Approximate Dynamic Programming for Revenue Management in Attended Home Delivery

Denis Lebedev

Mansfield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Engineering Science

Michaelmas Term 2020

To Anni
for being my strongest support
in times of greatest need.

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. Kostas Margellos and Prof. Paul Goulart for their continuous guidance and invaluable technical expertise throughout this project. I thank you for all the time that you dedicated to my project and all the obstacles that you helped me overcome on this exciting journey.

I also thank all current and past members of the Control Group for the inspiring discussions that we shared over the years. In particular, I would like to thank Michael Garstka, Xiaonan Lu and Licio Romao, who I am grateful to call not only my colleagues but some of the best friends that I have made during my time in Oxford. I will miss our weekly lunch breaks and the positive atmosphere that you brought into the office every day.

Generous financial support from SIA Food Union Management is gratefully acknowledged and I would like to extend my special thanks to Andrey Beskhmelnitsky for making this project happen. Furthermore, I would like to thank all current and past members of Food Union who I had the privilege to work with and learn from over the course of the project.

I would like to thank Anni and her family for their emotional support. Little did we know about how the COVID-19 crisis would impact our lives and I am deeply grateful for their hospitality in the past months, in which I also wrote this thesis.

Last but not least, I would like to thank my family for their unconditional love and continuous encouragement. Thank you for always being within my reach, no matter how far away from home I am.

Denis Lebedev
Dortmund, Germany
September 2020

Abstract

We consider the problem of finding profit-maximising prices for delivery time slots in the context of attended home delivery. This is a multi-stage optimal control problem that admits a dynamic programming formulation. However, this formulation is intractable for realistic problem sizes due to the so-called “curse of dimensionality”. To facilitate an approximation of this problem, we characterise some mathematical properties of the dynamic program: We show that the underlying Bellman operator has a unique fixed point and provide a closed-form expression for it. Moreover, we show that – under certain technical assumptions – the value function, which has a discrete domain and a continuous codomain, admits a continuous extension, which is a finite-valued, concave function of its state variables, at every time step. Motivated by these results, we develop an approximate dynamic programming algorithm for a more general class of problems, characterised by value functions that are concave extensible and submodular in their state-space. Our new algorithm computes deterministic upper and stochastic lower bounds of the value function in the realm of dual dynamic programming. We show that the proposed algorithm terminates after a finite number of iterations. Moreover, we also derive probabilistic guarantees on the value accumulated under the associated policy for a single realisation of the dynamic program and for its expected value. Furthermore, we compare our algorithm with two other approximate dynamic programming solutions proposed in the literature. Our analysis entails a theoretical examination from a control-theoretical perspective and a parametric numerical case study. Our experiments are 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 analysis shows that the proposed algorithm dominates the others with respect to computation time and profit-generation capabilities of the delivery slot pricing policies that it generates.

Contents

List of Figures viii

List of Tables x

Notation xi

1 Introduction 1
- 1.1 Background 1
- 1.2 Revenue management in attended home delivery 3
- 1.3 Challenges of revenue management in attended home delivery 5
  - 1.3.1 Delivery cost approximations 5
  - 1.3.2 Customer choice modelling 6
  - 1.3.3 Incentive mechanism design 6
  - 1.3.4 Optimal decision making 7
- 1.4 Research hypothesis 7
- 1.5 Outline and contributions of the thesis 8

2 A Dynamic Programming Framework 10
- 2.1 Introduction 11
- 2.2 Problem formulation 11
  - 2.2.1 A control view of dynamic delivery slot pricing 11
  - 2.2.2 Multi-stage optimal control problem formulation 13
  - 2.2.3 Dynamic programming formulation 16
- 2.3 Infinite and finite time horizon results 19
  - 2.3.1 Infinite time horizon result 19
  - 2.3.2 Finite time horizon results 20
- 2.4 Proofs of main results 25
  - 2.4.1 Proof of infinite time horizon theorem 25
  - 2.4.2 Proof of finite time horizon theorem 26
  - 2.4.3 Proof of Corollary 2.4 29
- 2.5 Numerical examples 30
  - 2.5.1 Increasing opportunity costs in an illustrative 2-slot example 30
  - 2.5.2 Monotonicity of prices in the illustrative 2-slot example 31

2.5.3 Exploiting concave extensibility in non-linear stochastic dual dynamic programming . . . 32
2.6 Conclusion . . . 35
2.A Proofs . . . 36
2.A.1 Proof of Lemma 2.5(i) . . . 36
2.A.2 Proof of Lemma 2.5(ii) . . . 37
2.A.3 Proof of Lemma 2.7 . . . 48
2.A.4 Proof of Lemma 2.8 . . . 50

3 Algorithmic Developments **51**
3.1 Introduction . . . 52
3.2 Value function approximation algorithm . . . 55
3.2.1 Initialisation . . . 56
3.2.2 “Forward sweep” . . . 57
3.2.3 “Backward sweep” . . . 58
3.3 Approximation algorithm properties . . . 58
3.4 Validation algorithm . . . 61
3.5 Validation algorithm properties . . . 62
3.5.1 Tail bounds . . . 63
3.5.2 Expectation bounds . . . 64
3.6 Numerical example . . . 66
3.6.1 Computation of the approximate value function . . . 67
3.6.2 Computational complexity analysis . . . 70
3.6.3 Computation of the bounds . . . 72
3.7 Conclusion . . . 76
3.A Proofs . . . 78
3.A.1 Proof of Proposition 3.1 . . . 78
3.A.2 Proof of Proposition 3.4 . . . 79
3.A.3 Proof of Proposition 3.5(i) . . . 80
3.A.4 Proof of Proposition 3.5(ii) . . . 82
3.A.5 Proof of Lemma 3.6 . . . 83
3.A.6 Proof of Proposition 3.7(ii) . . . 83
3.A.7 Proof of Proposition 3.8 . . . 84
3.B Synthetic problem data for Section 3.6.1 . . . 85

4 Comparative Study **86**
4.1 Introduction . . . 86
4.2 General sample-based, iterative approximation algorithm . . . 89
4.2.1 Affine value function approximation update . . . 91
4.2.2 Non-linear stochastic dual dynamic programming update . . 92
4.2.3 Gradient-bounded dynamic programming update . . . 95
4.3 Case study . . . 95
4.3.1 Exact model analysis . . . 96
4.3.2 Parameter sensitivity analysis . . . 102
4.4 Conclusion . . . 107
4.A Problem reformulation and table of parameters . . . 108
4.A.1 Problem reformulation for non-linear stochastic dual dynamic programming . . . 108
4.A.2 Table of parameters for Section 4.3.2 . . . 110

5 Conclusions **111**
5.1 Conclusions of this dissertation . . . 111
5.1.1 Dynamic programming framework . . . 111
5.1.2 Algorithmic developments . . . 112
5.1.3 Comparative study . . . 112
5.2 Directions for future research . . . 113
5.2.1 Directions for future research on the dynamic programming framework . . . 113
5.2.2 Directions for future research on algorithmic developments . 114
5.2.3 Directions for future research on the comparative study . . . 114

References **116**

List of Figures

1.1 Schematic structure of the procedural context of revenue management in attended home delivery. 4

2.1 A feedback control view of the revenue management problem in attended home delivery. 12
2.2 Two equivalent interpretations of the concave closure $\tilde{f}$ (blue, dotted line) of a function f (black dots): by minimising hyperplanes (green, solid lines, Definition 2.1) and by maximising convex combinations (red, shaded area, Proposition 2.2). Arrows include the directions of optimisation. 22
2.3 Illustrative example of a 2-slot problem. 31
2.4 Monotonicity of prices. 32
2.5 Plots of sample profits (grey dots), upper bound on expected profit (blue solid line) and cumulative moving average of sample profits, i.e. the stochastic lower bound on expected profit (red dashed line). . . 34
2.6 Green, dotted line segment between y and y' and red, dotted line segment between $y + \sum_{s \in \sigma} 1_s$ and $y + 1_1$. By (2.72) we have that at the intersection of the two lines, the interpolation of V_t between the two extreme points of the red line lies above the interpolation of V_t between the two extreme points of the green line. The light blue, solid line box spans all the points in $B(y)$. 47

3.1 Schematic grouping of selected approximate DP literature based on how general-purpose or problem-specific the approaches are and if they are limited to a discrete or continuous state space. 52
3.2 Illustration how the approximate value function is refined in the "backward sweeps" of gradient-bounded DP. 59
3.3 Plots of sample profits (grey dots), upper bound $u(i)$ (blue solid line, Corollary 3.2) and stochastic lower bound $l(i)$ (red dashed line, Proposition 3.3) on expected profit. 68
3.4 Histograms of 1,000 sample profits after 1 (a) and 100 (b) iterations of Algorithm 3.1. 69

3.5 Computation time of Algorithm 3.1 in seconds against number of iterations of the 17-slot example from Section 3.6.1 70
3.6 Computation time for 100 iterations of Algorithm 3.1 in seconds against varying problem dimensionality. 71
3.7 Probabilistic bounds – empirical Cantelli bound (green dashed line, Proposition 3.5(i)) and Dvoretzky-Kiefer-Wolfowitz tail bound (blue dotted line, Proposition 3.5(ii)) – on the profit obtained from a single validation sample as functions of various significance levels for 10 (a), 100 (b) and 1000 (c) validation samples. 74
3.8 Probabilistic bounds – Gaussian bound (green dashed line, (3.11)), expectation Dvoretzky-Kiefer-Wolfowitz bound (blue dotted line, Proposition 3.7(ii)) and empirical Bernstein bound (orange dash-dotted line, Proposition 3.7(i)) – on the expectation of the profit obtained as functions of various significance levels and for 10 (a), 100 (b) and 1000 (c) validation samples. 75

4.1 Delivery sub-area schematic with length L and width W. Crosses indicate customer locations served in a particular delivery time slot. 97
4.2 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. 99
4.3 Expected profits of non-linear stochastic dual DP for demand factor 8 and slot capacity 12 decrease as more iterations are added over time. 101
4.4 Expected profits with 99% confidence of the three algorithms against demand factor and for all delivery slot capacities. 101
4.5 Expected profits with 99% confidence for affine value function approximation (a), non-linear stochastic dual DP (b) and gradient-bounded DP (c) under perturbed customer arrival rate. Shaded regions indicate how expected profit with 99% confidence increases when the customer arrival rate increases from 0.6 to 1. The lines correspond to an arrival rate of 0.8. 105
4.6 Expected profits with 99% confidence for affine value function approximation (a), non-linear stochastic dual DP (b) and gradient-bounded DP (c) under perturbed customer choice model parameters from Table 4.2. Lines indicate performance at $\sigma^2 = 0.01$ and shaded regions indicate performance as σ^2 increases to 0.1 and 1. 106

List of Tables

2.1 The parameters of the numerical example. 30
2.2 The parameters of the non-linear stochastic dual DP example. 33

3.1 Re-statement of numerical example parameters from Section 2.5.3. . 66
3.2 Synthetically generated customer choice value parameters for $s > 17$. 85

4.1 Exact model parameters. 96
4.2 Estimation error-corrupted customer choice parameters. 110

Notation

Scalar sets

$\mathbb{N}$	the natural numbers
$\mathbb{R}$	the real numbers
$\mathbb{R}_{+(+)}$	the non-negative (positive) real numbers
$\mathbb{R}_{-(-)}$	the non-positive (negative) real numbers
$\mathbb{Z}$	the whole numbers

Definitions and inequalities

$A := B$	A is defined by B
$A =: B$	B defines A
$A \leq B$	element-wise inequality between A and B
$A < B$	strict element-wise inequality between A and B
$\lfloor A \rfloor$	rounding *down* the elements of A to the nearest integer

Convex sets

$\mathrm{conv}(C)$	convex hull of set C
∂C	boundary of the set C

Vectors and matrices

$\mathbf{1}$	vector of ones of appropriate dimension, i.e. $[1, \ldots, 1]^\intercal$
$\langle x, y \rangle$	standard inner product of vectors x and y
$A^\intercal$	transpose of the matrix A
x_i	i^{th} element of vector x
1_i	vector of zeros of appropriate dimension with 1 as the i^{th} element
$A_{i,j}$	element in row i and column j of the matrix A

Set operations

$\lvert A \rvert$	cardinality of set $\lvert A \rvert$, i.e. the number of its elements
$A \cup B$	union of sets A and B
$A \setminus B$	set of elements in A but not in B

Probability

$\Pr(C)$	probability of event C

$\mathbb{E}[x]$	expected value of vector x

Notation specific to revenue management in attended home delivery

$T := \{1, 2, \ldots, \bar{t}\}$	finite and discrete time horizon
n	number of delivery time slots
$S := \{1, 2, \ldots, n\}$	set of delivery time slots
$x_t \in \mathbb{Z}^n$	state vector at time $t \in T \cup \{\bar{t} + 1\}$
$\bar{x}$	maximum feasible state vector
$X := \{x_t \in \mathbb{Z}^n \mid 0 \leq x_t \leq \bar{x}\}$	set of feasible states
$F(x_t) := \{s \in S \mid x_t + 1_s \in X\}$	set of feasible slots in state x_t
$d_t := [d_{1,t}, d_{2,t}, \ldots, d_{n,t}]^\intercal$	vector of delivery slot charges
$\underline{d} \in \mathbb{R}$	minimum admissible delivery charge
$\bar{d} \in \mathbb{R}$	maximum admissible delivery charge
$D := \Big\{d_t \Big\vert d_{s,t} \in [\underline{d}, \bar{d}] \cup \{\infty\},$ for all $s \in S\Big\}$	set of admissible delivery charge vectors
$p_s(d_t)$	probability of a customer choosing slot s when offered delivery slot prices d_t
$\beta_c \in \mathbb{R}$	multinomial logit parameter for a constant offset
$\beta_d \in \mathbb{R}_{--}$	multinomial logit parameter for price sensitivity
$\{\beta_s\}_{s \in S} \in \mathbb{R}^n$	multinomial logit parameters for slot preferences
$r \in \mathbb{R}$	average revenue per order
$\lambda \in (0, 1]$	expected rate of customer arrival on the booking system per time step
$C : \mathbb{Z}^n \to \mathbb{R} \cup \infty$	approximate delivery cost function
$\xi_t \in \{0, 1\}^n$	stochastic state transition vector
$V_t : \mathbb{Z}^n \to \mathbb{R} \cup -\infty$	value function of the dynamic program; measures expected profit-to-go at time t.

Acronyms

DP	dynamic program(ming)

1
Introduction

Contents

1.1 Background **1**
1.2 Revenue management in attended home delivery . . . **3**
1.3 Challenges of revenue management in attended home delivery **5**
1.3.1 Delivery cost approximations 5
1.3.2 Customer choice modelling 6
1.3.3 Incentive mechanism design 6
1.3.4 Optimal decision making 7
1.4 Research hypothesis **7**
1.5 Outline and contributions of the thesis **8**

1.1 Background

Data-driven optimisation and control pertain a diverse set of engineering problems. Some prominent examples include control of congestion on the internet [1], control and incentive mechanisms in bicycle sharing systems [2] and pricing of uncertain electricity from renewable energy sources [3]. In all these problems, decisions need to be made over time and under uncertainty to reach a desired behaviour. Similarly, the problem under study outlined below also displays such features.

This thesis focuses on revenue management in the context of attended home delivery. In general, revenue management deals with how to sell goods and services in a profit-maximising way, subject to a number of resource constraints [4]. One of the classical examples of revenue management is the optimal pricing of tickets for seats on an aircraft; over a finite-time booking horizon, prices must be adjusted dynamically according to stochastic customer buying behaviour in order to maximise profits [5]. Over time, revenue management has been applied to a variety of applications, one of which is attended home delivery.

Attended home delivery differentiates itself from ordinary delivery, e.g. mail delivery, by imposing the additional requirement that customers must be present upon delivery of the goods or services. This is typically the case for the delivery of food, other perishable goods or maintenance services. In practice, this means that the delivery firm and the customers must pre-agree delivery times [6].

Recently, the practical relevance of attended home delivery has increased substantially. For example, U.S. households are predicted to spend over $100 billion per year on online grocery shopping in 2022 according to the Food Marketing Institute [7] and up to $133.8 billion per year according to GlobalData [8]. Similar developments are taking place in Europe where, e.g. in the Netherlands, online grocery sales are predicted to grow annually by 14% to reach a share of 16% of all grocery sales by 2030 [9]. The prevalent COVID-19 crisis is likely to accelerate this growth even further [10].

One of the main impediments to the growth of online supermarkets is the increased cost of home delivery compared with the logistics of brick-and-mortar supermarkets [11], not least due to the requirement to fulfil attended home delivery orders. From the side of the product-selling company, this poses an array of questions about how the distribution of food directly to end-customers could be optimised. This has been recognised as an important problem by Food Union[1]. This approach questions the *status quo* solution, which often solves the distribution of goods problem indirectly, e.g. through supermarkets in the food industry.

[1]Research resulting in this thesis has been supported by SIA Food Union Management.

The means by which the distribution process can be optimised in a "direct" way stems from the assumption that customers have a certain degree of flexibility and generally varying preferences with respect to the timing of deliveries. These preferences – stochastic from the point of view of the firm – have to be exploited through some mechanism, e.g. differentiated pricing of delivery options, in order to steer customer choice behaviour in a desired way, allowing orders to be fulfilled efficiently. This approach can outperform the indirect distribution chain if the additional margin from cutting out the intermediate agent, e.g. the supermarket, is sufficient to cover the increased cost of end-customer deliveries as well as any possible costs of incentive mechanisms, e.g. discounts on delivery prices of unpopular delivery time slots.

In the remainder of this introduction, we show in Section 1.2 how revenue management in attended home delivery fits procedurally between inventory control and vehicle routing problems, which are all subjects of operations research and fall into the broad class of control and optimisation problems. In Section 1.3, we present the problems surrounding revenue management that have been studied in the application-focused literature. Section 1.4 states the main hypothesis of this dissertation. Finally, Section 1.5 outlines the structure of the remainder of this thesis in support of our overall argument.

1.2 Revenue management in attended home delivery

Revenue management is part of a larger set of logistics and operations research problems. With the particular application of attended home delivery in mind, the value chain features a number of distinct challenges both upstream and downstream of the revenue management problem as illustrated in Fig. 1.1.

Procedurally upstream of the revenue management problem, the task is to have enough stock produced and stored in the production facility or, in the case of services, to have enough staff on duty. Excessively high levels of stock generate unnecessary costs, whereas excessively low levels of stock result in customer fulfilment problems,

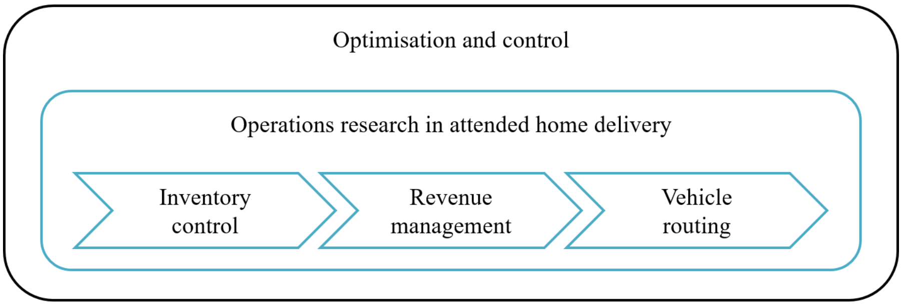


Figure 1.1: Schematic structure of the procedural context of revenue management in attended home delivery.

i.e. customers are unable to purchase a product or service. Finding the optimal inventory level is the objective of *inventory control*.

Theoretical developments in this field mainly involve methods from dynamic programming theory and cover a broad range of product and customer types; see [12] for a textbook reference. Current research focuses on specific scenarios, where standard inventory control methods fail. For example, [13] proposes how to cope with inventory control challenges posed by low-consumption items.

Procedurally downstream of the revenue management problem, the main question is how a fixed list of orders should be fulfilled. This question is closely linked to a class of mathematical problems known as *vehicle routing problems*, which were initially defined by [14]. The vehicle routing problem extends the well-known travelling salesman problem in [15] by considering multiple delivery vehicles instead of one "salesman". In its standard form, however, the vehicle routing problem is still too specific to capture critical elements of many real-world applications.

This justifies the interest in developing theory for many extensions including various types of additional constraints, e.g. limits on delivery vehicle capacity or acceptable delivery time windows for individual orders. The authors of [16] provide a meta-analysis of the various settings that have been studied in the past.

Vehicle routing problems in attended home delivery feature both capacity and delivery time constraints, which require solution of a general and hence complicated form of the vehicle routing problem known as the capacitated vehicle routing problem with time windows. For industry-sized applications with hundreds of customers, this problem cannot be solved exactly, but approximate algorithms have been developed which allow trade-offs between numerical accuracy and computational cost. References [17–19] have compiled overviews of approximate solution algorithms.

1.3 Challenges of revenue management in attended home delivery

The *revenue management problem* in attended home delivery sits between the inventory control problem and the vehicle routing problem. It assumes that enough stock is available to cover profit-maximising customer demands for a product and that a vehicle routing problem solver can allocate delivery routes to vehicles after the sales process generates a list of orders. While these problems are addressed in the related literature discussed above, a number of challenges faced by revenue management remain. We address them in the following sections.

1.3.1 Delivery cost approximations

The first challenge of revenue management is finding an approximate delivery cost function. Even if a vehicle routing problem solver can provide a near-optimal solution given *one* list of orders, for the purposes of revenue management, we would ideally need the total delivery cost for *any* possible list of orders. Intuitively, this information helps the algorithm decide whether customers should be encouraged or discouraged to purchase a particular delivery option. Due to the large number of possible combinations of orders, it is computationally infeasible to calculate the exact cost for all potential delivery schedules. That is why revenue management needs to find approximations to expected delivery cost functions, which can be evaluated in near-real time. This is different from the vehicle routing problem, because we do not require a firm allocation of vehicles and definition of routes, but simply an approximation of total delivery costs.

In practice, we need to find approximations sufficiently close to the actual delivery cost in fractions of a second, typically less than 100 ms [20]. Approximations that can satisfy this time constraint were devised, for example, by [21]. They assume that customers are clustered in rectangular sub-areas as samples of a continuous uniform distribution, which makes it possible to express an expected delivery cost function as a function of the number of orders placed in every delivery time slot.

1.3.2 Customer choice modelling

The second challenge of revenue management is the modelling of customer choice probabilities, i.e. the probabilities that customers choose a delivery option given some incentives. These incentives can be, for example, the different prices for alternative delivery slots. Many different types of models, both parametric and non-parametric, are considered in the literature; see [22] for a survey. They explain how many parametric models build on the flexible and well-understood multinomial logit model and how recent research has been exploring more complicated models, capturing the dynamics of specific applications.

In addition to parametric customer choice models, [22] also present research on non-parametric, rank-based models, i.e. models that do not capture customer choice behaviour in a particular functional form, but as an ordered list. While this provides more adaptability, it loses the possibility of exploiting structural properties of parametric models. In this thesis, we eschew non-parametric models, since exploiting the structural properties of the customer choice model will be a key to our solution approach as discussed further below.

1.3.3 Incentive mechanism design

The third challenge of revenue management is the design of incentive mechanisms that steer customer choice in a desired way. One of the simplest and most frequently used mechanisms is differentiated pricing of delivery slots, the intuition being that customers can be incentivised to choose a particular slot by reducing delivery prices and, *vice versa*, dis-incentivised by increasing delivery prices. This mechanism has been studied extensively, e.g. in [23–25]. Since this is the most widely used mechanism in practice, this dissertation focuses on pricing as the sole incentive mechanism for revenue management in attended home delivery.

However, other mechanisms have also been investigated. For example, [26] studied how changing the duration of delivery slots can be more effective to generate fulfillable delivery schedules than pricing of delivery slots. Another direction that might attract further research activities in the future is pointed out by [27], who investigate optimal marketing strategies (which are not restricted to the context of attended home delivery) over social networks, i.e. when the buying behaviour of some customers affects the behaviour of others.

1.3.4 Optimal decision making

The final challenge of revenue management integrates the results of the previous challenges and seeks to find the optimal strategy of the firm given an incentive mechanism, customer choice model and routing cost approximation. For example, it seeks to determine what prices should be charged for all delivery options if customers are assumed to purchase based on a multinomial logit model and the delivery cost is assumed to be an affine function of the accepted number of orders, irrespective of the location of customers. This is precisely what [20] and [28] provide by unifying the revenue management problem into a single dynamic program (DP).

1.4 Research hypothesis

Although accuracy of predicted optimal prices is important, the literature is concentrated on approximating the value function of the underlying DP formulation as an affine function [20]. We suspect that this restriction significantly limits the expressivity of the model and hence the accuracy of the predicted optimal prices. Therefore, we believe that closer approximations can be found if we better understand the properties of the value function, including its behaviour in time and across states. This would allow informed choices about approximation methods to be made, yielding more accurate results at a sufficiently low computational cost. To this end, we studied the DP as proposed by [28] and analysed some of its mathematical properties. Based on these results, we derived an approximate DP algorithm, termed gradient-bounded DP, and derived methods for quantifying its random performance probabilistically. Finally, we compared the algorithm with approaches from the literature in a detailed parametric case study. The outline and main contributions of this thesis follow this line of argument, as described in the next section.

1.5 Outline and contributions of the thesis

In support of the research hypothesis outlined above, this doctoral thesis entails three main sets of contributions, which naturally form its three main technical chapters. We detail them in the outline of the thesis below.

In Chapter 2, we lay the theoretical groundwork for our results. For a particular set-up of the revenue management problem in attended home delivery, we consider a DP formulation, which is intractable for realistic problem sizes. We characterise key mathematical properties of the value function of the DP, which make systematic and efficient approximations possible. These developments largely contributed to the development of the following two publications:

- 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: *18th European Control Conference, ECC 2019, Naples, Italy, June 25-28, 2019.* 2019, pp. 999–1004.
- D. Lebedev, P. Goulart, and K. Margellos. “A dynamic programming framework for optimal delivery time slot pricing”. In: *European Journal of Operational Research* 292.2 (2021), pp. 456–468.

In Chapter 3, we exploit the theoretical results from Chapter 2 to develop an approximate DP algorithm. The algorithm generates deterministic upper and stochastic lower bounds on the expected profit obtained by the product-selling company, similarly to other dual DP algorithms. The algorithm is new because it can be applied to value functions defined over integer domains and because it does not rely on the solution of non-convex optimisation problems. Furthermore, since the performance of this algorithm is random, we derive probabilistic bounds on the tails and the expectations of the distribution of this random variable, making it possible to quantify and to compare the performance of this and similar algorithms. These developments largely contributed to the development of the following two publications:

- D. Lebedev, P. Goulart, and K. Margellos. "Gradient-Bounded Dynamic Programming with Submodular and Concave Extensible Value Functions". In: *21st IFAC World Congress 2020*. Available: `https://arxiv.org/pdf/2005.11213.pdf`. 2020.
- D. Lebedev, P. Goulart, and K. Margellos. *Gradient-Bounded Dynamic Programming for Submodular and Concave Extensible Value Functions with Probabilistic Performance Guarantees*. Tech. rep. Submitted for peer review at Automatica. Available: `https://arxiv.org/pdf/2006.02910.pdf`. 2020.

In Chapter 4, we derive a benchmark case study to compare our novel approximate DP algorithm with two other solutions proposed in the literature. We identify the limitations of the other two solutions from a control theoretical perspective and explain how our novel approximate DP algorithm can overcome these. Finally, we verify our theoretical analysis on a multi-scenario case study and parameter sensitivity tests. These developments largely contributed to the development of the following publication:

- D. Lebedev, K. Margellos, and P. Goulart. *Approximate Dynamic Programming for Delivery Time Slot Pricing: a Sensitivity Analysis*. Tech. rep. Submitted for peer review at IEEE Transactions on Control Systems Technology. Available: `https://arxiv.org/pdf/2008.00780.pdf`. 2020.

In Chapter 5, we conclude the thesis and discuss some possible directions for future research.

2

A Dynamic Programming Framework

for Revenue Management in Attended Home Delivery

Contents

2.1 Introduction **11**
2.2 Problem formulation **11**
2.2.1 A control view of dynamic delivery slot pricing 11
2.2.2 Multi-stage optimal control problem formulation 13
2.2.3 Dynamic programming formulation 16
2.3 Infinite and finite time horizon results **19**
2.3.1 Infinite time horizon result 19
2.3.2 Finite time horizon results 20
2.4 Proofs of main results **25**
2.4.1 Proof of infinite time horizon theorem 25
2.4.2 Proof of finite time horizon theorem 26
2.4.3 Proof of Corollary 2.4 29
2.5 Numerical examples **30**
2.5.1 Increasing opportunity costs in an illustrative 2-slot example 30
2.5.2 Monotonicity of prices in the illustrative 2-slot example 31
2.5.3 Exploiting concave extensibility in non-linear stochastic dual dynamic programming 32
2.6 Conclusion **35**
2.A Proofs **36**
2.A.1 Proof of Lemma 2.5(i) 36
2.A.2 Proof of Lemma 2.5(ii) 37
2.A.3 Proof of Lemma 2.7 48
2.A.4 Proof of Lemma 2.8 50

2.1 Introduction

In this chapter, we lay the theoretical groundwork for a solution of the revenue management problem in attended home delivery, which we introduced informally in the previous chapter. In particular, we derive a DP formulation of the revenue management problem in attended home delivery and characterise key mathematical properties of its value function, which allow computation of efficient approximations. This will be the basis for the subsequent algorithmic developments in Chapter 3 and the numerical analysis in Chapter 4.

To this end, the structure of this chapter is as follows: In Section 2.2, we introduce and formally define the revenue management problem in attended home delivery, starting from an optimal control perspective, and derive multiple equivalent DP formulations that will be instrumental in the subsequent sections and chapters. In Section 2.3, we present the main characterisations of the value function of the derived DP both for infinite and finite time horizons. We then prove these results in Section 2.4, where we also develop a result on the monotonicity of prices with respect to the number of placed orders. In Section 2.5, we present a numerical illustration of our theoretical results on a low-dimensional example and on an industry-sized problem, while Section 2.6 concludes the chapter and suggests directions for future research. Appendix 2.A at the end of this chapter contains the proofs of results not included in the main body of this chapter.

2.2 Problem formulation

2.2.1 A control view of dynamic delivery slot pricing

Broadly speaking, and independently of any specific problem characteristics, 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 behaviour by keeping track of placed orders, which we can use as *state feedback*. 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. 2.1 below.

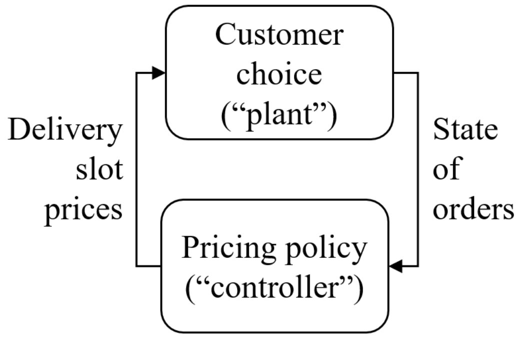


Figure 2.1: A feedback control view of the revenue management problem in attended home delivery.

In principle, the exact state of orders is high-dimensional. For example, it would need to represent locations of all customers and their chosen delivery time slot, if any. For realistically-sized problems the number of states becomes prohibitively large to compute the optimal pricing policy exactly for all states. Therefore, several model simplifications have been proposed in the literature; see [22] for an overview.

In this thesis, we focus on the state-space representation of [20, 28]. 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 dimensionality of the state is the number of delivery time slots of any delivery day. The state of orders can then be thought of as a vector whose entries represent the number of deliveries for every delivery time slot in a particular sub-area.

Notice that this model neglects the location of customers, which is a reasonable simplification assuming that customers are evenly distributed in the delivery sub-area and that sufficiently many customers order in every time slot, such that most of the delivery sub-area needs to be travelled to deliver all orders in any time slot. In practice, one can achieve such conditions approximately by dividing the delivery area in appropriate delivery sub-areas (see e.g. [20]). We formally define the revenue management problem in attended home delivery in the next section based on this model.

2.2.2 Multi-stage optimal control problem formulation

Revenue management in attended home delivery can be formulated as a multi-stage optimal control problem for any delivery sub-area served by a single delivery vehicle. Customers are assumed 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 $T := \left\{1, 2, \ldots, \bar{t}\right\}$. There is an additional time step $\bar{t}+1$, beyond which no bookings happen any more, which we will use to define the terminal condition of the problem.

The probability of a customer arriving (on the booking system) in any of these time steps is time-independent and denoted by $\lambda \in (0, 1]$. Notice that we do not assume that all time steps $t \in T$ cover the same period passing in real time, which makes it possible to model non-homogeneous customer arrival processes (see [28, Section 4.3]), e.g. the effect that usually more customers arrive closer to the end than to the start of the booking horizon.

Suppose that the delivery day is split into n delivery time slots. Denote the set of delivery time slots by $S := \{1, 2, ..., n\}$. We focus on an aggregated state-space representation, where for any time step $t \in T \cup \left\{\bar{t}+1\right\}$ 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 a non-negative hypercube bounded by the maximum state vector $\bar{x}$, i.e. $X := \{x_t \in \mathbb{Z}^n \mid 0 \leq x_t \leq \bar{x}\}$. We also define the set of feasible slots $F(x_t) := \{s \in S \mid x_t + 1_s \in X\}$. For any $t \in T$, we define the delivery charge vector $d_t := [d_{1,t}, d_{2,t}, \ldots, d_{n,t}]^\intercal$. Denote the minimum admissible delivery charge for any slot by $\underline{d} \in \mathbb{R}$ and the maximum admissible delivery charge by $\bar{d} \in \mathbb{R}$, such that $\bar{d} \geq \underline{d}$. From a theoretical standpoint, it is not necessary to assume that both of these will be positive, but in practice they usually are. Let the set of admissible delivery charge vectors be $D := \left\{d_t \;\middle|\; d_{s,t} \in \left[\underline{d}, \bar{d}\right] \cup \{\infty\} \text{ for all } s \in S\right\}$.

The role of infinite prices is to mathematically indicate unfeasible slots, which are hence not offered to the customer. We subsequently ensure that the customer choice model is designed that if the delivery slot charge for any slot $s \in S$ is infinite, then the probability of the customer selecting that slot is 0. We make these assumptions explicit in Section 2.3.

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 also 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 requirement 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. Furthermore, we restrict $p_s(d_t) = 0$ if $x_t + 1_s \notin X$, i.e. we do not allow unfeasible orders. Such models are typical for order-taking processes (see [20, 23, 28, 34]). Notice that in this setting we neglect that customers might be allowed to cancel their orders or to change their preferred delivery time slot. This is reasonable to assume if such requests happen sufficiently rarely. For the purpose of some results and numerical examples, we will assume that the *customer choice model* follows a multinomial logit model, like in [20, 25, 28], i.e.

$$p_s(d_t) := \frac{\exp(\beta_c + \beta_s + \beta_d d_{s,t})}{\sum_{k \in S} \exp(\beta_c + \beta_k + \beta_d d_{k,t}) + 1}, \tag{2.1}$$

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 of the s^{th} slot and $\beta_d < 0$ is a parameter for the price sensitivity. Note that the no-purchase utility is normalised to zero, i.e. 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 (2.1) 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 $\{\beta_s\}_{s \in S}$ parameters. However, β_c is often kept in practice to normalise one of the $\{\beta_s\}_{s \in S}$ parameters to zero (see e.g. [20]). Define the vector of customer choice probabilities by $p(d_t) := [p_s(d_t)]_{s \in S}$ and denote the set of values that $p(d_t)$ can take for all $d_t \in D$ as P, i.e. $P := \{p(d_t) \mid d_t \in D\}$.

