THE UNIVERSITY OF CHICAGO

ESSAYS ON APPLIED OPTIMIZATION MODELS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE UNIVERSITY OF CHICAGO
BOOTH SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
ZUGUANG GAO

CHICAGO, ILLINOIS
JUNE 2023

To my parents, for their love and support.

The music is not in the notes, but in the silence between.

— Unknown, often attributed to Mozart

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, my deepest gratitude goes to Professor John Birge, who has always been incredibly supportive and encouraging throughout this journey. I am truly privileged to have the opportunity to learn stochastic optimization from his course and his book, and later conduct research under his guidance. His great wisdom and sharp intuition have shaped my research significantly. His advising has been in a perfect balance that leaves me the freedom with what to pursue while also provides me with his critical insights and points me to the right direction whenever I needed. This dissertation would not have been possible without him being my advisor.

Not only has Professor Birge distinguished himself as an accomplished academic, but also as a person who upholds the highest level of integrity, honesty, and kindness - ethics he does not usually preach, but always practice. Throughout the years, I have also learned the lifelong lesson from him to be a righteous person, both in academia and in all aspects of life. He is the example that I aspire to follow in the years to come.

I would also like to express my sincere appreciation to Professor Varun Gupta, who has not only taught me to develop my own research vision but also inspired me to aim high and develop my own criteria for what constitutes good research, rather than solely relying on published results. I am also grateful for his open-door policy, which allowed me to walk into his office and find him anytime to bounce off ideas and discuss any concerns regarding my research or personal life.

I cannot forget to thank Professor Tamer Başar at my alma mater, University of Illinois Urbana-Champaign. My research career started with him, and I would never have achieved what I have today without his continued support. Memories of old times often emerge in my mind when I look south from Chicago, and it is always fulfilling to know that the Coordinated Science Laboratory is just 2 hours away. Throughout my PhD, I have kept him updated with my progress, and he has provided me with proper encouragement and valuable

suggestions. In the times when I felt lost in research or in life, my Illinois connection through his research group has saved me and helped me back on track. Words cannot describe how much I appreciate his guidance over the years.

There are also many mentors and coauthors that I am forever in debt to. Dr. Khaled Alshehri has always been generous with his time to have many hour-long discussions with me. Prof. Xudong Chen has been a great mentor and friend, who helped me tremendously in jumpstarting my research. Dr. Qianqian Ma has provided me with a great deal of support. Prof. Nur Sunar has been a respected mentor and collaborator, guiding me in various aspects of research. It has been an honor to work with these great minds.

I would also like to thank many other faculty members at Booth: Professors René Caldentey, Levi DeValve, Rad Niazadeh and Amy Ward. They have provided me with valuable comments and insightful feedback on my research and presentations.

I would also like to express my gratitude to the managers and mentors I had during my internships: Richard Chen and Maurice Cheung at Flexport, as well as Adam Schultz and Cem Randa at Uber. They have offered me the opportunity to learn and implement real world problems, and helped me understand how academia and industry can inspire each other.

Life in Chicago has been enriched by many fellows at Chicago: Deniz Akturk, Amir Alwan, Lisa Hillas, Ebru Kaşıkaralar, Çağla Keçeli, Robert Montgomery (Monty), Gizem Yılmaz, and many others. Each of them has shared with me part of this journey, and I am delighted to have known them and been friends with them. I would also thank my long-time friends outside of Chicago: Guochao Sun, Xiaobin Gao, and Yi Ding. They have been incredibly caring, supportive, and have lifted me up in every way possible.

I would like to convey my deep gratitude to the musicians at the Chicago Symphony Orchestra and Chicago Symphony Chorus, led by Maestro Riccardo Muti, as well as the numerous concerto soloists. I have always said that when I leave Chicago, one thing I will

miss the most is the sound of this city. I will never forget the endless wonder from the concert hall, brought by these talented musicians. I also want to express my appreciation to Johann Sebastian Bach, Franz Joseph Haydn, Wolfgang Amadeus Mozart, Ludwig van Beethoven, among others, for the peace and joy they have brought to me in those hard-working days and nights, through their eternal masterpieces.

Finally, this dissertation is dedicated to my mother, Wen Zhao, and my father, Suowen Gao. Over the past 30 years, they have worked tirelessly to help me grow in every dimension. Their unconditional love and unwavering support have shaped me into the person I am today, and I am incredibly grateful to have them as my parents. Life is short, and I will make every effort to spend as much time as possible with them in the coming years.

ABSTRACT

This dissertation studies several different optimization problems, and consists of five chapters. In Chapter 1, we consider a multiperiod binary Knapsack problem, along with several extensions. We propose fully polynomial time approximation schemes to these problems where applicable. We also prove the performance guarantee of some greedy algorithms, and propose parameterized approximation algorithms for some extensions. In Chapter 2, we propose several models to aggregate the distributed energy resources, where the aggregator can be profit-seeking or regulated. We design the two-part pricing mechanism for the aggregator to achieve full market efficiency. In Chapter 3, we analyze the sample complexity of decentralized Q-learning algorithms for stochastic games, in both the tabular case and the case with linear function approximation. In Chapter 4, we design a power purchase agreement (PPA), where a firm signs a long term contract with a renewable energy generator. The contract specifies a one-time transfer payment by the firm to the renewable energy generator, as an investment to build new renewable energy facilities. The firm then owns all the generation from these facilities for an extended period of time. We formulate the firm's decision on when to sign the PPA as an optimal stopping problem, and analyze the firm's optimal policies. Chapter 5 is motivated by an application in freight forwarding, where we formulate the freight forwarder's decisions on the assignment of shipments to containers as an integer program, which turns out to be a combination of the bin packing problem and the generalized assignment problem. We propose several heuristics for this problem and run numerical experiments on their performances.

In the following, we provide more detailed abstracts for each chapter of this dissertation.

Chapter 1 studies the approximation schemes of multi-period Knapsack problems, and is based on Gao et al. [90]. An instance of the multiperiod binary Knapsack problem (MPBKP) is given by a horizon length T, a non-decreasing vector of knapsack sizes $(c_1, \ldots, c_T)$ where c_t denotes the cumulative size for periods $1, \ldots, t$, and a list of n items. Each item is a triple

(r, q, d) where r denotes the reward or value of the item, q its size, and d denotes its time index (or, deadline). The goal is to choose, for each deadline t, which items to include to maximize the total reward, subject to the constraints that for all $t = 1, \ldots, T$, the total size of selected items with deadlines at most t does not exceed the cumulative capacity of the knapsack up to time t. We also consider the multiperiod binary knapsack problem with soft capacity constraints (MPBKP-S) where the capacity constraints are allowed to be violated by paying a penalty that is linear in the violation. The goal of MPBKP-S is to maximize the total profit, which is the total reward of the selected items less the total penalty. Finally, we consider the multiperiod binary knapsack problem with soft stochastic capacity constraints (MPBKP-SS), where the non-decreasing vector of knapsack sizes $(c_1, \ldots, c_T)$ follow some arbitrary joint distribution but we are given access to the profit as an oracle, and we must choose a subset of items to maximize the total expected profit, which is the total reward less the total expected penalty.

For MPBKP, we exhibit a fully polynomial-time approximation scheme that achieves $(1+\epsilon)$ approximation with runtime $\tilde{\mathcal{O}}\left(\min\left\{n + \frac{T^{3.25}}{\epsilon^{2.25}}, n + \frac{T^2}{\epsilon^3}, \frac{nT}{\epsilon^2}, \frac{n^2}{\epsilon}\right\}\right)$; for MPBKP-S, the $(1+\epsilon)$ approximation can be achieved in $\mathcal{O}\left(\frac{n \log n}{\epsilon} \cdot \min\left\{\frac{T}{\epsilon}, n\right\}\right)$. To the best of our knowledge, our algorithms are the first FPTAS for any multiperiod version of the Knapsack problem since its study began in 1980s. For MPBKP-SS, we prove that a natural greedy algorithm is a 2-approximation when items have the same size. We also provide parameterized approximation algorithms for MPBKP-S and MPBKP-SS. Our algorithms also provide insights on how other multiperiod versions of the knapsack problem may be approximated.

Chapter 2 studies the aggregation of distributed energy resources, and is based on Gao et al. [88, 89, 92]. The rapid expansion of distributed energy resources (DERs) is one of the most significant changes to electricity systems around the world. Examples of DERs include solar panels, electric storage, thermal storage, combined heat and power plants, etc. Due to the small supply capacities of these DERs, it is impractical for them to participate directly

in the wholesale electricity market. We study in this chapter the question of how to integrate these DER supplies into the electricity market, with the objective of achieving full market efficiency. Specifically, we study three aggregation models, where there is an aggregator who procures electricity from DERs, and sells them in the wholesale market. In the first aggregation model, a profit-maximizing aggregator announces a differential two-part pricing policy to the DER owners. We show that this model preserves full market efficiency, i.e., the social welfare achieved by the aggregation model is the same as that when DERs participate directly in the wholesale market. In the second aggregation model, the profit-seeking aggregator is forced to impose a uniform two-part pricing policy to prosumers from the same location, and we numerically show the efficiency loss of this model. In the third aggregation model, a uniform two-part pricing policy is applied to DER owners, while the aggregator becomes fully regulated but is guaranteed positive profit. It is shown that this third model again achieves full market efficiency. Furthermore, we show that DER aggregation also leads to a reduction on the market power of conventional generators. DER aggregation via profit-seeking and/or regulated aggregators have been investigated by CAISO and NYISO, among others, and the recent FERC Order No. 2222 paved the way for aggregators to bid in the wholesale market. Our efficient aggregation models may settle the debate on how DERs should be included in the wholesale electricity market.

Chapter 3 studies the sample complexity of decentralized Q-learning algorithms for stochastic games, and is based on Gao et al. [91, 94]. Learning in stochastic games is arguably the most standard and fundamental setting in multi-agent reinforcement learning (MARL). In this chapter, we consider decentralized MARL in stochastic games in the non-asymptotic regime. In particular, we establish the finite-sample complexity of fully decentralized Q-learning algorithms in a significant class of general-sum stochastic games (SGs) – weakly acyclic SGs, which includes the common cooperative MARL setting with an identical reward to all agents (a Markov team problem) as a special case. We focus on the

practical while challenging setting of *fully decentralized* MARL, where neither the rewards nor the actions of other agents can be observed by each agent. In fact, each agent is completely oblivious to the presence of other decision makers. Both the tabular and the linear function approximation cases have been considered. In the tabular setting, we analyze the sample complexity for the decentralized Q-learning algorithm in Arslan and Yüksel (2016) to converge to a Markov perfect equilibrium (Nash equilibrium). With linear function approximation, the results are for convergence to a linear approximated equilibrium – a new notion of equilibrium that we propose – which describes that each agent's policy is a best reply (to other agents) within a linear space. Numerical experiments are also provided for both settings to demonstrate the results.

Chapter 4 studies the design of a power purchase agreement (PPA) where the firm agrees to make a certain transfer payment to the renewable generator, and the generator invests that payment to build new renewable energy facilities. The firm will then have access to all electricity generation from the new facilities for a long-term period. The firm may dynamically decide when to start the PPA on an ongoing basis, based on the evolving market conditions, and the transfer payment (amount of investment) is also specified by the firm. The firm's objective is to maximize its long-term discounted benefit (total savings) from signing the PPA. We mathematically formulate the firm's decision problem as an optimal stopping problem and provide analytical solutions. We also provide insights on how the firm's investment capacity, expected saving, and the total new generation due to the PPA change with respect to different problem parameters.

Chapter 5 defines and studies the (ocean) freight consolidation problem (FCP), which plays a crucial role in solving today's supply chain crisis. Roughly speaking, every day and every hour, a freight forwarder sees a set of shipments and a set of containers at the origin port. There is a shipment cost associated with assigning each shipment to each container. If a container is assigned any shipment, there is also a procurement cost for that container.

The FCP aims to minimize the total cost of fulfilling all the shipments, subject to capacity constraints of the containers. In this chapter, we show that no constant factor approximation exists for FCP, and propose a series of greedy based heuristics for solving the problem. We also test our heuristics with simulated data and show that our heuristics achieve small optimality gaps. This chapter is based on Gao et al. [93].

CHAPTER 1

APPROXIMATION ALGORITHMS FOR MULTIPERIOD BINARY KNAPSACK PROBLEMS

1.1 Introduction

Knapsack problems are a classical category of combinatorial optimization problems, and have been studied for more than a century (Mathews [158]). They have found wide applications in various fields (Kellerer et al. [127]), such as selection of investments and portfolios, selection of assets, finding the least wasteful way to cut raw materials, etc. One of the most commonly studied problem is the so-called *0-1 knapsack problem*, where a set of n items are given, each with a reward and a size, and the goal is to select a subset of these items to maximize the total reward, subject to the constraint that the total size may not exceed some knapsack capacity. It is well-known that the 0-1 knapsack problem is NP-complete. However, the problem was shown to possess *fully polynomial-time approximation schemes (FPTAS's)*, i.e., there are algorithms that achieve $(1+\epsilon)$ factor of the optimal value for any $\epsilon \in (0,1)$, and take polynomial time in n and $1/\epsilon$.

In this chapter, we study three extensions of the 0-1 knapsack problem. First, we consider a multiperiod version of the 0-1 knapsack problem, which we call the *multiperiod binary knapsack problem (MPBKP)*. There is a horizon length T and a vector of knapsack sizes $(c_1, \ldots, c_T)$, where c_t is the cumulative size for periods $1, \ldots, t$ and is non-decreasing in t. We are also given a list of n items, each associated with a triple (r, q, d) where r denotes the reward or value of the item, q its size, and d denotes its time index (or, deadline). The goal is to choose a reward maximizing set of items to include such that for any $t = 1, \ldots, T$, the total size of selected items with deadlines at most t does not exceed the cumulative capacity of the knapsack up to time t. The application that motivates this problem is a seller who produces $(c_t - c_{t-1})$ units of a good in time period t, and can store unsold goods for selling

later. The seller is offered a set of bids, where each bid includes a price (r), a quantity demanded (q), and a time at which this quantity is needed. The problem of deciding the revenue maximizing subset of bids to accept is exactly MPBKP.

The second extension we consider is the *multiperiod binary knapsack problem with soft capacity constraints (MPBKP-S)* where at each period the capacity constraint is allowed to be violated by paying a penalty that is linear in the violation. The goal of MPBKP-S is then to maximize the total profit, which is the total reward of the selected items less the total penalty. In this case, the seller can procure goods from outside at a certain rate if its supply is not enough to fulfill the bids it accepts, and wants to maximize its profit.

The third extension we consider is the *multiperiod binary knapsack problem with soft stochastic capacity constraints (MPBKP-SS)* where the non-decreasing vector of knapsack sizes $(c_1, \ldots, c_T)$ follows some arbitrary joint distribution given as the set of sample paths of the possible realizations and their probabilities. We select the items *before* realizations of any of these random incremental capacities to maximize the total *expected* profit, which is the total reward of selected items less the total expected penalty. In this case, the production of the seller at each time is random, but it has to select a subset of bids before realizing its supply. Again, the seller can procure capacity from outside at a certain rate if its realized supply is not enough to fulfill the bids it accepts, and wants to maximize its expected profit.

1.1.1 Literature Review

The first published FPTAS for the 0-1 knapsack problem was due to Ibarra and Kim [113] where they achieve a time complexity $\widetilde{\mathcal{O}}\left(n + (1/\epsilon^4)\right)$ by dividing the items into a class of "large" items and a class of "small" items. The problem is first solved for large items only, using the dynamic program approach, with rewards rounded down using some discretization quantum (chosen in advance), and the small items are added later. Lawler [135] proposed a more nuanced discretization method to improve the polylogarithmic factors. Since then,

improvements have been made on the dynamic program for large items, where Kellerer and Pferschy [126] proposed an algorithm with runtime $\tilde{\mathcal{O}}\left(n + 1/\epsilon^3\right)$, and Rhee [180] achieved $\tilde{\mathcal{O}}\left(n + 1/\epsilon^{2.5}\right)$ with a randomized algorithm. Most recently, Chan [46] proposed a deterministic algorithm, achieving the same $\tilde{\mathcal{O}}\left(n + 1/\epsilon^{2.5}\right)$ runtime, which has subsequently been improved to $\tilde{\mathcal{O}}\left(n + (1/\epsilon)^{9/4}\right)$ in Jin [121].

We note that MPBKP is also related to a number of other multiperiod versions of the knapsack problem in literature. The multiperiod knapsack problem (MPKP) proposed by Faaland [73] has the same structure as MPBKP, except that in Faaland [73], each item can be repeated multiple times, i.e., the decision variables for each item is not binary, but any nonnegative integer (in the single-period case, this is called the unbounded knapsack problem (Andonov et al. [13])). To the best of our knowledge, there has been no further studies on MPKP since Faaland [73]. In the multiple knapsack problem (MKP), there are m knapsacks, each with a different capacity, and items can be inserted to any knapsacks (subject to its capacity constraints). MKP is a special case of the generalized assignment problem (GAP), where each item has different reward and size when being put into different knapsacks. Shmoys and Tardos [192] proposed a 2-approximation for GAP. Chekuri and Khanna [48] proved that GAP is APX-hard and does not admit an FPTAS. Later, Jansen [119] proved that MKP admits an efficient polynomial time approximation scheme (EPTAS), with runtime depending polynomially on n but exponentially on $1/\epsilon$.

The incremental knapsack problem (IKP) is another multiperiod version of the knapsack problem (Hartline and Sharp [103]), where the knapsack capacity increases over time, and each selected item generates a reward on every period after its insertion, but this reward is discounted over time. Unlike MPBKP, items do not have deadlines and can be selected anytime throughout the T periods. When the discount factors are 1, it is called the time invariant incremental knapsack problem (IIKP). A PTAS for IIKP is proposed in Bienstock et al. [37] under the assumption that $T = \mathcal{O}\left(\sqrt{\log n}\right)$, and it has been shown that IIKP

is strongly NP-hard. Later, Faenza and Malinovic [78] proposed the first PTAS for IIKP regardless of T, and Della Croce et al. [64] proposed an PTAS for IKP when T is a constant. Most recent developments of IKP include Aouad and Segev [18], Faenza et al. [79]. Other similar problems and/or further extensions include the multiple-choice multiperiod knapsack problem (Randeniya [179], Lin and Wu [145], Lin and Chen [144]), the multiperiod multi-dimensional knapsack problem (Lau and Lim [133]), the multiperiod precedence-constrained knapsack problem (Moreno et al. [163], Samavati et al. [185]), to name a few.

1.1.2 Our Contributions

Our main contributions of this chapter are two-fold. First, from the perspective of model formulation, we propose the MPBKP and its generalized versions MPBKP-S and MPBKP-SS. Despite the fact that there are a number of multiperiod/multiple versions of knapsack problems, including those mentioned above (many of which are strongly NP-hard), the MPBKP and MPBKP-S we proposed here are the first to admit an FPTAS among any multiperiod versions of the classical knapsack problem since their initiation back in 1980s. With these results, it is thus interesting to see where the boundary lies between these multiperiod problems that admit an FPTAS and those problems that do not admit an FPTAS. Second, the algorithms we propose for both MPBKP and MPBKP-S are generalized from the ideas of solving 0-1 knapsack problems, but with nontrivial modifications as we will address in the remaining of this chapter. For MPBKP-S and MPBKP-SS, we also propose parameterized approximation algorithms, under the assumption that the (expected) total penalty on the optimal solution is no greater than β-fraction of the total reward. For MPBKP-SS, we also propose a greedy algorithm that achieves 2-approximation for the special case when all items have the same size. The comparison of the performance of greedy algorithms for this special case is also provided. Our results are summarized in Table 1.1 and Table 1.2.

Approximation	MPBKP	MPBKP-S	MPBKP-SS
$1+\epsilon$	$\tilde{\mathcal{O}}\left(\min\left\{n+\frac{T^{3.25}}{\epsilon^{2.25}}, n+\frac{T^2}{\epsilon^3}, \frac{nT}{\epsilon^2}, \frac{n^2}{\epsilon}\right\}\right)$	$\mathcal{O}\left(\frac{n\log n}{\epsilon}\cdot\min\left\{\frac{T}{\epsilon}, n\right\}\right)$	-
$1+\frac{\epsilon}{1-\beta}$	-	$\tilde{\mathcal{O}}\left(n+\frac{T^{3.25}}{\epsilon^{2.25}}\right)$	$\tilde{\mathcal{O}}\left(n+\frac{1}{\epsilon^T}\right)$

Table 1.1: Summary of runtime results for three multiperiod Knapsack problems

Algorithm	MPBKP, $q_i = 1$	MPBKP-S, $q_i = 1$	MPBKP-SS, $q_i = 1$
Greedy	Optimal	Optimal	2-approximation

Table 1.2: Summary of approximation results of greedy algorithms on special cases

1.1.3 Organization

The rest of this chapter is organized as follows. In Section 1.2 we formally write the three problems in mathematical programming form. Two FPTAS for MPBKP is proposed in Section 1.3. An FPTAS for MPBKP-S, as well as a parameterized approximation for MPBKP-S, are proposed in Section 1.4.1. An alternative FPTAS for MPBKP-S is provided in Appendix 1.7.4. A greedy algorithm for a special case of MPBKP-SS, as well as a parameterized approximation for MPBKP-SS, are proposed in Section 1.5. Some other results on special cases of these problems are supplemented in Appendix 1.7.5. We also provide pseudo-polynomial time algorithms in Appendix 1.7.6. Most proofs are left to Appendices 1.7.1, 1.7.2, and 1.7.3, but we provide proof ideas in the main body.

1.2 Problem Formulation and Main Results

In this section, we formally introduce the Multiperiod Binary Knapsack Problem (MPBKP), as well as the two generalized versions of it, namely, the (deterministic) Multiperiod Binary Knapsack Problem with Soft capacity constraints (MPBKP-S), and the Multiperiod Binary Knapsack Problem with Soft Stochastic capacity constraints (MPBKP-SS).

1.2.1 Multiperiod Binary Knapsack Problem (MPBKP)

An instance of MPBKP is given by a set of n items, each associated with a triple (r_i, q_i, d_i), and a sequence of knapsack capacities $\{c_1, \ldots, c_T\}$. For each item i, we get reward r_i if and only if i is included in the knapsack by time d_i. We assume that $r_i \in \mathbb{N}$, $q_i \in \mathbb{N}$ and $d_i \in [T] := \{1, \ldots, T\}$. The knapsack capacity at time t is c_t, and by convention $c_0 = 0$. The MPBKP can be written in the integer program (IP) form:

$$\max_x z = \sum_{i=1}^{n} r_i x_i \tag{1.1a}$$

$$\text{s.t.} \sum_{j:d_j \le t} q_j x_j \le c_t, \quad \forall t = 1, \ldots, T \tag{1.1b}$$

$$x_i \in \{0, 1\}, \quad \forall i = 1, \ldots, n \tag{1.1c}$$

where x_i's are binary decision variables, i.e., x_i is 1 if item i is included in the knapsack and is 0 otherwise. In (1.1), we aim to pick a subset of items to maximize the objective function, which is the total reward of picked items, subject to the constraints that by each time t, the total size of picked items with deadlines up to t does not exceed the knapsack capacity at time t, which is c_t. For each $t \in [T]$, let $\mathcal{I}(t) := \{i \in [n] \mid d_i = t\}$ denote the set of items with deadline t. Note that without loss of generality, we may assume that $\mathcal{I}(t) \neq \emptyset, \forall t$ and $c_t > 0$. We further note that the decision variables x_i's in (1.1) are binary, but if we relax this to any nonnegative integers, the problem becomes the so-called multiperiod knapsack problem (MPKP) as in Faaland [73]. Indeed, we can write (1.1) equivalently as

$$\max_{x \in \{0,1\}^n} z = \sum_{j \in \mathcal{I}(1)} r_j x_j + \sum_{j \in \mathcal{I}(2)} r_j x_j + \cdots + \sum_{j \in \mathcal{I}(T)} r_j x_j \tag{1.2a}$$

$$\text{s.t.} \sum_{j \in \mathcal{I}(1)} q_j x_j \le c_1 \tag{1.2b}$$

$$\sum_{j\in\mathcal{I}(1)} q_j x_j + \sum_{j\in\mathcal{I}(2)} q_j x_j \le c_2 \tag{1.2c}$$

$$\vdots \qquad\qquad \vdots \tag{1.2d}$$

$$\sum_{j\in\mathcal{I}(1)} q_j x_j + \sum_{j\in\mathcal{I}(2)} q_j x_j + \cdots + \sum_{j\in\mathcal{I}(T)} q_j x_j \le c_T \tag{1.2e}$$

$$x_j \in \{0,1\}, \quad \forall j \in [n], \tag{1.2f}$$

and if the decision variables are changed from binary to all nonnegative integers in (1.2), the problem becomes exactly the same as the multiperiod knapsack problem (MPKP) in [73].

Our first main result is the following theorem on MPBKP.

Theorem 1.1. *An FPTAS exists for MPBKP. Specifically, there exists a deterministic algorithm that achieves* $(1+\epsilon)$*-approximation in* $\tilde{\mathcal{O}}\left(\min\left\{n+\frac{T^{3.25}}{\epsilon^{2.25}}, n+\frac{T^2}{\epsilon^3}, \frac{nT}{\epsilon^2}, \frac{n^2}{\epsilon}\right\}\right)$.

We note that when $T=1$, MPBKP (1.1) becomes the traditional 0-1 Knapsack problem. In Section 1.3, we will provide an approximation algorithm for MPBKP with runtime $\tilde{\mathcal{O}}\left(n+\frac{T^{3.25}}{\epsilon^{2.25}}\right)$, as well as an alternative algorithm with runtime $\tilde{\mathcal{O}}\left(n+\frac{T^2}{\epsilon^3}\right)$. As we will see shortly, MPBKP can be viewed as a special case of MPBKP-S. In Section 1.4.1, we will provide an approximation algorithms for MPBKP-S with runtime $\tilde{\mathcal{O}}\left(\frac{nT}{\epsilon^2}\right)$, while another approximation algorithm with runtime $\tilde{\mathcal{O}}\left(\frac{n^2}{\epsilon}\right)$ is provided in Appendix 1.7.4, both of which are also applicable to MPBKP. A pseudo-polynomial time algorithm which achieves the exact optimal solution is also provided in Appendix 1.7.6.1. For the special case that $q_i = 1$ for all $i \in [n]$, the greedy algorithm also achieves exact optimality (see Appendix 1.7.5.1).

1.2.2 Multiperiod Binary Knapsack Problem with Soft Capacity Constraints (MPBKP-S)

In MPBKP-S, the capacity constraints in (1.1) no longer exist, i.e., the total size of selected items at each time step is allowed to be greater than the total capacity up to that time,

however, there is a penalty rate $B_t \in \mathbb{N}$ for each unit of overflow at period t. We assume that $B_t > \max_{i\in[n]:d_i\le t} \frac{r_i}{q_i}$ to avoid trivial cases (any item with $\frac{r_i}{q_i} \ge B_t$ and $d_i \le t$ will always be added to generate more profit). In the IP form, MPBKP-S can be written as

$$\begin{aligned}\max_{x\in\{0,1\}^n} z := \sum_{i=1}^{n} r_i x_i - \Bigg\{ & B_1 \cdot \Big[\sum_{j\in\mathcal{I}(1)} q_j x_j - c_1 \Big]^+ \\ & + B_2 \cdot \Big[\sum_{j\in\mathcal{I}(2)} q_j x_j - \Big(c_1 - \sum_{j\in\mathcal{I}(1)} q_j x_j\Big)^+ - (c_2 - c_1) \Big]^+ + \cdots \Bigg\}\end{aligned} \tag{1.3}$$

where $[a]^+$ is the maximum of a and 0. In the objective function, $\sum_{j\in\mathcal{I}(1)} q_j x_j$ is the total size of selected items with deadline 1, and c_1 is the capacity for time 1, thus $B_1 \cdot \Big[\sum_{j\in\mathcal{I}(1)} q_j x_j - c_1\Big]^+$ is the penalty generated at time 1. Similarly, $\sum_{j\in\mathcal{I}(2)} q_j x_j$ is the total size of selected items with deadline 2, $c_2 - c_1$ is the incremental capacity from time 1 to time 2, and $\Big(c_1 - \sum_{j\in\mathcal{I}(1)} q_j x_j\Big)^+$ is the leftover capacity (if any) carried from time 1, thus $B_2 \cdot \Big[\sum_{j\in\mathcal{I}(2)} q_j x_j - \Big(c_1 - \sum_{j\in\mathcal{I}(1)} q_j x_j\Big)^+ - (c_2 - c_1)\Big]^+$ is the penalty generated at time 2. We continue this pattern and write out the penalties generated at each time.

An equivalent expression of (1.3) is the following.

$$\begin{aligned}\max_{x\in\{0,1\}^n} z(x) := & \sum_{i=1}^{n} r_i x_i \\ & - \sum_{t=1}^{T} B_t \left[\sum_{j\in\mathcal{I}(t)} q_j x_j - \max_{0\le t'<t} \left\{ c_t - c_{t'} - \sum_{j\in\mathcal{S}:t'+1\le d_j\le t-1} q_j x_j \right\} \right]^+ .\end{aligned} \tag{1.4}$$

Further, if we add decision variables $y_t, t = 1, \ldots, T$, which represents the overflow at time t, and let $a_t := c_t - c_{t-1}$ be the incremental capacity at time t, then the problem can be written as

$$\max_{x,y} \sum_{i\in[n]} r_i x_i - \sum_{t=1}^{T} B_t y_t \tag{1.5a}$$

$$\text{s.t.} \sum_{i\in\mathcal{I}(1)\cup\cdots\cup\mathcal{I}(t)} q_i x_i - \sum_{s=1}^{t} y_s \le \sum_{s=1}^{t} a_t = c_t, \quad \forall t : 1 \le t \le T \tag{1.5b}$$

$$x_i \in \{0,1\}, \quad y_t \ge 0. \tag{1.5c}$$

The objective is to choose a subset of the n items to maximize the total profit, which is the sum of the rewards of the selected items deducted by the sum of penalty paid at each period, and the constraints enforce that the total size of accepted items by the end of each period must not exceed the sum of the cumulative capacity and the units of overflow.

We will consider approximation algorithms for MPBKP-S. Moreover, for those "good" instances, we will also provide a parameterized approximation algorithm. Specifically, we consider those instances that satisfy the following assumption.

Assumption 1.1. *In the optimal solution of MPBKP-S, the total penalty is at most β fraction of the total rewards, i.e., $\sum_{t=1}^{T} B_t y_t^*(\omega) \le \beta \cdot \sum_{i\in[n]} r_i x_i^*$, where x^* and y^* constitute the optimal solution of* (1.5).

Our second main result is the following theorem on MPBKP-S. The theorem has two parts: the first part asserts an approximation algorithm for general instances of MPBKP-S, while the second part claims a parameterized algorithm under Assumption 1.1.

Theorem 1.2. *We have the following algorithms for MPBKP-S.*

1. *An FPTAS exists for MPBKP-S. Specifically, there exists an algorithm which achieves $(1+\epsilon)$-approximation in $\mathcal{O}\left(\frac{n\log n}{\epsilon} \cdot \min\left\{\frac{T}{\epsilon}, n\right\}\right)$.*
2. *Under Assumption 1.1, there is an algorithm that achieves $\left(1+\frac{\epsilon}{1-\beta}\right)$-approximation in $\tilde{\mathcal{O}}\left(n + \frac{T^{3.25}}{\epsilon^{2.25}}\right)$.*

In Section 1.4.1, we will present an $\mathcal{O}\left(\frac{nT\log n}{\epsilon^2}\right)$ approximation algorithm for solving MPBKP-S. An alternative FPTAS with runtime $\mathcal{O}\left(\frac{n^2\log n}{\epsilon}\right)$ is provided in Appendix 1.7.4. In Section 1.4.2, we will present the parameterized approximation algorithm with $\left(1+\frac{\epsilon}{1-\beta}\right)$-approximation factor and $\tilde{\mathcal{O}}\left(n+\frac{T^{3.25}}{\epsilon^{2.25}}\right)$ runtime. For the ease of presentation, our algorithms and analysis are presented for the case $B_t = B$, but they can be generalized to the heterogeneous $\{B_1, \ldots, B_T\}$ in a straightforward manner. It is worth noting that the algorithm for MPBKP that we introduce in Section 1.3 does not extend to MPBKP-S, and we will make this clear in the beginning of Section 1.4.1. A pseudo-polynomial time algorithm which achieves the exact optimal solution is also provided in Appendix 1.7.6.2. For the special case that $q_i = 1$ for all $i \in [n]$, we prove that the greedy algorithm achieves exact optimality (see Appendix 1.7.5.2).

1.2.3 Multiperiod Binary Knapsack Problem with Soft Stochastic Capacity Constraints (MPBKP-SS)

The MPBKP-SS formulation is similar to (1.5), except that the vector of knapsack sizes $(c_1, \ldots, c_T)$ follows some arbitrary joint distribution given to the algorithm as the set of possible sample path (realization) of knapsack sizes and the probability of each sample path. We use ω to index sample paths which we denote by $\{c_t(\omega)\}$, $p(\omega)$ as the probability of sample path ω, and Ω as the set of possible sample paths. The goal is to pick a subset of items before the realization of ω so as to maximize the expected total profit, which is the sum of the rewards of the selected items deducted by the total (expected) penalty. For a sample $\omega \in \Omega$ let $y_t(\omega)$ be the overflow at time t. Then, we can write the problem as:

$$\max_{x,y} \sum_{i\in[n]} r_i x_i - \mathbb{E}_\omega \left[B_t \cdot \sum_{t=1}^{T} y_t(\omega) \right] \tag{1.6a}$$

$$\text{s.t.} \sum_{i\in\mathcal{I}(1)\cup\cdots\cup\mathcal{I}(t)} q_i x_i - \sum_{s=1}^{t} y_s(\omega) \le c_t(\omega), \quad \forall \omega \in \Omega, 1 \le t \le T \tag{1.6b}$$

$$x_i \in \{0,1\}, \quad y_t \ge 0. \tag{1.6c}$$

Furthermore, we make the following assumption.

Assumption 1.2. *In the optimal solution of MPBKP-SS, the total expected penalty is at most β fraction of the total rewards, i.e., $\mathbb{E}_\omega \left[\sum_{t=1}^{T} B_t y_t^*(\omega)\right] \le \beta \cdot \sum_{i\in[n]} r_i x_i^*$, where x^* and y^* constitute the optimal solution of (1.6).*

Our third main result is the following theorem on MPBKP-SS. The theorem has two parts: the first part asserts a greedy algorithm for the special case when all items are of unit size, and the second part claims a parameterized approximation algorithm for the general MPBKP-SS.

Theorem 1.3. *We have the following algorithms.*

1. *If $q_i = q$ for all $i \in [n]$, then, there exists a greedy algorithm with runtime $\mathcal{O}\left(n^2 T|\Omega|\right)$ that achieves 2-approximation for MPBKP-SS.*

2. *Under Assumption 1.2, there is an algorithm to MPBKP-SS that achieves $\left(1+\frac{\epsilon}{1-\beta}\right)$-approximation in $\tilde{\mathcal{O}}\left(n + 1/\epsilon^T\right)$.*

In Section 1.5.1, we will present a 2-approximation greedy algorithm for solving MPBKP-SS in the special case that all items have size 1. In Section 1.5.2, we will address the difficulties of the general MPBKP-SS, and present a parameterized approximation algorithm with $\left(1+\frac{\epsilon}{1-\beta}\right)$-approximation factor and $\tilde{\mathcal{O}}\left(n + \frac{1}{\epsilon^T}\right)$ runtime. For the special case that $T = 1$, i.e., a single period knapsack problem with stochastic capacity, we provide an FPTAS with runtime $\mathcal{O}\left(\frac{n^2 \log n}{\epsilon}\right)$ (see Appendix 1.7.5.3).

1.3 Approximation Algorithms for MPBKP

In the following two subsections, we provide two approximation algorithms for solving MPBKP. The first one is a $(1+\epsilon)$-approximation algorithm with runtime $\tilde{\mathcal{O}}\left(n + T^{3.25}/\epsilon^{2.25}\right)$, and the second one gives the same approximation factor with runtime $\tilde{\mathcal{O}}\left(n + T^2/\epsilon^3\right)$.

1.3.1 A $(1+\epsilon)$-approximation in $\tilde{\mathcal{O}}\left(n + T^{3.25}/\epsilon^{2.25}\right)$

In this subsection, we provide an FPTAS for the MPBKP, which has time complexity $\tilde{\mathcal{O}}\left(n + \frac{T^{3.25}}{\epsilon^{2.25}}\right)$. We will apply the "functional approach" as used in Chan [46]. The main idea is to use the results on function approximations (Chan [46], Jin [121]) as building blocks – for each period we approximate one function that gives, for every choice of available capacity, the maximum reward obtainable by selecting items in that period. We then combine "truncated" version of these functions using $(\max, +)$-convolution. This idea, despite its simplicity, allows us to obtain an FPTAS for MPBKP. Such a result should not be taken as granted – as we will see in the next section, this method does not apply for MPBKP-S, even though it is just a slight generalization of MPBKP.

We begin with some preliminary definitions and notations. For a given set of item rewards and sizes, $\mathcal{I} = \{(r_1, q_1), \ldots, (r_{n'}, q_{n'})\}$, define the function

$$f_{\mathcal{I}}(c) := \max_{x_1,\ldots,x_{n'}} \left\{ \sum_{i\in\mathcal{I}} r_i x_i \;:\; \sum_{i\in\mathcal{I}} q_i x_i \le c,\; x_1,\ldots,x_{n'} \in \{0,1\} \right\} \tag{1.7}$$

for all $c \ge 0$, and $f_{\mathcal{I}}(c) := -\infty$ for $c < 0$. The function $f_{\mathcal{I}}$ is a nondecreasing step function, and the number of steps is called the *complexity* of that function. Further, for any $\mathcal{I} = \mathcal{I}_1 \sqcup \mathcal{I}_2$, i.e., $\mathcal{I}$ being a disjoint union of $\mathcal{I}_1$ and $\mathcal{I}_2$, we have that $f_{\mathcal{I}} = f_{\mathcal{I}_1} \oplus f_{\mathcal{I}_2}$, where $\oplus$ denotes the $(\max, +)$-*convolution*: $(f \oplus g)(c) = \max_{c'\in\mathbb{R}} \left(f(c') + g(c - c')\right)$.

