Convexity and Feedback in Approximate Dynamic Programming for Delivery Time Slot Pricing

Denis Lebedev, Kostas Margellos, *Member, IEEE*, and Paul Goulart, *Member, IEEE*

***Abstract*— We consider the revenue management problem of finding profit-maximizing prices for delivery time slots in the context of attended home delivery. This multistage optimal control problem admits a dynamic programming (DP) formulation that is intractable for realistic problem sizes due to the so-called "curse of dimensionality." We therefore study three approximate DP algorithms both from a numerical and control-theoretical perspective. Our analysis is based on real-world data, from which we generate multiple scenarios to stress-test the robustness of the pricing policies to errors in model parameter estimates. Our theoretical analysis and numerical benchmark tests indicate that one of these algorithms, namely gradient-bounded DP, dominates the others with respect to computation time and profit-generation capabilities of the pricing policies that it generates.**

***Index Terms*— Approximate dynamic programming (DP), attended home delivery, revenue management.**

I. Introduction

ONLINE grocery sales have been on the rise for the past few years. U.S. households are predicted to spend up to $133.8 billion per year on online grocery shopping according to GlobalData [13]. However, one of the main impediments for growth is the increased cost of home delivery compared with the logistics of brick-and-mortar supermarkets. A further logistical problem for online supermarkets is that they have to fulfill attended home delivery, that is, to deliver groceries to customers in pre-agreed delivery time windows. To this end, customers are asked to select a delivery time window as part of their purchase on a sales website. From the viewpoint of the company, this poses an optimization question: How should one adjust prices for delivery slots over time to maximize profits, taking into account how customers respond to price changes and how customer choice affects delivery costs? We call this the revenue management problem in attended home delivery.

Broadly speaking, revenue management problems in attended home delivery can be viewed as optimal control problems. The dynamics of customers choosing delivery time windows on the booking website form the plant that we seek to control. We can measure the customer choice behavior by keeping track of placed orders, which we treat as *states*. An optimal control law would then use information from the states to update delivery slot prices, which serve as control inputs to our plant, as shown schematically in Fig. 1. In principle, the exact state of orders is high-dimensional, for example, since it represents locations of all customers and their chosen delivery time slot. For industry-sized problems, the number of states becomes prohibitively large to compute the optimal pricing policy exactly. Therefore, simplified models have been proposed in the literature (see [15] for an overview).

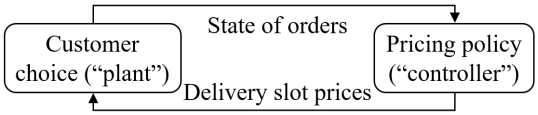

Fig. 1. Feedback control view of the pricing problem.

In this brief, we focus on the state-space representation of [18] and [19]. In this model, we split the delivery area into several sub-areas, each of which is served by a single delivery vehicle. We can then solve the optimal delivery slot pricing problem for each delivery sub-area separately. In that case, the dimension of the state is the number of delivery time slots of any delivery day. The state of orders can then be modeled as a vector whose entries represent the number of deliveries for every delivery time slot in a particular sub-area.

For this problem, Yang and Strauss [18] proposes a dynamic programming (DP) formulation and an approximate DP scheme. Their algorithm approximates the exact value function of the DP as an affine function in the vector of states. In this brief, we show that this method results in an open loop controller, thus motivating the use of nonlinear value function approximations, which provide state feedback, as suggested in [6] and [18]. The recently developed approximate DP algorithm in [20] was derived outside the attended home delivery literature and it provides such nonlinear approximate value functions. However, we show that this algorithm imposes severe practical difficulties in the computation of the optimal control policy, since it requires solving non-convex optimization problems.

Finally, we show our gradient-bounded DP algorithm from [9] that overcomes the limitations of the other two algorithms, since it provides nonlinear value function approximations, which can be computed using convex optimization. We demonstrate its efficacy in numerical examples based on data from [18], in which we benchmark the performance of all three algorithms in the case of both exact and incomplete knowledge (due to estimation errors) of model parameter values. This comparative analysis is new for this problem formulation and hence complements the numerical studies on attended home delivery conducted on different formulations, for example, in [6]. Heuristic algorithms are also examined in the literature, for example, [18, Sec. 7] implements a myopic

Manuscript received April 2, 2021; accepted June 4, 2021. Date of publication July 15, 2021; date of current version February 10, 2022. Manuscript received in final form June 28, 2021. This work was supported by the SIA Food Union Management. Recommended by Associate Editor W. Zhang. *(Corresponding author: Denis Lebedev.)*

The authors are with the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, U.K. (e-mail: denis.lebedev@eng.ox.ac.uk; kostas.margellos@eng.ox.ac.uk; paul.goulart@eng.ox.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCST.2021.3093648.

Digital Object Identifier 10.1109/TCST.2021.3093648

policy, which maximizes the stage revenue under simplified opportunity costs.

The structure of the remaining brief is as follows: We define the optimization problem and its DP formulation in Section II. Section III presents a prototype, sample-based approximate DP algorithm and we explain how the three algorithms considered in this brief are special cases of it. In that section, we also elaborate the main theoretical limitations of the first two algorithms and how the third overcomes these. Since the profits generated by all three algorithms are random variables, we explain how we quantify their profit-generation performance in Section IV. We then provide numerical evidence on how the gradient-bounded DP algorithm outperforms the two other algorithms in terms of both its profit-generation capabilities and computation time in Section V. Finally, we conclude this brief and provide directions for future research in Section VI.

Notation: Given some $s \in \mathbb{N}$, let 1_s be a column vector of all zeros apart from the sth entry, which equals 1. We adopt the convention that 1_0 is a vector of zeros. Let $\mathbf{1}$ denote a vector of 1's. Let $\langle \cdot, \cdot \rangle$ denote the standard inner product. Let $\mathbb{E}$ denote the expectation operator. Let $\Pr(\cdot)$ denote the probability of its argument. Let $\mathbb{1}(\cdot)$ denote the indicator function.