While the multinomial logit model is a very specific choice and some of our results depend on its functional form, it is a broadly-used and well-studied model with some desirable properties (see [35, Chapter 3] for a textbook reference). For example, its parameters can be estimated using maximum likelihood estimation, where it is convenient that the log-likelihood function is concave in the parameters [36]. Furthermore, we exploit the fact that one can obtain the inverse customer choice model, i.e. prices in terms of choice probabilities, further below in Section 2.4.2, similarly to [25]. Finally, if the multinomial customer choice model cannot represent true customer choice behaviour exactly, for example due to exogenous effects like changing prices of competitors, we demonstrate numerically in Section 4.3.2 that even substantial estimation errors in the multinomial logit choice model only have a moderate effect on the performance of the pricing policy using the algorithm, which we will derive in Chapter 3 based on the results from this chapter.

Next, we define an average revenue per order $r \in \mathbb{R}$, which is usually but not mathematically necessarily positive, an expected customer arrival rate (on the booking system) per time step $\lambda \in (0, 1]$ and an approximate delivery cost function $C : \mathbb{Z}^n \to \mathbb{R} \cup \infty$, such that

$$C(x) := \begin{cases} C_+(x), & \text{if } x \in X \\ \infty & \text{otherwise,} \end{cases} \tag{2.2}$$

where $C_+ : \mathbb{R}^n \to \mathbb{R}$ is a given Lipschitz continuous function. The role of infinite delivery costs is to indicate infeasible states, i.e. undeliverable combinations of orders. Let $\langle ., . \rangle$ denote the standard inner product of its arguments. We construct a multi-stage optimal control problem of the following form:

$$\begin{aligned} &\max_{\{d_t \in D\}_{t=1}^{\bar{t}}} \mathbb{E}\left[-C(x_{\bar{t}+1}) + \sum_{t \in T} \langle x_{t+1} - x_t, d_t + r \rangle \right], \\ &\text{subject to } x_{t+1} = x_t + \xi_t, \text{ for all } t \in T, \quad x_1 = 0, \end{aligned} \tag{2.3}$$

where $\mathbb{E}$ is the expectation operator that is 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 in (2.3) is the total expected operational contribution margin, i.e. revenue from sales and delivery charges minus delivery costs. For simplicity, we will refer to this objective as simply the *expected profit* that we seek to maximise.

2.2.3 Dynamic programming formulation

The above multi-stage optimal control problem is stage-wise independent, which makes it possible to derive a DP recursion. Depending on the particular situation, some DP formulations may be preferred over others. Hence, we introduce the formulations that are most convenient for our analysis in this section. It is important to note that all formulations introduced further below describe the *same* problem and they are mathematically *identical*.

The first DP formulation is obtained, in a manner analogously to [20], 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 $(x,t) \in X \times T$ and by expanding the expectation in (2.3) as follows:

$$V_t(x) := \max_{d \in D} \left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + V_{t+1}(x + 1_s) \right) + \left(1 - \lambda \sum_{s \in S} p_s(d) \right) V_{t+1}(x) \right\}$$

$$\forall (x,t) \in X \times T, \text{ where } V_{\bar{t}+1}(x) = -C(x) \quad \forall x \in X. \tag{2.4}$$

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 mathematically undefined in (2.4), but it is zero in practice. Note that – similar to [20] – we ignore any vehicle load capacity constraints in the problem, as they are much less restricting than the time constraints on the delivery slots. Therefore, including the vehicle load capacity constraints would only increase computational costs, but would not substantially improve the decision policy.

We refer to (2.4) as the *standard* DP formulation of the revenue management problem in attended home delivery. An equivalent representation of this DP, which is helpful for some of the subsequent analysis, also provides further economic insight into the profit-to-go function. By grouping the summations in (2.4), we can express the standard DP as

$$V_t(x) := \max_{d \in D} \left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + V_{t+1}(x + 1_s) - V_{t+1}(x) \right) + V_{t+1}(x) \right\}$$
$$\forall (x, t) \in X \times T, \text{ where } V_{\bar{t}+1}(x) = -C(x) \quad \forall x \in X. \tag{2.5}$$

The difference $V_{t+1}(x) - V_{t+1}(x + 1_s)$ in (2.5) represents the value foregone by accepting an additional (discrete spatial) order, which in economic terms is the *opportunity cost* of an order. Hence, we refer to (2.5) as the *opportunity cost* formulation of the revenue management problem in attended home delivery.

Furthermore, we state a third formulation of the problem, by noticing that (2.5) can be equivalently expressed as a so-called *stochastic shortest path problem*, a special version of an undiscounted, finite-state, discrete-time Markov decision problem. We follow the notation of [37, Chapter 3] to construct a problem of the following form:

$$V_t(x) := \max_{d \in D} \sum_{y \in Y_+(x)} \pi_{x,y}(d) [\hat{\rho}(x, y, d) + V_{t+1}(y)]$$
$$\forall (x, t) \in X \times T, \text{ where } V_{\bar{t}+1}(x) = -C(x) \quad \forall x \in X, \tag{2.6}$$

where $Y_+(x) := \{y \mid y = x + 1_s \quad \forall s \in F(x)\}$. The modified state transitions $\pi_{x,y}(d)$ and stage revenue function $\hat{\rho}(x, d, y)$ are found as follows to establish equivalence with (2.4): If $y = x \in X$, then $\pi_{x,y}(d) := 1 - \lambda \sum_{s \in S} p_s(d)$ and for all $s \in S$, if $y = x + 1_s \in X$, then $\pi_{x,y}(d) = \lambda p_s(d)$. Note that the first case is associated with no transition, i.e. the last term in the first line of (2.4), and the second case is a valid, i.e. unit-sized, feasible order, i.e. the first summation over $s \in S$ in the first line of (2.4).

In a similar way, for all $s \in S$, if $y = x + 1_s \in X$, then $\hat{\rho}(x, d, y) = d_s + r$ and in all other cases, $\hat{\rho}(x, d, y) = 0$. For convenience, similarly to [37, Chapter 3], we define $\rho(x,d) := \sum_{y \in Y_+(x)} \pi_{x,y}(d)\hat{\rho}(x,y,d)$ to simplify the Bellman equation in (2.6), which yields

$$V_t(x) = \max_{d \in D} \left\{ \rho(x,d) + \sum_{y \in Y_+(x)} \pi_{x,y}(d) V_{t+1}(y) \right\}. \tag{2.7}$$

Finally, to represent the DP in (2.4) in a more compact form, we define the Bellman operator $\mathcal{T}$ through the relationship

$$V_t = \mathcal{T} V_{t+1}, \text{ for all } t \in T. \tag{2.8}$$

In summary, (2.4)–(2.8) are mathematically identical DP formulations of the revenue management problem in attended home delivery. For simplicity, we will always refer to the classical formulation in (2.4) when the particular formulation of the DP is irrelevant for our argument, but all subsequent results hold irrespective of the particular formulation due to their equivalence.

Since for realistic problem instances the DP formulation in (2.4) cannot be solved by direct computation of the exact value function due to a prohibitively large number of states, one needs to approximate the value function of the DP. However, given *any* approximate value function, finding approximately optimal prices is relatively easy, provided that evaluating the approximate value function locally is not overly costly computationally. For example, for the multinomial logit model, [25] determine approximately optimal prices using a simple Newton root search. Hence, the main focus of our theoretical results, which we present in the next section, is concerned with understanding mathematical properties of the value function to facilitate its approximation. Furthermore, our mathematical results have immediate implications on the monotonicity of (approximately) optimal prices with respect to changes in the number of placed orders, which we also characterise in this chapter. This analysis complements the research on the price-inventory relationship under multinomial logit customer choice (see e.g. [34, 38, 39]).

2.3 Infinite and finite time horizon results

2.3.1 Infinite time horizon result

We first consider the infinite horizon case, i.e. the limiting case of going backwards infinitely many time steps in problem (2.4) (without any additional discount factor). In this scenario, we can find a fixed point of the DP described by (2.4) based on the following assumptions.

Assumption 2.1. *The marginal cost of an additional, feasible order is always smaller than the maximum marginal profit, i.e.* $C(x + 1_s) - C(x) \leq \bar{d} + r$*, for all* $(x, s) \in X \times F(x)$*.*

Recall that the components of the delivery charge vector d can take real-valued and infinite values and thus $p(\cdot)$ is an extended function, whose domain contains infinite-valued points.

Assumption 2.2. *We assume that the transition probability density function has the following properties. For any* $s \in S$*:*

(a) $p_s(d) > 0$*, if* $d_s \in [\underline{d}, \bar{d}\,]$ *and* $d_{s'} \in [\underline{d}, \bar{d}\,] \cup \{\infty\}$ *for all* $s' \in S$*, such that* $s' \neq s$*.*

(b) $p_s(d)d_s = 0$*, if* $d_s = \infty$ *and* $d_{s'} \in [\underline{d}, \bar{d}\,] \cup \{\infty\}$ *for all* $s' \in S$*, such that* $s' \neq s$*.*

Assumption 2.1 is not restrictive, since it offers the means to ensure that every additional, feasible order can generate profit. Otherwise, the delivery slot price, which maximises (2.4), would be $d_s = \infty$ for some $s \in S$, even if additional orders would still be feasible for that slot. Assumption 2.2(a) is not restrictive either, since we can change $\bar{d}$ to a value for which the choice probability for all slots is modelled to be arbitrarily small, yet positive. Assumption 2.2(b) is also not restrictive, since it effectively ensures that infinite prices do not generate infinite expected returns. Furthermore, it is easy to show that Assumption 2.2(b) holds for the adopted multinomial logit model. Based on the aforementioned definitions and Assumptions 2.1 and 2.2, we formulate our first result, proof of which is deferred to Section 2.4.1.

Theorem 2.1. *Under Assumptions 2.1 and 2.2, the unique fixed point of* (2.4) *is given by*

$$V^*(x) := (\bar{d} + r)\mathbf{1}^\intercal(\bar{x} - x) - C(\bar{x}), \text{ for all } x \in X. \tag{2.9}$$

There is a natural interpretation of this perhaps surprisingly compact result. The fixed point of the DP is a hyperplane in x, where each element of the gradient of V^* is equal to $-(\bar{d} + r)$, or equivalently, the opportunity cost of an order is $V_t(x) - V_t(x + 1_s) = \bar{d} + r$ for all $(x, s) \in X \times F(x)$. Therefore, the only optimal selection of delivery slot prices is to choose $\bar{d}$ for all delivery slots $s \in F(x)$. For any other choice the opportunity costs would be larger than the revenue generated by any order.

This result makes intuitive sense as in the limit as $t \to -\infty$, there will always arrive enough customers who will be willing to pay $\bar{d}$ for a delivery. Therefore, in the infinite time horizon case, it is best to always try to charge the maximum admissible delivery charge. The value obtained in this case is $\bar{d} + r$ multiplied by the number of residual feasible orders $\mathbf{1}^\intercal(\bar{x} - x)$, from which the terminal cost $C(\bar{x})$ is then subtracted. Note that this result is not of direct practical use as a pricing policy. However, it serves as a useful upper bound to the value function for all time steps, which we exploit as an initial value function approximation in the numerical example in Section 2.5.3, to speed up computation.

2.3.2 Finite time horizon results

For finite $\bar{t}$, we establish a geometric property of the value function V_t, for all $t \in T$, related to concavity of a continuous function. Since the domain of V_t is discrete, it is not possible to establish this property from standard convexity theory.

We first provide some definitions before stating our main result. Let the opportunity cost of an order in slot s at time t be denoted by

$$\gamma_{s,t}(x) := V_t(x) - V_t(x + 1_s) \text{ for all } (x, s, t) \in X \times S \times T. \tag{2.10}$$

Let $\gamma_t(x) := [\gamma_{1,t}(x), \gamma_{2,t}(x), \ldots, \gamma_{n,t}(x)]^\intercal$ for all $(x, t) \in X \times T$. Define the set of stochastic vectors in a set A as

$$\mathcal{V}_A := \left\{ v \in \mathbb{R}_+^{|A|} \mid \mathbf{1}^\intercal v = 1 \right\}, \tag{2.11}$$

where v_i denotes the i-th component of v. Let $a \in \mathbb{Z}^n$ and let $E \subseteq \mathbb{Z}^n$ be a finite set. Then E is defined to be an *enclosing set* of a if $a \in \mathrm{conv}(E)$. We define $\mathcal{E}(a)$ as the set of all sets E enclosing a. The following two definitions are frequently used in discrete convex analysis:

Definition 2.1 (see [40, (2.1)])**.** *The* concave closure $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ *of a function* $f : \mathbb{Z}^n \to \mathbb{R} \cup \{-\infty\}$ *is defined as*

$$\tilde{f}(x) := \inf_{a,b} \left\{ a^\intercal x + b \;\middle|\; a^\intercal y + b \geq f(y) \quad \forall y \in \mathbb{Z}^n \right\}. \tag{2.12}$$

Definition 2.2 (see [40, Lemma 2.3])**.** *A function* $f : \mathbb{Z}^n \to \mathbb{R} \cup \{-\infty\}$ *is* concave extensible *if and only if the evaluations of* f *coincide with the evaluations of its concave closure* $\tilde{f}$*, i.e.* $f(x) = \tilde{f}(x)$ *for all* $x \in \mathbb{Z}^n$*.*

An equivalent characterisation of concave extensibility for functions constrained to a compact domain, X in our case, is given by the following result.

Proposition 2.2. *A function* $f : X \to \mathbb{R} \cup \{-\infty\}$ *is* concave extensible *if and only if, for all* $x \in X$ *and for all* $E \in \mathcal{E}(x)$*, the evaluation of* f *at* x *does not lie below any possible convex combination of* f *on the points* $e \in E$*, i.e. for all* $x \in X$*, for all* $E \in \mathcal{E}(x)$ *and for all* $\mu \in \mathcal{V}_E$*, such that* $x = \sum_{e \in E} \mu_e e$*, it holds that*

$$f(x) \geq \sum_{e \in E} \mu_e f(e). \tag{2.13}$$

Proof. The proof is based on the discussion around (2.2)–(2.3) in [40] on a similar result. Proposition 2.2 is shown by equating $f(x) = \tilde{f}(x)$ for all $x \in X$ and exploiting the linear programming duality (see [41]), which yields

$$\begin{aligned} f(x) &= \tilde{f}(x) \\ &= \inf_{a,b} \left\{ a^\intercal x + b \;\middle|\; a^\intercal y + b \geq f(y) \quad \forall y \in X \right\} \\ &= \sup_{\tilde{\mu}} \left\{ \sum_{y \in X} \tilde{\mu}_y f(y) \;\middle|\; \sum_{y \in X} \tilde{\mu}_y y = x, \tilde{\mu} \in \mathcal{V}_X \right\} \\ &\geq \sup_{\mu} \left\{ \sum_{e \in E} \mu_e f(e) \;\middle|\; \sum_{e \in E} \mu_e e = x, \mu \in \mathcal{V}_E \right\}, \end{aligned} \tag{2.14}$$

where the inequality is achieved by constraining the set to any arbitrary $E \subseteq X$. Since the inequality holds for the supremum, it also holds for all other choices of μ, as required to show the result. $\square$

The geometric intuition behind this result is shown in Fig. 2.2: Each point of the concave closure$\tilde{f}$ can be found by either finding the smallest hyperplane that is an upper bound to f or by computing the largest value from all convex combinations of f on enclosing points. Based on these definitions, we impose the

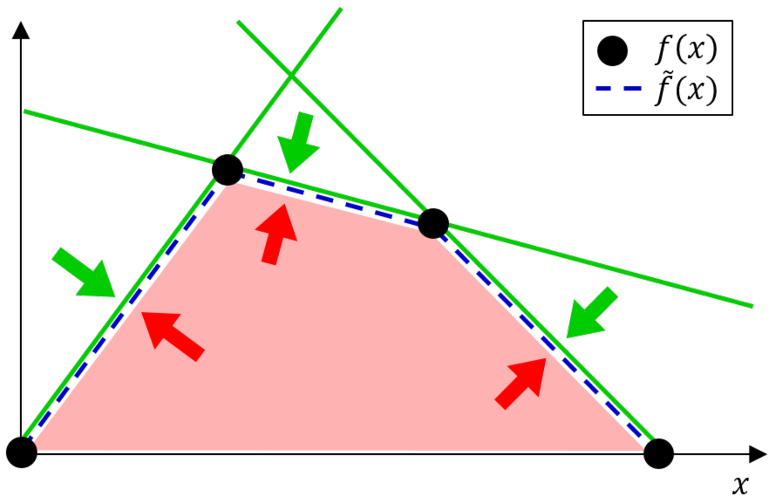


Figure 2.2: Two equivalent interpretations of the concave closure $\tilde{f}$ (blue, dotted line) of a function f (black dots): by minimising hyperplanes (green, solid lines, Definition 2.1) and by maximising convex combinations (red, shaded area, Proposition 2.2). Arrows include the directions of optimisation.

following assumptions on our finite time horizon result.

Assumption 2.3. *We assume that the opportunity cost at the terminal condition* $\gamma_{s,\bar{t}+1}$ *of all orders is increasing in* x *for all unit steps in* X*, i.e.* $\gamma_{s,\bar{t}+1}(x) < \gamma_{s,\bar{t}+1}(x + 1_{s'})$ *for all* $(x, s, s') \in X \times F(x) \times F(x)$*, such that* $s \neq s'$*.*

Assumption 2.4. *The function* $-C$ *is concave extensible.*

Assumption 2.5. *The customer choice probability distribution follows a multinomial logit model, i.e. it can be expressed as stated in* (2.1)*.*

Assumption 2.3 is satisfied if $V_{\bar{t}+1}$ is *strictly submodular*. This follows directly from the definition of (strict) submodularity: A function $f : \mathbb{Z}^n \to \mathbb{R}$ is *(strictly) submodular* if we have

$$f(\max(y,z)) + f(\min(y,z)) \underset{(<)}{\leq} f(y) + f(z) \tag{2.15}$$

for all y and $z \in \mathrm{dom}(f)$, where the maximum and minimum are taken component-wise (e.g. see [42, Definition 3.2]). This means that f has increasing opportunity costs, since, for all $(x, s, s') \in X \times S \times S$, such that $s \neq s'$, we can set the terms in (2.15) to be $f = V_{\bar{t}+1}$, $y = x + 1_s$, $z = x + 1_{s'}$, which yields the desired inequality:

$$\begin{aligned} V_{\bar{t}+1}(x + 1_s + 1_{s'}) + V_{\bar{t}+1}(x) &< V_{\bar{t}+1}(x + 1_s) + V_{\bar{t}+1}(x + 1_{s'}) \\ \iff V_{\bar{t}+1}(x) - V_{\bar{t}+1}(x + 1_s) &< V_{\bar{t}+1}(x + 1_{s'}) - V_{\bar{t}+1}(x + 1_s + 1_{s'}) \\ \iff \gamma_{s,\bar{t}+1}(x) &< \gamma_{s,\bar{t}+1}(x + 1_{s'}), \end{aligned} \tag{2.16}$$

where the last reformulation is due to the definition of opportunity costs in (2.10). Since $V_{\bar{t}+1}$ needs to be strictly submodular, this requires that C is strictly supermodular as $V_{\bar{t}+1}(x) = -C(x)$ for all $x \in X$.

This is not the case for all C used in the literature. For example, [20] use an affine cost function, which is supermodular, but not strictly so. However, our results are also relevant for situations with affine cost functions, since – as we show numerically in Section 2.5 – the value function can reach a state where Assumption 2.3 is satisfied in a small number of iterations of the Bellman operator. Assumption 2.4 is weak as it is satisfied by any convex cost function, which also includes the aforementioned affine cost functions.

Assumption 2.5 indirectly assumes that the customer choice model satisfies the so-called *assumption on independence of irrelevant alternatives*, which states that the presence (or absence) of a slot does not change the ratio of choice probabilities of any other pair of slots [22], as one can see by using (2.1) and evaluating

$$\frac{p_s(d)}{p_{s'}(d)} = \frac{\exp(\beta_c + \beta_s + \beta_d d_s)}{\exp(\beta_c + \beta_{s'} + \beta_d d_{s'})} \tag{2.17}$$

for all $(s, s') \in S \times S$, such that $s \neq s'$. We observe that (2.17) only depends on d_s and $d_{s'}$, but not the other components of d, such that assumption on independence of irrelevant alternatives holds. While the multinomial customer choice model may not be the most accurate representation of the actual customer behaviour, it is still popularly used in the literature, see e.g. [20, 28], since optimal prices under this model are the unique solution of a convex optimisation problem [25].

Theorem 2.3. *Under Assumptions 2.1 and 2.3–2.5, there exists a sufficiently small* $\lambda > 0$, *such that* V_t *is finite-valued, concave extensible in* x *for all* $t \in T$.

Since there is only one part in the subsequent proof of Theorem 2.3 where we use Assumption 2.3 we can, as an alternative to Theorem 2.3, relax the underlying assumption on the terminal condition of the DP in exchange for a different, albeit also restrictive assumption as follows:

Assumption 2.6. *Fix any* $t \in T$. *If* V_t *is concave extensible in* x, *then for all*[1] $z := (x, p) \in X \times P$, *for all* $E \in \mathcal{E}(z)$ *with elements* $(e^{(x)}, e^{(p)}) = e \in E$ *and for all* $\mu \in \mathcal{V}_E$, *such that* $z = \sum_{e \in E} \mu_e e$, *it is true that*

$$\begin{aligned} &\sum_{s \in S} p_s \left[V_t(x + 1_s) - V_t(x)\right] + V_t(x) \\ &\geq \sum_{e \in E} \mu_e \left\{ \sum_{s \in S} e_s^{(p)} \left[V_t\left(e^{(x)} + 1_s\right) - V_t\left(e^{(x)}\right)\right] + V_t\left(e^{(x)}\right)\right\}. \end{aligned} \tag{2.18}$$

While Assumption 2.6 is strong in general, it exhibits a geometric intuition. Fix any $z \in X \times P$ and consider the following special case: Suppose that, for some enclosing set $E \in \mathcal{E}(z)$, the condition for concave extensibility (2.13) is satisfied with equality and all the points lie on a hyperplane. Then $V_t(x + 1_s) - V_t(x) = V_t(e + 1_s) - V_t(e)$ for all $(x, E, s) \in X \times E \times S$. Therefore, (2.13) holds with equality, independently of the particular choice of λ.

Despite not being necessarily satisfied for all conceivable cases, Assumption 2.6 allows us to relax Assumption 2.3 and thus extends our analysis for more general approximate delivery cost functions, including the affine cost functions proposed in [20, 28], by means of the following result.

[1] Section 2.4.2 details why we have dropped the argument d in $p(d)$ and $p_s(d)$ and write simply p and p_s instead.

Corollary 2.4. *Under Assumptions 2.1 and 2.4–2.6, there exists a sufficiently small $\lambda > 0$, such that V_t is finite-valued, concave extensible in x for all $t \in T$.*

In the next section, we present the proofs of Theorems 2.1 and 2.3, as well as Corollary 2.4.

2.4 Proofs of main results

2.4.1 Proof of infinite time horizon theorem

To prove Theorem 2.1, we first note that the DP in (2.4) can be reformulated as a stochastic shortest path problem (2.6). The Bellman operator mapping of this class of problems is known to be contractive (see [37, Chapters 1 and 3] and [29, Lemma 5]). Therefore, the DP in (2.4) admits a unique fixed point. We start with the necessary and sufficient condition for the Bellman operator $\mathcal{T}$ to have a fixed point V^*, i.e. a point satisfying $V^* = \mathcal{T}V^*$. Setting $V_t(x) = V_{t+1}(x) = V^*(x)$ in the opportunity cost formulation of the DP in (2.5) yields

$$\max_{d \in D} \left\{ \lambda \sum_{s \in S} p_s(d)[r + d_s + V^*(x + 1_s) - V^*(x)] \right\} = 0. \tag{2.19}$$

Substituting the candidate V^* from (2.9) into (2.19) then results in

$$\max_{d \in D} \left\{ \lambda \sum_{s \in S} p_s(d)[d_s - \bar{d}\,] \right\} = 0. \tag{2.20}$$

Fix any $s \in S$ and consider the following two possible cases:

Case I: Suppose that $d_s \in [\underline{d}, \bar{d}]$. Then by Assumption 2.2(a), $p_s(d) > 0$. Furthermore, the value of $[d_s - \bar{d}\,]$ is non-positive and 0 only if $d_s = \bar{d}$. Hence, the maximum value that $p_s(d)[d_s - \bar{d}]$ can take in this case is 0, namely when $d_s = \bar{d}$.

Case II: Suppose that $d_s = \infty$. Then by Assumption 2.2(b), $p_s(d)[d_s - \bar{d}] = 0$.

Since in both cases the maximum attainable value of each term in the sum over s is 0, the equality in (2.20) holds. Finally, notice that $V_t(\bar{x}) = C(\bar{x})$ for all $t \in T$. Since the candidate fixed point satisfies $V^*(\bar{x}) = V_{\bar{t}+1}(\bar{x}) = -C(\bar{x})$, V^* is a fixed point of $\mathcal{T}$ for all $x \in X$. □

2.4.2 Proof of finite time horizon theorem

In this section, we prove Theorem 2.3. We start by reformulating (2.4) as a maximisation over $p \in P$ instead of $d \in D$. As shown by [25], this is possible under Assumption 2.5, since, for all $s \in S$, the following unique mapping between p and d exists:

$$\frac{p_s}{p_0} = \exp(\beta_c + \beta_s + \beta_d d_s), \tag{2.21}$$

where we recall that $p_0 > 0$, since we have defined $\sum_{s \in S} p_s < 1$ and $p_0 := 1 - \sum_{s \in S} p_s$. Notice that we have dropped the argument of $p(d)$ and now simply write p, since we now view p and d not as function and variable, respectively, but as related elements of the two sets P and D. We solve this equation with respect to d_s to obtain

$$d_s = \beta_d^{-1} \left[\ln\left(\frac{p_s}{p_0}\right) - \beta_c - \beta_s \right]. \tag{2.22}$$

We will prove the theorem by induction. To this end, we fix an arbitrary $t \in (T \cup \{\bar{t}+1\}) \setminus \{1\}$, assume for an induction hypothesis that V_t is concave extensible in x and now show that $V_{t-1} = \mathcal{T} V_t$ is also concave extensible. Note that the base case in our induction proof is captured by Assumption 2.4. By substituting (2.22) into (2.5) we obtain

$$\begin{aligned} \mathcal{T} V_t(x) &= \max_{p \in P} \lambda \sum_{s \in S} p_s \left\{ r + \beta_d^{-1} \left[\ln\left(\frac{p_s}{p_0}\right) - \beta_c - \beta_s \right] + V_t(x + 1_s) - V_t(x) \right\} + V_t(x) \\ &= \max_{p \in P} \{ f(p) + g_t(x,p) \}, \end{aligned} \tag{2.23}$$

where we have defined

$$\begin{aligned} f(p) :=& \lambda \sum_{s \in S} p_s \left\{ r + \beta_d^{-1} \left[\ln\left(\frac{p_s}{p_0}\right) - \beta_c - \beta_s \right] \right\}, \\ g_t(x,p) :=& \lambda \sum_{s \in S} p_s \left\{ V_t(x + 1_s) - V_t(x) \right\} + V_t(x) \end{aligned} \tag{2.24}$$

for all $(x,p) \in X \times P$. This allows us to formulate the following result.

Lemma 2.5. *For all $t \in (T \cup \{\bar{t}+1\}) \setminus \{1\}$, the functions f and g_t have the following properties:*

(i) The function f is concave in p.

(ii) Under Assumptions 2.3 and 2.5 and if V_t is concave extensible in x, there exists a sufficiently small $\lambda > 0$ such that the function g_t is concave extensible in (x, p).

The proof of Lemma 2.5(i) is given in the appendix of this chapter, i.e. Section 2.A.1. In Lemma2.5(ii), we assume that V_t is concave extensible in x as this is embedded within our induction proof for Theorem 2.3, where this corresponds to our induction hypothesis. The proof of Lemma 2.5(ii) depends on the following self-contained result. Let us consider a relaxation of the constraint on the optimisation variable d in (2.4) and optimise over $\mathbb{R}^n$ instead of D. We refer to this problem as the *unconstrained* DP as opposed to the original, *constrained* DP.

Proposition 2.6. *The constrained and unconstrained version of the DP share the following property:*

(i) Consider the unconstrained DP. Under Assumption 2.3, the opportunity cost $\gamma_{s,t}$ of all orders is increasing for all unit steps in X, i.e.

$$\gamma_{s,t}(x + 1_{s'}) > \gamma_{s,t}(x) \tag{2.25}$$

for all $(x, s, s', t) \in X \times F(x) \times F(x) \times T$, such that $s \neq s'$.

(ii) The property in (2.25) also holds for the constrained DP.

We prove the two parts of Proposition 2.6 in Appendix 2.A.2 and use these results to prove Lemma 2.5(ii) in the same appendix. An interesting implication of Proposition 2.6 is that increasing opportunity costs in x cause unconstrained optimal prices to exhibit monotonic behaviour in x:

Lemma 2.7. *Under Assumptions 2.1 and 2.3–2.5, there exists a sufficiently small $\lambda > 0$, such that for all slots $s \in S$, the optimal price in the unconstrained DP for this slot, for any state $x \in X$, denoted by d_s^* is non-decreasing in x_s and non-increasing in $x_{s'}$, for all $s' \in S \setminus \{s\}$.*

We prove this result in 2.A.3. Lemma 2.7 also complements the research on the price-inventory relationship under multinomial logit customer choice, i.e. the function describing prices in terms of remaining order capacity, i.e. d^* as a function of $\bar{x} - x$, for all time steps in the booking horizon (see [39] for a review).

For any given customer choice model, it is desirable to understand if the price-inventory function exhibits monotonic behaviour, e.g. if prices for a slot increase or decrease monotonically as orders for that slot (or another slot) increase. For the case of multinomial logit customer choice, such statements cannot be made without additional assumptions. For example, [38, Section 6.2] show that prices for a slot are not necessarily non-increasing as orders in a different slot increase. For the two-dimensional case, i.e. $n = 2$ in our notation, [34, Proposition 2] show that prices for a slot are non-decreasing in the number of orders for *that* slot. The prices for the *other* slot are not monotonic in general (see [34, Fig. 1(a)] for an example). Our results, however, provide sufficient assumptions to make monotonicity statements for the multi-dimensional case. We illustrate this numerically in the example of Section 2.5.2.

By Lemma 2.5, there exists a sufficiently small λ such that g_t has a continuous extension $\tilde{g}_t$, which is jointly concave in (x, p). By inspection, f is only a function of the continuously-valued variable p. Therefore, $f(p) + \tilde{g}_t(x, p)$ is also jointly concave in (x, p). We need one final auxiliary result to complete the proof.

Lemma 2.8. *Under Assumption 2.5 and for all $x \in X$, the set*

$$\hat{P} := \left\{ [p_s(d)]_{s \in F(x)} \;\middle|\; d_s \in [\underline{d}, \bar{d}\,] \quad \forall s \in F(x) \right\} \tag{2.26}$$

is a convex set.

We prove Lemma 2.8 in Appendix 2.A.4. We define

$$\begin{aligned} U(x) &:= \max_{p \in P} \{ f(p) + \tilde{g}_t(x, p) \} \\ &= \max_{p \in P} \left\{ \lambda \sum_{s \in S} p_s \left\{ r + \beta_d^{-1} \left[\ln \left(\frac{p_s}{p_0} \right) - \beta_c - \beta_s \right] + \tilde{V}_t(x + 1_s) - \tilde{V}_t(x) \right\} \right. \\ &\qquad \left. + \tilde{V}_t(x) \right\} \end{aligned}$$

$$= \max_{p \in P} \left\{ \lambda \sum_{s \in F(x)} p_s \left\{ r + \beta_d^{-1} \left[\ln\left(\frac{p_s}{p_0}\right) - \beta_c - \beta_s \right] + \tilde{V}_t(x + 1_s) - \tilde{V}_t(x) \right\} \right.$$
$$\left. + \tilde{V}_t(x) \right\}$$
$$= \max_{p \in \hat{P}} \left\{ \lambda \sum_{s \in F(x)} p_s \left\{ r + \beta_d^{-1} \left[\ln\left(\frac{p_s}{p_0}\right) - \beta_c - \beta_s \right] + \tilde{V}_t(x + 1_s) - \tilde{V}_t(x) \right\} \right.$$
$$\left. + \tilde{V}_t(x) \right\}, \tag{2.27}$$

where the second equality follows from (2.23), the third equality from the observation that terms summed over unfeasible slots evaluate to 0 and the last equality from noticing that the expression is independent of p_s for all $s \notin F(x)$.

By [43, Proposition 2.22] or [44, Section 3.2.5], partial maximisation with respect to some variables of a continuous multivariate function that is jointly concave in all its variables, preserves concavity in the resulting function if the optimisation variable is constrained to a convex set. Since $\hat{P}$ is convex by Lemma 2.8, U is a concave function of x.

It remains to show that $U(x) = \mathcal{T}V_t(x)$ for all discrete-valued $x \in X$. Repeating the same calculation, now with the discrete $f(p) + g_t(x, p)$ in place of $f(p) + \tilde{g}_t(x, p)$, i.e. $\mathcal{T}V_t(x) = \max_{p \in P} \{f(p) + g_t(x, p)\}$, note that $\tilde{f}(p) + \tilde{g}_t(x, p) = f(x) + g_t(x, p)$ for all $x \in X$ by Definition 2.2(a). Therefore, $\mathcal{T}V_t(x) = U(x)$ for all $x \in X$. This shows that U is a continuous extension of $\mathcal{T}V_t$, which is concave in x. Hence, $\mathcal{T}V_t$ is concave extensible in x. This concludes our induction argument and shows that the value function V_t is concave extensible in x for all $t \in T$. □

2.4.3 Proof of Corollary 2.4

The proof follows the previous proof of Theorem 2.3 exactly, up to the proof of its supporting Lemma 2.5(ii), which is now trivially true by Assumption 2.6. Hence Lemma 2.5(ii) holds for this case without invoking Assumption 2.3, as it was needed for the proof of Theorem 2.3. The remainder of the proof is then equivalent to the rest of the previous proof of Theorem 2.3. □

2.5 Numerical examples

In the following two sections, we illustrate the validity and practical utility of our results. We first present a low-dimensional numerical example of a 2-slot problem and then the application of a non-linear stochastic dual DP algorithm to a 17-slot problem.

2.5.1 Increasing opportunity costs in an illustrative 2-slot example

To illustrate our results, consider a 2-slot problem, where we fix the parameters listed in Table 2.1(a) and vary the terminal conditions as listed in Table 2.1(b).

Table 2.1: The parameters of the numerical example.

(a) Fixed parameters.

