routing problems, which do not necessarily include demand management. They propose a novel MDP modeling framework and show its benefits for informing the design of solution approaches. Our work also aims to establish connections between modeling and solving i-DMVRPs but, different from Ulmer et al. (2020), with a focus on the demand control subproblem rather than on the vehicle routing subproblem.

In summary, our work closes existing research gaps in two ways: First, we provide a theoretical foundation for the existing qualitative reasoning and computational results. Second, our analysis provides the basis for developing algorithmic approaches that have not been considered in existing works, which we demonstrate in a computational study.

3. Opportunity Cost in Integrated Demand Management and Vehicle Routing Problems

In this section, we adapt and discuss the concept of opportunity cost specifically for i-DMVRPs. We first introduce a generic problem definition, on which we base our discussion throughout the whole section. For didactical reasons, we consider a problem as generic as possible and show later on (Section 4.5) how our insights can be transferred to more specific i-DMVRPs. We model the prototypical problem as an MDP (Section 3.1) and show the relevance of opportunity cost for solving the introduced MDP (Section 3.2). Then, we discuss the structural differences of opportunity cost in i-DMVRPs compared with those in traditional revenue management applications and formalize a unified definition of opportunity cost specifically tailored to i-DMVRPs (Section 3.3). For ease of readability, we provide a list of the notation used throughout this paper in Online Appendix A.

3.1. Generic Problem Definition and Modeling

We discuss the concept of opportunity cost in the light of i-DMVRPs for the following generic problem: In each stage $t = 1, \ldots, T$ within a finite booking horizon, at most one *customer request of type* $c \in C$ can arrive with a certain *arrival rate* λ_c^t. A customer request of type c is characterized by the locations of its *origin and destination*, stored by parameter l_c, and *revenue* r_c. Without loss of generality, we assume that multiple individual customer requests can arrive from the same location with the same revenue, such that the arrival rates λ_c^t are independent of whether an individual customer request of type c has already realized before or not. Because the customer requests arrive sequentially, we can distinguish individual customer requests by their *request time* τ.

The provider offers each arriving customer an offer set, which is a subset of a set of fulfillment options, for example, different time windows for order delivery. Once a customer books definitively, their request turns into a confirmed *customer order* that requires a certain amount of fulfillment resources, such as driving time or physical space in a vehicle, depending on the characteristics of the corresponding fulfillment option stored in parameter o. Requests may arrive in parallel to fulfillment operations, that is, the booking and the finite service horizon overlap. During the service horizon, all

confirmed customer orders are served, and the provider incurs the resulting fulfillment cost defined as variable overhead cost arising from the execution of planned routes.

Because an individual opportunity cost value is associated with each (potential) order, that is, with a certain fulfillment option, modeling only a single fulfillment option is sufficient to generally analyze the concept of opportunity cost. By omitting explicitly modeling multiple fulfillment options, we obtain a much simpler model because the provider's decision space for demand control reduces from all possible offer sets to an accept/reject decision per request. Hence, for the sake of simplicity, we consider a single-option model in the remainder of this work. However, we will explain how to generalize the results of our discussion for problems requiring the explicit modeling of multiple fulfillment options in Section 4.5. In the following, we state the corresponding MDP model:

Decision epoch: A decision epoch defines the beginning of a stage of the MDP. In the considered problem, such stages correspond to (constant) time steps $t = 1, \ldots, T$. These time steps are sufficiently small to ensure that at most one customer request arrives between two decision epochs. Hence, λ_c^t can well approximate the probability for observing exactly one request per time interval (Lee and Hersh 1993, Subramanian, Stidham, and Lautenbacher 1999).

State: The state of the system consists of all information that is known so far and relevant for decision making. In i-DMVRPs with overlapping booking and service horizons, information about orders' and vehicles' statuses stored in two separate components are part of the state definition. First, state s_t at decision epoch t stores all confirmed customer orders for which fulfillment has not yet started as tuples (l_c, τ, o) in set $\mathcal{C}_t$. The second component is the overall tour plan at decision epoch t, denoted by ϕ_t. It contains the currently running tours θ_t^v for every vehicle $v \in \mathcal{V}$. Thus, we define the state as $s_t = (\mathcal{C}_t, \phi_t)$. It is important to note that we construct the MDP around the postdecision state because this is the "natural" formulation when making decisions for an observed arriving request (Powell 2022, p. 490). Hence, s_t consistently refers to a postdecision state. An important consequence of this is that a decision in decision epoch t is based on the information stored in state s_{t-1}.

Action: In an MDP model, the action a_t taken at decision epoch t corresponds to the realization of a certain decision. In the considered problem, at most one customer request, denoted by its request type c, can arrive in any stage between $t = 1$ and $t = T$. In the case of a request arrival, the provider must integratively make a demand control decision $g_t \in \mathcal{G}(s_{t-1}, c) \subseteq \{0, 1\}$, that is, accept or reject the arriving request of type c, and a tour planning decision $\phi_t(g_t) \in \Phi(s_{t-1}, c, g_t)$ depending on whether the arriving request must be served

according to the demand control decision. A tour planning decision is a decision on an update of the tour plan stored in the system state, which can also be the decision to leave the tour plan unchanged. The set $\Phi(s_{t-1}, c, g_t)$ defines all potential tour plans that are feasible given the preceding (postdecision) state s_{t-1} and the demand control decision g_t for the arriving customer request of type c. In general, the tour plan must allow for the duly fulfillment of all confirmed customer orders. However, the precise definition of any feasible tour plan, that is, of the action space for the tour planning decision, depends on the specific problem. If no customer request arrives, we set $g_t = 0$, and thus, only a tour planning decision $\phi_t(0) \in \Phi(s_{t-1}, 0)$ not including a new request is required. In summary, the action a_t is formally defined for three distinct cases:

$$a_t = \begin{cases} (0, \phi_t(0)) & \text{for } t = 1, \ldots, T, \text{ if there is no} \\ & \quad \text{customer request arrival} \\ (0, \phi_t(0)) & \text{for } t = 1, \ldots, T, \text{ if the current} \\ & \quad \text{customer request is rejected} \\ (1, \phi_t(1)) & \text{for } t = 1, \ldots, T, \text{ if the current} \\ & \quad \text{customer request is accepted.} \end{cases}$$ (1)

The corresponding action space $\mathcal{A}_t(s_{t-1}, c)$ at a decision epoch t when being in state s_{t-1} and receiving a customer request of type c comprises the two, above introduced components, that is, the demand control component $\mathcal{G}(s_{t-1}, c)$ and the tour planning component $\Phi(s_{t-1}, c, g_t)$. Thus, the action space is formally defined as $\mathcal{A}_t(s_{t-1}, c) = \{(g_t, \phi_t(g_t)) : g_t \in \mathcal{G}(s_{t-1}, c), \phi_t(g_t) \in \Phi(s_{t-1}, c, g_t)\}$. Thereby, the action space for the demand control decision depends on tour planning, because accepting a request of type c given the relevant state s_{t-1} is only feasible if at least one feasible tour plan exists, that is, if $\Phi(s_{t-1}, c, 1) \neq \emptyset$. However, for better readability, we omit this dependency as well as the dependency of the demand control decision g_t on the request type c in the notation. In case there is no customer request arrival at decision epoch t, the respective action space is defined as $\mathcal{A}_t(s_{t-1}) = \Phi(s_{t-1}, 0)$.

Rewards: The demand-control-related rewards r_c are received with actions $g_t = 1$ for $t = 1, \ldots, T$. They are positive and equal the revenue of the customer request type c that is accepted at t, respectively. Demand control actions $g_t = 0$ for $t = 1, \ldots, T$ entail no rewards. Because we assume that the triangle inequality holds, the reward accrued with a tour planning decision is either zero (in case no new fulfillment vehicle operation is triggered) or negative (otherwise). We call those rewards logistics-related rewards and denote them formally by $r_{\phi_t(g_t)}$. They equal the negative of all fulfillment cost that are newly triggered with a decision $\phi_t(g_t)$, that is, the variable overhead cost of all new fulfillment operations that are executed definitively.

Transition: As a consequence of actions and stochasticity, the MDP transitions from a given state s_{t-1} to a successor state s_t. The second state component changes due to the execution of fulfillment operations according to the tour planning decision $\phi_t(g_t)$ in action a_t. In the absence of stochastic elements in the fulfillment operations, such as stochastic travel times, this transition is purely deterministic. Thus, for state s_t, ϕ_t is set to $\phi_t(g_t)$ from a_t. The first state component $\mathcal{C}_{t-1}$, that is, the set of confirmed customer orders for which fulfillment has not yet started, changes as follows: First, the stochasticity of the i-DMVRP influences the transition of the first state component in the form of the potential request arrival according to time-dependent arrival rates λ_c^t. If a customer request arrives and turns into a customer order, it is added. We denote this particular order by $c_t = (l_c^t, \tau^t, o^t)$. Second, the subset of orders $\Psi(\phi_t)$ for which the fulfillment process has started according to the new tour plan ϕ_t are removed. The transitions of the state components can be formalized as follows:

$$\phi_t = \phi_t(g_t), \tag{2}$$

$$\mathcal{C}_t = \begin{cases} \mathcal{C}_{t-1} \setminus \Psi(\phi_t), & \text{if there is no customer} \\ & \quad \text{request arrival or if } g_t = 0 \\ (\mathcal{C}_{t-1} \cup \{c_t\}) \setminus \Psi(\phi_t), & \text{if } g_t = 1. \end{cases} \tag{3}$$