II. Problem Statement

A. Multistage Optimal Control Problem Formulation

Revenue management in attended home delivery can be formulated as the following multistage optimal control problem for any delivery sub-area served by a single delivery vehicle: Customers are assumed to be allowed to make bookings in a finite time horizon and there are only a finite number of times that the online vendor can change delivery slot prices. Therefore, we consider a finite and discrete time horizon $T := \{1, 2, \ldots, \bar{t}\}$. There is an additional time step $\bar{t} + 1$, at which no bookings happen anymore, which we will use to define the terminal condition of the problem. Suppose that the delivery day is split into n delivery time slots. Denote the set of delivery time slots by $S := \{1, 2, \ldots, n\}$. As mentioned in Section I, we focus on an aggregated state-space representation, where for any time step $t \in T \cup \{\bar{t} + 1\}$, we define a state vector $x_t \in X \subset \mathbb{Z}^n$, whose entries are the number of orders placed in the respective delivery time slots. The set X is defined by the maximum state vector $\bar{x}$, that is, $X := \{x_t \in \mathbb{Z}^n \mid 0 \leq x_t \leq \bar{x}\}$. For any $t \in T$, we define the delivery charge vector $d_t := [d_{1,t}, d_{2,t}, \ldots, d_{n,t}]^\intercal$. Let the set of admissible delivery charge vectors be $D := \{d_t \mid d_{s,t} \in [\underline{d}, \bar{d}\,] \cup \{\infty\} \text{ for all } s \in S\}$.

For any $s \in S$, define the transition probability between two states x_t and $x_{t+1} = x_t + 1_s$ under delivery price vector d_t as $P_s(d_t)$, where we require $P_s(d_t) \geq 0$ for all $(s, d_t) \in S \times D$. We require that $\sum_{s \in S} P_s(d_t) < 1$, such that the probability of the customer not choosing *any* slot is defined as $P_0(d_t) = 1 - \sum_{s \in S} P_s(d_t)$. This implies that transitions from x_t to x_{t+1} are only possible in the positive direction and by at most a unit step along one dimension. We restrict $P_s(d_t) = 0$ if $x_t + 1_s \notin X$, that is, we do not allow infeasible orders. Such models are typical for order-taking processes (see [1], [16], [18], and [19]). Note that multi-product pricing, for example, as in [5], is stated in terms of inventory which depletes over time (down to 0), rather than orders which accumulate (up to $\bar{x}$). We assume that customer choice follows a multinomial logit model, like in [5], [18], and [19], that is,

$$P_s(d_t) := \frac{\exp(\beta_c + \beta_s + \beta_d d_{s,t})}{\sum_{k \in S} \exp(\beta_c + \beta_k + \beta_d d_{k,t}) + 1} \tag{1}$$

for all $(s, d) \in S \times D$, where $\beta_c \in \mathbb{R}$ denotes a constant offset, $\beta_s \in \mathbb{R}$ represents a measure of the popularity for all delivery slots, and $\beta_d < 0$ is a parameter for the price sensitivity. Note that the no-purchase utility is normalized to zero, that is, for the no-purchase "slot" $s = 0$, we have a no-delivery "charge" $d_{0,t} = 0$, such that $\beta_c + \beta_0 + \beta_d d_{0,t} = \beta_c + \beta_0 = 0$ and hence, the 1 in the denominator of (1) arises from $\exp(\beta_c + \beta_0) = 1$. Furthermore, note that the constant offset β_c is not necessary, since it can be absorbed in the $\{\beta_s\}_{s \in S \cup \{0\}}$ parameters. However, β_c is often kept in practice to normalize one of the $\{\beta_s\}_{s \in S \cup \{0\}}$ parameters to zero (see, e.g. [18]). Finally, we define an average revenue per order $r \in \mathbb{R}$, an expected customer arrival rate (on the booking system) per time step $\lambda \in (0, 1]$, and an approximate delivery cost function $C : \mathbb{Z}^n \to \mathbb{R} \cup \infty$, which we assume is Lipschitz continuous. Infinite delivery costs indicate infeasible states. We construct a multistage optimal control problem in the form

$$\max_{\{d_t \in D\}_{t=1}^{\bar{t}}} \mathbb{E}\left[-C(x_{\bar{t}+1}) + \sum_{t \in T} \langle x_{t+1} - x_t, d_t + r \rangle \right]$$
$$\text{subject to } x_{t+1} = x_t + \xi_t, \quad \text{for all } t \in T, \; x_1 = 0 \tag{2}$$

where $\mathbb{E}$ is the expectation operator associated with the probability distribution of $\xi_t \in \{0, 1\}^n$, defined as follows: For all $(s, t) \in S \times T$, $\xi_t = 1_s$ with probability $\lambda P_s(d_t)$ and $\xi_t = 0$ with probability $1 - \sum_{s \in S} \lambda P_s(d_t)$. From an economic perspective, the objective value is the total expected operational contribution margin, that is, revenue from sales and delivery charges minus delivery costs. For simplicity, we will refer to this as the expected profit that we seek to maximize.

B. DP Formulation

The objective function in (2) is separable across stages and there is stage-wise coupling of the states x_t for all $t \in T$. These can only be decoupled if $\bar{x}$ is sufficiently large, such that $x_t < \bar{x}$, for all $t \in T$, and if C is a linear function. In the general case, however, we can derive the following DP recursion, analogous to [18], by introducing the value function $V_t : \mathbb{Z}^n \to \mathbb{R} \cup -\infty$, which represents the expected profit-to-go for any state–time pair $(x, t) \in X \times T$, as shown in the following equation:

$$V_t(x) := \max_{d \in D} \Bigg\{ \lambda \sum_{s \in S} P_s(d)(r + d_s + V_{t+1}(x + 1_s)) + \left(1 - \lambda \sum_{s \in S} P_s(d)\right) V_{t+1}(x) \Bigg\} \quad \forall (x, t) \in X \times T$$
$$V_{\bar{t}+1}(x) = -C(x) \quad \forall x \in X. \tag{3}$$

Algorithm 1 Prototype, Sample-Based Approximation

1: Initialise parameters: $X, D, P_s, T, r, \lambda, C$ and $i_{\max}$
2: Initialise $Q_t^0(x) \leftarrow (\bar{d}+r)\langle \mathbf{1}, \bar{x}-x\rangle - C(\bar{x})$, for all $(x,t) \in X \times T$
3: Initialise $Q_{\bar{t}+1}^0(x) \leftarrow -C(x)$, for all $x \in X$
4: **for** $i \in \{1,2,\ldots,i_{\max}\}$ **do**
5: $\quad x_1^i \leftarrow \mathbf{0}$
6: $\quad$ **for** $t \in \{1,2,\ldots,\bar{t}\}$ **do** ▷ "Forward sweep"
7: $\quad\quad d_t^i \leftarrow d^*(x_t^i)$, the solution of (4)
8: $\quad\quad x_{t+1}^i \leftarrow x_t^i + \underset{x_{t+1}^i}{\text{sample}}\{P_s(d_t^i)\}$
9: $\quad$ **end for**
10: $\quad$ **for** $t \in \{\bar{t}, \bar{t}-1,\ldots,1\}$ **do** ▷ "Backward sweep"
11: $\quad\quad Q_t^i \leftarrow$ update Q_t^{i-1}
12: $\quad$ **end for**
13: **end for**
14: **return** $\{Q_t^{i_{\max}} \mid t \in \{1,2,\ldots,\bar{t}\}\}$