$(\lambda, \bar{t}, n)$	$(0.5, 200, 2)$
$\left(\underline{d}, \bar{d}, r\right)$	$(0, 2, 2)$
$\bar{x}$	$[4, 4]^\intercal$
$(\beta_c, \beta_d, \beta_1, \beta_2)$	$(0, -1, 1, -1)$

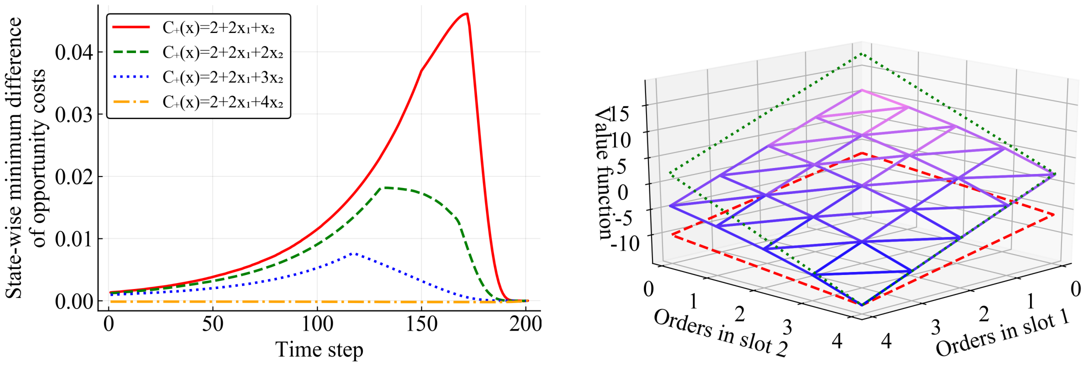
(b) Variable terminal conditions.

Terminal condition 1	$C_+(x) = 2 + 2x_1 + x_2$
Terminal condition 2	$C_+(x) = 2 + 2x_1 + 2x_2$
Terminal condition 3	$C_+(x) = 2 + 2x_1 + 3x_2$
Terminal condition 4	$C_+(x) = 2 + 2x_1 + 4x_2$

Notice that all terminal conditions violate Assumption 2.3. However, after a few iterations of the Bellman operator, the opportunity costs become strictly increasing by inspection for all terminal conditions, thus satisfying Assumption 2.3 if $\bar{t}$ is set to that time instance. This can be seen in Fig.2.3(a), where we plot the quantity

$$\min_{(x,s,s') \in X \times F(x) \times F(x),\, s \neq s'} \gamma_{s,t}(x + 1_{s'}) - \gamma_{s,t}(x) \tag{2.28}$$

for all time steps $t \in T$. Observing that this quantity is always non-negative shows numerically that the result in Proposition 2.6 with non-strict inequality holds if the terminal condition has non-decreasing opportunity costs, i.e. constant in our case, since all C_+ are affine functions of x. Furthermore, it is easy to verify that the resulting value function is concave extensible by computing the concave closure of V_t for all $t \in T$ and checking that the concave closure $\tilde{V}_t(x) = V_t(x)$ for all $x \in X$. For this example, we can verify by direct observation that the function is concave extensible by plotting V_t against (x_1, x_2) as shown in Fig. 2.3(b) for the first terminal condition in Table 2.1(b) at time step $t = \bar{t} - 10$.



(a) The state-wise minimum difference of opportunity costs is non-negative for all time steps t in the booking horizon and for all four terminal conditions of Table 2.1(b).

(b) The value function at $t = \bar{t}+1$ (red, dashed line), at the fixed point $t = -\infty$ (green, dotted line) and at $t = \bar{t} - 10$ (blue/violet colour gradient).

Figure 2.3: Illustrative example of a 2-slot problem.

In Fig.2.3(b), we also include the terminal condition (red, dashed line)

$$V_{\bar{t}+1}(x) := -C_+(x) = -2 - 2x_1 - x_2 \tag{2.29}$$

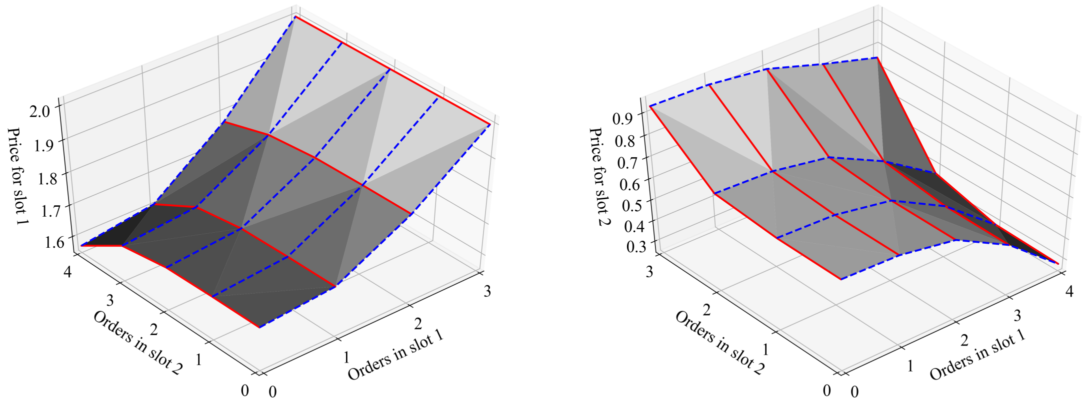
and, from Theorem 2.1, we also plot the fixed point (green, dotted line)

$$\begin{aligned} V^*(x) &:= (\bar{d} + r)\mathbf{1}^\intercal(\bar{x} - x) - C(\bar{x}) \\ &= 18 - 4(x_1 + x_2), \end{aligned} \tag{2.30}$$

which corresponds to $V^* = \lim_{t\to-\infty} V_t$. When it comes to approximating V_t, e.g. by means of basis functions, we can use the observation that V_t always lies between the terminal condition and the fixed point to limit the range of basis functions, such that the approximated version of V_t also falls between these lower and upper bounds. Also notice that the value function at $t = \bar{t} - 10$ is concave extensible and has increasing opportunity costs.

2.5.2 Monotonicity of prices in the illustrative 2-slot example

We now illustrate the results of Lemma 2.7, namely that there exists a small enough $\lambda > 0$ such that, for all time steps $t \in T$, the optimal price of a slot is increasing in the number of orders of that slot and decreasing in the number of orders of any other slot. Consider the parameters of the numerical example from the previous section, Table 2.1(a) and the first terminal condition of Table 2.1(b). We obtain optimal prices for both slots by direct computation of the value function of the DP in 2.4. We show these prices for the time step $t = \bar{t} - 5$ in Fig. 2.4.



(a) Optimal price of slot 1 at time $t = \bar{t} - 5$ is non-decreasing in orders in slot 1 (indicated by the blue, dashed lines) and non-increasing in orders in slot 2 (indicated by the red, solid lines).

(b) Optimal price of slot 2 at time $t = \bar{t} - 5$ is non-decreasing in orders in slot 2 (indicated by the red, solid lines) and non-increasing in orders in slot 1 (indicated by the blue, dashed lines).

Figure 2.4: Monotonicity of prices.

It should be remarked that the monotonicity property of Lemma 2.7 was shown to hold for unconstrained problems. However, it appears to hold for this numerical example even in the presence of constraints on admissible prices. Proving such a property for constrained problems constitutes a direction of current research.

2.5.3 Exploiting concave extensibility in non-linear stochastic dual dynamic programming

We now consider a more realistic problem described by the parameters in Table 2.2, which we have adapted from a real-world data, multi sub-area case study from [20] to a single sub-area. Notice that due to the large state space, $|X| = (\bar{x} + 1)^n \approx 8.65 \times 10^{18}$, it is impossible to compute the value function by direct computation.

Table 2.2: The parameters of the non-linear stochastic dual DP example.

$(\lambda, \bar{t}, n)$	$(0.8, 65, 17)$
$\left(\underline{d}, \bar{d}, r\right)$	$(0, 10, 34.53)$
$\bar{x}$	$[12, \dots, 12]^\intercal$
(β_c, β_d)	$(-2.5087, -0.0766)$
$\{\beta_s\}_{s\in S}$	$\{-1.0305, -0.3591, 0.3107, 0.5922, 0.6154, 0.0796, 0.5356, -0.2415,$ $-0.6286, -1.6736, -0.4351, -0.161, 0, 0.2533, 0.0736, 0.562, 0.2346\}$
$C_+(x)$	$0.1736 \times \mathbf{1}^\intercal x$

Based on the ideas of [45] for DP problems with binary state variables, [46, Algorithm 3] have developed a stochastic dual DP algorithm for multi-stage stochastic mixed-integer non-linear optimisation problems. The origins of this technique go back to the original stochastic dual DP approach for linear optimisation problems (see [47, 48]). We implement this algorithm for the above problem parameters.

For this example, it suffices to state that the algorithm produces a value function approximation of the form of the pointwise minimum of hyperplanes in x for all $t \in T$. Since this approximate function is concave extensible in x, the algorithm can only converge to the exact value function if the original value function is also concave extensible. Through an iterative procedure, hyperplanes are added to refine the representation of the value function. Approximation progress across iterations is quantified by means of:

1. The algorithm produces a deterministic upper bound u on the total expected profit, i.e. $u \geq V_1(0)$.

2. The algorithm produces a stochastic lower bound l on the total expected profit, such that $\mathbb{E}l \leq V_1(0)$, which is computed from sample profits obtained in simulations of state transitions forward in time, while pricing is based on the approximate value function instead of the (unavailable) exact value function. Note that $\mathbb{E}$ denotes the associated expected value operator. The stochastic lower bound is then computed simply as the cumulative moving average of the sample profits obtained in all previous iterations.

Fig. 2.5 below shows how these bounds converge for the simulated scenario.

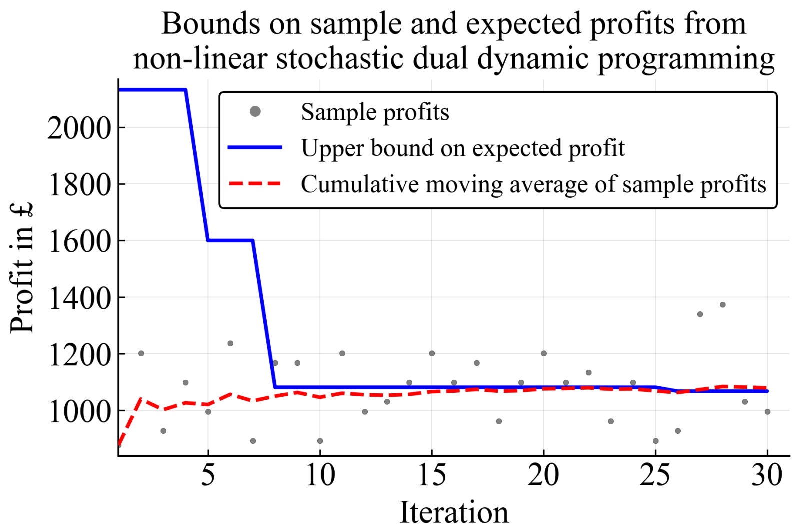


Figure 2.5: Plots of sample profits (grey dots), upper bound on expected profit (blue solid line) and cumulative moving average of sample profits, i.e. the stochastic lower bound on expected profit (red dashed line).

The deterministic upper and stochastic lower bounds of the algorithm only converge if the exact value function is concave extensible. Our theoretical analysis, which guarantees concave extensibility, allows us to employ this algorithm and provides theoretical support for its convergent behaviour. Hence, the concave extensibility preserving properties of the DP in (2.4), even in cases where we cannot numerically verify them due to the prohibitive problem size, open the road to employ techniques derived from stochastic dual DP theory.

Furthermore, since this approach is quite fast – we compute 30 iterations on an i7-8565U CPU with 1.80 GHz base frequency and with 16 GB in 4 minutes, 48 seconds – it is also possible to update the delivery cost function and hence the terminal condition of the DP using additional information like location of customers that have already placed an order. One could then re-run the DP to obtain approximately optimal prices that have adjusted to the perturbed delivery cost function before the next customer places an order.

The exact details of this algorithm are beyond the scope of this chapter and we explore them in greater detail in Section 4.2.2, where we introduce a slightly simpler variant of the algorithm in [46]. As a final remark, we also use the infinite time horizon result from Theorem 2.1 in the initialisation of this non-linear stochastic dual DP algorithm. The user-defined, initial value function approximation needs to be an upper bound to the exact value function. We can use the result of Theorem 2.1 for this purpose, since the fixed point provides a non-trivial upper bound on the value function for all time steps $t \in T$ and can easily be implemented, due to its low complexity, i.e. only a single hyperplane is needed.

2.6 Conclusion

In this chapter, we have studied the mathematical properties of the value function of an exact DP formulation of the revenue management problem in attended home delivery. We have shown that the recursive DP mapping has a unique, finite-valued fixed point and concavity-preserving properties. Hence, we have derived our main result stating that – under certain assumptions – for all time steps in the DP, the value function admits a continuous extension, which is a finite-valued, concave function of its state variables. We have illustrated our findings on a low-dimensional numerical example and an industry-sized problem.

Recent approaches have estimated V_t as an affine function of x for each $t \in T$ [20]. Based on our result, we believe that closer approximations can be found by pursuing different approximation strategies. One such strategy would be to adapt approximate DP algorithms from stochastic dual DP [47, 48]. We pursue this direction in the next chapter, where we tailor an algorithm based on the ideas of stochastic dual DP to this specific problem by exploiting the mathematical properties that we have derived in this chapter.

2.A Proofs

This appendix contains mathematical proofs omitted in the main body of Chapter 2.

2.A.1 Proof of Lemma 2.5(i)

It is shown in [25] that a structurally similar function to f is concave in its variables. We adopt a similar approach – computing the Hessian and showing that it is negative definite – to verify that f is jointly concave in all components of the vector p. We first compute the first-order partial derivatives of f:

$$\begin{aligned}\frac{\partial f}{\partial p_i} &= \lambda[r + \beta_d^{-1}(\ln(p_i/p_0) - \beta_c - \beta_i)] + \lambda\beta_d^{-1}, \text{ for all } i \in S,\\ \frac{\partial f}{\partial p_0} &= \lambda \sum_{s \in S} -p_s \beta_d^{-1} p_0^{-1}.\end{aligned} \tag{2.31}$$

The second-order partial derivatives are:

$$\begin{aligned}\frac{\partial^2 f}{\partial p_i^2} &= \lambda\beta_d^{-1}p_i^{-1}, \text{ for all } i \in S,\\ \frac{\partial^2 f}{\partial p_0^2} &= \lambda \sum_{s \in S} p_s \beta_d^{-1} p_0^{-2},\\ \frac{\partial^2 f}{\partial p_i \partial p_0} &= -\lambda\beta_d^{-1}p_0^{-1}, \text{ for all } i \neq 0,\\ \frac{\partial^2 f}{\partial p_i \partial p_j} &= 0, \text{ for all } (i,j) \in S \times S, \text{ such that } i \neq j.\end{aligned} \tag{2.32}$$

The resulting Hessian H of f with its second partial derivatives with respect to $\{p_i\}$ for all $i \in S \cup \{0\}$ is:

$$\begin{aligned}H &:= \lambda\beta_d^{-1}\begin{bmatrix} p_1^{-1} & \dots & 0 & -p_0^{-1}\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & p_n^{-1} & -p_0^{-1}\\ -p_0^{-1} & \dots & -p_0^{-1} & p_0^{-2}\sum_s p_s\end{bmatrix}\\ &=: \begin{bmatrix} A & B\\ B^\intercal & C\end{bmatrix},\end{aligned} \tag{2.33}$$

where we have defined block sub-matrices $A, B, B^\intercal$ and C of appropriate dimension, such that C is a scalar corresponding to the last entry of H. Note that A is negative definite, because $p_s \geq 0$ for all $s \in S \cup \{0\}$, $\lambda > 0$ and $\beta_d^{-1} < 0$. We compute the Schur complement of A in H:

$$C - B^\intercal A^{-1} B = \lambda \beta_d^{-1} \left(\sum_{s\in S} p_s p_0^{-2} - p_0^{-2} \sum_{s \in S} p_s \right) = 0. \tag{2.34}$$

As a result of $\lambda > 0$ and $\beta_d < 0$, A is negative definite and as H/A is non-positive, H is negative semi-definite (see e.g. [44, Appendix A.5.5]). This implies that f is concave in p. $\square$

2.A.2 Proof of Lemma 2.5(ii)

The proof of Lemma 2.5(ii) requires several intermediate results which we present in the following sections before returning to the proof of Lemma 2.5(ii).

Auxiliary function definitions and properties

In this section, we define some auxiliary functions and establish some of their properties that are needed in the subsequent sections. We define $W : \mathbb{R}_+ \mapsto \mathbb{R}_+$ as the inverse function of $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$, such that $f(x) := x \exp(x)$, i.e. implicitly defined through the relationship

$$x = W(x)\mathrm{e}^{W(x)}. \tag{2.35}$$

Note that W, as defined above, is the principal branch of the so-called Lambert W function, which is uniquely defined over the non-negative real numbers. To simplify notation in the following proof, we define two more functions: We define $\psi_s : \mathbb{R} \mapsto \mathbb{R}_{++}$ as

$$\psi_s(z) := \exp(\beta_c + \beta_s + \beta_d(z - r) - 1) \tag{2.36}$$

for all $s \in S$. We define the function $\phi : \mathbb{R}^n \mapsto \mathbb{R}_{++}$ as

$$\phi(z) := -\lambda \beta_d^{-1} W \left(\sum_{s \in S} \psi_s(z_s) \right), \tag{2.37}$$

where z_s indicates the s^{th} component of z. We can now establish some properties of ϕ that are instrumental for the subsequent proof of Lemma 2.5(ii).

Lemma 2.9. *The function ϕ has the following properties:*

(i) It is decreasing over its domain.

(ii) It satisfies the inequality:

$$\begin{aligned}&\phi(\gamma_t(x+1_{s'})) - \phi(\max\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\}) + \phi(\gamma_t(x+1_s))\\ \geq&\phi(\min\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\})\end{aligned} \tag{2.38}$$

for all $(x, s, s') \in X \times S \times S$, such that $s \neq s'$.

Proof. (i) It is useful to state the first derivative of W, which is

$$\frac{\mathrm{d}W}{\mathrm{d}y}(y) = \frac{W(y)}{y(1+W(y))}. \tag{2.39}$$

It suffices to show that the first partial derivative of W with respect to the components z_i for all $i \in S$ is negative. To this end, fix any $i \in S$. Setting $y = \sum_{s\in S} \psi_s(z_s)$, gives

$$\begin{aligned}\frac{\partial\phi}{\partial z_i}(z) =& - \lambda\beta_d^{-1}\frac{\mathrm{d}W}{\mathrm{d}y}(y)\frac{\partial y}{\partial z_i}(z)\\ =& - \lambda\beta_d^{-1}\frac{W(\sum_{s\in S}\psi_s(z_s))}{[\sum_{s\in S}\psi_s(z_s)]\,(1+W\,(\sum_{s\in S}\psi_s(z_s))}\\ &\times \beta_d\exp(\beta_c + \beta_i + \beta_d(z_i - r) - 1)\\ =& - \lambda\frac{W(\sum_{s\in S}\psi_s(z_s))}{1+W(\sum_{s\in S}\psi_s(z_s))}\frac{\psi_i(z_i)}{\sum_{s\in S}\psi_s(z_s)},\end{aligned} \tag{2.40}$$

where the last equality follows from the definition of ψ_i in (2.36). The customer arrival rate $\lambda \in (0, 1)$. The first fraction in (2.40) lies in $(0, 1)$ as $W(y) \geq 0$ for all $y \in \mathrm{dom}(W)$, while the second one lies in $(0, 1]$ as $\psi_s(y) > 0$ for all $y \in \mathrm{dom}(\psi_s)$ for all $s \in S$ and $n \geq 1$. Therefore, $\partial\phi/\partial z_i(z) \in (-1, 0)$ for all $i \in S$ and hence, it is negative.

(ii) Fix any $(i, s, s', x) \in S \times S \times S \times X$, such that $s \neq s'$. Let $\alpha_{i,t} := \max\{\gamma_{i,t}(x + 1_s), \gamma_{i,t}(x + 1_{s'})\}$ and $\beta_{i,t} := \min\{\gamma_{i,t}(x + 1_s), \gamma_{i,t}(x + 1_{s'})\}$ and notice that $\alpha_{i,t} \geq \beta_{i,t}$. Let us distinguish two cases.

Case I: Suppose that $\gamma_{i,t}(x + 1_s) \geq \gamma_{i,t}(x + 1_{s'})$. For all $j \in S, j \neq i$, define $\epsilon^{\alpha}_{j,t} := \gamma_{j,t}(x+1_s)$, $\epsilon^{\beta}_{j,t} := \gamma_{j,t}(x+1_{s'})$ and $\epsilon^{\alpha\beta}_{j,t} := \max\{\gamma_{j,t}(x + 1_s), \gamma_{j,t}(x + 1_{s'})\}$.

Under these assignments, the left-hand side of (2.38) can be equivalently written as

$$\begin{aligned}
&\phi(\gamma_t(x+1_{s'})) - \phi(\max\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\}) + \phi(\gamma_t(x+1_s)) \\
&= \phi(\epsilon^{\beta}_{1,t}, \ldots, \beta_{i,t}, \ldots, \epsilon^{\beta}_{n,t}) - \phi(\epsilon^{\alpha\beta}_{1,t}, \ldots, \alpha_{i,t}, \ldots, \epsilon^{\alpha\beta}_{n,t}) \\
&\quad + \phi(\epsilon^{\alpha}_{1,t}, \ldots, \alpha_{i,t}, \ldots, \epsilon^{\alpha}_{n,t}). \qquad (2.41)
\end{aligned}$$

Define the scalar function $f_\theta : A \mapsto \mathbb{R}$, where A contains all real numbers no smaller than $\beta_{i,t}$ and $\theta := \{\{\epsilon^{\alpha}_{j,t}, \epsilon^{\beta}_{j,t}, \epsilon^{\alpha\beta}_{j,t}\}_{j\neq i}, \beta_{i,t}\}$, such that

$$\begin{aligned}
f_\theta(\alpha_{i,t}) = \phi(\epsilon^{\beta}_{1,t}, \ldots, \beta_{i,t}, \ldots, \epsilon^{\beta}_{n,t}) - \phi(\epsilon^{\alpha\beta}_{1,t}, \ldots, \alpha_{i,t}, \ldots, \epsilon^{\alpha\beta}_{n,t}) \\
+ \phi(\epsilon^{\alpha}_{1,t}, \ldots, \alpha_{i,t}, \ldots, \epsilon^{\alpha}_{n,t}). \qquad (2.42)
\end{aligned}$$

Consider the derivative of f_θ, given by

$$\frac{\mathrm{d}f_\theta}{\mathrm{d}\alpha_{i,t}}(\alpha_{i,t}) = -\frac{\partial\phi}{\partial\alpha_{i,t}}(z^{\alpha\beta}) + \frac{\partial\phi}{\partial\alpha_{i,t}}(z^{\alpha}), \qquad (2.43)$$

where we have defined $z^{\alpha\beta} := [\epsilon^{\alpha\beta}_{1,t}, \ldots, \alpha_{i,t}, \ldots, \epsilon^{\alpha\beta}_{n,t}]$ as well as $z^{\alpha} := [\epsilon^{\alpha}_{1,t}, \ldots, \alpha_{i,t}, \ldots, \epsilon^{\alpha}_{n,t}]$. We compute the derivative of ϕ from the first equation in (2.40) to arrive at

$$\frac{\mathrm{d}f_\theta}{\mathrm{d}\alpha_{i,t}}(\alpha_{i,t}) = \lambda\beta_d^{-1}\frac{\mathrm{d}W}{\mathrm{d}y}(y^{\alpha\beta})\frac{\partial y^{\alpha\beta}}{\partial\alpha_{i,t}}(z^{\alpha\beta}) - \lambda\beta_d^{-1}\frac{\mathrm{d}W}{\mathrm{d}y}(y^{\alpha})\frac{\partial y^{\alpha}}{\partial\alpha_{i,t}}(z^{\alpha}), \qquad (2.44)$$

where $y^{\alpha\beta} := \sum_{s\in S}\psi_s(z^{\alpha\beta}_s)$ and $y^{\alpha} := \sum_{s\in S}\psi_s(z^{\alpha}_s)$. Substituting for the derivatives of $y^{\alpha\beta}$ and y^{α} and simplifying yields

$$\frac{\mathrm{d}f_\theta}{\mathrm{d}\alpha_{i,t}}(\alpha_{i,t}) = \lambda\psi_i(\alpha_{i,t})\left[\frac{\mathrm{d}W}{\mathrm{d}z}(z^{\alpha\beta}) - \frac{\mathrm{d}W}{\mathrm{d}z}(z^{\alpha})\right] \geq 0. \qquad (2.45)$$

The inequality follows from noting that $z^{\alpha} \geq z^{\alpha\beta}$ and that W has a negative second derivative over its domain. We conclude that f_θ is non-decreasing in $\alpha_{i,t}$, which means that f_θ is non-increasing by decreasing $\alpha_{i,t}$ to its minimum value $\alpha_{i,t} = \beta_{i,t}$. Repeating this minimisation for all $i \in S$, we obtain the following bound:

$$\phi(\gamma_t(x+1_{s'})) - \phi(\max\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\}) + \phi(\gamma_t(x+1_s))$$

$$\geq \phi(\min\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\}) - \phi(\min\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\})$$
$$+ \phi(\min\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\})$$
$$= \phi(\min\{\gamma_t(x+1_s), \gamma_t(x+1_{s'})\}), \tag{2.46}$$

as required.

Case II: Suppose that the roles of s and s' are now reversed, i.e. $\gamma_{i,t}(x+1_{s'}) \geq \gamma_{i,t}(x+1_s)$. Via symmetry, we reach the same conclusion as in Case I.

As both cases reach the same conclusion and collectively exhaust all possibilities, this concludes our proof and shows that (2.38) holds.

□

Proof of Proposition 2.6(i)

We start by reformulating the DP in (2.4) in terms of ϕ. Fix any $t \in T \setminus \{1\} \cup \{\bar{t}+1\}$. The unique optimisers of the unconstrained optimisation problem at time $t-1$, denoted by $d_s^*(x)$ for all $(x,s) \in X \times S$, is given by [20] based on the development of [25] as

$$d_s^*(x) = \gamma_{s,t}(x) - r - \beta_d^{-1} h_t(x) \tag{2.47}$$

for all $s \in S$, where $h_t(x)$ is the unique solution of

$$(h_t(x) - 1)\exp\left(h_t(x)\right) = \sum_{s\in S} \exp\left(\beta_c + \beta_s + \beta_d(\gamma_{s,t}(x) - r)\right). \tag{2.48}$$

We rewrite (2.48) equivalently as

$$\iff (h_t(x) - 1)\exp\left(h_t(x) - 1\right) = \sum_{s\in S} \exp\left(\beta_c + \beta_s + \beta_d(\gamma_{s,t}(x) - r) - 1\right)$$
$$\iff (h_t(x) - 1)\exp\left(h_t(x) - 1\right) = \sum_{s\in S} \psi_s(\gamma_{s,t}(x)), \tag{2.49}$$

where we have used ψ_s from (2.36). By the definition of W in (2.35), we obtain

$$h_t(x) = 1 + W\left(\sum_{s\in S} \psi_s(\gamma_{s,t}(x))\right). \tag{2.50}$$

Now, we can substitute (2.50) into (2.47):

$$d_s^*(x) = \gamma_{s,t}(x) - r - \beta_d^{-1}\left[1 + W\left(\sum_{s\in S} \psi_s(\gamma_{s,t}(x))\right)\right]. \tag{2.51}$$

Finally we can substitute (2.51) into the unconstrained version of (2.4) to obtain

$$\begin{aligned}V_{t-1}(x) =&\lambda \sum_{s\in S} p_s(d^*(x)) \left\{ r + \gamma_{s,t}(x) - r - \beta_d^{-1}\left[1 + W\left(\sum_{s'\in S}\psi_s(\gamma_{s',t}(x))\right)\right]\right.\\ &\left. - \gamma_{s,t}(x)\right\} + V_t(x)\\ =&\lambda \sum_{s\in S} p_s(d^*(x)) \left\{-\beta_d^{-1}\left[1 + W\left(\sum_{s'\in S}\psi_s(\gamma_{s',t}(x))\right)\right]\right\} + V_t(x),\end{aligned} \tag{2.52}$$

where we have defined $d^*(x) := [d_1^*(x), d_2^*(x), \ldots, d_n^*(x)]^\intercal$. We now substitute the customer choice model p evaluated at the optimiser $d^*(x)$ into (2.52):

$$\begin{aligned}V_{t-1}(x) = \lambda \sum_{s\in S} &\frac{\exp(\beta_c + \beta_s + \beta_d d_s^*(x))}{\sum_{s''\in S}\exp(\beta_c + \beta_{s''} + \beta_d d_{s''}^*(x)) + 1}\\ &\times \left\{-\beta_d^{-1}\left[1 + W\left(\sum_{s'\in S}\psi_s(\gamma_{s',t}(x))\right)\right]\right\} + V_t(x).\end{aligned} \tag{2.53}$$

Note that – using the definitions of d_s^* and h_t – the following relationship holds:

$$\begin{aligned}&\sum_{s\in S}\exp(\beta_c + \beta_s + \beta_d d_s^*(x))\\ =&\sum_{s\in S}\exp\left(\beta_c + \beta_s + \beta_d\left(\gamma_{s,t}(x) - r - \beta_d^{-1}h_t(x)\right)\right)\\ =&\sum_{s\in S}\exp\left(\beta_c + \beta_s + \beta_d\left(\gamma_{s,t}(x) - r\right)\right)\exp(-h_t(x))\\ =&\,(h_t(x) - 1)\exp(h_t(x))\exp(-h_t(x))\\ =&h_t(x) - 1,\end{aligned} \tag{2.54}$$

where the third equality follows from (2.48). Substituting (2.54) into (2.53), we obtain

$$V_{t-1}(x) = \lambda\frac{h_t(x) - 1}{h_t(x)}\left\{-\beta_d^{-1}\left[1 + W\left(\sum_{s\in S}\psi_s(\gamma_{s,t}(x))\right)\right]\right\} + V_t(x). \tag{2.55}$$

Substituting for $h_t(x)$ using (2.50) yields the desired expression:

$$\begin{aligned}V_{t-1}(x) =& \frac{\lambda W\left(\sum_{s\in S}\psi_s(\gamma_{s,t}(x))\right)}{1 + W\left(\sum_{s\in S}\psi_s(\gamma_{s,t}(x))\right)}\left\{-\beta_d^{-1}\left[1 + W\left(\sum_{s\in S}\psi_s(\gamma_{s,t}(x))\right)\right]\right\} + V_t(x)\\ =& -\lambda\beta_d^{-1}W\left(\sum_{s\in S}\psi_s(\gamma_{s,t}(x))\right) + V_t(x)\\ =& \,\phi(\gamma_t(x)) + V_t(x),\end{aligned} \tag{2.56}$$

where the last inequality follows from the definition of ϕ in (2.37). By Assumption 2.3, the terminal condition of the DP satisfies the property of increasing opportunity costs.

We can now prove Proposition 2.6(i) by showing that a monotonic mapping exists between $\gamma_{s,t-1}(x+1_{s'})-\gamma_{s,t-1}(x)$ and $\gamma_{s,t}(x+1_{s'})-\gamma_{s,t}(x)$ for all $(x,s,s',t)\in X\times F(x)\times F(x)\times(T\cup\{\bar{t}+1\}\setminus\{1\})$.

To this end, fix any $(x,s,t)\in X\times F(x)\times(T\cup\{\bar{t}+1\}\setminus\{1\})$. By using the definition of opportunity costs and (2.56) we can write the opportunity cost in state x with respect to an arbitrary slot $s\in S$ at time $t-1$ as

$$\gamma_{s,t-1}(x)=\gamma_{s,t}(x)+\phi(\gamma_t(x))-\phi(\gamma_t(x+1_s)). \tag{2.57}$$

Notice that $\gamma_{s,t-1}(x)\geq\gamma_{s,t}(x)$ for all $(x,s,t)\in X\times F(x)\times(T\cup\{\bar{t}+1\}\setminus\{1\})$, since ϕ is decreasing in its argument and $\gamma_t(x)\leq\gamma_t(x+1_s)$ for all $(x,s,t)\in X\times F(x)\ \times(T\cup\{\bar{t}+1\}\setminus\{1\})$. Fix any $s'\in F(x)$, such that $s'\neq s$. To prove the theorem, we require

$$\gamma_{s,t-1}(x+1_{s'})-\gamma_{s,t-1}(x)>0 \tag{2.58}$$

for all $t\in T\cup\{\bar{t}+1\}\setminus\{1\}$. Substitute (2.57) into the left-hand side of (2.58) to obtain

$$\begin{aligned}&\gamma_{s,t-1}(x+1_{s'})-\gamma_{s,t-1}(x)\\&=\gamma_{s,t}(x+1_{s'})-\gamma_{s,t}(x)+\phi(\gamma_t(x+1_{s'}))\\&\quad-\phi(\gamma_t(x+1_s+1_{s'}))-\phi(\gamma_t(x))+\phi(\gamma_t(x+1_s)).\end{aligned} \tag{2.59}$$

First note that $\gamma_t(x+1_s+1_{s'})\geq\gamma_t(x+1_s)$ and $\gamma_t(x+1_s+1_{s'})\geq\gamma_t(x+1_{s'})$. As $\phi(\gamma_t(x+1_s+1_{s'}))$ is decreasing in its argument by Lemma 2.9(i) and since it is subtracted on the right-hand side of the above equation, we can create the following lower bound:

$$\begin{aligned}&\gamma_{s,t-1}(x+1_{s'})-\gamma_{s,t-1}(x)\\&\geq\gamma_{s,t}(x+1_{s'})-\gamma_{s,t}(x)+\phi(\gamma_t(x+1_{s'}))\\&\quad-\phi(\max\{\gamma_t(x+1_s),\gamma_t(x+1_{s'})\})-\phi(\gamma_t(x))+\phi(\gamma_t(x+1_s)).\end{aligned} \tag{2.60}$$

Using Lemma 2.9(ii), we bound (2.60) from below by

$$\gamma_{s,t-1}(x+1_{s'})-\gamma_{s,t-1}(x)\geq\phi(\min\{\gamma_t(x+1_s),\gamma_t(x+1_{s'})\})-\phi(\gamma_t(x))$$

$$+\gamma_{s,t}(x+1_{s'})-\gamma_{s,t}(x), \tag{2.61}$$

where the minimum is taken element-wise. Since ϕ is positive by definition, we construct a lower bound on (2.61) by dropping the $\phi(\min\{\gamma_t(x+1_s),\gamma_t(x+1_{s'})\})$ term on the right-hand side. Furthermore, since ϕ is decreasing in its argument by Lemma 2.9(i), we create another lower bound on (2.61) from $\gamma_t(x)\geq\gamma_{\bar{t}+1}(0)$. This is possible since opportunity costs are increasing in x at $t=\bar{t}+1$ by Assumption 2.3 and since opportunity costs are non-increasing in time as discussed below (2.57). Hence, we obtain

$$\begin{aligned}
&\gamma_{s,t-1}(x+1_{s'})-\gamma_{s,t-1}(x)\\
&> -\phi(\gamma_{s,\bar{t}+1}(0))+\gamma_{s,t}(x+1_{s'})-\gamma_{s,t}(x)\\
&=\lambda\beta_d^{-1}W\left(\sum_{s\in S}\psi_s(\gamma_{\bar{t}+1}(0))\right)+\gamma_{s,t}(x+1_{s'})-\gamma_{s,t}(x).
\end{aligned} \tag{2.62}$$

Since only β_d^{-1} is negative, $\lambda\beta_d^{-1}W\left(\sum_s\psi_s(\gamma_{s,\bar{t}+1}(0))\right)<0$, independent of the choice of (s,s',t,x). Therefore, the above inequality describes a monotonic mapping from $\gamma_{s,t}(x+1_s')-\gamma_{s,t}(x)$ to $\gamma_{s,t-1}(x+1_s')-\gamma_{s,t-1}(x)$ for all $(s,s',t-1,x)\in F(x)\times F(x)\times T\times X$, such that $s\neq s'$. The bound on $\gamma_{s,t}(x+1_s')-\gamma_{s,t}(x)$ will therefore decrease as t decreases. Hence, $\gamma_{s,t}(x+1_s')-\gamma_{s,t}(x)$ will be minimal at $t=1$. Using the monotonicity of this mapping, we can find a λ for which $\gamma_{s,1}(x+1_s')-\gamma_{s,1}(x)>0$ by repetitively applying the above equation starting from the terminal condition at $t=\bar{t}+1$ to obtain

$$\gamma_{s,1}(x+1_{s'})-\gamma_{s,1}(x)\geq\bar{t}\lambda\beta_d^{-1}W\left(\sum_{s\in S}\psi_s(\gamma_{s,\bar{t}+1}(0))\right)+\gamma_{s,\bar{t}+1}(x+1_{s'})-\gamma_{s,\bar{t}+1}(x), \tag{2.63}$$

where the right-hand side is positive if

$$\begin{aligned}
&0<\bar{t}\lambda\beta_d^{-1}W\left(\sum_{s\in S}\psi_s(\gamma_{s,\bar{t}+1}(0))\right)+\gamma_{s,\bar{t}+1}(x+1_{s'})-\gamma_{s,\bar{t}+1}(x)\\
\iff &\lambda<-\beta_d\frac{\gamma_{s,\bar{t}+1}(x+1_{s'})-\gamma_{s,\bar{t}+1}(x)}{\bar{t}W\left(\sum_{s\in S}\psi_s(\gamma_{s,\bar{t}+1}(0))\right)}
\end{aligned} \tag{2.64}$$

for all $(x,s,s')\in X\times F(x)\times F(x)$, such that $s\neq s'$. The right-hand side of the above expression is strictly positive, because opportunity costs are increasing at the terminal condition by Assumption 2.3 and therefore, the numerator of the fraction is positive, W only takes positive values, $\bar{t} > 0$ and $\beta_d < 0$. As both X and $F(x)$ are finite sets, a small enough $\lambda > 0$ can be found that satisfies all inequalities described by (2.64). Therefore, a $\lambda > 0$ exists, such that $\gamma_{s,t}(x + 1_{s'}) > \gamma_{s,t}(x)$ for all $(x, s, s', t) \in X \times F(x) \times F(x) \times T$, such that $s \neq s'$. □

Proof of Proposition 2.6(ii)

Fix any $(x, t) \in X \times T$. Define $u := \gamma_t(x)$ and consider the function $w : \mathbb{R}^n \mapsto \mathbb{R}^n$ defined by

$$w(u) := u - r - \beta_d^{-1} \left[1 + W \left(\sum_{s \in S} \psi_s(u_s) \right) \right]. \tag{2.65}$$

In comparison with (2.51), notice that $w(u)$ is mathematically identical to $d^*(x) = [d_1^*(x), d_2^*(x), \ldots, d_n^*(x)]^\intercal$. Furthermore, W and ψ_s for all $s \in S$ are invertible functions. In particular, their inverses in the domain of interest are

$$\begin{aligned} x =& W(x) \exp\left(W(x)\right) \text{ and} \\ u =& \left(\ln\left(\psi_s(u)\right) - \beta_c - \beta_s + 1\right) \beta_d^{-1} + r \end{aligned} \tag{2.66}$$

for all $s \in S$. Therefore, w is a composition of invertible functions and hence, the mapping between w and u is bijective. Therefore, the mapping between $d^*(x)$ and $\gamma_t(x)$, equivalent to the mapping between w and u, is bijective.