Objective: The objective of the generic i-DMVRP is maximizing the expected profit across all decision epochs starting in state s_0. Thus, we aim at determining a policy x, with $a_t^x(s_{t-1}, c_t) = (g_t^x(s_{t-1}, c_t), \phi_t^x(g_t^x(s_{t-1}, c_t)))$ denoting the action selected by the policy x at decision epoch t, according to the following objective function:

$$\max_x \; \mathbb{E}\left(\sum_{t=1}^{T} (r_{c_t} \cdot g_t^x(s_{t-1}, c_t) + r_{\phi_t^x(g_t^x(s_{t-1}, c_t))}) \,|\, s_0 \right). \tag{4}$$

3.2. State Values and Opportunity Cost

We now represent the previously introduced objective of the generic i-DMVRP (4) by the corresponding value function that equals the well-known Bellman equation. It captures the value of being in a given state and can be applied to find an optimal policy for the MDP model (Powell 2022, p. 46). Specified for the generic model, the value function explicitly models the mutual temporal interdependencies of the two integrated decisions, that is, the demand control decision and the tour planning decision:

$$V_{t-1}(s_{t-1}) = \sum_{c \in C} \lambda_c^t \cdot \max_{g_t \in \mathcal{G}(s_{t-1}, c)} \left(g_t \cdot r_c + \max_{\phi_t(g_t) \in \Phi(s_{t-1}, c, g_t)} (r_{\phi_t(g_t)} \right.$$
$$+ V_t(s_t \,|\, s_{t-1}, \phi_t(g_t))) \Big)$$
$$+ \left(1 - \sum_{c \in C} \lambda_c^t \right) \cdot \max_{\phi_t(0) \in \Phi(s_{t-1}, 0)} (r_{\phi_t(0)} + V_t(s_t \,|\, s_{t-1}, \phi_t(0))), \tag{5}$$

with boundary condition:

$$V_T(s_T) = 0. \tag{6}$$

In i-DMVRPs, an action can comprise two types of integrated decisions, namely, demand control and tour planning decisions. In this work, the effects of a demand control decision are of interest. Thus, it is the target to calculate opportunity cost from comparing state values that reflect such effects separated from potential effects of tour planning decisions at the same decision epoch. Therefore, we introduce a fictive state for each decision epoch $t = 1, \ldots, T$. We refer to it as the *interim state* and denote it as $s_t' \,|\, s_{t-1}, c, g_t$. Technically, c and g_t are captured in additional state dimensions of the interim state. The interim state describes the state that is reached if the provider accepts ($g_t = 1$) a customer request of type c starting in state s_{t-1} or rejects it ($g_t = 0$). In other words, the state is measured after the demand control decision but before the integrated tour planning decision. The idea behind it is comparable to the idea of the *postdecision state* introduced by Powell (2011) (p. 129) with the aim of isolating different effects of decisions and information on the state variable. However, the postdecision state separates the *deterministic effect* of a decision from the *stochastic effect* of the *same* decision in order to ease decision making. The interim state, instead, separates the effects of two *different* decisions, that is, the effects of the *demand control decision* from the effects of a *tour planning decision* taken in the same decision epoch. Figure 1 illustrates the interim state within the decision process and its components.

We denote the value of interim state $s_t' \,|\, s_{t-1}, c, g_t$ by $V_t'(s_t' \,|\, s_{t-1}, c, g_t)$. Generally, it can be calculated as the sum of the succeeding postdecision state's value, that is, of state $s_t \,|\, s_{t-1}, \phi_t^*(g_t)$, and the logistics-related rewards of decision epoch t:

$$V_t'(s_t' \,|\, s_{t-1}, c, g_t)$$
$$= \max_{\phi_t(g_t) \in \Phi(s_{t-1}, c, g_t)} (r_{\phi_t(g_t)} + V_t(s_t \,|\, s_{t-1}, \phi_t(g_t)))$$
$$= r_{\phi_t^*(g_t)} + V_t(s_t \,|\, s_{t-1}, \phi_t^*(g_t)), \tag{7}$$

with $\phi_t^*(g_t)$ denoting the optimal tour planning decision given demand control decision g_t at decision epoch t. Note that this simplification of notation will be used repeatedly throughout the remainder of the paper. Based on interim state values, we can formulate a simplified variant of the value function (5) isolating the demand control decision:

$$V_{t-1}(s_{t-1}) = \sum_{c \in C} \lambda_c^t \cdot \max_{g_t \in \mathcal{G}(s_{t-1}, c)} (g_t \cdot r_c + V_t'(s_t' \,|\, s_{t-1}, c, g_t))$$
$$+ \left(1 - \sum_{c \in C} \lambda_c^t \right) \cdot V_t'(s_t' \,|\, s_{t-1}, 0). \tag{8}$$

In the remainder of our discussion, we denote interim states $s_t' \,|\, s_{t-1}, c, 1$ by $s_t'(c)$ and interim states $s_t' \,|\, s_{t-1}, c, 0$

Figure 1. (Color online) Overview of the MDP Model of the i-DMVRP Booking and Fulfillment Process Including the Interim State
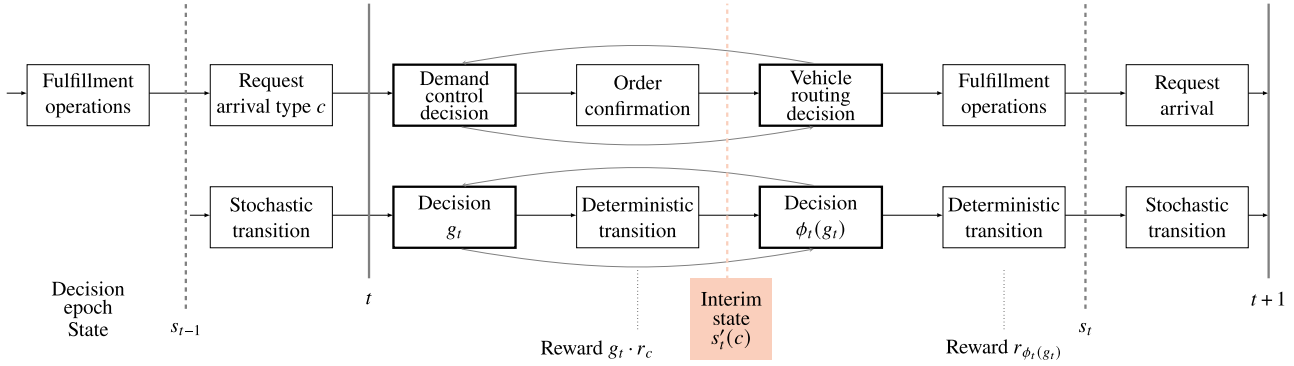
(or $s_t'|s_{t-1}, 0$ in case there is no customer request) by $s_t'(0)$ for ease of presentation. Based on the interim state, the following definition formalizes the concept of opportunity cost for solving the demand control problem of our generic i-DMVRP.

Definition 1. The *opportunity cost* $\Delta V_t(s_{t-1}, c)$ of accepting a customer request of type c in a certain state s_{t-1} is calculated as the difference of the values of the following two interim states: (1) the interim state following the rejection of customer request c and (2) the interim state following the acceptance of c. Thus, it is defined as

$$\Delta V_t(s_{t-1}, c) = V_t'(s_t'(0)) - V_t'(s_t'(c)). \tag{9}$$

This opportunity cost is then used as input to solve the demand control problem, which can be illustrated by the following reformulation of the value function (8). This reformulation is typical in the revenue management literature (Strauss, Klein, and Steinhardt 2018) and yields

$$V_{t-1}(s_{t-1}) = \sum_{c \in C} \lambda_c^t \cdot \underbrace{\max_{g_t \in \mathcal{G}(s_{t-1}, c)} (g_t \cdot (r_c - \Delta V_t(s_{t-1}, c)))}_{\text{Demand control subproblem}} + V_t'(s_t'(0)).$$

$$\tag{10}$$

Because the provider only takes a demand control decision when a certain customer request arrives, the probability $\sum_{c \in C} \lambda_c^t$ is not relevant for decision making. Also, the second summand of Equation (10), that is, $V_t'(s_t'(0))$, is not relevant as it is a constant and independent of the decision. Further, the provider knows r_c. Thus, given the opportunity cost of a customer request of type c, $\Delta V_t(s_{t-1}, c)$, it is possible to solve the demand control problem as a deterministic subproblem. This is why, for industry-sized problems, it is necessary to find accurate and efficient approximation approaches for opportunity cost, that is, for the value function (Strauss, Klein, and Steinhardt 2018). This motivates a deeper understanding of opportunity cost

and of its peculiarities and properties in i-DMVRPs. Knowing certain properties enables exploiting them to accelerate and enhance approaches to approximate opportunity cost as discussed in Section 2.2. Consequently, in the following, we compare opportunity cost in traditional settings, that is, in revenue management problems, and opportunity cost in i-DMVRPs and carve out decisive differences.

3.3. Generalization of the Concept of Opportunity Cost for i-DMVRPs

In traditional revenue management applications, the concept of opportunity cost bases on two main assumptions (Weatherford and Bodily 1992) that cause the opportunity cost to be equivalent to displacement cost (DPC). Those are defined as "[…] the expected loss in future revenue from using the capacity now rather than reserving it for future use" (Talluri and Van Ryzin 2004, p. 33). In the following, we show that this definition cannot be transferred to i-DMVRPs by stating each of the underlying assumptions and investigating it in the respective context.

Assumption 1. *Supply is inflexible, that is, resource capacities are fixed.*

In i-DMVRPs, either driver working times, fleet sizes, or loads represent resources with fixed capacities. As expected, such limited resources may cause a displacement of demand (see Example 1 in Online Appendix B.1). Thus, in most i-DMVRPs, the first assumption is valid.