We assume that $V_t(x) = -\infty$ for all infeasible states $x \notin X$. Notice that we have dropped subscripts t for x and d to simplify notation and since the time step is evident from the value function. Furthermore, we adopt the convention that when, for any $s \in S$, it happens that $d_s = \infty$ and $V_{t+1}(x+1_s) = -\infty$, we have $P_s(d)(r + d_s + V_{t+1}(x+1_s)) = 0$. This corresponds to the additional profit of accepting an unavailable slot, which is undefined in (3), but practically it is zero. To represent the DP in (3) more compactly, we define the Bellman operator $\mathcal{T}$ through the relationship $V_t = \mathcal{T}V_{t+1}$, for all $t \in T$. It is not possible to solve the DP in (3) by direct computation of the value function for realistic problem instances due to the prohibitively large number of states. However, given an approximate value function, one can find approximately optimal prices at relatively low computational cost for the multinomial logit model using Newton's method [5].

III. Prototype, Sample-Based Approximation

A key to solving this revenue management problem is to approximate the value function in (3) effectively. A popular strategy, not only for this problem (see [6] and [18]), but also for other stochastic multistage problems (see [12] and [14]), is to use a sample-based approach and to refine the value function along states that are likely to be visited under the approximately optimal decision policy. Approximations can then be improved by iterating between generating samples and refining the value function along the sample paths obtained.

We first present a prototype, sample-based, iterative approximate DP procedure in Algorithm 1. The three algorithms that we investigate are special cases of this prototype algorithm and differ only in step 11. We detail how this step is computed for each of the individual specific algorithms in the sequel.

We first initialize all parameters of the DP in (3) (see step 1). Denote the maximum number of iterations by $i_{\max} \in \mathbb{N}$ and let $I := \{0,1,\ldots,i_{\max}\}$. Let the value function approximation be denoted by Q_t^i for all $(i,t) \in I \times T$. We could initialize Q_t^0 to any value as long as it does not violate any assumptions on the approximation algorithm used, as discussed further below. However, one effective way to satisfy the assumptions of all three algorithms considered, and to speed up computation, is to initialize Q_t^0 for all $t \in T$ using the unique fixed point of DP, V^* (see step 2). In [7], it is shown that under mild technical assumptions the fixed point is

$$V^*(x) := (\bar{d}+r)\langle \mathbf{1}, \bar{x}-x\rangle - C(\bar{x}), \quad \text{for all } x \in X. \tag{4}$$

Note that V^* is an upper bound to V_t for all $t \in T$, since $\mathcal{T}$ is a monotone operator (see [3, Ch. 3]). Furthermore, notice that the fixed point is affine in x and that the components of the gradient are given by $\bar{d}+r$. Using (3), we can compute the optimal delivery slot prices at the fixed point, which is $\bar{d}$ for all feasible slots. The intuition behind this is that the fixed point corresponds to the limit of the value function as t tends to $-\infty$. Hence, going backwards infinitely many time steps, the probability of selling out the entire delivery capacity tends to 1 for all prices $d \in D$. To maximize profits over an infinite horizon, one would charge customers the maximum admissible delivery charge $\bar{d}$ for all delivery time slots. Finally, we also initialize $Q_{\bar{t}+1}^i(x) := V_{\bar{t}+1}(x) = -C(x)$ for all $(x,i) \in X \times I$ (see step 3).

Now fix any iteration $i \in I \setminus \{0\}$. In each "forward sweep," we solve an approximate version of (3) forward in time by replacing V_t with its approximation Q_t^{i-1} (see step 7), that is,

$$d^*(x_t^i) := \underset{d \in D}{\operatorname{argmax}} \Bigg\{ \lambda \sum_{s \in S} P_s(d)\big(r + d_s + Q_{t+1}^{i-1}(x_t^i + 1_s)\big) + \Bigg(1 - \lambda \sum_{s \in S} P_s(d)\Bigg) Q_{t+1}^{i-1}(x_t^i) \Bigg\}. \tag{5}$$

Notice that [5, Th. 1] shows that for the multinomial choice model, the maximizers are unique for the above expression, hence we use equality in (5). Using (5), we compute a suboptimal d_t^i for all $t \in T$ and simulate state transitions by sampling from the transition probability distribution given the approximately optimal decisions (see step 8). This defines a sample path x_t^i for all $t \in T \cup \{\bar{t}+1\}$.

In each "backward sweep," we update the approximation using one of the three algorithms in this brief (see step 11). These two sequences—"forward sweep" and "backward sweep"—are repeated for $i_{\max}$ iterations. In Sections III-A–III-C, we describe the exact mechanisms of the three algorithms considered in this brief, making step 11 in Algorithm 1 explicit.

A. Affine Value Function Approximation Update

This approach, proposed in [18], approximates the value function by an affine function of the form

$$Q_t^i(x) := \gamma_0^i + (\bar{t}+1-t)\theta^i - \sum_s \gamma_s^i x_s \tag{6}$$

for all $(x,i,t) \in X \times I \times T$ and where γ_s^i, for all $s \in S \cup \{0\}$ and θ^i are real scalar parameters, for all $i \in I$. The updating rule in step 11 of Algorithm 1 is a gradient descent step to

minimize $(Q_t^i(x_{t+1}^i) - \mathcal{T}Q_{t+1}^i(x_{t+1}^i))^2$, which thus becomes

$$\begin{aligned}\gamma_0^i &= \gamma_0^{i-1} - \alpha_1\big(Q_t^{i-1}(x_{t+1}^i) - \mathcal{T}Q_{t+1}^{i-1}(x_{t+1}^i)\big)\\ \gamma_s^i &= \gamma_s^{i-1} - \alpha_2\big(Q_t^{i-1}(x_{t+1}^i) - \mathcal{T}Q_{t+1}^{i-1}(x_{t+1}^i)\big)x_{s,t+1}^i \quad \forall s \in S\\ \theta^i &= \theta^{i-1} - \alpha_3\big(Q_t^{i-1}(x_{t+1}^i) - \mathcal{T}Q_{t+1}^{i-1}(x_{t+1}^i)\big)(\bar{t}+1-t) \end{aligned} \tag{7}$$

where α_1, α_2, and α_3 are (positive) step sizes, chosen to be sufficiently small for convergence of the above iterative procedure (see, e.g. [2, Lemma 8.2]).