We define the *truncated function* $f_{\mathcal{I}}^{c'}$ as follows:

$$f_{\mathcal{I}}^{c'}(c) = \begin{cases} f_{\mathcal{I}}(c) & c \le c', \\ -\infty & c > c'. \end{cases} \tag{1.8}$$

Recall that we denote the set of items with deadline t by $\mathcal{I}(t)$. We next define the function f_t as follows:

$$f_t := \begin{cases} f_{\mathcal{I}(1)}^{c_1} & t = 1, \\ \left(f_{t-1} \oplus f_{\mathcal{I}(t)}\right)^{c_t} & t \ge 2. \end{cases} \tag{1.9}$$

In other words, f_t's are defined recursively: for $t = 1$, let $f_1 := f_{\mathcal{I}(1)}^{c_1}$; for $t \ge 2$, we define $f_t = \left(f_{t-1} \oplus f_{\mathcal{I}(t)}\right)^{c_t}$. Each function value of $f_t(c)$ corresponds to a feasible, in fact an optimal, solution x for items with deadline at most t, as shown in the following proposition.

Proposition 1.1. *Let x^* be the optimal solution for MPBKP* (1.1)*. We have that the optimal value of* (1.1)*, $\sum_{i\in[n]} r_i x_i^*$, satisfies $\sum_{i\in[n]} r_i x_i^* = f_T(c_T)$.*

Proposition 1.1 implies that, in order to obtain an approximately optimal solution for MPBKP (1.1), it is sufficient to have a good approximation for the function

$$f_T = \left(\cdots\left(\left(f_{\mathcal{I}(1)}^{c_1} \oplus f_{\mathcal{I}(2)}\right)^{c_2} \oplus f_{\mathcal{I}(3)}\right)^{c_3} \cdots \oplus f_{\mathcal{I}(T)}\right)^{c_T}. \tag{1.10}$$

We say that a function $\tilde{f}$ approximates the nonnegative function f with factor $1+\epsilon$ if $\tilde{f}(c) \le f(c) \le (1+\epsilon)\tilde{f}(c)$ for all $c \in \mathbb{R}$. It should be clear that if $\tilde{f}$ approximates f with factor $1+\epsilon$ and $\tilde{g}$ approximates g with factor $1+\epsilon$, then $\tilde{f} \oplus \tilde{g}$ approximates $f \oplus g$ with factor $1+\epsilon$. We then introduce the following result from Jin [121] for the 0-1 Knapsack problem.

Lemma 1.1 (Jin [121])**.** *Given a set $\mathcal{I} = \{(r_1, q_1), \ldots, (r_n, q_n)\}$, we can obtain $\tilde{f}_{\mathcal{I}}$ that approximates $f_{\mathcal{I}}$ (defined in* (1.7)*) with factor $1+\epsilon$ and complexity $\tilde{O}\left(\frac{1}{\epsilon}\right)$ in $\tilde{O}\left(n + (1/\epsilon)^{2.25}\right)$.*

With the above lemma, we present Algorithm 1.1 for MPBKP.

Algorithm 1.1 FPTAS for MPBKP in $\tilde{\mathcal{O}}\left(n + T^{3.25}/\epsilon^{2.25}\right)$

Input: $\epsilon, [n], c_1, \ldots, c_T$ ▷ Set of items to be packed, cumulative capacities
Output: $\tilde{f}_t$ ▷ Approximation of function f_t

Discard all items with $r_i \le \frac{\epsilon}{n} \max_j r_j$ and relabel the items
$r_0 \leftarrow \min_i r_i$ ▷ Lower bound of solution value
$m \leftarrow \left\lceil \log_{1+\epsilon} \frac{n^2}{\epsilon} \right\rceil$ ▷ Number of distinct rewards to be considered, each in the form $r_0 \cdot (1+\epsilon)^k$
Obtain $\tilde{f}_{\mathcal{I}(1)}$ that approximates $f_{\mathcal{I}(1)}$ with factor $(1+\epsilon)$ using Lemma 1.1
$\tilde{f}_1 := \tilde{f}_{\mathcal{I}(1)}^{c_1}$ ▷ $\tilde{f}_1$ has complexity at most $m = \tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$
for $t = 2, \ldots, T$ **do**
    Obtain $\tilde{f}_{\mathcal{I}(t)}$ that approximates $f_{\mathcal{I}(t)}$ with factor $(1+\epsilon)$ using Lemma 1.1
    $l \leftarrow$ complexity of $\tilde{f}_{\mathcal{I}(t)}$ ▷ $l = \tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$
    Compute (all breakpoints and their values of) $\hat{f}_t := \left(\tilde{f}_{t-1} \oplus \tilde{f}_{\mathcal{I}(t)}\right)^{c_t}$, taking $m \cdot l$ time
    ▷ $\hat{f}_t$ has complexity $\tilde{\mathcal{O}}\left(\frac{1}{\epsilon^2}\right)$
    $\tilde{f}_t := r_0 \cdot (1+\epsilon)^{\left\lfloor \log_{1+\epsilon}\left(\frac{\hat{f}_t}{r_0}\right)\right\rfloor}$ ▷ Round $\hat{f}_t$ down to the nearest $r_0 \cdot (1+\epsilon)^k$ for $k = 0, \ldots, m$.
    ▷ Now $\tilde{f}_t$ has complexity at most $m = \tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$
end for

We now describe the intuition behind Algorithm 1.1. We first discard all items with reward $r_i \le \frac{\epsilon}{n} \max_j r_j$. The maximum we could lose is $n \cdot \frac{\epsilon}{n} \max_j r_j = \epsilon \max_j r_j$, which is at most ϵ fraction of the optimal value. We next obtain all $\tilde{f}_{\mathcal{I}(t)}$, for all $t = 1, \ldots, T$, that approximate $f_{\mathcal{I}(t)}$ (as defined in (1.7)) within a $(1+\epsilon)$ factor. These functions $\tilde{f}_{\mathcal{I}(t)}$ have complexity $\tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$. We start with combining the functions of period 1 and period 2 using $(\max, +)$-convolution. To enforce the constraint that the total size of selected items in period 1 does not exceed the capacity of period 1, we truncate $\tilde{f}_{\mathcal{I}(1)}$ by c_1 (so that any solution using more capacity in period 1 results in $-\infty$ reward) and do the convolution on the truncated function $\tilde{f}_1$. Since both functions are step functions with complexity $\tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$, the $(\max, +)$ convolution can be done in time $\mathcal{O}\left(\frac{1}{\epsilon^2}\right)$. The resulting function $\hat{f}_2$ would have complexity $\mathcal{O}\left(\frac{1}{\epsilon^2}\right)$. To avoid inflating the complexity throughout different periods (which increases computation complexity), the function $\hat{f}_2$ is rounded down to the nearest $r_0 \cdot (1+\epsilon)^k$, where $r_0 := \min_j r_j$ and k is some nonnegative integer. Note that r_0 is a lower bound of any solution value. After discarding small-reward items, we have that $\frac{\max_j r_j}{r_0} \le \frac{n}{\epsilon}$, which implies that $n \max_j r_j = \frac{n^2}{\epsilon} r_0$ is an upper bound for the optimal solution value.

Therefore, after rounding down the function values of $\hat{f}_2$ and obtaining $\tilde{f}_2$, there are at most $\log_{1+\epsilon} \frac{n^2}{\epsilon} \approx \frac{1}{\epsilon} \log \frac{n^2}{\epsilon}$ different values on $\tilde{f}_2$. Now we have brought down the complexity of $\tilde{f}_2$ again to $\tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$, at an additional $(1+\epsilon)$ factor loss in the approximation error. We then move to period 3 and continue this pattern of $(\max, +)$-convolution, truncation, and rounding down. In the end when we reach period T, $\tilde{f}_T$ will only contain feasible solutions to (1.1), and approximate f_T with total approximation factor of $(1+\epsilon)^T \approx (1+T\epsilon)$. Formally, we have the following lemma which shows the approximation factor of $\tilde{f}_t$ for f_t.

Lemma 1.2. *Let $\tilde{f}_t$ be the functions obtained from Algorithm 1.1, and let f_t be defined as in* (1.9)*. Then, $\tilde{f}_t$ approximates f_t with factor $(1+\epsilon)^t$, i.e., $\tilde{f}_t(c) \le f_t(c) \le (1+\epsilon)^t \tilde{f}_t(c)$ for all $0 \le c \le c_t$.*

Lemma 1.2 and Proposition 1.1 together imply that $\tilde{f}_T(c_T)$, obtained from Algorithm 1.1, approximates the optimal value of MPBKP (1.1) by a factor of $(1+\epsilon)^T \approx (1+T\epsilon)$. In Algorithm 1.1, obtaining $\tilde{f}_{\mathcal{I}(t)}$ for all $t = 1, \ldots, T$ takes time $\tilde{O}\left(n + T/\epsilon^{2.25}\right)$; computing the $(\max, +)$-convolution on $\tilde{f}_{t-1} \oplus \tilde{f}_{\mathcal{I}(t)}$ for all t take time $T \cdot m \cdot l = \tilde{O}\left(T/\epsilon^2\right)$. Therefore, Algorithm 1.1 has runtime $\tilde{O}\left(n + T/\epsilon^{2.25}\right)$. As a result, we have the following proposition.

Proposition 1.2. *Taking $\epsilon' = T\epsilon$, Algorithm 1.1 achieves $(1+\epsilon')$-approximation for MPBKP in $\tilde{O}\left(n + \frac{T^{3.25}}{\epsilon'^{2.25}}\right)$.*

1.3.2 A $(1+\epsilon)$-approximation in $\tilde{\mathcal{O}}\left(n + T^2/\epsilon^3\right)$

In this subsection, we introduce another FPTAS for MPBKP which has time complexity $\tilde{\mathcal{O}}\left(n + \frac{T^2}{\epsilon^3}\right)$. To roughly describe the main idea, we will again adopt the functional approach to approximate (1.9). Instead of having an approximation of $f_{\mathcal{I}(t)}$ for each t directly from Lemma 1.1, we further partition $\mathcal{I}(t)$ into $m+1$ subsets (m being specified later), i.e., $\mathcal{I}(t) := \mathcal{I}(t)_0 \sqcup \mathcal{I}(t)_1 \sqcup \cdots \sqcup \mathcal{I}(t)_m$, where items in each subset have approximately the same reward. Then, we have that $f_{\mathcal{I}(t)} = f_{\mathcal{I}(t)_0} \oplus f_{\mathcal{I}(t)_1} \oplus \cdots \oplus f_{\mathcal{I}(t)_m} := \oplus_{j=0}^{m} f_{\mathcal{I}(t)_j}$, and by

noting that the $(\max, +)$-convolution $\oplus$ is commutative, the function f_t as defined in (1.9) can be computed as

$$f_t := \begin{cases} f_{\mathcal{I}(1)}^{c_1} & t = 1, \\ \left(f_{t-1} \oplus f_{\mathcal{I}(t)}\right)^{c_t} = \left(f_{t-1} \oplus f_{\mathcal{I}(t)_0} \oplus f_{\mathcal{I}(t)_1} \oplus \cdots \oplus f_{\mathcal{I}(t)_m}\right)^{c_t} & t \geq 2, \end{cases} \tag{1.11}$$

and (1.11) can be computed more efficiently due to some special properties of $f_{\mathcal{I}(t)_j}$.

Before proceeding to the actual algorithm, we first have some preliminaries. A monotone step function $f_{\mathcal{I}}(c)$ with steps at $c_1, c_2, \ldots, c_l$ is called *r-uniform* if it satisfies both of the following conditions:

1. $\forall c \in \mathbb{R}_+$, $f_{\mathcal{I}}(c) = kr$ for some nonnegative integer k,

2. $\exists c_j$ s.t. $f_{\mathcal{I}}(c_j) = kr \Longrightarrow \exists c_{j'}$ s.t. $f_{\mathcal{I}}(c_{j'}) = k'r, \forall k' \leq k$ nonnegative integers.

The monotone step function $f_{\mathcal{I}}(c)$ with steps at $c_1, c_2, \ldots, c_l$ is called *pseudo-concave* if $c_{j+2} - c_{j+1} \geq c_{j+1} - c_j, \forall j = 1, \ldots, l-2$. The *range* of a function f is the set of all possible function values. We then introduce the following lemma from Chan [46] for approximating $f \oplus g$ when g is r-uniform and pseudo-concave.

Lemma 1.3 (Chan [46])**.** *Let f and g be monotone step functions with total complexity l and ranges contained in $\{-\infty, 0\} \cup \{A, B\}$. Then we can compute a monotone step function that approximates $f \oplus g$ with factor $1 + \mathcal{O}(\epsilon')$ and complexity $\tilde{\mathcal{O}}\left(\frac{1}{\epsilon'}\right)$ in $\mathcal{O}(l) + \tilde{\mathcal{O}}\left(\frac{1}{\epsilon'}\right)$ time if g is r-uniform and pseudo-concave.*

With the above lemma, we present Algorithm 1.2 for MPBKP.

In Algorithm 1.2. We first discard all items with reward $r_i \leq \frac{\epsilon}{n} \max_j r_j$. The maximum we could lose is $n \cdot \frac{\epsilon}{n} \max_j r_j = \epsilon \max_j r_j$, which is at most ϵ fraction of the optimal value. We next round down the rewards of all remaining items to the nearest $r_0 \cdot (1+\epsilon)^k$, where $r_0 := \min_j r_j$ and k is some nonnegative integer, so we lose at most a fraction of $(1+\epsilon)$ in the

Algorithm 1.2 FPTAS for MPBKP in $\tilde{\mathcal{O}}\left(n + T^2/\epsilon^3\right)$

Input: $[n], c_1, \ldots, c_T$ ▷ Set of items to be packed, cumulative capacities up to each time t
Output: $\tilde{f}_t$ ▷ Approximation of function f_t

Discard all items with $r_i \leq \frac{\epsilon}{n} \max_j r_j$ and relabel the items
$r_0 \leftarrow \min_i r_i$ ▷ Lower bound of solution value
$\hat{r}_i \leftarrow r_0 \cdot (1+\epsilon)^{\left\lfloor \log_{1+\epsilon}\left(\frac{r_i}{r_0}\right)\right\rfloor}$ ▷ Round down the reward of each item
$m \leftarrow \left\lceil \log_{1+\epsilon} \frac{n^2}{\epsilon} \right\rceil$ ▷ Number of distinct rewards to be considered, each in the form $r_0 \cdot (1+\epsilon)^k$
$\tilde{f}_0 \leftarrow -\infty$
for $t = 1, \ldots, T$ **do**
  $\hat{f}_t \leftarrow \tilde{f}_{t-1}$
  for $j = 0, \ldots, m$ **do**
    $\mathcal{I}(t)_j = \left\{ i \in \mathcal{I}(t) \mid \hat{r}_i = r_0 \cdot (1+\epsilon)^j \right\}$ ▷ Items in each $\mathcal{I}(t)_j$ has the same rounded reward
    $\hat{\mathcal{I}}(t)_j = \{(\hat{r}_i, q_i) \mid i \in \mathcal{I}(t)_j\}$ and obtain $f_{\hat{\mathcal{I}}(t)_j}$
    ▷ Using items with rounded rewards, build the function $f_{\hat{\mathcal{I}}(t)_j}$
    Approximately compute $\hat{f}_t = \hat{f}_t \oplus f_{\hat{\mathcal{I}}(t)_j}$ using Lemma 1.3
  end for
  $\tilde{f}_t = \hat{f}_t^{c_t}$ ▷ $\tilde{f}_t$ is an approximation of f_t
end for

rounding, and the number of distinct rounded rewards is bounded by $m = \left\lceil \log_{1+\epsilon} \frac{n^2}{\epsilon} \right\rceil = \tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$. We begin with initializing $\tilde{f}_0 = -\infty$. Then, for period $t = 1$, we partition $\mathcal{I}(1) = \sqcup_{j=0}^{m} \mathcal{I}(1)_j$ where all items in $\mathcal{I}(1)_j$ have rounded reward $r_0 \cdot (1+\epsilon)^j$. Denote by $\hat{\mathcal{I}}(1)_j$ these items with rounded rewards, and by adding these items greedily in nonincreasing order of their sizes, we obtain $f_{\hat{\mathcal{I}}(1)_j}$, which is a $(1+\epsilon)$ approximation of $f_{\mathcal{I}(1)_j}$, and is $r_0 \cdot (1+\epsilon)^j$-uniform and pseudo-concave. By applying Lemma 1.3 for $m+1$ times (with ϵ' to be specified later), we obtain $\tilde{f}_0 \oplus f_{\hat{\mathcal{I}}(1)_0} \oplus f_{\hat{\mathcal{I}}(1)_1} \oplus \cdots \oplus f_{\mathcal{I}(1)_m} = \tilde{f}_0 \oplus f_{\hat{\mathcal{I}}(1)}$, which approximates $f_0 \oplus f_{\mathcal{I}(1)}$ with an accumulative approximation factor $(1+\epsilon)(1+\epsilon')^{m+1}$, and is computed in total time $\mathcal{O}(n_1) + \tilde{\mathcal{O}}\left(\frac{m+1}{\epsilon'}\right)$. Then, to ensure feasibility, $\tilde{f}_1$ is obtained by taking truncation c_1 on $\tilde{f}_0 \oplus f_{\hat{\mathcal{I}}(1)}$, which becomes a $(1+\epsilon)(1+\epsilon')^{m+1}$ approximation of f_1. We then move to period 2 and continue this pattern of partition, convolutions, and truncation. In the end as we reach period T, $\tilde{f}_T$ would only contain feasible solutions to (1.1), and approximates f_T with accumulated approximation factor $(1+\epsilon)(1+\epsilon')^{(m+1)T} \approx (1+\epsilon)(1+(m+1)T\epsilon')$. Formally, we have the following lemma which shows the approximation factor of $\tilde{f}_t$ to f_t.

Lemma 1.4. *Let $\tilde{f}_t$ be the functions obtained from Algorithm 1.2, and let f_t be defined as*

in (1.9). *Then,* $\tilde{f}_t$ *approximates* f_t *with factor* $(1+\epsilon)(1+\epsilon')^{(m+1)t}$*, i.e.,* $\tilde{f}_t(c) \leq f_t(c) \leq (1+\epsilon)(1+\epsilon')^{(m+1)t}\tilde{f}_t(c)$ *for all* $0 \leq c \leq c_t$.

The proof of Lemma 1.4 relies on the following fact.

Lemma 1.5. *At any period* t*, after running the inner "for" loop of Algorithm 1.2, we have that* $(1+\epsilon')^{m+1}\hat{f}_t \geq \tilde{f}_{t-1} \oplus f_{\hat{\mathcal{I}}(t)_0} \oplus f_{\hat{\mathcal{I}}(t)_1} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t)_m}$.

Proof of Lemma 1.5. We prove by induction on $j = 0, 1, \ldots, m$. Base case is when $j = 0$, i.e., after the first round of the inner "for" loop, by Lemma 1.3, we have that $(1+\epsilon')\hat{f}_t \geq \tilde{f}_{t-1} \oplus f_{\hat{\mathcal{I}}(t)_0}$. For the induction step, assume that after j rounds of the inner "for" loop, $(1+\epsilon')^j\hat{f}_t \geq \tilde{f}_{t-1} \oplus f_{\hat{\mathcal{I}}(t)_0} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t)_{j-1}}$, we show that after $j+1$ rounds, $(1+\epsilon')^{j+1}\hat{f}_t \geq \tilde{f}_{t-1} \oplus f_{\hat{\mathcal{I}}(t)_0} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t)_j}$. As a notation, we denote by $\hat{f}_t^{\text{old}}$ the $\hat{f}_t$ right before the $(j+1)$th round of the inner "for" loop, and by $\hat{f}_t^{\text{new}}$ the $\hat{f}_t$ right after the $(j+1)$th round of the inner "for" loop. Then, from Lemma 1.3 we have that $(1+\epsilon')\hat{f}_t^{\text{new}} \geq \hat{f}_t^{\text{old}} \oplus f_{\hat{\mathcal{I}}(t)_j}$, which implies that

$$(1+\epsilon')^{j+1}\hat{f}_t^{\text{new}} \geq (1+\epsilon')^j\hat{f}_t^{\text{old}} \oplus f_{\hat{\mathcal{I}}(t)_j} \geq \tilde{f}_{t-1} \oplus f_{\hat{\mathcal{I}}(t)_0} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t)_{j-1}} \oplus f_{\hat{\mathcal{I}}(t)_j},$$

where the second inequality follows from the induction assumption. This finishes the induction step, and thus the proof of the lemma. □

With Lemma 1.5 at hand, we now prove Lemma 1.4.

Proof of Lemma 1.4. By the construction of $\tilde{f}_t$, it should be clear that $\tilde{f}_t \leq f_t$. We prove that $(1+\epsilon)(1+\epsilon')^{(m+1)t}\tilde{f}_t \geq f_t$ by induction on t. Base case is when $t = 1$, we have that $(1+\epsilon)(1+\epsilon')^{m+1}\tilde{f}_1 = (1+\epsilon)(1+\epsilon')^{m+1}\hat{f}_1^{c_1} \geq (1+\epsilon)\left(\tilde{f}_0 \oplus f_{\hat{\mathcal{I}}(1)}\right)^{c_1} = (1+\epsilon)f_{\hat{\mathcal{I}}(1)}^{c_1} \geq f_{\mathcal{I}(1)}^{c_1} = f_1$, where the first inequality follows from Lemma 1.3, and the second inequality follows from the rounding of the rewards. For the induction step, assume that $(1+\epsilon)(1+\epsilon')^{(m+1)t}\tilde{f}_t \geq f_t$, we show that $(1+\epsilon)(1+\epsilon')^{(m+1)(t+1)}\tilde{f}_{t+1} \geq f_{t+1}$.

After partitioning $\mathcal{I}(t+1) = \mathcal{I}(t+1)_0 \sqcup \mathcal{I}(t+1)_1 \sqcup \cdots \sqcup \mathcal{I}(t+1)_m$, for any item $i \in \mathcal{I}(t+1)$, by the rounding down, we have that $(1+\epsilon)\hat{r}_i \geq r_i \geq \hat{r}_i$, which further implies that $(1+\epsilon) f_{\hat{\mathcal{I}}(t+1)_j} \geq f_{\mathcal{I}(t+1)_j} \geq f_{\hat{\mathcal{I}}(t+1)_j}, \forall j = 0, 1, \ldots, m$. Thus,

$$\begin{aligned}(1+\epsilon)\left(f_{\hat{\mathcal{I}}(t+1)_0} \oplus f_{\hat{\mathcal{I}}(t+1)_1} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t+1)_m}\right) &\geq f_{\mathcal{I}(t+1)} \\ &\geq f_{\hat{\mathcal{I}}(t+1)_0} \oplus f_{\hat{\mathcal{I}}(t+1)_1} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t+1)_m},\end{aligned}$$

which, together with the induction assumption, implies that

$$(1+\epsilon)(1+\epsilon')^{(m+1)t}\left(\tilde{f}_t \oplus f_{\hat{\mathcal{I}}(t+1)_0} \oplus f_{\hat{\mathcal{I}}(t+1)_1} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t+1)_m}\right) \geq f_t \oplus f_{\mathcal{I}(t+1)}.$$

By Lemma 1.5, after the inner "for" loop in Algorithm 1.2, we have that $(1+\epsilon')^{m+1}\hat{f}_{t+1} \geq \tilde{f}_t \oplus f_{\hat{\mathcal{I}}(t+1)_0} \oplus f_{\hat{\mathcal{I}}(t+1)_1} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t+1)_m}$, which implies that

$$\begin{aligned}&(1+\epsilon)(1+\epsilon')^{(m+1)(t+1)}\hat{f}_{t+1} \\ \geq\ &(1+\epsilon)(1+\epsilon')^{(m+1)(t+1)}\left(\tilde{f}_t \oplus f_{\hat{\mathcal{I}}(t+1)_0} \oplus f_{\hat{\mathcal{I}}(t+1)_1} \oplus \cdots \oplus f_{\hat{\mathcal{I}}(t+1)_m}\right) \\ \geq\ &f_t \oplus f_{\mathcal{I}(t+1)}.\end{aligned}$$

Taking truncation on both sides, we conclude that

$$(1+\epsilon)(1+\epsilon')^{(m+1)(t+1)}\tilde{f}_{t+1} = (1+\epsilon)(1+\epsilon')^{(m+1)(t+1)}\hat{f}_{t+1}^{c_{t+1}} \geq \left(f_t \oplus f_{\mathcal{I}(t+1)}\right)^{c_{t+1}} = f_{t+1}.$$

This finishes the induction step, and thus the proof of the lemma. □

Lemma 1.4 and Proposition 1.1 together imply that $\tilde{f}_T(c_T)$, obtained from Algorithm 1.2, approximates the optimal value of MPBKP (1.1) by a factor of $(1+\epsilon)(1+\epsilon')^{(m+1)T} \approx (1+\epsilon+mT\epsilon')$. In Algorithm 1.2, during each of the periods $t = 1, \ldots, T$, approximately computing the $(\max, +)$-convolutions on $\hat{f}_t \oplus f_{\hat{\mathcal{I}}(t)_j}$ for all $j = 0, 1, \ldots, m$ take total time

$\tilde{O}\left(n_t + (m+1)/\epsilon'\right)$. Therefore, Algorithm 1.2 has total runtime $\tilde{O}\left(n + (m+1)T/\epsilon'\right)$. As a result, we have the following proposition.

Proposition 1.3. *Taking $\epsilon = mT\epsilon'$ and $m = \tilde{\mathcal{O}}(1/\epsilon)$, Algorithm 1.2 achieves $(1+\epsilon)$-approximation for MPBKP in $\tilde{O}\left(n + \frac{T^2}{\epsilon^3}\right)$.*

1.4 Approximation Algorithms for MPBKP-S

In this section, we provide two approximation algorithms for MPBKP-S. The first one is an FPTAS with time complexity $\mathcal{O}\left(\frac{Tn\log n}{\epsilon^2}\right)$, and the second one is a parameterized approximation algorithm that (under Assumption 1.1) achieves $\left(1+\frac{\epsilon}{1-\beta}\right)$ approximation factor with time complexity $\tilde{\mathcal{O}}\left(n + \frac{T^{3.25}}{\epsilon^{2.25}}\right)$.

1.4.1 FPTAS for MPBKP-S

In this subsection, we provide an FPTAS for the MPBKP-S with time complexity $\mathcal{O}\left(\frac{Tn\log n}{\epsilon^2}\right)$. An alternative FPTAS with time complexity $\mathcal{O}\left(\frac{n^2\log n}{\epsilon}\right)$ is provided in Appendix 1.7.4. Combining the two, we show that our algorithms achieve $(1+\epsilon)$ approximation ratio in time $\mathcal{O}\left(\frac{n\log n}{\epsilon}\cdot\min\left\{\frac{T}{\epsilon}, n\right\}\right)$, which proves Theorem 1.2. We should note that the algorithm in the previous section does not apply here: We could similarly define a function which gives the maximum profit (= reward−penalty) under a given capacity constraint, but the main obstacle is on the $(\max,+)$-convolution, because profit does not "add up". In other words, the total profit we earn by selecting items in the set $\mathcal{S}_1 \cup \mathcal{S}_2$ is not the sum of the profits we earned by selecting $\mathcal{S}_1$ and $\mathcal{S}_2$ separately. For this reason, we can no longer rely on the techniques used in function approximation and $(\max,+)$-convolution as in Chan [46], Jin [121]. Instead, our main idea is motivated by the techniques that originated from earlier papers (Ibarra and Kim [113], Lawler [135]), but adapting their technique to MPBKP-S requires significant modifications as we show in this section. We restrict our presentation to

the case $B_t = B$ for readability, but our algorithms and analysis generalize in a straightforward manner when the penalties for buying capacity are heterogeneous $\{B_1, \ldots, B_T\}$ (by replacing B with $\min_{\tau \le t} B_\tau$ in the calculations of profit/penalty at period t on line 7 of Algorithm 1.3).

Preliminaries: We first introduce some notation. From now on, let $\mathcal{R}(\mathcal{S}) := \sum_{i \in \mathcal{S}} r_i$. The optimal solution set to (1.5) is denoted by $\mathcal{S}^*$. The total profit earned can be expressed as a function of the solution set $\mathcal{S}$:

$$\mathcal{P}(\mathcal{S}) = \mathcal{R}(\mathcal{S}) - B \cdot \sum_{t=1}^{T} \left[\sum_{j \in \mathcal{S} \cap \mathcal{I}(t)} q_j - \max \left\{ c_t - \sum_{j \in \mathcal{S}, d_j \le t} q_j,\ c_t - c_{t-1} \right\} \right]^+ . \tag{1.12}$$

Let p_i be the profit of item i, which is defined as the profit earned if we select only i, i.e., $p_i = r_i - B \cdot \left(q_i - c_{d_i}\right)^+$. Without loss of generality, we assume that each item i is by itself profitable, i.e., $p_i \ge 0$, so one profitable solution would be $\{i\}$. Let $P := \max_i p_i$ and $\bar{P} := \sum_{i \in [n]} p_i$. The following bounds on $\mathcal{P}(\mathcal{S}^*)$ follow:

$$P \le \mathcal{P}(\mathcal{S}^*) \le \bar{P} \le nP. \tag{1.13}$$

Partition of items: We partition the set of items $[n]$ into two sets: a set of "large" items $\mathcal{I}_L$ and a set of "small" items $\mathcal{I}_S$ such that we can bound the number of large items in any optimal solution. The main idea is to use dynamic programming to pick the large items in the solution, and a greedy heuristic for 'padding' this partial solution with small items. The criterion for small and large items is based on balancing the permissible error $\epsilon \mathcal{P}(\mathcal{S}^*)$ equally in filling large items and filling small items. Instead of first packing all large items and then all small items, we consider items in the order of their deadlines, and for each deadline t, the large items are selected first and then the small items are selected greedily in order of their reward densities. As a result, the approximation error due to large items overall will be $\frac{1}{2}\epsilon \mathcal{P}(\mathcal{S}^*)$ and the error due to the small items with each deadline will be $\frac{1}{2T}\epsilon \mathcal{P}(\mathcal{S}^*)$. This

gives a total approximation error of $\frac{1}{2}\epsilon\mathcal{P}(\mathcal{S}^*) + T \cdot \frac{1}{2T}\epsilon\mathcal{P}(\mathcal{S}^*) = \epsilon\mathcal{P}(\mathcal{S}^*)$.

Suppose that we can find some P_0 that satisfies (1.14).

$$P_0 \leq \mathcal{P}(\mathcal{S}^*) \leq 2P_0. \tag{1.14}$$

Then, the set of items is partitioned as follows.

$$\mathcal{I}_L := \left\{ i \in [n] \mid p_i \geq \frac{1}{2T}\epsilon P_0 \right\}; \qquad \mathcal{I}_S := \left\{ i \in [n] \mid p_i < \frac{1}{2T}\epsilon P_0 \right\}. \tag{1.15}$$

This partition is computed in $\mathcal{O}(n)$ time and is not the dominant term in time complexity. Let $n_L = |\mathcal{I}_L|$ and $n_S = |\mathcal{I}_S|$, so that $n_L + n_S = n$. Further, let

$$\mathcal{I}_L(t) := \{i \in \mathcal{S}_L \mid d_i = t\}, \quad \text{and} \quad \mathcal{I}_S(t) := \{i \in \mathcal{S}_S \mid d_i = t\}$$

denote the set of large and small items, respectively, with deadline t. We will assume that the items in $\mathcal{I}_L$ are indexed in non-decreasing order of their deadlines, i.e., $\forall i, j \in \mathcal{I}_L$ such that $j \geq i$, we have that $d_i \leq d_j$. Denote by $I_L(t)$ as the index of the last item with deadline t, i.e., $I_L(t) := \max_{i \in \mathcal{S}_L \cap \mathcal{I}_L(t)} i$. For each time t, we will also sort the small items in $\mathcal{I}_S(t)$ according to their reward densities, i.e., $\forall i < j$ and $i, j \in \mathcal{I}_S(t)$, $\frac{r_i}{q_i} \geq \frac{r_j}{q_j}$. This sorting only takes place once for each guess P_0, and does not affect our overall time complexity result.

Algorithm overview: Our FPTAS is given in Algorithm 1.6 which uses a doubling trick to guess the value of P_0 satisfying (1.14), and for each guess uses Algorithm 1.5 as a subroutine. Algorithm 1.5 is the main algorithm for MPBKP-S, which first selects the items with deadline 1, then the items with deadline 2, and so on. For each deadline t, we maintain two sets of partial solutions: the first, $\widetilde{A}_t(p)$, corresponds to an approximately optimal (in terms of leftover capacity carried forward to time $t+1$) subset of large and small items with deadline at most t and some *rounded profit* p; and the second, $\hat{A}_t(p)$, corresponds to

Algorithm 1.3 DP on large items for MPBKP-S

Input: $\mathcal{I}_L, \Delta c$, ▷ Set of (large) items to be packed, additional capacity available for packing
$\widetilde{A}(p)$ for all $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$ ▷ A set of partial solutions
Output: $\hat{A}(I_L, p)$ for all $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$ ▷ Set of partial solutions after packing $\mathcal{I}_L$

Initialize $\forall p \ : \ \hat{A}(0, p) := \widetilde{A}(p) + \Delta c$
for $i = 1, \ldots, I_L$ **do**
  for $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$ **do**
    $\hat{A}(i, p) := \hat{A}(i-1, p)$ ▷ If reject item i
  end for
  for $\bar{p} = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$ **do**
    $p = \bar{p} + \hat{r}_i - \left\lceil B\left(q_i - \max\left\{0, \hat{A}(i-1, \bar{p})\right\}\right)^+ \right\rceil_\kappa$
    $\hat{A}(i, p) = \max\left\{\hat{A}(i, p), \hat{A}(i-1, \bar{p}) - q_i\right\}$ ▷ Accept i
  end for
  for $p = \left\{\left\lceil \frac{16T}{\epsilon^2} \right\rceil, \left\lceil \frac{16T}{\epsilon^2} \right\rceil - 1, \ldots, 1\right\} \cdot \kappa$ **do**
    if $\hat{A}(i, p-\kappa) < \hat{A}(i, p)$ **then**
      $\hat{A}(i, p-\kappa) = \hat{A}(i, p)$
    end if
  end for
end for

the optimal appending of large items with deadline t to the approximately optimal set of solutions corresponding to $\widetilde{A}_{t-1}$.

Given $\widetilde{A}_{t-1}$, we first select large items from $\mathcal{I}_L(t)$ using dynamic programming to obtain $\hat{A}_t$, which is done in Algorithm 1.3. In other words, *given* the partial solutions $\widetilde{A}_{t-1}(\bar{p})$ for all $\bar{p} \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$, $\hat{A}_t(p)$ is the maximum capacity left when earning *rounded profit* (precise definition given in (1.19)) p by adding items in $\mathcal{I}_L(t)$. We then use a greedy heuristic to pick small items from $\mathcal{I}_S(t)$ to obtain $\widetilde{A}_t$, which is done in Algorithm 1.4. Specifically, our goal in Algorithm 1.4 is to obtain the partial solutions $\widetilde{A}_t(\cdot)$ given the partial solutions $\hat{A}_t(\cdot)$ by packing the small items $\mathcal{I}_S(t)$. We initialize $\widetilde{A}_t(\bar{p})$ with $\hat{A}_t(\bar{p})$, and for each $\bar{p}$ we try to augment the solution corresponding to $\hat{A}_t(\bar{p})$ using a subset $\widetilde{\mathcal{I}}_S(t) \subseteq \mathcal{I}_S(t)$ defined as

$$\widetilde{\mathcal{I}}_S(t) := \{i \in \mathcal{I}_S(t) \mid q_i \leq \hat{A}_t(\bar{p})\}.$$

The small items in $\widetilde{\mathcal{I}}_S(t)$ are sorted according to their reward densities, and are added to the solution of $\hat{A}_t(\bar{p})$ one by one. After each addition of a small item, if the new total rounded

Algorithm 1.4 Greedy on small items for MPBKP-S

Input: $\mathcal{I}_S$, $\hat{A}(p)$ for all $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$. ▷ Set of small items, set of partial solutions
Output: $\widetilde{A}(p)$ for all $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$ ▷ Set of partial solutions after packing $\mathcal{I}_S$

Initialize $\forall p \ : \ \widetilde{A}(p) = \hat{A}(p)$
for $\bar{p} = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$ **do**
`// Filter out small items with size larger than` $\hat{A}(p)$
  $\widetilde{\mathcal{I}}_S \leftarrow \emptyset$
  for $i \in \mathcal{I}_S$ **do**
    if $\hat{A}(\bar{p}) \geq q_i$ **then**
      $\widetilde{\mathcal{I}}_S \leftarrow \widetilde{\mathcal{I}}_S \cup \{i\}$
    end if
  end for
  $\tilde{\mathcal{R}}_{0'} = 0, \tilde{q}_{0'} = 0$, and relabel the items in $\widetilde{\mathcal{I}}_S$ as $\left\{1', \ldots, |\widetilde{\mathcal{I}}_S|'\right\}$ (in decreasing order of reward density)
  for $i' = 1', \ldots, |\widetilde{\mathcal{I}}_S|'$ **do**
    $\tilde{\mathcal{R}}_{i'} = \tilde{\mathcal{R}}_{(i-1)'} + r_{i'}$
    $\tilde{q}_{i'} = \tilde{q}_{(i-1)'} + q_{i'}$
  end for
  `// Add small items using Greedy algorithm`
  for $i' = 1', \ldots, |\widetilde{\mathcal{I}}_S|'$ **do**
    if $\tilde{q}_{i'} \leq \hat{A}(\bar{p})$ **then**
      $p = \left\lfloor \bar{p} + \tilde{\mathcal{R}}_{i'} \right\rfloor_\kappa$
      $\widetilde{A}(p) = \max\left\{\widetilde{A}(p), \hat{A}(\bar{p}) - \tilde{q}_{i'}\right\}$
    end if
  end for
end for

reward is p, we compare the leftover capacity with current $\widetilde{A}_t(p)$, and update $\widetilde{A}_t(p)$ with the new solution if it has more leftover capacity. We continue this add-and-compare (and possibly update) until we reach the situation where adding the next small item overflows the available capacity.