Assumption 2. *Variable cost associated with the usage of capacity are either negligible or at least directly attributable to individual orders.*

This does not hold in most i-DMVRPs, which can be shown by considering fuel cost as an example: Because the fuel consumption of a fulfillment tour depends on the specific combination of customer locations in the tour, there is no way to calculate and attribute the share

and the resulting cost of each individual customer location (Vinsensius et al. 2020). Further, in i-DMVRPs, such variable overhead cost are not negligible. The exact same combination of state and customer request of the same problem instance can yield different optimal demand control decisions depending on whether fulfillment cost are taken into consideration or are neglected (see Examples 2a and 2b in Online Appendix B.2).

Consequently, we must adapt the traditional concept of opportunity cost, which equalizes opportunity cost and expected displacement cost (Talluri and Van Ryzin 2004, p. 33), for i-DMVRPs. More precisely, a concept is needed that explicitly takes into account variable overhead cost related to order fulfillment: In the literature on i-DMVRPs, some authors already explicitly consider the marginal increase of variable overhead cost caused by the acceptance of a customer request and refer to it as *marginal cost-to-serve (MCTS)* (Vinsensius et al. 2020, Strauss, Gülpinar, and Zheng 2021). For myopic decision making, that is, when neglecting future orders, we can calculate a request's MCTS by optimizing the tour plan for all accepted customer orders including the current request and comparing its variable fulfillment cost with the cost of the optimal tour plan without the current request (see Example 2b in Online Appendix B.2). However, such myopic MCTS are not sufficient for optimal decision making (see Example 3 in Online Appendix B.3). In summary, for optimal decision making, opportunity cost for i-DMVRPs cannot only be revenue related in the form of expected DPC, but also have to take cost-related effects into account in the form of expected MCTS. Correspondingly, we amend the definition of opportunity cost as follows.

Definition 2. In i-DMVRPs with variable overhead cost that are not directly attributable to customer requests, opportunity cost comprises two components: DPC as the difference of cumulative expected future revenue caused by accepting a customer request and MCTS as the difference of expected future fulfillment cost caused by accepting a customer request.

4. Properties and Analytical Discussion of Opportunity Cost for i-DMVRPs

We showed in Section 3 that opportunity cost is calculated as the difference of two value functions. Hence, opportunity cost, as well as its components DPC and MCTS, are recursive functions that are intractable for realistic-sized i-DMVRPs. Consequently, solving i-DMVRPs requires accurate approximations of value functions or opportunity cost. To support the development and selection of respective approximation approaches, we discuss four central properties of the generic i-DMVRP, or more precisely, of the corresponding value function (5) and the derived opportunity cost values. Those are as follows.

1. Decomposability into DPC and MCTS

2. Potential negativity of DPC and MCTS
3. Nonnegativity of opportunity cost
4. Monotonicity of the value function

Please note, for ease of readability, we move the minor mathematical proofs to Online Appendix C and only state the final proofs of the central properties throughout our discussions.

4.1. Decomposability into Displacement Cost and Marginal Cost-to-Serve

To prove the decomposability of opportunity cost into DPC and MCTS, we first define both terms formally, starting with a definition of *expected future revenue* and *expected future fulfillment cost* of a certain interim state s'_{t-1}.

Definition 3. The expected future revenue $R'_{t-1}(s'_{t-1})$ of a given interim state s'_{t-1} at decision epoch $t - 1$ is defined as

$$R'_{t-1}(s'_{t-1}) = \sum_{c \in C} \lambda_c^t \cdot (g_t^* \cdot r_c + R'_t(s'_t | s'_{t-1}, \phi_{t-1}^*(g_{t-1}), g_t^*))$$
$$+ \left(1 - \sum_{c \in C} \lambda_c^t\right) \cdot R'_t(s'_t | s'_{t-1}, \phi_{t-1}^*(g_{t-1}), 0),$$
(11)

with boundary condition

$$R'_T(s'_T) = 0,$$
(12)

and g_t^* denoting the optimal demand control decision, given that a customer request of type c arrives in state s_t.

Definition 4. The expected future fulfillment cost $F'_{t-1}(s'_{t-1})$ of a given interim state s'_{t-1} at decision epoch $t - 1$ is defined as

$$F'_{t-1}(s'_{t-1}) = r_{\phi_{t-1}^*(g_{t-1})} + \sum_{c \in C} \lambda_c^t \cdot (F'_t(s'_t | s'_{t-1}, \phi_{t-1}^*(g_{t-1}), g_t^*))$$
$$+ \left(1 - \sum_{c \in C} \lambda_c^t\right) \cdot F'_t(s'_t | s'_{t-1}, \phi_{t-1}^*(g_{t-1}), 0),$$
(13)

with boundary condition

$$F'_T(s'_T) = r_{\phi_T^*(g_T)}.$$
(14)

Based on Definitions 3 and 4, we now formally define DPC and MCTS.

Definition 5. DPC of accepting a customer request of type c at decision epoch t and state s_{t-1} is defined as

$$\Delta R_t(s_{t-1}, c) = R'_t(s'_t(0)) - R'_t(s'_t(c)).$$
(15)

Definition 6. MCTS of accepting a customer request of type c at decision epoch t and state s_{t-1} is defined as

$$\Delta F_t(s_{t-1}, c) = F'_t(s'_t(0)) - F'_t(s'_t(c)).$$
(16)

DPC and MCTS both depend on the state of the system and all consecutive decisions and transitions.

Thus, both suffer from the curse of dimensionality (Powell, Simao, and Bouzaiene-Ayari 2012) in that the number of potential tour planning decisions that must be evaluated is intractable for realistic-sized instances.

Now, we show that there is a valid decomposition of the value function (7) for interim states into two components. In other words, the *value* of any interim state, that is, $V'_t(s'_t)$, equals the sum of expected future revenue, $R'_t(s'_t)$, and expected future fulfillment cost, $F'_t(s'_t)$. This leads to the following lemma, based on which we then define the first property.

Lemma 1. *The value (function) of an interim state s'_t can be decomposed into two additive components: one capturing expected future revenue and one capturing expected future fulfillment cost:*

$$V'_t(s'_t) = R'_t(s'_t) + F'_t(s'_t). \tag{17}$$

Property 1. Opportunity cost can be decomposed into DPC and MCTS:

$$\Delta V_t(s_{t-1}, c) = \Delta R_t(s_{t-1}, c) + \Delta F_t(s_{t-1}, c). \tag{18}$$

Proof. To prove Property 1, we substitute Lemma 1 into Equation (9). Further, we substitute Definitions 5 and 6, which results in

$$
\begin{aligned}
\Delta V_t(s_{t-1}, c) &= V'_t(s'_t(0)) - V'_t(s'_t(c)) \\
&= (R'_t(s'_t(0)) + F'_t(s'_t(0))) \\
&\quad - (R'_t(s'_t(c)) + F'_t(s'_t(c))) \\
&= R'_t(s'_t(0)) - R'_t(s'_t(c)) + F'_t(s'_t(0)) - F'_t(s'_t(c)) \\
&= \Delta R_t(s_{t-1}, c) + \Delta F_t(s_{t-1}, c). \quad \square
\end{aligned}
$$
$$\tag{19}$$

Please note, despite that DPC and MCTS can be expressed as two separate terms, the decisions they stem from are still interconnected. More precisely, to optimally calculate one of the components, the other one has to be taken into account because both Equations (11) and (13) incorporate optimal decisions that can only be determined based on the original value function including DPC and MCTS.

4.2. Potential Negativity of DPC and MCTS
Contrary to demand management problems of traditional revenue management applications in which DPC can only be nonnegative (Talluri and Van Ryzin 2004, p. 217), in i-DVMRPs, DPC and MCTS can be negative, which is the next property we discuss.

Negative DPC: The intuition behind negative DPC, as they occur in Example 5 in Online Appendix B.5, is the following: turning the considered customer request into a customer order enables accepting one or more expected future customer requests in its vicinity that otherwise would not be profitable regarding their fulfillment cost and revenue.

Negative MCTS: The intuition behind negative MCTS, as they occur in Example 4 in Online Appendix B.4, is the following: Accepting a corresponding customer request and following the subsequent optimal decisions leads to expected future fulfillment cost that is lower than the cost generated by optimal decisions following the rejection of the same customer request. In other words, accepting a certain customer request inhibits the acceptance of one or more future customer requests, which would otherwise be accepted with optimal decisions and would lead to a longer tour, that is, larger fulfillment cost.

This is a decisive difference between the traditional concept of opportunity cost and the newly derived concept for i-DMVRPs.

Property 2. DPC and MCTS can both be negative.

Proof. By Example 4 and Example 5 in Online Appendix B. $\square$

4.3. Nonnegativity of Opportunity Cost
Despite the finding that both DPC and MCTS can be negative, we can show that for the generic MDP model with the value functions defined by (5) or (8), opportunity cost, that is, the sum of DPC and MCTS, is always nonnegative. To prove this property, we show that the value of the interim state following the acceptance of a customer request c_t by action $g_t = 1$ cannot be greater than the value of the interim state following a rejection of the same customer request c_t by action $g_t = 0$. The corresponding proof builds on three lemmas (Lemma 2, Lemma 3, and Lemma 4), which formalize characteristics that are valid for the i-DMVRP model defined in Section 3.1. First, Lemma 2 concerns the stochastic transition probabilities, that is, the arrival rates λ^t_c in a stage t.