One important observation from a control perspective is that a value function approximation that is affine in x implies that the pricing control will have no state feedback for all states x such that $x + 1_s \in X$ for all $s \in S$. To see this, notice that (5) can be rewritten in terms of differences $Q_{t+1}^{i-1}(x_t^i) - Q_{t+1}^{i-1}(x_t^i + 1_s)$ for all $s \in S$ as follows:

$$\begin{aligned}&\operatorname*{argmax}_{d\in D}\Bigg\{\lambda\sum_{s\in S}P_s(d)(r + d_s\\ &\qquad + Q_{t+1}^{i-1}(x_t^i + 1_s) - Q_{t+1}^{i-1}(x_t^i)) + Q_{t+1}^{i-1}(x_t^i)\Bigg\}\\ &= \operatorname*{argmax}_{d\in D}\Bigg\{\lambda\sum_{s\in S}P_s(d)\big(r + d_s - \gamma_s^{i-1}\big) + Q_{t+1}^{i-1}(x_t^i)\Bigg\}\end{aligned}$$

for all $x \in X$, such that $x + 1_s \in X$ for all $s \in S$, and where we have first substituted for Q_{t+1}^{i-1} from (6). Notice that $Q_{t+1}^{i-1}(x_t^i)$ is independent of d and thus irrelevant for the argmax operator. Hence, the approximately optimal pricing policy does not depend on the state x_t^i for all $(x,t) \in X \times T$ such that $x + 1_s \in X$ for all $s \in S$. This ultimately means that the affine value function approximation generates a *feedforward* pricing policy, which is incapable of adjusting prices based on changes in the vector of orders. This insight also provides theoretical support for the suggestions of [6] and [18] to explore nonlinear value function approximations: The preceding discussion shows that nonlinear value functions are necessary in order to enable state feedback in the pricing policy.

B. Nonlinear Stochastic Dual DP Update

In contrast to the affine value function update above, nonlinear stochastic dual DP generates nonlinear value function approximations that make it possible to include state feedback in the pricing policy. Similar to [20] and [21], this update is computed in step 11 of Algorithm 1 as $Q_t^i \leftarrow \min\{H^*, Q_t^{i-1}\}$, where the minimum is taken pointwise and the so-called Lagrange dual cut H^* is defined as $H^*(x) := v^* - \langle \mu^*, x_{t+1}^i - x\rangle$, for all $x \in X$ and where

$$\begin{aligned}v^* := \min_{\mu\in\mathcal{M}}\max_{d\in D, z\in X}\Bigg\{&\lambda\sum_{s\in S}P_s(d)\big(r + d_s + Q_{t+1}^{i-1}(z + 1_s)\big)\\ &+\Bigg(1 - \lambda\sum_{s\in S}P_s(d)\Bigg)Q_{t+1}^{i-1}(z)\\ &+\big\langle\mu, x_{t+1}^i - z\big\rangle\Bigg\}\end{aligned} \tag{8}$$

and where μ^* is the minimizer of (8). In [20, Proposition 5], it is shown that the resulting value function approximation is an upper bound to the exact value function if we additionally assume that the initial approximation Q_t^0 is an upper bound to V_t for all $t \in T$. The formulation in [20] differs from (8), since it includes an additional regularization term. See [10] for an equivalent proof for this formulation.

From a control perspective, the benefit of having a nonlinear value function approximation, in comparison with the affine value function approximation from Section III-A, comes at a different cost: The problem of finding the optimal cut coefficients μ in (8) is a non-convex optimization problem and there are consequently no guarantees that it can be solved to global optimality. For the particular form of the problem in this brief, we can find the cut coefficients from a reformulation of the problem, resulting in a bi-concave objective function that can be exploited to solve this problem as outlined in [10, Appendix A]. Nevertheless, since global optimality is required to ensure that the approximate value function constitutes an upper bound on the exact value function, in practice we cannot easily guarantee that all approximations are indeed upper bounds on the exact value function. We illustrate how this results in computational problems in Section V.

C. Gradient-Bounded DP Update

Gradient-bounded DP was introduced in [9] for the specific application of revenue management in attended home delivery. This method makes two assumptions on the exact value function of the DP.

First, the value function must be concave extensible. This means that its concave closure $\tilde{V}_t : \mathbb{R}^n \to \mathbb{R}$, defined as the smallest concave upper bound on the exact value function, coincides with the exact value function for all state–time pairs, that is, $\tilde{V}_t(x) = V_t(x)$ for all $(x,t) \in X \times (T \cup \{\bar{t}+1\})$. Second, the exact value function must be submodular, that is,

$$V_t(\max(y_1, y_2)) + V_t(\min(y_1, y_2)) \le V_t(y_1) + V_t(y_2) \tag{9}$$

for all $(y_1, y_2, t) \in X \times X \times (T \cup \{\bar{t}+1\})$.

These two assumptions result in a particular segmentation of the convex hull of V_t: For any $t \in T$, construct the unique hyperplane H through the set of pairs $(y, V_t(y))$ for all $y \in Y_+(x_{t+1}^i) := \{x_{t+1}^i + 1_s\}_{s\in(S\cup\{0\})}$. Then H is a separating hyperplane, that is, $H(x) \ge V_t(x)$ for all $x \in X$, where the inequality holds with equality for all $y \in Y_+(x)$. The gradient-bounded DP algorithm exploits this property. We refer the interested reader to [9] for details on the above-mentioned assumptions and to [11, Th. 2 and Proposition 4(ii)] for proofs that these assumptions hold for the revenue management problem under study.

For gradient-bounded DP, let the value function approximation Q_t^i for all $(i,t) \in I \times T$ be the pointwise minimum of a finite number of affine functions, that is, $Q_t^i(x) := \min_{j\in\{0,1,\ldots,i\}} H_t^j(x)$, for all $x \in X$, where $H_t^j : X \mapsto \mathbb{R}$ describes a hyperplane, that is, $H_t^j(x) := \langle a_t^j, x\rangle + b_t^j$, for all $x \in X$, with $a_t^j \in \mathbb{R}^n, b_t^j \in \mathbb{R}$ for all $(t,j) \in T \times I$. Furthermore, this approximation is an upper bound on the exact value function, that is, $Q_t^i(x) \ge V_t(x)$ for all $(x,t,i) \in X \times T \times I$. To this end, it is important to initialize Q_t^0 for all $t \in T$ at an upper bound. The gradient-bounded DP update then