If we constrain $d^*(x)$ to $D \subset \mathbb{R}^n$, we can conclude that there still exists a bijective mapping between D and (an unknown) $\Gamma \subset \mathbb{R}_+^n$ corresponding to the range of values that $\gamma_t(x)$ can take in this constrained scenario. Conversely, this means that no matter which $d^*(x) \in D$ maximises the *constrained* stage optimisation problem, there exists $\hat{\gamma}_t(x) \in \Gamma$, which is linked to the same $d^*(x)$ in the *unconstrained* problem. In other words, there exists $\hat{\gamma}_t(x) \in \Gamma$, which produces the same $V_{t-1}(x)$ in the *unconstrained* problem as $\gamma_t(x)$ does in the *constrained* problem. Due to this equivalence, the following statement is a necessary and sufficient condition for Proposition 2.6(i) to hold: Evaluating the *unconstrained* problem at $\hat{\gamma}_t(x)$, i.e. $V_{t-1}(x) = \phi(\hat{\gamma}_t(x)) + V_t(x)$ for all $x \in X$, there exists a sufficiently small $\lambda > 0$ that yields a value function at time $t-1$, whose opportunity cost is increasing in x. From

Proposition 2.6(i), we know that this statement holds true for all opportunity costs $\hat{\gamma}_t(x) \in \mathbb{R}_+^n$ under Assumption 2.3. Since $\hat{\gamma}_t(x) \in \Gamma \subset \mathbb{R}_+^n$, there exists a sufficiently small $\lambda > 0$ such that the opportunity cost will also be increasing in the *constrained* DP. $\square$

Completing the proof of Lemma 2.5(ii)

From Definition 2.2(b), g_t is concave extensible in (x, p) if

$$g_t(x,p) \geq \sum_{e \in E} \mu_e g_t\left(e^{(x)}, e^{(p)}\right) \tag{2.67}$$

for all $(x, p) \in X \times P$ and for all enclosing sets E, such that $[x,p]^\intercal = \sum_{e \in E} \mu_e \left[e^{(x)}, e^{(p)}\right]^\intercal$. We show that this inequality holds by starting from the right-hand side and substituting for $g_t(e^{(x)}, e^{(p)})$:

$$\begin{aligned}
\sum_{e \in E} \mu_e g_t(e^{(x)}, e^{(p)}) &= \sum_{e \in E} \mu_e \left[\lambda \sum_{s \in S} e_s^{(p)} \left\{V_t(e^{(x)} + 1_s) - V_t(e^{(x)})\right\} + V_t(e^{(x)})\right] \\
&= \sum_{e \in E} \mu_e \Bigg[\lambda \sum_{s \in F(e^{(x)})} e_s^{(p)} V_t(e^{(x)} + 1_s) \\
&\qquad + \Big(1 - \lambda \sum_{s \in F(e^{(x)})} e_s^{(p)}\Big) V_t(e^{(x)})\Bigg],
\end{aligned} \tag{2.68}$$

where we note that each of the summed terms in square brackets is a convex combination on the set $A(e^{(x)}) := \{e^{(x)} \cup \{e^{(x)} + 1_s\}_{s \in F(e^{(x)})}\}$, where the set of indices s is $F(e^{(x)})$ instead of S, since $e_s^{(p)} = 0$ for all $s \notin F(e^{(x)})$, i.e. assigning zero transition probability to infeasible slots. To show that (2.67) holds, we now derive the supporting result that there exists a small enough $\lambda > 0$, such that the concave closure $\tilde{V}_t(y)$, evaluated at any point $y \in X$, is a hyperplane on the set $A(y)$, which will make it possible to simplify (2.68) further.

Fix any $y \in X$ and consider the unit hypercube in the positive direction of y, which we define as $B(y) := \left\{z \;\middle|\; y \leq z \leq y + \sum_{s \in F(y)} 1_s\right\}$. We will show that only points in $A(y)$ form the concave closure around y by demonstrating that for all $y' \in B(y) \setminus A(y)$, the line segment between $(y, V_t(y))$ and $(y', V_t(y'))$ lies below a second line segment between two other points $(z, V_t(z))$ and $(z', V_t(z'))$ for some $(z, z') \in B(y) \times B(y)$, i.e. we will show that there exists some $(z, z') \in B(y) \times B(y)$, such that

$$\alpha_1 V_t(y) + (1 - \alpha_1) V_t(y') < \alpha_2 V_t(z) + (1 - \alpha_2) V_t(z'), \tag{2.69}$$

where $0 \leq \alpha_1 \leq 1$ and $0 \leq \alpha_2 \leq 1$, such that $\alpha_1 y + (1 - \alpha_1) y' = \alpha_2 z + (1 - \alpha_2) z'$ for all $y' \in B(y) \setminus A(y)$. This means that the line segments cannot be part of the concave closure $\tilde{V}_t$. We show this result by induction on $|F(y)|$. Consider the base case when $|F(y)| = 1$, then (2.69) is trivially satisfied as there is only a single element $s \in F(y)$, meaning that the set $B(y) \setminus A(y) = \emptyset$. Let $m := |F(y)|$. Suppose by means of an induction hypothesis that (2.69) holds for all cardinalities of $F(y)$ up to and including $m - 1$. Then the only line segment that we need to consider is the one connecting y and $y' = y + \sum_{s \in F(y)} 1_s$, because otherwise, we are in a lower-dimensional case, for which (2.69) holds by the induction hypothesis. For this choice of (y, y'), we can find a quadruple $(z, z', \alpha_1, \alpha_2)$ that satisfies (2.69) by repetitively invoking Proposition 2.6(ii) as follows: There exists a sufficiently small $\lambda > 0$, such that

$$\begin{aligned}
&\gamma_1(y) \\
&< \gamma_1(y + 1_2) \\
&< \gamma_1(y + 1_2 + 1_3) \\
&\vdots \\
&< \gamma_1\Big(y + \sum_{s \in \sigma} 1_s\Big),
\end{aligned} \tag{2.70}$$

where we have defined $\sigma := F(y) \setminus \{1\}$, which – strictly speaking – depends on y and $s = 1$, but we neglect this to ease notation. We can expand the first and last line of the above chained inequality by using the definition of opportunity costs:

$$\begin{aligned}
&\gamma_1(y) < \gamma_1\Big(y + \sum_{s \in \sigma} 1_s\Big) \\
\iff\ & V_t(y) - V_t(y + 1_1) < V_t\Big(y + \sum_{s \in \sigma} 1_s\Big) - V_t\Big(y + \sum_{s \in F(y)} 1_s\Big) \\
\iff\ & V_t(y) - V_t(y + 1_1) < V_t\Big(y + \sum_{s \in \sigma} 1_s\Big) - V_t(y'),
\end{aligned} \tag{2.71}$$

where we have used $y' = y + \sum_{s \in F(y)} 1_s$ as previously. Rearranging and multiplying both sides by $1/2$ yields

$$\iff \frac{1}{2}V_t(y) + \frac{1}{2}V_t(y') < \frac{1}{2}V_t\Big(y + \sum_{s\in\sigma} 1_s\Big) + \frac{1}{2}V_t(y + 1_1). \tag{2.72}$$

Comparing with (2.69), we see that $\alpha_1 = \alpha_2 = 1/2$, $z = y + \sum_{s\in\sigma} 1_s$ and $z' = y + 1_1$. To illustrate this for the case when $m = 3$, we plot the points $y, y', y + \sum_{s\in\sigma} 1_s$ and $y + 1_1$ as well as the line segments between the former and latter pair of points in Fig. 2.6 below.

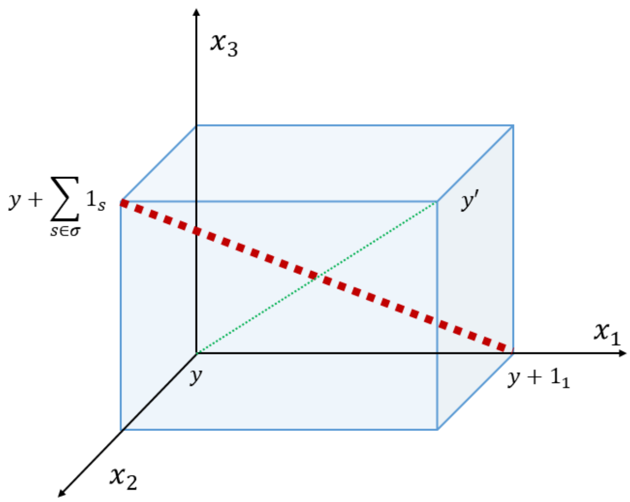


Figure 2.6: Green, dotted line segment between y and y' and red, dotted line segment between $y + \sum_{s\in\sigma} 1_s$ and $y + 1_1$. By (2.72) we have that at the intersection of the two lines, the interpolation of V_t between the two extreme points of the red line lies above the interpolation of V_t between the two extreme points of the green line. The light blue, solid line box spans all the points in $B(y)$.

Since we have found a quadruple $(z, z', \alpha_1, \alpha_2)$ that satisfies (2.69), we conclude that no $y' \in B(y) \setminus A(y)$ can be part of the concave closure $\tilde{V}_t$ for all $y \in B(y)$ and for all m. Therefore, only points in $A(y)$ can be part of the concave closure $\tilde{V}_t$ for all $y \in B(y)$ and, in fact, they all are, since the $|F(y)| + 1$ points in $A(y)$ uniquely define the support vectors of a hyperplane through $\{(y, V_t(y))\}_{y\in A(y)}$, i.e. each $y \in A(y)$ giving rise to one linearly independent

equality constraint, such that all $|F(y)|$ gradients and the offset of the hyperplane are uniquely defined. This holds true for unit hypercubes $B(y)$ for all $y \in X$

and hence, there exists a small enough $\lambda > 0$, such that $\tilde{V}_t$ is a hyperplane on the set $A(y)$ for all $y \in X$.

Returning to (2.68), we notice that the expression in square brackets is a convex combination on the set $A(e^{(x)})$, which means that $\tilde{V}_t$ is a hyperplane on this set. Hence, we can rewrite (2.68) with equality as

$$\begin{aligned}
\sum_{e\in E} \mu_e g_t(e^{(x)}, e^{(p)}) &= \sum_{e\in E} \mu_e \tilde{V}_t\left(\lambda \sum_{s\in S} e_s^{(p)} 1_s + e^{(x)}\right) \\
&\leq \tilde{V}_t\left(\sum_{e\in E} \mu_e \left[\lambda \sum_{s\in S} e_s^{(p)} 1_s + e^{(x)}\right]\right) \\
&= \tilde{V}_t\left(\lambda \sum_{s\in S} 1_s \sum_{e\in E} \mu_e e_s^{(p)} + x\right) \\
&= \tilde{V}_t\left(\lambda \sum_{s\in S} 1_s p_s + x\right) \\
&= \lambda \sum_{s\in S} p_s \left\{V_t(x+1_s) - V_t(x)\right\} + V_t(x) \\
&= g_t(x,p), \qquad (2.73)
\end{aligned}$$

where the second-last equality follows from the observation that the convex combination of V_t is evaluated on a set $A(x)$, for which $\tilde{V}_t$ is again a hyperplane, and the inequality is obtained by noticing that $\tilde{V}_t$ is concave in x and therefore, Jensen's inequality holds. From the above derivation, we conclude that $g_t(x,p) \geq \sum_{e\in E} \mu_e g_t(e^{(x)}, e^{(p)})$, as required. Therefore, there exists a sufficiently small $\lambda > 0$, for which g_t is concave extensible in (x,p) if V_t is concave extensible in x. $\square$

2.A.3 Proof of Lemma 2.7

Fix any $t \in T$. We prove the result by first showing that opportunity costs γ_t are non-decreasing in x and then showing that optimal prices exhibit the desired monotonicity property with respect to opportunity costs.

By Proposition 2.6(i), the inequality $\gamma_{s,t}(x + 1_{s'}) > \gamma_{s,t}(x)$, for all $(x, s, s') \in X \times F(x) \times F(x)$, such that $s \neq s'$, holds in the unconstrained DP. We now show that this inequality also holds for the case when $s = s'$ with non-strict inequality.

Theorem 2.3 states that under Assumptions 2.1 and 2.3–2.5, there exists a sufficiently small $\lambda > 0$, such that V_t is concave extensible in x and hence

$$\begin{aligned} 2V_t(x+1_s) &\geq V_t(x) + V_t(x+1_s+1_s) && \text{(2.74a)}\\ \iff V_t(x+1_s) - V_t(x+1_s+1_s) &\geq V_t(x) - V_t(x+1_s) \\ \iff \gamma_{s,t}(x+1_s) &\geq \gamma_{s,t}(x), && \text{(2.74b)} \end{aligned}$$

for all $(x,s) \in X \times F(x)$, where (2.74a) follows directly from the definition of concavity and the fact that V_t and its concave closure $\tilde{V}_t$ agree on $\{x, x+1_s, x+1_s+1_s\}$. Note that (2.74b) follows from the definition of $\gamma_{s,t}$. Taking both cases together, we conclude that there exists a sufficiently small $\lambda > 0$, such that $\gamma_{s,t}(x+1_{s'}) \geq \gamma_{s,t}(x)$, for all $(x,s,s') \in X \times F(x) \times F(x)$. Hence opportunity costs are monotonically non-decreasing in x.

We now show that the desired monotonicity property holds with respect to opportunity costs, which in turn implies that the property holds with respect to x. Recall from (2.51) that

$$d_s^*(x) = \gamma_{s,t}(x) - r - \beta_d^{-1}\left[1 + W\left(\sum_{s'\in S}\psi_{s'}(\gamma_{s',t}(x))\right)\right], \tag{2.75}$$

where W and ψ_s, for all $s \in S$ are defined in 2.A.2. Let $u := \gamma_t(x)$. By (2.40), we have that

$$\frac{\partial d_s^*}{\partial u_{s'}} = \begin{cases} 1 - \frac{W\left(\sum_{s\in S}\psi_s(u_s)\right)}{1+W\left(\sum_{s\in S}\psi_s(u_s)\right)}\frac{\psi_s(u_s)}{\sum_{s\in S}\psi_s(u_s)}, & \text{for } s = s' \\ -\frac{W\left(\sum_{s\in S}\psi_s(u_s)\right)}{1+W\left(\sum_{s\in S}\psi_s(u_s)\right)}\frac{\psi_{s'}(u_{s'})}{\sum_{s'\in S}\psi_s(u_s)}, & \text{for } s \neq s'. \end{cases} \tag{2.76}$$

The value of the product of the fractions lies in $(0,1)$, since the value of W is non-negative and the value of ψ_s is positive. Hence, we have

$$\frac{\partial d_s^*}{\partial u_s} > 0 \quad \text{and} \quad \frac{\partial d_s^*}{\partial u_{s'}} < 0, \text{ for all } s' \in S \setminus \{s\}. \tag{2.77}$$

Noting that $u = \gamma_t(x)$ and that $\gamma_t(x)$ is monotonically non-decreasing in x (under Assumption 2.3 due to Proposition 2.6(i)) yields the desired property. □

2.A.4 Proof of Lemma 2.8

Fix any $x \in X$ and recall that $\hat{P} := \left\{ [p_s(d)]_{s \in F(x)} \;\middle|\; d_s \in [\underline{d}, \bar{d}] \quad \forall s \in F(x) \right\}$. To show that $\hat{P}$ is convex under Assumption 2.5, i.e. under the multinomial customer choice model in (2.1), we will show that p_s is related to a convexity-preserving function, the so-called *perspective function* [44, Section 2.3.3]. We do this by means of the following reformulation.

Denote $m := |F(x)|$ and define $\delta_s(d_s) := \exp(\beta_c + \beta_s + \beta_d d_s) \in \mathbb{R}_{++}$ for all $s \in F(x)$ and $\delta(d) := [\delta_s(d_s)]_{s \in F(x)} \in \mathbb{R}^m_{++}$. Denote the image of $\hat{D} := [\underline{d}, \bar{d}]^m$ obtained through this transformation by Δ, i.e. $\Delta := \left\{ \delta(d) \;\middle|\; d \in \hat{D} \right\} \subseteq \mathbb{R}^m_{++}$. Since $\hat{D}$ is a hypercube and since the mapping is monotonic and separable for all $s \in S$, Δ is a hypercuboid, which is a convex set. We can thus express the mapping between $p(d)$ and $\delta(d)$ as

$$p(d) = \frac{\delta(d)}{\mathbf{1}^\intercal \delta(d) + 1}, \text{ for all } d \in \hat{D}. \tag{2.78}$$

Furthermore, we express $\hat{P}$ equivalently in terms of Δ as

$$\hat{P} := \left\{ \frac{\delta}{\mathbf{1}^\intercal \delta + 1} \;\middle|\; \delta \in \Delta \right\}. \tag{2.79}$$

Now consider the pair of variables $(y, z) \in G \subseteq \mathbb{R}^m \times \mathbb{R}_{++}$ and the *perspective* function $f : \mathbb{R}^m \times \mathbb{R}_{++} \to \mathbb{R}^m$, such that $f(y, z) := y/z$. As shown in [44, Section 2.3.3], the image $f(G) := \{f(y, z) \mid (y, z) \in G\}$ is a convex set if G is a convex set. We can see that the mapping in (2.79) is a perspective function by setting $y = \delta$ and $G = \{(\delta, \mathbf{1}^\intercal \delta + 1) \mid \delta \in \Delta\}$, which is still convex, since Δ is a convex set and the affine mapping $z = \mathbf{1}^\intercal \delta + 1$ preserves convexity. Having shown that the mapping in (2.79) can be represented as a perspective function and thus preserves convexity, shows that $\hat{P}$ is a convex set and thus concludes our proof. $\square$

3

Algorithmic Developments

in Approximate Dynamic Programming for Revenue Management in Attended Home Delivery

Contents

3.1 Introduction **52**
3.2 Value function approximation algorithm **55**
3.2.1 Initialisation 56
3.2.2 “Forward sweep” 57
3.2.3 “Backward sweep” 58
3.3 Approximation algorithm properties **58**
3.4 Validation algorithm **61**
3.5 Validation algorithm properties **62**
3.5.1 Tail bounds 63
3.5.2 Expectation bounds 64
3.6 Numerical example **66**
3.6.1 Computation of the approximate value function 67
3.6.2 Computational complexity analysis 70
3.6.3 Computation of the bounds 72
3.7 Conclusion **76**
3.A Proofs **78**
3.A.1 Proof of Proposition 3.1 78
3.A.2 Proof of Proposition 3.4 79
3.A.3 Proof of Proposition 3.5(i) 80
3.A.4 Proof of Proposition 3.5(ii) 82
3.A.5 Proof of Lemma 3.6 83
3.A.6 Proof of Proposition 3.7(ii) 83
3.A.7 Proof of Proposition 3.8 84
3.B Synthetic problem data for Section 3.6.1 . . . . **85**

3.1 Introduction

As we have seen in the previous chapter, DP is an established tool to model multi-stage optimal control problems, such as the revenue management problem in attended home delivery. However, one typical limitation of DP is that it scales poorly with the dimensionality of the problem. In this chapter, we exploit the mathematical properties derived in the previous chapter to take a step towards alleviating this so-called "curse of dimensionality". Over time, different approximation approaches have been proposed ranging from general-purpose techniques to methods exploiting certain structural features of the DP problem at hand; see Fig. 3.1.

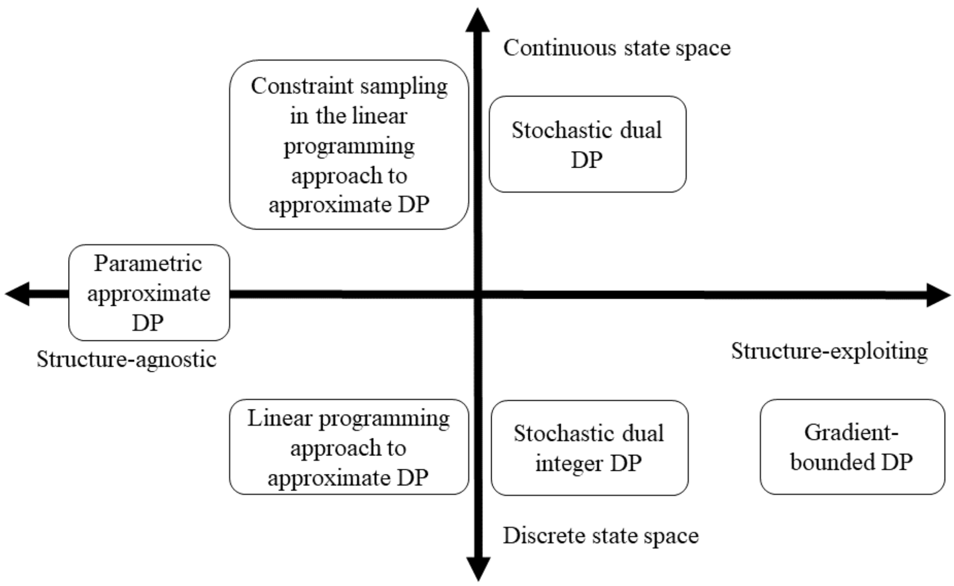


Figure 3.1: Schematic grouping of selected approximate DP literature based on how general-purpose or problem-specific the approaches are and if they are limited to a discrete or continuous state space.

One of the most general and widely used approaches is to model the value function of the DP as a linear combination of basis functions, which are sometimes also called features [49, Chapter 8.2]. Approximation of the value function then reduces to the problem of estimating the weight coefficients of the linear combination. While this approach is very flexible, it might not be obvious how to choose the number and algebraic form of basis functions in an informed way for a particular problem at hand.

A more structured way to solve such an approximate DP formulation is given by the linear programming approach to approximate DP [50], which was originally devised for finite state and action spaces, but is readily extended to continuous states and actions [51, 52]. The general idea is to change the Bellman equation of the DP by an epigraphic reformulation into a linear program, where the Bellman equation appears as a constraint for each state and action pair. Consequently, the number of constraints is often prohibitively large even for problems with finite state and action spaces. However, it may be possible to create a simpler problem by neglecting some of these constraints. If sufficiently many constraints are still considered it may be possible that the solution to this approximate problem provides a good enough solution to the original problem. Still, the quality of the approximation once again hinges on the need to choose suitable basis functions.

Certain structures of the value function can simplify the process of selecting basis functions. For example, if the value function of a DP over continuous states is convex[1], then stochastic dual DP provides an alternative to the above-mentioned approximate DP techniques [47, 48]. The main objective is to under-approximate the value function by the pointwise maximum of a finite number of hyperplanes in the state space for all time steps in the DP. These hyperplanes are added iteratively by first generating an approximately optimal sample path of states forward in time and then adding a hyperplane at each time step along this sample path backward in time, which refines the value function along this sample path.

Variants of this algorithm have been developed for problems with quadratic stage costs [53] as well as extensions to robust optimisation settings [54]. However, extensions to problems with discrete state spaces have attracted only limited interest to date [45, 46]. Furthermore, while exploiting convex problem structure and bypassing the need to choose basis functions, these approaches still suffer from some limitations, such as being constrained to linear systems [45] or having to solve

[1]The value function needs to be convex if costs are minimised or alternatively concave if rewards are maximised.

a non-convex optimisation problem to add hyperplanes to the representation of the approximate value function at each time step [46].

In this chapter, we extend the stochastic dual DP approach to value functions that are both submodular (see (2.15)) and concave extensible (see Definition 2.2) over a discrete domain. This class of problems encompasses the revenue management problem introduced in Chapter 2 as well as other revenue management problems in a variety of applications like transportation, hospitality and appointment scheduling [55–57].

For any DP whose value function possesses these properties, we present a new algorithm, which we term gradient-bounded DP. This algorithm computes deterministic upper bounds and stochastic lower bounds to the exact value function in the spirit of stochastic dual DP theory [47, 48]. Moreover, we also derive probabilistic guarantees on the value accumulated under the associated policy for a single realisation of the DP and for the expectation of this value. Furthermore, we provide a numerical example with problem parameters that stem from this application to show the effectiveness of our proposed algorithm. Our approach also extends the approximate DP approach of [20], whose value function is approximated by an affine function, however, without being accompanied by guarantees to generate a certain fraction of the optimal profit.

The remainder of this chapter is structured as follows: In Section 3.2, we present a novel algorithm to compute approximately optimal policies for value functions over discrete state spaces under assumptions on submodularity and concave extensibility. Section 3.3 derives deterministic upper bounds and stochastic lower bounds to the exact value function and shows convergence of the algorithm in a finite number of iterations. In Section 3.4, we present an algorithm that validates the policy obtained in Section 3.2 by computing sample profits, their empirical mean and standard deviation. Section 3.5 details our theoretical results on tail and expectation bounds of the sample profits obtained in Section 3.4. In Section 3.6, we present a numerical example on a high-dimensional problem that is unsolvable by direct computation. Finally, we conclude in Section 3.7 and provide directions for future research.

3.2 Value function approximation algorithm

We consider the stochastic shortest path DP formulation introduced in (2.6) and impose the following assumptions:

Assumption 3.1. *The function $-C$ is submodular and concave extensible in x.*

Assumption 3.2. *We assume that the functions D, $\hat{\rho}$ and $\pi_{x,y}$ for all $(x, y) \in X \times X$ as well as the time horizon T in* (2.6) *are chosen such that the Bellman operator $\mathcal{T}$ preserves concave extensibility and submodularity of any concave extensible and submodular value function, i.e. if V_{t+1} is submodular and concave extensible, then $V_t = \mathcal{T} V_{t+1}$ also has these properties for all $t \in T$.*

As shown in Chapter 2, the DP formulation of the revenue management problem in attended home delivery in (2.4), which is mathematically identical to (2.6), satisfies these assumptions. To be precise, Theorem 2.3 shows that the value function is concave extensible over its domain for all time steps and Proposition 2.6(ii) shows that the value function is strictly submodular over its domain for all time steps, since strictly increasing opportunity costs imply strict submodularity as shown in (2.16). Notice that Assumptions 3.1 and 3.2 do not require *strict* submodularity and thus, the class of problems targeted in this Chapter is larger than the problems considered in Chapter 2.

Based on these assumptions, we now state our proposed approximation procedure in Algorithm 3.1 below and subsequently detail the individual algorithm steps. Inspired by stochastic dual DP techniques [48], the main idea of our algorithm is to alternate between generating sample paths in "forward sweeps" and refining the value function in "backward sweeps". We term our approximation algorithm "gradient-bounded DP", since it exploits properties of the gradient of the approximate value function, namely submodularity and concave extensibility, to compute an upper bound to the exact value function of the DP. The following sections describe this procedure in detail.

Algorithm 3.1 Gradient-bounded dynamic programming

1: Initialise parameters: $X, D, \pi_{x,y}, T, \hat{\rho}, C$ and $i_{\max}$
2: Initialise $Q_t^0(x) \leftarrow \infty$, 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: $\quad x_1^i \leftarrow 0$
6: $\quad$**for** $t \in \{1, 2, \ldots, \bar{t}\}$ **do** ▷ "Forward sweep"
7: $\quad\quad d_t^i \leftarrow d^* \in \underset{d \in D}{\operatorname{argmax}} \left\{ \sum_{x_{t+1}^i \in Y_+(x_t^i)} \pi_{x_t^i, x_{t+1}^i}(d) \left(\hat{\rho}(x_t^i, x_{t+1}^i, d) + Q_{t+1}^{i-1}(x_{t+1}^i) \right) \right\}$
8: $\quad\quad x_{t+1}^i \leftarrow x_t^i + \underset{x_{t+1}^i}{\operatorname{sample}} \left\{ \pi_{x_t^i, x_{t+1}^i}(d_t^i) \right\}$
9: $\quad$**end for**
10: $\quad l(i) \leftarrow \sum_{t=1}^{\bar{t}} \hat{\rho}(x_t^i, x_{t+1}^i, d_t^i) - C(x_{\bar{t}+1})$
11: $\quad$**for** $t \in \{\bar{t}, \bar{t}-1, \ldots, 1\}$ **do** ▷ "Backward sweep"
12: $\quad\quad Z(x_{t+1}^i) \leftarrow \{x_{t+1}^i + 1_s + 1_{s'}\}_{s \in S \cup \{0\}, s' \in S \cup \{0\}}$
13: $\quad\quad$**if** Q_{t+1}^{i-1} is submodular on $Z(x_{t+1}^i)$ **then**
14: $\quad\quad\quad H^* \leftarrow$ unique hyperplane through $\left\{ \left(y, (\mathcal{T} Q_{t+1}^{i-1})(y) \right) \right\}_{y \in Y_+(x_{t+1}^i)}$
15: $\quad\quad$**else**
16: $\quad\quad\quad j^* \in \underset{j \in J_{t+1}^{i-1}(x_{t+1}^i)}{\operatorname{argmin}} \left\{ \left(\mathcal{T} H_{t+1}^{j-1} \right) \left(x_{t+1}^i \right) \right\}$
17: $\quad\quad\quad H^* \leftarrow \mathcal{T} H_{t+1}^{j^*-1}$
18: $\quad\quad$**end if**
19: $\quad\quad Q_t^i \leftarrow \min \left\{ H^*, Q_t^{i-1} \right\}$
20: $\quad$**end for**
21: $\quad u(i) \leftarrow Q_1^i(0)$
22: **end for**

3.2.1 Initialisation

We first initialise all parameters of the DP in (2.6) (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 Q_t^i for all $(i,t) \in I \times T$ be the pointwise minimum of a finite number of affine functions, i.e.

$$Q_t^i(x) := \min_{j \in \{0,1,\ldots,i\}} H_t^j(x), \text{ for all } x \in X, \tag{3.1}$$

where $H_t^j : X \mapsto \mathbb{R}$ describes a hyperplane, i.e.

$$H_t^j(x) := \langle a_t^j, x\rangle + b_t^j, \text{ for all } x \in X, \tag{3.2}$$

with $a_t^j \in \mathbb{R}^n, b_t^j \in \mathbb{R}$ for all $(t, j) \in T \times I$. We characterise the set of supporting hyperplanes at x as

$$J_t^i(x) := \underset{j\in\{0,1,\ldots,i\}}{\operatorname{argmin}} \left\{\langle a_t^j, x\rangle + b_t^j\right\} \tag{3.3}$$

for all $(x, i, t) \in X \times I \times T$. Notice that the above-mentioned functions are defined for all states $x \in X$. In practice, the number of states may be prohibitively large to compute these functions for all states, however, for our purposes, we will only ever evaluate these functions locally for some $x \in X$, which is possible since the maximum number of hyperplanes $i_{\max}$ will be relatively moderate.

We construct Q_t^i as a successively tighter upper bound of V_t (as i increases), i.e. $V_t(x) \le Q_t^i(x) \le Q_t^{i-1}(x)$ for all $(x, i, t) \in X \times (I \setminus \{0\}) \times T$. In the i-th "backward sweep", H_t^i is added to Q_t^{i-1} for all $t \in T$ to form Q_t^i. To initialise Q_t^0, one could simply set Q_t^0 to be a single affine function with $a_t^0 = 0$ and $b_t^0 = \infty$, such that Q_t^0 is indeed an upper bound to V_t for all $t \in T$ (step 2). We discuss the possibility of closer initialisations in the context of our example in Section 3.6. We also initialise $Q_{\bar{t}+1}^i(x) := V_{\bar{t}+1}(x) = -C(x)$ for all $(x, i) \in X \times I$, which is a tight upper bound by the construction of the DP in (2.6) (step 3).

3.2.2 "Forward sweep"

Fix any iteration $i \in I \setminus \{0\}$. In each "forward sweep", we solve an approximate version of the Bellman equation in (2.6) forward in time, i.e. by replacing V_t with its approximation Q_t^{i-1} (step 7). Hence, we compute 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 (step 8). This defines a *sample path* $\{x_t^i\}_{t\in T\cup\{\bar{t}+1\}}$. At the end of each "forward sweep", we compute a stochastic lower bound on the total expected profit $V_1(0)$, which we denote by $l(i)$ for all $i \in I \setminus \{0\}$ (step 10). The total expected profit is given by $V_1(0)$ because it is the expected profit-to-go at the first step of the time horizon and at the initial state $x_1^i = 0$ (see (2.3)). We show in Section 3.3 that $l(i)$ is indeed a stochastic lower bound to $V_1(0)$ for all $i \in I \setminus \{0\}$.