Lemma 2. *Stochastic transition probabilities are independent of the set of already confirmed customer orders:*

$$\forall t \in 1, \ldots, T, c \in C : \lambda^t_c \text{ independent of } \mathcal{C}_{t-1}. \tag{20}$$

Second, Lemma 3 concerns the relationship of the action spaces at decision epoch t when starting in any two states s_{t-1} and $\hat{s}_{t-1}$ that only differ in that the latter contains exactly one additional customer order, denoted by $\hat{c}$, that is, $\hat{\mathcal{C}}_{t-1} = \mathcal{C}_{t-1} \cup \{\hat{c}\}$. Starting in those two states, the action space resulting from the latter is a subset of the action space resulting from the former.

Lemma 3. *The action space resulting from any state $\hat{s}_{t-1} = (\mathcal{C}_{t-1} \cup \{\hat{c}\}, \phi_{t-1})$ is a subset of the action space resulting from a corresponding state $s_{t-1} = (\mathcal{C}_{t-1}, \phi_{t-1})$:*

$$\forall t \in 1, \ldots, T, c \in C, \hat{c} \in C : \mathcal{A}(\hat{s}_{t-1}, c) \subseteq \mathcal{A}(s_{t-1}, c). \tag{21}$$

Third, Lemma 4 concerns the state space of an i-DMVRP MDP model. More precisely, it claims that, for any state

$\hat{s}_t = (\hat{\mathcal{C}}_t, \phi_t)$, there exists a state $s_t = (\mathcal{C}_t, \phi_t)$, which only differs in that it does not include a certain customer order $\hat{c}$.

Lemma 4. *For every state* $\hat{s}_t = (\hat{\mathcal{C}}_t, \phi_t)$, *there exists a state* $s_t = (\mathcal{C}_t, \phi_t)$ *with* $\mathcal{C}_t = \hat{\mathcal{C}}_t \setminus \{\hat{c}\}$:

$$\forall \hat{s}_t \text{ with } t \in 1, \ldots, T : \exists s_t : \mathcal{C}_t = \hat{\mathcal{C}}_t \setminus \{\hat{c}\}. \qquad (22)$$

We now consider a certain decision sequence, denoted as $\pi = (\phi_t(g_t), a_{t+1}, a_{t+2}, \ldots, a_T)$, and apply it to a certain sample path $\omega = (c_t, c_{t+1}, c_{t+2}, \ldots, c_T)$. Both start in an interim state s'_t. A sample path is a specific sequence of stochastic realizations throughout the decision epochs. Thus, for this sample path, the respective request arrival probabilities in Equation (5) equal one and the probabilities of other realizations $\omega' \neq \omega$ equal zero.

Given Lemmas 2–4, two further lemmas regarding the resulting *revenue*, denoted by $R_t^{'\pi\omega}(s'_t)$, and regarding the resulting *fulfillment cost*, denoted by $F_t^{'\pi\omega}(s'_t)$, can be derived: More precisely, Lemma 5 states that, applying π to ω, assuming it starts in the interim state $s'_t(c_t)$, results in the same cumulative revenue as assuming ω starts in the corresponding interim state $s'_t(0)$. This is because of the circumstance that in the interim state the revenue of c_t has already been collected. Lemma 6 states that, applying π to ω, assuming it starts in the interim state $s'_t(c_t)$, results in higher or equal fulfillment cost as assuming ω starts in the corresponding interim state $s'_t(0)$.

Lemma 5. *Applying decision sequence π to sample path ω, assuming it starts in interim state $s'_t(c_t)$, results in the same cumulative revenue as assuming ω starts in the interim state $s'_t(0)$:*

$$R_t^{'\pi\omega}(s'_t(c_t)) = R_t^{'\pi\omega}(s'_t(0)). \qquad (23)$$

Lemma 6. *Applying decision sequence π to sample path ω, assuming it starts in the interim state $s'_t(c_t)$, results in higher or equal fulfillment cost as assuming ω starts in the corresponding interim state $s'_t(0)$:*

$$F_t^{'\pi\omega}(s'_t(c_t)) \leq F_t^{'\pi\omega}(s'_t(0)). \qquad (24)$$

Combining Lemmas 5 and 6 shows that applying a decision sequence π to sample path ω starting in interim state $s'_t(c_t)$ cannot result in a greater objective value than starting in interim state $s'_t(0)$. We formalize this in the following lemma.

Lemma 7. *If the same decision sequence π is applied to sample path ω starting in interim state $s'_t(c_t)$, it cannot yield a greater value than it does when starting in interim state $s'_t(0)$:*

$$V_t^{'\pi\omega}(s'_t(c_t)) \leq V_t^{'\pi\omega}(s'_t(0)). \qquad (25)$$

With this in mind, we can formally prove the third opportunity cost property.

Property 3. Opportunity cost is generally nonnegative:

$$\forall c \in C, t = 1, \ldots, T : \Delta V_t(s_{t-1}, c) \geq 0. \qquad (26)$$

Proof. The proof is by contradiction. For a sample path ω, starting in an interim state $s'_t(c_t)$, the optimal sequence of decisions, denoted by $\pi^*(c_t)$, results in value $V_t^{'\pi^*(c_t)\omega}(s'_t(c_t))$. We now assume that this value is higher than any value that we can accrue on the same sample path starting in interim state $s'_t(0)$. However, with Lemmas 2–4, we can feasibly apply $\pi^*(c_t)$ to the sample path starting in interim state $s'_t(0)$ and, with Lemma 7, this results in at least the same value. The original assumption is proven wrong. Hence, the following holds:

$$V_t^{'\omega}(s'_t(c_t)) = V_t^{'\pi^*(c_t)\omega}(s'_t(c_t)) \leq V_t^{'\pi^*(c_t)\omega}(s'_t(0))$$

$$\leq V_t^{'\pi^*(0)\omega}(s'_t(0)) = V_t^{'\omega}(s'_t(0)), \qquad (27)$$

with $\pi^*(0)$ being the optimal sequence of decisions for sample path ω, starting in interim state $s'_t(0)$. This proof by contradiction can be replicated for every sample path $\omega \in \Omega$. Then, because the Bellman function represents the expected value over all possible sample paths $\omega \in \Omega$, following their respective optimal decision sequences π^*, it holds that

$$V'_t(s'_t(c_t)) = V_t^{'\pi^*(c_t)}(s'_t(c_t)) \leq V_t^{'\pi^*(0)}(s'_t(0)) = V'_t(s'_t(0)), \qquad (28)$$

and substituted in Equation (9), this proves that $\Delta V_t(s_{t-1}, c_t) \geq 0$. $\square$

4.4. Monotonicity of the Value Function

We now investigate the monotonicity of the value function in confirmed customer orders, and across decision epochs, as typically done in traditional revenue management (Adelman 2007, Gallego and Topaloglu 2019, p. 10). Property 3 directly implies that the value function is monotonically decreasing in the confirmed customer orders for which fulfillment has not yet started $c \in \mathcal{C}_{t-1}$ at a certain decision epoch t, that is, the following holds:

$$\Delta V_t(s_{t-1}, c) = V'_t(s'_t(0)) - V'_t(s'_t(c)) \geq 0 \Longleftrightarrow V'_t(s'_t(0))$$

$$\geq V'_t(s'_t(c)). \qquad (29)$$

Analogously, we can investigate monotonicity in time across decision epochs. More precisely, we consider the monotonicity of state values of *consecutive* states s_t and $s_{t'}$, with $t' > t$. For two states s_t and $s_{t'}$ to be consecutive, there must exist a sequence of (potentially multiple) stochastic transitions, that is, request arrivals, such that making optimal decisions $a_t^*(s_t, c_t) = (g_t^*, \phi_t^*(g_t^*))$ causes the decision process to transition from s_t to $s_{t'}$ in a finite number of decision epochs. Then, the value function (5) of the generic MDP model presented in Section 3.1 is clearly not monotonically decreasing across consecutive states. This is because of the negative logistics-related

rewards arising throughout the decision process. Consider, for example, a decision epoch at which the routing constraints do not allow feasibly accepting any new order until the end of the service horizon. In this case, no future revenue will be collected. However, the tour planning decisions required for the fulfillment of the confirmed orders cause future logistics-related rewards, which are negative. Over the remaining service horizon, these negative rewards realize and the state value increases, that is, becomes less negative, over the remaining decision epochs until it equals zero in the terminal state.

Despite this finding, it is possible to achieve value function monotonicity across consecutive states by modifying the generic MDP model as shown in Online Appendix D, such that it still models the exact same problem, that is, such that the modified model is mathematically equivalent to the original model. Formally, the modified model differs from the original model regarding *cost realization* and *cost modeling*. Cost realization concerns the point of time in which the cost is incurred in the real application. Cost modeling concerns the decision epoch in which the corresponding cost is taken into account within the MDP model as a negative reward. In the original model, cost realization and cost modeling match. In the modified model, we delay cost modeling. To this end, we augment the state of the modified model and adapt the transition and the value function accordingly, whereas all other model components remain unaltered:

State: For the modified model, we add a third state component denoted by $r_t^{l\,cum}$. It captures the cumulative logistics-related rewards, that is, the negative of the cumulative fulfillment cost that realized before or at decision epoch t. Thus, we define the state as $s_t = (\mathcal{C}_t, \phi_t, r_t^{l\,cum})$. The state space comprises all combinations of possible customer requests and arrival times with potential tour plans and potential cumulative logistics-related rewards.

Transition: The transition of the additional state component $r_t^{l\,cum}$ equals $r_t^{l\,cum} = r_{t-1}^{l\,cum} + r_{\phi_t(g_t)}$. The transition of all other components remains unaltered as introduced in Section 4.