Algorithm 2 Gradient-Bounded DP Update

1: $Z(x^i_{t+1}) \leftarrow \{x^i_{t+1} + 1_s + 1'_s\}_{s\in S\cup\{\mathbf{0}\}, s'\in S\cup\{\mathbf{0}\}}$
2: **if** Q^{i-1}_{t+1} is submodular on $Z(x^i_{t+1})$ **then**
3: $H^* \leftarrow$ unique hyperplane through $\left\{\left(y, (\mathcal{T}Q^{i-1}_{t+1})(y)\right)\right\}_{y\in Y_+(x^i_{t+1})}$
4: **else**
5: $j^* \in \underset{j\in J^{i-1}_{t+1}}{\operatorname{argmin}} \left\{\left(\mathcal{T}H^{j-1}_{t+1}\right)(x^i_{t+1})\right\}$
6: $H^* \leftarrow \mathcal{T}H^{j^*-1}_{t+1}$
7: **end if**
8: $Q^i_t \leftarrow \min\left\{H^*, Q^{i-1}_t\right\}$

ensures that the approximate value functions remain upper bounds to the exact value function for all iterations $i \in I$, as shown in [9, Proposition 1]. Notice that both nonlinear stochastic dual DP and gradient-bounded DP are similar in that they compute approximate value functions as the pointwise minima of a finite number of hyperplanes, which are upper bounds to the exact value function. In step 11 of Algorithm 1, gradient-bounded DP generates updates for the approximate value function, as shown in Algorithm 2 and explained further below.

Fix any iteration $i \in I$. We first check if Q^{i-1}_{t+1} is submodular [see (9)] on the set $Z(x^i_{t+1}) := \{x^i_{t+1} + 1_s + 1_{s'}\}$, for all $(s, s') \in (S \cup \{0\}) \times (S \cup \{0\})$, that is, if and only if $0 \leq Q^{i-1}_{t+1}(y_1) + Q^{i-1}_{t+1}(y_2) - Q^{i-1}_{t+1}(\min\{y_1, y_2\}) - Q^{i-1}_{t+1}(\max\{y_1, y_2\})$ holds for all $(y_1, y_2) \in Z(x^i_{t+1}) \times Z(x^i_{t+1})$ (see steps 1 and 2). Note that this is not necessarily the case for the approximate value function, even if the exact value function is submodular. We then distinguish between two cases:

Case I: If Q^{i-1}_{t+1} is submodular on $Z(x_{t+1^i})$, we locally compute the exact DP stage problem on the set $Y_+(x_t + 1)^{i-1}$, that is, $\{\mathcal{T}Q^{i-1}_{t+1}(y)\}_{y\in Y_+(x^i_{t+1})}$, to construct the hyperplane through $\{(y, (\mathcal{T}Q^{i-1}_{t+1})(y))\}_{y\in Y_+(x^i_{t+1})}$ (see step 3).

Case II: If Q^{i-1}_{t+1} is *not* submodular on $Z(x_{t+1^i})$, we need to compute a submodular upper bound on Q^{i-1}_{t+1}, which is readily given by the hyperplanes from which Q^{i-1}_{t+1} is constructed. Therefore, we select the hyperplane $H^{j^*-1}_{t+1}$ that minimizes the value at the evaluation point x^i_t, that is, $Q^i_t = \min\{\mathcal{T}H^{j^*-1}_{t+1}, Q^{i-1}_t\}$, where $j^* \in \operatorname{argmin}_{j\in J^{i-1}_{t+1}}\{(\mathcal{T}H^{j-1}_{t+1})(x^i_{t+1})\}$ and where J^{i-1}_{t+1} is the set of supporting hyperplanes, that is, $J^{i-1}_{t+1}(x) := \operatorname{argmin}_{j\in\{0,1,\ldots,i-1\}} H^j_{t+1}(x)$, for all $(i, t, x) \in I \times T \times X$ (see steps 5 and 6). Therefore, this creates the locally tightest upper bound. Finally, we take the pointwise minimum of the approximate value function at the previous iteration Q^{i-1}_t and the newly created hyperplane H^* to obtain the new value function approximation (see step 8).

Similar to the nonlinear stochastic dual DP update from Section III-B and in contrast to the affine value function approximation update from Section III-A, the approximate value function generated by the gradient-bounded DP update is nonlinear in x as it is given by the pointwise minimum of affine functions in x. Assuming that the initializer is an upper bound on the exact value function, which can be satisfied if we choose the fixed point for this purpose (as discussed at the beginning of Section III), the approximate value function is an upper bound to the exact value function, as shown in [9, Proposition 1]. Finally, the advantage of gradient-bounded DP over nonlinear stochastic dual DP is that only convex optimization problems must be solved to compute the update, which makes gradient-bounded DP more resilient against computational stability problems than nonlinear stochastic dual DP, as shown in Section V.

IV. Profit-Generation Performance Criterion

Since the profits that all three algorithms generate are random variables, we can quantify their performance with probabilistic guarantees by performing validation runs, that is, by simulating customer decisions forward in time and pricing based on the most refined approximate value function. Let the profit that we obtain in each of $k_{\max}$ validation runs be $l_v(k)$ for all $k \in K := \{1, \ldots, k_{\max}\}$. Let $[l_-, l_+]$ denote the (finite) support of the distribution of $l_v(k)$ for any $k \in K$. In our case, $l_+ = V^*(0) = (\bar{d} + r)\langle \mathbf{1}, \bar{x}\rangle - C(\bar{x})$, where we use the fixed point from (4) and $l_- = -C(0)$. We then compute the empirical mean as $\bar{l}_v := k^{-1}_{\max}\sum_{k=1}^{k_{\max}} l_v(k)$ and empirical standard error as $\sigma_v := [k^{-1}_{\max}\sum_{k=1}^{k_{\max}}(l_v(k) - \bar{l}_v)^2]^{1/2}$. For *any* of the three algorithms considered, we can then quantify the performance of a pricing policy using the maximum of the two following bounds, presented in [8, Proposition 7], which state the expected profit which can be guaranteed with confidence $(1-\alpha) \in (0, 1)$ after observing $k_{\max}$ validation samples. Recall that $\mathbb{E}$ denotes the expectation operator and that $\Pr(\cdot)$ denotes the probability of its argument.

Proposition 1: Fix any significance level $\alpha \in (0, 1)$. Then $\Pr(\mathbb{E}\bar{l}_v \geq l^*) \geq 1 - \alpha$, for all $l^* \in \{l_A, l_B\}$, where

$$l_A := \bar{l}_v - \sqrt{\frac{2\sigma_v \ln\left(\frac{2}{\alpha}\right)}{k_{\max}}} - \frac{7(l_+ - l_-)\ln\left(\frac{2}{\alpha}\right)}{3(k_{\max} - 1)} \tag{10a}$$

and

$$l_B := \int_{l=0}^{\infty} 1 - \min\left\{1, F_K(l) + \sqrt{\frac{\ln\left(\frac{1}{\alpha}\right)}{2k_{\max}}}\right\} dl \tag{10b}$$

where F_K denotes the empirical cumulative distribution function of $\{l_v(k)\}_{k\in K}$, that is, $F_K(l) := k^{-1}_{\max}\sum_{k\in K}\mathbb{1}(l_v(k) \geq l)$, where $\mathbb{1}$ denotes the indicator function.