3.2.3 "Backward sweep"

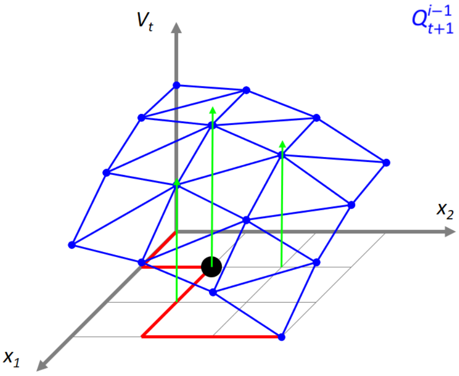
Fix any iteration $i \in I$. In each "backward sweep", we first check if Q_{t+1}^{i-1} is submodular on $Z(x_{t+1}^i)$ by computing the sign of (2.15) for all possible pairs of points $(y, y') \in Z(x_{t+1}^i) \times Z(x_{t+1}^i)$, such that $y \neq y'$ (step 12). If the inequality in (2.15) holds for all these points, we refine the value function approximation Q_{t+1}^{i-1} as shown in Fig. 3.2 and described further below.

First, we locally compute the exact DP stage problem on the set $Y_+(x_{t+1}^i)$, i.e. $\{\mathcal{T}Q_{t+1}^{i-1}(y)\}_{y \in Y_+(x_{t+1}^i)}$ (Fig. 3.2(a)), to construct the unique hyperplane through $\left\{\left(y, (\mathcal{T}Q_{t+1}^{i-1})(y)\right)\right\}_{y \in Y_+(x_{t+1}^i)}$ (step 14, Fig. 3.2(b)). Then, the resulting added hyperplane is an upper bound to $V_t(x)$ for all $x \in X$, as shown in Section 3.3.

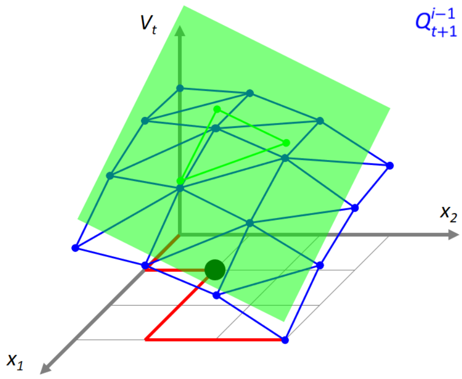
In the opposite case, we need to compute a submodular upper bound on Q_{t+1}^{i-1}, which is readily given by the hyperplanes from which Q_{t+1}^{i-1} is constructed. Therefore, we select the hyperplane $H_{t+1}^{j^*-1}$ that minimises the value at the evaluation point x_t^i, which therefore locally creates the tightest upper bound (step 16). It may be possible to construct other submodular upper bounds to non-submodular Q_{t+1}^{i-1}, however, steps 16 and 17 of Algorithm 3.1 offer a simple implementation. Finally, we update the value function approximation as the pointwise minimum of the approximation from the previous iteration (Fig. 3.2(c)) and the newly constructed hyperplane (step 19, Fig. 3.2(d)). We also compute an upper bound, $u(i)$ for all $i \in I \setminus \{0\}$, on the total expected profit $V_1(0)$ (step 21). We show that this is indeed an upper bound in Section 3.3.

3.3 Approximation algorithm properties

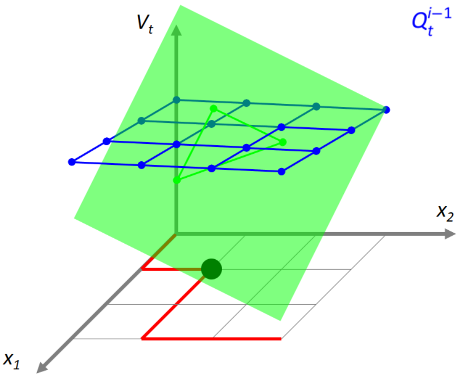
In this section, we show our main theoretical results on bounds on the exact value function and convergence properties of Algorithm 3.1. Proofs not included in this section can be found in Appendix 3.A.



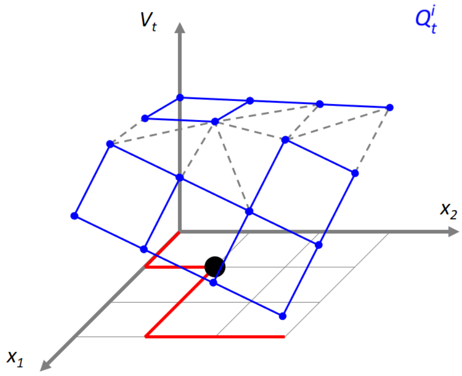
(a) Using the approximate value function Q_{t+1}^{i-1} (blue), we compute one step of the Bellman operator locally at the point x_{t+1}^{i} (black circle) on the sample path (red line) as well as on the other points in $Y_+(x_{t+1}^{i})$ (green arrows).



(b) The newly computed points (green triangle) define a unique hyperplane (green shaded plane).



(c) The constructed hyperplane (green shaded plane) is shown together with the approximate value function at the previous time step Q_t^{i-1} (blue), which is a hyperplane for illustrative purposes here, but in general, it is the pointwise minimum of a finite number of hyperplanes.



(d) Finally, the refined value function approximation Q_t^i (blue) is found by computing the pointwise minimum between the constructed hyperplane and the approximate value function at the previous time step Q_t^{i-1}.

Figure 3.2: Illustration how the approximate value function is refined in the "backward sweeps" of gradient-bounded DP.

Proposition 3.1. *Under Assumptions 3.1 and 3.2, the approximate value function is an upper bound to the exact finite horizon value function, i.e.* $Q_t^i(x) \geq V_t(x)$ *for all* $(x, i, t) \in X \times I \times T$.

Corollary 3.2. *Under Assumptions 3.1 and 3.2, the value of* $u(i)$ *is an upper bound to the exact total expected profit, i.e.* $u(i) \geq V_1(0)$ *for all* $i \in I \setminus \{0\}$.

Proof. This result follows immediately from Proposition 3.1 and by observing that $u(i) = Q_1^i(0)$ for all $i \in I \setminus \{0\}$ from step 21 of Algorithm 3.1. $\square$

Proposition 3.3. *The expected value of $l(i)$ is a stochastic lower bound to the expected total profit, i.e. $\mathbb{E}[l(i)] \leq V_1(0)$ for all $i \in I \setminus \{0\}$.*

Proof. For any $i \in I \setminus \{0\}$, the value of $l(i)$ is obtained from suboptimal decisions d_t^i for all $t \in T$, due to the use of Q_{t+1}^{i-1} instead of the exact (yet unavailable) V_{t+1} in step 7 of Algorithm 3.1. It follows that d_t^i is not a maximiser of the exact DP in (2.6) which, by the principle of optimality, implies that the expected value accumulated under this suboptimal policy will not be greater than the value obtained under the optimal policy. Hence, $\mathbb{E}[l(i)] \leq V_1(0)$ for all $i \in I \setminus \{0\}$. $\square$

The stochastic dual DP algorithm converges asymptotically in i to the exact value function [48]. We can strengthen this result for our algorithm by exploiting the fact that the set of states X is finite (but potentially very large). Hence, the proposed algorithm converges in a finite number of steps under the following minor modification to Algorithm 3.1.

Algorithm 3.2 Resampling procedure replacing step 8 of Algorithm 3.1

$m \leftarrow \lfloor (i-1)/|X| \rfloor$
$x_{t+1}^i \leftarrow x_t^i + \underset{x_{t+1}^i}{\text{sample}} \left\{ \pi_{x_t^i, x_{t+1}^i}(d_t^i) \right\}$
if $t = \bar{t} - m$ and $x_{t+1}^i \in \left\{ x_{t+1}^j \;\middle|\; m|X| + 1 \leq j < i \right\}$ **then**
  $x_{t+1}^i \leftarrow$ sample with uniform probability from $X \setminus \left\{ x_{t+1}^j \;\middle|\; m|X| + 1 \leq j < i \right\}$
end if

Assumption 3.3. *Step 8 of Algorithm 3.1 is replaced by Algorithm 3.2.*

Proposition 3.4. *Under Assumptions 3.1–3.3, the gap $u(i) - \mathbb{E}[l(i)]$ for all $i \in I \setminus \{0\}$ converges to 0 in at most $\bar{t}|X|$ iterations of Algorithm 3.1.*

Note that the number of states, $|X| = (\mathbf{1} + \bar{x})^n$, depends exponentially on the dimensionality of the problem, n. Hence, it is likely to take an unacceptably large number of iterations for the algorithm to converge to the exact value function due to the enormous number of states $|X|$. Since the value function is computationally expensive to calculate for all states, we seek to generate closer approximations at points that are likely to be visited, i.e. points on the sample path, and to use this information to save on approximation accuracy for less likely samples. Furthermore, note that the resampling procedure in Algorithm 3.2 ensures that every $|X|$ *iterations* of Algorithm 3.1, all states $x \in X$ are visited at a particular time step $\bar{t} - m$ exactly once. The choice of m in step 1 of Algorithm 3.2 ensures that all states $x \in X$ are visited in an order that allows the exact computation of the DP backwards in time, i.e. starting at the end of the time horizon and progressing backwards to $t = 1$.

Our ultimate objective is to solve problems with large state spaces ($|X| \approx 10^{20}$) and long time horizons ($|T| \approx 10^4$). In such scenarios, the need to resample the state as detailed in Assumption 3.3 becomes negligible, because the required number of iterations to reach convergence is much larger than the maximum acceptable number of iterations. Therefore, from a practical point of view, we do not resample to satisfy Assumption 3.3. In this case, the proposed algorithm only converges asymptotically to the exact value function instead of in a finite number of steps, just as in stochastic dual DP [48].

3.4 Validation algorithm

As noted in the previous section, absolute convergence of the approximate value function to the exact value function cannot be achieved for industry-sized problems due to the "curse of dimensionality". The performance of the algorithm, i.e. how close the stochastic lower bound $l(i)$ is to the deterministic upper bound $u(i)$ for any $i \in I \setminus \{0\}$, can only be validated statistically to a certain level of probabilistic confidence. To this end, we will generate a set of validation samples as described in Algorithm 3.3 and detailed further below.

We first compute the approximation obtained in Algorithm 3.1 (step 1). We denote the maximum number of validation samples by $k_{\max} \in \mathbb{N}$ and let $K := \{1, 2, \ldots, k_{\max}\}$ (step 2). We then compute $k_{\max}$ "forward validation sweeps", where in each of them we use our most refined estimate, $Q_{t+1}^{i_{\max}}$ as our approximate value

Algorithm 3.3 Validation algorithm

1: Compute approximation: $Q_t^{i_{\max}}$, for all $t \in T \setminus \{0\} \cup \{\bar{t}+1\}$
2: Initialise number of validation samples $k_{\max}$
3: **for** $k \in K := \{1, 2, \dots, k_{\max}\}$ **do**
4: $\quad x_1^k \leftarrow 0$
5: $\quad$ **for** $t \in \{1, 2, \dots, \bar{t}\}$ **do** ▷ "Forward validation sweep"
6: $\quad\quad d_t^k \leftarrow d^* \in \operatorname*{argmax}_{d \in D} \left\{ \sum_{x_{t+1}^k \in Y_+(x_t^k)} \pi_{x_t^k, x_{t+1}^k}(d) \left(\hat{\rho}(x_t^k, x_{t+1}^k, d) + Q_{t+1}^{i_{\max}}(x_{t+1}^k) \right) \right\}$
7: $\quad\quad x_{t+1}^k \leftarrow x_t^k + \operatorname*{sample}_{x_{t+1}^k} \left\{ \pi_{x_t^k, x_{t+1}^k} \left(d_t^k \right) \right\}$
8: $\quad$ **end for**
9: $\quad l_{\mathrm{v}}(k) \leftarrow \sum_{t=1}^{T} \hat{\rho}(x_t^k, x_{t+1}^k, d_t^k) - C(x_{\bar{t}+1})$
10: **end for**
11: $\bar{l}_{\mathrm{v}} \leftarrow k_{\max}^{-1} \sum_{k=1}^{k_{\max}} l_{\mathrm{v}}(k)$
12: $\sigma_{\mathrm{v}} \leftarrow \sqrt{(k_{\max} - 1)^{-1} \sum_{k=1}^{k_{\max}} \left(l_{\mathrm{v}}(k) - \bar{l}_{\mathrm{v}} \right)^2}$

function (steps 5–8). After each sweep $k \in K$, we compute the stochastic lower bound $l_{\mathrm{v}}(k)$ on the total expected profit, similarly to $l(i)$ for all $i \in I \setminus \{0\}$ in Algorithm 3.1 (step 9). We then compute the sample mean profit $\bar{l}_{\mathrm{v}}$ and unbiased empirical standard deviation σ_{v} of the set of sampled lower bounds $\{l_{\mathrm{v}}(k)\}_{k \in K}$ (steps 11–12). As detailed in the next section, these quantities will be used to generate one-sided confidence intervals, quantifying the performance of the decision policy associated with the approximate value function $Q_{t+1}^{i_{\max}}$.

3.5 Validation algorithm properties

In this section we state the main theoretical properties of our validation procedure. The proofs can be found in Appendix 3.A. We use $\{l_{\mathrm{v}}(k)\}_{k \in K}$, $\bar{l}_{\mathrm{v}}$ and σ_{v} from Algorithm 3.3 to derive two different measures for the performance guarantee. The first is a probabilistic bound on the tail of the distribution of a single lower bound sample, i.e. a value for $l(k_{\max} + 1)$ that is reached or exceeded with $1 - \alpha$ confidence for a user-defined $\alpha \in (0, 1)$. As we will see later in Section 3.6, this bound is not necessarily indicative of the expectation of $\bar{l}_{\mathrm{v}}$, since even under the profit-maximising decision policy, some variance will persist in $l(k_{\max}+1)$ from the randomness of the state transitions. Therefore, we also derive a bound on the expectation of the empirical sample mean $\bar{l}_{\mathrm{v}}$ that holds with confidence $1-\alpha^{\mathbb{E}}$, where $\alpha^{\mathbb{E}} \in (0,1)$ can be chosen by the user.

3.5.1 Tail bounds

In this section, we present two tail bounds of the distribution of $l_{\mathrm{v}}(k_{\max}+1)$. Let $[l_-, l_+]$ denote the (finite) support of the distribution of $l_{\mathrm{v}}(k)$ for any $k \in K \cup \{k_{\max}+1\}$ and let F_K denote the empirical cumulative distribution function of $\{l_{\mathrm{v}}(k)\}_{k\in K}$, i.e. $F_K(l) := k_{\max}^{-1} \sum_{k\in K} \mathbb{1}(l_{\mathrm{v}}(k) \leq l)$. We derive two tail bounds of the distribution of $l_{\mathrm{v}}(k_{\max}+1)$ with a given confidence level $(1-\alpha) \in (0,1)$, which is mildly restricted for the first bound due to the next assumption.

Assumption 3.4. *Assume that* $\alpha > \theta_{\mathrm{C}} := \Pr(\sigma_{\mathrm{v}} = 0)$.

The value of θ_{C} will often be negligibly small, since $l_{\mathrm{v}}(k)$ for all $k \in K$ is highly unlikely to take identical values due to the typically high-dimensional state space and long time horizon. We show this later in Section 3.6.3.

Proposition 3.5. *The inequality* $\Pr(l_{\mathrm{v}}(k_{\max}+1) > l^*) \geq 1-\alpha$ *holds*
(i) *under Assumption 3.4, if* $\alpha \in (\theta_{\mathrm{C}}, 1)$ *and* $l^* = l_{\mathrm{C}}$*, the empirical Cantelli bound given by*

$$l_{\mathrm{C}} := \bar{l}_{\mathrm{v}} - \sigma_{\mathrm{v}} \sqrt{\frac{(1-\alpha)(k_{\max}-1)}{(\alpha-\theta_{\mathrm{C}})k_{\max}}}, \quad \textit{or} \tag{3.4}$$

(ii) *if* $\alpha \in (0,1)$ *and* $l^* = l_{\mathrm{D}}$*, the Dvoretzky-Kiefer-Wolfowitz bound given by*

$$l_{\mathrm{D}} := \sup \left\{ l \in [l_-, l_+] \;\middle|\; F_K(l) \leq \alpha - \theta_{\mathrm{D}} - \sqrt{\frac{\ln(1/\theta_{\mathrm{D}})}{2k_{\max}}} \right\}, \tag{3.5}$$

where $\theta_{\mathrm{D}} \in (0,\alpha)$ *is a user-defined parameter.*

For l_{D}, we find the θ_{D}, which maximises the value of the bound, from the so-called Lambert W function.

Definition 3.1. *Let the Lambert W function be implicitly defined as* $W_i : \mathbb{R} \to \mathbb{R}$*, such that* $W_i(x)\exp(W_i(x)) = x$ *for* $i \in \{0, -1\}$*, where* $W_0(x) > -1$ *is called the principal branch and* $W_{-1}(x) \leq -1$ *is called the lower branch.*

Lemma 3.6. *For any* $\alpha \in (0, 1)$*, the value of* l_{D} *is maximised at*

$$\theta_{\mathrm{D}} = \min\left\{\alpha, \sqrt{\exp\left(W_{-1}\left(\frac{-1}{4k_{\max}}\right)\right)}\right\}. \tag{3.6}$$

The bounds l_{C} and l_{D}, are termed after Cantelli's inequality [58] and the Dvoretzky-Kiefer-Wolfowitz [59] inequality, respectively. These inequalities are critical for showing that the bounds are indeed reached or exceeded with probability $1 - \alpha$. By Proposition 3.5, we can always choose the tighter, i.e. greater, of the two bounds and we will see later in Section 3.6 that the selection of α and $k_{\max}$ influences which bound is preferred.

3.5.2 Expectation bounds

Similarly to the tail bounds, we now state our theoretical results on two bounds on the expectation of $\bar{l}_{\mathrm{v}}$, denoted by $\mathbb{E}\bar{l}_v$.

Proposition 3.7. *Fix any significance level* $\alpha^{\mathbb{E}} \in (0, 1)$*. Then* $\Pr(\mathbb{E}\bar{l}_{\mathrm{v}} > l^*) \geq 1 - \alpha^{\mathbb{E}}$*, for all* $l^* \in \{l_{\mathrm{B}}^{\mathbb{E}}, l_{\mathrm{D}}^{\mathbb{E}}\}$*, where:*

(i) $l_{\mathrm{B}}^{\mathbb{E}}$ *is the empirical Bernstein bound given by*

$$l_{\mathrm{B}}^{\mathbb{E}} := \bar{l}_{\mathrm{v}} - \sqrt{\frac{2\sigma_{\mathrm{v}}^2 \ln(2/\alpha^{\mathbb{E}})}{k_{\max}}} - \frac{7(l_+ - l_-)\ln(2/\alpha^{\mathbb{E}})}{3(k_{\max} - 1)} \quad \textit{and} \tag{3.7}$$

(ii) $l_{\mathrm{D}}^{\mathbb{E}}$ *is the expectation Dvoretzky-Kiefer-Wolfowitz bound given by*

$$\begin{aligned} l_{\mathrm{D}}^{\mathbb{E}} := & \int_{l=\max\{0,l_-\}}^{\max\{0,l_+\}} 1 - \min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\} \mathrm{d}l + \max\{0, l_-\} \\ & - \int_{l=\min\{0,l_-\}}^{\min\{0,l_+\}} \min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\} \mathrm{d}l - \min\{0, l_+\}. \end{aligned} \tag{3.8}$$

The bounds $l_{\mathrm{B}}^{\mathbb{E}}$ and $l_{\mathrm{D}}^{\mathbb{E}}$ are termed after the empirical Bernstein [60] and Dvoretzky-Kiefer-Wolfowitz [59] inequalities, respectively. The proof of Proposition 3.7(i) is given in [60]. Notice that to compute $l_{\mathrm{D}}^{\mathbb{E}}$ it suffices to solve one of the two integrals if l_- and l_+ have the same sign, e.g. if both are non-negative, then the second integral evaluates to 0. Furthermore, it can be shown that $l_{\mathrm{D}}^{\mathbb{E}}$ is at least as tight as Hoeffding's concentration bound, given by

$$l_{\mathrm{H}}^{\mathbb{E}} := \bar{l}_{\mathrm{v}} - (l_+ - l_-)\sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\mathrm{max}}}}. \tag{3.9}$$

In fact, under an additional technical assumption, we show that the expectation Dvoretzky-Kiefer-Wolfowitz bound is strictly better than Hoeffding's bound.

Assumption 3.5. *We assume that α and k_{max} are chosen to satisfy*

$$\sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\mathrm{max}}}} > k_{\mathrm{max}}^{-1}. \tag{3.10}$$

Assumption 3.5 is very mild, since even for only a single observation $k_{\mathrm{max}} = 1$, the critical value of $\alpha^{\mathbb{E}}$ would be $e^{-2} \approx 13.5\%$, which is much larger than typical significance levels, e.g. 5% or 1%. Taking any smaller value of $\alpha^{\mathbb{E}}$ than the critical value will ensure that Assumption 3.5 is always satisfied. Furthermore, for $k_{\mathrm{max}} > 1$, the constraint on $\alpha^{\mathbb{E}}$ is even less restrictive.

Proposition 3.8. *Under Assumption 3.5, the expectation Dvoretzky-Kiefer-Wolfowitz bound is strictly tighter than Hoeffding's concentration bound, i.e. $l_{\mathrm{D}}^{\mathbb{E}} > l_{\mathrm{H}}^{\mathbb{E}}$ for all $\alpha^{\mathbb{E}} \in (0, 1)$.*

Finally, we note that other bounds have also been proposed in the literature, e.g. [48] assumes that the distribution of $\bar{l}_{\mathrm{v}}$ is Gaussian and determines confidence intervals based on the corresponding standard score, i.e. a Gaussian lower bound on the expectation of $\bar{l}_{\mathrm{v}}$ would be

$$l_{\mathrm{G}}^{\mathbb{E}} := \bar{l}_{\mathrm{v}} - z(\alpha^{\mathbb{E}})\frac{\sigma_{\mathrm{v}}}{\sqrt{k_{\mathrm{max}}}}, \tag{3.11}$$

where $z(\alpha^{\mathbb{E}})$ is the standard score of the Gaussian distribution (in fact, Student's t-distribution, especially for small sample sizes k_{max}, since the true variance of the underlying distribution of $\bar{l}_{\mathrm{v}}$ is approximated by σ_{v}^2). We compare this with our proposed bounds in Section 3.6.

3.6 Numerical example

We demonstrate our algorithm on an example of the so-called revenue management problem in attended home delivery introduced in Section 2.2. The DP in our numerical example takes the same parameters as the non-linear stochastic dual DP example in Section 2.5.3. We re-state these parameters in Table 3.1 for convenience. Recall that these parameters are our adaptation of the real-world, multi sub-area case study by [20] to a single delivery sub-area scenario.

Table 3.1: Re-statement of numerical example parameters from Section 2.5.3.

$(\lambda, \bar{t}, n)$	$(0.8, 65, 17)$
$\left(\underline{d}, \bar{d}, r\right)$	$(0, 10, 34.53)$
$\bar{x}$	$[12, \ldots, 12]^\intercal$
(β_c, β_d)	$(-2.5087, -0.0766)$
$\{\beta_s\}_{s\in S}$	$\{-1.0305, -0.3591, 0.3107, 0.5922, 0.6154, 0.0796, 0.5356, -0.2415,$ $-0.6286, -1.6736, -0.4351, -0.161, 0, 0.2533, 0.0736, 0.562, 0.2346\}$
$C_+(x)$	$0.1736 \times \mathbf{1}^\intercal x$

We choose $C(0) = 0$, i.e. we ignore fixed costs, which have no effect on the pricing policy.

Recall from Theorem 2.3 that the Bellman operator preserves strict submodularity for this type of DP, i.e. the condition in (2.15) holds with strict inequality, if a small enough $\lambda > 0$ is chosen. However, in this problem the terminal condition C is submodular in x, but not strictly so. In fact, it is an affine function of x and hence modular. We assume that $\lambda = 0.8$ from Table 3.1 is small enough to satisfy Assumption 3.2 in this problem. We note however, that in the absence of strict submodularity, we compromise on the theoretical guarantee that $u(i)$ (Corollary 3.2) is an upper bound to the exact value function. Furthermore, we will see that larger values for λ do not negatively affect the profit-generation performance of the algorithm in the numerical analysis in Chapter 4.

To speed up computation, we initialise Q_t^0 for all $t \in T$ using the fixed point of DP, V^*, which is a known upper bound to the exact value function at any $(x,t) \in X \times T$, i.e. $V^*(x) \geq V_t(x)$. This is always the case when $\mathcal{T}$ in (2.8) is a monotone operator [37, Chapter 3]. Recall from Theorem 2.1 that the fixed point is given analytically as

$$V^*(x) := (\bar{d} + r)\langle \mathbf{1}, \bar{x} - x\rangle - C(\bar{x}), \text{ for all } x \in X. \tag{3.12}$$

Hence, we use this result to set $Q_t^0(x) = V^*(x)$ instead of ∞ for all $(x,t) \in X \times T$. Note that the fixed point in (4.1) is an affine function, so the initialiser has low complexity, i.e. only one affine function describes Q_t^0.

3.6.1 Computation of the approximate value function

We implement Algorithm 3.1 in Julia [61] and run $i_{\max} = 100$ iterations on an i7-8565U CPU at 1.80 GHz processor base frequency and with 16GB RAM. The run time on our machine is 9 seconds. In each iteration $i \in \{1, 2, \ldots i_{\max}\}$, we compute the upper bound on the expected profit $u(i)$ (Corollary 3.2) and the stochastic lower bound $l(i)$ (Proposition 3.3), corresponding to the sample profit obtained in a single "forward sweep" of Algorithm 3.1, as discussed in Section 3.2. We also compute the cumulative moving average of the sample profits, i.e. $i^{-1} \sum_{j=1}^{i} l(j)$, which tends to the expected value of the stochastic lower bound $l(i)$ (Proposition 3.3) as i increases. Fig. 3.3 shows how these bounds develop over 100 iterations. We make the following observations:

(1) The upper bound converges within 10 iterations.

(2) The cumulative moving average of the sample profits converges after about 100 iterations. We conjecture that this is due to the fact that Algorithm 3.1 refines the value function approximation iteratively for all time steps. However, not all refinements propagate through all time steps of the DP. Therefore, each iteration step has a direct influence on the stochastic lower bound, which depends on the value function approximation at all time steps, while the upper bound might be unchanged as it only depends on the approximation at the first time step.

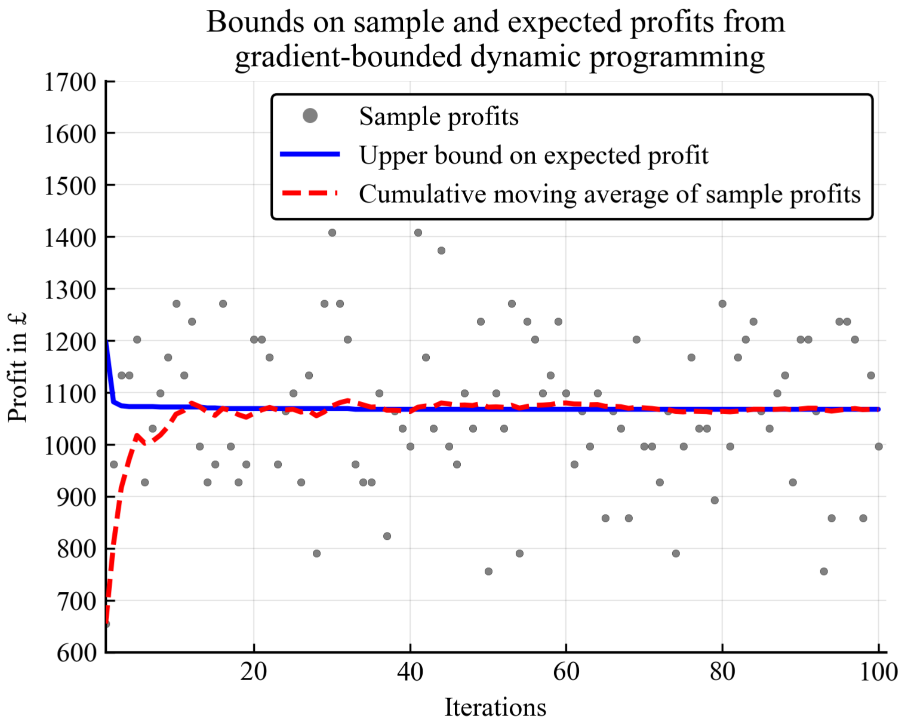


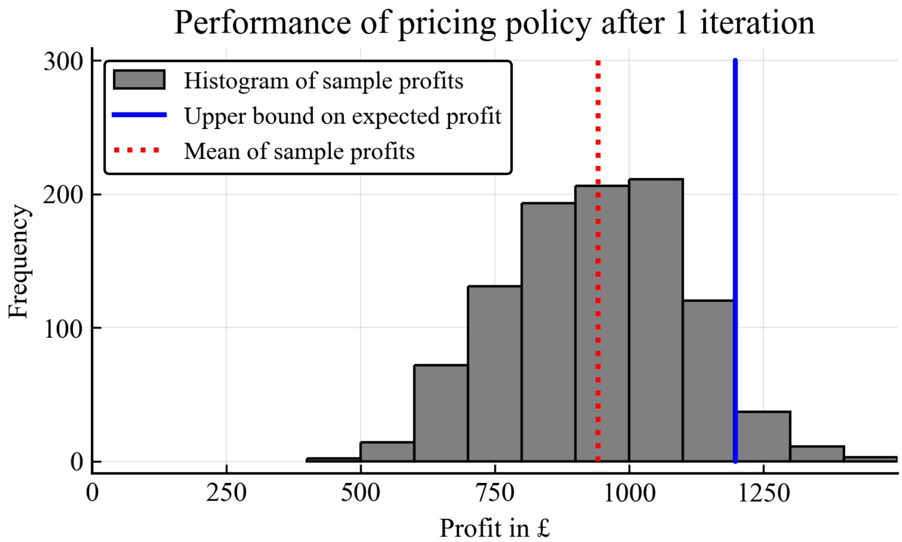
Figure 3.3: Plots of sample profits (grey dots), upper bound $u(i)$ (blue solid line, Corollary 3.2) and stochastic lower bound $l(i)$ (red dashed line, Proposition 3.3) on expected profit.

(3) The sample profits have high variance at all iterations since the customer choice process is random.

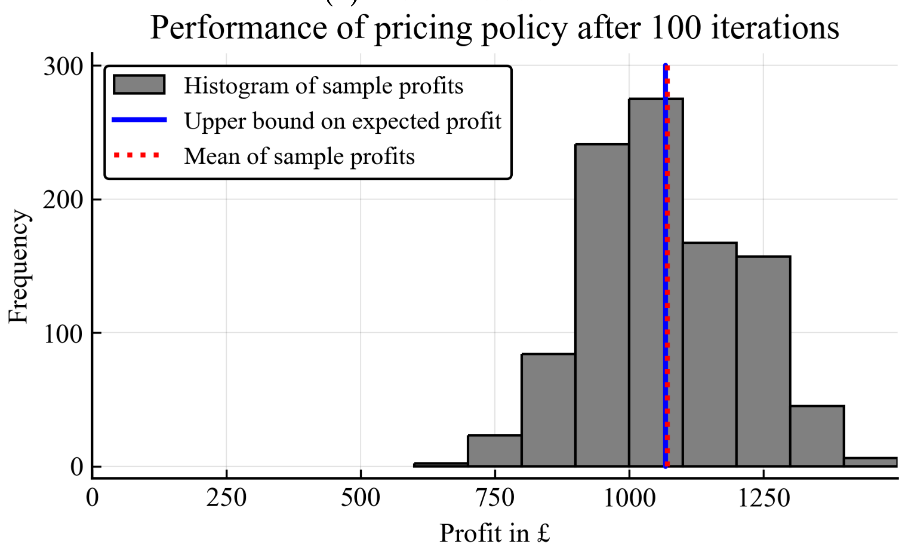
We investigate the influence of additional iterations, by comparing the performance of the pricing policies obtained after 1 and after 100 iterations of Algorithm 3.1. In particular, for both iteration counts, we simulate 1,000 booking periods and compute the sample profits obtained in each period by using Algorithm 3.3. Notice that we need to generate new samples from the transition probability distribution for these simulations and that these samples are independent from the samples that were used to compute the approximate value function in Algorithm 3.1. The resulting histogram of sample profits is shown in Fig. 3.4 below, where we observe the following:

(1) The mean sample profit increases from £942 after 1 iteration by 13.1% to £1,064 after 100 iterations.

(2) The gap between upper bound on expected profit, $u(i)$, and the empirical mean of 1,000 samples of $l(i)$, decreases from £256 to approximately £0.



(a) After 1 iteration.



(b) After 100 iterations.

Figure 3.4: Histograms of 1,000 sample profits after 1 (a) and 100 (b) iterations of Algorithm 3.1.

(3) The histograms show that the sample profits are more concentrated around their empirical mean after 100 iterations than after 1 iteration. This indicates that the variance of the sample profits can be decreased by increasing the number of algorithm iterations in this particular example, however, this is not a general conclusion.

We now validate our results by obtaining probabilistic performance guarantees for the approximate policy obtained after 100 iterations of Algorithm 3.1. In particular, we compute confidence intervals on the stochastic lower bound using validation samples $\{l_{\mathrm{v}}(k)\}$ for all $k \in K$ and the two pairs of bounds defined in Algorithm 3.3.