Value function: For the modified model, we delay cost modeling to decision epoch T. Consequently, during the decision epochs $t = 1, \ldots, T$, only rewards r_c are considered in the value function, which is hence defined as

$$\tilde{V}_{t-1}(s_{t-1})$$
$$= \sum_{c \in C} \lambda_c^t \cdot \max_{g_t \in \mathcal{G}(s_{t-1}, c)} \left(g_t \cdot r_c + \max_{\phi_t(g_t) \in \Phi(s_{t-1}, c, g_t)} \tilde{V}_t(s_t | s_{t-1}, \phi_t(g_t)) \right)$$
$$+ \left(1 - \sum_{c \in C} \lambda_c^t \right) \cdot \max_{\phi_t(0) \in \Phi(s_{t-1}, 0)} \tilde{V}_t(s_t | s_{t-1}, \phi_t(0)). \tag{30}$$

We only consider rewards $r_{\phi_t(g_t)}$ in the boundary condition such that the salvage value equals the respective

state component:

$$\tilde{V}_T(s_T) = r_T^{l\,cum}. \tag{31}$$

In the following, we show that the value function of the modified model (30), denoted by $\tilde{V}_t(s_t)$, is monotonically decreasing across consecutive states, that is, $\tilde{V}_t(s_t) \geq \tilde{V}'_{t+1}(s'_{t+1}) \geq \tilde{V}_{t'}(s_{t'})$, with $\tilde{V}'_{t+1}(s'_{t+1})$ denoting the value of interim state s'_{t+1} at decision epoch $t + 1$ in the modified model. Thus, we must show that $\tilde{V}_t(s_t) \geq \tilde{V}'_{t+1}(s'_{t+1})$ and $\tilde{V}'_{t+1}(s'_{t+1}) \geq \tilde{V}_{t+1}(s_{t+1})$ hold for any pair of consecutive states s_t and s_{t+1} as this directly implies that $\forall t < t' \leq T : \tilde{V}'_t(s'_t) \geq \tilde{V}_{t'}(s_{t'})$.

Property 4. The value function of the modified model is monotonically decreasing in the course of decision epochs for consecutive states s_t and s_{t+1}:

$$\forall t = 0, \ldots, T - 1 : \tilde{V}_t(s_t) \geq \tilde{V}_{t+1}(s_{t+1}). \tag{32}$$

Proof. The fact that $\tilde{V}_t(s_t) \geq \tilde{V}'_{t+1}(s'_{t+1})$ holds, follows directly from Equation (10) with the following line of reasoning: Starting in a certain postdecision state s_t in decision epoch t means that there is one more customer request c_{t+1} potentially contributing to the state value compared with starting in the resulting interim state s'_{t+1}. If all potentially arriving requests are not profitable based on s_t, the optimal demand control decision is the rejection in any case, and $\tilde{V}_t(s_t) = \tilde{V}'_{t+1}(s'_{t+1})$ holds, as no revenue can be collected in $t + 1$. Otherwise, if there is a nonempty subset of potentially arriving requests that is profitable, it is optimal to accept these requests and the associated expected revenue positively contributes to $\tilde{V}_t(s_t)$. Then, $\tilde{V}_t(s_t) > \tilde{V}'_{t+1}(s'_{t+1})$ holds; $\tilde{V}'_{t+1}(s'_{t+1}) \geq \tilde{V}_{t+1}(s_{t+1})$ directly follows from Equation (7) because for the considered value function, $\forall t \in 0, \ldots, T - 1 : r_{\phi^*_{t+1}(g_{t+1})} = 0$ holds by definition of the modified MDP formulation. Thus, with $\forall t = 0, \ldots, T - 1$, it holds that $\tilde{V}'_{t+1}(s'_{t+1}) = \tilde{V}_{t+1}(s_{t+1})$. $\square$

Please note, Properties 1–3 as well as the monotonicity of the value function in confirmed customer orders also hold for the modified model. The respective proofs are straightforward with $\tilde{F}'_{t-1}(s'_{t-1}) = F'_{t-1}(s'_{t-1}) + r_t^{l\,cum}$ (analogously to the transformation of $\tilde{V}'_{t-1}(s'_{t-1})$ as previously described and proven in Online Appendix D). Because OC, DPC, and MCTS are defined as the differences of the respective value functions (7), (11), and (13) for different interim states on the same stage t, the constant $r_t^{l\,cum}$ cancels in Equations (9), (15), and (16). Thus, for the modified model, our proofs can be applied as conducted for the original model.

4.5. Generalization for Multiple Fulfillment Options

All properties formulated in Section 4 are also valid for i-DMVRPs with multiple fulfillment options. Compared

with accept/reject control, the difference in this case is that the demand control decision consists of selecting an offer set of feasible fulfillment options. Thus, there is an interim state for each fulfillment option the customer can possibly choose, preceded by an additional stochastic transition according to the customer's purchase choice probabilities. Opportunity cost is then defined separately for each fulfillment option as the difference of the interim state value for the customer choosing the particular option compared with choosing the no-purchase option, as shown in Online Appendix E.

Despite these differences, the two types of rewards, revenue and cost, can still be separated, such that Lemma 1 remains valid and the proof for Property 1 can be conducted as presented for a single fulfillment option. As the multioption case generalizes the single-option case, Example 4 and Example 5 also prove Property 2 for multiple options. Likewise, modeling multiple options does not affect the validity of Lemma 2 because the additional stochastic transition reflecting the customer's purchase choice is also independent of the set of already confirmed customer orders. Lemma 3 also holds for multiple fulfillment options because the feasibility of each fulfillment option depends on the tour planning component in the same way as described for the single-option case. Based on the basic lemmas, the remainder of the proof for Property 3 can be conducted in a similar way as presented above for the single-option case. Finally, the line of reasoning in the proof of Property 4 can also be made based on profitable fulfillment options instead of customer requests.

5. Computational Study

The theoretical results obtained in Sections 3 and 4 contribute to a deeper understanding of i-DMVRPs' mathematical structure, which is useful in itself. On top of that, they can also be of direct practical use because they represent domain knowledge that can be exploited by a variety of solution approaches. In the case of the nonnegativity of opportunity cost (Property 3) and the monotonicity of the value function (Property 4), first promising results in this regard are found by Koch and Klein (2020). They consider a problem with disjoint booking and service horizons and apply a statistical learning approach that imposes structural constraints on policy updates to preserve both properties, which facilitates the learning process. Because we prove that these properties hold in general, such an exploitation is possible for any i-DMVRP, including problems with overlapping horizons based on the modified model (Online Appendix D). Likewise, our results imply that the opportunity cost approximation approach by Adelman (2007), which is based on linear programming and includes constraints exploiting domain knowledge, can also be validly transferred to i-DMVRP solution approaches.

In contrast to Property 3 and Property 4, to the best of our knowledge, the targeted algorithmic exploitation of the decomposability of opportunity cost into DPC and MCTS (Property 1 and Property 2) has not yet been proposed by existing research. Hence, in this computational study, we systematically explore the potential of three general approaches for exploiting the decomposability. In Section 5.1, we first present the three approaches. In Section 5.2, we then describe the design of the computational study; that is, we define the specific i-DVPRP and the settings we consider. Finally, in Section 5.3, we discuss the computational results and derive insights regarding the potential of each of the three approaches.

5.1. General Algorithmic Approaches Exploiting Decomposability

In the following, we explain the algorithmic approaches we analyze in the computational experiments. Because we are interested in the general potential of exploiting decomposability in a certain way rather than in a specific (heuristic) algorithm design, we define the algorithms by formulating the respective variant of the Bellman equation and solve it by means of backward recursion. Thus, we obtain idealized algorithms by combining each of the approximation approaches with an exact method to compute the values of the approximation. For an in-depth review on existing heuristic solution approaches for specific i-DMVRPs, see Fleckenstein, Klein, and Steinhardt (2023).

5.1.1. Single-Component Approximations. This approach is based on the idea that, in certain problem settings, one of the opportunity cost components may generally have a considerably greater absolute value than the other. This suggests that completely neglecting the other component in the opportunity cost approximation should be possible with only a small deterioration of performance.

We derive an idealized *DPC-based approximation* $\Delta\tilde{R}_t(s_{t-1}, c)$ from the following Bellman equation:

$$\tilde{R}_{t-1}(s_{t-1}) = \sum_{c \in C} \lambda_c^t \cdot \max_{g_t \in \mathcal{G}(s_{t-1}, c)} (g_t \cdot (r_c - \Delta\tilde{R}_t(s_{t-1}, c))) + \tilde{R}_t'(s_t'(0)), \tag{33}$$

with

$$r_{\phi_t(s_t')} = 0 \quad \forall t \in \{1, \ldots, T\}. \tag{34}$$

Here, we set the logistics-related rewards equal to zero while retaining the revenue as the immediate reward of an acceptance decision. Thus, decisions are optimized comparing the immediate reward, that is, the potential immediate revenue, with the exact future revenue impact of the acceptance decision neglecting its future cost impact.

Similarly, an idealized *MCTS-based approximation* $\Delta \tilde{F}_t(s_{t-1}, c)$ results from

$$\tilde{F}_{t-1}(s_{t-1}) = \sum_{c \in C} \lambda_c^t \cdot \underset{g_t \in \mathcal{G}(s_{t-1}, c)}{\operatorname{argmax}} (g_t \cdot (r_c - \Delta \tilde{F}_t(s_{t-1}, c)))$$

$$\cdot (-\Delta \tilde{F}_t(s_{t-1}, c)) + \tilde{F}'_t(s'_t(0)). \tag{35}$$

By using an $argmax(\cdot)$ operator in Equation (35), we still consider the revenue for decision making but at the same time prevent it from entering the value function as an actual immediate reward. Thereby, we make sure that $\Delta \tilde{F}_t(s_{t-1}, c)$ only captures logistics-related rewards and thus correctly quantifies the future cost impact of an acceptance decision. This cost impact is then compared with the revenue to derive the currently optimal decision making in each stage.