The proof can be found in [8, Appendix A.6]. Strictly speaking, we assume non-positive fixed costs $C(0)$ for the bounds to hold (see [8, Assumption 5]). Hence, we set $l_- = -C(0) = 0$ for our case study. However, since the fixed costs do not impact the pricing policy of any of the algorithms considered, these become irrelevant for our analysis.

V. Numerical Examples

In Sections V-A and V-B, we present a numerical analysis that compares the three value function approximation algorithms stated and analyzed in Section III. To this end, we generate particular instances of the revenue management problem in attended home delivery presented in Section II. We use the parameter values in [18] as a base case and

TABLE I
PARAMETERS OF THE EXACT MODEL

r	$\underline{d}$	$\bar{d}$	n	λ	β_c	β_d
£34.53	£0	£10	17	0.8	-2.5087	-0.0766
$\{\beta_s\}_{s\in S}$	$\{-1.0305, -0.3591, 0.3107, 0.5922, 0.6154, 0.0796, 0.5356, -0.2415, -0.6286, -1.6736, -0.4351, -0.161, 0, 0.2533, 0.0736, 0.562, 0.2346\}$					

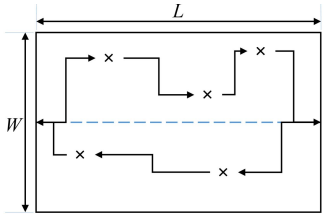

Fig. 2. Delivery sub-area with length L and width W. Crosses indicate customer locations served in a particular time slot.

modify these parameters to simulate various scenarios and conduct a sensitivity analysis. In Section V-A, we analyze the performance of the three algorithms under the assumption that the model parameters are known accurately. Then, in Section V-B, we simulate how well the algorithms perform when they are trained on the data in V-A, but tested on scenarios where the customer choice parameters differ from the model.

A. Exact Model Analysis

In this section, we adapt the numerical case study parameters from [18] to arrive at the setup defined in Table I. We use the same step sizes for the affine value function update (see (7) in Section III-A) as in [18], that is, $\alpha_1 := 0.0001$, $\alpha_2 := 0.00025$, and $\alpha_3 := 0.00014$. Additionally, we consider two parameters, which we vary as described further below.

First, we vary the delivery capacity of each time slot by varying the size of the delivery sub-area under consideration to simulate urban, suburban, and rural scenarios. Each scenario has a different value of capacity per delivery time slot $\bar{x}$, which influences the variable delivery cost. In practice, the mapping between the characteristics of the delivery sub-area and the delivery capacity for all delivery time slots depends on many additional factors, including infrastructure, traffic, and weather conditions. However, for the purpose of our numerical analysis, we use a simplified model from [4] and [18], which derives the delivery slot capacity as follows: Suppose that the delivery sub-area is rectangular and has length L and width W, as shown in Fig. 2.

Furthermore, suppose an average delivery truck velocity of $\omega = 25$ mph and a cost per mile of $\xi =$ £0.25. We assume that in each delivery time slot, the truck travels back and forth along the length L of the delivery sub-area; along the half-width $[0, W/2]$ of the delivery sub-area in one direction and along the other half-width $[W/2, W]$ in the other direction. We then assume that customer locations are random, uniformly distributed in the delivery sub-area and that the truck travels Manhattan distances. This implies that the average distance traveled between two customers along the axis aligned with the width of the sub-area is one-third times the half-width $W/2$.

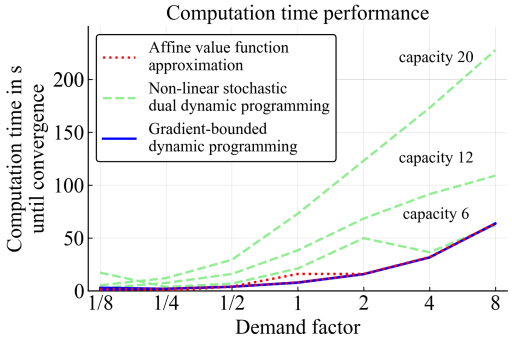

Fig. 3. Computational time to reach 95% of the maximum expected profit with 99% confidence for each of the three algorithms against demand factor and delivery slot capacities.

This results in a variable delivery cost of $c_{\text{var}} := \xi \times W/6$, as shown in [4]. We find W from the condition that, in any delivery slot, the delivery truck must be able to make $\bar{x}$ deliveries and an additional assumption that $L = 2W$. The last choice is arbitrary and our results do not change qualitatively for other ratios L/W. The total traveling distance in every delivery time slot thus becomes $\omega \times 1\text{h} = 2L + \bar{x}W/6 \Rightarrow W = \omega/(4 + \bar{x}/6)$. This finally implies that $c_{\text{var}} := \xi\omega/(24 + \bar{x})$.

Second, we vary the expected demand, that is, the expected number of customer arrivals on the booking website, given by $\lambda\bar{t}$. Since it is reasonable to keep $\lambda \approx 0.8$ for customer choice parameter estimation purposes (see [19]), we fix $\lambda = 0.8$ and vary $\bar{t}$ to achieve a total demand level corresponding to $\phi n\bar{x}$, where $n\bar{x}$ is the total delivery capacity for all slots and $\phi \in \mathbb{R}$ is a demand factor, such that $\phi \in \Phi := \{1/8, 1/4, 1/2, 1, 2, 4, 8\}$. Hence, $\bar{t} \approx \phi n\bar{x}/\lambda$, for all $\phi \in \Phi$ and where the approximation comes from rounding $\bar{t}$ to the nearest integer. For all scenarios, we compute the profit that is reached in expectation with confidence 99%, by computing 100 validation samples for each scenario and each algorithm and using the tighter of the two bounds from Section IV.

In general, we observe that the nonlinear stochastic dual DP algorithm produces higher expected profits than the affine value function approximation algorithm, while taking significantly more time to compute a good solution. However, the gradient-bounded DP algorithm exhibits the strengths of other algorithms: very similar profit generation performance to nonlinear stochastic dual DP and similar speed to the affine value function approximation algorithm. For example, Fig. 3 shows the computation time that it takes for the three algorithms to reach at least 95% of their maximum expected profit with 99% confidence for various demand factors and delivery time slot capacities. Nonlinear stochastic dual DP always takes longest to compute out of the three algorithms. Computation time also tends to increase for nonlinear stochastic dual DP as demand factor or slot capacity increase. For capacity 20, it takes about four times longer to compute the solution for demand factor 8, compared with the other algorithms. This time factor increases to about 10 as the demand factor decreases to 1/8.