Intuitively, for any amount of capacity available to be filled by small items, and a minimum increase in profit, the optimal solution either packs a single item from $\mathcal{I}_S(t) \setminus \widetilde{\mathcal{I}}_S(t)$ in which case the loss by ignoring items in this set is bounded by the maximum reward of any small item, or the optimal solution only contains items from $\widetilde{\mathcal{I}}_S(t)$ in which case the space used by this optimal set of items is lower bounded by the a fractional packing of the highest density items in $\widetilde{I}_S(t)$. During Algorithm 1.4, one of the solutions we would consider would be the integral items of this fractional solution, and lose at most $\frac{1}{2T}\epsilon P_0$ in profit, and

Algorithm 1.5 DP on large items and Greedy on small items for MPBKP-S

Define $\kappa = \frac{\epsilon^2 P_0}{8T}$
Define $\hat{r}_i = \lfloor r_i \rfloor_\kappa$ ▷ Round down reward
// $\widetilde{A}_t(p)$ = leftover capacity for the algorithm's partial solution when earning (rounded) profit p using items with deadlines at most t (small and large)
// $\hat{A}_t(p)$ = capacity left for the algorithm's partial solution when earning (rounded) profit p by selecting large items in $\mathcal{I}_L(t)$ with rounded down rewards $\hat{r}$, given the partial solutions $\widetilde{A}_{t-1}(p)$
Initialize $\hat{A}(0,p) = \widetilde{A}_0(p) = \begin{cases} 0 & p = 0, \\ -\infty & p > 0. \end{cases}$
for $t = 1, \ldots, T$ **do**
    Run Algorithm 1.3 with $\mathcal{I}_L = \mathcal{I}_L(t)$, $\Delta c = c_t - c_{t-1}$, and $\widetilde{A}(p) = \widetilde{A}_{t-1}(p)$ for all $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$, and obtain $\hat{A}_t(p) := \hat{A}(I_L, p)$ for all p.
    Run Algorithm 1.4 with $\mathcal{I}_S = \mathcal{I}_S(t)$ and $\hat{A}(p) = \hat{A}(I_L(t), p)$ for all $p = \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$, and obtain $\widetilde{A}_t(p) := \widetilde{A}(p)$ for all p.
end for

Algorithm 1.6 FPTAS for MPBKP-S in $\mathcal{O}(Tn \log n/\epsilon^2)$

$P_0 \leftarrow \bar{P}$
$p^* \leftarrow 0$
while $p^* < (1-\epsilon)P_0$ **do**
    $P_0 \leftarrow \frac{P_0}{2}$
    Run Algorithm 1.5 with the current P_0.
    $p^* \leftarrow \max \left\{ p \in \left\{0, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa \;\middle|\; \widetilde{A}_T(p) > -\infty \right\} p$
end while

obtain a solution with still smaller space used (more leftover capacity) than the fractional solution. Accumulation of these errors for t periods then will give us the invariant: the partial solution $\widetilde{A}_t(p)$ obtained as above has more leftover capacity than any solution obtained by selecting items from $\cup_{t'=1}^{t} \mathcal{I}_L(t')$ with rounded rewards and rounded penalties, and items from $\cup_{t'=1}^{t} \mathcal{I}_S(t')$ with original (unrounded) rewards such that the rounded total profit is at least $p + \frac{1}{2T}\epsilon P_0 t + \kappa t$.

Our main theorem for the approximation ratio for MPBKP follows.

Theorem 1.4 (Partially restating Theorem 1.2)**.** *Algorithm 1.6 is a fully polynomial approximation scheme for the MPBKP-S, which achieves* $(1+\epsilon)$ *approximation ratio with running time* $\mathcal{O}\left(\frac{Tn \log n}{\epsilon^2}\right)$.

Remark 1.1. *Theorem 1.4, together with Theorem 1.8, implies that we can obtain a* $(1+\epsilon)$ *approximate solution for the MPBKP-S in* $\mathcal{O}\left(\frac{n\log n}{\epsilon}\cdot\min\left\{\frac{T}{\epsilon},n\right\}\right)$*, where Algorithm 1.12 is used when* $T/\epsilon \gg n$ *and Algorithm 1.6 is used when* $T/\epsilon \ll n$*.*

Remark 1.2. *One may question if it is possible to achieve* $\tilde{\mathcal{O}}\left(n+T^{\alpha}/\epsilon^{\beta}\right)$ *for some* α,β*, as in the 0-1 Knapsack problem. We note that using a finer rounding technique as in Lawler [135], the number of large items can be further bounded from* $\mathcal{O}(n)$ *to* $\mathcal{O}\left(\frac{T}{\epsilon^2}\right)$*, which would reduce the runtime of the DP (for large items) from* $\mathcal{O}\left(nT/\epsilon^2\right)$ *to* $\tilde{\mathcal{O}}\left(n+T^2/\epsilon^4\right)$*. However, the small items still have to be added one by one to the solutions of* $\hat{A}_t(\bar{p})$*, which in the worst case takes* $\mathcal{O}\left(\frac{nT}{\epsilon^2}\right)$*. We cannot first group the small items into sets of partial solutions and do* $(\max,+)$ *convolution with solution sets of* $\hat{A}_t(p)$*, which would take* $\mathcal{O}\left(\frac{T^2}{\epsilon^4}\right)$ *(similar to Ibarra and Kim [113], Lawler [135]), because again the profits of two sets do not add up when we take the union of these two sets. Therefore, we do not further bound the number of large items as it does not improve the overall asymptotic time complexity (since the bottleneck is on packing small items).*

1.4.2 Parameterized Approximation for MPBKP-S

In this subsection, we provide a parameterized approximation algorithm with $\left(1+\frac{\epsilon}{1-\beta}\right)$-approximation factor and $\tilde{\mathcal{O}}\left(n+\frac{T^{3.25}}{\epsilon^{2.25}}\right)$ runtime. We will again use the results of the 0-1 Knapsack problem [121] as building blocks. However, unlike in Section 1.3.1, where we first obtained the approximated function that gives maximum reward (by selecting items in that period) on every capacity, and then combined the "truncated" version of these functions using the $(\max,+)$-convolution, in MPBKP-S where the capacity constraints can be violated by paying penalties, the total profit of two sets of items do not equal to the sum of the profits of each set separately. As a result, we cannot rely on the $(\max,+)$ convolution to combine the sets of items from different periods.

We propose, in this subsection, an algorithm that builds a dynamic programming on the

approximately maximum leftover capacities for each period and each profit, based upon the function approximation results on the (approximated) maximum reward for each capacity. This algorithm differentiates itself from the one in Section 1.4.1 in the sense that items are no longer divided into "large" ones and "small" ones. Further, instead of keeping track of the additional profit obtained from each item, we use the approximated function which focuses on the rewards. For each period we obtain a number of candidate solution sets, each earning the approximately maximum reward given its total size. We then consider adding each set (with all items in the set together) and see how much additional profit we could earn. By focusing only on rewards while building the item sets for each period, we are able to utilize the results from Jin [121] and the algorithm runs in $\tilde{\mathcal{O}}\left(n + \frac{T^{3.25}}{\epsilon^{2.25}}\right)$, which in general performs better than $\mathcal{O}\left(\frac{nT \log n}{\epsilon^2}\right)$ when n is large. In the end, nevertheless, the approximation factors of these rewards (and penalties) will need to be converted back to profits, and that is where we need the additional Assumption 1.1.

We first review some notation that was introduced in the previous sections. Recall that the function $f_{\mathcal{I}}$ for $\mathcal{I} = \{(r_1, q_1), \ldots, (r_{n'}, q_{n'})\}$ is defined as in (1.7): For all $c \in \mathbb{R}$,

$$f_{\mathcal{I}}(c) = \max_{x_1,\ldots,x_{n'}} \left\{ \sum_{i \in \mathcal{I}} r_i x_i \ : \ \sum_{i \in \mathcal{I}} q_i x_i \leq c, \ x_1, \ldots, x_{n'} \in \{0,1\} \right\},$$

while $\tilde{f}_{\mathcal{I}}$ is a $(1+\epsilon)$ approximation of $f_{\mathcal{I}}$, with complexity $\tilde{\mathcal{O}}\left(1/\epsilon\right)$, that can be obtained in $\tilde{\mathcal{O}}\left(n + 1/\epsilon^{2.25}\right)$ by Lemma 1.1. The set of items with deadline t is denoted by $\mathcal{I}(t)$. For any set of items $\mathcal{S}$, we let $\mathcal{P}(\mathcal{S})$ be the profit of set $\mathcal{S}$, and let $p_i := \mathcal{P}(\{i\})$ and $p_0 := \min_{i \in [n]} p_i$. With these preliminaries, we present our parameterized algorithm for MPBKP-S.

We now describe the intuition behind Algorithm 1.7, with the rigorous proofs left to Appendix. At the beginning, we discard all items with $p_i \leq \frac{\epsilon}{n} \max_j p_j$. The maximum total profit we could lose is bounded by $n \cdot \frac{\epsilon}{n} \max_j p_j = \epsilon \max_j p_j$, which is at most ϵ fraction of the optimal profit. For each time t, we obtain partial solutions $\tilde{A}(t, p)$ that corresponds to

Algorithm 1.7 Parameterized approximation for MPBKP-S

Input: $[n], a_1, \ldots, a_T$ ▷ Set of items to be packed, incremental capacities at each time t
Output: $\tilde{A}(T, p)$

Discard all items with $p_i \leq \frac{\epsilon}{n} \max_j p_j$ and relabel the items ▷ p_i is the profit earned with item i itself
$p_0 \leftarrow \min_i p_i$ ▷ Lower bound of optimal profit
$m \leftarrow \left\lceil \log_{1+\epsilon} \frac{n^2}{\epsilon} \right\rceil$ ▷ number of distinct p values, each being p_0 times a power of $(1+\epsilon)$
// $\tilde{A}(t,p)$ = `maximum capacity left at time` t `when earning (rounded) profit at least` p `using items in` $\cup_{t'=1}^{t} \mathcal{I}(t')$ `with approximated functions` $\tilde{f}_{\mathcal{I}(t)}$
Initialize $\tilde{A}(0,p) = \begin{cases} 0 & p = 0, \\ -\infty & p > 0. \end{cases}$
for $t = 1, \ldots, T$ **do**
  Obtain $\tilde{f}_{\mathcal{I}(t)}$ that approximates $f_{\mathcal{I}(t)}$ with factor $1+\epsilon$ using Lemma 1.1
  $l \leftarrow$ complexity of $\tilde{f}_{\mathcal{I}(t)}$ ▷ $l = \tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$
  // `In reality,` $\tilde{f}_{\mathcal{I}(t)}$ `is saved as` l `pairs` $\{(C_1^t, R_1^t), (C_2^t, R_2^t), \ldots, (C_l^t, R_l^t))\}$
  for $p = 0,\ p_0 \cdot (1+\epsilon)^{\{-1,0,1,\ldots,m\}}$ **do**
    $\tilde{A}(t,p) := \tilde{A}(t-1,p) + a_t$
  end for
  for $\bar{p} = 0,\ p_0 \cdot (1+\epsilon)^{\{-1,0,1,\ldots,m\}}$ **do**
    for $k = 1, \ldots, l$ **do**
      if $\bar{p} + R_k^t - B\left(C_k^t - \max\left\{0, \tilde{A}(t-1,\bar{p})\right\} - a_t\right)^+ \geq \frac{p_0}{1+\epsilon}$ **then**
        $p = p_0 \cdot (1+\epsilon)^{\left\lfloor \log_{1+\epsilon}\left(\frac{\bar{p} + R_k^t - B\left(C_k^t - \max\left\{0, \tilde{A}(t-1,\bar{p})\right\} - a_t\right)^+}{p_0}\right)\right\rfloor}$
        $\tilde{A}(t,p) = \max\left\{\tilde{A}(t,p), \tilde{A}(t-1,\bar{p}) - C_{tk} + a_t\right\}$
      end if
    end for
  end for
  for $p = p_0 \cdot (1+\epsilon)^{\{m, m-1, \ldots, 1\}}$ **do**
    if $\tilde{A}(t, p/(1+\epsilon)) < \tilde{A}(t,p)$ **then**
      $\tilde{A}(t, p/(1+\epsilon)) = \tilde{A}(t,p)$
    end if
  end for
end for

an approximate optimal (in terms of leftover capacity carried forward to time $t+1$) set of items with deadline at most t and total *rounded profit* (precise definition given in (1.22)) at least p. Specifically, we first obtain $\tilde{f}_{\mathcal{I}(t)}$ that approximates $f_{\mathcal{I}(t)}$ (as defined in (1.7)) with factor $1+\epsilon$. By Lemma 1.1, $\tilde{f}_{\mathcal{I}(t)}$ is a step function with at most $l = \tilde{\mathcal{O}}\left(1/\epsilon\right)$ "steps", i.e., the function $\tilde{f}_{\mathcal{I}(t)}$ can be fully characterized as l size-reward pairs, each corresponding to one "step" of the function: $\{(C_1^t, R_1^t), (C_2^t, R_2^t), \ldots, (C_l^t, R_l^t))\}$.

Then, we consider adding the sets of items corresponding to each of these pairs, all of which having deadline t, to the existing partial solutions that include items with deadlines up

to $t-1$, via dynamic program. As we will prove later, *given* the partial solutions $\tilde{A}(t-1,p)$, the $\tilde{A}(t,p)$ we obtained from Algorithm 1.7 has the maximum leftover capacity when earning rounded profit p by adding items in $\mathcal{I}(t)$. To limit the number of p's to be considered in the dynamic program, after trying to add each set of items, we always round the total profit down to the nearest $p_0(1+\epsilon)^k$ for some integer $k \geq -1$. Note that p_0 is a lower bound of profit for any solution, and $p_0/(1+\epsilon)$ is the lower bound of the approximated profit for any solution. After discarding small-reward items, we have that $\frac{\max_j p_j}{p_0} \leq \frac{n}{\epsilon}$, which implies that $n \max_j p_j = \frac{n^2}{\epsilon} p_0$ is an upper bound for the optimal profit. Therefore, with p being the rounded down profit, there are at most $\log_{1+\epsilon} \frac{n^2}{\epsilon(1+\epsilon)} \approx \frac{1}{\epsilon} \log \frac{n^2}{\epsilon}$ different values of p in $\tilde{A}(t,p)$. In the end, we obtain the solutions corresponding to $\tilde{A}(T,p)$ and find the one with largest p while keeping $\tilde{A}(T,p) \neq -\infty$. The profit of this solution, after all the rounding downs, will have accumulated approximation factor $\left(1+\frac{\epsilon T}{1-\beta}\right)$, and the total runtime will be $\tilde{\mathcal{O}}\left(n + T/\epsilon^{2.25}\right)$. After scaling ϵ properly, we have the following theorem.

Theorem 1.5. *Under Assumption 1.1, Algorithm 1.7 achieves* $\left(1+\frac{\epsilon}{1-\beta}\right)$*-approximation factor for MPBKP-S with runtime* $\tilde{\mathcal{O}}\left(n + \frac{T^{3.25}}{\epsilon^{2.25}}\right)$.

The formal proof of Theorem 1.5 is left to Appendix 1.7.2.2.

1.5 Approximation Algorithms for MPBKP-SS

In this section, we consider the MPBKP-SS as defined in (1.6). We first show in Section 1.5.1 that when all items have size 1, a greedy algorithm achieves 2-approximation. Then, in Section 1.5.2, we provide a parameterized approximation algorithm that achieves $\left(1+\frac{\epsilon}{1-\beta}\right)$-approximation with time complexity $\tilde{\mathcal{O}}\left(n + 1/\epsilon^T\right)$, where we also address the difficulty of this problem by its nature.

1.5.1 A Greedy Algorithm for a Special Case of MPBKP-SS

In this subsection, we consider the special case of MPBKP-SS when all items have the same size, i.e., $q_i = q, \forall i \in [n]$. We again only present for the case $B_t = B, \forall t \in [T]$. The problem is written as

$$\max_{x,y} \sum_{i\in[n]} r_i x_i - \mathbb{E}_\omega \left[B \cdot \sum_{t=1}^{T} y_t(\omega) \right] \tag{1.16a}$$

$$\text{s.t.} \sum_{i\in\mathcal{I}(1)\cup\cdots\cup\mathcal{I}(t)} x_i - \sum_{s=1}^{t} y_s(\omega) \le \sum_{s=1}^{t} a_t(\omega) = c_t(\omega), \quad \forall \omega \in \Omega, 1 \le t \le T \tag{1.16b}$$

$$x_i \in \{0,1\}, \quad y_t \ge 0. \tag{1.16c}$$

As we point out in Appendix 1.7.5, in the deterministic problems (MPBKP or MPBKP-S), when items all have sizes $q_i = 1$, greedily adding items one by one in decreasing order of their rewards leads to the optimal solution. For MPBKP-SS, as the incremental capacities are now stochastic, we wonder if there is any greedy algorithm performs well. We propose Algorithm 1.8, where we start with an empty set, and greedily insert the item that brings the maximum increment on expected profit, and we stop if adding any of the remaining items does not increase the expected profit.

Algorithm 1.8 Greedy algorithm according to profit change

$\mathcal{S} \leftarrow \emptyset$
$s \leftarrow 1$
while $s == 1$ **do**
    $i^* \leftarrow \arg\max_{i\notin\mathcal{S}} \{\mathcal{P}(\mathcal{S}\cup\{i\}) - \mathcal{P}(\mathcal{S})\}$
    if $\mathcal{P}(\mathcal{S}\cup\{i^*\}) - \mathcal{P}(\mathcal{S}) \ge 0$ **then**
        $\mathcal{S} \leftarrow \mathcal{S}\cup\{i^*\}$
    else
        $s \leftarrow 0$
    end if
end while
$\mathcal{S}_p \leftarrow \mathcal{S}$
Return $\mathcal{S}_p$

Let $\mathcal{S}^*$ be an optimal solution, i.e., $\mathcal{S}^* \in \arg\max_{\mathcal{S}\subseteq[n]} \mathcal{P}(\mathcal{S}) := \mathcal{R}(\mathcal{S}) - B \cdot \Phi(\mathcal{S})$, where

$$\Phi(\mathcal{S}) := \mathbb{E}\left\{\sum_{t=1}^{T}\left[\sum_{j\in\mathcal{I}(t)\cap\mathcal{S}} q_j - \max_{0\le t'<t}\left\{c_t - c_{t'} - \sum_{j\in\mathcal{S}:t'+1\le d_j\le t-1} q_j\right\}\right]^+\right\}$$

is the expected quantity of overflow on set $\mathcal{S}$, and let $\mathcal{S}_p$ be the set output by Algorithm 1.8. Then, we have the following theorem.

Theorem 1.6 (Restating Theorem 1.3)**.** *Algorithm 1.8 achieves 2-approximation factor for MPBKP-SS when items have the same size, i.e.,* $\mathcal{P}(\mathcal{S}_p) \ge \frac{1}{2}\mathcal{P}(\mathcal{S}^*)$ *in* $\mathcal{O}\left(n^2T|\Omega|\right)$.

The proof of the 2-approximation could be more nontrivial than one may think. The idea is to look at the greedy solution set $\mathcal{S}_p$ and the optimal solution set $\mathcal{S}^*$, where we will use the dual to characterize the optimal solution on each sample path. By swapping each item in $\mathcal{S}_p$ to $\mathcal{S}^*$ in replacement of the same item or two other items, we construct a sequence of partial solutions of the greedy algorithm as well as modified optimal solution set, while maintaining the invariant that the profit of $\mathcal{S}^*$ is bounded by the sum of two times the profit of items in $\mathcal{S}_p$ swapped into $\mathcal{S}^*$ so far and the additional profit of remaining items in the modified optimal solution set. We leave the formal proof of Theorem 1.6 to Appendix 1.7.3.1.

1.5.2 Parameterized Approximation for MPBKP-SS

In this subsection, we consider the most general problem as defined in (1.6). Now that the incremental capacities for each period are stochastic, the problem becomes even more challenging. Neither algorithms we provided for MPBKP-S could apply. To see this, recall that both the FPTAS and the parameterized approximation algorithm we provided in Section 1.4 rely on the dynamic program that keeps tracking on the *maximum leftover capacity* to earn at least some level of profit p using items with deadlines no later than t. Unfortunately, such dynamic program could not be naively applied to solve the problem MPBKP-SS. Specifically,

in MPBKP-S, for different sets of partial solutions at time t (sets of items with deadlines up to t) that earn the same (rounded) profit, we only need to consider the one with the *maximum* leftover capacity when moving to $t+1$. Here the *leftover capacities* are real numbers and have a total ordering. For MPBKP-SS, if we set up the DP in a similar manner, we would want to keep only the solution with the *best distribution* of leftover capacities. For any two distributions of leftover capacities $\mathcal{F}_1$ and $\mathcal{F}_2$, we would prefer $\mathcal{F}_1$ over $\mathcal{F}_2$ if and only if there is a binary relation $\mathcal{F}_1 \succeq \mathcal{F}_2$, which is defined as the following.

Definition 1.1. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two probability distributions over $\mathbb{R}$. We say that $\mathcal{F}_1 \succeq \mathcal{F}_2$ if $\mathbb{P}_{C_1 \sim \mathcal{F}_1}(C_1 \geq c) \geq \mathbb{P}_{C_2 \sim \mathcal{F}_2}(C_2 \geq c), \forall c \in \mathbb{R}$.*

It should be clear that, the binary relation $\succeq$ as defined above is in general a partial order, but not a total order, on the set of the distributions of leftover capacities. We then have the problem of which partial solution (and therefore which distribution of leftover capacity) should be kept when moving to the next period. To guarantee the approximation factor, it is likely that we would want to keep more than one (or even infinitely many) partial solutions on the Pareto frontier with respect to the partial order, all of which earn profit p at time t. The total number of candidate solutions could easily grow superexponentially in T.

Given these difficulties, we only focus on rewards and not profits when building candidate solutions. Each candidate solution is composed of T subsets of items where items in each subset have the same deadline. For every period $t \in [T]$, by adopting the result on 0-1 Knapsack problem, we obtain $\tilde{\mathcal{O}}(1/\epsilon)$ number of sets, each achieving approximately best reward with minimum total size. A candidate solution is then built by taking one set of items in each period. As a result, there are $\tilde{\mathcal{O}}\left(1/\epsilon^T\right)$ number of candidate solutions. Assuming that for any set of items $\mathcal{S}$, its profit $\mathcal{P}(\mathcal{S})$ can be returned immediately. We simply find the best solution among the $\tilde{\mathcal{O}}\left(1/\epsilon^T\right)$ candidates that returns the highest profit. We present this algorithm as Algorithm 1.9, and the approximation result as Theorem 1.7.

Algorithm 1.9 Parameterized approximation for MPBKP-SS

Input: $[n], \mathcal{P}(\cdot)$ ▷ Set of items to be packed, profit function that returns the profit of any given set
Output: $\mathcal{S}'$

for $t = 1, \ldots, T$ **do**
  Obtain $\tilde{f}_{\mathcal{I}(t)}$ that approximates $f_{\mathcal{I}(t)}$ with factor $1+\epsilon$ using Lemma 1.1
  $l_t \leftarrow$ complexity of $\tilde{f}_{\mathcal{I}(t)}$ ▷ $l_t = \tilde{\mathcal{O}}\left(\frac{1}{\epsilon}\right)$
  Save the "steps" of $\tilde{f}_{\mathcal{I}(t)}$ as l_t pairs $\{(C_1^t, R_1^t), (C_2^t, R_2^t), \ldots, (C_l^t, R_{l_t}^t))\}$
  $\mathcal{S}_0^t \leftarrow \emptyset$
  for $k = 1, \ldots, l_t$ **do**
    $\mathcal{S}_k^t :=$ set of items corresponding to (C_k^t, R_k^t)
  end for
end for
$\mathcal{S}' \leftarrow \arg\max_{\left\{ \substack{\mathcal{S}=\cup_{s=1}^{t} \mathcal{S}_{k_s}^s \\ k_s \in \{0,1,\ldots,l_s\}} \right\}} \mathcal{P}(\mathcal{S})$

Theorem 1.7. *Under Assumption 1.2, Algorithm 1.9 achieves* $\left(1+\frac{\epsilon}{1-\beta}\right)$*-approximation factor for MPBKP-SS with runtime* $\tilde{\mathcal{O}}\left(n + 1/\epsilon^T\right)$.

The formal proof of Theorem 1.7 is left to Appendix 1.7.3.2.

1.6 Comments and Future Directions

The current work represents to the best of our knowledge the first FPTAS and theoretical guarantees for multi-period variants of the classical knapsack problem. For MPBKP, we obtained the runtime $\tilde{\mathcal{O}}\left(n + (T^{3.25}/\epsilon^{2.25})\right)$. This was done via the function approximation approach, where a function is approximated for each period. The runtime increases in T since we conduct T number of rounding downs, one after each $(\max, +)$-convolution. An alternative algorithm with runtime $\tilde{\mathcal{O}}\left(n + \frac{T^2}{\epsilon^3}\right)$ is also provided. For MPBKP-S, we obtained an FPTAS with runtime $\mathcal{O}\left(\frac{nT\log n}{\epsilon^2}\right)$, as well as a parameterized algorithm with approximation factor $\left(1+\frac{\epsilon}{1-\beta}\right)$ and runtime $\tilde{\mathcal{O}}\left(n + (T^{3.25}/\epsilon^{2.25})\right)$. For MPBKP-SS, the same approximation factor $\left(1+\frac{\epsilon}{1-\beta}\right)$ can be achieved in $\tilde{\mathcal{O}}\left(n + (1/\epsilon^T)\right)$, and a greedy algorithm achieves 2-approximation for the special case that all items have size 1.

There are a number of research directions in this area that could be pursued in the future, and we mention here a few of them. For MPBKP, note that the function we approximated

is in the same form as used in the 0-1 knapsack problem [46]. It is thus interesting to ask if we could instead directly approximate the following function:

$$f_{\mathcal{I}}(c) = \max_{x} \left\{ \sum_{i \in \mathcal{I}} r_i x_i \ : \sum_{i \in \cup_{t'=1}^{t} \mathcal{I}(t')} q_i x_i \leq c_t, \forall t \in [T], \ x \in \{0,1\}^n \right\},$$

where $\mathcal{I} = \cup_{t=1}^{T} \mathcal{I}(t)$ and $c = \{c_1, \ldots, c_T\}$ is a T-dimensional vector. Here we impose all T constraints in the function. The hope is that, if the above function could be approximated, and if we could properly define the $(\max, +)$-convolution on T dimensional vectors (and have a fairly easy computation of it), then we may get an algorithm that depends more mildly on T.

For MPBKP-S and MPBKP-SS, there seems to be less we can do without further assumptions. One direction to explore is parameterized approximation schemes: assuming that in the optimal solution, the total (expected) penalty is at most β fraction of the total reward. Then we may just focus on rewards. Our ongoing work suggests that an approximation factor of $\left(1 + \frac{\epsilon}{1-\beta}\right)$ may be achieved in $\tilde{\mathcal{O}}\left(n + (T^{3.25}/\epsilon^{2.25})\right)$ for MPBKP-S, and the same approximation factor in $\tilde{\mathcal{O}}\left(n + \frac{1}{\epsilon^T}\right)$ for MPBKP-SS.

We further note that the objective function for the three multiperiod variants are in fact submodular (but not non-negative, or monotone). Whether we can get a constant competitive solution in time $\widetilde{\mathcal{O}}(n)$, using approaches in submodular function maximization, is also an intriguing open problem.

Finally, motivated by applications, one natural extension that the authors are working on now is when there is a general non-decreasing cost function $\phi_t(\Delta c)$ for procuring capacity Δc at time t, and the goal is to admit a profit maximizing set of items when the unused capacity can be carried forward. Another extension is when there is a bound on the leftover capacity that can be carried forward, and we wonder how much the current results would change. All of these are interesting directions that may be worth exploring.

1.7 Appendix

1.7.1 Proofs for Section 1.3

Proof of Proposition 1.1. We show that the solution corresponding to $f_T(c)$ is optimal for $c_T = c$ among all solutions feasible to (1.1). We prove by induction on T. Base case is $T = 1$, this reduces to 0-1 Knapsack problem, and by definition, the solution corresponding to $f_{\mathcal{I}(1)}(c)$ is the optimal feasible solution when the Knapsack capacity is c. For the induction step, assume that the solution of $f_{T-1}(c')$ is the optimal feasible solution to (1.1) for the $T-1$ period problem and $c_{T-1} = c'$, we show that the solution corresponding to $f_T(c)$ is also the optimal feasible solution to (1.1) for the T period problem and $c_T = c$.

By definition,

$$f_T(c) = \left(\left(f_{T-1} \oplus f_{\mathcal{I}(T)}\right)(c)\right)^{c_T} = \left(\max_{c' \in \mathbb{R}} \left(f_{T-1}(c') + f_{\mathcal{I}(T)}(c - c')\right)\right)^{c}.$$

We first show that $f_T(c)$ is at least the optimal value of (1.1) when $c_T = c$. Suppose that, in the optimal solution of (1.1), the total size of accepted items up to time $T-1$ is $\hat{c}$ with $\hat{c} < c$, then the optimal value is $f_{T-1}(\hat{c}) + f_{\mathcal{I}(T)}(c - \hat{c})$ since $f_{T-1}(\hat{c})$ is the maximum achievable reward with $c_{T-1} = \hat{c}$ (by induction assumption) and $f_{\mathcal{I}(T)}(c-\hat{c})$ is the maximum achievable reward using items from $\mathcal{I}(T)$ with space constraint $c - \hat{c}$. Thus, we have that the optimal value $f_{T-1}(\hat{c}) + f_{\mathcal{I}(T)}(c - \hat{c}) \leq \left(\max_{c' \in \mathbb{R}} \left(f_{T-1}(c') + f_{\mathcal{I}(T)}(c - c')\right)\right)^{c} = f_T(c)$.

We next show the other direction: the optimal value of (1.1) for the T period problem with $c_T = c$ is at least $f_T(c)$. It suffices to show that every possible solution considered in $f_T(c)$ satisfies the feasibility constraints in (1.1). By induction assumption, every solution of $f_{T-1}(c')$ satisfies the constraints up to time $T-1$. When computing $f_T(c)$, we note that since $f_{T-1}(c')$ is a function truncated at c_{T-1}, which implies that $f_{T-1}(c') = -\infty$ for any $c' > c_{T-1}$. Therefore, any $c' > c_{T-1}$ must not be in the solution of $\max_{c' \in \mathbb{R}} \left(f_{T-1}(c') + f_{\mathcal{I}(T)}(c - c')\right)$. As a result, every solution of $f_T(c)$ is enforcing that

$c' \leq c_{T-1}$, and satisfies the feasibility constraints up to time T.

Combining both directions, we conclude the induction step, and thus the proof of the proposition. □

Proof of Lemma 1.2. By the construction of $\tilde{f}_t$, it should be clear that $\tilde{f}_t \leq f_t$. We prove that $(1+\epsilon)^t \tilde{f}_t \geq f_t$ by induction on t. Base case is when $t = 1$, we have that $(1+\epsilon)\tilde{f}_1 = (1+\epsilon)\tilde{f}_{\mathcal{I}(1)}^{c_1} \geq f_{\mathcal{I}(1)}^{c_1} = f_1$, where the inequality follows from Lemma 1.1. As for the induction step, assume that $(1+\epsilon)^{t-1}\tilde{f}_{t-1} \geq f_{t-1}$, we show that $(1+\epsilon)^t \tilde{f}_t \geq f_t$. Again, by Lemma 1.1 we have that

$$(1+\epsilon)^{t-1}\tilde{f}_{\mathcal{I}(t)} \geq (1+\epsilon)\tilde{f}_{\mathcal{I}(t)} \geq f_{\mathcal{I}(t)}.$$

Combined with the induction hypothesis, we have that

$$(1+\epsilon)^{t-1}\left(\tilde{f}_{t-1} \oplus \tilde{f}_{\mathcal{I}(t)}\right) = \left((1+\epsilon)^{t-1}\tilde{f}_{t-1}\right) \oplus \left((1+\epsilon)^{t-1}\tilde{f}_{\mathcal{I}(t)}\right) \geq f_{t-1} \oplus f_{\mathcal{I}(t)}.$$

Taking truncation on both sides, we have that

$$(1+\epsilon)^{t-1}\hat{f}_t = (1+\epsilon)^{t-1}\left(\tilde{f}_{t-1} \oplus \tilde{f}_{\mathcal{I}(t)}\right)^{c_t} \geq \left(f_{t-1} \oplus f_{\mathcal{I}(t)}\right)^{c_t} = f_t.$$

Because of rounding down, we have that $(1+\epsilon)\tilde{f}_t \geq \hat{f}_t$. Therefore,

$$(1+\epsilon)^t \tilde{f}_t \geq (1+\epsilon)^{t-1}\hat{f}_t \geq f_t.$$

This concludes the induction step, and thus the proof of the lemma. □

1.7.2 Proofs for Section 1.4

This subsection consists of two parts. The first part is devoted to the proof of Theorem 1.4, while the second part is devoted to the proof of Theorem 1.5.

1.7.2.1 Proof of Theorem 1.4

In this part, we prove Theorem 1.4. To proceed, we first present the following result on Algorithm 1.3.

Lemma 1.6. *Given a set of partial solutions with leftover capacities $\widetilde{A}(p)$ for all $p \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil \right\} \cdot \kappa$, the additional capacity available for packing Δc, and the set of large items to be added $\mathcal{I}_L := \{1, \ldots, I_L\}$, the output of Algorithm 1.3, $\hat{A}(I_L, p)$, satisfies:*

$$\hat{A}(I_L, p) = \max_{\left\{ \begin{array}{c} \mathcal{I}', \bar{p} \,:\, \mathcal{I}' \subseteq \mathcal{I}_L \\ \Delta\hat{\mathcal{P}}(\mathcal{I}', \widetilde{A}(\bar{p}) + \Delta c) \geq p - \bar{p} \\ \bar{p} \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil \right\} \cdot \kappa \end{array} \right\}} \widetilde{A}(\bar{p}) + \Delta c - \mathcal{Q}(\mathcal{I}'), \quad \forall p. \tag{1.17}$$

That is, $\hat{A}(I_L, p)$ is the maximum leftover capacity for any solution with (rounded) profit at least p obtained by adding items in $\mathcal{I}_L$ to the solutions corresponding to $\widetilde{A}(\cdot)$.

Proof of Lemma 1.6. We will prove a more general result than (1.17), i.e.,

$$\hat{A}(i, p) = \max_{\left\{ \begin{array}{c} \mathcal{I}', \bar{p} \,:\, \mathcal{I}' \subseteq \{1, \ldots, i\} \\ \Delta\hat{\mathcal{P}}(\mathcal{I}', \widetilde{A}(\bar{p}) + \Delta c) \geq p - \bar{p} \\ \bar{p} \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil \right\} \cdot \kappa \end{array} \right\}} \widetilde{A}(\bar{p}) + \Delta c - \mathcal{Q}(\mathcal{I}'), \quad \forall p \tag{1.18}$$

We prove this by induction. The base case ($i = 0$) is vacuously true. Now we assume that (1.18) holds for all $p \in \left\{0, 1, \ldots, \lceil 16T/\epsilon^2 \rceil\right\} \kappa$ and for all $k \in [i-1]$. Consider some $p \in \left\{0, 1, \ldots, \lceil 16T/\epsilon^2 \rceil\right\} \kappa$, and let $\mathcal{I}^*$ be any set achieving the maximum in (1.18) so that $\hat{P}(\mathcal{I}^*) \geq p - \bar{p}$ for some $\bar{p} \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil \right\} \cdot \kappa$. We will show that $\hat{A}(i, p)$ is at least the leftover capacity under solution $\mathcal{I}^*$ via case analysis:

- Case $i \notin \mathcal{I}^*$: In this case, the leftover capacity under $\mathcal{I}^*$ is the leftover capacity by d_i, which is the sum of leftover capacity in $\mathcal{I}^*$ by d_{i-1} and $c_{d_i} - c_{d_{i-1}}$. By induction hypothesis, $\hat{A}(i-1, p)$ is no less than the leftover capacity of $\mathcal{I}^*$ by d_{i-1}, and therefore,

by lines 4 and 8, $\hat{A}(i,p) \geq \hat{A}(i-1,p) + c_{d_i} - c_{d_{i-1}}$ which in turn is no less than the leftover capacity under $\mathcal{I}^*$ by d_i. By optimality of $\mathcal{I}^*$, all the inequalities must be equalities.