5.1.2. Hybrid Reward Approximations. Compared with single-component policies, an approximation aiming at both components has structural advantages if both MCTS and DPC have nonnegligible values. Many existing approximation approaches of this type are based on value function approximations that return an aggregated estimate such that the components are not distinguishable (Ulmer 2020). Alternatively, the decomposability property allows hybrid approximation design. The underlying idea is to compute a separate estimate for DPC and MCTS, aggregate the estimates, and therewith, use both as an input for demand control decision making. A new sampling-based solution method recently presented by Abdollahi et al. (2023) constitutes a first non-learning-based step in this direction. They use tentative route planning with sampled orders to compute a DPC estimate and an MCTS estimate separately.

In the class of learning-based approaches, the equivalent are hybrid reward architectures that have been successfully applied to classical reinforcement learning problems (Van Seijen et al. 2017), but, to the best of our knowledge, have not yet been applied to i-DMVRPs. In such approaches, a separate value function approximation is computed for revenue and routing cost, that is, two single-component opportunity cost approximations result, each of which can depend on different features such that learning is expedited. To aggregate the two estimates, several strategies can be used, like a simple additive aggregation (Van Seijen et al. 2017) or a more complex delegation architecture (Russell and Zimdars 2003).

Because Property 1 suggests that i-DMVRPs lend themselves to hybrid reward architectures, we evaluate the potential of such opportunity cost approximation approaches for i-DMVRPs by applying two idealized implementations: The first, rather naive one, bases on the aggregation of the two single-component approximations by simply summing up their value functions

(33) and (35), that is, the corresponding opportunity cost approximation is set to $\Delta \tilde{R}_t(s_{t-1}, c) + \Delta \tilde{F}_t(s_{t-1}, c)$. We refer to this approach as *naive hybrid reward approximation (naive HR approximation)*. The other more sophisticated approach relies on the additive aggregation of an offline learned component and an online learned component. More precisely, we investigate whether it is promising to approximate the DPC offline and the MCTS online or vice versa. In the existing literature, offline-online learning as a general concept has already been successfully applied (Ulmer et al. 2019). However, no approach has been proposed yet that specifically takes advantage of opportunity cost decomposability. For the offline approximation, we draw on the (disaggregate) approximation of the respective single-component approximation ((33) and (35)) for each state and store them in a look-up table by averaging the respective estimates over a two-dimensional state space representation. The first dimension measures the remaining time in the booking process, and the second dimension represents the remaining logistical capacity. For the online approximation, we use the disaggregate DPC (MCTS) value from the original model ((15) and (16)). Then, we set the corresponding opportunity cost approximation equal to $\Delta \tilde{F}_t^{aggr}(s_{t-1}, c) + \Delta R_t(s_{t-1}, c)$ (or $\Delta \tilde{R}_t^{aggr}(s_{t-1}, c) + \Delta F_t(s_{t-1}, c)$, respectively). We refer to this approach as *offline-online hybrid reward approximation (offline-online HR approximation)* and distinguish its variants by using the term *DPC (MCTS)* when the DPC (MCTS) are approximated online.

5.1.3. Compact Overview. In summary, this computational study compares decision making based on the following approximations, which include the ones described above and the true opportunity cost derived from solving the dynamic program (9) as the benchmark:
- DPC-based approximation: $\Delta \tilde{V}_t(s_{t-1}, c) = \Delta \tilde{R}_t(s_{t-1}, c)$
- MCTS-based approximation: $\Delta \tilde{V}_t(s_{t-1}, c) = \Delta \tilde{F}_t(s_{t-1}, c)$
- Naive hybrid reward approximation: $\Delta \tilde{V}_t(s_{t-1}, c) = \Delta \tilde{R}_t(s_{t-1}, c) + \Delta \tilde{F}_t(s_{t-1}, c)$
- Offline-online HR approximation (DPC-based): $\Delta \tilde{V}_t(s_{t-1}, c) = \Delta \tilde{F}_t^{aggr}(s_{t-1}, c) + \Delta R_t(s_{t-1}, c)$
- Offline-online HR approximation (MCTS-based): $\Delta \tilde{V}_t(s_{t-1}, c) = \Delta \tilde{R}_t^{aggr}(s_{t-1}, c) + \Delta F_t(s_{t-1}, c)$
- True opportunity cost (benchmark): $\Delta \tilde{V}_t(s_{t-1}, c) = \Delta V_t(s_{t-1}, c)$

5.2. Study Design and Methodology

To evaluate the approaches introduced above, we apply them for decision making in various different settings. All settings have in common that they reflect an i-DMVRP with the same problem structure as introduced in Section 3.1 and the following additional assumptions: disjoint booking and service horizons, a single fulfillment vehicle, pure accept/reject decisions,

and a booking horizon of $T = 10$ potential decision epochs. With that, we ensure that the instance size is sufficiently small for being computationally tractable without overly compromising complexity.

The settings differ with regard to four parameters whose realizations emulate characteristics of i-DMVRPs from typical application areas. Two of those parameters define a setting's *customer distribution*, one sets the general *profitability* of a setting, and the last defines the type of *capacity consumption* considered in a setting. In the following, we further describe those parameters and motivate our choice of their realizations.

5.2.1. Customer Distribution. The customer distribution of a setting is characterized by two parameter values, which are the customers' *location distribution* and their *revenue distribution*, and is further defined by the mutual interplay of those two parameters.

Location distribution: To vary the level of difficulty of demand consolidation, we draw the customer locations l_c from two different customer distributions on a line segment of 50 *length units* (*LU*) in the interval $[−25, 25]$, with a centrally located depot. The first customer distribution follows a uniform distribution over the entire interval and hence mimics an urban area. The second customer distribution is drawn from two truncated normal distributions with means $−10$ and 20 and the same standard deviation of $2.5LU$, from which we draw 50% of the customer locations each. Therewith, we obtain two clusters that could represent two villages in a rural area.

Revenue distribution: The customers' revenue distribution is an important characteristic that influences displacement effects, both in a spatial and a temporal sense. Thus, on the one hand, we consider *homogeneous revenues* (no additional displacement effects) with a value of $r_c = 15$ *monetary units* (*MU*). This corresponds to, for example, next-day parcel delivery with a static, uniform delivery fee. On the other hand, we consider heterogeneous settings with 70% low-revenue customers ($r_c = 15MU$) and 30% high-revenue customers ($r_c = 25MU$). We vary their distribution over time as follows: The sequence of customer arrivals either follows a strict *high-before-low* sorting (low displacement effects), a *random* sorting (medium displacement effects), or a strict *low-before-high* sorting (high displacement effects). Such variations over time can occur, for

example, as a result of markup pricing or markdown pricing. In addition, for each of those three schemes, we consider a customer setting in which the high-revenue customers only originate from the distant cluster to mimic a *distance-based* pricing scheme, which is common, for example, in mobility-on-demand applications. We refer to this special combination of location distribution and revenue distribution as *clustered sorted*.

Overall, these two parameters and the combinations of their potential realizations yield 11 different customer settings, which are marked with a ✓ in Table 1. As mentioned before, for all those 11 customer settings, we vary two more parameters that address the general profitability of a setting, as well as the capacity consumption.

5.2.2. Profitability. To vary the general level of profitability of a problem setting, we adjust the relation between revenues and cost by solving all previously mentioned settings for three different routing cost factors. More precisely, we consider low-cost settings with cost of $0.2MU$ per LU traveled, medium-cost settings with $0.6MU$, and high-cost settings with $1MU$. The different profitability settings represent different fields of application. Low-cost settings are dominant in attended home delivery due to minimum order values that cause high revenue relative to cost. On the contrary, high-cost settings occur in mobility-on-demand applications with comparatively low revenues equal to the fare the provider charges.

5.2.3. Capacity Consumption. To consider the impact of the marginal capacity consumption per order for all our settings, we additionally assume another parameter. Its value either reflects *route length constraints*, or *physical capacity constraints* such that only one of them is restrictive in a setting. First, for the settings with restrictive physical capacity, we set the maximum capacity to three orders and assume unit demand for all orders. Hence, the marginal capacity consumption is both uniform and known a priori. A typical application that is represented by these settings is attended home delivery of bulky goods. Second, for settings with a restrictive route length, we set the maximum capacity to $50\,LU$. With this type of constraint, the marginal capacity consumption is variable among the orders and is unknown until the final routing decision. This is typical for

Table 1. Customer Distribution Parameters

	Location distribution		
Revenue distribution	Uniform (unif)	Clustered (clust)	Clustered sorted (clust_sort)
Homogeneous (homog)	✓	✓	
High-before-low (h-b-l)	✓	✓	✓
Low-before-high (l-b-h)	✓	✓	✓
Random (rand)	✓	✓	✓

applications that are mainly time constrained such as same-day delivery.

Because we aim at a full-factorial analysis of our approaches, we solve all the 11 customer settings from Table 1 for all six combinations of parameter values defining the profitability and capacity consumption. Hence, we test our approaches in 66 settings. For each setting, we draw 50 instances. Thereby, for each instance, we draw a fixed (deterministic) customer stream of 10 customers from the setting-dependent customer distribution. Then, for each customer, we assume a probability of $\lambda_0^t = 0.5$ that the respective request does not arrive, which is the only source of stochasticity, once a setting is defined.