Affine value function approximation and gradient-bounded DP take similar time to converge to their respective optimal solutions. Computation time does not vary across slot

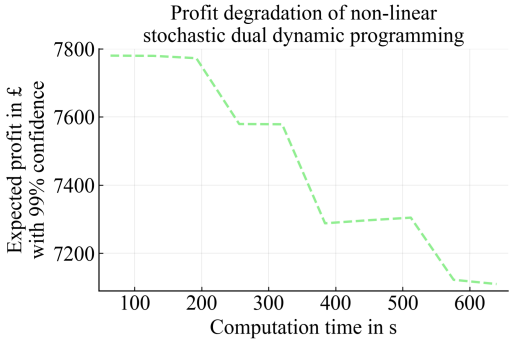

Fig. 4. Expected profits of nonlinear stochastic dual DP (demand factor 8, slot capacity 12) decrease over time.

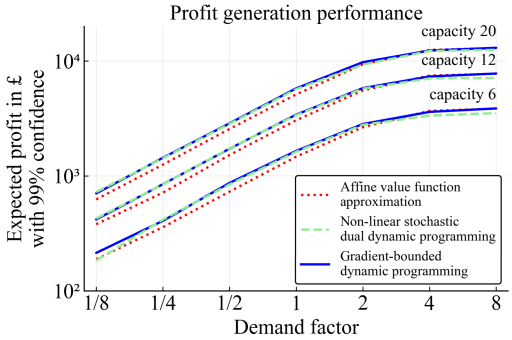

Fig. 5. Expected profits with 99% confidence of the three algorithms against demand factor, for all delivery slot capacities.

capacities for these algorithms for all but one scenario: For demand factor 1 and slot capacity 6, affine value function approximation takes twice as long to converge compared with gradient-bounded DP. A possible explanation is that, for this particular scenario, it might be computationally involved to find the optimal affine value function approximation since for a medium demand factor it is difficult to find a single affine value function approximation that works well for all sample paths. Some slots might sell out, some might not, which increases the need for a more flexible solution that gradient-bounded DP can provide.

Another issue observed is that the nonlinear stochastic dual DP algorithm becomes computationally unstable under certain conditions. For example, for demand factor 8 and slot capacity 12, its profit-generation performance decreases over time as can be seen in Fig. 4. This might appear counter-intuitive at first, but is in line with our theoretical analysis from Section III-B: We conjecture that this is due to the difficulty of finding global maxima of non-convex optimization problems. If the algorithm converges to a local maximum, the value function approximation is no longer guaranteed to be an upper bound on the exact value function. Over time, this then leads to a compounding of errors caused by suboptimality, that is, instead of increasing, the expected profit decreases as more cuts are added to the approximate value function. A practical way to circumvent this problem is to compute the expected profit with 99% confidence after each iteration and to pick the iteration which produces the maximum expected profit with 99% confidence. In the example of Fig. 4, the best solution is found after the first iteration—the optimal policy is dominated by pricing all slots at the maximum charge $\bar{d}$ for all time steps, since the high demand factor 8 almost guarantees that all slots will be sold out for any choice of admissible prices. Hence, over time invalid cuts accumulate, which results in a degradation of the profit performance.

Comparing the expected profits obtained between the three algorithms, we observe that gradient-bounded DP either generates the highest expected profit with 99% confidence or is within 1% of the optimal value, when the demand factor is so high that demand saturates and all three algorithms perform very similarly. This saturation behavior can be seen in Fig. 5, where we also show that for demand factors 1 and lower, gradient-bounded DP produces between 10% and 15% more expected profit with 99% confidence than affine value function approximation. At the same time, gradient-bounded DP performs similar to nonlinear stochastic dual DP in most scenarios. However, gradient-bounded DP generates up to 10% more profit than nonlinear stochastic dual DP for small demand factor $1/8$ and capacity 6 as well as for large demand factor 8 across all slot capacities.

Overall, we conclude that gradient-bounded DP performs best in this exact model experiment, because it outperforms affine value function approximation in terms of profit generation while being similarly fast and at the same time, gradient-bounded DP is more than four times faster than nonlinear stochastic dual DP while generating very similar profit.

B. Parameter Sensitivity Analysis

We assume in Section V-A that the customer choice model parameters are known exactly, which is not the case in practice. Hence, we now investigate how well the pricing policies obtained by the three algorithms in Section V-A perform on perturbed models. To this end, we corrupt the parameter estimates β_c, β_d, and $\{\beta_s\}_{s\in S\cup\{0\}}$ by additive Gaussian noise. This choice of distribution is justified because, in the limit as the number of data points used for estimating the customer choice parameters tends to infinity, the error between estimated and true customer choice parameter value vector is a Gaussian with zero mean [17, Ch. 8.6].

We consider three scenarios in which we set the variance of the Gaussian to $\sigma^2 \in \{0.01, 0.1, 1\}$. With these noise levels, we sample customer choice parameters, which we hold fixed for all validation runs. Note that we do not have to worry about normalizing the probability distribution, since the multinomial choice model is normalized for all possible parameter values. The numerical values used in our analysis are documented in [10, Appendix B]. We show how the profit generation performance of the three algorithms degrades compared with the ideal scenario in Section V-A in Fig. 6.

As we see in Fig. 6(b) and (c), nonlinear stochastic dual DP and gradient-bounded DP are both robust against model uncertainty. Only for $\sigma^2 = 1$, there is a substantial degradation in profit-generation performance. In contrast, Fig. 6(a) shows that even small uncertainties in the customer choice model have substantial negative impact on the affine value function approximation algorithm, decreasing expected profit with 99% confidence by about an order of magnitude for $\sigma^2 = 1$. We believe that this is due to the lack of state feedback in the affine value function approximation solution as detailed in