- Case $i \in \mathcal{I}^*$: Let $\mathcal{I}' = \mathcal{I}^* \setminus \{i\}$, and let $p' = \hat{\mathcal{P}}(\mathcal{I}')$ be its rounded profit. Then by induction hypothesis, $\hat{A}(i-1,p')$ is no less than the leftover capacity under $\mathcal{I}'$ by d_{i-1}. Further, by packing item i in the solution corresponding to $\hat{A}(i-1,p')$, the change in profit is larger than by packing item i in $\mathcal{I}'$ (the penalty is no less under $\mathcal{I}'$ since it has weakly smaller leftover capacity). Therefore, packing item i in the solution corresponding to $\hat{A}(i-1,p')$ gives a solution with at least as large a rounded profit as p and at least as much leftover capacity by d_i as $\mathcal{I}^*$. Therefore, in turn $\hat{A}(i,p)$ is at least as much as the leftover capacity in $\mathcal{I}^*$. Since we assume $\mathcal{I}^*$ to have the largest leftover capacity with profit at least p, all the inequalities must be equalities.

This completes the induction step, and thus the proof of the lemma. □

Next, we have the following Lemma as a preparation for our result on $\widetilde{A}(p)$ of Algorithm 1.4.

Lemma 1.7. *Given some capacity c and a set of small items $\mathcal{I}_S$ with $p_{max} := \max_{i \in \mathcal{I}_S} p_i$, let $\mathcal{S}^*$ be the profit-optimal subset, i.e., $\mathcal{S}^* = \arg\max_{\mathcal{S} \subseteq \mathcal{I}_S} \mathcal{P}(\mathcal{S}) = \mathcal{R}(\mathcal{S}) - B\,(\mathcal{Q}(\mathcal{S}) - c)^+$. Further, let $\widetilde{\mathcal{I}}_S := \{i \in \mathcal{I}_S \mid q_i \leq c\}$ and relabel the items in $\widetilde{\mathcal{I}}_S$ as $\left\{1', \ldots, |\widetilde{\mathcal{I}}_S|'\right\}$ (in decreasing order of reward density r_i/q_i). Let i' be such that $\sum_{j'=1'}^{i'} q_{j'} \leq c$ and $\sum_{j'=1'}^{(i+1)'} q_{j'} > c$. Then, the solution $\mathcal{S}' := \{1', \ldots, i'\}$ satisfies*

- $\mathcal{Q}(\mathcal{S}') \leq \mathcal{Q}(\mathcal{S}^*)$,
- $\mathcal{P}(\mathcal{S}') \geq \mathcal{P}(\mathcal{S}^*) - p_{max}$.

Proof of Lemma 1.7. The first item can be shown by contradiction. Suppose that to the contrary $\mathcal{Q}(\mathcal{S}') > \mathcal{Q}(\mathcal{S}^*)$, that is, $\mathcal{S}'$ uses more space than $\mathcal{S}^*$. Since the items in $\mathcal{S}'$ have

the highest reward densities, it is in fact the optimal solution which uses space $\mathcal{Q}(\mathcal{S}') < c$. Since the optimal profit is non-decreasing in the capacity c, this violates optimality of $\mathcal{S}^*$.

To see the second item, we look at two different cases. First, if $\mathcal{S}^* \cap \left(\mathcal{I}_S \setminus \widetilde{\mathcal{I}_S}\right) \neq \emptyset$, i.e., the optimal packing $\mathcal{S}^*$ includes some item i^* with $q_{i^*} > c$, then, there should be only one item in $\mathcal{S}^*$, i.e., $\mathcal{S}^* = \{i^*\}$. In this case, $\mathcal{P}(\mathcal{S}^*) = p_{i^*} = p_{max}$ and thus $\mathcal{P}(\mathcal{S}') \geq \mathcal{P}(\emptyset) = 0 = \mathcal{P}(\mathcal{S}^*) - p_{max}$.

Second, if $\mathcal{S}^* \cap \left(\mathcal{I}_S \setminus \widetilde{\mathcal{I}_S}\right) = \emptyset$, then $\mathcal{S}^* = \arg\max_{\mathcal{S} \subseteq \widetilde{I}_S} \mathcal{P}(\mathcal{S})$. Note that $\mathcal{P}(\mathcal{S}^*)$ is upper bounded by the reward for the fractional packing: $\mathcal{P}(\mathcal{S}^*) \leq \mathcal{R}_{LP} := \mathcal{R}(\mathcal{S}') + r_{(i+1)'} \cdot \frac{c - \mathcal{Q}(\mathcal{S}')}{q_{(i+1)'}} \leq \mathcal{R}(\mathcal{S}') + r_{(i+1)'} = \mathcal{P}(\mathcal{S}') + p_{(i+1)'} \leq \mathcal{P}(\mathcal{S}') + p_{max}$.

In either cases, we conclude that $\mathcal{P}(\mathcal{S}') \geq \mathcal{P}(\mathcal{S}^*) - p_{max}$. □

Before presenting our result on $\widetilde{A}_t(p)$, we will need the following definitions. For a solution $\mathcal{S} = \mathcal{S}(1) \cup \mathcal{S}(2) \cup \cdots \cup \mathcal{S}(T)$ with $\mathcal{S}(t) = \mathcal{S}_L(t) \cup \mathcal{S}_S(t)$, denoting the items with deadline t in $\mathcal{S}$, let the large items be indexed as $\mathcal{S}_L(t) = (i_1^{(t)}, \ldots, i_{L_t}^{(t)})$ in the order in which Algorithm 1.3 considers them, and the small items be indexed arbitrarily $\mathcal{S}_S(t) = \left(j_1^{(t)}, \ldots, j_{S_t}^{(t)}\right)$. Let $\mathcal{S}_L := \mathcal{S}_L(1) \cup \cdots \cup \mathcal{S}_L(T)$ and $\mathcal{S}_S := \mathcal{S}_S(1) \cup \cdots \cup \mathcal{S}_S(T)$ denote the large and small items in $\mathcal{S}$, respectively (this depends on the choice of P_0 but we suppress the dependence for brevity). We define the *rounded profit* of $\mathcal{S}$ as:

$$\tilde{\mathcal{P}}(\mathcal{S}) = \hat{\mathcal{R}}(\mathcal{S}_L) - \sum_{t=1}^{T} \sum_{k=1}^{L_t} \left\lceil B \left(\sum_{\ell \leq k} q_{i_\ell^{(t)}} - \max_{0 \leq t' < t} \left\{ c_t - c_{t'} - \sum_{t'+1 \leq \tau < t} \mathcal{Q}(\mathcal{S}(\tau)) \right\} \right)^+ \right\rceil_\kappa + \sum_{t=1}^{T} \left\lfloor \mathcal{R}(\mathcal{S}_S(t)) - B \left(\mathcal{Q}(\mathcal{S}(t)) - \max_{0 \leq t' < t} \left\{ c_t - c_{t'} - \sum_{t'+1 \leq \tau < t} \mathcal{Q}(\mathcal{S}(\tau)) \right\} \right)^+ \right\rfloor_\kappa . \tag{1.19}$$

That is, we add the rounded rewards of the large items, and for small items, we first group the small items by their deadlines, and for each deadline we round the sum of unrounded

rewards of small item. Further, let

$$\widetilde{C}_t(p) := \max_{\{\mathcal{S} \subseteq \bigcup_{t'=1}^{t} \mathcal{I}(t') \,:\, \tilde{\mathcal{P}}(\mathcal{S}) \geq p\}} \max_{0 \leq t' < t} \left\{ c_t - c_{t'} - \sum_{t'+1 \leq \tau \leq t} \mathcal{Q}(\mathcal{S}(\tau)) \right\}$$

denote the feasible partial solution with largest leftover capacity at time t and rounded total profit at least p. Then, we have the following lemma.

Lemma 1.8. *For any $t = 1, \ldots, T$ and any $p' \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$, we have that $\widetilde{A}_t(p) \geq \widetilde{C}_t(p')$ for some $p \geq p' - \frac{1}{2T}\epsilon P_0 t - \kappa t \geq r' - \frac{1}{2T}\epsilon(1-\epsilon/4)P_0 t$. That is, for any rounded total profit p' by time t, there exists some partial solution $\widetilde{A}_t$ of Algorithm 1.5 which has at least as much leftover capacity at time t the optimal solution $\widetilde{C}_t(p')$, and has rounded profit p not too much smaller than p'.*

Proof of Lemma 1.8. We prove by induction on t. Base case is when $t = 1$. Let $\mathcal{S}'$ be the solution corresponding to $\widetilde{C}_1(p')$, i.e., $\mathcal{S}' := \arg\max_{\left\{\begin{smallmatrix} \mathcal{S} \subseteq \mathcal{I}(1) \\ \tilde{\mathcal{P}}(\mathcal{S}) \geq p' \end{smallmatrix}\right\}} c_1 - \mathcal{Q}(\mathcal{S})$, and let $\mathcal{S}'_L = \mathcal{S}' \cap \mathcal{I}_L$, $\mathcal{S}'_S = \mathcal{S}' \cap \mathcal{I}_S$. Then $\tilde{\mathcal{P}}(\mathcal{S}'_L) = \hat{\mathcal{P}}(\mathcal{S}'_L)$. By Lemma 1.6, $\hat{A}(I_L(1), \tilde{\mathcal{P}}(\mathcal{S}'_L))$ is the maximum leftover capacity using items in $\mathcal{I}_L(1)$ earning rounded profit $\tilde{\mathcal{P}}(\mathcal{S}'_L)$. Thus, $\hat{A}_1(\tilde{\mathcal{P}}(\mathcal{S}'_L)) = \hat{A}(I_L(1), \tilde{\mathcal{P}}(\mathcal{S}'_L)) \geq c_1 - \mathcal{Q}(\mathcal{S}'_L)$. Let $\mathcal{S}''_L$ be the solution corresponding to $\hat{A}_1(\tilde{\mathcal{P}}(\mathcal{S}'_L))$, and thus $\mathcal{Q}(\mathcal{S}''_L) \leq \mathcal{Q}(\mathcal{S}'_L)$. Consider appending the partial solution $\mathcal{S}''_L$ using items from $\mathcal{I}_S(1)$. Let $\mathcal{S}''_S$ be the small item set obtained by adding small items greedily in their reward densities, subject to the constraint that $\mathcal{Q}(\mathcal{S}''_S) \leq \mathcal{Q}(\mathcal{S}'_S)$. Then, by Lemma 1.7, with $\mathcal{S}''_S$ being the greedy solution, $\mathcal{Q}(\mathcal{S}'_S)$ being the capacity constraint and $\mathcal{S}'_S$ being the optimal filling of small items in $\mathcal{I}_S(1)$, we conclude that

$$\mathcal{P}(\mathcal{S}''_S) \geq \mathcal{P}(\mathcal{S}'_S) - \frac{1}{2T}\epsilon P_0.$$

Therefore, $p' = \tilde{\mathcal{P}}(\mathcal{S}') = \tilde{\mathcal{P}}(\mathcal{S}'_L \cup \mathcal{S}'_S) = \tilde{\mathcal{P}}(\mathcal{S}'_L) + \Delta\tilde{\mathcal{P}}(\mathcal{S}'_S, c_1 - \mathcal{Q}(\mathcal{S}'_L)) \leq \tilde{\mathcal{P}}(\mathcal{S}''_L) + \Delta\tilde{\mathcal{P}}(\mathcal{S}''_S, c_1 - \mathcal{Q}(\mathcal{S}'_L)) + \frac{1}{2T}\epsilon P_0 + \kappa \leq \tilde{\mathcal{P}}(\mathcal{S}''_L) + \Delta\tilde{\mathcal{P}}(\mathcal{S}''_S, c_1 - \mathcal{Q}(\mathcal{S}''_L)) + \frac{1}{2T}\epsilon P_0 + \kappa = \tilde{\mathcal{P}}(\mathcal{S}''_L \cup \mathcal{S}''_S) + \frac{1}{2T}\epsilon P_0 + \kappa.$

Let $p = \tilde{\mathcal{P}}(\mathcal{S}''_L \cup \mathcal{S}''_S)$. From Algorithm 1.4, we know that since $\mathcal{S}''_S$ includes the small items in $\widetilde{\mathcal{I}}_S(1)$ with the highest reward densities, the solution $\mathcal{S}''_L \cup \mathcal{S}''_S$ is one feasible solution for $\widetilde{A}_1(p)$. We thus have that

$$\widetilde{A}_1(p) \geq c_1 - \mathcal{Q}(\mathcal{S}''_L \cup \mathcal{S}''_S) \geq c_1 - \mathcal{Q}(\mathcal{S}') = \widetilde{C}_1(p'),$$

where $p \geq p' - \frac{1}{2T}\epsilon R_0 - \kappa$, and the second inequality follows from the facts that $\mathcal{Q}(\mathcal{S}''_L) \leq \mathcal{Q}(\mathcal{S}'_L)$ and $\mathcal{Q}(\mathcal{S}''_S) \leq \mathcal{Q}(\mathcal{S}'_S)$.

For the induction step, assume that for all $p'' \in \left\{0, 1, \ldots, \left\lceil \frac{16T}{\epsilon^2} \right\rceil\right\} \cdot \kappa$, we have that $\widetilde{A}_{t-1}(p) \geq \widetilde{C}_{t-1}(p'')$ for some $p \geq p'' - \frac{1}{2T}\epsilon P_0(t-1) - \kappa(t-1)$. We want to show that for all p', $\widetilde{A}_t(p) \geq \widetilde{C}_t(p')$ for some $p \geq p' - \frac{1}{2T}\epsilon P_0 t - \kappa t$. Let $\mathcal{S}'$ be the solution corresponding to $\widetilde{C}_t(p')$, i.e.,

$$\mathcal{S}' := \arg \max_{\{\mathcal{S} \subseteq \bigcup_{t'=1}^{t} \mathcal{I}(t') \,:\, \tilde{\mathcal{P}}(\mathcal{S}) \geq p'\}} \max_{0 \leq t' < t} \left\{ c_t - c_{t'} - \sum_{t'+1 \leq \tau \leq t} \mathcal{Q}(\mathcal{S}(\tau)) \right\},$$

and let $\mathcal{S}'_L = \mathcal{S}' \cap \mathcal{I}_L$, $\mathcal{S}'_S = \mathcal{S}' \cap \mathcal{I}_S$. Let $\mathcal{S}'(t) := \{i \in \mathcal{S}' \mid d_i = t\}$ and consider the partial solution $\cup_{t'=1}^{t-1} \mathcal{S}'(t')$. By induction assumption, there exists some partial solution $\cup_{t'=1}^{t-1} \mathcal{S}''(t')$ which satisfies $\mathcal{Q}\left(\cup_{t'=1}^{t-1} \mathcal{S}''(t')\right) \leq \mathcal{Q}\left(\cup_{t'=1}^{t-1} \mathcal{S}'(t')\right)$, and $\tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1} \mathcal{S}''(t')\right) \geq \tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1} \mathcal{S}'(t')\right) - \frac{1}{2T}\epsilon P_0(t-1) - \kappa(t-1)$.

First, we fill the partial solution $\cup_{t'=1}^{t-1} \mathcal{S}''(t')$ using items from $\mathcal{I}_L(t)$ according to Algorithm 1.3. Note that one feasible solution is $\mathcal{S}'_L(t)$ which results in $\cup_{t'=1}^{t-1} \mathcal{S}''(t') \cup \mathcal{S}'_L(t)$. This keeps $\mathcal{Q}\left(\cup_{t'=1}^{t-1} \mathcal{S}''(t') \cup \mathcal{S}'_L(t)\right) \leq \mathcal{Q}\left(\cup_{t'=1}^{t-1} \mathcal{S}'(t') \cup \mathcal{S}'_L(t)\right)$ while we still have that $\tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1} \mathcal{S}''(t') \cup \mathcal{S}'_L(t)\right) \geq \tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1} \mathcal{S}'(t') \cup \mathcal{S}'_L(t)\right) - \frac{1}{2T}\epsilon P_0(t-1) - \kappa(t-1)$. Suppose that after filling items from $\mathcal{I}_L(t)$ using DP in Algorithm 1.3, the resulting set corresponding to $\hat{A}_t\left(\tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1} \mathcal{S}''(t') \cup \mathcal{S}'_L(t)\right)\right)$ is $\widetilde{\mathcal{S}}$, then this $\widetilde{\mathcal{S}}$ would only use less space and earn more

profit, i.e.,

$$\begin{aligned}
&\mathcal{Q}\left(\widetilde{\mathcal{S}}\right) \leq \mathcal{Q}\left(\cup_{t'=1}^{t-1}\mathcal{S}''(t') \cup \mathcal{S}'_L(t)\right) \leq \mathcal{Q}\left(\cup_{t'=1}^{t-1}\mathcal{S}'(t') \cup \mathcal{S}'_L(t)\right), \\
&\tilde{\mathcal{P}}\left(\widetilde{\mathcal{S}}\right) \geq \tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1}\mathcal{S}''(t') \cup \mathcal{S}'_L(t)\right) \\
&\qquad \geq \tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1}\mathcal{S}'(t') \cup \mathcal{S}'_L(t)\right) - \frac{1}{2T}\epsilon P_0(t-1) - \kappa(t-1).
\end{aligned}$$

Next, consider filling the partial solution $\widetilde{\mathcal{S}}$ using items from $\mathcal{I}_S(t)$. Let $\mathcal{S}''_S(t)$ be the small item set obtained by adding small items greedily in their reward densities, subject to the constraint that $\mathcal{Q}\left(\mathcal{S}''_S(t)\right) \leq \mathcal{Q}\left(\mathcal{S}'_S(t)\right)$. Then, by Lemma 1.7, with $\mathcal{S}''_S(t)$ being the greedy solution, $\mathcal{Q}(\mathcal{S}'_S(t))$ being the capacity constraint and $\mathcal{S}'_S(t)$ being the optimal filling of small items in $\mathcal{I}_S(t)$, we conclude that

$$\mathcal{P}(\mathcal{S}''_S(t)) \geq \mathcal{P}(\mathcal{S}'_S(t)) - \frac{1}{2T}\epsilon P_0.$$

Therefore,

$$\begin{aligned}
p' = \tilde{\mathcal{P}}(\mathcal{S}') &= \tilde{\mathcal{P}}\left(\cup_{t'=1}^{t-1}\mathcal{S}'(t') \cup \mathcal{S}'_L(t) \cup \mathcal{S}'_S(t)\right) \\
&\leq \tilde{\mathcal{P}}\left(\widetilde{\mathcal{S}} \cup \mathcal{S}''_S(t)\right) + \frac{1}{2T}\epsilon P_0(t-1) + \kappa(t-1) + \frac{1}{2T}\epsilon P_0 + \kappa \\
&\leq \tilde{\mathcal{P}}\left(\widetilde{\mathcal{S}} \cup \mathcal{S}''_S(t)\right) + \frac{1}{2T}\epsilon P_0 t + \kappa t.
\end{aligned}$$

Let $p = \tilde{\mathcal{P}}\left(\widetilde{\mathcal{S}} \cup \mathcal{S}''_S(t)\right)$. From Algorithm 1.4, we know that since $\mathcal{S}''_S(t)$ includes the small items in $\widetilde{\mathcal{I}}_S(t)$ with the highest reward densities, the solution $\widetilde{\mathcal{S}} \cup \mathcal{S}''_S(t)$ is one feasible solution for $\widetilde{A}_t(p)$. We thus have that

$$\widetilde{A}_t(p) \geq \max_{0 \leq t' < t} \left\{ c_t - c_{t'} - \sum_{t'+1 \leq \tau \leq t} \mathcal{Q}\left(\left(\widetilde{\mathcal{S}} \cup \mathcal{S}''_S(t)\right)(\tau)\right) \right\}$$

$$\geq \max_{0\leq t'<t} \left\{ c_t - c_{t'} - \sum_{t'+1\leq\tau\leq t} \mathcal{Q}(\mathcal{S}'(\tau)) \right\} = \widetilde{C}_t(p'),$$

where $p \geq p' - \frac{1}{2T}\epsilon P_0 t - \kappa t$. This finishes the induction step, and thus the proof of the lemma. □

Using the above lemmas, we prove the following approximation result.

Proposition 1.4. *Let $\mathcal{S}'$ denote the optimal solution set by Algorithm 1.5, i.e., $\mathcal{S}'$ is the solution set corresponding to $\widetilde{A}_T(p^*)$ where p^* is the maximum p such that $\widetilde{A}_T(p) > -\infty$. Let $\mathcal{S}^*$ be the optimal solution set to the original MPBKP-S. Then,*

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq (1 - \epsilon - 3\epsilon^2/8)\mathcal{P}(\mathcal{S}^*).$$

Proof. Note that $\tilde{\mathcal{P}}(\mathcal{S}') = p^*$. Lemma 1.8 implies that

$$\widetilde{A}_T(p^*) \geq \widetilde{C}_T\left(p^* + \frac{1}{2T}\epsilon P_0 T + \kappa T\right) = \widetilde{C}_T\left(p^* + \frac{1}{2}\epsilon P_0 + \kappa T\right).$$

Since $\widetilde{C}_T(\tilde{\mathcal{P}}(\mathcal{S}^*)) > -\infty$, we have that $\widetilde{A}_T\left(\tilde{\mathcal{P}}(\mathcal{S}^*) - \frac{1}{2}\epsilon P_0 - \kappa T\right) \geq \widetilde{C}_T(\tilde{\mathcal{P}}(\mathcal{S}^*)) > -\infty$. Therefore,

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq \tilde{\mathcal{P}}(\mathcal{S}^*) - \frac{1}{2}\epsilon P_0 - \kappa T.$$

By the definition of $\tilde{\mathcal{P}}$ as in (1.19), for each large item, the reward is rounded down by at most κ and the penalty is rounded up by at most κ, and all small items are together rounded down by at most κT. Note that each large items earns profit p_i unless it is paying more penalty than it would be by itself, which happens at most once at each period. Thus, there are at most $\frac{2P_0}{\frac{1}{2T}\epsilon P_0} + T = \frac{4T}{\epsilon} + T$ number of large items, and thus the total number of rounding downs (for both large and small items) is bounded by $\frac{4T}{\epsilon} + 2T$. Therefore, we

have that $\mathcal{P}(\mathcal{S}^*) \leq \tilde{\mathcal{P}}(\mathcal{S}^*) + \left(\frac{4T}{\epsilon} + 2T\right)\kappa$. In conclusion,

$$\begin{aligned}\mathcal{P}(\mathcal{S}') \geq p^* &\geq \tilde{\mathcal{P}}(\mathcal{S}^*) - \frac{1}{2}\epsilon P_0 - \kappa T \\ &\geq \mathcal{P}(\mathcal{S}^*) - \left(\frac{4T}{\epsilon} + 2T\right)\kappa - \frac{1}{2}\epsilon P_0 - T\kappa = \mathcal{P}(\mathcal{S}^*) - \epsilon P_0 - 3T\kappa \\ &\geq \left(1 - \epsilon - 3\epsilon^2/8\right)\mathcal{P}(\mathcal{S}^*).\end{aligned}$$

□

It remains to validate Algorithm 1.6 in the search of P_0 which satisfies (1.14). When Algorithm 1.6 terminates, it returns the last p^* and the solution set $\mathcal{S}'$ corresponding to $\widetilde{A}_T(p^*)$. We then have the following lemmas.

Lemma 1.9. *Algorithm 1.6 terminates within* $\log n$ *iterations of the "while" loop (line 3).*

Proof of Lemma 1.9. When P_0 satisfies (1.14), by Proposition 1.4 we have that

$$p^* \geq (1 - \epsilon)\mathcal{P}(\mathcal{S}^*) \geq (1 - \epsilon)P_0.$$

Thus, the "while" loop terminates when P_0 satisfies (1.14), if not before P_0 satisfies (1.14). When P_0 satisfies (1.14), we would also have $\mathcal{P}(\mathcal{S}^*)/2 \leq P_0 \leq \mathcal{P}(\mathcal{S}^*)$. Therefore, the number of iterations is upper bounded by

$$\text{number of iterations} \leq \log \frac{\bar{P}/2}{\mathcal{P}(\mathcal{S}^*)/2} \leq \log n,$$

where we have used the fact that $\bar{P} \leq nP \leq n\mathcal{P}(\mathcal{S}^*)$. □

Lemma 1.10. *After running Algorithm 1.6, suppose* $\mathcal{S}'$ *is the solution set corresponding to* $\widetilde{A}_T(p^*)$*, and* $\mathcal{S}^*$ *is the optimal solution set to the original MPBKP-S. Then,*

$$\mathcal{P}(\mathcal{S}') \geq (1 - \epsilon)\mathcal{P}(\mathcal{S}^*).$$

Proof of Lemma 1.10. If the "while" loop terminates when $P_0 > \mathcal{P}(\mathcal{S}^*)$, i.e., it stops before P_0 falls below $\mathcal{P}(\mathcal{S}^*)$, then we have that

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq (1-\epsilon)P_0 > (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

Otherwise, from the proof of Lemma 1.9 we know that the "while" loop must terminate when P_0 first falls below $\mathcal{P}(\mathcal{S}^*)$, which implies that the last P_0 satisfies (1.14). Then by Proposition 1.4 we again have that

$$\mathcal{P}(\mathcal{S}') \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

In either case, the solution we obtained from Algorithm 1.6 achieves $(1-\epsilon)$ optimal. □

With the above Lemmas, we are in a position to prove Theorem 1.4.

Proof of Theorem 1.4. By Lemma 1.10, the solution found is within $(1-\epsilon)$ factor of $\mathcal{P}(\mathcal{S}^*)$. Since the running time of the algorithm is $\mathcal{O}\left(n \cdot \left\lceil \frac{16T}{\epsilon^2} \right\rceil \cdot \log n\right) = \mathcal{O}\left(\frac{Tn \log n}{\epsilon^2}\right)$, which is polynomial in n and $1/\epsilon$, the theorem follows. □

1.7.2.2 Proof of Theorem 1.5

In this part, we prove Theorem 1.5. Recall that $\mathcal{S}(t) := \{i \in \mathcal{S} \mid d_i = t\}$, and $\mathcal{R}(\mathcal{S}) := \sum_{i \in \mathcal{S}} r_i$ is the total reward of set $\mathcal{S}$. Let $\Phi(\mathcal{S}) := \mathcal{R}(\mathcal{S}) - \mathcal{P}(\mathcal{S})$ be the total penalty on $\mathcal{S}$. For simplicity, we let $\mathcal{R}_t(\mathcal{S}) := \mathcal{R}(\mathcal{S}(t))$, $\mathcal{Q}_t(\mathcal{S}) := \mathcal{Q}(\mathcal{S}(t))$, $\mathcal{P}_t(\mathcal{S}) := \mathcal{P}\left(\cup_{t'=1}^{t} \mathcal{S}(t')\right)$. Further, we define the *incremental penalty* as follows:

$$\Phi_t(\mathcal{S}) := \begin{cases} \Phi(\mathcal{S}(1)), & t = 1, \\ \Phi\left(\cup_{t'=1}^{t} \mathcal{S}(t')\right) - \Phi\left(\cup_{t'=1}^{t-1} \mathcal{S}(t')\right), & t \geq 2. \end{cases} \tag{1.20}$$

In other words, $\Phi_t(\mathcal{S})$ is the additional penalty paid when adding $\mathcal{S}(t)$, items with deadline t, to $\cup_{t'=1}^{t-1}\mathcal{S}(t')$, items with deadlines before t. It should then be clear that

$$\mathcal{P}(\mathcal{S}) = \mathcal{R}(\mathcal{S}) - \Phi(\mathcal{S}) = \mathcal{R}_1(\mathcal{S}) - \Phi_1(\mathcal{S}) + \mathcal{R}_2(\mathcal{S}) - \Phi_2(\mathcal{S}) + \cdots + \mathcal{R}_T(\mathcal{S}) - \Phi_T(\mathcal{S}). \tag{1.21}$$

Given p_0, we let $\lfloor a \rfloor_{(1+\epsilon)} := p_0 \cdot (1+\epsilon)^{\left\lfloor \log_{1+\epsilon} \frac{a}{p_0} \right\rfloor}$ if $a \geq \frac{p_0}{1+\epsilon}$. We next define the *(exponentially) rounded profit up to time t* on set $\mathcal{S}$ as follows:

$$\hat{\mathcal{P}}_t(\mathcal{S}) := \begin{cases} \lfloor \mathcal{R}_1(\mathcal{S}) - \Phi_1(\mathcal{S}) \rfloor_{(1+\epsilon)} & t = 1, \\ \left\lfloor \hat{\mathcal{P}}_{t-1}(\mathcal{S}) + \mathcal{R}_t(\mathcal{S}) - \Phi_t(\mathcal{S}) \right\rfloor_{(1+\epsilon)} & t \geq 2. \end{cases} \tag{1.22}$$

In other words, $\hat{\mathcal{P}}_t(\mathcal{S})$'s are defined recursively: for $t = 1$, $\hat{\mathcal{P}}_1(\mathcal{S}) := \lfloor \mathcal{R}_1(\mathcal{S}) - \Phi_1(\mathcal{S}) \rfloor_{(1+\epsilon)}$; for $t \geq 2$, we define $\hat{\mathcal{P}}_t(\mathcal{S}) \left\lfloor \hat{\mathcal{P}}_{t-1}(\mathcal{S}) + \mathcal{R}_t(\mathcal{S}) - \Phi_t(\mathcal{S}) \right\rfloor_{(1+\epsilon)}$. The *rounded profit* on set $\mathcal{S}$ is then $\hat{\mathcal{P}}(\mathcal{S}) := \hat{\mathcal{P}}_T(\mathcal{S})$.

In Algorithm 1.7, we have used the result from Lemma 1.1 to obtain $\tilde{f}_{\mathcal{I}(t)}$ for all $t \in [T]$, each of which is a step function. Let the complexity (number of "steps") of $\tilde{f}_{\mathcal{I}(t)}$ be l_t. Then, $\tilde{f}_{\mathcal{I}(t)}(c)$ can be expressed as

$$\tilde{f}_{\mathcal{I}(t)}(c) = \begin{cases} 0, & c < C_1^t \\ R_k^t, & C_k^t \leq c < C_{(k+1)}^t, \ \forall k = 1, \ldots, l_t - 1, \\ R_l^t, & c \geq C_{l_t}^t. \end{cases}$$

Given the structure of $\tilde{f}_{\mathcal{I}(t)}(c)$, each function can be fully characterized with l_t number of size-reward pairs $\{(C_1^t, R_1^t), (C_2^t, R_2^t), \ldots, (C_{l_t}^t, R_{l_t}^t))\}$, where each pair corresponds to a "step" of the function. Lemma 1.1 implies that there exists some set of items with deadline t, which we denote by $\mathcal{S}_k^t$, such that $\mathcal{Q}(\mathcal{S}_k^t) = C_k^t$ and $R_k^t \leq \mathcal{R}(\mathcal{S}_k^t) \leq (1+\epsilon)R_k^t$, for all $k \in [l_t]$ and

$t \in [T]$. Further, let $\mathcal{S}_0^t = \emptyset$ for all $t \in [T]$.

We consider the problem of selecting one set from $\{\mathcal{S}_0^t, \ldots, \mathcal{S}_{l_t}^t\}$, denoted by $\mathcal{S}_{k_t}^t$, for each $t \in [T]$, such that $\hat{\mathcal{P}}(\cup_{t=1}^T \mathcal{S}_{k_t}^t) \geq p$ and that $\cup_{t=1}^T \mathcal{S}_{k_t}^t$ has the largest leftover capacity. Formally, we define

$$A_t(p) := \max_{\substack{k_s \in \{0,1,\ldots,l_s\},\ \forall s \in [t] \\ \mathcal{S} = \cup_{s=1}^t \mathcal{S}_{k_s}^s\ :\ \hat{\mathcal{P}}(\mathcal{S}) \geq p}} \max_{0 \leq t' < t} \left\{ c_t - c_{t'} - \sum_{t'+1 \leq \tau \leq t} \mathcal{Q}_\tau(\mathcal{S}). \right\} \tag{1.23}$$

Then, we have the following lemma as a result of the dynamic program in Algorithm 1.7.

Lemma 1.11. *For any $t = 1, \ldots, T$ and any $p = 0, p_0 \cdot (1+\epsilon)^{\{-1,0,1,\ldots,m\}}$, we have that*

$$\tilde{A}(t,p) = A_t(p). \tag{1.24}$$

Proof of Lemma 1.11. We prove by induction on t. Base case is $t = 1$. We have that

$$A_1(p) = \max_{\left\{ \substack{k \in \{0,1,\ldots,l_1\}, \\ \hat{\mathcal{P}}(\mathcal{S}_k^1) \geq p} \right\}} c_1 - \mathcal{Q}_1\left(\mathcal{S}_{k_1}^1\right).$$

For each $p = 0, p_0 \cdot (1+\epsilon)^{\{-1,0,1,\ldots,m\}}$, since the solution set corresponding to $\tilde{A}(1,p)$ is exactly one of $\mathcal{S}_k^1$ for $k \in \{0,1,\ldots,l_1\}$, it follows that $\tilde{A}(1,p) \leq A_1(p)$. On the other hand, let $\mathcal{S}_{k_1}^1$ be the set corresponding to $A_1(p)$ and $\mathcal{S}_{k'}^1$ be the one corresponding to $\tilde{A}(1,p)$. Note that $k_1 \in \{0,1,\ldots,l_1\}$, which implies that during the "for" loop of line 12 to line 17 of Algorithm 1.7, k_1 has been visited. Since $\hat{\mathcal{P}}_1(\mathcal{S}_{k_1}^1) = \hat{\mathcal{P}}_1(\mathcal{S}_{k'}^1) = p$, by line 15, we have that $\tilde{A}(1,p) \geq c_1 - \mathcal{Q}_1\left(\mathcal{S}_{k_1}^1\right) = A_1(p)$. Therefore, we conclude that $\tilde{A}(1,p) = A_1(p)$.

As for the induction step, assume that $\tilde{A}(t',p) = A_{t'}(p)$ for all $p = 0, p_0 \cdot (1+\epsilon)^{\{-1,0,1,\ldots,m\}}$ at all $t' < t$, we show that $\tilde{A}(t,p) = A_t(p)$. Consider an arbitrary p, and notice that the solution corresponding to $\tilde{A}(t,p)$, denoted by $\mathcal{S}$, can be written as $\mathcal{S} = \cup_{s=1}^t \mathcal{S}_{k_s}^s$ with $k_s \in \{0,1,\ldots,l_s\},\ \forall s \in [t]$, which falls into the feasible range of (1.23). Thus, we have that

$\tilde{A}(t,p) \leq A_t(p)$. On the other hand, let $\mathcal{S}' = \cup_{s=1}^{t} \mathcal{S}_{k'_s}^s$ be the set corresponding to $A_t(p)$. Then, we have that

$$p = \hat{\mathcal{P}}(\mathcal{S}') = \left\lfloor \hat{\mathcal{P}}_{t-1}\left(\mathcal{S}'\right) + \mathcal{R}_t(\mathcal{S}') - \Phi_t(\mathcal{S}') \right\rfloor_{(1+\epsilon)},$$

where $\hat{\mathcal{P}}_{t-1}(\mathcal{S}')$ is the rounded profit of $\cup_{t'=1}^{t-1} \mathcal{S}'(t') = \cup_{s=1}^{t-1} \mathcal{S}_{k'_s}^s$. Note that from the definition (1.23), $\cup_{s=1}^{t-1} \mathcal{S}_{k'_s}^s$ must have the leftover capacity $A_{t-1}\left(\hat{\mathcal{P}}_{t-1}(\mathcal{S}')\right)$ at time $t-1$, because otherwise we could let $\bar{\mathcal{S}}$ be the one that have leftover capacity $A_{t-1}\left(\hat{\mathcal{P}}_{t-1}(\mathcal{S}')\right)$, and then $\bar{\mathcal{S}} \cup \mathcal{S}_{k'_t}^t$ would have more leftover capacity than $A_t(p)$, which is a contradiction. Therefore, by the induction assumption, $\cup_{s=1}^{t-1} \mathcal{S}_{k'_s}^s$ is also the set corresponding to $\tilde{A}\left(t-1, \hat{\mathcal{P}}_{t-1}(\mathcal{S}')\right)$. Since $k_t \in \{0, 1, \ldots, l_t\}$, during the "for" loop of line 12 to line 17 of Algorithm 1.7, k_t has been visited. We should have that $\tilde{A}(t,p) \geq A_t(p)$, and so $\tilde{A}(t,p) = A_t(p)$. This concludes the induction step, and thus the proof of the lemma. □

Let $\mathcal{S}^*$ be the optimal solution for the MPBKP-S. By Lemma 1.1, we can find sets $\mathcal{S}_{k_t}^t$ for all $t \in [T]$, which consists of only items with deadline t, with $k_t \in \{0, 1, \ldots, l_t\}$, such that $\mathcal{Q}\left(\mathcal{S}_{k_t}^t\right) \leq \mathcal{Q}_t(\mathcal{S}^*)$ and that $\mathcal{R}\left(\mathcal{S}_{k_t}^t\right) \leq \mathcal{R}(\mathcal{S}^*) \leq (1+\epsilon)\mathcal{R}\left(\mathcal{S}_{k_t}^t\right)$. Then, we have the following lemma.