5.3. Performance Evaluation

To evaluate the performance of the considered approaches, for each of the 66 settings, we calculate the mean objective value, that is, profit after fulfillment, over the 50 different instances drawn. In Figure 2 and Online Appendix F, we report the results and discuss them in the following.

5.3.1. Single-Component Approximation.

The comparison between the two variants of this approximation approach reveals that the DPC-based variant outperforms the MCTS-based variant in high-profitability settings and most route length-constrained settings. Especially in the latter settings, it also achieves very small optimality gaps and is competitive with the more sophisticated approximations. In turn, the MCTS-based variant is among the best-performing approaches in low-profitability, physical capacity-constrained settings. On the downside, the performance of both variants fluctuates strongly across different settings. For example, we observe a very bad performance of the MCTS-based variant for low-profitability settings with distance-dependent revenues and even negative objective values for the DPC-based variant in some low-profitability, physical capacity-constrained settings. However, despite these issues with solution quality fluctuations, single-component approximations can be a viable approach because their practical implementations usually require less computational effort, and if the right variant is applied in the right setting, it can offer competitive solution quality.

5.3.2. Naive Hybrid Reward Approximation.

The naive hybrid reward approximation is among the best-performing policies in many high-profitability settings and, with some outliers, also in medium-profitability settings. Hence, in these settings, ignoring the interdependency between DPC and MCTS values does not appear particularly harmful to the solution quality. However, it performs weaker and shows severe outliers in low-profitability settings, especially if capacity is route length constrained. Because of this lack of robustness

and the small gains relative to single-component approximations, the additional computational effort may only be justified in a few specific settings.

5.3.3. Offline-Online HR Approximation.

Comparing the two variants of the offline-online HR approximation, the MCTS-based variant performs superior in physical capacity-constrained settings and in low-profitability settings. With a few exceptions, it is even the best-performing approach for these settings and shows a very robust performance overall. The DPC-based variant is slightly better in some route length-constrained settings with medium or high profitability. However, it can be considered inferior overall because its performance is also more variable. A possible explanation for this result is that the MCTS show a stronger variation over similar states, whereas the differences in DPC are smaller for states with the same decision epoch and remaining capacity. If this is indeed the case, anticipating the MCTS online in disaggregate form should yield more accurate opportunity cost estimates. Overall, the results are very promising for the MCTS-based variant, even though its practical implementations are expected to require the highest computational effort. We can conclude that none of the presented approaches exploiting the decomposability of opportunity cost is strictly dominated such that all are worth being investigated further. Our results can serve as rough guidance as to which approach is most promising in a certain setting. Interestingly, we also observe that the relevance of DPC and MCTS correlates with the performance of the approaches. As an example, the MCTS gain relevance with decreasing profitability because the fulfillment cost becomes larger relative to the revenues, and thus, we would expect a relative performance gain of the MCTS-based approaches. This gain can indeed be observed for both the MCTS-based single component approximation and the MCTS-based offline-online HR approximation.

6. Conclusion and Future Research Opportunities

This work constitutes the first formal generic analysis of opportunity cost in i-DMVRPs. We showed that the original interpretation of opportunity cost from traditional revenue management applications cannot be transferred to i-DMVRPs and therefore generalized its definition. Further, we analytically investigated opportunity cost properties with the central property being the decomposability into DPC and MCTS. Finally, we conducted a computational study and applied previously unconsidered approximation approaches as a proof of concept for that the properties can be directly exploited in algorithm design. In the following, we first briefly summarize our theoretical and computational

Figure 2. (Color online) Objective Values Resulting from the Different Opportunity Cost Approximations Averaged Across 50 Instances per Setting

(a) Physical capacity-constrained settings

(b) Route length-constrained settings
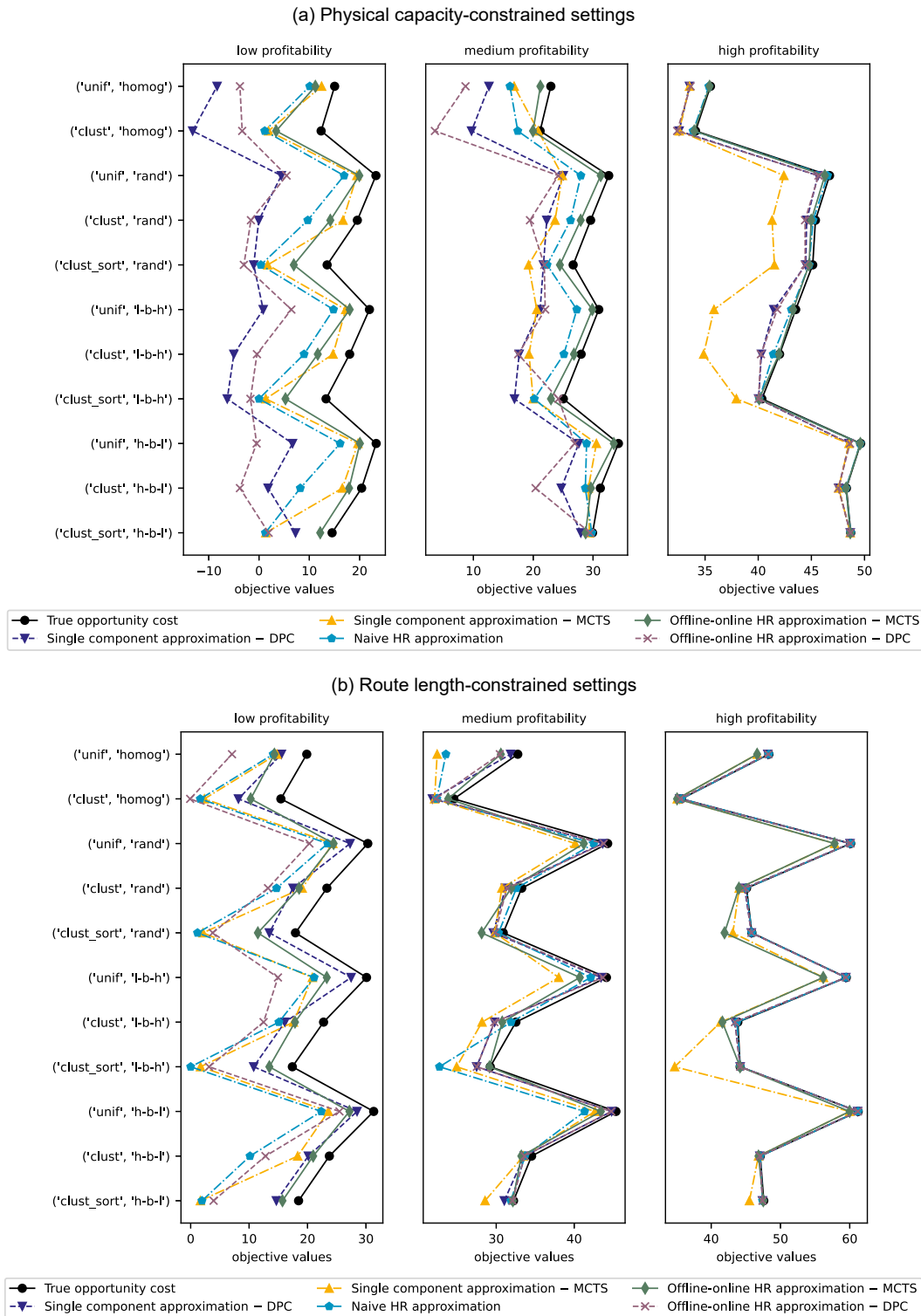
results. Second, we discuss the future research opportunities emerging from our work.

In existing works, insights are mainly derived either from qualitative reasoning and computational studies

(Mackert 2019, Vinsensius et al. 2020) or from analytical analyses for specific i-DMVRPs (Asdemir, Jacob, and Krishnan 2009, Lebedev, Margellos, and Goulart 2020). In contrast, our approach is both analytical and generic,

which is why our results apply to the whole family of the most common types of i-DMVRPs. By showing that it is possible to substantiate existing observations at the modeling level, we not only confirm their general validity but also highlight the importance of modeling frameworks for understanding and exploiting the problem structure of i-DMVRPs when designing solution concepts. The following summary captures the essence of our theoretical framework.

For i-DMVRPs, opportunity cost measures the impact of selling an order as a consequence of a demand control decision, on both (expected) future revenues, in the form of DPC, and (expected) future cost of fulfillment, in the form of MCTS (Yang and Strauss 2017). With our analysis, we can precisely express this impact in a formal way. First, we show that the impact of a demand control decision can be isolated from the impact of the subsequent routing decision. Second, we show that both revenue impacts and cost impacts can be formally expressed and thus mathematically decomposed from each other (Property 1). Third, we find that both impacts can have a positive or a negative sign (Property 2). Fourth, although we can measure the impacts in isolation through MCTS and DPC, the corresponding values are still nonseparable and, in sum, nonnegative, which leads to the general nonnegativity of opportunity cost (Property 3). Finally, the value function decreases monotonically in an increasing set of accepted but not yet served customer orders and can also be transformed to decrease monotonically in time (Property 4).

In addition to these theoretical findings, our computational study shows that exploiting the decomposability of opportunity cost allows transferring new algorithmic approaches to i-DMVRPs that have been shown to be beneficial in related fields. This includes different hybrid reward approximation approaches, which are already established in reinforcement learning (Van Seijen et al. 2017). The computational evaluation of idealized implementations of these approaches illustrates their potential in the context of i-DMVRPs. At the same time, we also observe that the relative performance of the approaches can vary considerably even among similar settings. This highlights the importance of continuously expanding the available toolbox of approaches such that a wide range of settings can be suitably tackled.