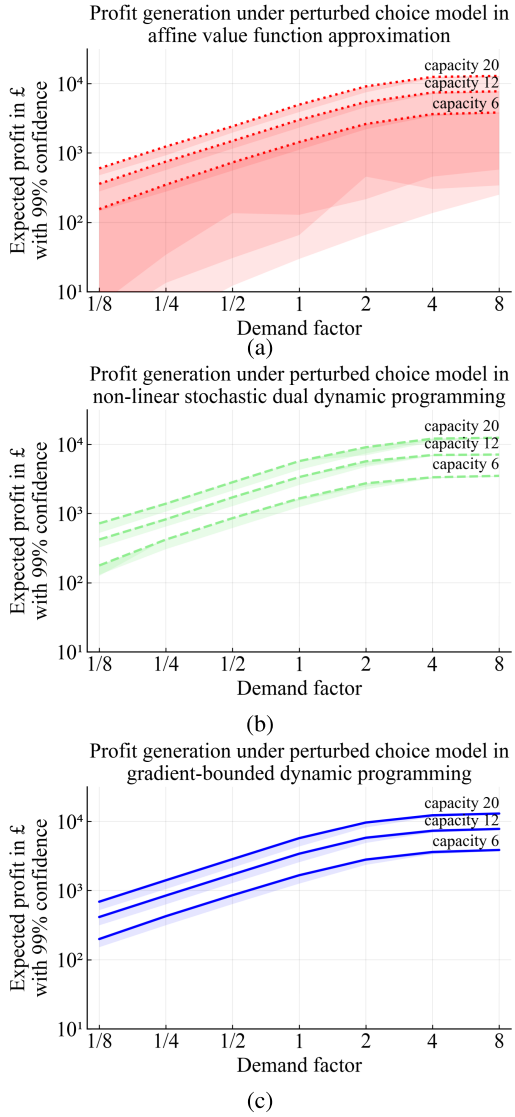

Fig. 6. Expected profits under perturbed model parameters. Lines: $\sigma^2 = 0.01$. Shaded regions: σ^2 increases to 0.1 and 1. (a) Affine value function approximation. (b) Non-linear stochastic dual dynamic programming. (c) Gradient-bounded dynamic programming.

Section III-A: For any $t \in T$, the suggested optimal slot price vector is identical for all x strictly inside the set of feasible states X, because the affine value function approximation has constant gradient for all these points. Since the other two algorithms both generate a piecewise affine approximate value function, gradients and hence optimal delivery prices vary depending on the particular state–time pair $(x, t) \in X \times T$.

We conclude that both gradient-bounded DP and nonlinear stochastic dual DP increase their relative profit-generation advantage over affine value function approximation under imperfect customer choice model parameter estimates.

VI. Conclusion and Future Work

In this brief, we analyzed three approximate DP algorithms to find approximately optimal delivery slot prices in the revenue management problem in attended home delivery. From a control-theoretical perspective, we identified limitations in the affine value function approximation algorithm and the nonlinear stochastic dual DP algorithm. We provided numerical evidence on how gradient-bounded DP can overcome these limitations. Possible directions for future work include investigating the numerical performance of these algorithms for other revenue management problems and extending the promising gradient-bounded DP approach to other customer decision models than multinomial logit.

References

[1] K. Asdemir, V. S. Jacob, and R. Krishnan, "Dynamic pricing of multiple home delivery options," *Eur. J. Oper. Res.*, vol. 196, no. 1, pp. 246–257, Jul. 2009.
[2] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2017.
[3] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 2, 4th ed. Nashua, NH, USA: Athena Scientific, 2012.
[4] C. F. Daganzo, "Modeling distribution problems with time windows: Part I," *Transp. Sci.*, vol. 21, no. 3, pp. 171–179, Aug. 1987.
[5] L. Dong, P. Kouvelis, and Z. Tian, "Dynamic pricing and inventory control of substitute products," *Manuf. Service Oper. Manage.*, vol. 11, no. 2, pp. 317–339, Apr. 2009.
[6] S. Koch and R. Klein, "Route-based approximate dynamic programming for dynamic pricing in attended home delivery," *Eur. J. Oper. Res.*, vol. 287, no. 2, pp. 633–652, Dec. 2020.
[7] D. Lebedev, P. Goulart, and K. Margellos, "A concave value function extension for the dynamic programming approach to revenue management in attended home delivery," in *Proc. 18th Eur. Control Conf. (ECC)*, Jun. 2019, pp. 999–1004.
[8] D. Lebedev, P. Goulart, and K. Margellos, "Gradient-bounded dynamic programming for submodular and concave extensible value functions with probabilistic performance guarantees," 2020, *arXiv:2006.02910*. [Online]. Available: http://arxiv.org/abs/2006.02910
[9] D. Lebedev, P. Goulart, and K. Margellos, "Gradient-bounded dynamic programming with submodular and concave extensible value functions," in *Proc. 21st IFAC World Congr.*, 2020, pp. 1–6. [Online]. Available: https://arxiv.org/pdf/2005.11213.pdf
[10] D. Lebedev, K. Margellos, and P. Goulart, "Approximate dynamic programming for delivery time slot pricing: A sensitivity analysis," 2020, *arXiv:2008.00780*. [Online]. Available: http://arxiv.org/abs/2008.00780
[11] D. Lebedev, P. Goulart, and K. Margellos, "A dynamic programming framework for optimal delivery time slot pricing," *Eur. J. Oper. Res.*, vol. 292, no. 2, pp. 456–468, Jul. 2021.
[12] M. V. F. Pereira and L. M. V. G. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Math. Program.*, vol. 52, nos. 1–3, pp. 359–375, May 1991.
[13] N. Saunders. (2018). *Online Grocery & Food Shopping Statistics*. Accessed: Jun. 16, 2020. [Online]. Available: https://www.onespace.com/blog/2018/08/online-grocery-food-shopping-stat%istics/
[14] A. Shapiro, "Analysis of stochastic dual dynamic programming method," *Eur. J. Oper. Res.*, vol. 209, no. 1, pp. 63–72, Feb. 2011.
[15] A. K. Strauss, R. Klein, and C. Steinhardt, "A review of choice-based revenue management: Theory and methods," *Eur. J. Oper. Res.*, vol. 271, no. 2, pp. 375–387, Dec. 2018.
[16] M. Suh and G. Aydin, "Dynamic pricing of substitutable products with limited inventories under logit demand," *IIE Trans.*, vol. 43, no. 5, pp. 323–331, Feb. 2011.
[17] K. E. Train, *Discrete Choice Methods With Simulation*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[18] X. Yang and A. K. Strauss, "An approximate dynamic programming approach to attended home delivery management," *Eur. J. Oper. Res.*, vol. 263, no. 3, pp. 935–945, Dec. 2017.
[19] X. Yang, A. K. Strauss, C. S. M. Currie, and R. Eglese, "Choice-based demand management and vehicle routing in E-fulfillment," *Transp. Sci.*, vol. 50, no. 2, pp. 473–488, May 2016.
[20] S. Zhang and X. A. Sun, "Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization," 2019, *arXiv:1912.13278*. [Online]. Available: http://arxiv.org/abs/1912.13278
[21] J. Zou, S. Ahmed, and X. A. Sun, "Stochastic dual dynamic integer programming," *Math. Program.*, vol. 175, nos. 1–2, pp. 461–502, May 2019.