Lemma 1.12. *Let $\tilde{f}_{\mathcal{I}(t)}$ be the approximated function obtained from Lemma 1.1, and let the complexity of $\tilde{f}_{\mathcal{I}(t)}$ be l_t. Further, let $\{(C_1^t, R_1^t), (C_2^t, R_2^t), \ldots, (C_l^t, R_{l_t}^t))\}$ be the l_t steps of the function. For S^* an optimal solution of MPBKP-S, let $k_t := \max_{\{k \in [l_t]:\ C_k^t \leq \mathcal{Q}_t(\mathcal{S}^*)\}} k$ and $\mathcal{S}_{k_t}^t$ be the set corresponding to $(C_{k_t}^t, R_{k_t}^t)$. If $C_k^t > \mathcal{Q}_t(\mathcal{S}^*)$ for all $k \in [l_t]$, let $\mathcal{S}_{k_t}^t := \emptyset$. Further, let $\mathcal{S}'' := \cup_{t=1}^{T} \mathcal{S}_{k_t}^t$, then, we have the following:*

$$(1+\epsilon)\hat{\mathcal{P}}(\mathcal{S}'') \geq \frac{\mathcal{R}(\mathcal{S}^*)}{(1+\epsilon)^T} - \Phi(\mathcal{S}^*). \tag{1.25}$$

Proof of Lemma 1.12. We will prove a more general result than (1.12), i.e.,

$$(1+\epsilon)\hat{\mathcal{P}}_t(\mathcal{S}'') \geq \frac{\sum_{t'=1}^{t} \mathcal{R}_{t'}(\mathcal{S}^*)}{(1+\epsilon)^t} - \sum_{t'=1}^{t} \Phi_{t'}(\mathcal{S}^*). \tag{1.26}$$

We prove this by induction on t. The base case is when $t = 1$, in which case we have that

$$(1+\epsilon)\hat{\mathcal{P}}_1(\mathcal{S}'') \geq \mathcal{P}_1(\mathcal{S}'') = \mathcal{R}_1(\mathcal{S}'') - \Phi_1(\mathcal{S}'') = \mathcal{R}\left(\mathcal{S}^t_{k_t}\right) - \Phi\left(\mathcal{S}^t_{k_t}\right) \geq \frac{\mathcal{R}_1(\mathcal{S}^*)}{1+\epsilon} - \Phi_1(\mathcal{S}^*),$$

where the first inequality follows from the definition in (1.22), and the second inequality follows from the approximation error of $\tilde{f}_{\mathcal{I}(1)}$ to $f_{\mathcal{I}(1)}$.

As for the induction step, assume that (1.26) holds for all $t = 1, \ldots, \bar{t}$, we prove that it also holds for $t = \bar{t}+1$. Again, following the definition in (1.22) as well as the approximation of $\tilde{f}_{\mathcal{I}(1)}$ to $f_{\mathcal{I}(1)}$, we have that

$$\begin{aligned}
(1+\epsilon)\hat{\mathcal{P}}_{\bar{t}+1}(\mathcal{S}'') &\geq \hat{\mathcal{P}}_{\bar{t}}(\mathcal{S}'') + \mathcal{R}_{\bar{t}+1}(\mathcal{S}'') - \Phi_{\bar{t}+1}(\mathcal{S}'') \\
&\geq \frac{\sum_{t'=1}^{\bar{t}} \mathcal{R}_{t'}(\mathcal{S}^*)}{(1+\epsilon)^{\bar{t}}} - \sum_{t'=1}^{\bar{t}} \Phi_{t'}(\mathcal{S}^*) + \frac{\mathcal{R}_{\bar{t}+1}(\mathcal{S}^*)}{1+\epsilon} - \Phi_{\bar{t}+1}(\mathcal{S}^*) \\
&\geq \frac{\sum_{t'=1}^{\bar{t}+1} \mathcal{R}_{t'}(\mathcal{S}^*)}{(1+\epsilon)^{\bar{t}+1}} - \sum_{t'=1}^{\bar{t}+1} \Phi_{t'}(\mathcal{S}^*).
\end{aligned}$$

This completes the induction step, and thus the proof of the lemma. □

With the above lemmas, we are now in a position to prove Theorem 1.5.

Proof of Theorem 1.5. **Approximation ratio:** Let $\mathcal{S}'$ be the solution that corresponds to $\tilde{A}(T, p^*)$ with $\hat{\mathcal{P}}(\mathcal{S}') = p^*$, where p^* is the maximum p such that $\tilde{A}(T, p) > -\infty$, i.e., $p^* = \max_{\tilde{A}(T,p)>-\infty} p$. Then $\mathcal{S}'$ is our solution set from Algorithm 1.7. Let $\mathcal{S}''$ be as defined in Lemma 1.12. Since $\mathcal{Q}_t(\mathcal{S}'') \leq \mathcal{Q}_t(\mathcal{S}^*)$, we know that the leftover capacity of $\mathcal{S}''$ at $t = T$ is not $-\infty$, and is upper bounded by $\tilde{A}\left(T, \hat{\mathcal{P}}(\mathcal{S}'')\right)$ by Lemma 1.11 and (1.23), which implies

that $\tilde{A}\left(T, \hat{\mathcal{P}}(\mathcal{S}'')\right) > -\infty$, and thus $\hat{\mathcal{P}}(\mathcal{S}') = p^* \geq \hat{\mathcal{P}}(\mathcal{S}'')$. This, together with Lemma 1.12, lead to the following:

$$\hat{\mathcal{P}}(\mathcal{S}') \geq \hat{\mathcal{P}}(\mathcal{S}'') \geq \frac{\mathcal{R}(\mathcal{S}^*)}{(1+\epsilon)^{T+1}} - \frac{\Phi(\mathcal{S}^*)}{1+\epsilon}.$$

Recall that in Assumption 1.1, we assume that $\Phi(\mathcal{S}^*) \leq \beta\mathcal{R}(\mathcal{S}^*)$, which implies that $\mathcal{P}(\mathcal{S}^*) = \mathcal{R}(\mathcal{S}^*) - \Phi(\mathcal{S}^*) \geq \mathcal{R}(\mathcal{S}^*) - \beta\mathcal{R}(\mathcal{S}^*) = (1-\beta)\mathcal{R}(\mathcal{S}^*)$. Therefore, we have that

$$\hat{\mathcal{P}}(\mathcal{S}') \geq \frac{\mathcal{R}(\mathcal{S}^*)}{(1+\epsilon)^{T+1}} - \frac{\Phi(\mathcal{S}^*)}{1+\epsilon} \geq \frac{\mathcal{R}(\mathcal{S}^*)}{(1+\epsilon)^{T+1}} - \Phi(\mathcal{S}^*) \approx \frac{\mathcal{R}(\mathcal{S}^*)}{1+T\epsilon} - \Phi(\mathcal{S}^*).$$

Taking $\epsilon' = T\epsilon$, we have that

$$\begin{aligned}\hat{\mathcal{P}}(\mathcal{S}') &\geq \frac{\mathcal{R}(\mathcal{S}^*)}{1+\epsilon'} - \Phi(\mathcal{S}^*) = \mathcal{R}(\mathcal{S}^*) - \frac{\epsilon'\mathcal{R}(\mathcal{S}^*)}{1+\epsilon'} - \Phi(\mathcal{S}^*) = \mathcal{P}(\mathcal{S}^*) - \frac{\epsilon'\mathcal{R}(\mathcal{S}^*)}{1+\epsilon'} \\ &\geq \mathcal{P}(\mathcal{S}^*) - \frac{\epsilon'\mathcal{P}(\mathcal{S}^*)}{(1+\epsilon')(1-\beta)} \geq \left(1 - \frac{\epsilon'}{1-\beta}\right)\mathcal{P}(\mathcal{S}^*).\end{aligned}$$

Time complexity: Obtaining all $\tilde{f}_{\mathcal{I}(t)}$ takes $\tilde{\mathcal{O}}\left(n + T/\epsilon^{2.25}\right)$. The rest of the algorithm takes $\mathcal{O}(Tml) = \tilde{\mathcal{O}}(T/\epsilon)$. By taking $\epsilon' = T\epsilon$, we conclude that the algorithm has runtime $\tilde{\mathcal{O}}\left(n + \frac{T^{3.25}}{\epsilon'^{2.25}}\right)$ to achieve at least $\left(1 - \frac{\epsilon'}{1-\beta}\right)$ approximation. □

1.7.3 Proofs for Section 1.5

This subsection consists of two parts. The first part is devoted to the proof of Theorem 1.6, while the second part is devoted to the proof of Theorem 1.7.

1.7.3.1 Proof of Theorem 1.6

In this part, we prove Theorem 1.6. The idea is to look at the greedy solution set $\mathcal{S}_p$ and the optimal solution set $\mathcal{S}^*$, and by swapping each item in $\mathcal{S}_p$ to $\mathcal{S}^*$ in replacement of the same item or two other items, we construct a sequence of partial solutions of the greedy algorithm

as well as modified optimal solution set, while maintaining the invariant that the profit of $\mathcal{S}^*$ is bounded by the sum of two times the profit of items in $\mathcal{S}_p$ swapped into $\mathcal{S}^*$ so far and the additional profit of remaining items in the modified optimal solution set. We will make this clear in the following.

We first introduce some notations. Let $\mathcal{S}_p = \{g_1, \ldots, g_l\}$ and $\mathcal{S}^* = \{o_1, \ldots, o_m\}$, i.e., the items in greedy solution is denoted by g_i's and the items in the optimal solution is denoted by o_i's. Further, for any two sets of items $\mathcal{S}_1$ and $\mathcal{S}_2$, we define the incremental profit of adding $\mathcal{S}_2$ to the set $\mathcal{S}_1$ as

$$\Delta\mathcal{P}(\mathcal{S}_1, \mathcal{S}_2) = \mathcal{P}(\mathcal{S}_1 \cup \mathcal{S}_2) - \mathcal{P}(\mathcal{S}_1). \tag{1.27}$$

Recall that $\Phi(\mathcal{S})$ is the expected number of units of overflows that penalties are paid, which will be referred as *overflow units* in the following. The incremental expected overflow units of adding $\mathcal{S}_2$ to the set $\mathcal{S}_1$ is defined as

$$\Delta\Phi(\mathcal{S}_1, \mathcal{S}_2) = \Phi(\mathcal{S}_1 \cup \mathcal{S}_2) - \Phi(\mathcal{S}_1). \tag{1.28}$$

On a sample path of incremental capacities $\omega = \{c_t\}_{t=1}^T$, let $a_t := c_t - c_{t-1}$. Let $\mathcal{P}_\omega$ and Φ_ω be the profit and overflow units function, respectively, and the incremental profit of adding $\mathcal{S}_2$ to the set $\mathcal{S}_1$ is

$$\Delta\mathcal{P}_\omega(\mathcal{S}_1, \mathcal{S}_2) = \mathcal{P}_\omega(\mathcal{S}_1 \cup \mathcal{S}_2) - \mathcal{P}_\omega(\mathcal{S}_1).$$

Similarly, on sample path ω, the incremental penalty of adding $\mathcal{S}_2$ to the set $\mathcal{S}_1$ is

$$\Delta\Phi_\omega(\mathcal{S}_1, \mathcal{S}_2) = \Phi_\omega(\mathcal{S}_1 \cup \mathcal{S}_2) - \Phi_\omega(\mathcal{S}_1).$$

Then, the relationship of $\Delta\mathcal{P}$ and $\Delta\Phi$ is:

$$\begin{aligned}\Delta\mathcal{P}(\mathcal{S}_1, \mathcal{S}_2) &= \mathcal{P}(\mathcal{S}_1 \cup \mathcal{S}_2) - \mathcal{P}(\mathcal{S}_1) = \mathcal{R}(\mathcal{S}_1 \cup \mathcal{S}_2) - \mathcal{R}(\mathcal{S}_1) - B \cdot \Phi(\mathcal{S}_1 \cup \mathcal{S}_2) + B \cdot \Phi(\mathcal{S}_1) \\ &= \mathcal{R}(\mathcal{S}_2) - B \cdot \Delta\Phi(\mathcal{S}_1, \mathcal{S}_2).\end{aligned}$$

Similarly, on a sample path, we have that $\Delta\mathcal{P}_\omega(\mathcal{S}_1, \mathcal{S}_2) = \mathcal{R}(\mathcal{S}_2) - B \cdot \Delta\Phi_\omega(\mathcal{S}_1, \mathcal{S}_2)$.

Let $\mathcal{S}(t) := \{j \in \mathcal{S} \mid d_j = t\}$. Given a (partial) solution $\mathcal{S}$ and a sample path of capacities $\omega = \{c_t\}_{t=1}^T \in \Omega$. We let $a_t := c_t - c_{t-1}$, and the available leftover capacity at time t (after including items in $\mathcal{S}(t)$) is

$$\max\left\{\sup_{t' \le t} \sum_{\tau=t'}^{t} a_\tau - \mathcal{Q}(S(\tau)), 0\right\} := \mathcal{C}_\omega^{\mathcal{S}}(t).$$

Then, overflow units at time t is

$$\max\left\{\sup_{t' \le t} \mathcal{Q}(S(\tau)) - \sum_{\tau=t'}^{t} a_\tau, 0\right\} := \Phi_\omega^{\mathcal{S}}(t),$$

and the total overflow units is $\Phi_\omega(\mathcal{S}) = \sum_{t=1}^T \Phi_\omega^{\mathcal{S}}(t)$.

With the above definitions, we first consider the calculation of overflows on a set $\mathcal{S}$ of items for a given sample path ω. This is done in Algorithm 1.10.

Algorithm 1.10 serves dual purpose – while calculating the overflow, it also implicitly finds an assignment of the items which do not suffer a penalty to supply units. The assignment of items to supply units can be non-unique, while Algorithm 1.10 identifies one way of matching. Intuitively, the algorithm assigns items to the latest available units, saving the earlier capacity for items with shorter deadlines. This allows us to find the total overflows by considering the items in an arbitrary order (instead of in increasing order of deadlines), which is in turn useful for finding incremental profit $\Delta\mathcal{P}$ when we add a set of requests to

Algorithm 1.10 OVERFLOW ASSIGNMENT

Parameters: Sample path of capacities $(c_1, \ldots, c_T) \in \mathbb{N}^T$, an arbitrary ordered list of requests $\mathcal{L} = (d_1, d_2, \ldots, d_n)$
Initialize: Remaining capacity $\mathbf{a}^r = (a_1^r, \ldots, a_T^r) \leftarrow (a_1, \ldots, a_T)$ ▷ $a_t = c_t - c_{t-1}$
Initialize: Units of overflow needing to pay penalty $\Phi \leftarrow 0$
$i \leftarrow 1$
while $i \leq n$ **do**
    $q^r \leftarrow q_i$
    $t_i = \max\{t \leq d_i : a_t^r > 0\}$
    while $q^r > 0$ **do**
        if $t_i < \infty$ and $t_i > 0$ **then**
            $a_{t_i}^r \leftarrow a_{t_i}^r - \min\left\{a_{t_i}^r, q^r\right\}$
            $q^r \leftarrow q^r - \min\left\{a_{t_i}^r, q^r\right\}$
            $t_i \leftarrow t_i - 1$
        else
            $\Phi \leftarrow \Phi + q^r$
            $q^r \leftarrow 0$
        end if
    end while
    $i \leftarrow i + 1$
end while
Return $(\mathbf{a}^r, \Phi)$

an existing set of accepted requests. We begin with the following lemma which proves that Algorithm 1.10 indeed finds the minimum overflow.

Lemma 1.13. *Given a sample path $\omega \in \mathbb{N}^T$ of supply, and a set $\mathcal{S}$ of items with general integer demands, let $\mathcal{L} = (d_1, \ldots, d_n)$ be an arbitrary ordering of the items in $\mathcal{S}$ (d_i denoting the deadlines). Then the overflow units Φ returned when executing Algorithm 1.10 (*OVERFLOW ASSIGNMENT*) on $(\omega, \mathcal{L})$ satisfies $\Phi = \Phi_\omega(\mathcal{S})$.*

Proof of Lemma 1.13. We will use LP duality to prove the Lemma. In a nutshell, we will use the the assignment created by Algorithm 1.10 to create a feasible solution to the dual LP such that the objective function of the dual matches the objective function penalty of the assignment. Since any feasible solution of the dual lower bounds the optimal, we would have thus demonstrated the optimality of the assignment and hence of the overflow units Φ.

(PRIMAL)		(DUAL)	
min	$\sum_{i=1}^{n} y_i$	max	$\sum_{i=1}^{n} q_i\gamma_i - \sum_t \lambda_t c_t$
s.t.		s.t.	
$\forall t \in [T]:$	$-\sum_{i:d_i \le t} x_i \ge -c_t$	$\forall i \in [n]:$	$\gamma_i \le 1$
$\forall i \in [n]:$	$x_i + y_i = q_i$	$\forall i \in [n]:$	$\gamma_i \le \sum_{t \ge d_i} \lambda_t$
	$x_i, y_i \ge 0$		$\lambda_t \ge 0$

To construct the dual solution, let $\tau = \min\{t : a_t^r > 0\}$. That is, τ is the first time at which there is some capacity remaining after the assignment of Overflow Assignment. By the nature of the algorithm, there are no items with $d_i \ge \tau$ for which penalty is paid, and in fact all items with $d_i \ge \tau$ are served with capacity that arrives at time τ or later. Therefore, the overflow units under the assignment is the total size of items with $d_i < \tau$ minus the capacity $c_{\tau-1}$ (since this capacity is only used by requests with $d_i < \tau$).

Now construct a dual solution as follows:

$$\lambda_t = \begin{cases} 1 & t = \tau - 1, \\ 0 & t \ne \tau - 1; \end{cases} \qquad \gamma_i = \begin{cases} 1 & d_i \le \tau - 1, \\ 0 & d_i \ge \tau. \end{cases}$$

It is easy to verify that this is a feasible dual solution. Further, the objective function value under this feasible dual is

$$\sum_{i:d_i \le \tau - 1} q_i - c_{\tau-1}$$

which is exactly the overflow units of the primal assignment. Therefore, the primal solution in fact attains the optimal objective. □

As a result of Algorithm 1.10 and Lemma 1.13, we have the following lemma.

Lemma 1.14. *Let $\mathcal{S}$ be a set of items disjoint with $\mathcal{S}_1$ and $\mathcal{S}_2$. If for some $\omega = \{c_t \mid t \in$*

$[T]\} \in \Omega$, *we have* $\mathcal{C}_\omega^{\mathcal{S}_1}(t) \geq \mathcal{C}_\omega^{\mathcal{S}_2}(t), \forall t \in [T]$, *then,* $\Delta\mathcal{P}_\omega(\mathcal{S}_1, \mathcal{S}) \geq \Delta\mathcal{P}_\omega(\mathcal{S}_2, \mathcal{S})$. *If this is true for all* $\omega \in \Omega$, *we further have that* $\Delta\mathcal{P}(\mathcal{S}_1, \mathcal{S}) \geq \Delta\mathcal{P}(\mathcal{S}_2, \mathcal{S})$.

Proof of Lemma 1.14. It suffices to show that $\Delta\Phi_\omega(\mathcal{S}_1, \mathcal{S}) \leq \Delta\Phi(\mathcal{S}_2, \mathcal{S})$. Note that

$$\Delta\Phi_\omega(\mathcal{S}_1, \mathcal{S}) = \Phi_{\omega'}(\mathcal{S}), \text{ where } \omega' = \left\{\mathcal{C}_\omega^{\mathcal{S}_1}(t) \mid t \in [T]\right\},$$
$$\Delta\Phi_\omega(\mathcal{S}_2, \mathcal{S}) = \Phi_{\omega''}(\mathcal{S}), \text{ where } \omega'' = \left\{\mathcal{C}_\omega^{\mathcal{S}_2}(t) \mid t \in [T]\right\}.$$

By Lemma 1.13, the ordering of items in $\mathcal{S}$ does not matter when computing the total overflow units, and we may apply Algorithm 1.10 to compute $\Phi_{\omega'}(\mathcal{S})$ and $\Phi_{\omega''}(\mathcal{S})$. Since $\mathcal{C}_\omega^{\mathcal{S}_1}(t) \geq \mathcal{C}_\omega^{\mathcal{S}_2}(t), \forall t \in [T]$, as we apply Algorithm 1.10, for any capacity in ω'' that is used to serve a unit of demand in $\mathcal{S}$, we have the same capacity in ω' that can be used to serve the same unit of demand in $\mathcal{S}$. It then follows that $\Phi_{\omega'}(\mathcal{S}) \leq \Phi_{\omega''}(\mathcal{S})$, which implies that $\Delta\Phi_\omega(\mathcal{S}_1, \mathcal{S}) \leq \Delta\Phi_\omega(\mathcal{S}_2, \mathcal{S})$. □

We next show the submodularity of $\mathcal{P}$.

Lemma 1.15. *For any* $\mathcal{S}_1 \subseteq \mathcal{S}_2$, *we have that* $\Delta\mathcal{P}(\mathcal{S}_1, \mathcal{S}_3) \geq \Delta\mathcal{P}(\mathcal{S}_2, \mathcal{S}_3)$.

Proof of Lemma 1.15. Since $\mathcal{S}_1 \subseteq \mathcal{S}_2$, in each realized sample path of capacities $\omega = \{c_t\}_{t=1}^T$, it should be clear that $\mathcal{C}_\omega^{\mathcal{S}_1}(t) \geq \mathcal{C}_\omega^{\mathcal{S}_2}(t), \forall t$, i.e., at each time period, the available remaining capacity on $\mathcal{S}_1$ is no less than the available remaining capacity on $\mathcal{S}_2$. Thus, by Lemma 1.14, the result follows. □

Lemma 1.14 and Lemma 1.15 showed the relationship of incremental profit change of adding a set of items on top of two other sets of items. Specifically, if one set always has more remaining capacity than the other set, then adding a third set to one generates more incremental profit than adding the same set to the other.

For the rest of this part, we impose the assumption that $q_i = q, \forall i \in [N]$. To simplify the presentation, we may without loss of generality assume that $q = 1$ by allowing $\{c_t\}$

to be nonintegers. We next have the following result which will serve as a key to prove Theorem 1.6.

Lemma 1.16. *Let $\mathcal{S}_1$ and $\mathcal{S}_2 = \mathcal{S}_2^- \sqcup \mathcal{S}_2^+$ be two disjoint set of items. Let i, j, k be three items not in either set such that:*

1. *$d_m \leq d_j \leq d_i$, for all items $m \in \mathcal{S}_2^-$,*
2. *$d_i \leq d_k \leq d_m$, for all items $m \in \mathcal{S}_2^+$.*

Then, we have that

$$\Delta\mathcal{P}\left(\mathcal{S}_1 \cup \{j,k\}, \mathcal{S}_2\right) \leq \Delta\mathcal{P}\left(\mathcal{S}_1 \cup \{i\}, \mathcal{S}_2\right). \tag{1.29}$$

Proof of Lemma 1.16. We begin with two observations.

Observation 1: Using Lemma 1.13, we can determine $\Delta\mathcal{P}\left(\mathcal{S}_1 \cup \{j,k\}, \mathcal{S}_2\right)$ as follows: We first fix an ordering of $\mathcal{S}_1$ and assign them using Algorithm 1.10. This gives some residual capacity vector $\mathbf{c}^r$. The problem of finding $\Delta\mathcal{P}(\mathcal{S}_1 \cup \{j,k\}, \mathcal{S}_2)$ under capacity vector ω now reduces to finding $\Delta\mathcal{P}(\{j,k\}, \mathcal{S}_2)$ under capacity vector $\mathbf{c}^r$. Similarly, finding $\Delta\mathcal{P}(\mathcal{S}_1 \cup \{i\}, \mathcal{S}_2)$ under capacity vector ω reduces to finding $\Delta\mathcal{P}(\{i\}, \mathcal{S}_2)$ under capacity vector $\mathbf{c}^r$.

Observation 2: It suffices to prove the Lemma for $|\mathcal{S}_2| = 1$.

We therefore consider two cases, based on whether the item m in S_2 has $d_m \leq d_j \leq d_i$ or $d_m \geq d_k \geq d_i$. Note that we have reduced to a case where we only need to worry about items i, j, k, m and capacity availability $\mathbf{c}^r$.

Case : $d_m \leq d_j \leq d_i$

To find incremental penalty:

$$\Phi_{\mathbf{c}^r}^{\{i,m\}} - \Phi_{\mathbf{c}^r}^{\{i\}}$$

we will first add item i and then m according to Algorithm 1.10. Similarly, for

$$\Phi_{\mathbf{c}^r}^{\{j,k,m\}} - \Phi_{\mathbf{c}^r}^{\{j,k\}}$$

we first add item j, then k and then m. We claim that if item m does not pay a penalty in the latter case (when added to $\{j,k\}$), then it does not pay a penalty when added to $\{i\}$. To see why, if m does not pay a penalty when added to $\{j,k\}$, then it must be that

$$\sum_{t \le d_j} c_t^r \ge 2, \quad \sum_{t \le d_m} c_t^r \ge 1.$$

In this case, when adding item i, there is still residual capacity left for matching m.

Case : $d_i \le d_k \le d_m$

In this we argue that if m pays a penalty when added to i, then it must pay a penalty when added to $\{j,k\}$. If m pays penalty for i, then:

$$\sum_{t \le d_i} c_t^r \le 1, \quad \sum_{d_i < t \le d_m} c_t^r = 0.$$

In this case when we first add k, it uses up any capacity $c_t^r \le d_i$, leaving m to pay a penalty.

Therefore, in either case, the incremental overflow units when adding item m to item i is at most the incremental overflow units when adding m to $\{j,k\}$. □

With the above lemmas, we are in a position to prove Theorem 1.6.

Proof of Theorem 1.6. First, suppose that without loss of generality, the items in $\mathcal{S}_p$ are added exactly in the order of $g_1, \ldots, g_l$. Our proof is done by defining G_i and $\mathcal{S}_i^*$ inductively, and show that

$$\mathcal{P}(\mathcal{S}^*) \le 2\mathcal{P}(G_i) + \Delta\mathcal{P}(G_i, \mathcal{S}_i^*), \quad \forall i \le \min\{l, m\} \text{ s.t. } \mathcal{S}_i^* \text{ is well-defined.}$$

Base Case. Let $G_1 = \{g_1\}$ and let $\mathcal{S}^* = \mathcal{S}^{*-} \sqcup \mathcal{S}^{*+}$ where $\mathcal{S}^{*-} := \{j \in \mathcal{S}^* \mid d_j < d_{g_1}\}$ and $\mathcal{S}^{*+} := \{j \in \mathcal{S}^* \mid d_j \geq d_{g_1}\}$. Define

$$\mathcal{S}_1^* = \begin{cases} \mathcal{S}^* \setminus \{g_1\}, & \text{if } g_1 \in \mathcal{S}^* \\ \mathcal{S}^* \setminus \{o', o''\}, & \text{if } g_1 \notin \mathcal{S}^* \end{cases}$$

where $o' \in \mathcal{S}^* : d_{i'} \leq d_{o'} \leq d_{g_1}, \forall i' \in \mathcal{S}^{*-}$, and $o'' \in \mathcal{S}^* : d_{g_1} \leq d_{o'} \leq d_{j'}, \forall j' \in \mathcal{S}^{*+}$, i.e., o' is an item in $\mathcal{S}^*$ with deadline no later than g_1 but no earlier than the deadlines of items in $\mathcal{S}^{*-}$, and o'' is an item in $\mathcal{S}^*$ with deadline no earlier than g_1 but no later than the deadlines of items in $\mathcal{S}^{*+}_i$ (if such o' or o'' does not exist, then simply ignore it). Then, we have the two cases:

- $g_1 \in \mathcal{S}^*$.

$$\begin{aligned} \mathcal{P}(\mathcal{S}^*) = \Delta\mathcal{P}\left(\emptyset, \mathcal{S}^*\right) &= \Delta\mathcal{P}\left(\emptyset, \{g_1\}\right) + \Delta\mathcal{P}\left(\{g_1\}, \mathcal{S}_1^*\right) \\ &\leq 2\mathcal{P}(G_1) + \Delta\mathcal{P}(G_1, \mathcal{S}_1^*) \end{aligned}$$

  where the inequality follows directly from the fact that $\mathcal{P}(G_1) = \Delta\mathcal{P}\left(\emptyset, \{g_1\}\right)$ is non-negative.

- $g_1 \notin \mathcal{S}^*$. First note that

$$\begin{aligned} \Delta\mathcal{P}\left(\emptyset, \left\{o', o''\right\}\right) &= \Delta\mathcal{P}\left(\emptyset, \{o'\}\right) + \Delta\mathcal{P}\left(\{o'\}, \{o''\}\right) \\ &\leq \Delta\mathcal{P}\left(\emptyset, \{o'\}\right) + \Delta\mathcal{P}\left(\emptyset, \{o''\}\right) \\ &\leq \Delta\mathcal{P}(\emptyset, \{g_1\}) + \Delta\mathcal{P}(\emptyset, \{g_1\}) = 2\Delta\mathcal{P}(\emptyset, \{g_1\}) = 2\mathcal{P}(G_1), \end{aligned}$$

  where the first inequality follows from Lemma 1.15 and the second inequality follows from the greedy algorithm that g_1 gives the greatest incremental profit.

On the other hand, by Lemma 1.16, we also have that

$$\Delta\mathcal{P}\left(\left\{o', o''\right\}, \mathcal{S}_1^*\right) \leq \Delta\mathcal{P}\left(G_1, \mathcal{S}_1^*\right).$$

Combining the above two inequalities, we conclude that

$$\begin{aligned}\mathcal{P}(\mathcal{S}^*) = \Delta\mathcal{P}\left(\emptyset, \mathcal{S}^*\right) &= \Delta\mathcal{P}\left(\emptyset, \left\{o', o''\right\}\right) + \Delta\mathcal{P}\left(\left\{o', o''\right\}, \mathcal{S}_1^*\right) \\ &\leq 2\mathcal{P}\left(G_1\right) + \Delta\mathcal{P}\left(G_1, \mathcal{S}_1^*\right)\end{aligned}$$

Induction Step. Assume that $\mathcal{P}(\mathcal{S}^*) \leq 2\mathcal{P}(G_i) + \Delta\mathcal{P}(G_i, \mathcal{S}_i^*)$, we define $G_{i+1} = G_i \cup \{g_{i+1}\}$ and let $\mathcal{S}_i^* = \mathcal{S}_i^{*-} \sqcup \mathcal{S}_i^{*+}$ where $\mathcal{S}_i^{*-} := \{j \in \mathcal{S}_i^* \mid d_j < d_{g_i}\}$ and $\mathcal{S}_i^{*+} := \{j \in \mathcal{S}_i^* \mid d_j \geq d_{g_i}\}$. Define

$$\mathcal{S}_{i+1}^* = \begin{cases} \mathcal{S}_i^* \setminus \{g_{i+1}\}, & \text{if } g_{i+1} \in \mathcal{S}_i^* \\ \mathcal{S}_i^* \setminus \{o', o''\}, & \text{if } g_{i+1} \notin \mathcal{S}_i^* \end{cases}$$

where where $o' \in \mathcal{S}_i^* : d_{i'} \leq d_{o'} \leq d_{g_i}, \forall i' \in \mathcal{S}_i^{*-}$, and $o'' \in \mathcal{S}_i^* : d_{g_i} \leq d_{o'} \leq d_{j'}, \forall j' \in \mathcal{S}_i^{*+}$, i.e., o' is an item in $\mathcal{S}_i^*$ with deadline no later than g_i but no earlier than the deadlines of items in $\mathcal{S}_i^{*-}$, and o'' is an item in $\mathcal{S}_i^*$ with deadline no earlier than g_i but no later than the deadlines of items in $\mathcal{S}^{*+}_i$ (if such o' or o'' does not exist, then simply ignore it). Then, we have in the two cases:

- $g_{i+1} \in \mathcal{S}_i^*$.

$$\begin{aligned}\mathcal{P}(\mathcal{S}^*) &\leq 2\mathcal{P}(G_i) + \Delta\mathcal{P}(G_i, \mathcal{S}_i^*) = 2\mathcal{P}(G_i) + \Delta\mathcal{P}\left(G_i, \{g_{i+1}\}\right) + \Delta\mathcal{P}\left(G_{i+1}, \mathcal{S}_{i+1}^*\right) \\ &\leq 2\mathcal{P}(G_i) + 2\Delta\mathcal{P}\left(G_i, \{g_{i+1}\}\right) + \Delta\mathcal{P}\left(G_{i+1}, \mathcal{S}_{i+1}^*\right) \\ &= 2\mathcal{P}(G_{i+1}) + \Delta\mathcal{P}\left(G_{i+1}, \mathcal{S}_{i+1}^*\right)\end{aligned}$$

where the first inequality follows from the induction assumption and the second inequality follows directly from the fact that $\Delta\mathcal{P}\left(G_i, \{g_{i+1}\}\right)$ is nonnegative.

- $g_{i+1} \notin \mathcal{S}_i^*$. First note that

$$\begin{aligned}\Delta\mathcal{P}\left(G_i, \left\{o', o''\right\}\right) &= \Delta\mathcal{P}\left(G_i, \{o'\}\right) + \Delta\mathcal{P}\left(G_i \cup \{o'\}, \{o''\}\right) \\ &\leq \Delta\mathcal{P}\left(G_i, \{o'\}\right) + \Delta\mathcal{P}\left(G_i, \{o''\}\right) \\ &\leq \Delta\mathcal{P}(G_i, \{g_{i+1}\}) + \Delta\mathcal{P}(G_i, \{g_{i+1}\}) = 2\Delta\mathcal{P}(G_i, \{g_{i+1}\}),\end{aligned}$$

where the first inequality follows from Lemma 1.15 and the second inequality follows from the greedy algorithm that g_{i+1} adds the greatest incremental profit to G_i.

On the other hand, by Lemma 1.16, we also have that

$$\Delta\mathcal{P}\left(G_i \cup \left\{o', o''\right\}, \mathcal{S}_{i+1}^*\right) \leq \Delta\mathcal{P}\left(G_{i+1}, \mathcal{S}_{i+1}^*\right).$$

Combining the above two inequalities, we conclude that

$$\begin{aligned}\mathcal{P}(\mathcal{S}^*) &\leq 2\mathcal{P}(G_i) + \Delta\mathcal{P}(G_i, \mathcal{S}_i^*) \\ &= 2\mathcal{P}(G_i) + \Delta\mathcal{P}\left(G_i, \left\{o', o''\right\}\right) + \Delta\mathcal{P}\left(G_i \cup \left\{o', o''\right\}, \mathcal{S}_{i+1}^*\right) \\ &\leq 2\mathcal{P}(G_i) + 2\Delta\mathcal{P}\left(G_i, \{g_{i+1}\}\right) + \Delta\mathcal{P}\left(G_{i+1}, \mathcal{S}_{i+1}^*\right) \\ &\leq 2\mathcal{P}\left(G_{i+1}\right) + \Delta\mathcal{P}\left(G_{i+1}, \mathcal{S}_{i+1}^*\right)\end{aligned}$$

This completes the induction step. Note that at each step, $\mathcal{S}_{i+1}^* \subsetneq \mathcal{S}_i^*$ and $G_i \subsetneq G_{i+1}$. In the end, we will reach some i' such that either $\mathcal{S}_{i'}^* = \emptyset$ or $G_{i'} = \mathcal{S}_p$ and $\mathcal{S}_{i'}^* \neq \emptyset$. In the first case, we have that

$$\mathcal{P}(\mathcal{S}^*) \leq 2\mathcal{P}(G_{i'}) + \Delta\mathcal{P}(G_{i'}, \mathcal{S}_{i'}^*) = 2\mathcal{P}(G_{i'}) + 0 \leq 2\mathcal{P}(\mathcal{S}_p).$$

In the second case, i.e., $G_{i'} = \mathcal{S}_p$ and $\mathcal{S}^*_{i'} \neq \emptyset$, we again have that $\mathcal{P}(\mathcal{S}^*) \leq 2\mathcal{P}(G_{i'}) + \Delta\mathcal{P}(G_{i'}, \mathcal{S}^*_{i'})$. Now if $\Delta\mathcal{P}(G_{i'}, \mathcal{S}^*_{i'}) > 0$, then we can add the items in $\mathcal{S}^*_{i'}$ to $\mathcal{S}_p$ and still increase the profit, which violates the greedy algorithm. Thus, it must be that $\Delta\mathcal{P}(G_{i'}, \mathcal{S}^*_{i'}) \leq 0$. Then we would have

$$\mathcal{P}(\mathcal{S}^*) \leq 2\mathcal{P}(G_{i'}) + \Delta\mathcal{P}(G_{i'}, \mathcal{S}^*_{i'}) \leq 2\mathcal{P}(\mathcal{S}_p).$$

In conclusion, we have that $\mathcal{P}(\mathcal{S}^*) \leq 2\mathcal{P}(\mathcal{S}_p)$, or equivalently $\mathcal{P}(\mathcal{S}_p) \geq \frac{1}{2}\mathcal{P}(\mathcal{S}^*)$. This completes the proof of Theorem 1.6. □

1.7.3.2 Proof of Theorem 1.7

Proof of Theorem 1.7. **Approximation ratio:** We let $\mathcal{S}^*$ be the optimal solution set for MPBKP-SS, and $\mathcal{S}'$ be the solution obtained from Algorithm 1.9. For each $t \in [T]$, we obtain from Algorithm 1.9 the sets $\mathcal{S}^t_k$ such that $\mathcal{S}^t_0 = \emptyset$, and for $k \geq 1$, $\mathcal{S}^t_k$ is the set of items such that $\mathcal{Q}(\mathcal{S}^t_k) = C^t_k$ and $\mathcal{R}(\mathcal{S}^t_k) = R^t_k$. By Lemma 1.1 and the approximation factor of $\tilde{f}_{\mathcal{I}(t)}$ to $f_{\mathcal{I}(t)}$, we have that given $\mathcal{S}^*(t)$ for each $t \in [T]$, there is some k_t such that $\mathcal{Q}(\mathcal{S}^t_{k_t}) \leq \mathcal{Q}_t(\mathcal{S}^*)$ and that $\mathcal{R}(\mathcal{S}^t_{k_t}) \leq \mathcal{R}_t(\mathcal{S}^*) \leq (1+\epsilon)\mathcal{R}(\mathcal{S}^t_{k_t})$. Let $\mathcal{S}'' := \cup_{t=1}^T \mathcal{S}^t_{k_t}$, then, we have that

$$\mathcal{R}(\mathcal{S}'') = \sum_{t=1}^T \mathcal{R}(\mathcal{S}^t_{k_t}) \leq \sum_{t=1}^T \mathcal{R}_t(\mathcal{S}^*) = \mathcal{R}(\mathcal{S}^*) \leq (1+\epsilon)\sum_{t=1}^T \mathcal{R}(\mathcal{S}^t_{k_t}) = (1+\epsilon)\mathcal{R}\left(\mathcal{S}''\right),$$

and that $\mathcal{Q}_t(\mathcal{S}'') = \mathcal{Q}(\mathcal{S}^t_{k_t}) \leq \mathcal{Q}_t(\mathcal{S}^*)$ for all $t \in [T]$. Now, let $\mathcal{P}(\mathcal{S})$ be the expected profit of $\mathcal{S}$, and $\Phi(\mathcal{S})$ be the expected total penalty on $\mathcal{S}$. Then $\mathcal{P}(\mathcal{S}) = \mathcal{R}(\mathcal{S}) - \Phi(\mathcal{S})$, and we have that

$$\begin{aligned}\mathcal{P}(\mathcal{S}'') = \mathcal{R}(\mathcal{S}'') - \Phi(\mathcal{S}'') &\geq \frac{1}{1+\epsilon}\mathcal{R}(\mathcal{S}^*) - \Phi(\mathcal{S}'') \\ &\geq \frac{1}{1+\epsilon}\mathcal{R}(\mathcal{S}^*) - \Phi(\mathcal{S}^*) = \mathcal{R}(\mathcal{S}^*) - \frac{\epsilon}{1+\epsilon}\mathcal{R}(\mathcal{S}^*) - \Phi(\mathcal{S}^*)\end{aligned}$$

$$\geq \mathcal{P}(\mathcal{S}^*) - \epsilon\mathcal{R}(\mathcal{S}^*),$$

where the second inequality follows from the fact that $\Phi(\mathcal{S}'') \leq \Phi(\mathcal{S}^*)$, since $\mathcal{Q}_t(\mathcal{S}'') \leq \mathcal{Q}_t(\mathcal{S}^*)$ for all $t \in [T]$. Further, from Assumption 1.2, we have that $\Phi(\mathcal{S}^*) \leq \beta\mathcal{R}(\mathcal{S})$, which implies that $\mathcal{P}(\mathcal{S}^*) = \mathcal{R}(\mathcal{S}^*) - \Phi(\mathcal{S}^*) \geq \mathcal{R}(\mathcal{S}^*) - \beta\mathcal{R}(\mathcal{S}^*) = (1-\beta)\mathcal{R}(\mathcal{S}^*)$. Therefore, we have that

$$\mathcal{P}(\mathcal{S}'') \geq \mathcal{P}(\mathcal{S}^*) - \frac{\epsilon}{1-\beta}\mathcal{P}(\mathcal{S}^*) = \left(1 - \frac{\epsilon}{1-\beta}\right)\mathcal{P}(\mathcal{S}^*).$$

Since $\mathcal{S}'' = \cup_{t=1}^{T} \mathcal{S}^t_{k_t}$ is one of the solutions considered in line 10 of Algorithm 1.9. We conclude that $\mathcal{P}(\mathcal{S}') \geq \mathcal{P}(\mathcal{S}'') \geq \left(1 - \frac{\epsilon}{1-\beta}\right)\mathcal{P}(\mathcal{S}^*)$.