For future research, we primarily see the opportunity to exploit the decomposability and the other properties in heuristic solution algorithms for specific i-DMVRPs, not only with the aim of improving solution quality but also for reducing runtimes. In particular, we believe that heuristic versions of the presented hybrid reward approximation approaches deserve further investigation. Another future research question arises in connection with Property 4. In the case of problems with overlapping horizons, it only holds for modified models

(Online Appendix D). Interestingly, this model transformation can be viewed as a form of reward shaping (Laud 2004), which establishes monotonicity at the cost of increasing the delay of rewards. To the best of our knowledge, reward shaping has not been applied to solve i-DMVRPs. Hence, although it is out of scope for the study at hand, our theoretical results suggest that there is potential to investigate its application.

Finally, we see the potential for doing similar theoretical research for other novel demand management problems that feature integrated combinatorial optimization problems to plan service fulfillment, such as scheduling problems (Xu, Wang, and Huang 2015).

Acknowledgments

References

Abdollahi M, Yang X, Nasri MI, Fairbank M (2023) Demand management in time-slotted last-mile delivery via dynamic routing with forecast orders. *Eur. J. Oper. Res.* 309(2):704–718.

Adelman D (2007) Dynamic bid prices in revenue management. *Oper. Res.* 55(4):647–661.

Agatz N, Campbell AM, Fleischmann M, Van Nunen J, Savelsbergh M (2013) Revenue management opportunities for Internet retailers. *J. Revenue Pricing Management* 12(2):128–138.

Akkerman F, Mes M, Lalla-Ruiz E (2022) Dynamic time slot pricing using delivery costs approximations. de Armas J, Ramalhinho H, Voß S, eds. *Computational Logistics*, Lecture Notes in Computer Science, vol. 13557 (Springer, Cham, Switzerland), 214–230.

Al-Kanj L, Nascimento J, Powell WB (2020) Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles. *Eur. J. Oper. Res.* 284(3):1088–1106.

Arian E, Bai X, Chen X (2022) Joint pricing and routing for a ridesharing platform in low-density rural areas. Working paper, University of Illinois at Urbana-Champaign, Urbana.

Asdemir K, Jacob VS, Krishnan R (2009) Dynamic pricing of multiple home delivery options. *Eur. J. Oper. Res.* 196(1):246–257.

Atasoy B, Ikeda T, Song X, Ben-Akiva ME (2015) The concept and impact analysis of a flexible mobility on demand system. *Transportation Res. Part C Emerging Tech.* 56:373–392.

Avraham E, Raviv T (2021) The steady-state mobile personnel booking problem. *Transportation Res. Part B Methodological* 154:266–288.

Azi N, Gendreau M, Potvin JY (2012) A dynamic vehicle routing problem with multiple delivery routes. *Ann. Oper. Res.* 199(1):103–112.

Bertsimas D, Jaillet P, Martin S (2019) Online vehicle routing: The edge of optimization in large-scale applications. *Oper. Res.* 67(1):143–162.

Campbell AM, Savelsbergh M (2005) Decision support for consumer direct grocery initiatives. *Transportation Sci.* 39(3):313–327.

Campbell AM, Savelsbergh M (2006) Incentive schemes for attended home delivery services. *Transportation Sci.* 40(3):327–341.

Fleckenstein D, Klein R, Steinhardt C (2023) Recent advances in integrating demand management and vehicle routing: A methodological review. *Eur. J. Oper. Res.* 306(2):499–518.

Gallego G, Topaloglu H (2019) *Revenue Management and Pricing Analytics* (Springer, New York).

Hosni H, Naoum-Sawaya J, Artail H (2014) The shared-taxi problem: Formulation and solution methods. *Transportation Res. Part B Methodological* 70:303–318.

Klein V, Steinhardt C (2023) Dynamic demand management and online tour planning for same-day delivery. *Eur. J. Oper. Res.* 307(2):860–886.

Klein R, Koch S, Steinhardt C, Strauss AK (2020) A review of revenue management: Recent generalizations and advances in industry applications. *Eur. J. Oper. Res.* 284(2):397–412.

Klein R, Mackert J, Neugebauer M, Steinhardt C (2018) A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum* 40(4):969–996.

Koch S (2017) Least squares approximate policy iteration for learning bid prices in choice-based revenue management. *Comput. Oper. Res.* 77:240–253.

Koch S, Klein R (2020) Route-based approximate dynamic programming for dynamic pricing in attended home delivery. *Eur. J. Oper. Res.* 287(2):633–652.

Kullman ND, Cousineau M, Goodson JC, Mendoza JE (2022) Dynamic ride-hailing with electric vehicles. *Transportation Sci.* 56(3):775–794.

Lang MA, Cleophas C, Ehmke JF (2021) Anticipative dynamic slotting for attended home deliveries. *Oper. Res. Forum* 2(4):1–39.

Laud AD (2004) Theory and application of reward shaping in reinforcement learning. PhD thesis, University of Illinois at Urbana-Champaign, Urbana.

Lebedev D, Goulart P, Margellos K (2021) A dynamic programming framework for optimal delivery time slot pricing. *Eur. J. Oper. Res.* 292(2):456–468.

Lebedev D, Margellos K, Goulart P 2020 Approximate dynamic programming for delivery time slot pricing: A sensitivity analysis. Working paper, University of Oxford, Oxford, UK.

Lee TC, Hersh M (1993) A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Sci.* 27(3):252–265.

Li Y, Archetti C, Ljubic I (2024) Emerging optimization problems for distribution in same-day delivery. Working paper, ESSEC Business School, Cergy-Pontoise, France.

Mackert J (2019) Choice-based dynamic time slot management in attended home delivery. *Comput. Industry Engrg.* 129:333–345.

Maddah B, Moussawi-Haidar L, El-Taha M, Rida H (2010) Dynamic cruise ship revenue management. *Eur. J. Oper. Res.* 207(1):445–455.

Powell WB (2011) *Approximate Dynamic Programming: Solving the Curses of Dimensionality* (John Wiley & Sons, Hoboken, NJ).

Powell WB (2022) *Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions* (John Wiley & Sons, Hoboken, NJ).

Powell WB, Simao HP, Bouzaiene-Ayari B (2012) Approximate dynamic programming in transportation and logistics: A unified framework. *EURO J. Transportation Logist.* 1(3):237–284.

Prokhorchuk A, Dauwels J, Jaillet P (2019) Stochastic dynamic pricing for same-day delivery routing. Working paper, Nanyang Technological University, Singapore.

Quante R, Fleischmann M, Meyr H (2009) A stochastic dynamic programming approach to revenue management in a make-to-stock production system. Working paper, Vienna University of Economics and Business, Vienna.

Russell SJ, Zimdars A (2003) Q-decomposition for reinforcement learning agents. Fawcett T, Mishra N, eds. *Proc. 20th Internat. Conf. Machine Learn.* (AAAI Press, Washington, DC), 656–663.

Snoeck A, Merchán D, Winkenbach M (2020) Revenue management in last-mile delivery: State-of-the-art and future research directions. *Transportation Res. Proc.* 46:109–116.

Soeffker N, Ulmer MW, Mattfeld DC (2022) Stochastic dynamic vehicle routing in the light of prescriptive analytics: A review. *Eur. J. Oper. Res.* 298(3):801–820.

Strauss AK, Gülpinar N, Zheng Y (2021) Dynamic pricing of flexible time slots for attended home delivery. *Eur. J. Oper. Res.* 294(3):1022–1041.

Strauss AK, Klein R, Steinhardt C (2018) A review of choice-based revenue management: Theory and methods. *Eur. J. Oper. Res.* 271(2):375–387.

Subramanian J, Stidham S Jr, Lautenbacher CJ (1999) Airline yield management with overbooking, cancellations, and no-shows. *Transportation Sci.* 33(2):147–167.

Talluri K, Van Ryzin G (2004) *The Theory and Practice of Revenue Management* (Springer, Boston).

Ulmer MW (2020) Dynamic pricing and routing for same-day delivery. *Transportation Sci.* 54(4):1016–1033.

Ulmer MW, Goodson JC, Mattfeld DC, Hennig M (2019) Offline–online approximate dynamic programming for dynamic vehicle routing with stochastic requests. *Transportation Sci.* 53(1):185–202.

Ulmer MW, Goodson JC, Mattfeld DC, Thomas BW (2020) On modeling stochastic dynamic vehicle routing problems. *EURO J. Transportation Logist.* 9(2):100008.

Van Seijen H, Fatemi M, Romoff J, Laroche R, Barnes T, Tsang J (2017) Hybrid reward architecture for reinforcement learning. Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Advance Neural Information Processing Systems* (Neural Information Processing Systems Foundation, Inc. (NeurIPS), San Diego, CA), 5392–5402.

Vinsensius A, Wang Y, Chew EP, Lee LH (2020) Dynamic incentive mechanism for delivery slot management in e-commerce attended home delivery. *Transportation Sci.* 54(3):567–587.

Waßmuth K, Köhler C, Agatz N, Fleischmann M (2023) Demand management for attended home delivery: A literature review. *Eur. J. Oper. Res.* 311(3):801–815.

Weatherford LR, Bodily SE (1992) A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking, and pricing. *Oper. Res.* 40(5):831–844.

Xu L, Wang Q, Huang S (2015) Dynamic order acceptance and scheduling problem with sequence-dependent setup time. *Internat. J. Production Res.* 53(19):5797–5808.

Yang X, Strauss AK (2017) An approximate dynamic programming approach to attended home delivery management. *Eur. J. Oper. Res.* 263(3):935–945.

Yang X, Strauss AK, Currie CS, Eglese R (2016) Choice-based demand management and vehicle routing in e-fulfillment. *Transportation Sci.* 50(2):473–488.