3.6.2 Computational complexity analysis

We now investigate numerically how the computation time of Algorithm 3.1 scales with the number of iterations performed and the dimensionality of the problem. Starting with the example from the previous section and the parameters from Table 3.1, we first examine how computation time varies with the number of iterations performed, as shown in Fig. 3.5. Despite the randomness in the algorithm, the computation time grows smoothly with iteration indices. The growth is approximately quadratic in i, which is to be expected: The cost of evaluating the approximate value function at any state-time pair $(x,t) \in X \times T$ and in any iteration $i \in I$ is linear in i, since one needs to find the minimum of the i hyperplanes that the approximate value function is composed of. The cumulative effect of this linearly growing evaluation cost in i is an overall approximately quadratic cost in i. We next examine how the computation time varies first with n. Since we

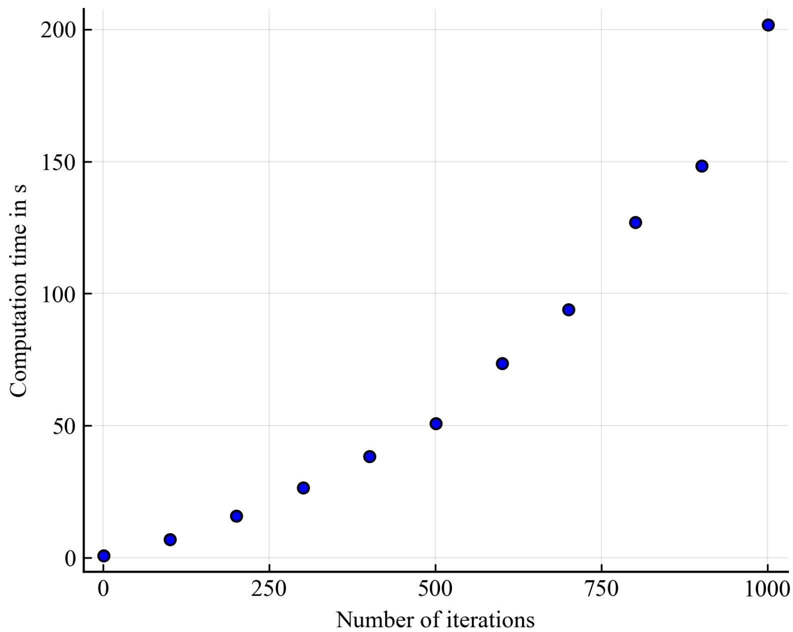


Figure 3.5: Computation time of Algorithm 3.1 in seconds against number of iterations of the 17-slot example from Section 3.6.1

want to simulate more slots than in the previous section, we need to artificially extend the customer choice model. In particular, we need to define synthetic values for β_s for all $s > 17$. We do this by taking samples from a uniform distribution with support $[-1.6736, 0.6154]$, which corresponds to the interval between the minimum and maximum value of β_s from Table 3.1. The sampled values can be found up to $n = 201$ in Table 3.2 in Appendix 3.B. We always compute 100 iterations of Algorithm 3.1 first for $n = 1$ and then we increase the number of slots by 10 each time, until we have computed the problem for $n = 201$. All other parameters, including admissible prices, maximum delivery capacity per slot and variable delivery cost are assumed to be identical for all n and correspond to the values found in Table 3.1. The resulting (random) computation times for all n are shown in Fig. 3.6. We observe that the computation time increases with the dimensionality of the problem, however the curve is less smooth than in the previous example, indicating a higher variance in the computation time estimates. We conjecture that different values of β_s for all $s \in S$ as n increases have different effects on the actual computation time in the stage optimisation problems that need to be solved in the forward and backward sweeps of Algorithm 3.1.

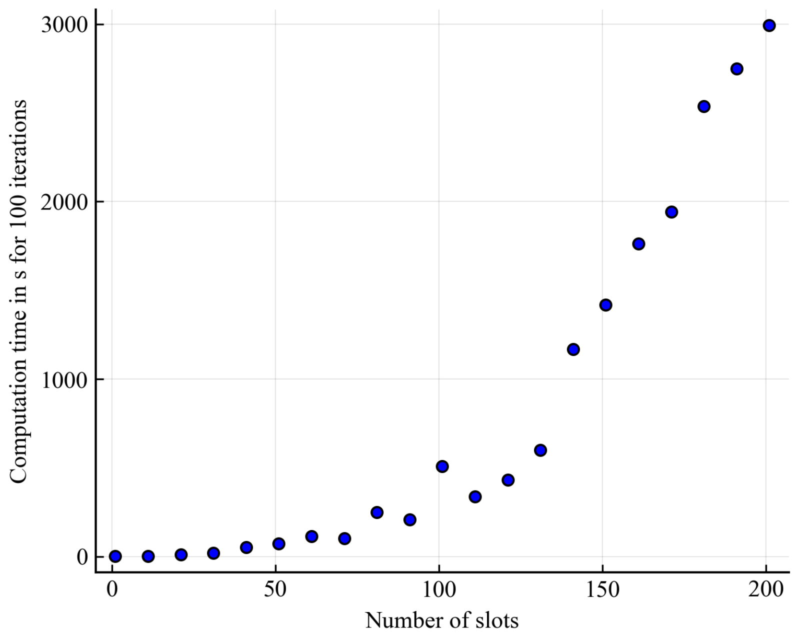


Figure 3.6: Computation time for 100 iterations of Algorithm 3.1 in seconds against varying problem dimensionality.

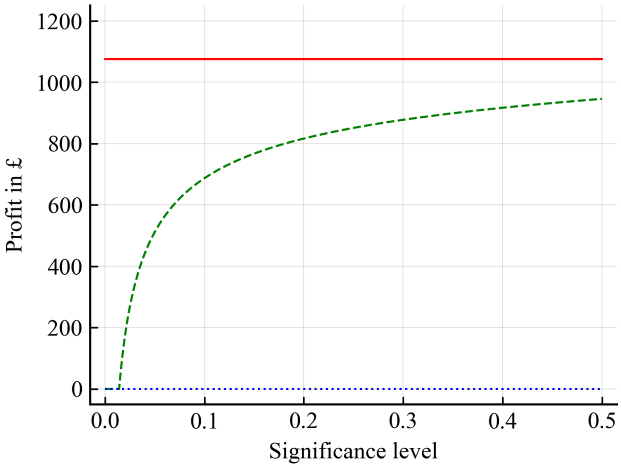
3.6.3 Computation of the bounds

We first compute the tail bounds on the value of profit obtained by a single sample under the approximate policy after 100 iterations of Algorithm 3.1. To this end, we compute the empirical Cantelli bound l_{C} from (3.4) and the Dvoretzky-Kiefer-Wolfowitz tail bound l_{D} from (3.5). Due to the 17-dimensional state space and time horizon of 65 steps, $\theta_{\mathrm{C}} \approx 0$. To see this, upper bound the probability of the most likely event at every stage, namely no order being placed by $\pi_{x,x}(d) \leq \pi_{x,x}(\mathbf{1}\bar{d}) \approx 0.6732$. Due to time-independence of the transition probabilities, we can exponentiate this number by the number of time steps in the DP to obtain the probability of 0 orders at the end of the booking period. This needs to happen for all $k_{\max}$ (independent) validation samples, hence we again exponentiate this number by $k_{\max}$, which we assume is at least 10. This gives us the probability of all validation samples having 0 orders. This is the most likely, but only one of $|X|$ states, so we multiply this number by $|X|$ to obtain $\Pr(\sigma_{\mathrm{v}} = 0) \leq |X| \pi_{x,x}(\mathbf{1}\bar{d})^{\bar{t}k_{\max}} \approx 1.9766 \times 10^{-112} \approx 0$. Notice that for larger values of $k_{\max}$, this approximation is even tighter.

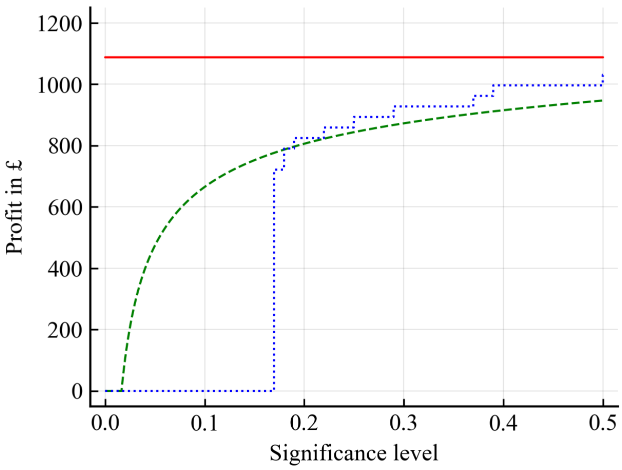
As we see in Fig. 3.7 at the end of this section, the tail bounds do not converge to the sample average, since there is an inherent variance in the customer choice model. This can be seen by inspecting the high variation of the sample profits in Fig. 3.3 for all iteration steps. In Fig. 3.7, notice that the choice of optimal bound depends on the sample sizes: For example, for $k_{\max} = 10$ iterations and with probability 90% the empirical Cantelli bound yields approximately £700, whereas the Dvoretzky-Kiefer-Wolfowitz tail bound only produces a trivial bound of £0. In contrast, for $k_{\max} = 1,000$ iterations and with probability 90% the Dvoretzky-Kiefer-Wolfowitz tail bound generates approximately £800 with optimal parameter $\theta_{\mathrm{D}} \approx \min\{\alpha, 4.8 \times 10^{-3}\}$ from Lemma 3.6, while the empirical Cantelli bound only certifies around £650.

To get more meaningful measures of the convergence of Algorithm 3.1, we now compute the bounds on the expectation of the profit obtained after 100 iterations of Algorithm 3.1. To this end, we compute the empirical Bernstein bound $l_{\mathrm{B}}^{\mathbb{E}}$ from (3.7) and the expectation Dvoretzky-Kiefer-Wolfowitz bound $l_{\mathrm{D}}^{\mathbb{E}}$ from (3.8). We also compute the Gaussian bound $l_{\mathrm{G}}^{\mathbb{E}}$ from (3.11), following [48]. This final bound relies on the additional assumption that the exact distribution of the mean sample profit $\bar{l}_{\mathrm{v}}$ is Gaussian. This is only asymptotically true by a Central Limit Theorem (see [62, Proposition 2.16]). As seen in Fig. 3.8 at the end of this section, the Gaussian bound always produces the largest expected profit for any given confidence level $\alpha^{\mathbb{E}}$, however it is not accompanied by theoretical guarantees. Moreover, for large validation samples sizes $k_{\max}$, the gap with the empirical Bernstein bound and with the expectation Dvoretzky-Kiefer-Wolfowitz bound is small.

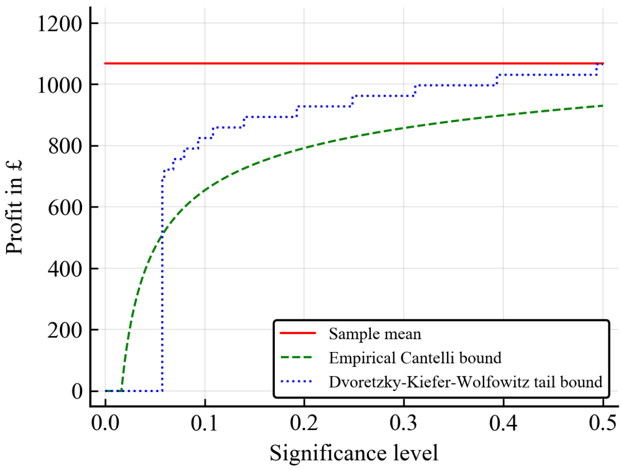
Therefore, we suggest to choose the two bounds presented in this chapter, which guarantee slightly smaller expected profits for any given $\alpha^{\mathbb{E}}$, but do not rely on the additional assumption of the Gaussian bound. Out of these two bounds, the expectation Dvoretzky-Kiefer-Wolfowitz bound tends to perform better than the empirical Bernstein bound in our numerical experiments. Note that we omit the empirical Bernstein bound in Fig. 3.8(a) since its negative values are not meaningful.



(a) 10 samples.

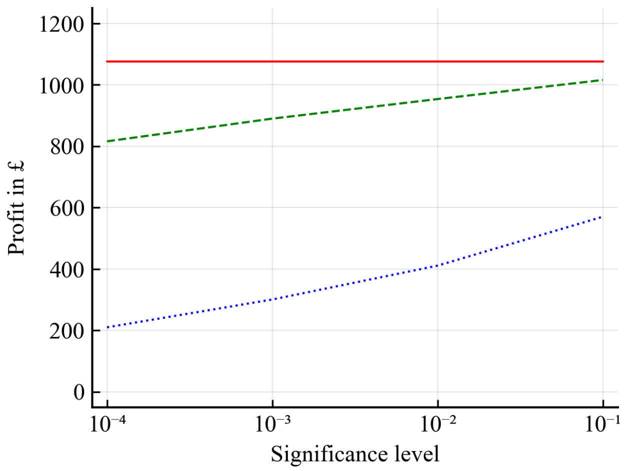


(b) 100 samples.

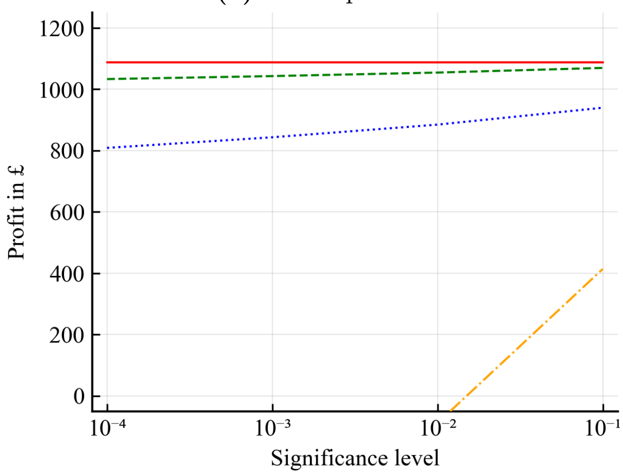


(c) 1000 samples.

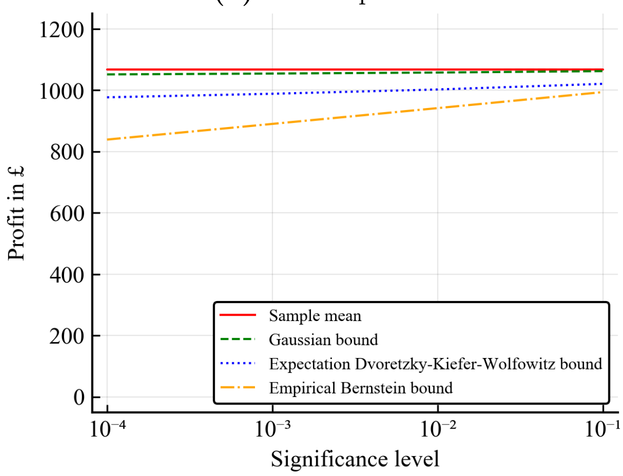
Figure 3.7: Probabilistic bounds – empirical Cantelli bound (green dashed line, Proposition 3.5(i)) and Dvoretzky-Kiefer-Wolfowitz tail bound (blue dotted line, Proposition 3.5(ii)) – on the profit obtained from a single validation sample as functions of various significance levels for 10 (a), 100 (b) and 1000 (c) validation samples.



(a) 10 samples.



(b) 100 samples.



(c) 1000 samples.

Figure 3.8: Probabilistic bounds – Gaussian bound (green dashed line, (3.11)), expectation Dvoretzky-Kiefer-Wolfowitz bound (blue dotted line, Proposition 3.7(ii)) and empirical Bernstein bound (orange dash-dotted line, Proposition 3.7(i)) – on the expectation of the profit obtained as functions of various significance levels and for 10 (a), 100 (b) and 1000 (c) validation samples.

3.7 Conclusion

In this chapter, we presented a new algorithm to compute approximate solutions to DP problems with submodular, concave extensible value functions. An example of such a DP is the revenue management problem in attended home delivery introduced in Section 2.2. We derived deterministic upper and stochastic lower bounds to the exact value function. We showed that our algorithm converged in a finite number of iterations and we derived tail and expectation bounds for the stochastic lower bound. Finally, we demonstrated our results in an example of the revenue management problem in attended home delivery.

Comparing deterministic upper with stochastic lower bound, we can quantify the profit generation efficiency of our algorithm. This serves as a benchmark for other algorithms, possibly with weaker theoretical guarantees, but better practical performance. Furthermore, the gap between deterministic upper and stochastic lower bound will allow to quantify and optimise the trade-off between quality of approximation and computational cost *dynamically* as an application runs. For example, in the revenue management problem in attended home delivery, the terminal condition of the DP is an approximation to the intractable capacitated vehicle routing problem with time windows [16], where capacitated refers to additional constraints describing that delivery vehicles can only transport a finite amount of goods. As time in the booking horizon progresses, orders come in and reveal the location of customers. This information could be used to obtain a closer estimate of the terminal condition of the DP. For example, any time an order is placed, one could update the terminal condition using location information of the most recent customer and then compute our algorithm again with the modified terminal condition. Pricing based on the associated updated policy would then take into account the newly available information.

Another possible direction for future research is to devise *deterministic* lower bounds for the gradient-bounded DP algorithm, instead of the proposed *stochastic* lower bounds in this chapter. Such deterministic lower bounds have already been devised by [63] for the classical stochastic dual DP algorithm from [47, 48].

Having derived an approximate DP algorithm for the revenue management problem in attended home delivery, the focus of the next chapter is to view the developed algorithm as well as other algorithms in the literature as instances of a generic algorithmic scheme. We will compare these algorithms in terms of computation time and profit generation performance, for which we will also use the probabilistic bounds derived in this chapter.

3.A Proofs

This appendix contains mathematical proofs omitted in the main body of Chapter 3.

3.A.1 Proof of Proposition 3.1

We show this result by induction on t. In the base case (the terminal condition), $Q^i_{\bar{t}+1}(x) := V_{\bar{t}+1}(x) = -C(x)$ for all $(x,i) \in X \times I$, which satisfies the proposition trivially by Assumption 3.1. Assume for an induction hypothesis that $Q^{i-1}_{t+1}(x) \geq V_{t+1}(x)$ for some $(i,t) \in I \setminus \{0\} \times T$ and for all $x \in X$. Fix any x in X and distinguish the two cases of the if-statement in step 12 of Algorithm 3.1.

Case I: Suppose that Q^{i-1}_{t+1} is submodular on $Z(x^i_{t+1})$. Then H^* is the unique hyperplane through the set $\{(y, (\mathcal{T}Q^i_{t+1})(y))\}_{y \in Y_+(x^i_{t+1})}$. By (3.1), Q^i_{t+1} is concave extensible since it is the pointwise minimum of a finite number of hyperplanes. Hence, we invoke Assumption 3.2 to conclude that $\mathcal{T}Q^{i-1}_{t+1}$ is concave extensible and submodular. As shown in Appendix 2.A.2, this implies that H^* is a separating hyperplane, i.e. $H^*(x) \geq \mathcal{T}Q^{i-1}_{t+1}(x)$ for all $x \in X$. Define d^V to be the maximiser of (2.4) and define d^Q to be the maximiser of (2.4) with $V_{t+1}(y)$ replaced by $Q^{i-1}_{t+1}(y)$. We now show that the Bellman operator of the DP preserves the inequality $Q^{i-1}_{t+1}(x) \geq V_{t+1}(x)$, i.e. $\mathcal{T}Q^{i-1}_{t+1}(x) \geq \mathcal{T}V_{t+1}(x)$. To this end, fix any $x \in X$ and consider

$$\begin{aligned}
(\mathcal{T}Q^{i-1}_{t+1})(x) &= \sum_{y \in Y_+(x)} \pi_{x,y}(d^Q) \left[\hat{\rho}(x,y,d^Q) + Q^{i-1}_{t+1}(y)\right] \\
&\geq \sum_{y \in Y_+(x)} \pi_{x,y}(d^V) \left[\hat{\rho}(x,y,d^V) + Q^{i-1}_{t+1}(y)\right] \\
&\geq \sum_{y \in Y_+(x)} \pi_{x,y}(d^V) \left[\hat{\rho}(x,y,d^V) + V_{t+1}(y)\right] \\
&= (\mathcal{T}V_{t+1})(x), \qquad (3.13)
\end{aligned}$$

where the first inequality follows from the supoptimality of d^V for $(\mathcal{T}Q^{i-1}_{t+1})(x)$ and the second inequality follows from the induction hypothesis.

Case II: Now consider the case when Q^{i-1}_{t+1} is not submodular on $Z(x^i_{t+1})$. Then $H^* \in \left\{\mathcal{T}H^{j-1}_{t+1} \;\middle|\; j \in J^{i-1}_{t+1}\right\}$. Furthermore, by (3.1) and the induction hypothesis,

$$H^{j-1}_{t+1}(x) \geq Q^{i-1}_{t+1}(x) \geq V_{t+1}(x), \text{ for all } (x,j) \in X \times J^{i-1}_{t+1}. \qquad (3.14)$$

We now show that all possible realisations of H^* constitute upper bounds on $\mathcal{T}V_{t+1}$. To this end, fix any $(x, j) \in X \times J^{i-1}_{t+1}$. Define d^H to be the maximiser of (2.6) with $V_{t+1}(y)$ replaced by $H^{j-1}_{t+1}(y)$. We can show that the Bellman operator of the DP preserves the inequality $Q^{i-1}_{t+1}(x) \geq V_{t+1}(x)$ using a similar argument as before:

$$(\mathcal{T}H^{j-1}_{t+1})(x) \geq (\mathcal{T}V_{t+1})(x), \tag{3.15}$$

which follows from the suboptimality of d^V (see Case I) for $(\mathcal{T}H^{j-1}_{t+1})(x)$ and the fact that $H^{j-1}_{t+1}(x) \geq V_{t+1}(x)$ (see (3.14)). Therefore, we conclude that $H^*(x) \geq \mathcal{T}V_t(x)$ for all $x \in X$ in the second case as well.

Since both cases lead to an upper bound, i.e. $H^*(x) \geq \mathcal{T}V_{t+1}(x)$ for all $x \in X$, we infer that

$$Q^i_t(x) = \min\left\{H^*(x), Q^{i-1}_{t+1}(x)\right\} \geq \mathcal{T}V_{t+1}(x) \tag{3.16}$$

for all $x \in X$. This concludes our induction argument and shows that $Q^i_t(x) \geq V_t(x)$ for all $(x, i, t) \in X \times I \times T$. □

3.A.2 Proof of Proposition 3.4

We will show the proposition by induction on t. Consider the base case, when $Q^0_{\bar{t}+1}(x) = V_{\bar{t}+1}(x)$ for all $x \in X$. Then notice that in the "backward sweep", the proposed algorithm computes the Bellman equation from $\bar{t}+1 \to \bar{t}$ exactly for every $x \in X$. This is because $Q^0_{\bar{t}+1}$ is submodular by Assumption 3.1 and hence, the if-statement in step 12 of Algorithm 3.1 is true. By Assumption 3.3, $x^i_{\bar{t}+1}$ is resampled if for the time step transition $\bar{t}+1 \to \bar{t}$, the algorithm has already sampled this state in a previous iteration. Therefore, the value function is computed exactly at all $x \in X$ for the time step transition $\bar{t}+1 \to \bar{t}$ after at most $|X|$ iterations of the proposed algorithm, i.e. $Q^{\hat{i}}_{\bar{t}}(x) = V_{\bar{t}}(x)$ for all $x \in X$, where $\hat{i} \leq |X|$.

Now suppose by means of an induction hypothesis that for some $(t, i) \in T \times I$, $Q^i_{t+1}(x) = V_{t+1}(x)$ for all $x \in X$. Then by Assumptions 3.1 and 3.2, V_{t+1} is submodular and hence, Q^i_{t+1} is also submodular. By a similar argument to the base case, the proposed algorithm computes the exact value function for the time step transition $t+1 \to t$ in another $\hat{i} \leq |X|$ iterations.

Hence, we conclude that for every time step transition, the proposed algorithm needs at most $|X|$ iterations to compute the exact value function for any one time step $t \in T$, which gives at most $\bar{t}|X|$ iterations for the total time horizon. Hence, after any $i \geq \bar{t}|X|$ iterations, $Q_t^i(x) = V_t(x)$ for all $(x,t) \in X \times T$. Therefore, both $\mathbb{E}[l(i)] = V_1(0)$ and $u(i) = Q_1^i(0) = V_1(0)$, which finally implies that $\mathbb{E}[l(i)] = u(i)$ for all $i \geq \bar{t}|X|$ iterations. $\square$

3.A.3 Proof of Proposition 3.5(i)

The proof is a finite sample adaptation of the one-sided Chebyshev's inequality, i.e. Cantelli's inequality [64, Theorem 1]. We distinguish the following two cases:

Case I: Suppose that $\sigma_{\mathrm{v}} \neq 0$. Fix any $k \in K$ and consider the conditional probability that $l_{\mathrm{C}} := l_{\mathrm{v}}(k_{\max}+1)$ is no greater than $\bar{l}_{\mathrm{v}} - m\hat{\sigma}$ for some $m > 0$:

$$\begin{aligned}
&\Pr(l_{\mathrm{C}} \leq \bar{l}_{\mathrm{v}} - m\sigma_{\mathrm{v}} | \sigma_{\mathrm{v}} \neq 0) \\
=& \Pr(m\sigma_{\mathrm{v}} \leq \bar{l}_{\mathrm{v}} - l_{\mathrm{C}} | \sigma_{\mathrm{v}} \neq 0) \\
=& \frac{1}{k_{\max}} \sum_{k\in K} \Pr(m\sigma_{\mathrm{v}} \leq \bar{l}_{\mathrm{v}} - l_{\mathrm{v}}(k) | \sigma_{\mathrm{v}} \neq 0) \\
=& \frac{1}{k_{\max}} \mathbb{E}\left(\sum_{k\in K} \mathbb{1}(m\sigma_{\mathrm{v}} \leq \bar{l}_{\mathrm{v}} - l_{\mathrm{v}}(k)) \middle| \sigma_{\mathrm{v}} \neq 0 \right),
\end{aligned} \tag{3.17}$$

where the second last equality follows from the observation that $l_{\mathrm{v}}(k)$ for all $k \in K \cup \{k_{\max}+1\}$ are independently and identically distributed. Next, we want to upper bound the indicator function in (3.17) by a quadratic function. A suitable expression is given for any $c > 0$ by

$$\begin{aligned}
&\Pr(l_{\mathrm{C}} \leq \bar{l}_{\mathrm{v}} - m\sigma_{\mathrm{v}} | \sigma_{\mathrm{v}} \neq 0) \\
\leq& \frac{1}{k_{\max}} \mathbb{E}\left(\sum_{k\in K} \frac{(\bar{l}_{\mathrm{v}} - l_{\mathrm{v}}(k) + c\sigma_{\mathrm{v}})^2}{(m\sigma_{\mathrm{v}} + c\sigma_{\mathrm{v}})^2} \middle| \sigma_{\mathrm{v}} \neq 0 \right).
\end{aligned} \tag{3.18}$$

Note that each element in the summation is always non-negative and no smaller than one if $m\sigma_{\mathrm{v}} \leq \bar{l} - l_{\mathrm{v}}(k)$ and hence is an upper bound to (3.17). We simplify this expression using the definitions of $\bar{l}_{\mathrm{v}}$ and σ_{v} from Algorithm 3.3 in Section 3.4 as

$$\begin{aligned}
&\frac{1}{k_{\max}}\mathbb{E}\left(\sum_{k\in K}\frac{(\bar{l}_{\mathrm{v}}-l_{\mathrm{v}}(k)+c\sigma_{\mathrm{v}})^2}{(m\sigma_{\mathrm{v}}+c\sigma_{\mathrm{v}})^2}\middle|\,\sigma_{\mathrm{v}}\neq 0\right)\\
=&\frac{1}{k_{\max}}\mathbb{E}\left(\frac{(k_{\max}-1)\sigma_{\mathrm{v}}^2+k_{\max}c^2\sigma_{\mathrm{v}}^2}{(m+c)^2\sigma_{\mathrm{v}}^2}\middle|\,\sigma_{\mathrm{v}}\neq 0\right)\\
=&\frac{1}{k_{\max}}\mathbb{E}\left(\frac{k_{\max}-1+k_{\max}c^2}{(m+c)^2}\right)\\
=&\frac{k_{\max}-1+k_{\max}c^2}{k_{\max}(m+c)^2},
\end{aligned}\tag{3.19}$$

where σ_{v} cancels, since $\sigma_{\mathrm{v}}\neq 0$, and the expectation operator drops, since its argument is a constant. We minimise (3.19) by considering its first order condition, i.e.

$$\begin{aligned}
0 &= \frac{\partial}{\partial c}\frac{k_{\max}-1+k_{\max}c^2}{k_{\max}(m+c)^2}\\
&= (k_{\max}(m+c)^2)^{-2}\Big\{2k_{\max}^2 c(m+c)^2\\
&\quad -(k_{\max}-1+k_{\max}c^2)2k_{\max}(m+c)\Big\}\\
\Rightarrow c &= \frac{k_{\max}-1}{k_{\max}m}.
\end{aligned}\tag{3.20}$$

The second-order condition shows that c minimises (3.19). Substituting c into (3.19) and simplifying gives:

$$\Pr(l_{\mathrm{C}}\leq \bar{l}_{\mathrm{v}}-m\sigma_{\mathrm{v}}|\sigma_{\mathrm{v}}\neq 0)\leq\frac{k_{\max}-1}{k_{\max}m^2+k_{\max}-1}.\tag{3.21}$$

Case II: Suppose that $\sigma_{\mathrm{v}}=0$. We repeat the derivation of Case I with $\Pr(l_{\mathrm{C}}\leq \bar{l}_{\mathrm{v}}-m\sigma_{\mathrm{v}}|\sigma_{\mathrm{v}}\neq 0)$ replaced by $\Pr(l_{\mathrm{C}}\leq \bar{l}_{\mathrm{v}}-m\sigma_{\mathrm{v}}|\sigma_{\mathrm{v}}=0)$ until (3.17), where we note that due to $\sigma_{\mathrm{v}}=0$, we have $l_{\mathrm{v}}(k)=l_{\mathrm{v}}(k')$ for all $(k,k')\in K\times K$ and hence, $\Pr(l_{\mathrm{C}}\leq \bar{l}_{\mathrm{v}}-m\sigma_{\mathrm{v}}|\sigma_{\mathrm{v}}=0)=1$.

Recalling that $\theta_{\mathrm{C}}:=\Pr(\sigma_{\mathrm{v}}=0)$ and taking both cases together, we obtain by the total probability theorem that

$$\Pr(l_{\mathrm{C}}\leq \bar{l}_{\mathrm{v}}-m\sigma_{\mathrm{v}})\leq\frac{(1-\theta_{\mathrm{C}})(k_{\max}-1)}{k_{\max}m^2+k_{\max}-1}+\theta_{\mathrm{C}}.\tag{3.22}$$

We want this probability to be at most the significance level α. Hence, we solve this expression for m, yielding

$$\begin{aligned}
&\alpha\geq(1-\theta_{\mathrm{C}})\frac{k_{\max}-1}{k_{\max}m^2+k_{\max}-1}+\theta_{\mathrm{C}}\\
\iff\ & m\geq\sqrt{\frac{(1-\alpha)(k_{\max}-1)}{(\alpha-\theta_{\mathrm{C}})k_{\max}}},
\end{aligned}\tag{3.23}$$

which is real-valued, since $\alpha > \theta_\mathrm{C}$ by Assumption 3.4. Substituting for m on the left-hand side of (3.22) gives the desired property:

$$\Pr\left(l_\mathrm{C} \leq \bar{l} - \sigma_\mathrm{v}\sqrt{\frac{(1-\alpha)(k_\mathrm{max}-1)}{(\alpha-\theta_\mathrm{C})k_\mathrm{max}}}\right) \leq \alpha. \tag{3.24}$$

□

3.A.4 Proof of Proposition 3.5(ii)

Fix any $\alpha \in (0,1)$, any $\theta_\mathrm{D} \in (0,\alpha)$ and compute

$$l_\mathrm{D} = \sup\left\{ l \in [l_-, l_+] \,\middle|\, F_K(l) \leq \alpha - \theta_\mathrm{D} - \sqrt{\frac{\ln(1/\theta_\mathrm{D})}{2k_\mathrm{max}}} \right\}. \tag{3.25}$$

We will now show that $l_\mathrm{v}(k_\mathrm{max}+1) > l_\mathrm{D}$ with probability at least $1-\alpha$. By the total probability theorem, we write

$$\Pr(l_\mathrm{v}(k_\mathrm{max}+1) > l_\mathrm{D}) = \Pr(B|E)\Pr(E) + \Pr(B|E^c)\Pr(E^c), \tag{3.26}$$

where B denotes the random event that $l_\mathrm{v}(k_\mathrm{max}+1) > l_\mathrm{D}$, i.e. $\Pr(B) = 1 - F(l_\mathrm{D})$, and E denotes the random event that

$$F(l_\mathrm{D}) \leq F_K(l_\mathrm{D}) + \sqrt{\frac{\ln(1/\theta_\mathrm{D})}{2k_\mathrm{max}}}, \tag{3.27}$$

which according to the Dvoretzky-Kiefer-Wolfowitz inequality [59] has probability $\Pr(E) \geq 1-\theta_\mathrm{D}$. E^c denotes the complementary event of E. Notice that $\Pr(E^c)$ and $\Pr(B|E^c)$ are non-negative and hence, we can create the following lower bound:

$$\Pr(l_\mathrm{v}(k_\mathrm{max}+1) > l_\mathrm{D}) \geq \Pr(B|E)\Pr(E). \tag{3.28}$$

Since we condition on E, we can lower bound $\Pr(B|E) \geq 1 - (F_K(l_\mathrm{D}) + \sqrt{\ln(1/\theta_\mathrm{D})/(2k_\mathrm{max})})$ according to (3.27), which yields

$$\begin{aligned}
\Pr(l_\mathrm{v}(k_\mathrm{max}+1) > l_\mathrm{D}) &\geq \left[1 - \left(F_K(l_\mathrm{D}) + \sqrt{\frac{\ln(1/\theta_\mathrm{D})}{2k_\mathrm{max}}}\right)\right](1-\theta_\mathrm{D}) \\
&> 1 - \left(F_K(l_\mathrm{D}) + \sqrt{\frac{\ln(1/\theta_\mathrm{D})}{2k_\mathrm{max}}}\right) - \theta_\mathrm{D} \\
&\geq 1 - (\alpha - \theta_\mathrm{D}) - \theta_\mathrm{D} = 1 - \alpha,
\end{aligned} \tag{3.29}$$

where the final inequality follows from the choice of l_D in (3.25), which thus concludes our proof. □

3.A.5 Proof of Lemma 3.6

Consider the first-order optimality condition, i.e.