Time complexity: Obtaining all $\tilde{f}_{\mathcal{I}(t)}$ takes $\tilde{\mathcal{O}}\left(n + T/\epsilon^{2.25}\right)$, while taking the maximum in line 10 costs $\mathcal{O}\left(\prod_{t=1}^{T} l_t\right) = \tilde{\mathcal{O}}\left(1/\epsilon^T\right)$. We conclude that the algorithm has runtime $\tilde{\mathcal{O}}\left(n + \frac{1}{\epsilon^T}\right)$ to achieve at least $\left(1 - \frac{\epsilon}{1-\beta}\right)$ approximation. □

1.7.4 An Alternative FPTAS for MPBKP-S

In this section, we provide an FPTAS for the MPBKP-S with time complexity $\mathcal{O}\left(\frac{n^2 \log n}{\epsilon}\right)$. Following the classical approach for "0-1" knapsack problems (see, e.g., [216]), we round down the reward of each item so that the optimal solution for the MPBKP under the new rounded rewards is upper bounded by some polynomial of n and $1/\epsilon$, and thus the naive pseudo-polynomial dynamic program becomes a polynomial time algorithm.

We assume that the items are initially sorted and relabeled in the increasing order of their deadlines, i.e., $d_1 \leq d_2 \leq \cdots \leq d_n$. Further, assume that we have a guess P_0 that satisfies (1.14). Then, we choose a discretization quantum $\kappa := \epsilon P_0/2n$ and define rounded rewards $\hat{r}_i := \left\lfloor \frac{r_i}{\kappa} \right\rfloor_\kappa$. We then have $\mathcal{P}(\mathcal{S}^*) \leq \frac{4n}{\epsilon}\kappa$.

For a solution $\mathcal{S} = \mathcal{S}(1) \cup \mathcal{S}(2) \cup \cdots \cup \mathcal{S}(T)$ where $\mathcal{S}(t)$ is the set of items with deadline t. Let the items in $\mathcal{S}(t)$ be indexed as $\mathcal{S}(t) = \left(i_1^{(t)}, \ldots, i_{S_t}^{(t)}\right)$ in the order in which Algorithm 1.11

considers them, we define the *rounded profit* of $\mathcal{S}$ as:

$$\hat{\mathcal{P}}(\mathcal{S}) = \hat{\mathcal{R}}(\mathcal{S}) - \sum_{t=1}^{T}\sum_{k=1}^{S_t}\left\lceil B\left(\sum_{\ell\le k} q_{i_\ell^{(t)}} - \max_{0\le t'<t}\left\{c_t - c_{t'} - \sum_{t'+1\le\tau<t}\mathcal{Q}(\mathcal{S}(\tau))\right\}\right)^+\right\rceil_\kappa . \tag{1.30}$$

Let us also define a single period change in rounded profit for a set of items $\mathcal{S} = (i_1, \ldots, i_S)$ with knapsack capacity c as:

$$\Delta\hat{\mathcal{P}}(\mathcal{S}, c) = \hat{\mathcal{R}}(\mathcal{S}) - \sum_{k=1}^{S}\left\lceil B\left(\sum_{\ell\le k} q_{i_\ell} - c\right)^+\right\rceil_\kappa . \tag{1.31}$$

Let $\hat{A}(i,p)$ be the maximum capacity left at time d_i when earning rounded profit at least p using items $\{1,\ldots,i\}$ with rounded down rewards $\hat{r}$, equivalently,

$$\hat{A}(i,p) := \max_{\substack{\mathcal{S}\subseteq\{1,\ldots,i\}\\ \hat{\mathcal{P}}(\mathcal{S})\ge p}}\ \max_{0\le t'<d_i}\left\{c_{d_i} - c_{t'} - \sum_{t'+1\le\tau\le d_i-1}\mathcal{Q}(\mathcal{S}(\tau))\right\}. \tag{1.32}$$

If it is not possible to earn profit p at time d_i using items $\{1,\ldots,i\}$ with rounded down rewards, i.e., no $\mathcal{S}\subseteq\{1,\ldots,i\}$ exists such that $\hat{\mathcal{P}}(\mathcal{S})\ge p$, then $\hat{A}(i,p)$ is labeled $-\infty$. The DP table runs for $i = 1,\ldots,n$ and $p = 0,\kappa,\ldots,\left\lceil\frac{4n}{\epsilon}\right\rceil\kappa$. We then have Algorithm 1.11, which returns an exact optimal solution of $\hat{\mathcal{P}}(\mathcal{S})$ under the rounded rewards and rounded penalties.

Proof of Correctness of Algorithm 1.11. We show that $\hat{A}(i,p)$ returned by the algorithm satisfies (1.32) by induction on i. The base case ($i = 0$) is vacuously true. Now we assume that (1.32) holds for all $p\in\{0,1,\ldots,\lceil 4n/\epsilon\rceil\}\kappa$ and for all $k\in[i-1]$. Consider some $p\in\{0,1,\ldots,\lceil 4n/\epsilon\rceil\}\kappa$, and let $\mathcal{S}^*$ be any set achieving the maximum in (1.32) so that $\hat{\mathcal{P}}(\mathcal{S})\ge p$. We will show that $\hat{A}(i,p)$ is at least the leftover capacity under solution $\mathcal{S}^*$ via

Algorithm 1.11 DP with rounded down rewards for MPBKP-S

1: **Define** $\kappa = \frac{\epsilon P_0}{2n}$
2: **Define** $\hat{r}_i = \kappa \left\lfloor \frac{r_i}{\kappa} \right\rfloor$ ▷ Round down reward
// $\hat{A}(i,p)$ = max capacity left at time d_i when earning (rounded) profit at least p by selecting items in $\{1,\ldots,i\}$ with rounded down rewards $\hat{r}$
3: Initialize $\hat{A}(0,p) = \begin{cases} 0 & p = 0, \\ -\infty & p > 0. \end{cases}$
4: **for** $t = 1,\ldots,T$ **do**
5: $i = I(t-1)+1$
6: **for** $p = \left\{0,1,\ldots,\left\lceil \frac{4n}{\epsilon} \right\rceil\right\} \cdot \kappa$ **do**
7: $\hat{A}(i,p) := \hat{A}(i-1,p) + c_t - c_{t-1}$ ▷ If reject request i
8: **end for**
9: **for** $\bar{p} = \left\{0,1,\ldots,\left\lceil \frac{4n}{\epsilon} \right\rceil\right\} \cdot \kappa$ **do**
10: $p = \bar{p} + \hat{r}_i - \left\lceil B(q_i - \max\{0, \hat{A}(i-1,\bar{p}) + (c_t - c_{t-1})\})^+ \right\rceil_\kappa$
11: $\hat{A}(i,p) = \max\{\hat{A}(i,p), \hat{A}(i-1,\bar{p}) + (c_t - c_{t-1}) - q_i\}$ ▷ Accept i
12: **end for**
13: **for** $p = \left\{\left\lceil \frac{4n}{\epsilon} \right\rceil, \left\lceil \frac{4n}{\epsilon} \right\rceil - 1,\ldots,1\right\} \cdot \kappa$ **do**
14: **if** $\hat{A}(i,p-\kappa) < \hat{A}(i,p)$ **then**
15: $\hat{A}(i,p-\kappa) = \hat{A}(i,p)$
16: **end if**
17: **end for**
18: **for** $i = I(t-1)+2,\ldots,I(t)$ **do**
19: **for** $p = \left\{0,1,\ldots,\left\lceil \frac{4n}{\epsilon} \right\rceil\right\} \cdot \kappa$ **do**
20: $\hat{A}(i,p) := \hat{A}(i-1,p)$ ▷ If reject request i
21: **end for**
22: **for** $\bar{p} = \left\{0,1,\ldots,\left\lceil \frac{4n}{\epsilon} \right\rceil\right\} \cdot \kappa$ **do**
23: $p = \bar{p} + \hat{r}_i - \left\lceil B(q_i - \max\{0, \hat{A}(i-1,\bar{p})\})^+ \right\rceil_\kappa$
24: $\hat{A}(i,p) = \max\{\hat{A}(i,p), \hat{A}(i-1,\bar{p}) - q_i\}$ ▷ Accept i
25: **end for**
26: **for** $p = \left\{\left\lceil \frac{4n}{\epsilon} \right\rceil, \left\lceil \frac{4n}{\epsilon} \right\rceil - 1,\ldots,1\right\} \cdot \kappa$ **do**
27: **if** $\hat{A}(i,p-\kappa) < \hat{A}(i,p)$ **then**
28: $\hat{A}(i,p-\kappa) = \hat{A}(i,p)$
29: **end if**
30: **end for**
31: **end for**
32: **end for**

case analysis:

- Case $i \notin \mathcal{S}^*$: In this case, the leftover capacity under $\mathcal{S}^*$ is the leftover capacity by d_i, which is the sum of leftover capacity in $\mathcal{S}^*$ by d_{i-1} and $c_{d_i} - c_{d_{i-1}}$. By induction hypothesis, $\hat{A}(i-1,p)$ is no less than the leftover capacity of $\mathcal{S}^*$ by d_{i-1}, and therefore, by lines (7,11) and (20,24), $\hat{A}(i,p) \geq \hat{A}(i-1,p) + c_{d_i} - c_{d_{i-1}}$ which in turn is no less

than the leftover capacity under $\mathcal{S}^*$ by d_i. By optimality of $\mathcal{S}^*$, all the inequalities must be equalities.

- Case $i \in \mathcal{S}^*$: Let $\mathcal{S}' = \mathcal{S}^* \setminus \{i\}$, and let $p' = \hat{\mathcal{P}}(\mathcal{S}')$ be its rounded profit. Then by induction hypothesis, $\hat{A}(i-1, p')$ is no less than the leftover capacity under $\mathcal{S}'$ by d_{i-1}. Further, by packing item i in the solution corresponding to $\hat{A}(i-1, p')$, the change in profit is larger than by packing item i in $\mathcal{S}'$ (the penalty is no less under $\mathcal{S}'$ since it has weakly smaller leftover capacity). Therefore, packing item i in the solution corresponding to $\hat{A}(i-1, p')$ gives a solution with at least as large a rounded profit as p and at least as much leftover capacity by d_i as $\mathcal{S}^*$. Therefore, in turn $\hat{A}(i, p)$ is at least as much as the leftover capacity in $\mathcal{S}^*$. Since we assume $\mathcal{S}^*$ to have the largest leftover capacity with profit at least p, all the inequalities must be equalities.

□

Our next result gives the approximation guarantee for Algorithm 1.11.

Lemma 1.17. *Let $\mathcal{S}^*$ be the optimal solution set to the original MPBKP-S, and P_0 satisfy* (1.14)*. Let $\mathcal{S}'$ denote the optimal solution set by Algorithm 1.11, i.e., $\mathcal{S}'$ is the solution set corresponding to $\hat{A}(n, p^*)$ where p^* is the maximum p such that $\hat{A}(n, p) > -\infty$. Then,*

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

Proof of Lemma 1.17. For any item i, because of rounding down, $\hat{r}_i$ is smaller than r_i. Also there are at most n rounding ups on the penalties in $\mathcal{S}^*$, each by not more than κ. Then,

$$\mathcal{P}(\mathcal{S}^*) - \hat{\mathcal{P}}(\mathcal{S}^*) \leq 2n\kappa.$$

The dynamic programming step must return a set, $\mathcal{S}'$, at least as good as $\mathcal{S}^*$ under the new

profit. Therefore,

$$\mathcal{P}(\mathcal{S}') \geq \hat{\mathcal{P}}(\mathcal{S}') = p^* \geq \hat{\mathcal{P}}(\mathcal{S}^*) \geq \mathcal{P}(\mathcal{S}^*) - 2n\kappa = \mathcal{P}(\mathcal{S}^*) - \epsilon P_0 \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*),$$

where first inequality follows because the rewards are rounded down and the penalties are rounded up in calculation of $\hat{\mathcal{P}}$, second inequality follows because $\mathcal{S}'$ is the optimal set for objective $\hat{\mathcal{P}}$, the third inequality follows because $|\mathcal{S}^*| \leq n$ and $T \leq n$, and the last inequality follows from (1.14) that $\mathcal{P}(\mathcal{S}^*) \geq P_0$. □

It remains to find P_0 which satisfies (1.14). Since $\mathcal{P}(\mathcal{S}^*) \leq \bar{P}$, we can enumerate P_0 from $\bar{P}/2, \bar{P}/4, \bar{P}/8, \ldots$, and one of them must satisfy (1.14). The FPTAS is presented as Algorithm 1.12.

Algorithm 1.12 FPTAS for MPBKP-S in $\mathcal{O}(n^2 \log n/\epsilon)$

1: $P_0 \leftarrow \bar{P}$
2: $p^* \leftarrow 0$
3: **while** $p^* < (1-\epsilon)P_0$ **do**
4: $\quad P_0 \leftarrow \frac{P_0}{2}$
5: $\quad$ Run Algorithm 1.11 with the current P_0.
6: $\quad p^* \leftarrow \max_{\substack{p \in \left\{0,\ldots,\left\lceil \frac{4n}{\epsilon} \right\rceil\right\} \cdot \kappa \\ \hat{A}(n,p) > -\infty}} p$
7: **end while**

Theorem 1.8. *Algorithm 1.12 is a fully polynomial approximation scheme for the MPBKP-S, which achieves* $(1-\epsilon)$ *factor of optimal with running time* $\mathcal{O}\left(\frac{n^2 \log n}{\epsilon}\right)$.

Proof of Theorem 1.8. **Time complexity:** When P_0 satisfies (1.14), by Lemma 1.17 we have that

$$p^* \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*) \geq (1-\epsilon)P_0.$$

Thus, the "while" loop terminates when P_0 satisfies (1.14), if not before P_0 satisfies (1.14). When P_0 satisfies (1.14), we would also have $\mathcal{P}(\mathcal{S}^*)/2 \leq P_0 \leq \mathcal{P}(\mathcal{S}^*)$. Therefore, the number

of iterations is upper bounded by

$$\text{number of iterations} \leq \log \frac{\bar{P}/2}{\mathcal{P}(\mathcal{S}^*)/2} \leq \log n,$$

where we have used the fact that $\bar{P} \leq nP \leq n\mathcal{P}(\mathcal{S}^*)$. Since each iteration takes time $\mathcal{O}\left(n \cdot \left\lceil \frac{4n}{\epsilon} \right\rceil\right)$ we get a total time complexity of $\mathcal{O}\left(\frac{n^2 \log n}{\epsilon}\right)$.

Approximation ratio: When Algorithm 1.12 terminates, it returns the last p^* and the solution set $\mathcal{S}'$ corresponding to $\hat{A}(n, p^*)$. If the "while" loop terminates when $P_0 > \mathcal{P}(\mathcal{S}^*)$, i.e., it stops before P_0 falls below $\mathcal{P}(\mathcal{S}^*)$, then we have that

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq (1-\epsilon)P_0 > (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

Otherwise, from the time complexity analysis, we know that the "while" loop must terminate when P_0 first falls below $\mathcal{P}(\mathcal{S}^*)$, which implies that the last P_0 satisfies (1.14). Then by Lemma 1.17 we again have that

$$\mathcal{P}(\mathcal{S}') \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

In either case, the solution we obtained from Algorithm 1.12 achieves $(1-\epsilon)$ optimal. $(1-\epsilon)$ factor of $\mathcal{P}(\mathcal{S}^*)$. □

1.7.5 Other Special Cases for MPBKP, MPBKP-S, and MPBKP-SS

While we have considered the special case of MPBKP-SS when all items have $q_i = 1$ in Section 1.5.1. In this subsection, we consider three other special cases: when all items have $q_i = 1$ in MPBKP, when all items have $q_i = 1$ in MPBKP-S, and when $T = 1$ in MPBKP-SS.

1.7.5.1 Special case for MPBKP when $q_i = 1$

Considering a special case of MPBKP where $q_i = 1$ for all $i \in [n]$, i.e., all items have size 1. In this case, the problem becomes

$$\begin{aligned} &\max_x \sum_{i=1}^{n} r_i x_i \\ &\text{s.t.} \sum_{j:d_j \le d_i} x_j \le c_i, \quad \forall i = 1, \dots, n \\ &\qquad x_i \in \{0,1\}, \quad \forall i = 1, \dots, n. \end{aligned} \tag{1.33}$$

We claim that in this special case, the optimal solution can be obtained by greedily picking the items according to their rewards. Formally,

Proposition 1.5. *For the special case where $q_i = 1$ for all i, the optimal algorithm is to sort the items in decreasing order of r_i, and greedily pick an item if it is feasible.*

The proof of Proposition 1.5 is straightforward and omitted. In Proposition 1.5, by "greedily pick an item if it is feasible", we mean that we begin with an empty solution, check the item with the largest reward and pick it if it is feasible to do so. We then check the item with the second largest reward, and pick it if adding it to the current solution is feasible. We continue this pattern until all items have been checked. To check feasibility, we maintain counters for number of items accepted with deadlines within each dyadic interval (that is intervals of the form $[j \cdot 2^k + 1, (j+1) \cdot 2^k]$ for all $k \in \{0, 1, \dots, \log_2 T\}$ and $j \in \{0, \dots, T/2^k\}$), which can be updated in $\mathcal{O}(\log T)$ time per item, and enable checking feasibility in time $\mathcal{O}(\log T)$ per item. The total time complexity of the greedy algorithm is therefore $\mathcal{O}(n \log n)$.

1.7.5.2 Special case for MPBKP-S when $q_i = 1$

Consider a special case of MPBKP-SS where $q_i = 1$ for all $i \in [n]$, i.e., all items have size 1. In this case, the problem becomes

$$\max_x z = \sum_{i=1}^{n} r_i x_i - B \cdot \sum_{t=1}^{T} \left[\sum_{j \in \mathcal{I}(t)} x_j - \max_{0 \le t' < t} \left\{ c_t - c_{t'} - \sum_{j \in \mathcal{S}: t'+1 \le d_j \le t-1} x_j \right\} \right]^+ \tag{1.34}$$

$$\text{s.t. } x_i \in \{0, 1\}, \quad \forall i \in [n]$$

We claim that in this special case, the optimal solution again can be obtained by greedily picking the items according to their rewards. Formally,

Proposition 1.6. *For the special case where $q_i = 1$ for all i, the optimal algorithm is to sort the items in decreasing order of r_i, and greedily pick an item if it is profitable.*

The proof of Proposition 1.6 is straightforward and omitted. In Proposition 1.6, by "greedily pick an item if it is feasible", we mean that we check the item with the largest reward, and pick it if the profit increases (compared to the profit before picking it). We then check the item with the second largest reward, and pick it if it is profitable (the profit increases by adding this item). We continue this pattern until all items have been checked. As in the previous section, we can compute the change in profit by maintaining the total size of items picked with deadlines within each dyadic interval, and therefore we can check profitability in $\mathcal{O}(\log T)$ time per item, or $\mathcal{O}(n \log n)$ in total.

1.7.5.3 Special case for MPBKP-SS when $T = 1$

In this subsection, we consider the special case when $T = 1$. This is the same as the 0-1 knapsack problem with random capacity and linear penalty for overflow, i.e., we solve the following stochastic program:

$$\max_{x\in\{0,1\}^n} z(x) := \sum_{i=1}^{n} r_i x_i - B\cdot\mathbb{E}\left[\sum_{j\in[n]} q_j x_j - C\right]^+, \tag{1.35}$$

where C is the random capacity following some known distribution $\mathcal{C}$. We again denote an optimal solution to (1.35) by x^*, and the corresponding item set by $\mathcal{S}^*$. Let p_i be the expected profit of item i, which is defined as the expected profit earned if we select only i, i.e., $p_i = r_i - B\cdot\mathbb{E}\left[(q_i - C)^+\right]$. Then, we have the following lemma.

Lemma 1.18. *If for item j we have $p_j < 0$, i.e., $r_j < B\cdot\mathbb{E}\left[\left(q_j - C\right)^+\right]$, then $j \notin \mathcal{S}^*$.*

Proof of Lemma 1.18. Suppose that to the contrary $j \in \mathcal{S}^*$. Then consider the set $\mathcal{S}' := \mathcal{S}^* \setminus \{j\}$. We have that

$$\begin{aligned}\mathcal{P}(\mathcal{S}^*) - \mathcal{P}(\mathcal{S}') &= r_j - B\cdot\mathbb{E}\left[(\mathcal{Q}(\mathcal{S}^*) - C)^+\right] + B\cdot\mathbb{E}\left[\left(\mathcal{Q}(\mathcal{S}') - C\right)^+\right]\\ &= r_j - B\cdot\left(\mathbb{E}\left[(\mathcal{Q}(\mathcal{S}^*) - C)^+\right] - \mathbb{E}\left[\left(\mathcal{Q}(\mathcal{S}') - C\right)^+\right]\right)\end{aligned}$$

Note that since $\mathcal{Q}(\mathcal{S}^*) = \mathcal{Q}(\mathcal{S}') + q_j$, we have that

$$E\left[(\mathcal{Q}(\mathcal{S}^*) - C)^+\right] \geq \mathbb{E}\left[\left(\mathcal{Q}(\mathcal{S}') - C\right)^+\right] + \mathbb{E}\left[\left(q_j - C\right)^+\right].$$

Therefore,

$$\mathcal{P}(\mathcal{S}^*) - \mathcal{P}(\mathcal{S}') \leq r_j - B\cdot\mathbb{E}\left[\left(q_j - C\right)^+\right] < 0,$$

which contradicts with the assumption that $\mathcal{S}^*$ is the optimal solution. □

From Lemma 1.18, we assume without loss of generality that each item i is by itself profitable in expectation, i.e., $r_i \geq B\cdot\mathbb{E}\left[(q_i - C)^+\right]$ for all $i \in [n]$, so one profitable solution

would be $\{i\}$. This assumption is natural as otherwise there exists some item i that will only bring down the expected profit if included in any solution, in which case we may simply eliminate that item when solving for the problem.

The problem of maximizing profit under stochastic capacity (1.35) can be equivalently expressed as finding the set $\mathcal{S}^*$:

$$\mathcal{S}^* \in \arg\max_{\mathcal{S}\subseteq[n]} \mathcal{P}(\mathcal{S}) := \mathcal{R}(\mathcal{S}) - B \cdot \mathbb{E}\left[(\mathcal{Q}(\mathcal{S}) - C)^+\right]. \tag{1.36}$$

We first observe that the penalty part of (1.36) is convex and increasing in $\mathcal{Q}(\mathcal{S})$. Formally, we have the following lemma.

Lemma 1.19. *The function $\phi(q) := B \cdot \mathbb{E}\left[(q - C)^+\right]$ is a convex, nondecreasing function of q.*

Proof of Lemma 1.19. We first show that $\phi(q)$ is nondecreasing in q. For any $q_2 \geq q_1 \geq 0$, we have that

$$\phi(q_2) - \phi(q_1) = B \cdot \mathbb{E}\left[(q_2 - C)^+\right] - B \cdot \mathbb{E}\left[(q_1 - C)^+\right] = B \cdot \mathbb{E}\left[(q_2 - C)^+ - (q_1 - C)^+\right].$$

For each realization of C, we have that $(q_2 - C)^+ - (q_1 - C)^+ \geq 0$. Therefore, since taking the expectation is equivalent to the convex sum of all realizations of C, we conclude that $\phi(q_2) - \phi(q_1) \geq 0$, and thus the function is nondecreasing in q.

We next show that $\phi(q)$ is a convex function, i.e., for any $0 \leq \theta \leq 1$,

$$B \cdot \mathbb{E}\left[(\theta q_1 + (1-\theta)q_2 - C)^+\right] \leq \theta \cdot B \cdot \mathbb{E}\left[(q_1 - C)^+\right] + (1-\theta) \cdot B \cdot \mathbb{E}\left[(q_2 - C)^+\right].$$

It suffices to show that for each realization of C,

$$(\theta q_1 + (1-\theta)q_2 - C)^+ \leq \theta \cdot (q_1 - C)^+ + (1-\theta) \cdot (q_2 - C)^+. \tag{1.37}$$

Without loss of generality, assume that $q_2 \geq q_1$, then

- If $q_2 \geq q_1 \geq C$, then (1.37) becomes $\theta q_1 + (1-\theta) q_2 - C \leq \theta \cdot (q_1 - C) + (1-\theta) \cdot (q_2 - C)$, which is true since we actually have $\theta q_1 + (1-\theta) q_2 - C = \theta \cdot (q_1 - C) + (1-\theta) \cdot (q_2 - C)$.
- If $C \geq q_2 \geq q_1$, then both sides of (1.37) are 0, and (1.37) holds.
- If $q_2 \geq C \geq q_1$ and we further have $\theta q_1 + (1-\theta) q_2 - C \leq 0$, then (1.37) becomes $0 \leq (1-\theta) \cdot (q_2 - C)$, which holds since $q_2 - C \geq 0$
- If $q_2 \geq C \geq q_1$ and we further have $\theta q_1 + (1-\theta) q_2 - C \geq 0$, then (1.37) becomes $\theta q_1 + (1-\theta) q_2 - C \leq (1-\theta) \cdot (q_2 - C)$. Note that $\theta q_1 + (1-\theta) q_2 - C = \theta (q_1 - C) + (1-\theta)(q_2 - C)$, and $q_1 - C \leq 0$, thus we have $\theta q_1 + (1-\theta) q_2 - C = \theta (q_1 - C) + (1-\theta)(q_2 - C) \leq (1-\theta) \cdot (q_2 - C)$.

This completes the proof of (1.37), and thus the convexity of $\phi(q)$. □

By observing that the function $\phi(q) := B \cdot \mathbb{E}\left[(q - C)^+\right]$ is a convex, nondecreasing function of q, we can write the profit function equivalently as the following:

$$\mathcal{P}(\mathcal{S}) := \mathcal{R}(\mathcal{S}) - \phi\left(\mathcal{Q}(\mathcal{S})\right) \tag{1.38}$$

where $\phi(\cdot)$ is nondecreasing, convex, non-negative penalty function with $\phi(0) = 0$. Note that in our problem $\phi(q)$ is in fact a piecewise linear convex function. We have thus converted the Knapsack problem with random capacity and linear penalties for overflow to a deterministic Knapsack problem with convex penalty function ϕ.

Up to this point, one may speculate that the algorithms in Section 1.4.1, with $T = 1$, could possibly be used to solve this problem. However, we note that this is not the case. One of the most important features of the algorithm for MPBKP-S that we introduced in Section 1.4.1, as well as those well-known algorithms for 0-1 Knapsack problem in literature [113, 125, 126, 135] is to divide items into large items and small items according to some threshold on reward (or profit), and that large items are added via dynamic program and small items

are added greedily in their reward densities. In the 0-1 Knapsack problem (or MPBKP-S), we could bound the loss from adding small items greedily in their reward densities by the reward (or profit) of one small item. Specifically, in MPBKP-S, the penalty for going above the capacity grows linearly with rate B, and we assumed without loss of generality that $B > r_i/q_i$ for all $i \in [n]$. This implies that there is at most one small item that could be added above the capacity – adding more small items would only bring down the profit, and thus the loss from small items is at most the profit of that small item, which is bounded by the threshold. In our current problem, when the penalty function becomes convex, its derivative *can* be smaller than some r_i/q_i, and there is no longer a bound for the difference on profits between the greedy padding of small items and the optimal padding of small items, i.e., we cannot have the result in analogous to Lemma 1.7.

Facing this problem, we turn to a more intuitive dynamic program, without partitioning items to large ones and small ones. We now describe the algorithm.

Suppose again that we can find some some P_0 such that

$$P_0 \leq \mathcal{P}\left(\mathcal{S}^*\right) \leq 2P_0 \tag{1.39}$$

Then, we choose a discretization quantum $\kappa := \epsilon P_0/2n$ and define rounded rewards $\hat{r}_i := \left\lfloor \frac{r_i}{\kappa} \right\rfloor \cdot \kappa = \lfloor r_i \rfloor_\kappa$. We then have $\mathcal{P}(\mathcal{S}^*) \leq \frac{4n}{\epsilon}\kappa$. We next define the *rounded profit* of $\mathcal{S} = \{1', 2', \ldots, k'\}$ as:

$$\begin{aligned}\hat{\mathcal{P}}(\mathcal{S}) = &\lfloor r_{1'} - \phi(q_{1'}) \rfloor_\kappa + \lfloor r_{2'} + \phi(q_{1'}) - \phi(q_{1'} + q_{2'}) \rfloor_\kappa + \cdots \\ &+ \left\lfloor r_{k'} + \phi(q_{1'} + \cdots + q_{(k-1)'}) - \phi(q_{1'} + \cdots + q_{k'}) \right\rfloor_\kappa .\end{aligned} \tag{1.40}$$

Let $\hat{A}(i, p)$ be the minimum total size of used capacity using items $\{1, \ldots, i\}$ when earning

rounded profit at least p, with rounded down rewards $\hat{r}$, equivalently,

$$\hat{A}(i,p) = \min_{\left\{\substack{\mathcal{S}\subseteq\{1,\ldots,i\} \\ \hat{\mathcal{P}}(\mathcal{S})\geq p}\right\}} \mathcal{Q}(\mathcal{S}). \tag{1.41}$$

Again, if it is not possible to earn rounded profit at least p using items $\{1,\ldots,i\}$ with rounded rewards and penalties, i.e., no $\mathcal{S}\subseteq\{1,\ldots,i\}$ exists such that $\hat{\mathcal{P}}(\mathcal{S})\geq p$, then $\hat{A}(i,p)$ is labeled ∞. We then have Algorithm 1.13, which returns an exact optimal solution of $\hat{\mathcal{P}}(\mathcal{S})$ under the rounded rewards and rounded penalties. The DP table runs for $i=1,\ldots,n$ and $p=0,\kappa,\ldots,\left\lceil\frac{4n}{\epsilon}\right\rceil\kappa$.

Algorithm 1.13 DP with rounded profits for the single period problem with convex penalty functions

Define $\kappa=\frac{\epsilon P_0}{2n}$
Define $\hat{r}_i=\lfloor r_i\rfloor_\kappa$
// $\hat{A}(i,p)$ = `min total size of subset of items` $\{1,\ldots,i\}$ `with total rounded down profit` p
for $p=\left\{0,1,\ldots,\lceil\frac{4n}{\epsilon}\rceil\right\}\cdot\kappa$ **do**
    Initialize $\hat{A}(0,p)=\begin{cases}0 & p=0\\ \infty & \text{otherwise}\end{cases}$
end for
for $i=1,2,\ldots,n$ **do**
    for $p=\left\{0,1,\ldots,\lceil\frac{4n}{\epsilon}\rceil\right\}\cdot\kappa$ **do**
        $\hat{A}(i,p)=\hat{A}(i-1,p)$
    end for
    for $\bar{p}=\left\{0,1,\ldots,\lceil\frac{4n}{\epsilon}\rceil\right\}\cdot\kappa$ **do**
        $p=\bar{p}+\left\lfloor r_i-\Phi\left(\hat{A}(i-1,\bar{p})+q_i\right)+\Phi\left(\hat{A}(i-1,\bar{p})\right)\right\rfloor_\kappa$
        $\hat{A}(i,p)=\min\left\{\hat{A}(i,p),\ \hat{A}(i-1,\bar{p})+q_i\right\}$
    end for
end for
Return $\max\left\{p\in\left\{0,1,\ldots,\lceil\frac{4n}{\epsilon}\rceil\right\}\cdot\kappa\middle|\ \hat{A}(n,p)<\infty\right\}$

Proof of Correctness of Algorithm 1.13. Note that each finite $\hat{A}(i,p)$ corresponds to a feasible solution to the problem of earning rounded profit p using items $\{1,\ldots,i\}$. We prove by induction that $\hat{A}(i,p)$ calculated from the algorithm indeed satisfies (1.41). At the beginning, as the base case, no items have been added, so $\hat{A}(0,0)=0$ and $\hat{A}(0,p)=\infty$ for any $p>0$. In the induction step, assume that (1.41) holds for all $p\in\{0,1,\ldots,\lceil 4n/\epsilon\rceil\}\,\kappa$ and

for all $k \in [i-1]$. Consider some $p \in \{0, 1, \ldots, \lceil 4n/\epsilon \rceil\}\,\kappa$, and let $\mathcal{S}^*$ be any set achieving the minimum in (1.41) so that $\hat{\mathcal{P}}(\mathcal{S}) \geq p$. We will show that $\hat{A}(i,p)$ is at most the total size under solution $\mathcal{S}^*$ via case analysis:

- Case $i \notin \mathcal{S}^*$: In this case, the total size of $\mathcal{S}^*$ is the same as the total size of $\mathcal{S}^* \cap \{1, \ldots, i-1\}$. By induction hypothesis, $\hat{A}(i-1,p)$ is no greater than the total size of $\mathcal{S}^* \cap \{1, \ldots, i-1\}$, and therefore, by lines (8,12), $\hat{A}(i,p) \leq \hat{A}(i-1,p)$, which in turn is no greater than the total size of $\mathcal{S}^* \cap \{1, \ldots, i-1\}$. By optimality of $\mathcal{S}^*$, all the inequalities must be equalities.

- Case $i \in \mathcal{S}^*$: Let $\mathcal{S}' = \mathcal{S}^* \setminus \{i\}$, and let $p' = \hat{\mathcal{P}}(\mathcal{S}')$ be its rounded profit. Then by induction hypothesis, $\hat{A}(i-1,p')$ is no greater than the total size of $\mathcal{S}'$. Further, by packing item i in the solution corresponding to $\hat{A}(i-1,p')$, the change in profit is larger than by packing item i in $\mathcal{S}'$ (the penalty is no less under $\mathcal{S}'$ since it has weakly greater total size). Therefore, packing item i in the solution corresponding to $\hat{A}(i-1,p')$ gives a solution with at least as large a rounded profit as p and at most as much total size as $\mathcal{S}^*$. Therefore, in turn $\hat{A}(i,p)$ is at most as much as the total size of $\mathcal{S}^*$. Since we assume $\mathcal{S}^*$ to have the min total size with rounded profit at least p, all the inequalities must be equalities.

□

Our next result gives the approximation guarantee for Algorithm 1.13.

Lemma 1.20. *Let $\mathcal{S}^*$ be the optimal solution set maximizing* (1.38)*, and P_0 satisfy* (1.39)*. Let $\mathcal{S}'$ denote the optimal solution set by Algorithm 1.13, i.e., $\mathcal{S}'$ is the solution set corresponding to $\hat{A}(n,p^*)$ where p^* is the maximum p such that $\hat{A}(n,p) < \infty$. Then,*

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

Proof of Lemma 1.20. For any item i, because of rounding down, $\hat{r}_i$ is smaller than r_i, but by no more than κ. Also there are at most n rounding ups on the penalties in $\mathcal{S}^*$, each by not more than κ. Then,

$$\mathcal{P}(\mathcal{S}^*) - \hat{\mathcal{P}}(\mathcal{S}^*) \leq 2n\kappa.$$

The dynamic programming step must return a set, $\mathcal{S}'$, at least as good as $\mathcal{S}^*$ under the new profit $\hat{\mathcal{P}}$. Therefore,

$$\mathcal{P}(\mathcal{S}') \geq \hat{\mathcal{P}}(\mathcal{S}') = p^* \geq \hat{\mathcal{P}}(\mathcal{S}^*) \geq \mathcal{P}(\mathcal{S}^*) - 2n\kappa = \mathcal{P}(\mathcal{S}^*) - \epsilon P_0 \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*),$$

where first inequality follows because the rewards are rounded down and the penalties are rounded up in calculation of $\hat{\mathcal{P}}$, second inequality follows because $\mathcal{S}'$ is the optimal set for objective $\hat{\mathcal{P}}$, the third inequality follows because $|\mathcal{S}^*| \leq n$, and the last inequality follows from (1.39) that $\mathcal{P}(\mathcal{S}^*) \geq P_0$. □

It remains to find P_0 that satisfies (1.39). Since $\mathcal{P}(\mathcal{S}^*) \leq \bar{P}$, we can enumerate P_0 from $\bar{P}/2, \bar{P}/4, \bar{P}/8, \ldots$, and one of them must satisfy (1.39). The FPTAS is presented as Algorithm 1.14.

Algorithm 1.14 FPTAS for the single period Knapsack with convex penalty functions in $\mathcal{O}(n^2 \log n/\epsilon)$

1: $P_0 \leftarrow \bar{P}$
2: $p^* \leftarrow 0$
3: **while** $p^* < (1-\epsilon)P_0$ **do**
4: $\quad P_0 \leftarrow \frac{P_0}{2}$
5: $\quad$ Run Algorithm 1.13 with the current P_0.
6: $\quad p^* \leftarrow \max \left\{ p \in \left\{0, \ldots, \left\lceil \frac{4n}{\epsilon} \right\rceil\right\} \cdot \kappa \atop \hat{A}(n,p) < \infty \right\} p$
7: **end while**

Theorem 1.9. *Algorithm 1.14 is a fully polynomial approximation scheme for the Knapsack problem with convex penalty functions* (1.36)*, which achieves* $(1-\epsilon)$ *factor of optimal with running time* $\mathcal{O}\left(\frac{n^2 \log n}{\epsilon}\right)$.

Proof of Theorem 1.9. **Approximation ratio:** When Algorithm 1.14 terminates, it returns the last p^* and the solution set $\mathcal{S}'$ corresponding to $\hat{A}(n, p^*)$. If the "while" loop terminates when $P_0 > \mathcal{P}(\mathcal{S}^*)$, i.e., it stops before P_0 falls below $\mathcal{P}(\mathcal{S}^*)$, then we have that

$$\mathcal{P}(\mathcal{S}') \geq p^* \geq (1-\epsilon)P_0 > (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

Otherwise, from the time complexity analysis, we know that the "while" loop must terminate when P_0 first falls below $\mathcal{P}(\mathcal{S}^*)$, which implies that the last P_0 satisfies (1.39). Then by Lemma 1.20 we again have that

$$\mathcal{P}(\mathcal{S}') \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*).$$

In either case, the solution we obtained from Algorithm 1.14 achieves $(1-\epsilon)$ factor of $\mathcal{P}(\mathcal{S}^*)$.

Time complexity: When P_0 satisfies (1.39), by Lemma 1.20 we have that

$$p^* \geq (1-\epsilon)\mathcal{P}(\mathcal{S}^*) \geq (1-\epsilon)P_0.$$

Thus, the "while" loop terminates when P_0 satisfies (1.39), if not before P_0 satisfies (1.39). When P_0 satisfies (1.39), we would also have $\mathcal{P}(\mathcal{S}^*)/2 \leq P_0 \leq \mathcal{P}(\mathcal{S}^*)$. Therefore, the number of iterations is upper bounded by

$$\text{number of iterations} \leq \log \frac{\bar{P}/2}{\mathcal{P}(\mathcal{S}^*)/2} \leq \log n,$$

where we have used the fact that $\bar{P} \leq nP \leq n\mathcal{P}(\mathcal{S}^*)$. Since each iteration takes time $\mathcal{O}\left(n \cdot \left\lceil \frac{4n}{\epsilon} \right\rceil\right)$ we get a total time complexity of $\mathcal{O}\left(\frac{n^2 \log n}{\epsilon}\right)$. □

1.7.6 Pseudo-Polynomial Time Algorithms for Exact Solutions

In this subsection, we provide pseudo-polynomial time algorithms for MPBKP and MPBKP-S, which return the exact optimal solutions.

1.7.6.1 A pseudo-polynomial time algorithm for MPBKP

In this subsection, we introduce a pseudo-polynomial time algorithm, Algorithm 1.15, which returns the exact optimal solution for MPBKP. The algorithm uses dynamic programming (DP) approach. Let $A(i,r)$ be the maximum capacity left at time d_i when earning reward r using items $\{1,\ldots,i\}$, equivalently,

$$A(i,r) := \max_{\left\{\substack{\mathcal{S}\subseteq\{1,\ldots,i\}\\ \mathcal{R}(\mathcal{S})=r\\ \mathcal{S}\text{ feasible to (1.1)}}\right\}} c_{d_i} - \mathcal{Q}(\mathcal{S}) \tag{1.42}$$

If it is not possible to earn reward r using items $\{1,\ldots,i\}$, i.e., no feasible $\mathcal{S}\subseteq\{1,\ldots,i\}$ exists such that $\mathcal{R}(\mathcal{S})=r$, then $A(i,r)$ is labeled $-\infty$. The DP runs for $i=1,\ldots,n$ and $r=0,\ldots,nR$, where $R:=\max_i r_i$.

Algorithm 1.15 Exact pseudo-polynomial time algorithm for MPBKP

// $A(i,r)$ = max capacity left at time d_i when earning reward r using items $\{1,\ldots,i\}$
for $r=0,\ldots,n\cdot R$ **do**
    Initialize $A(0,r)=\begin{cases}0 & r=0,\\ -\infty & r\geq 1.\end{cases}$
end for
for $i=1,2,\ldots,n$ **do**
    for $r=0,\ldots,n\cdot R$ **do**
        $A_0(i,r) := A(i-1,r) + c_{d_i} - c_{d_{i-1}}$ //If reject item i
        $A_1(i,r) := \begin{cases}A(i-1,r-r_i) + c_{d_i} - c_{d_{i-1}} - q_i, & \text{if } \geq 0,\\ -\infty, & \text{otherwise.}\end{cases}$ //If accept item i
        $A(i,r) = \max\{A_0(i,r), A_1(i,r)\}$
    end for
end for

Proof of Correctness of Algorithm 1.15. Note that by definition of $A(i,r)$, for $i\geq 1$ and

$r \geq 0$, each nonnegative $A(i,r)$ corresponds to a feasible solution to the problem of earning reward r using items $\{1,\ldots,i\}$. At the beginning, no items have been added, so $A(0,r) = 0$. It then suffices to show that $A(i,r) = \max\{A_0(i,r), A_1(i,r)\}$, i.e., the $A(i,r)$ given by the recursion in line 8 of the algorithm is indeed (1.42), which is the maximum remaining capacity at time d_i when earning reward r using items $\{1,\ldots,i\}$.

We first show that $A(i,r) \geq \max\{A_0(i,r), A_1(i,r)\}$. If both $A_0(i,r)$ and $A_1(i,r)$ are $-\infty$, then since $A(i,r) \geq -\infty$, the inequality holds. If at least one of $A_0(i,r)$ and $A_1(i,r)$ is not $-\infty$, then it is nonnegative, and $\max\{A_0(i,r), A_1(i,r)\} \geq 0$ and gives a feasible solution. By definition $A(i,r)$ is the optimal (maximum) capacity left, thus

$$A(i,r) \geq \max\{A_0(i,r), A_1(i,r)\}. \tag{1.43}$$

We next show that $A(i,r) \leq \max\{A_0(i,r), A_1(i,r)\}$. The result trivially holds if $A(i,r) = -\infty$. If $A(i,r) \neq -\infty$, then $A(i,r) \geq 0$, which means it is feasible to earn reward r using items $\{1,\ldots,i\}$. We look at the solution corresponding to $A(i,r)$.

- If item i is rejected in this solution, then the capacity left when earning reward r using $\{1,\ldots,i\}$ would be the capacity left when earning reward r using $\{1,\ldots,i-1\}$ plus $(c_{d_i} - c_{d_{i-1}})$. Since $A(i-1,r)$ is the max capacity left when earning reward r using $\{1,\ldots,i-1\}$, we have that $A(i,r) \leq A(i-1,r) + (c_{d_i} - c_{d_{i-1}}) = A_0(i,r)$.

- If item i is accepted in this solution, then the capacity left when earning reward r using $\{1,\ldots,i\}$ would be the capacity left when earning reward $(r - r_i)$ using $\{1,\ldots,i-1\}$, plus $(c_{d_i} - c_{d_{i-1}})$, minus q_i. Since $A(i-1, r-r_i)$ is the max capacity left when earning reward $r - r_i$ using $\{1,\ldots,i-1\}$, we have that $A(i,r) \leq A(i-1,r-r_i) + (c_{d_i} - c_{d_{i-1}}) - q_i = A_1(i,r)$

In the solution corresponding to $A(i,r)$, item i is either accepted or rejected. Therefore, we

have that

$$A(i,r) \leq \max\{A_0(i,r), A_1(i,r)\}. \tag{1.44}$$

Combining (1.43) and (1.44), we conclude that $A(i,r) = \max\{A_0(i,r), A_1(i,r)\}$. □

After running Algorithm 1.15, we have $A(n,r)$ which gives the optimal solution for earning reward r using $\{1,\ldots,n\}$. Let r^* be the maximum r such that $A(n,r) \geq 0$. Then the optimal solution to the problem is the solution corresponding to $A(n,r^*)$. The running time for Algorithm 1.15 is $\mathcal{O}\left(n^2R\right)$, which is polynomial in n and R.

Remark 1.3. *Note that in Algorithm 1.15, we have used nR as the upper bound for the optimal value. This leads to the running time being $\mathcal{O}\left(n^2R\right)$. We can instead use the upper bound $\bar{R} := \sum_i r_i$, in which case the "for" loops in line 1 and line 5 run as $r = 0,\ldots,\bar{R}$. The time complexity of the algorithm then becomes $\mathcal{O}\left(n\bar{R}\right)$.*

1.7.6.2 A pseudo-polynomial time algorithm for MPBKP-S

In this subsection, we introduce the following pseudo-polynomial time algorithm, Algorithm 1.16, which returns the exact optimal solution for the MPBKP-S. The algorithm uses dynamic programming approach. Let $A(i,p)$ be the maximum capacity left at time d_i when earning profit p using items $\{1,\ldots,i\}$. If it is not possible to earn profit p at time d_i using items $\{1,\ldots,i\}$, $A(i,p)$ is labeled $-\infty$. The DP table runs for $i = 1,\ldots,n$ and $p = 0,\ldots,nP$, where $P := \max_i \mathcal{P}(\{i\})$.

Proof. **Proof of Correctness of Algorithm 1.16** Note that by definition of $A(i,p)$, for $i > 0$ and $p \geq 0$, each nonnegative $A(i,p)$ corresponds to a feasible solution to the problem of earning profit p using items $\{1,\ldots,i\}$. At the beginning, there is no inventory, so $A(0,r) = 0$. It then suffices to show that the recursion gives optimal $A(i,p)$.

Algorithm 1.16 Exact Pseudo-polynomial time algorithm for MPBKP-S

// $A(i,p)$ = `max capacity left at time` d_i `when earning profit` p `using items` $\{1,\ldots,i\}$

for $r = 0,\ldots,n\cdot P$ **do**

  Initialize $A(0,p) = \begin{cases} 0 & p = 0, \\ -\infty & p \geq 1. \end{cases}$

end for

for $i = 1,2,\ldots$ **do**

  for $r = \{0,\ldots,nP\}$ **do**

    $A_0(i,p) := A(i-1,p) + (c_{d_i} - c_{d_{i-1}})$ ▷ If reject request i

    $T_y(i,p) := \begin{cases} A(i-1, p-r_i+(q_i-y)B) + (c_{d_i} - c_{d_{i-1}}) - y & \text{if } \geq 0, \\ -\infty & \text{otherwise.} \end{cases}$ ▷ If accept request i and serve y units

    $A(i,p) = \max\left\{A_0(i,p), \max_{y\in\{1,\ldots,q_i\}} A_y(i,p)\right\}$

  end for

end for

We first show that $A(i,p) \geq \max\{A_0(i,p), \max_{y\in\{1,\ldots,q_i\}} A_y(i,p)\}$. If both $A_0(i,p)$ and $\max_{y\in\{1,\ldots,q_i\}} A_y(i,p)$ are $-\infty$, then since $A(i,r) \geq -\infty$, the inequality holds. If at least one of $A_0(i,r)$ and $\max_{y\in\{1,\ldots,q_i\}} A_y(i,p)$ is not $-\infty$, then it is nonnegative, and $\max\{A_0(i,r), \max_{y\in\{1,\ldots,q_i\}} A_y(i,p)\} \geq 0$ and gives a feasible solution. By definition $A(i,p)$ is the optimal (maximum) capacity left, thus

$$A(i,p) \geq \max\left\{A_0(i,p), \max_{y\in\{1,\ldots,q_i\}} A_y(i,p)\right\}. \tag{1.45}$$

We next show that $A(i,p) \leq \max\{A_0(i,p), \max_{y\in\{1,\ldots,q_i\}} A_y(i,p)\}$. The result trivially holds if $A(i,p) = -\infty$. If $A(i,p) \neq -\infty$, then $A(i,p) \geq 0$, which means it is feasible to earn profit p using items $\{1,\ldots,i\}$. We look at the solution corresponding to $A(i,p)$.

- If item i is rejected in this solution, then the capacity left when earning profit p using $\{1,\ldots,i\}$ would be the capacity left when earning profit p using $\{1,\ldots,i-1\}$ plus $(c_{d_i} - c_{d_{i-1}})$. Since $A(i-1,p)$ is the max capacity left when earning profit p using $\{1,\ldots,i-1\}$, we have that $A(i,p) \leq A(i-1,p) + (c_{d_i} - c_{d_{i-1}}) = A_0(i,p)$.

- If item i is accepted in this solution, then if we serve y units for item i from the capacity, the capacity left when earning profit p using $\{1,\ldots,i\}$ would be the capacity left when

earning profit $p - r_i + (q_i - y)B$ using $\{1, \ldots, i-1\}$, plus the supply from $d_{i-1}+1$ to d_i, minus y. Since $A(i-1, p - r_i + (q_i - y)B)$ is the max capacity left when earning profit $p - r_i + (q_i - y)B$ using $\{1, \ldots, i-1\}$ when serving y units for item i. Since we can only serve the units y from 1 to q_i, by maximizing over y we have that $A(i,p) \leq \max_{y \in \{1,\ldots,q_i\}} A(i-1, p - r_i + (q_i - y)B) + (c_{d_i} - c_{d_{i-1}}) - y = \max_{y \in \{1,\ldots,q_i\}} A_y(i,p)$.

In the solution corresponding to $A(i,p)$, item i is either accepted or rejected. Therefore, we have that

$$A(i,p) \leq \max \left\{ A_0(i,p), \max_{y \in \{1,\ldots,q_i\}} A_y(i,p) \right\}. \tag{1.46}$$

Combining (1.45) and (1.46), we conclude that $A(i,p) = \max\{A_0(i,r), \max_{y \in \{1,\ldots,q_i\}} A_y(i,p)\}$. Thus, $A(n,p)$ gives the optimal solution for earning profit p using $\{1, \ldots, n\}$. Let $\hat{p}$ be the maximum p such that $A(n,p) \geq 0$. Then the optimal solution to the problem is the solution corresponding to $A(n, \hat{p})$. □

After running Algorithm 1.16, we have $A(n,p)$ which gives the optimal solution for earning profit p using $\{1, \ldots, n\}$. Let p^* be the maximum p such that $A(n,p) > -\infty$. Then the optimal solution to the problem is the solution corresponding to $A(n, p^*)$. The running time for Algorithm 1.15 is $\mathcal{O}\left(n^2 P \cdot \max_i q_i\right)$, which is polynomial in n, P, and the sizes q.

Remark 1.4. *Note that in Algorithm 1.16, we have used nP as the upper bound for the optimal value. This leads to the running time being $\mathcal{O}\left(n^2 P \cdot \max_i q_i\right)$. We can instead use the upper bound $\bar{P} := \sum_{i \in [n]} p_i$, in which case the "for" loops in line 1 and line 5 run as $r = 0, \ldots, \bar{P}$. The time complexity of the algorithm then becomes $\mathcal{O}\left(n\bar{P} \cdot \max_i q_i\right)$.*

CHAPTER 2

AGGREGATING DISTRIBUTED ENERGY RESOURCES: EFFICIENCY AND MARKET POWER

2.1 Introduction

Distributed energy resources (DERs) such as solar photovoltaics, electric vehicles, and batteries, are small-scale resources located at the end-consumers level in distribution power systems. Under appropriate market rules, DERs enable end-consumers to become *prosumers*, i.e., if their DER supply exceeds their demand, they can sell the excess energy back to the grid. From an independent market operators' (ISOs) perspective, electric power demand is largely assumed to be inelastic. However, the presence of DERs, coupled with the recent developments in demand response programs, causes a fundamental shift where the demand becomes elastic (Adelman and Uçkun [4], Bertsimas et al. [35], Gan and Litvinov [87], Litvinov et al. [149], Zheng and Litvinov [234], Zhao et al. [233]), which challenges fundamental assumptions in existing electricity market's design and operation (Ritzenhofen et al. [182]). This fundamental shift calls for revisiting current wholesale markets design (Parag and Sovacool [171], Anjos and Gómez [16]). This is not an easy task for ISOs, as they do not have oversight over the distribution power network, and hence cannot include DER owners as market participants. Furthermore, even if ISOs can oversee the distribution power system, it would be a significant burden and impractical to include DER supply directly into the wholesale electricity market operations through ISOs, due to the communication, computational, and operational complexity. Different models have been proposed to include DER supply into wholesale electricity markets. One possibility is to have a Distribution System Operator (DSO) acting as a market manager at the distribution level, and finding socially-optimal dispatch, similar to ISOs (Lian et al. [143], Ntakou and Caramanis [168], Manshadi and Khodayar [153], Sotkiewicz and Vignolo [199], Terra et al. [211], Huang et al. [112]). An-

other possibility is to have fully-distributed electricity market designs, where end-consumers can trade among themselves (Moret and Pinson [164], Rahimi and Ipakchi [178]). The third model, which we adopt here, is to have an aggregator who collects energy from DER owners and sells in the wholesale market as a producer of electricity.

DER aggregation via profit-seeking intermediaries has been adopted by California ISO and New York ISO (Gundlach and Webb [98], Lavillotti [134]), and seems to be realistic for practical implementation in other markets. The recent FERC Order No. 2222 in September, 2020 (FERC [81]) enables DERs to be aggregated, in order to satisfy minimum size and performance requirements that each may not be able to meet individually, and participate alongside conventional generators in the wholesale markets, which is opened to new sources of energy and grid services. As an example, in California, there are now seven DER aggregators, and four of them are not conventional utilities (see https://www.caiso.com). The aggregator here buys DER supply from their owners, and bids directly to the ISO similar to generating companies. The relationship between the aggregator and prosumers in the same geographical footprint is naturally monopolistic (Cook et al. [58]), where aggregators become price-making in retail electricity markets as they can send price offers to prosumers in order to collect DER supply. Such price offers need to be high-enough so that DER owners are attracted to sell, but small-enough so that the aggregator can maximize its profits. This profit-seeking behavior can impact the overall electricity market efficiency, but, at the same time, due to the impracticality of direct DER participation into wholesale markets, aggregators are necessary and important players. This gives rise to the following important question: *In the presence of a profit-seeking and a monopolistic aggregator, is there an aggregation model that can attain a socially optimal (efficient) market outcome?* The presence of such a mechanism can in fact be significant. First, in reality, aggregators are mostly profit-seeking, and are often monopolistic, which makes the markets prone to efficiency losses (Alshehri et al. [10, 11]). Second, it is infeasible for DER owners to participate in wholesale markets; the presence

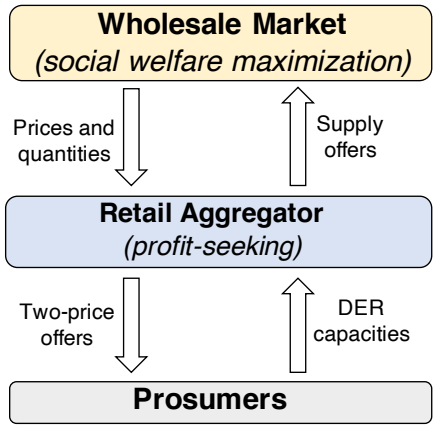

Figure 2.1: Overall interactions in the proposed efficient aggregation model

of such intermediaries is inevitable. Third, if such a mechanism can be designed, it would address various debates surrounding whether DER aggregation needs to be done by profit-making entities or social intermediaries.

In this chapter, we address the above question by first proposing a DER aggregation model that yields efficient market outcomes, even when a DER aggregator is *unregulated*, monopolistic and profit-seeking. We show that, with our proposed model, the optimal DER capacities being integrated through an aggregator are equivalent to a benchmark (ideal, not realistic) case where DER owners directly participate in the wholesale market. Briefly, our aggregation model utilizes two-price offers from the aggregator to the prosumers, where one price corresponds to a fixed DER owner participation cost (connection charges), and the other price is for marginal acquisition of DER capacities. We remark here that without the participation cost, having only a one-price offer would not yield efficient market outcomes (Sun et al. [202], Alshehri et al. [10, 11]). For an overall illustration, refer to Figure 2.1. Our results show that for the unregulated aggregator, while full market efficiency can be attained, the aggregator naturally discriminates among different prosumers, i.e., offering each prosumer a specific price pair. Interestingly, it has been previously observed that discriminatory policies are necessary for socially desirable outcomes (Singh and Scheller-Wolf [195]). Because electricity is a basic commodity, discriminatory policies might pose legal challenges. As a

remedy, we next consider a DER aggregation model with a profit-seeking aggregator who is enforced to offer uniform prices to prosumers from the same location. While it is difficult to analytically quantify the efficiency loss with the uniform price offerings, we propose an algorithm that numerically solves the aggregator's profit-maximization problem. Through an illustrative example, we show that the efficiency loss seems to be mild. Then, we also propose another aggregation model where the aggregator is *regulated* but guaranteed positive profit. With the regulated model, full market efficiency is achieved and the two-part pricing is only location-dependent, which is consistent with current market designs. Also, it is guaranteed for the regulated model that no prosumer is worse-off, compared to the case in which no DER participation is allowed, though some prosumers' payoffs might increase/decrease compared to the unregulated model. We also note that the analysis of this chapter does not consider the impact that DERs may have on the distribution network.

Finally, we shift our attention to market power and address the following question: *When DERs are aggregated, can the market power of conventional power generators be mitigated? If yes, to what extent?* When generators bid strategically, they might influence the market prices and thus negatively affect the social welfare (Al-Gwaiz et al. [7]). In this chapter, we demonstrate that our aggregation models are efficient, and can mitigate the market power of conventional generators. In particular, we prove that under strategic bidding, the social welfare is higher under our aggregation models, compared to the case when DERs are not integrated. We also prove that the welfare gap between truthful bidding and strategic bidding is smaller when DERs are aggregated, compared to the case where there are no DERs. Quantification of such differences are also provided.

The rest of this chapter is organized as follows: Section 2.2 discusses the benchmark ideal model. In Section 2.3, we propose an efficient *unregulated* DER aggregation model, which is discriminatory. Next, in Section 2.4, we numerically analyze the efficiency loss if no discriminatory prices are allowed with a profit-seeking aggregator. Then, in Section 2.5, we

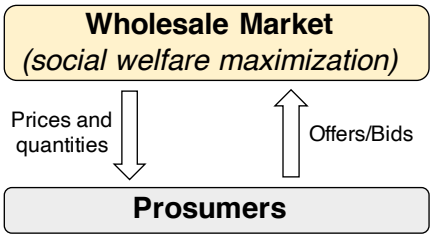

Figure 2.2: Benchmark model

propose a non-discriminatory *regulated* DER aggregation model, which also attains full market efficiency. Results on market power mitigation of conventional generators are provided in Section 2.6. Conclusions and future directions are provided in Section 2.7. All proofs can be found in the Appendix 2.8.

2.2 Direct Prosumer Participation Model (Benchmark)

In this section, we introduce an ideal model, termed *direct participation model*, where prosumers can participate (buy or sell) directly in the wholesale market, and no aggregator is present. This model, though not realistic, serves as a benchmark for evaluating the market efficiency of our following models. There are three parties in this model: prosumers who can buy and sell energy directly in the wholesale market, conventional generators who generate and sell electricity in the wholesale market, and an independent system operator (ISO) who clears the wholesale market. In the following, we describe the optimization problems solved by each of these three parties. Figure 2.2 illustrates the interactions between prosumers and the ISO.

2.2.1 Prosumer's Problem

Consider a power network with n nodes (locations). At each location k, there are n_k number of prosumers. As the focus of this chapter is on the aggregation mechanism, for simplicity, we assume that each prosumer i at location k can purchase energy at the wholesale price λ^k.

Furthermore, prosumer i at location k is endowed with a capacity $C_i^k \geq 0$ of power production from a collection of resources, such as solar panels, wall-mounted batteries, and plug-in electric vehicles. In this chapter, we assume that all prosumer capacities are deterministic. The power produced can be consumed locally by prosumer i or sold back to the wholesale market, again at the wholesale price λ^k. The prosumer observes λ^k and decides the amount of energy to buy/sell. Let u_i^k be prosumer i's utility of power consumption. We make the following assumption on prosumers' utility of consumption.

Assumption 2.1. *Each prosumer's utility u_i^k is increasing, strictly concave, and differentiable. Furthermore, the domain of u_i^k is $[0, Z]$ where $Z > C_i^k$ is some (large) upper bound of the amount of energy a prosumer consumes. We assume that $\frac{\partial u_i^k(z)}{\partial z} \to \infty$ as $z \to 0$, and $\frac{\partial u_i^k(z)}{\partial z} \to 0$ as $z \to Z$.*

We can then write prosumer i's optimization problem as:

$$\begin{aligned} \max_{z_i^k} \ & \pi_i^k(z_i^k) := \lambda^k(C_i^k - z_i^k) + u_i^k(z_i^k) \\ \text{s.t.} \ & \quad 0 \leq z_i^k \leq Z, \end{aligned} \tag{2.1}$$

where z_i^k is the amount of energy prosumer i actually consumes. Since prosumers at location k can buy and sell energy at the same price λ^k, prosumer i essentially has z_k^i as the single decision to make in solving (2.1). For selling $(C_i^k - z_i^k)$ at the wholesale price, prosumer i receives $\lambda^k(C_i^k - z_i^k)$; for buying $(z_i^k - C_i^k)$ at the wholesale price, it is charged $\lambda^k(z_i^k - C_i^k)$. The prosumer's utility from consumption would be $u_i^k(z_i^k)$. While Assumption 2.1 imposes strict concavity of prosumers' utilities, our analysis throughout this chapter remains largely applicable to generic concave utilities, but strict concavity allows us to derive unique analytical solutions and gain deep insights.

Lemma 2.1. *Under Assumption 2.1, given any λ^k, there exists a unique optimal solution*

z_i^{k*} *for the prosumer's problem* (2.1) *which satisfies* $\left.\frac{\partial u_i^k(z)}{\partial z}\right|_{z=z_i^{k*}} = \lambda^k$.

The above lemma directly follows from the properties of problem (2.1). We denote by x_i^k the actual amount of energy prosumer i sells, and by d_i^k the actual amount of energy prosumer i buys. Without loss of generality, and for ease of exposition, we restrict our attention to the case in which at most one of x_i^k and d_i^k can be nonzero. We first solve (2.1) for z_i^{k*}. Then, we let $x_i^{k*} = \left[z_i^{k*} - C_i^k\right]^+$ and $d_i^{k*} = \left[C_i^k - z_i^{k*}\right]^+$. We also use the notation $x_i^{k*}(\lambda^k)$ and $d_i^{k*}(\lambda^k)$ to denote the optimal response of prosumer i at location k for a given wholesale market price λ^k.

2.2.2 Generator's Problem

Let N_k denote the number of conventional generators at location k. Generator j at location k chooses to supply $y_j^k \in \left[\underline{y}_j^k, \overline{y}_j^k\right]$. Let $c_j^k(y_j^k)$ be the production cost with the following assumption.

Assumption 2.2. *Each generator's cost function* c_j^k *is increasing, strictly convex, and differentiable in* $\left[\underline{y}_j^k, \overline{y}_j^k\right]$. *Furthermore, we let* $\frac{\partial c_j^k(y_j^k)}{\partial y_j^k} \to 0$ *as* $y_j^k \to \underline{y}_j^k$ *and* $\frac{\partial c_j^k(y_j^k)}{\partial y_j^k} \to \infty$ *as* $y_j^k \to \overline{y}_j^k$.

By selling y_j^k, generator j earns a compensation $\lambda^k y_j^k$. Given a wholesale price λ^k, generator j maximizes its payoff by solving:

$$\max_{y_j^k \in \left[\underline{y}_j^k, \overline{y}_j^k\right]} \hat{\pi}_j^k(y_j^k) := \lambda^k y_j^k - c_j^k(y_j^k). \tag{2.2}$$

Lemma 2.2. *Under Assumption 2.2, given any* λ^k, *there exists a unique optimal solution* y_j^{k*} *for the generator's problem* (2.2) *which satisfies* $\left.\frac{\partial c_j^k(y_j^k)}{\partial y_j^k}\right|_{y_j^k = y_j^{k*}} = \lambda^k$.

The above lemma directly follows from the properties of problem (2.2). We use the notation $y_j^{k*}(\lambda^k)$ to denote the optimal response of generator j at location k at a given

wholesale price λ^k.

2.2.3 The Economic Dispatch Problem

Many wholesale electricity markets in the United States and other countries are managed by independent system operators (ISOs) (Greer [97]). An ISO clears the market by matching supply and demand via social welfare maximization (in practice, this is often done by production cost minimization to meet fixed system demands (Gan and Litvinov [87], Litvinov et al. [149], Zheng and Litvinov [234])), while ensuring that the power flows satisfy the network and line capacity constraints. Specifically, let $X^k = \sum_{i\in[n_k]} x_i^k$ be the total power supply at node k from prosumers; let $Y^k = \sum_{j\in[N_k]} y_j^k$ be the total power supply at node k from conventional generators; and let $D^k = \sum_{i\in[n_k]} d_j^k$ be the total demand (load) at node k. Furthermore, we let $\mathbf{Bh} \leq \mathbf{f}$ be the network constraints that are resolved from the DC approximation of the AC network, where $\mathbf{f}$ is the vector of capacities of transmission lines in the power network, and the system operator chooses a vector $\mathbf{h}$, where each element h^k is the net injection to node k. In summary, we have the following constraints:

$$\mathbf{h} = \mathbf{D} - \mathbf{Y} - \mathbf{X}, \qquad \mathbf{1}^T\mathbf{h} = 0, \qquad \mathbf{Bh} \leq \mathbf{f}. \tag{2.3}$$

We note that the first constraint ensures the total supply matches the total demand at each node (in this and the following models, we assume there is no transmission line loss); the next two constraints ensure that the total net injection by the system operator is zero over the power network (here, $\mathbf{1}$ is a vector of ones), and the total power transmission at each line does not exceed its capacity. In addition to the network constraints, the ISO also needs to consider all participant-specific constraints described earlier in problems (2.1) and (2.2):

$$\mathbf{C} - \mathbf{Z} \leq \mathbf{x} - \mathbf{d} \leq \mathbf{C}, \qquad \underline{\mathbf{y}} \leq \mathbf{y} \leq \overline{\mathbf{y}}. \tag{2.4}$$

The objective of the system operator is to maximize the social welfare, which includes the prosumer surplus (PS), generator surplus (GS), and merchandizing surplus (MS):

$$\mathsf{PS} := \sum_{k\in[n]} \left(\sum_{i\in[n_k]} u_i^k(d_i^k - x_i^k + C_i^k) - \lambda^k(h^k + Y^k) \right), \tag{2.5a}$$

$$\mathsf{GS} := \sum_{k\in[n]} \sum_{j\in[N_k]} \left(\lambda^k y_j^k - c_j^k(y_j^k) \right), \tag{2.5b}$$

$$\mathsf{MS} := \sum_{k\in[n]} \lambda^k h^k. \tag{2.5c}$$

where we have imposed the relationship (2.3) in deriving (2.5a). The social welfare that the system operator optimizes is the sum of PS, GS, and MS. After canceling terms, the social welfare can be written as

$$\mathcal{W}_B := \mathsf{PS} + \mathsf{GS} + \mathsf{MS} = \sum_{k\in[n]} \left(\sum_{i\in[n_k]} u_i^k(d_i^k + C_i^k - x_i^k) - \sum_{j\in[N_k]} c_j^k(y_j^k) \right). \tag{2.6}$$

The system operator's economic dispatch problem is then:

$$\begin{aligned} \max \quad & \mathcal{W}_B(\mathbf{h}, \mathbf{x} - \mathbf{d}, \mathbf{y}) \\ \text{subject to} \quad & (2.3) - (2.4). \end{aligned} \tag{2.7}$$

Assumption 2.3. *The system operator's economic dispatch problem* (2.7) *is feasible.*

Proposition 2.1 (Competitive Equilibrium)**.** *Under Assumptions 2.1-2.3, there exists a unique optimal solution* $(\mathbf{h}^*, (\mathbf{x}-\mathbf{d})^*, \mathbf{y}^*)$ *to* (2.7)*. Denote the optimal Lagrange multipliers of the first constraints of* (2.3) *by* $\boldsymbol{\lambda}$*. Then,* $(\mathbf{C} - \mathbf{x} + \mathbf{d})^* = \mathbf{z}^*$ *and* $\boldsymbol{\lambda}$ *satisfy Lemma 2.1;* $\mathbf{y}^*$ *and* $\boldsymbol{\lambda}$ *satisfy Lemma 2.2.*

The above proposition states that solving the system operator's problem (2.7) leads to a competitive equilibrium. From the perspective of prosumer i at node k, this means that

given the wholesale market price λ^k (the optimal Lagrange multiplier for the same node), the corresponding solution to its problem, which satisfies Lemma 2.1, is the same as the optimal decision made by the ISO via solving (2.7). This is true for all other prosumers and generators. Having all market participants being satisfied with the competitive equilibrium, it serves as a good benchmark for market efficiency.

Remark 2.1. *We note that in the case when prosumers are not allowed to sell back to the grid, each prosumer's x_i^k will be set to* 0*. Equivalently, there will be additional constraints $z_i^k = C_i^k - x_i^k + d_i^k \geq C_i^k$ for all prosumers in all locations, in both the prosumer's problem* (2.1) *and the system operator's economic dispatch problem* (2.7)*. With the additional constraints, we call this the* no participation model*, and its analysis will be similar to the direct participation model. Specifically, there will still be a competitive equilibrium, where we denote by $\hat{\mathbf{d}}$ the optimal amount of purchase of prosumers, $\hat{\mathbf{y}}$ the optimal production of the generators, $\hat{\mathbf{h}}$ the optimal injections by the system operator, $\hat{\boldsymbol{\lambda}}$ the corresponding market prices, and $\widehat{\mathcal{W}}_B^*$ the corresponding social welfare. Then, it follows that $\hat{\boldsymbol{\lambda}} \geq \boldsymbol{\lambda}$, $\hat{\mathbf{d}} \leq \mathbf{d}^*$, $\hat{\mathbf{y}} \leq \mathbf{y}^*$, and $\widehat{\mathcal{W}}_B^* \leq \mathcal{W}_B^*$, where $\mathcal{W}_B^*$ is the optimal social welfare given by* (2.7)*.*

2.3 Efficient Aggregation Model with an Unregulated Aggregator

The direct participation model introduced in the previous section is a benchmark: prosumers are allowed to participate directly and sell their production in the wholesale market. In reality, the supply capacities of prosumers are typically too small for consideration in the wholesale market. Also, computing the dispatch and settlement for a large number of prosumers raises an untenable computational burden on the system operator. The presence of DER aggregators brings benefits to the system, as they open the door for DER owners to participate and bring more flexibility to the grid. However, the profit-seeking nature of these aggregators can cause efficiency losses (Alshehri et al. [10, 11]). To resolve this, we propose in this section an *unregulated aggregation model* which achieves full market efficiency

under two-part pricing. In this model, prosumers sell part of their DER supply productions to an aggregator $\mathcal{A}$, based on the price offers made by $\mathcal{A}$. The interactions between $\mathcal{A}$ and prosumers are modeled as a Stackelberg game (Başar and Olsder [26]). The aggregator acts as a leader and announces a price pair (P_i^k, p_i^k) for each prosumer i at location k. We call the aggregator *unregulated* since $\mathcal{A}$ may announce any price pair to any prosumer for its own benefit. The prosumer follows by choosing the amount of energy to sell. If the prosumer decides to sell a nonzero fraction of its capacity to the aggregator, it pays the aggregator a participation fee P_i^k, and earns the price p_i^k for each unit of energy sold. The aggregator $\mathcal{A}$ then sells all the procured capacity to the wholesale market at the wholesale price λ^k. The goal of the aggregator $\mathcal{A}$ is to choose prices (P_i^k, p_i^k) that maximize its profit, while anticipating how DER owners would respond. Note that DER owners have access only to one aggregator; so $\mathcal{A}$ is in fact monopolistic, which further signifies the importance of our mechanism as it yields socially-optimal outcomes.