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta_{\mathrm{D}}} \left\{ \alpha - \theta_{\mathrm{D}} - \sqrt{\frac{\ln(1/\theta_{\mathrm{D}})}{2k_{\max}}} \right\} \\ \Rightarrow 0 &= -1 + \frac{1}{2\theta_{\mathrm{D}} \sqrt{-2k_{\max} \ln(\theta_{\mathrm{D}})}}. \end{aligned} \tag{3.30}$$

Since $\theta_{\mathrm{D}} > 0$, we can simplify to arrive at

$$\begin{aligned} \theta_{\mathrm{D}}^2 \ln(\theta_{\mathrm{D}}^2) &= \frac{-1}{4k_{\max}} \\ \Rightarrow \theta_{\mathrm{D}}^2 &= \exp\left(W_i \left(\frac{-1}{4k_{\max}} \right) \right), \end{aligned} \tag{3.31}$$

where $i \in \{0, -1\}$ and W_i is the Lambert W function (see Definition 3.1). The second-order conditions show that $i = -1$ gives a local maximum. Hence, we take the square root and note that θ_{D} is bounded from above by α to arrive at the desired result. □

3.A.6 Proof of Proposition 3.7(ii)

We start by writing the expectation of the random variable l in terms of its exact yet unknown cumulative distribution function F (see [65, Definition 3.7.1]), i.e.

$$\begin{aligned} \mathbb{E}l &:= \int_{l=0}^{\infty} 1 - F(l)\mathrm{d}l - \int_{l=-\infty}^{0} F(l)\mathrm{d}l. \\ &= \int_{l=0}^{\infty} 1 - F(l)\mathrm{d}l + \int_{l=-\infty}^{0} -F(l)\mathrm{d}l \\ &= \int_{l=\max\{0,l_-\}}^{\max\{0,l_+\}} 1 - F(l)\mathrm{d}l + \max\{0, l_-\} + \int_{l=\min\{0,l_-\}}^{\min\{0,l_+\}} -F(l)\mathrm{d}l - \min\{0, l_+\}, \end{aligned} \tag{3.32}$$

where we changed the limits of integration, since F has the finite support $[l_-, l_+]$. The max- and min-operators in the integration limits ensure that equality holds independent of the sign of either l_- or l_+. By the Dvoretzky-Kiefer-Wolfowitz inequality [59], we can write with confidence $1 - \alpha^{\mathbb{E}}$ that

$$F(l) \leq \min \left\{ 1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}} \right\}$$

$$\iff \; -F(l) \geq -\min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\}$$
$$\iff \; 1 - F(l) \geq 1 - \min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\}. \tag{3.33}$$

Using these expressions for $-F(l)$ and $1 - F(l)$, we lower bound (3.32) by

$$\begin{aligned}\mathbb{E}l \geq & \int_{l=\max\{0,l_-\}}^{\max\{0,l_+\}} 1 - \min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\} \mathrm{d}l + \max\{0, l_-\} \\ & - \int_{l=\min\{0,l_-\}}^{\min\{0,l_+\}} \min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\} \mathrm{d}l - \min\{0, l_+\}.\end{aligned} \tag{3.34}$$

Finally, $\mathbb{E}\bar{l}_{\mathrm{v}} = \mathbb{E}l$, since $\mathbb{E}$ is a linear operator and $l_{\mathrm{v}}(k)$ for all $k \in K$ are independent, thus concluding the proof. □

3.A.7 Proof of Proposition 3.8

We can express the empirical mean in Hoeffding's bound [66] as an integral over the empirical cumulative distribution function:

$$\begin{aligned}l_{\mathrm{H}}^{\mathbb{E}} &:= \bar{l}_{\mathrm{v}} - (l_+ - l_-)\sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}} \\ &= \int_{l=0}^{\infty} 1 - F_K(l)\mathrm{d}l + \int_{l=-\infty}^{0} -F_K(l)\mathrm{d}l - (l_+ - l_-)\sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}} \\ &= \int_{l=\max\{0,l_-\}}^{\max\{0,l_+\}} 1 - \left(F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right) \mathrm{d}l + \max\{0, l_-\} \\ &\quad + \int_{l=\min\{0,l_-\}}^{\min\{0,l_+\}} -\left(F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right) \mathrm{d}l - \min\{0, l_+\} \\ &< \int_{l=\max\{0,l_-\}}^{\max\{0,l_+\}} 1 - \min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\} \mathrm{d}l + \max\{0, l_-\} \\ &\quad + \int_{l=\min\{0,l_-\}}^{\min\{0,l_+\}} -\min\left\{1, F_K(l) + \sqrt{\frac{\ln(1/\alpha^{\mathbb{E}})}{2k_{\max}}}\right\} \mathrm{d}l - \min\{0, l_+\} \\ &= l_{\mathrm{D}}^{\mathbb{E}},\end{aligned} \tag{3.35}$$

where the third equality follows from the fact that the (finite) support of $F_K(l)$ is $[l_-, l_+]$ and the last inequality is strict by Assumption 3.5, since F_K is a stair function with step height $1/k_{\max}$ and $\sqrt{\ln(1/\alpha^{\mathbb{E}})/(2k_{\max})} > 1/k_{\max}$ implies that $F_K(l^*) + \sqrt{\ln(1/\alpha^{\mathbb{E}})/(2k_{\max})} > 1$ for some $l^* < l_+$. □

3.B Synthetic problem data for Section 3.6.1

Table 3.2: Synthetically generated customer choice value parameters for $s > 17$.

s	β_s	s	β_s	s	β_s	s	β_s
18	-1.4948	64	-0.1006	110	-0.3293	156	-1.0428
19	-0.3982	65	-1.1569	111	0.3511	157	0.5934
20	-1.5235	66	-0.6968	112	0.5280	158	-1.3042
21	-1.2132	67	-0.0643	113	-1.0187	159	-1.3412
22	0.1257	68	-1.0906	114	-0.6402	160	-1.4544
23	0.0717	69	0.1411	115	-0.4081	161	-0.8428
24	-1.0794	70	-1.4845	116	-0.9579	162	-1.2215
25	0.6062	71	-0.8621	117	-0.7033	163	-0.4735
26	-0.2441	72	-0.6315	118	-1.2536	164	0.4132
27	-0.2363	73	-1.6359	119	0.3055	165	0.4442
28	0.1553	74	-1.6201	120	0.5313	166	-0.5653
29	-1.0293	75	-1.0187	121	-0.1290	167	-1.0691
30	-1.6489	76	-0.3666	122	-1.3215	168	-0.7470
31	-1.1742	77	0.5036	123	-0.3422	169	-1.0176
32	-0.6005	78	-0.4982	124	-0.7985	170	-0.1399
33	-0.8926	79	0.4254	125	-0.6893	171	-0.2895
34	-1.4193	80	-0.2841	126	-1.2393	172	-0.3356
35	-1.5926	81	-0.2815	127	0.5629	173	-0.7097
36	-1.1868	82	0.1361	128	-0.4196	174	-1.1692
37	-0.4536	83	-1.2532	129	-1.6022	175	-0.4171
38	-0.1797	84	-1.1356	130	-0.7021	176	-0.2733
39	0.2971	85	-1.0231	131	0.1679	177	-0.2412
40	-0.5314	86	-0.4525	132	-1.2592	178	-0.8127
41	0.2137	87	0.3880	133	-0.9011	179	0.5041
42	-1.6574	88	0.6077	134	-1.3455	180	-0.2414
43	-0.2698	89	-0.0412	135	-1.5326	181	-0.9836
44	-0.4528	90	-0.0333	136	0.2803	182	-1.2706
45	0.1892	91	0.1488	137	-0.8387	183	-0.7475
46	-1.2540	92	0.1707	138	-0.6186	184	0.0243
47	0.2366	93	-0.1440	139	-1.5771	185	-0.3349
48	0.1301	94	-0.0318	140	-0.1335	186	-1.4789
49	-1.4047	95	-0.2758	141	-1.2653	187	-1.3107
50	-0.5401	96	-0.7432	142	-0.4687	188	-0.7177
51	-1.4222	97	-0.2195	143	0.0370	189	-0.3111
52	0.0181	98	-1.5782	144	-0.6932	190	-0.5097
53	-0.2588	99	-0.2688	145	-1.4828	191	-1.4646
54	-1.5921	100	-1.5123	146	-1.6513	192	-0.1555
55	-1.4008	101	0.2595	147	-0.6886	193	0.0559
56	-0.6508	102	-1.5315	148	-0.3817	194	0.6151
57	-0.1053	103	-1.5131	149	0.6125	195	0.0129
58	-0.2780	104	-1.2932	150	-0.6743	196	-1.0818
59	-0.2188	105	-1.1281	151	0.5344	197	-0.1089
60	-1.2916	106	0.1615	152	-0.2535	198	-0.3505
61	-0.4877	107	0.3028	153	0.0625	199	-0.1765
62	0.3306	108	-1.0030	154	-1.0588	200	-1.3724
63	0.4735	109	-0.2251	155	0.4163	201	0.1678

4

Comparative Study

of Approximate Dynamic Programming Algorithms
for Revenue Management in Attended Home Delivery

Contents

4.1 Introduction **86**

4.2 General sample-based, iterative approximation algorithm **89**

4.2.1 Affine value function approximation update 91

4.2.2 Non-linear stochastic dual dynamic programming update 92

4.2.3 Gradient-bounded dynamic programming update 95

4.3 Case study **95**

4.3.1 Exact model analysis 96

4.3.2 Parameter sensitivity analysis 102

4.4 Conclusion **107**

4.A Problem reformulation and table of parameters **108**

4.A.1 Problem reformulation for non-linear stochastic dual dynamic programming 108

4.A.2 Table of parameters for Section 4.3.2 110

4.1 Introduction

In the previous chapter, we derived our own approximate DP algorithm, termed gradient-bounded DP, to solve the revenue management problem in attended home delivery introduced in Chapter 2. In this chapter, we view gradient-bounded

DP as well as two other algorithms in the literature as instances of a generic algorithmic scheme. Using this framework, we examine the theoretical and practical performance of these algorithms.

For the problem under study, [20] proposes an algorithm that approximates the exact value function of the DP as an affine function in the vector of states. In its conclusions, [20] suggests that a possible direction for further research would be to explore non-linear approximate value functions. This direction is also proposed for future exploration in the conclusions of [67]. The recently developed non-linear stochastic dual DP algorithm in [46] was derived outside the attended home delivery literature and it provides such non-linear approximate value functions and we have already applied it for numerical validation of our theoretical results of Chapter 2 in Section 2.5.

However, in this chapter we view these algorithms under a control theoretic lens and show that they both exhibit certain limitations. These can be overcome by the gradient-bounded DP algorithm presented in the previous chapter. The main contributions of this chapter are thus:

1. We explain the theoretical limitations of the affine value function approximation algorithm from [20] by showing that it results in an open loop controller, thus motivating the use of non-linear value function approximations, which provide state feedback. The intuition behind this observation is that the control input in our problem setting depends on the gradient of the (approximate) value function. Since the gradient of an affine value function approximation is constant, it will generate identical control inputs independent of the state. This motivates the use of a non-linear value function approximation, whose gradient varies across states and hence provides state feedback.

2. We adapt the non-linear stochastic dual DP algorithm from [46] to the revenue management problem in attended home delivery and show that we can use a simplified version of the algorithm, which neglects a regularisation term that is redundant for the problem under study in this dissertation. However, despite

it being able to work as a closed loop controller, we show that the non-linear stochastic dual DP imposes difficulties in the computation of the optimal control policy since it requires the solution of a non-convex optimisation problem.

3. We then show that our gradient-bounded DP algorithm from Chapter 3 can overcome the limitations of the other two algorithms, since it provides non-linear value function approximations that can, however, be computed using convex optimisation. We demonstrate the efficacy of this approach by means of a multi-scenario case study based on data from [20], in which we benchmark the performance of all three algorithms when the model parameters are known exactly. Furthermore, we include a detailed model parameter sensitivity analysis, which quantifies the robustness of the algorithms against modelling errors, thus seeking to provide a balanced view on how these algorithms are likely to perform in practice. Such a comparative numerical case study is new for the problem under study and hence complements the numerical studies on attended home delivery conducted on other problem formulations, e.g. in [67].

The structure of the remaining chapter is as follows: Section 4.2 presents a general, sample-based approximate DP algorithm and we explain how the three algorithms considered in this chapter are special cases of this general algorithm. In that section, we also elaborate on the main theoretical limitations of the first two algorithms and how gradient-bounded DP overcomes these. We then demonstrate how the gradient-bounded DP algorithm dominates the two other algorithms by means of a multi-scenario numerical case study and robustness to uncertainty analysis in Section 4.3. Finally, we conclude the chapter and provide some directions for future research in Section 4.4.

4.2 General sample-based, iterative approximation algorithm

As mentioned in the introduction to our theoretical developments in Section 2.1, a key to solving the revenue management problem in attended home delivery is to approximate the value function of the DP in (2.4) effectively. A popular strategy, not only for this problem (see [20, 67]), but also for other stochastic multi-stage problems (see [47, 48]), 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 obtained sample paths.

We state a general, sample-based, iterative approximate DP procedure in Algorithm 4.1 below. The three algorithms that we investigate in this chapter are special cases of this general algorithm and they differ only in its 11$^{\text{th}}$ step. We detail how this step is computed for each of the individual specific algorithms further below.

Algorithm 4.1 General sample-based, iterative approximation algorithm

1: Initialise parameters: $X, D, p_s, T, r, \lambda, C$ and $i_{\max}$
2: Initialise $Q^0_t(x) \leftarrow (\bar{d} + r)\langle \mathbf{1}, \bar{x} - x\rangle - C(\bar{x})$, for all $(x, t) \in X \times T$
3: Initialise $Q^0_{\bar{t}+1}(x) \leftarrow -C(x)$, for all $x \in X$
4: **for** $i \in \{1, 2, \ldots, i_{\max}\}$ **do**
5: $\quad x^i_1 \leftarrow \mathbf{0}$
6: $\quad$ **for** $t \in \{1, 2, \ldots, \bar{t}\}$ **do** $\qquad \triangleright$ "Forward sweep"
7: $\quad\quad d^i_t \leftarrow d^*(x^i_t)$, the solution of (4.2)
8: $\quad\quad x^i_{t+1} \leftarrow x^i_t + \underset{x^i_{t+1}}{\text{sample}}\{p_s(d^i_t)\}$
9: $\quad$ **end for**
10: $\quad$ **for** $t \in \{\bar{t}, \bar{t} - 1, \ldots, 1\}$ **do** $\qquad \triangleright$ "Backward sweep"
11: $\quad\quad Q^i_t \leftarrow$ update Q^{i-1}_t
12: $\quad$ **end for**
13: **end for**

We first initialise all parameters of the DP in (2.4) (step 1). Similarly to our notation in the previous chapter, 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 initialise Q_t^0 to any value as long as that 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 initialise Q_t^0 for all $t \in T$ using the unique fixed point of DP, V^* (step 2) (see also the discussion in the numerical examples of the previous chapters in Sections 2.5.3 and 3.6). Theorem 2.1 shows that the fixed point is given analytically (under mild technical assumptions) as

$$V^*(x) := (\bar{d} + r)\langle \mathbf{1}, \bar{x} - x \rangle - C(\bar{x}), \text{ for all } x \in X. \tag{4.1}$$

Recall that V^* is an upper bound to V_t for all $t \in T$, since the DP operator $\mathcal{T}$ is monotone (see [37, Chapter 3]). Finally, we also initialise $Q_{\bar{t}+1}^i(x) := V_{\bar{t}+1}(x) = -C(x)$ for all $(x, i) \in X \times I$ (step 3).

For the next steps, fix any iteration $i \in I \setminus \{0\}$. In each "forward sweep", we simulate an approximate version of the DP (2.4) forward in time by replacing V_t with its approximation Q_t^{i-1} (step 7), i.e.

$$\begin{aligned} d^*(x_t^i) := \operatorname*{argmax}_{d \in D} \Bigg\{ & \lambda \sum_{s \in S} p_s(d) \left(r + d_s + Q_{t+1}^{i-1}(x_t^i + 1_s) \right) \\ & + \left(1 - \lambda \sum_{s \in S} p_s(d) \right) Q_{t+1}^{i-1}(x_t^i) \Bigg\}. \end{aligned} \tag{4.2}$$

Notice that [25, Theorem 1] have shown that for the multinomial logit choice model, the maximisers are unique for the above expression, hence we use equality in (4.2) instead of specifying $d^*(x_t^i)$ as one element of a set. We will use this choice model for our numerical case study in Section 4.3. By (4.2), 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 (step 8). This defines a *sample path* $\{x_t^i\}_{t \in T \cup \{\bar{t}+1\}}$.

In each "backward sweep", we update the value function approximation using the particular mechanism of the chosen algorithm (step 11). These two sequences – "forward sweep" and "backward sweep" – are repeated for $i_{\max}$ iterations. In the next sections, we describe the exact mechanisms of the three algorithms considered in this chapter, making step 11 in Algorithm 4.1 explicit.

4.2.1 Affine value function approximation update

This approach is proposed in [20]. The idea is to approximate the value function by an affine function of the form

$$Q_t^i(x) := \theta_0^i + (\bar{t} + 1 - t)\theta^i - \sum_s \theta_s^i x_s, \tag{4.3}$$

for all $(x, i, t) \in X \times I \times T$ and where θ_s^i, for all $s \in S \cup \{0\}$ and θ^i are scalar, real-valued parameters, for all $i \in I$. Then, the updating rule in step 11 of Algorithm 4.1 is a gradient descent step to bring Q_t^i closer to $\mathcal{T}Q_{t+1}^i$. In particular, we seek to minimise $\left(Q_t^i(x_{t+1}^i) - \mathcal{T}Q_{t+1}^i(x_{t+1}^i)\right)^2$. Therefore, the steps to update the parameters are

$$\begin{aligned}
\theta_0^i &= \theta_0^{i-1} - \alpha_1 \left(Q_t^{i-1}(x_{t+1}^i) - \mathcal{T}Q_{t+1}^{i-1}(x_{t+1}^i)\right) \\
\theta_s^i &= \theta_s^{i-1} - \alpha_2 \left(Q_t^{i-1}(x_{t+1}^i) - \mathcal{T}Q_{t+1}^{i-1}(x_{t+1}^i)\right) x_{s,t+1}^i, \text{ for all } s \in S \\
\theta^i &= \theta^{i-1} - \alpha_3 \left(Q_t^{i-1}(x_{t+1}^i) - \mathcal{T}Q_{t+1}^{i-1}(x_{t+1}^i)\right) (\bar{t} + 1 - t),
\end{aligned} \tag{4.4}$$

where the (positive) step sizes α_1, α_2 and α_3 are chosen to be sufficiently small for convergence of the above iterative procedure (see e.g. [68, Lemma 8.2]).

One important observation from a control perspective is that a value function approximation that is affine in x implies that the approximately optimal pricing control obtained by solving (4.2) will have *no state feedback* for all states $x \in X$ such that $x + 1_s \in X$ for all $s \in S$. To see this, notice that (4.2) can be re-written 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:

$$\begin{aligned}
d^*(x_t^i) &= \underset{d \in D}{\operatorname{argmax}} \left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + Q_{t+1}^{i-1}(x_t^i + 1_s) - Q_{t+1}^{i-1}(x_t^i) \right) + Q_{t+1}^{i-1}(x_t^i) \right\} \\
&= \underset{d \in D}{\operatorname{argmax}} \left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s - \theta_s^{i-1} \right) + Q_{t+1}^{i-1}(x_t^i) \right\} \\
&= \underset{d \in D}{\operatorname{argmax}} \left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s - \theta_s^{i-1} \right) \right\},
\end{aligned} \tag{4.5}$$

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 (4.3) and then cancelled the term that is independent of d since it is 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 [20] and [67] to explore non-linear value function approximations: The preceding discussion shows that allowing the value function to be non-linear makes it possible to include state feedback in the pricing policy.

4.2.2 Non-linear stochastic dual dynamic programming update

In contrast to the affine value function update above, the non-linear stochastic dual DP update generates non-linear value function approximations, which by the above discussion make it possible to include state feedback in the pricing policy. Similarly to [45] and [46], this update is computed in step 11 of Algorithm 4.1 as

$$Q_t^i \leftarrow \min\{H^*, Q_t^{i-1}\}, \tag{4.6}$$

where the minimum is taken pointwise and the so-called Lagrange dual cut H^* is defined as

$$H^*(x) := v^* - \langle \mu^*, x_{t+1}^i - x \rangle, \text{ for all } x \in X, \text{ and} \tag{4.7a}$$

$$\begin{aligned} v^* := \min_{\mu \in \mathcal{M}} \max_{d \in D, z \in X} \Bigg\{ &\lambda \sum_{s \in S} p_s(d) \left(r + d_s + Q_{t+1}^{i-1}(z + 1_s) \right) \\ &+ \left(1 - \lambda \sum_{s \in S} p_s(d) \right) Q_{t+1}^{i-1}(z) + \langle \mu, x_{t+1}^i - z \rangle \Bigg\}, \end{aligned} \tag{4.7b}$$

where μ^* is the minimiser of (4.7b). We now formally establish that the Lagrange dual cut from (4.7) generates an upper bound on the exact value function. Notice that this means that the entire approximate value function Q_t^i is an upper bound to the exact value function V_t if we assume that the initialiser Q_t^0 is an upper bound.

Using the unique fixed point of the DP for Q_t^0 at the initialising step, as discussed at the beginning of Section 4.2, satisfies this assumption.

Proposition 4.1. *Fix any $(i,t) \in (I \setminus \{0\}) \times T$. The Lagrange dual cut H^*, generated at a sample x_{t+1}^i, produces an upper bound to the exact value function, i.e.*

$$H^*(x) \geq V_t(x), \text{ for all } x \in X. \tag{4.8}$$

Proof. Note that in the base case at $i = 0$, we have $Q_t^i(x) \geq V_t(x)$ for all $(x,t) \in X \times T$, since Q_t^0 is initialised at the fixed point V^*, which is an upper bound to V_t. Suppose by means of an induction hypothesis that for some $(i,t) \in (I \setminus \{0\}) \times T$, $Q_{t+1}^{i-1}(x) \geq V_{t+1}(x)$ for all $x \in X$. Fix any $x \in X$ and any compact set $\mathcal{M} \subset \mathbb{R}^n$. Then we can upper bound the right-hand-side of the Bellman equation in (2.4) by introducing an auxiliary variable z and by replacing $V_{t+1}(y)$ by $Q_{t+1}^{i-1}(y)$ for all $y \in \{z + 1_s\}_{S \cup \{0\}}$, which yields

$$\begin{aligned}
V_t(x) \leq \max_{d \in D, z \in X} &\left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + V_{t+1}(z + 1_s)\right) \right. \\
&\left. + \left(1 - \lambda \sum_{s \in S} p_s(d)\right) V_{t+1}(z) + \langle \mu, x_{t+1}^i - z \rangle \right\} - \langle \mu, x_{t+1}^i - x \rangle \\
\leq \max_{d \in D, z \in X} &\left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + Q_{t+1}^{i-1}(z + 1_s)\right) \right. \\
&\left. + \left(1 - \lambda \sum_{s \in S} p_s(d)\right) Q_{t+1}^{i-1}(z) + \langle \mu, x_{t+1}^i - z \rangle \right\} - \langle \mu, x_{t+1}^i - x \rangle,
\end{aligned} \tag{4.9}$$

for all $\mu \in \mathcal{M}$. Note that the first inequality holds since we maximise over $z \in X$, but choosing $z = x \in X$, results in equality. To see this, notice that setting $z = x$ yields the original Bellman equation in (2.4) because the added and subtracted inner product terms $\langle \mu, x_{t+1}^i - z \rangle$ cancel. The second inequality in (4.9) holds due to the induction hypothesis that $Q_{t+1}^{i-1}(x) \geq V_{t+1}(x)$ for all $x \in X$. Since the above expression holds for all $\mu \in \mathcal{M}$, we can minimise over $\mu \in \mathcal{M}$ to arrive at

$$\begin{aligned}
V_t(x) \leq \min_{\mu \in \mathcal{M}} \max_{d \in D, z \in X} &\left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + Q_{t+1}^{i-1}(z + 1_s)\right) \right. \\
&\left. + \left(1 - \lambda \sum_{s \in S} p_s(d)\right) Q_{t+1}^{i-1}(z) + \langle \mu, x_{t+1}^i - z \rangle \right\} - \langle \mu^*, x_{t+1}^i - x \rangle \\
= v^* - \langle \mu^*, x_{t+1}^i - x \rangle = H^*(x) &
\end{aligned} \tag{4.10}$$

where $\mu^* \in \mathcal{M}$ denotes the minimiser of the maximised expression in curly brackets above, and where v^* as well as H^* are found from (4.7), as required. □

The proof of this proposition is based on a similar result in [46, Proposition 5], where an additional regularisation term is included in the approximate version of the forward problem in (4.2). The regularisation is needed for general multi-stage non-linear stochastic problems to guarantee that the approximate value function is Lipschitz continuous. For the specific revenue management problem under study, [29] have shown that the DP can be re-written as a so-called stochastic shortest path problem (see [37, Chapter 3]), where the Bellman operator $\mathcal{T}$ is known to be monotonic. Since the terminal condition of the DP and the fixed point are Lipschitz continuous functions, this implies that the value function has a Lipschitz continuous extension for all time steps $t \in T \cup \{\bar{t} + 1\}$ and that we do not need regularisation in the cut definition. Hence, our proof shows that the regularisation term can be dropped for the problem under study in this chapter, which simplifies the computation of the Lagrange dual cut H^* slightly.

From a control perspective, the benefit of having a non-linear value function approximation for the DP under study – in comparison with the affine value function approximation from Section 4.2.1 – comes at a different cost: The problem of finding the optimal cut coefficients μ in (4.7b) is a non-convex optimisation problem and there are consequently no guarantees that it can be solved to global optimality. For the particular form of the problem in this chapter, we can find the cut coefficients from a reformulation of the problem that results in a bi-concave objective function, which can be exploited to solve this problem as outlined in Appendix 4.A.1. Nevertheless, since global optimality is required to ensure that the approximate value function constitutes an upper bound to the exact value function in accordance with Proposition 4.1, we cannot guarantee that a cut is indeed an upper bound to the exact value function. We illustrate how this may result in computational problems in Section 4.3.

4.2.3 Gradient-bounded dynamic programming update

The gradient-bounded DP update is identical to the procedure introduced in Section 3.2, i.e. it follows steps 12–19 of Algorithm 3.1. The sample Algorithm 4.1 can be seen to be identical to Algorithm 3.1 by noting the equivalence between the DP formulations in (2.4) and (2.6). Note that gradient-bounded DP still relies on Assumptions 3.1 and 3.2, which restricts the set of applicable problems relative to the other two algorithms introduced in the previous two sections, which do not require these assumptions.

Similarly to the non-linear stochastic dual DP update from Section 4.2.2 and in contrast to the affine value function approximation update from Section 4.2.1, the approximate value function generated by the gradient-bounded DP update is non-linear as it is given by the pointwise minimum of affine functions in x. Assuming that the initialiser is an upper bound to the exact value function, which can be satisfied if we choose the fixed point for this purpose, the approximate value function is an upper bound to the exact value function (as discussed at the beginning of Section 4.2). Finally, the advantage of gradient-bounded DP over non-linear stochastic dual DP is that only convex optimisation problems need to be solved to compute the update, which makes gradient-bounded DP more resilient against computational stability problems than non-linear stochastic dual DP as shown in the next section.

4.3 Case study

In the following three sections, we present a numerical analysis that compares the three value function approximation algorithms stated and analysed in Section 4.2. To this end, we generate particular instances of the revenue management problem in attended home delivery presented in Section 2.2. We use the parameter values in [20] as a base case and modify these parameters to simulate various scenarios and conduct a sensitivity analysis. To quantify the performance of the algorithms with respect to their ability to generate profit, we use the probabilistic bounds on the expectation of the profit obtained under the approximately optimal policy, which we derived in Section 3.5.2.

In Section 4.3.1, we analyse the performance of the three algorithms under the assumption that the model parameters are known accurately. We then simulate how well the algorithms perform when they are trained on the data in 4.3.1, but being tested on scenarios, where the customer choice model parameters or the expected demand differs from the model in Section 4.3.2.

4.3.1 Exact model analysis

In this section, we adapt the numerical case study parameters from [20], which we have already used for the numerical examples in Sections 2.5.3 and 3.6, to a multi-scenario case study. The parameters that are fixed in all cases are stated in Table 4.1.

Table 4.1: Exact model parameters.

S	$\{1, 2, \ldots, 17\}$
λ	0.8
$\left[\underline{d}, \bar{d}\right]$	[£0, £10]
β_c, β_d	$-2.5087, -0.0766$
$\{\beta_s\}_{s \in S}$	$\{-1.0305, -0.3591, 0.3107, 0.5922, 0.6154, 0.0796, 0.5356, -0.2415,$ $-0.6286, -1.6736, -0.4351, -0.161, 0, 0.2533, 0.0736, 0.562, 0.2346\}$
r	£34.53

Furthermore, we use the same step sizes for the affine value function update (see (4.4) in Section 4.2.1) as in [20], i.e. $\alpha_1 := 0.0001, \alpha_2 := 0.00025, \alpha_3 := 0.00014$. In addition to these fixed parameters, we consider two more parameters that we will 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 an urban, suburban and rural setting. Each setting 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 may depend strongly on factors like infrastructure, traffic and weather conditions. However, for the purpose of our case study, we use a simplified model from [20, 21] that 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. 4.1 below.

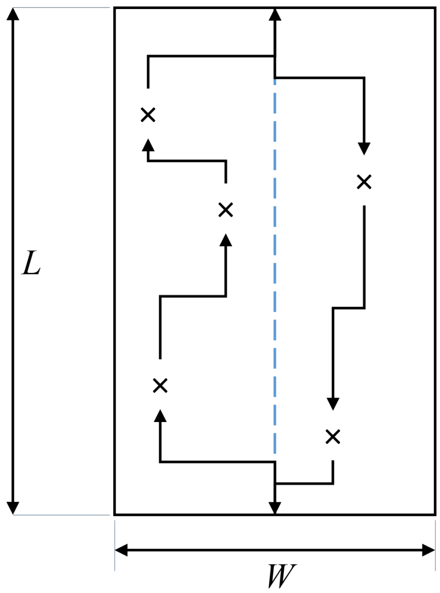


Figure 4.1: Delivery sub-area schematic with length L and width W. Crosses indicate customer locations served in a particular delivery time slot.

We further suppose an average delivery truck velocity of $\omega = 25$ mph and a cost per mile of $\kappa = £0.25$. 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. Since the average distance of two point sampled from the uniform distribution on $[0, 1]$ is $1/3$, this implies that the average distance travelled between two customers along the axis aligned with the width W of the sub-area is $1/3$ times the half-width $W/2$. This results in an expected variable delivery cost of

$$c_{\text{var}} := \kappa \times \frac{W}{6}, \tag{4.11}$$

as shown in [21], since the distance travelled along the axis aligned with the length L of the sub-area is always assumed to be $2L$, independent of the number of orders. 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 between L and W. The total travelling distance in every delivery time slot $s \in S$ thus becomes

$$\begin{aligned} \omega \times 1\mathrm{h} &= 2L + \bar{x}_s \frac{W}{6} \\ \Rightarrow W &= \frac{\omega}{4 + \bar{x}_s/6}. \end{aligned} \tag{4.12}$$

This finally implies that

$$c_{\mathrm{var}} := \frac{\kappa\omega}{24 + \bar{x}_s}. \tag{4.13}$$

Notice that c_{var} is independent of the particular slot, since we set the slot capacities equal for all slots, i.e. $\bar{x}_s = \bar{x}_{s'}$ for all $(s, s') \in S \times S$.

Second, we vary the expected demand, i.e. the expected number of customer arrivals. This quantity is given by $\lambda\bar{t}$. Since it is reasonable to keep $\lambda \approx 0.8$ for customer choice parameter estimation purposes (see [28]), 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, we find $\bar{t}$ as

$$\bar{t} \approx \frac{\phi n\bar{x}}{\lambda}, \tag{4.14}$$

for all $\phi \in \Phi$ and where the approximation comes from rounding $\bar{t}$ to the nearest integer. For all these 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 expectation bounds from Section 3.5.2.

In general, we observe that the non-linear stochastic dual DP algorithm produces higher expected profits than the affine value function approximation algorithm, while taking more time to compute a good solution. However, the gradient-bounded DP algorithm exhibits the strengths of both other algorithms: very similar profit generation performance to non-linear stochastic dual DP and similar speed to the affine value function approximation algorithm. For example, Fig. 4.2 below 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. Notice that this computation time is random, since the problem is stochastic. However, the curves in Fig. 4.2 are relatively smooth indicating a small variance around the mean. Non-linear stochastic dual DP always takes longest to compute out of the three algorithms. Computation time also tends to increase for non-linear stochastic dual DP as demand factor or slot capacity increase. For capacity 20, it takes about 4 times longer to compute the solution for demand factor 8, compared with the other two algorithms. This time factor increases to about 10 as we decease the demand factor to 1/8.

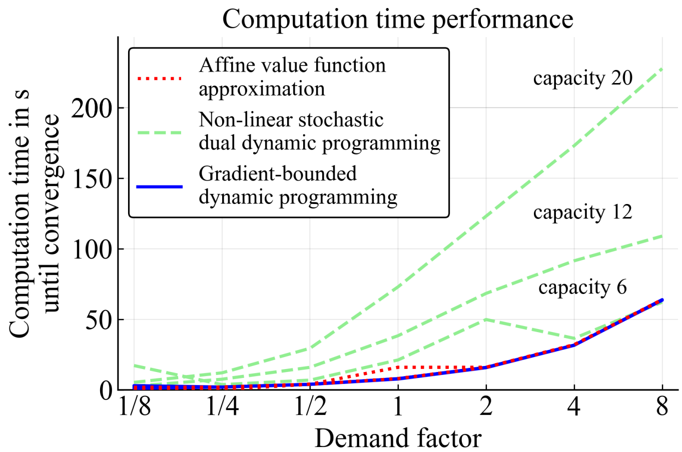


Figure 4.2: Computational time to reach 95% of the maximum expected profit with 99% confidence for each of the three algorithms against demand factor and delivery slot capacities.

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 capacities for these algorithms except in 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. One possible explanation for this 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 non-linear 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. 4.3 below. This might appear counter-intuitive at first, but is in line with our theoretical analysis from Section 4.2.2: We conjecture that this is due to the difficulty of finding global maxima of non-convex optimisation problems. If the algorithm converges to a local maximum, the value function approximation is no longer guaranteed to be an upper bound to the exact value function (see Proposition 4.1). Over time, this then leads to a compounding of errors caused by suboptimality, i.e. 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 that produces the maximum expected profit with 99% confidence. In the example of Fig. 4.3, 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, invalid cuts accumulate over time, which results in a degradation of the profit performance.

Comparing the expected profits obtained between the three algorithms, we observe that the gradient-bounded DP algorithm always 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 behaviour can be seen in Fig. 4.4 below, 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 the affine value function approximation algorithm. At the same time, gradient-bounded DP performs similarly to non-linear stochastic dual DP in most scenarios. However, gradient-bounded DP generates up to 10% more profit than non-linear stochastic dual DP for small demand factor 1/8 and capacity 6 as well as for large demand factor 8 across all slot capacities.

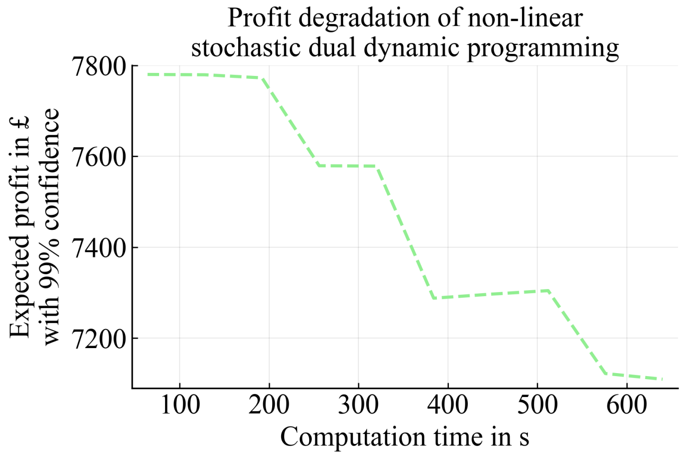


Figure 4.3: Expected profits of non-linear stochastic dual DP for demand factor 8 and slot capacity 12 decrease as more iterations are added over time.

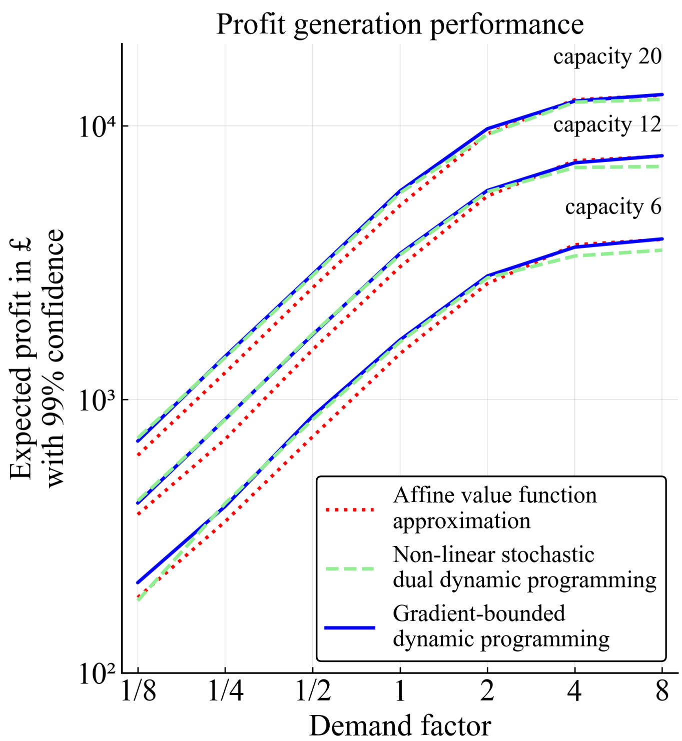


Figure 4.4: Expected profits with 99% confidence of the three algorithms against demand factor and for all delivery slot capacities.

Overall, we conclude that the gradient-bounded DP algorithm performs best in this experiment, where we assume exact knowledge of the model, i.e. exact knowledge of the customer arrival rate λ and the customer choice model parameters β_c, β_d and $\{\beta_s\}_{s\in S}$. Gradient-bounded DP outperforms the affine value function approximation algorithm in terms of profit generation while being similarly fast and at the same time, gradient-bounded DP is more than four times faster than non-linear stochastic dual DP while generating a very similar profit.

4.3.2 Parameter sensitivity analysis

We assumed in the previous section that the parameters, in particular the customer arrival rate λ and the customer choice model parameters β_c, β_d and $\{\beta_s\}_{s\in S}$, are known exactly, which is not the case in practice. Hence, we now investigate the performance of the three algorithms when we simulate how well the pricing policies obtained in the previous section perform on perturbed models.

Uncertain demand analysis

Consider the case that we derive value function approximations on the assumption that $\lambda = 0.8$, but the true value differs, i.e. more or less customers arrive on the booking website than anticipated. We model one scenario where demand is 25% lower, hence $\lambda = 0.6$, and a second scenario where demand is 25% higher, hence $\lambda = 1$. The resulting behaviour is very similar for all algorithms, as can be seen in Fig. 4.5 below.

The numerical results are almost identical to the ones obtained in the exact model analysis experiment in Section 4.3.1. The relative performance in terms of profit generation and computational time across all three algorithms is very similar. The only difference to the results from Section 4.3.1 are the absolute profit levels obtained for demand factors 2 and smaller, which scale proportionally with the customer arrival rate λ, as one would expect intuitively. For larger demand factors, the expected profit saturates across all $\lambda \in \{0.6, 0.8, 1\}$. This is because almost all slots are expected to sell out at the maximum delivery charge before the end of the time horizon. In conclusion, all three algorithms are equally capable of compensating for variations in the customer arrival rate and their relative performance remains unchanged in comparison with the exact model experiment in Section 4.3.1.

Uncertain customer choice parameters

Consider the case that customer preferences across slots and prices are misspecified. To model this, we now corrupt the parameter estimates β_c, β_d and $\{\beta_s\}_{s \in S}$ 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 [35, Section 8.6].

We consider three scenarios in which we vary the level of estimation error by setting the variance of the Gaussian to $\sigma^2 \in \{0.01, 0.1, 1\}$. With these three noise levels, we sample the sets of customer choice parameters, which we hold fixed for all validation runs in this experiment. Note that we do not have to worry about normalising the probability distribution, since the multinomial choice model is normalised for all possible parameter values. The numerical values used in our analysis are documented in Table 4.2 in Appendix 4.A.2. In practice, we could estimate the value of σ^2 from the data to test the robustness of candidate value function approximation algorithms with respect to uncertainty in the customer choice model. We document how the profit generation performance of the three algorithms degrades in comparison with the ideal scenario in the previous section in Fig. 4.6 at the end of this section.

As we see in Fig. 4.6(b) and (c), non-linear stochastic dual DP and gradient-bounded DP are both very robust against uncertainty in the customer choice model. Only for $\sigma^2 = 1$ is there a substantial degradation in profit generation performance. In contrast, Fig. 4.6(a) shows that even small uncertainties in the customer choice model have substantial negative impact on the profit generation performance of the affine value function approximation algorithm, decreasing expected profit with 99% confidence by about an order of magnitude for $\sigma^2 = 1$. We conjecture that this is due to the lack of state feedback in the affine value function approximation solution as detailed in Section 4.2.1. 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 $(x, t) \in X \times T$.

Overall, we conclude that both gradient-bounded DP and non-linear stochastic dual DP increase their relative profit-generation advantage over affine value function approximation when the parameter estimates of the customer choice model are not known exactly.

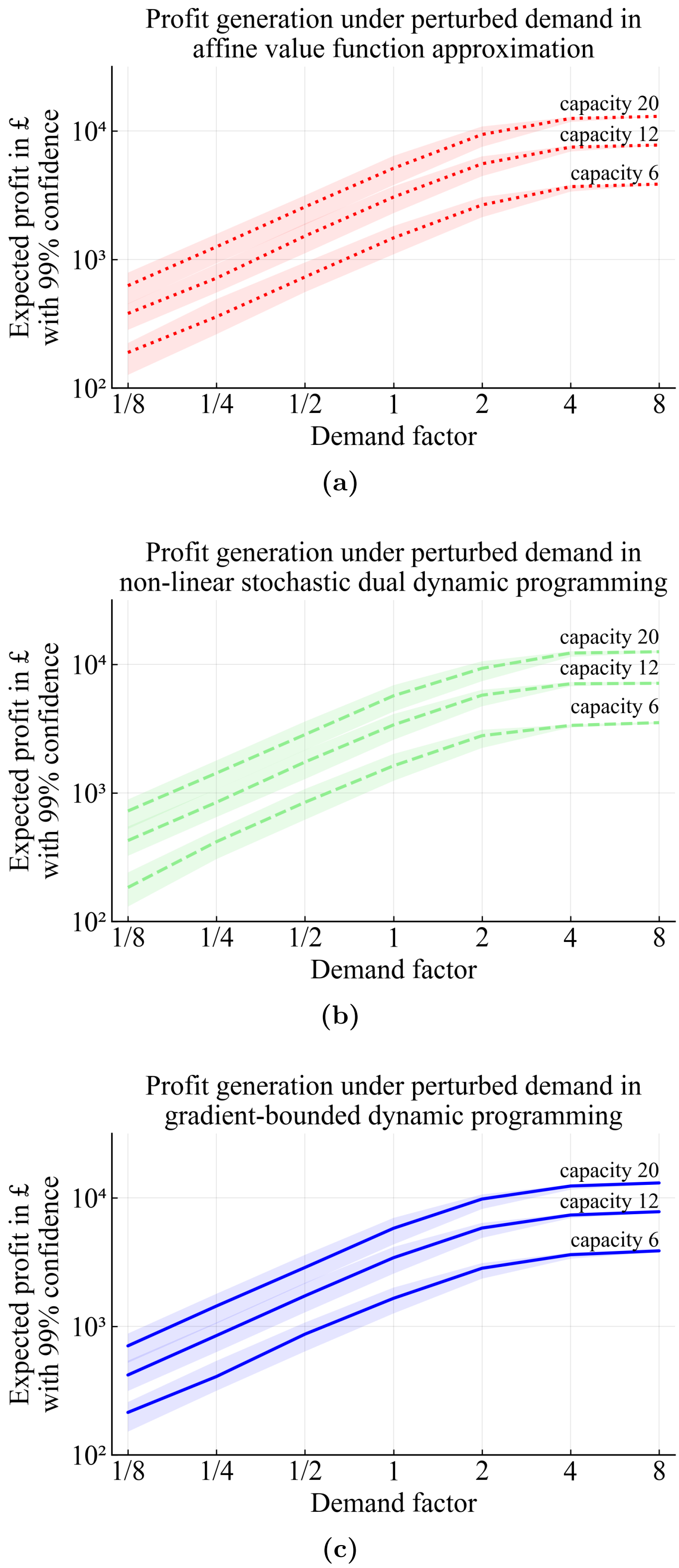
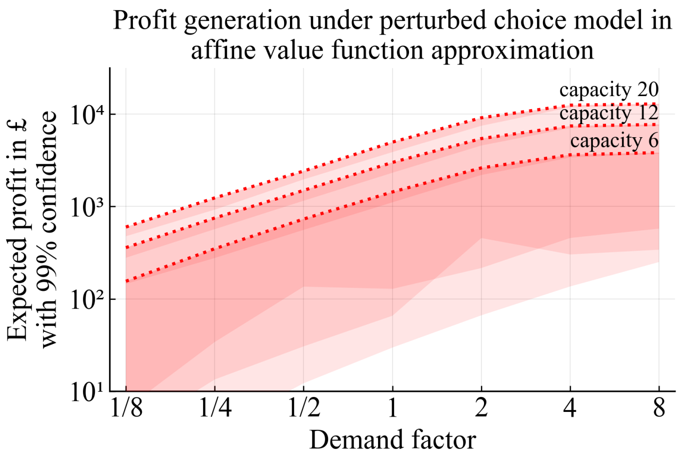
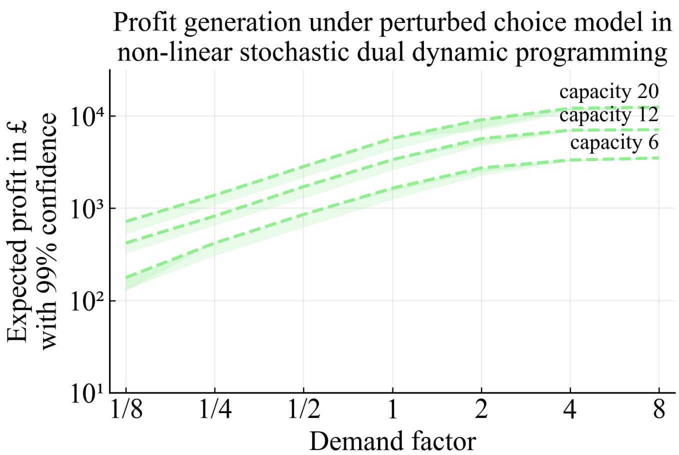


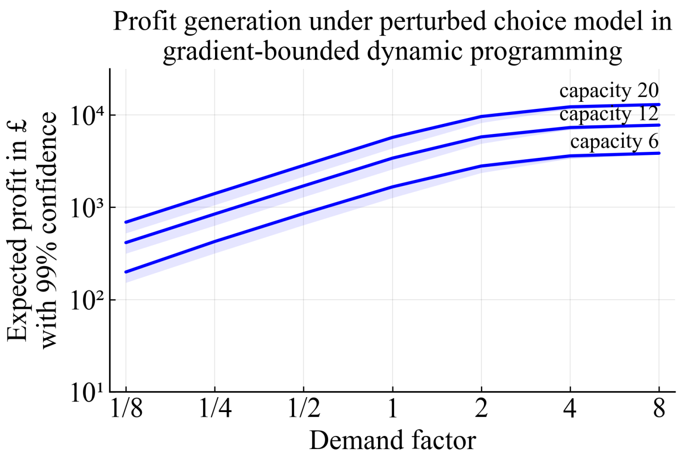
Figure 4.5: Expected profits with 99% confidence for affine value function approximation (a), non-linear stochastic dual DP (b) and gradient-bounded DP (c) under perturbed customer arrival rate. Shaded regions indicate how expected profit with 99% confidence increases when the customer arrival rate increases from 0.6 to 1. The lines correspond to an arrival rate of 0.8.



(a)



(b)



(c)

Figure 4.6: Expected profits with 99% confidence for affine value function approximation (a), non-linear stochastic dual DP (b) and gradient-bounded DP (c) under perturbed customer choice model parameters from Table 4.2. Lines indicate performance at $\sigma^2 = 0.01$ and shaded regions indicate performance as σ^2 increases to 0.1 and 1.

4.4 Conclusion

In this chapter, we analysed – theoretically and numerically – three approximate DP algorithms to find approximately optimal delivery slot prices in the revenue management problem in attended home delivery. From a control-theretical perspective, we identified limitations in the affine value function approximation algorithm and the non-linear stochastic dual DP algorithm. Through our numerical analysis, we showed how gradient-bounded DP can overcome these limitations. In our case study, we compared the performance of all three algorithms, i.e. profit-generation capabilities and computational time, in a number of scenarios. Overall, our numerical analysis shows that the gradient-bounded DP algorithm exhibits superior performance, since the affine value function approximation algorithm cannot reach its profit-generation capabilities and since the non-linear stochastic dual DP algorithm cannot match its computational speed and computational stability properties.

Possible directions for future work include investigating the numerical performance of these algorithms for other network revenue management problems and extending the promising gradient-bounded DP approach to other customer decision models than multinomial logit.

4.A Problem reformulation and table of parameters

4.A.1 Problem reformulation for non-linear stochastic dual dynamic programming

In this section, we show that we can re-write (4.7b), a non-convex optimisation problem, such that the resulting objective function is biconcave in two newly introduced variables as follows.

Equation (4.7b) depends on z, which is defined to be integer-valued. Since Q_{t+1}^{i-1} is the pointwise minimum of a finite number of affine functions in z, we can make the objective function in (4.7b) concave in a new variable $y \in \mathbb{R}^n$, which we restrict to be in the convex hull of the original state-space $\mathrm{conv}(X)$, i.e.

$$v^* = \min_{\mu \in \mathcal{M}} \max_{\substack{d \in D, \\ y \in \mathrm{conv}(X)}} \left\{ \lambda \sum_{s \in S} p_s(d) \left(r + d_s + Q_{t+1}^{i-1}(y + 1_s) \right) \right. \\ \left. + \left(1 - \lambda \sum_{s \in S} p_s(d) \right) Q_{t+1}^{i-1}(y) + \langle \mu, x_{t+1}^i - y \rangle \right\}. \tag{4.15}$$

This does not change the optimal solution since optimality can only be reached when $y = x_{t+1}^i$, such that $\langle \mu, x_{t+1}^i - y \rangle = 0$ for all $\mu \in \mathcal{M}$. Moreover, by a similar argument $z = x_{t+1}^i$ for optimality in (4.7b). Hence, introducing y does not change the overall solution. Furthermore, as shown in [25] and Section 2.4.2 of this dissertation, the maximisation over d can be expressed as a maximisation over transition probabilities, denoted by $p \in [0, 1)^n$, by inverting the function $p_s(d)$ for all $(x, s, d) \in X \times S \times D$. Writing $p_s(d)$ now as the elements p_s of the variable p and replacing d by the inversion of $p_s(d)$ as a function of p_s for all $s \in S$, results in an objective function that is concave in p, i.e.

$$v^* = \min_{\mu \in \mathcal{M}} \max_{\substack{p \in P, \\ y \in \mathrm{conv}(X)}} \left\{ \lambda \sum_{s \in S} p_s \left(r + \frac{1}{\beta_d} \left(\ln \left(\frac{p_s}{p_0} \right) - \beta_c - \beta_s \right) + Q_{t+1}^{i-1}(y + 1_s) \right) \right. \\ \left. + \left(1 - \lambda \sum_{s \in S} p_s \right) Q_{t+1}^{i-1}(y) + \langle \mu, x_{t+1}^i - y \rangle \right\} \tag{4.16}$$

$$= \min_{\mu \in \mathcal{M}} \max_{\substack{p \in \hat{P}, \\ y \in \mathrm{conv}(X)}} \left\{ \lambda \sum_{s \in F(x_{t+1}^i)} p_s \left(r + \frac{1}{\beta_d} \left(\ln \left(\frac{p_s}{p_0} \right) - \beta_c - \beta_s \right) + Q_{t+1}^{i-1}(y + 1_s) \right) \right.$$

$$+\left(1-\lambda\sum_{s\in F(x^i_{t+1})}p_s\right)Q^{i-1}_{t+1}(y)+\langle\mu,x^i_{t+1}-y\rangle\Bigg\}, \tag{4.17}$$

where we drop summations over unfeasible slots since they evaluate to 0.

Recall that $\hat{P} := \left\{[p_s(d)]_{s\in F(x)} \;\middle|\; d_s \in [\underline{d},\bar{d}] \quad \forall s\in F(x)\right\}$ is a convex set by Lemma 2.8. Notice that for any fixed $p\in\hat{P}$, the objective function is concave in y, because y only appears in Q^{i-1}_{t+1}, which is concave in y and in the affine term $\langle\mu, x^i_{t+1}-y\rangle$. Similarly to [25, theorem 1] and Appendix 2.A.1 of this dissertation, it can be shown that for any $y\in\operatorname{conv}(X)$, the objective function is concave in p. It follows that the resulting objective function is biconcave in (p,y) and that the constraints are convex in (p,y).

In general, this is a difficult problem. However, since the optimal solution is at $y=x^i_{t+1}$, we can fix y to this value and find the optimal value of p in the maximisation in (4.16). We then use these values for y and p to initialise a non-linear solver, which produces quite reliable results as shown in the numerical example in Section 4.3.

4.A.2 Table of parameters for Section 4.3.2

Table 4.2: Estimation error-corrupted customer choice parameters.

σ^2	β_c	β_d	$\{\beta_s\}_{s\in S}$
0.01	-2.5143	-0.0821	$\{-1.0232, -0.3546, 0.3127, 0.5860, 0.6211, 0.1056, 0.5179, -0.2563, -0.6204, -1.6762, -0.4313, -0.1621, -0.0037, 0.2453, 0.0732, 0.5590, 0.2192\}$
0.1	-2.5078	-0.1337	$\{-1.0866, -0.3378, 0.3152, 0.6729, 0.6802, 0.1164, 0.5365, -0.3083, -0.4778, -1.7180, -0.3431, -0.0674, -0.0291, 0.1821, 0.0411, 0.6579, 0.2281\}$
1	-2.6071	-1.2039	$\{-0.2242, -0.3338, -0.6643, 0.9198, -0.1940, -1.068, -0.5910, 0.9239, -1.0473, -1.8225, -0.6845, 1.8887, 0.7644, -1.4806, 0.4730, 0.2706, 1.1856\}$

5

Conclusions

Contents

5.1 Conclusions of this dissertation 111
5.1.1 Dynamic programming framework 111
5.1.2 Algorithmic developments 112
5.1.3 Comparative study 112
5.2 Directions for future research 113
5.2.1 Directions for future research on the dynamic programming framework 113
5.2.2 Directions for future research on algorithmic developments 114
5.2.3 Directions for future research on the comparative study 114

In this chapter, we summarise the main contributions of this dissertation and provide some directions for future research.

5.1 Conclusions of this dissertation

Following the main Chapters 2–4 of this dissertation, our key conclusions follow three themes, which we summarise in the next sections.

5.1.1 Dynamic programming framework

In Chapter 2, we analysed a dynamic programming framework for the revenue management problem in attended home delivery. Although the exact formulation is intractable for realistic problem sizes, the dynamic programming view of the problem allowed us to derive key mathematical characterisations of the underlying value function, which represents the expected profit-to-go for all states of orders and all time steps in the finite booking horizon. In particular, up to this time, the adoption of algorithms based on non-affine value function approximations has been *ad hoc*. Hence, in the wider context of this thesis, our findings provide theoretical support for the development of algorithms that leverage our mathematical characterisations of the value function.

5.1.2 Algorithmic developments

In Chapter 3, we developed a new approximate dynamic programming algorithm, termed *gradient-bounded dynamic programming*, for problems with submodular, concave extensible value functions. Theoretical and numerically verified convergence of the algorithm provide evidence for the effectiveness of the overall approximation strategy of tailoring a sample-based approximation algorithm to a dynamic program based on mathematical characterisations of its value function. Since the lower bound on the value accrued under the approximately optimal policy is random for gradient-bounded dynamic programming, we derive probabilistic bounds on the tail and the expectation of this quantity. Such probabilistic quantifications are also applicable to other algorithms whose performance criterion is random with a fixed but unknown probability distribution, such as in stochastic dual dynamic programming [47, 48] and in approximate dynamic programming with affine value function approximations [20].

5.1.3 Comparative study

In Chapter 4, we compared our gradient-bounded dynamic programming algorithm with two other approximate dynamic programming solutions from the literature. We analysed these algorithms from a control theoretic perspective and in a parameter sensitivity case study based on real-world data. We showed that gradient-bounded dynamic programming dominates the other algorithms in terms of both profit generation performance and computation time. Therefore, this constitutes theoretical and empirical evidence in support of the overarching research hypothesis outlined in Section 1.4 that more effective pricing policies can be derived by exploiting structural properties of the value function of the DP formulation of the revenue management problem in attended home delivery.

5.2 Directions for future research

Similarly to our conclusions, we suggest directions for future research, classified according to the themes of Chapters 2–4.

5.2.1 Directions for future research on the dynamic programming framework

Directions for future research on the dynamic programming framework include two main paths: First, the current results are mostly theoretical and do not offer a constructive way to verify if a particular set of problem parameters will produce a value function that is concave extensible and submodular. More specifically, our results state that there exists a small enough customer arrival rate for this to be the case, but the bound that we compute is loose and does not provide the means to check if the customer arrival rate is indeed small enough or not before checking this numerically. It would be of interest to determine the minimal set of assumptions needed to impose on the set of DP parameters for the value function to be concave extensible and submodular independent of the customer arrival rate.

Second, the setting for which our theoretical results hold is quite specific: We consider a multinomial customer choice model and supermodular expected delivery cost functions. Extending our research to more general discrete choice models (see e.g. [35]) and more general expected delivery cost functions could make the overall approach to revenue management described in this thesis applicable to a much larger set of revenue management problems.

5.2.2 Directions for future research on algorithmic developments

On the algorithmic side of research on revenue management in attended home delivery, we envisage the following directions for future research: First, it would be interesting to compare how concave extensibility and submodularity could be exploited to derive other approximation procedures. Our gradient-bounded DP algorithm refines the value function along a sample path, i.e. for any time step around one state and then it moves on to the next time step. This procedure is then repeated until convergence. An alternative to our approach could be to fix a particular time step in the horizon of the dynamic program, sample a large number of states and thus compute a very detailed approximate value function for this time step before moving on to the next time step (or previous time step if working backwards through the time horizon). This alternative approach would then result in solving a convex function fitting problem for every time step in the horizon of the dynamic program (see e.g. [69],[70]).

Second, our derived probabilistic bounds could be used to quantify the performance of other algorithms applied to other problems. The advantage of the proposed bounds is that they do not rely on any particular distribution of the random variable whose value they certify probabilistically. For example, this makes it immediately possible to analyse the performance of stochastic bounds in classical stochastic dual dynamic programming applications. There, it is often assumed that the random variable follows a normal distribution (see [48]), which may not even be true asymptotically in the limit as an increasing number of samples of the random variable are drawn, e.g. if the state or action space is discrete.

5.2.3 Directions for future research on the comparative study

Future research on comparative studies of gradient-bounded dynamic programming entails the following directions: First, one could further loosen the assumptions on the true customer choice model and delivery cost in comparison with the model that gradient-bounded DP uses to compute approximately optimal delivery slot prices. For example, one could consider more general discrete customer choice models than multinomial logit (see e.g. [35]). Since gradient-bounded DP uses state feedback, it might still produce reliable results even if the model is misspecified even more than in our parametric case study.

A second direction for future research entails exploring the relative performance of gradient-bounded DP for solutions of other network revenue management problems. Depending on the properties of the value function, gradient-bounded DP could produce at least a lower bound, which would serve as a benchmark for other algorithms, or possibly be a competitive solution on its own.

References

[1] R. Srikant. *The mathematics of Internet congestion control*. Springer Science & Business Media, 2012.

[2] J. Pfrommer et al. "Dynamic Vehicle Redistribution and Online Price Incentives in Shared Mobility Systems". In: *IEEE Transactions on Intelligent Transportation Systems* 15.4 (2014), pp. 1567–1578.

[3] E. Bitar et al. "Selling Random Wind". In: *2012 45th Hawaii International Conference on System Sciences*. 2012, pp. 1931–1937.

[4] K. T. Talluri and G. J. Van Ryzin. *The theory and practice of revenue management*. International Series in Operations Research & Management Science. Springer Science & Business Media, 2006.

[5] J. I. McGill and G. J. van Ryzin. "Revenue Management: Research Overview and Prospects". In: *Transportation Science* 33.2 (1999), pp. 233–256.

[6] N. Agatz et al. "Challenges and opportunities in attended home delivery". In: *The vehicle routing problem: Latest advances and new challenges*. Springer, 2008, pp. 379–396.

[7] Food Marketing Institute. *Digital shopper*. `https://www.fmi.org/digital-shopper/`. Accessed on 04/10/2019. 2018.

[8] N. Saunders. *Online Grocery & Food Shopping Statistics*. `https://www.onespace.com/blog/2018/08/online-grocery-food-shopping-statistics/`. Accessed on 16/06/2020. 2018.

[9] A. Belderok et al. *RETAILes: The Dutch grocery sector in 2030*. `https://www.rolandberger.com/publications/publication_pdf/roland_berger_online_grocery_shopping.pdf`. Accesssed on 16/06/2020. 2019.

[10] R. Redman. *Online grocery sales to grow 40% in 2020*. `https://www.supermarketnews.com/online-retail/online-grocery-sales-grow-40-2020`. Accessed on 16/06/2020. 2020.

[11] N. Galante, E. G. López, and S. Monroe. "The future of online grocery in Europe". In: *McKinsey & Company* (2013). Available: `https://www.mckinsey.com/~/media/McKinsey/Industries/Retail/Our%20Insights/The%20future%20of%20online%20grocery%20in%20Europe/The_future_of_online_grocery.ashx`, pp. 22–31.

[12] A. Bensoussan. *Dynamic Programming and Inventory Control*. Studies in probability, optimization, and statistics. IOS Press, 2011.

[13] P. Wanke. "A conceptual framework for inventory management: focusing on low-consumption items". In: *Production and Inventory Management Journal* 49.1 (2014), pp. 9–26.

[14] G. B. Dantzig and J. H. Ramser. “The Truck Dispatching Problem”. In: *Management Science* 6.1 (1959), pp. 80–91.

[15] K. Menger. “Das Botenproblem”. In: *Ergebnisse eines Mathematischen Kolloquiums*. Ed. by K. Menger. Vol. 2. Leipzig: Teubner, 1932, pp. 11–12.

[16] P. Toth and D. Vigo. *Vehicle Routing*. Ed. by D. Vigo and P. Toth. `https://doi.org/10.1137/1.9781611973594`. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2014.

[17] O. Bräysy and M. Gendreau. “Vehicle Routing Problem with Time Windows, Part I: Route Construction and Local Search Algorithms”. In: *Transportation Science* 39.1 (2005), pp. 104–118.

[18] O. Bräysy and M. Gendreau. “Vehicle Routing Problem With Time Windows, Part II: Metaheuristics”. In: *Transportation Science* 39.1 (2005), pp. 119–139.

[19] N. A. El-Sherbeny. “Vehicle routing with time windows: An overview of exact, heuristic and metaheuristic methods”. In: *Journal of King Saud University – Science* 22.3 (2010), pp. 123–131.

[20] X. Yang and A. K. Strauss. “An approximate dynamic programming approach to attended home delivery management”. In: *European Journal of Operational Research* 263.3 (2017), pp. 935–945.

[21] C. F. Daganzo. “Modeling Distribution Problems with Time Windows: Part I”. In: *Transportation Science* 21.3 (1987), pp. 171–179.

[22] A. K. Strauss, R. Klein, and C. Steinhardt. “A review of choice-based revenue management: Theory and methods”. In: *European Journal of Operational Research* (2018).

[23] K. Asdemir, V. S. Jacob, and R. Krishnan. “Dynamic pricing of multiple home delivery options”. In: *European Journal of Operational Research* 196.1 (2009), pp. 246–257.

[24] M. Ulmer. “Dynamic Pricing for Same-Day Delivery Routing”. Unpublished. Available: `https://www.researchgate.net/publication/317278950`. 2017.

[25] L. Dong, P. Kouvelis, and Z. Tian. “Dynamic Pricing and Inventory Control of Substitute Products”. In: *Manufacturing & Service Operations Management* 11.2 (2009), pp. 317–339.

[26] A. M. Campbell and M. Savelsbergh. “Incentive Schemes for Attended Home Delivery Services”. In: *Transportation Science* 40.3 (2006), pp. 327–341.

[27] J. Hartline, V. Mirrokni, and M. Sundararajan. “Optimal Marketing Strategies over Social Networks”. In: *Proceedings of the 17th International Conference on World Wide Web*. WWW ’08. Beijing, China: ACM, 2008, pp. 189–198.

[28] X. Yang et al. “Choice-Based Demand Management and Vehicle Routing in E-Fulfillment”. In: *Transportation Science* 50.2 (2016), pp. 473–488.

[29] 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: *18th European Control Conference, ECC 2019, Naples, Italy, June 25-28, 2019*. 2019, pp. 999–1004.

[30] D. Lebedev, P. Goulart, and K. Margellos. “A dynamic programming framework for optimal delivery time slot pricing”. In: *European Journal of Operational Research* 292.2 (2021), pp. 456–468.

[31] D. Lebedev, P. Goulart, and K. Margellos. “Gradient-Bounded Dynamic Programming with Submodular and Concave Extensible Value Functions”. In: *21st IFAC World Congress 2020*. Available: `https://arxiv.org/pdf/2005.11213.pdf`. 2020.

[32] D. Lebedev, P. Goulart, and K. Margellos. *Gradient-Bounded Dynamic Programming for Submodular and Concave Extensible Value Functions with Probabilistic Performance Guarantees*. Tech. rep. Submitted for peer review at Automatica. Available: `https://arxiv.org/pdf/2006.02910.pdf`. 2020.

[33] D. Lebedev, K. Margellos, and P. Goulart. *Approximate Dynamic Programming for Delivery Time Slot Pricing: a Sensitivity Analysis*. Tech. rep. Submitted for peer review at IEEE Transactions on Control Systems Technology. Available: `https://arxiv.org/pdf/2008.00780.pdf`. 2020.

[34] M. Suh and G. Aydin. “Dynamic pricing of substitutable products with limited inventories under logit demand”. In: *IIE Transactions* 43.5 (2011), pp. 323–331.

[35] K. E. Train. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press, 2009.

[36] D. McFadden. “Conditional Logit Analysis of Qualitative Choice Behaviour”. In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York, NY, USA: Academic Press New York, 1973, pp. 105–142.

[37] D. P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. II*. 4th. Athena Scientific, 2012.

[38] Y. Akçay, H. P. Natarajan, and S. H. Xu. “Joint Dynamic Pricing of Multiple Perishable Products Under Consumer Choice”. In: *Management Science* 56.8 (2010), pp. 1345–1361.

[39] M. Chen and Z.-L. Chen. “Recent Developments in Dynamic Pricing Research: Multiple Products, Competition, and Limited Demand Information”. In: *Production and Operations Management* 24.5 (2015), pp. 704–731.

[40] K. Murota and A. Shioura. “Relationship of M-/L-convex functions with discrete convex functions by Miller and Favati–Tardella”. In: *Discrete Applied Mathematics* 115.1 (2001). First Japanese-Hungarian Symposium for Discrete Mathematics and its Applications, pp. 151–176.

[41] D. Gale, H. W. Kuhn, and A. W. Tucker. “Linear programming and the theory of games”. In: *Activity Analysis of Production and Allocation*. Ed. by T. C. Koopmans. Vol. 13. Cowles Commission for Research in Economics Monographs. John Wiley & Sons, 1951. Chap. 19, pp. 317–335.

[42] D. Bertsimas and R. Weismantel. *Optimization over Integers*. Jan. 2005.

[43] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Heidelberg, Berlin, New York: Springer Verlag, 1998.

[44] S. Boyd and L. Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[45] J. Zou, S. Ahmed, and X. A. Sun. “Stochastic dual dynamic integer programming”. In: *Mathematical Programming* 175.1 (May 2019), pp. 461–502.

[46] S. Zhang and X. A. Sun. *Stochastic Dual Dynamic Programming for Multistage Stochastic Mixed-Integer Nonlinear Optimization*. Tech. rep. Available: `https://arxiv.org/pdf/1912.13278.pdf`. 2019.

[47] M. V. F. Pereira and L. M. V. G. Pinto. “Multi-stage stochastic optimization applied to energy planning”. In: *Mathematical Programming* 52.1 (May 1991), pp. 359–375.

[48] A. Shapiro. “Analysis of stochastic dual dynamic programming method”. In: *European Journal of Operational Research* 209.1 (Feb. 2011), pp. 63–72.

[49] W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics)*. New York, NY, USA: Wiley-Interscience, 2007.

[50] D. P. de Farias and B. V. Roy. “The Linear Programming Approach to Approximate Dynamic Programming”. In: *Operations Research* 51.6 (2003). `https://doi.org/10.1287/opre.51.6.850.24925`, pp. 850–865.

[51] D. P. de Farias and B. V. Roy. “On Constraint Sampling in the Linear Programming Approach to Approximate Dynamic Programming”. In: *Mathematics of Operations Research* 29.3 (2004), pp. 462–478.

[52] P. N. Beuchat, A. Georghiou, and J. Lygeros. “Performance Guarantees for Model-Based Approximate Dynamic Programming in Continuous Spaces”. In: *IEEE Transactions on Automatic Control* 65.1 (2020), pp. 143–158.

[53] J. Warrington, P. N. Beuchat, and J. Lygeros. “Generalized Dual Dynamic Programming for Infinite Horizon Problems in Continuous State and Action Spaces”. In: *IEEE Transactions on Automatic Control* 64.12 (2019), pp. 5012–5023.

[54] A. Georghiou, A. Tsoukalas, and W. Wiesemann. “Robust dual dynamic programming”. In: *Operations Research* 67.3 (2019), pp. 813–830.

[55] D. Zhang and D. Adelman. “An Approximate Dynamic Programming Approach to Network Revenue Management with Customer Choice”. In: *Transportation Science* 43.3 (2009), pp. 381–394.

[56] J. Meissner and A. Strauss. “Network revenue management with inventory-sensitive bid prices and customer choice”. In: *European Journal of Operational Research* 216.2 (2012), pp. 459–468.

[57] A. Sauré et al. “Dynamic multi-appointment patient scheduling for radiation therapy”. In: *European Journal of Operational Research* 223.2 (2012), pp. 573–584.

[58] F. P. Cantelli. “Sui confini della probabilita”. In: *Atti del Congresso Internazional del Matematici, Bologna*. Vol. 6. 1928, pp. 47–59.

[59] P. Massart. “The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality”. In: *Ann. Probab.* 18.3 (July 1990), pp. 1269–1283.

[60] A. Maurer and M. Pontil. *Empirical Bernstein Bounds and Sample Variance Penalization*. Tech. rep. `https://arxiv.org/pdf/0907.3740.pdf`. July 2009.

[61] J. Bezanson et al. “Julia: A Fresh Approach to Numerical Computing”. In: *SIAM Review* 59.1 (2017), pp. 65–98.

[62] B. Hajek. *Random Processes for Engineers.* Cambridge University Press, 2015.

[63] V. Leclère et al. "Exact Converging Bounds for Stochastic Dual Dynamic Programming via Fenchel Duality". In: *SIAM J. Optim.* 30 (2020), pp. 1223–1250.

[64] B. K. Ghosh. "Probability Inequalities Related to Markov's Theorem". In: *The American Statistician* 56.3 (2002), pp. 186–190.

[65] M. J. Evans and J. S. Rosenthal. *Probability and statistics: The science of uncertainty.* Macmillan, 2004.

[66] W. Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables". In: *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30.

[67] S. Koch and R. Klein. "Route-based approximate dynamic programming for dynamic pricing in attended home delivery". In: *European Journal of Operational Research* (2020).

[68] A. Beck. *First-Order Methods in Optimization.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 2017.

[69] L. A. Hannah and D. B. Dunson. *Bayesian nonparametric multivariate convex regression.* Tech. rep. Available: `https://arxiv.org/abs/1109.0322`. 2011.

[70] L. A. Hannah and D. B. Dunson. "Multivariate Convex Regression with Adaptive Partitioning". In: *J. Mach. Learn. Res.* 14.1 (2013), pp. 3261–3294.