We also note that differential pricing is allowed in this model, i.e., the price pair (P_i^k, p_i^k) can be set differently for different prosumers. While it is reasonable to have varying prices depending on locations (Birge et al. [40]), there exist some debates on whether differential pricing should be allowed for prosumers from the same location. The legal issue is not the focus of this chapter. While we assumed differential pricing in the model, the equilibrium in the end has the same marginal price at each location, i.e., $p_i^k = \lambda^k,\ \forall i \in [n_k]$. The participation fee can be differentiated by, for example, mailing different coupons to different prosumers to encourage their participations, which is arguably more justifiable. In the remainder of this section, we show that this unregulated aggregation model achieves the same socially-optimal market outcomes as in the direct participation model.

2.3.1 Prosumer's Problem

Consider prosumer i at location k. Upon seeing the prices (P_i^k, p_i^k) announced by $\mathcal{A}$, prosumer i decides if it would sell part of its capacity to $\mathcal{A}$. If it chooses so, it would pay a fee P_i^k to $\mathcal{A}$, and receives p_i^k for each unit of energy sold. We may write prosumer i's payoff as

$$\pi_i^k(x_i^k, d_i^k) := \begin{cases} p_i^k x_i^k - P_i^k + u_i^k\left(d_i^k + C_i^k - x_i^k\right) - \lambda^k d_i^k, & \text{if } x_i^k > 0, \\ u_i^k\left(d_i^k + C_i^k\right) - \lambda^k d_i^k, & \text{if } x_i^k = 0. \end{cases} \tag{2.8}$$

Given $(P_i^k, p_i^k, \lambda^k)$, prosumer i solves: $\max_{x_i^k \in [0, C_i^k], d_i^k \in [0, Z - C_i^k + x_i^k]} \pi_i^k(x_i^k, d_i^k)$, where d_i^k is the amount of energy prosumer i purchases at wholesale market price λ^k, and x_i^k is the amount of energy it sells to the aggregator. For buying d_i^k at the wholesale price, prosumer i is charged $\lambda^k d_i^k$. If the prosumer does not sell ($x_i^k = 0$), it has a total of $d_i^k + C_i^k$ to consume, and its utility from consumption would be $u_i^k(d_i^k + C_i^k)$. If the prosumer chooses to sell $x_i^k > 0$ to $\mathcal{A}$, it is charged a participation fee P_i^k, and receives a compensation $p_i^k x_i^k$. The prosumer would have $d_i^k + C_i^k - x_i^k$ to consume in this case and its utility from consumption would be $u_i^k(d_i^k + C_i^k - x_i^k)$.

Let $x_i^{k*}(P_i^k, p_i^k, \lambda^k)$ and $d_i^{k*}(P_i^k, p_i^k, \lambda^k)$ denote the optimal response of prosumer j given aggregator's announced prices (P_i^k, p_i^k) and the wholesale market price λ^k. (Sometimes, we drop arguments for simplicity.) Note that if $p_i^k > \lambda^k$, the prosumer can arbitrage by buying at the price λ^k and selling at a higher price p_i^k. This will result in the prosumer's earning infinite payoff and the aggregator's losing infinite profit, which would be avoided by the aggregator. Therefore, we may without loss of generality restrict our discussion to the case when $p_i^k \leq \lambda^k, \forall k \in [n], i \in [n_k]$. In the case $p_i^k = \lambda^k$, similar to the direct participation model, we may enforce that x_i^k and d_i^k cannot both be nonzero. We then have the following lemma on the optimal response of prosumers.

Lemma 2.3. *Consider an arbitrary prosumer i at location k. Let (z_λ, z_p) be such that*

$$\frac{\partial u_i^k(z)}{\partial z}\Big|_{z=z_\lambda} = \lambda^k, \qquad \frac{\partial u_i^k(z)}{\partial z}\Big|_{z=z_p} = p_i^k. \tag{2.9}$$

Then, under Assumption 2.1, both (z_λ, z_p) exist and are unique. Furthermore, prosumer i's optimal response can be described as follows: If $C_i^k \le z_p$, then, we have $d_i^{k*} = \left[z_\lambda - C_i^k\right]^+$, $x_i^{k*} = 0$. If $C_i^k > z_p$, then, we have $x_i^{k*} = \left(C_i^k - z_p\right) \cdot \mathbb{1}\{\mathcal{X}\}$, $d_i^{k*} = 0$, where $\mathcal{X} := \left\{P_i^k \le p_i^k\left(C_i^k - z_p\right) + u_i^k\left(z_p\right) - u_i^k\left(C_i^k\right)\right\}$.

The above lemma states that if the capacity exceeds a certain value z_p (z_p is the value at which the marginal utility of consumption is equal to the aggregator's marginal price offer p_i^k) and the upfront fee P_i^k is not too high, then prosumers have an incentive to sell DER supply. If the capacity is small or the upfront fee is too high, then prosumers would not sell and prefer to consume locally.

2.3.2 Aggregator's Problem

The DER aggregator $\mathcal{A}$ collects power from prosumers and sells it in the wholesale market. By offering the prices (P_i^k, p_i^k) to each prosumer i at location k, $\mathcal{A}$ procures a total capacity of $\sum_{i\in[n_k]} x_i^{k*}(P_i^k, p_i^k)$ from location k. $\mathcal{A}$ then sells it at the wholesale market price λ^k. Given the wholesale market price λ^k, the aggregator's profit from prosumer i at location k is $\Pi_i^k\left(P_i^k, p_i^k\right) := P_i^k \mathbb{1}\left\{x_i^{k*}(P_i^k, p_i^k) > 0\right\} + (\lambda^k - p_i^k) x_i^{k*}(P_i^k, p_i^k)$. Anticipating the response functions $x_i^{k*}(P_i^k, p_i^k)$'s, it seeks to maximize its overall profit:

$$\max_{\mathbf{P}\ge\mathbf{0},\mathbf{p}\ge\mathbf{0}} \underbrace{\sum_{k\in[n]} \sum_{i\in[n_k]} \Pi_i^k\left(P_i^k, p_i^k\right)}_{=:\hat{\Pi}(\mathbf{P},\mathbf{p})}, \tag{2.10}$$

Aggregator $\mathcal{A}$'s profit is composed of two parts: the total participation fees charged to prosumers who sell positive amount of energy and the profits earned by reselling the procured energy in the wholesale market. We note that $\mathcal{A}$'s profit from prosumer i depends only on the response and the prices of prosumer i, and there is no coupling among prosumers' problems. It should be clear that the aggregator's problem (2.10) can be decomposed to optimizing (P_i^k, p_i^k) for each prosumer:

$$\max_{P_i^k \geq 0, p_i^k \geq 0} \Pi_i^k \left(P_i^k, p_i^k \right). \tag{2.11}$$

Solving (2.11) for each $k \in [n]$ and $i \in [n_k]$ would lead to the vectors $(\mathbf{P}^*, \mathbf{p}^*)$, which constitute an optimal solution for (2.10). We have the following result on the optimal decisions of the aggregator.

Lemma 2.4. *Under Assumption 2.1, consider an arbitrary prosumer i at location k, with z_λ as in* (2.9)*, and wholesale market price λ^k, When $z_\lambda < C_i^k$, the aggregator's optimal pricing decision is*

$$p_i^{k*} = \lambda^k, \qquad P_i^{k*} = \lambda^k (C_i^k - z_\lambda) + u_i^k(z_\lambda) - u_i^k(C_i^k). \tag{2.12}$$

When $z_\lambda \geq C_i^k$, prosumer i will not sell DER supply, i.e., $\Pi_i^k \left(P_i^k, p_i^k \right) = 0$, for any $(P_i^k, p_i^k) \in \mathbb{R}_+^2$.

Lemma 2.4 provides an optimal solution (P_i^{k*}, p_i^{k*}) to (2.11), and the collection $(\mathbf{P}^*, \mathbf{p}^*)$ form an optimal solution to (2.10). We also note that the optimal solution may not be unique in general: $\mathcal{A}$ may deviate from (2.12) by further decreasing p_i^{k*} and increasing P_i^{k*} to earn the same profit, while keeping the response of the prosumer i unchanged. This lemma states that there is an optimal pricing scheme which sets the marginal price p_i^k to be the wholesale market price λ^k, and all prosumers at the same location will be offered the

same $p_i^k = \lambda^k$. The participation fees P_i^k, however, will be charged differently for prosumers with different utility functions. Specifically, prosumer i sells $(C_i^k - z_\lambda)$ amount of energy, and earns $\lambda^k(C_i^k - z_\lambda)$ from selling. Its remaining capacity is then z_λ, from which its utility of consumption is $u_i^k(z_\lambda)$. The last term $u_i^k(C_i^k)$ is the utility that prosumer i retains if it does not sell any of its DER capacity and consumes all its production locally. Therefore, P_i^{k*} is charged as the *additional prosumer surplus from selling.* A keen reader would observe that in view of Lemmas 2.1 and 2.3, the optimal DER supply to the wholesale market is the same with and without the aggregator $\mathcal{A}$, thus making this unregulated aggregation model economically efficient with socially optimal market outcomes.

2.3.3 Aggregator-Prosumers Interaction as a Stackelberg Game

Let $\mathcal{G}(\boldsymbol{\lambda})$ denote the Stackelberg game among the aggregator and the prosumers for a given vector of wholesale market prices $\boldsymbol{\lambda}$. Aggregator $\mathcal{A}$ acts as a leader and sets the prices $(\mathbf{P}, \mathbf{p})$. The prosumer follows by responding with $x_i^{k*}(P_i^k, p_i^k, \lambda^k)$. We now define the equilibrium of the game.

Definition 2.1. $(\mathbf{P}^*, \mathbf{p}^*, \mathbf{x}^*(\mathbf{P}^*, \mathbf{p}^*))$ *constitutes a Stackelberg equilibrium of the game* $\mathcal{G}(\boldsymbol{\lambda})$ *if:*

- *Prosumers: For any* $x_i^k \in \left[0, C_i^k\right]$ *and for all* $k \in [n]$ *and* $i \in [n_k]$*, we have that*

$$\pi_i^k \left(x_i^{k*}(P_i^k, p_i^k), P_i^k, p_i^k, \lambda^k\right) \geq \pi_i^k \left(x_i^k, P_i^{k*}, p_i^{k*}, \lambda^k\right).$$

- *Aggregator: For all* $\mathbf{P} \geq \mathbf{0}, \mathbf{p} \geq \mathbf{0}$*, we have that*

$$\hat{\Pi}(\mathbf{P}^*, \mathbf{p}^*, \mathbf{x}^*(\mathbf{P}^*, \mathbf{p}^*), \boldsymbol{\lambda}) \geq \hat{\Pi}(\mathbf{P}, \mathbf{p}, \mathbf{x}^*(\mathbf{P}, \mathbf{p}), \boldsymbol{\lambda}).$$

We then have the following Stackelberg equilibrium for the game $\mathcal{G}(\boldsymbol{\lambda})$, which follows

directly from the prosumer's optimal response (Lemma 2.3) and the aggregator's optimal pricing (Lemma 2.4).

Proposition 2.2. *Under Assumption 2.1, the game* $\mathcal{G}(\boldsymbol{\lambda})$ *admits a Stackelberg equilibrium that satisfies* $P_i^{k*} = \left[\lambda^k(C_i^k - z_i^k) + u_i^k\left(z_i^k\right) - u_i^k(C_i^k)\right]^+$, $p_i^{k*} = \lambda^k$, $x_i^{k*}(P_i^{k*}, p_i^{k*}) = \left[C_i^k - z_i^k\right]^+$, *for each prosumer* i *at each location* k, *where* z_i^k *satisfies* $\frac{\partial u_i^k(z)}{\partial z}\Big|_{z=z_i^k} = \lambda^k$, *and* $d_i^{k*} = \left[z_i^k - C_i^k\right]^+$.

We note that the game $\mathcal{G}(\boldsymbol{\lambda})$ may admit other Stackelberg equilibria, but Proposition 2.2 provides the most economically efficient one. In case of non-uniqueness, this economically efficient equilibrium corresponds to the case where prosumers slightly prefer participation, i.e., if selling x_i^k amount of energy earns the prosumer i the same π_i^k as not selling, then it chooses to sell this x_i^k (alternatively, one can impose a slight perturbation of P_i^{k*} to $P_i^{k*} - \epsilon$, where $\epsilon > 0$, to ensure maximum DER supply). The above equilibrium is quite intuitive: $\mathcal{A}$ passes the location marginal price λ^k obtained from wholesale market outcomes as is to prosumers; so, $\mathcal{A}$ has no marginal profits from $\mathbf{p}^*$. Instead, $\mathcal{A}$ makes all of the profits from the upfront participation fees $\mathbf{P}^*$.

2.3.4 Generator's Problem

For a given wholesale market price λ^k, the conventional generators solve the same problem as described in Section 2.2.2, and the result of Lemma 2.2 still applies.

2.3.5 The Economic Dispatch Problem

The system operator solves an optimization problem similar to that in the direct participation model as described in Section 2.2.3. The network constraints (2.3) remain valid. Besides,

the ISO also considers participant-specific constraints:

$$\mathbf{0} \leq \mathbf{x} \leq \mathbf{C}, \quad \mathbf{0} \leq \mathbf{d} \leq \mathbf{Z} - \mathbf{C} + \mathbf{x}, \quad \underline{\mathbf{y}} \leq \mathbf{y} \leq \overline{\mathbf{y}}. \tag{2.13}$$

The objective of the system operator is to maximize the social welfare, which includes the prosumer surplus (PS), aggregator surplus (AS), generator surplus (GS), and merchandizing surplus (MS):

$$\begin{aligned}
\mathsf{PS} &:= \sum_{k\in[n]} \sum_{i\in[n_k]} \left(u_i^k(d_i^k - x_i^k + C_i^k) - \lambda^k d_i^k + p_i^k x_i^k - P_i^k \mathbb{1}\left\{x_i^k > 0\right\} \right), \\
\mathsf{AS} &:= \sum_{k\in[n]} \sum_{i\in[n_k]} \left(P_i^k \mathbb{1}\left\{x_i^k > 0\right\} + \lambda^k x_i^k - p_i^k x_i^k \right), \\
\mathsf{GS} &:= \sum_{k\in[n]} \sum_{j\in[N_k]} \left(\lambda^k y_j^k - c_j^k(y_j^k) \right), \\
\mathsf{MS} &:= \sum_{k\in[n]} \lambda^k h^k.
\end{aligned}$$

The social welfare is the sum of the above four terms. By the supply-demand balance $\mathbf{h} = \mathbf{D} - \mathbf{Y} - \mathbf{X}$, and after canceling terms, we write the social welfare as

$$\mathcal{W}_A := \mathsf{PS} + \mathsf{AS} + \mathsf{GS} + \mathsf{MS} = \sum_{k\in[n]} \left(\sum_{i\in[n_k]} u_i^k(d_i^k + C_j^i - x_i^k) - \sum_{j\in[N_k]} c_j^k(y_j^k) \right),$$

which is the same as $\mathcal{W}_B$. The system operator's economic dispatch problem is then:

$$\begin{aligned}
\max \quad & \mathcal{W}_A(\mathbf{h}, \mathbf{x}, \mathbf{d}, \mathbf{y}) \\
\text{subject to} \quad & (2.3),\ (2.13).
\end{aligned} \tag{2.15}$$

Assumption 2.4. *The system operator's economic dispatch problem* (2.15) *is feasible.*

As the system operator solves (2.15), the wholesale market prices $\boldsymbol{\lambda}$ are given by the

optimal Lagrange multiplier of the constraint (2.3). We then have the following proposition.

Proposition 2.3 (Competitive Equilibrium)**.** *Under Assumptions 2.1, 2.2, 2.4, there exists an optimal solution* $(\mathbf{h}^*, \mathbf{x}^*, \mathbf{d}^*, \mathbf{y}^*)$ *to* (2.15)*. Let* $\boldsymbol{\lambda}$ *denote the corresponding optimal Lagrange multipliers of constraints* (2.3)*. Then,* $(\mathbf{x}^*, \mathbf{d}^*)$ *are consistent with Lemma 2.3, given* $(\mathbf{P}^*, \mathbf{p}^*)$ *and* $\boldsymbol{\lambda}$*;* $(\mathbf{P}^*, \mathbf{p}^*)$ *are consistent with Lemma 2.4, given* $\mathbf{x}^*$ *and* $\boldsymbol{\lambda}$*;* $\mathbf{y}^*$ *is consistent with Lemma 2.2, given* $\boldsymbol{\lambda}$*.*

We now present the following theorem, which states that our proposed unregulated aggregation model achieves the same market efficiency as the benchmark direct participation model.

Theorem 2.1. *Let* $\mathcal{W}_A^*$ *be the optimal social welfare of* (2.15)*, and let* $\mathcal{W}_B^*$ *be the optimal social welfare of* (2.7)*. Then, we have that* $\mathcal{W}_A^* = \mathcal{W}_B^*$*. Further, we have that the optimal* $\mathbf{x}^*, \mathbf{d}^*, \mathbf{y}^*$ *(from Proposition 2.3) solving* (2.15) *are the same as those solving* (2.7)*, and the wholesale market price (optimal Lagrange multipliers of constraints* (2.3)*)* $\boldsymbol{\lambda}$ *in the unregulated aggregation model is the same as that in the direct participation model.*

In summary, under the proposed unregulated aggregation model, the aggregator procures energy from prosumers using two-part pricing, the aggregator would optimally pay the wholesale market price to the prosumers for each unit of energy procured, while the participation fee is differently charged to each prosumer as the additional consumer surplus when it sells this energy compared with not selling (which is dependent on its utility of consumption). Theorem 2.1, together with Proposition 2.3, implies that under the aggregator's two-part differential pricing scheme, the prosumers' optimal buying and selling behavior is exactly the same as those in the direct participation model. As a result, the social welfare achieved under the unregulated aggregation model matches that of the direct participation model, i.e., there is no loss of efficiency from the aggregation.

Remark 2.2. *The significance of Theorem 2.1 follows from the fact that via our unregulated*

aggregation model, DER aggregation through a profit-seeking aggregator $\mathcal{A}$ is equivalent to solving a socially-optimal economic dispatch model where $\mathcal{A}$ is absent. Hence, the potential efficiency loss due to the presence of a monopolistic profit-seeking aggregator $\mathcal{A}$ is off-set by two-part pricing. This is not possible via one-part pricing, as demonstrated in Alshehri et al. [10, 11] and in Section 2.3.6, where efficiency loss arises from the profit-seeking behavior of $\mathcal{A}$.

2.3.6 Numerical Example

In this subsection, we consider a stylized example to reveal some key insights. For ease of exposition, we consider a power system with one node, and there are no network constraints. We remark that, however, our theoretical results still hold for any power network. We assume that there is a fixed demand $\bar{D}$, and the first equality of (2.3) becomes $D + \bar{D} - Y - X = 0$. We consider one conventional generator with cost: $c(y) = \alpha y^2 + \beta y$. With one prosumer, we consider the isoelastic utility function u, with risk-aversion parameter η (Ljungqvist and Sargent [151]):

$$u(z) = \begin{cases} \frac{z^{1-\eta}-1}{1-\eta} & \eta \geq 0, \eta \neq 1, \\ \ln(z) & \eta = 1. \end{cases}$$

When $\eta = 0$, the prosumer is risk-neutral, for $\eta > 0$, the prosumer is risk-averse, and increasing η implies more risk-aversion. For $C > p^{-1/\eta}$, from Proposition 2.2, the Stackelberg equilirium prices of the game $\mathcal{G}(\lambda)$ are

$$p^* = \lambda, \quad P^* = \left[\lambda\left(C - \lambda^{-1/\eta}\right) + u\left(\lambda^{-1/\eta}\right) - u(C)\right]^+,$$

and the equilibrium response is $x^*(P^*, p^*) = C - \lambda^{-1/\eta}$. We note that by Lemma 2.3, the case when $C \leq \lambda^{-1/\eta}$ corresponds to $x^* = 0$ and $d^* = \lambda^{-1/\eta} - C$. The ISO's problem for

this single-prosumer case is then:

$$\begin{aligned} \max \quad & \mathcal{W}_A(x,d,y) := u(C-x+d) - c(y) \\ \text{subject to} \quad & \bar{D} + d - x - y = 0, \quad 0 \le y \le \bar{y}, \quad 0 \le x \le C, \quad 0 \le d \le Z - C + x. \end{aligned} \tag{2.16}$$

We compare the optimal value of (2.16) (optimal social welfare) with the following alternative models.

1) *No DER participation model*

In this model, the prosumer is restricted to not selling its energy, i.e., $x = 0$. The ISO's problem is the same as (2.16) with the additional constraint that $x = 0$, i.e.,

$$\begin{aligned} \max \quad & \mathcal{W}_N(d,y) := u(C+d) - c(y) \\ \text{subject to} \quad & \bar{D} + d - y = 0, \quad 0 \le y \le \bar{y}, \quad 0 \le d \le Z - C. \end{aligned} \tag{2.17}$$

2) *One-part pricing model ($P = 0$)*

In this model, the prosumer may sell its energy to the aggregator $\mathcal{A}$ at a fixed marginal price p set by $\mathcal{A}$, with no participation fees. In this model, the prosumer solves

$$\begin{aligned} & \max_{x,d} \pi(x,d) = u(C+d-x) - \lambda d + px \\ & \text{s.t.} \quad 0 \le x \le C, \quad 0 \le d \le Z - C + x. \end{aligned} \tag{2.18}$$

With the isoelastic utility function, an optimal response of the prosumer is given by $d^* = \left[\lambda^{-1/\eta} - C\right]^+$ and $x^* = \left[C - p^{-1/\eta}\right]^+$. Note that if $C \le \lambda^{-1/\eta}$, then $d^* \ge 0$ and $x^* = 0$, which leads to zero profit of the aggregator for any $0 \le p \le \lambda$, leading to no DER dispatch at the wholesale market level. If $C > \lambda^{-1/\eta}$, the aggregator would choose p such that $C > p^{-1/\eta}$, and solve $\max_{0 \le p \le \lambda} \Pi = (\lambda - p)x^*(p) = (\lambda - p)\left(C - p^{-1/\eta}\right)$. The Stackelberg equilibrium of the aggregator-prosumer game $\hat{\mathcal{G}}(\lambda)$ (a different game from $\mathcal{G}(\lambda)$, but similarly defined) is given by $(x^*(p^*), p^*)$ that satisfy $(1-\eta)p^* + \eta C p^{*1+1/\eta} = \lambda$, $x^*(p^*) = C - p^{-1/\eta}$.

If we let $\eta = 1$ (logarithmic utility), we have $p^*(\lambda) = \sqrt{\frac{\lambda}{C}}$, $x^*(p^*) = C - p^{-1}$.

The next question is how we define the corresponding social welfare. Note that the Stackelberg equilibrium $(x^*(p^*), p^*)$ does not yield an efficient market outcome; so, instead of using prosumer's utility in ISO's economic dispatch problem, we need to construct an induced function. Note that $x^*(p^*(\lambda)) = C - \sqrt{\frac{C}{\lambda}}$, and thus the inverse supply function for the prosumer is $p_A(x) = \frac{C}{(C-x)^2}$.

The system operator solves the economic dispatch problem given by

$$\begin{aligned} \max \qquad & \mathcal{W}_O(x, y) := -c(y) - \int_0^x p_A(\hat{x}) d\hat{x} \\ \text{subject to} \qquad & \bar{D} - x - y = 0, \quad 0 \le y \le \bar{y}, \quad 0 \le x \le C. \end{aligned} \tag{2.19}$$

Figure 2.3 illustrates the market outcomes for the above models. In our numerical study, we vary the capacity C from 0 to $\bar{D} = 100$. One can think of this as a proxy for renewable energy integration, $C = 0$ implies 0% renewable integration, and $C = \bar{D}$ implies 100% of the total inflexible load might be met exclusively with DER capacity. We also assume the following parameters: $\eta = 1,\ \alpha = 0.01,\ \beta = 1,\ Z = 1000,\ \bar{y} = 1000$. These parameters are picked such that it is more expensive to use the conventional generator than DERs, and conventional generators can fully meet the total system's demand. DER aggregation is expected to improve the social welfare, and this is demonstrated in Figure 2.3 (**Left**). Since our model is efficient, the improvements shown are the maximum possible ones. We then increase the number of prosumers to two, having the same capacity and η, and observe that the efficient welfare further improves. This is natural: the more DER capacities available, the more cheaper resources are available to the ISO. Next, we fix the number of prosumers to one, and study other aspects. The one-part pricing model is also expected to improve the social welfare, compared to no DER participation, but remains inefficient, as Figure 2.3 (**Middle**) shows. We quantify the efficiency loss by $\mathcal{W}_A^* - \mathcal{W}_O^*$ and plot it in Figure 2.3 (**Right**).

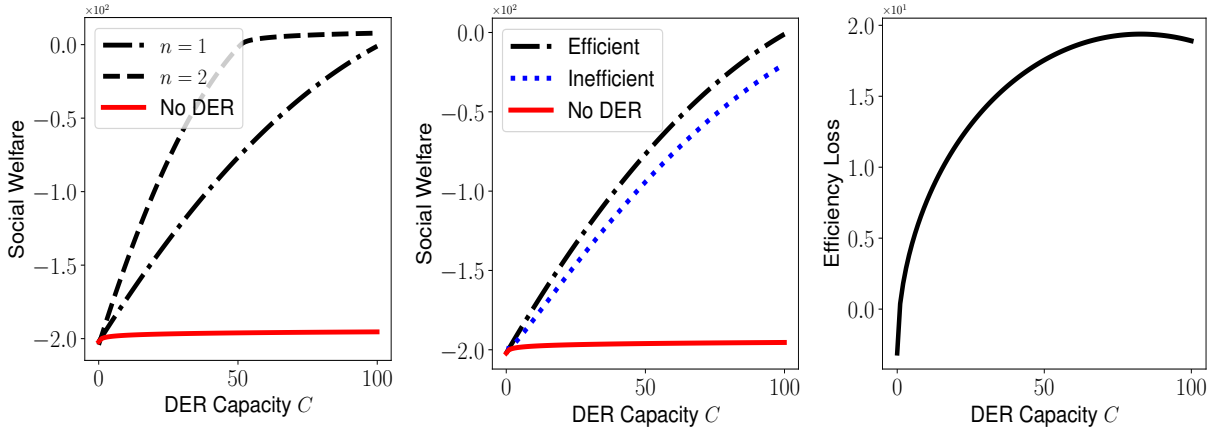

Figure 2.3: **Left:** Efficient aggregation vs. no DER integration. Adding more prosumers attains a higher social welfare. **Middle:** Comparison between the two extremes (efficient aggregation vs. no DER) and the one-part pricing model (inefficient). **Right:** Quantifying efficiency loss for the one-part pricing model.

2.4 Inefficient Aggregation Model with an Unregulated Aggregator

The unregulated aggregation model we proposed in Section 2.3 achieves full market efficiency. We have shown that, under the aggregator's two-part differential pricing scheme, the competitive equilibrium in Proposition 2.3 is achieved when the aggregator passes the wholesale market price λ^k to prosumers in location k, and charges a participation fee P_i^k to the prosumer i. As a result, all prosumers and generators' decisions are the same as those in the benchmark direct participation model (Section 2.2), which lead to the same social welfare being achieved. While our aggregation model is efficient, we recognize that for practical implementation, the differential pricing policy (the participation fee is prosumer-specific) might pose legal concerns. One may naturally ask the following question: what will happen if we have an unregulated profit-seeking aggregator who must impose a uniform two-part pricing policy? To this end, it is expected that the market outcome will no longer be socially optimal, and therefore, we term this model as *inefficient aggregation model*, but

it remains unclear how much efficiency loss one should expect. In this section, we make our attempt to address this question. For simplicity, we consider a single location with a single conventional generator (for multiple generators, we can equivalently consider their combined cost function), and assume that all prosumers' utility functions of consumption, as well as the generator's cost function, are quadratic. We show that, even with this specified type of utility and cost functions, it is difficult to obtain analytical results that quantify the social welfare. However, we provide an algorithm for solving the aggregator's profit-maximization problem with given specific parameters. Through numerical studies, we show that while the socially optimal outcome is no longer achieved, the reduction in social welfare (efficiency loss) seems to be mild.

2.4.1 Prosumer's Problem

Consider prosumer i who sees the price pair (P, p), and decides the amount of energy it buys and sells to maximize its overall payoff. In this section, we let $z_i := C_i + d_i - x_i$ be the actual amount of energy that prosumer i consumes, and assume that prosumer i's utility of consumption is $u_i(z_i) = a_i z_i^2 + b_i z_i$, with $a_i < 0$ and $b_i \gg 0$. Upon seeing the price pair (P, p) and the market price λ, the prosumer's problem may be written as:

$$\max_{z_i \geq 0} \pi_i(z_i) := u_i(z_i) - \lambda[z_i - C_i]^+ + p[C_i - z_i]^+ - P \cdot \mathbb{1}\{z_i < C_i\} \tag{2.20}$$

Since the prosumer's objective is the same as that in the unregulated aggregation model, we may apply Lemma 2.3 to obtain the optimal response:

$$z_i^* = \begin{cases} \frac{\lambda - b_i}{2a_i} & \text{if } > C_i, \\ \frac{p - b_i}{2a_i} & \text{if } < C_i \text{ and } P \leq \frac{(b_i - p)^2}{-4a_i} + pC_i - a_i C_i^2 - b_i C_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.21}$$

We say that a prosumer has the *intention to sell* if $\frac{p-b_i}{2a_i} - C_i < 0$. From (2.21), we may have four types of prosumers given a set of (λ, P, p):

- **Type A.** Prosumers with $\frac{p-b_i}{2a_i} - C_i < 0$ and $P \leq \frac{(b_i-p)^2}{-4a_i} + pC_i - a_iC_i^2 - b_iC_i$. Type A prosumers have the intention to sell, and are actually selling part of their capacities in the optimal response.

- **Type B.** Prosumers with $\frac{p-b_i}{2a_i} - C_i < 0$ and $P > \frac{(b_i-p)^2}{-4a_i} + pC_i - a_iC_i^2 - b_iC_i$. Type B prosumers have the intention to sell, but are *not* actually selling due to the high P.

- **Type C.** Prosumers with $\frac{\lambda-b_i}{2a_i} - C_i \leq 0 \leq \frac{p-b_i}{2a_i} - C_i$. Type C prosumers do not have the intention to sell, and are not buying either.

- **Type D.** Prosumers with $\frac{\lambda-b_i}{2a_i} - C_i \geq 0$. Type D prosumers are buying in the optimal response.

2.4.2 Generator's Problem

We assume in this section that the conventional generator's cost function is $c(y) = \alpha y^2 + \beta y$, with $\alpha > 0$ and $\beta \geq 0$. Given wholesale market price λ, the generator solves $\max_y \lambda y - c(y)$. The optimality condition, together with the supply-demand balance constraint $y = \sum_i z_i$, implies that

$$\lambda = \beta + 2\alpha y^* = \beta + 2\alpha \sum_i z_i^*. \tag{2.22}$$

2.4.3 Aggregator's Problem

In the inefficient aggregation model, the DER aggregator $\mathcal{A}$ offers the prices (P, p) to all prosumers, procures a capacity of $\sum_i [z_i^*(P,p)]^+$, and sells it at the wholesale market price λ. The aggregator now expects the market price to be affected by $\mathcal{A}$'s decisions, i.e., we shall

write the market price as $\lambda(P,p)$. Anticipating the response functions $z_i^*(P,p)$, the aggregator's objective now becomes

$$\max_{P\geq 0, p\geq 0} \sum_i P\cdot \mathbb{1}\{z_i^*(P,p)<C_i\} + (\lambda(P,p)-p)\sum_i (C_i - z_i^*(P,p))\cdot \mathbb{1}\{z_i^*(P,p)<C_i\}, \tag{2.23}$$

where $\lambda(P,p)$ and $z_i^*(P,p)$ jointly solve (2.21) and (2.22).

While our goal is to compute the social welfare (by summing over the prosumer surplus, generator surplus, and aggregator surplus), it is necessary to first solve (2.23) to obtain the aggregator's optimal (P,p) and thus λ, which will be used to compute the surplus of each party. This, however, creates significant difficulties since the objective (2.23) is generally not concave and not continuous, due to the fact that the discrete set of sellers/buyers changes with (P,p). It is thus impractical to derive analytical expressions on the optimal solution of (2.23). As a compromise, we provide Algorithm 2.1 that numerically finds the optimal (P,p) under certain assumptions, which are summarized below.

Assumption 2.5. *The generator's cost function is given by $c(y)=\alpha y^2+\beta y$ with $\alpha>0$ and $\beta>0$. Prosumer i's utility of consumption is given by $u_i(z_i)=a_iz_i^2+b_iz_i$ with $a_i<0$ and $b_i\gg 0$. Furthermore, for any two prosumers i and i', if $2a_iC_i+b_i\leq 2a_{i'}C_{i'}+b_{i'}$, then $a_i\geq a_{i'}$.*

In Assumption 2.5, we specify that the generator's cost function and the prosumers' utility of consumption are both quadratic. Moreover, prosumer i has the intention to sell if $2a_iC_i+b_i<p$; whether it actually sells, however, would also depend on if $P\leq \frac{(b_i-p)^2}{-4a_i}+pC_i-a_iC_i^2-b_iC_i$. The last part of Assumption 2.5 guarantees that, if prosumer i has more intention to sell than prosumer i', i.e., $2a_iC_i+b_i\leq 2a_{i'}C_{i'}+b_{i'}$, then, the thresholds on P for them to actually sell satisfy $\frac{(b_i-p)^2}{-4a_i}+pC_i-a_iC_i^2-b_iC_i\geq \frac{(b_{i'}-p)^2}{-4a_{i'}}+pC_{i'}-a_{i'}C_{i'}^2-b_{i'}C_{i'}$ for all $p\geq 2a_{i'}C_{i'}+b_{i'}$. In other words, for any given (P,p), if prosumer i is Type A, then

prosumer i' can be of any type; if prosumer i is Type B, then prosumer i' can only be Type B, Type C, or Type D, but *not* Type A.

Algorithm 2.1 Solving aggregator's problem

Sort the prosumers, and relabel them such that $2a_1C_1 + b_1 \le 2a_2C_2 + b_2 \le \cdots \le 2a_nC_n + b_n$.
Solution-Set $\leftarrow \emptyset$
for $k = 1, 2, \ldots, n-1$ **do**
Consider $p \in (2a_kC_k + b_k, 2a_{k+1}C_{k+1} + b_{k+1}]$. Then, prosumers $[1, k]$ have the intention to sell.
    for $t = 1, \ldots, k$ **do**
    Set $P = -\frac{(b_t - p)^2}{4a_t} + pC_t - a_tC_t^2 - b_tC_t$. Then, prosumers $[1, t]$ actually sell.
        for $s = k+1, \ldots, n$ **do**
        Let the set of buyers be $[s, n]$, i.e., $\{i \in [n] \mid 2a_iC_i + b_i > \lambda\} = [s, n]$.
            Solve the maximization problem (using any quadratic programming solver):

$$\max_{P,p,\lambda} t \cdot P + (\lambda - p) \cdot \sum_{i \in [1,t]} \left(C_i - \frac{p - b_i}{2a_i} \right) \tag{2.24a}$$

$$\text{s.t. } 2a_kC_k + b_k < p \le 2a_{k+1}C_{k+1} + b_{k+1} \tag{2.24b}$$

$$P = -\frac{(b_t - p)^2}{4a_t} + pC_t - a_tC_t^2 - b_tC_t \tag{2.24c}$$

$$2a_{s-1}C_{s-1} + b_{s-1} \le \lambda < 2a_sC_s + b_s \tag{2.24d}$$

$$\lambda = \beta + 2\alpha \sum_{i \in [1,t]} \left(\frac{p - b_i}{2a_i} - C_i \right) + 2\alpha \sum_{i \in [s,n]} \left(\frac{\lambda - b_i}{2a_i} - C_i \right) \tag{2.24e}$$

$$\frac{p - b_i}{2a_i} \ge C_i, \quad \forall i \in [1, t] \tag{2.24f}$$

$$\frac{\lambda - b_i}{2a_i} \le \frac{b_i}{-2a_i}, \quad \forall i \in [s, n] \tag{2.24g}$$

$$\sum_{i \in [1,t]} \frac{p - b_i}{2a_i} + \sum_{i \in [s,n]} \frac{\lambda - b_i}{2a_i} \ge 0 \tag{2.24h}$$

$$P, p, \lambda \ge 0 \tag{2.24i}$$

            if there exists a solution to (2.24) **then**
                add the solution (including the optimal value) to Solution-Set.
            end if
        end for
    end for
end for
Return the best solution (one with the highest objective value) in the Solution-Set.

We now explain the intuition behind Algorithm 2.1. Since main difficulty of solving (2.23) comes from the discrete set of sellers/buyers, we divide the problem (2.23) into subproblems by enumerating the possible combinations of the seller set and the buyer set. Specifically, we first note that, after sorting the prosumers by $2a_iC_i + b_i$, there are $(n-1)$ regions that p may

lie in. The outer "for loop" of Algorithm 2.1 (line 4) restricts p in one region such that only prosumers $[1, k]$ have the intention to sell. Second, we also note that in the optimal solution, $P = \frac{(b_i - p)^2}{-4a_i} + pC_i - a_i C_i^2 - b_i C_i$ for some i, i.e., P must be exactly on the threshold of some prosumer, because otherwise, the aggregator may increase P to the minimum threshold of the current sellers without changing any seller/buyer's behavior, thus strictly improving the aggregator's profit. In the middle "for loop" (line 5), we enumerate P on its thresholds for prosumers $[1, k]$, and by Assumption 2.5, we know that the ordering of these thresholds is the same as the ordering of $2a_i C_i + b_i$. Thus, by setting P to be prosumer t's threshold, prosumers $[1, t]$ actually sell (Type A), and prosumers $[t+1, k]$ have the intention to sell, but do not actually sell (Type B). Third, after restricting $p \in (2a_k C_k + b_k, 2a_{k+1} C_{k+1} + b_{k+1}]$, there are $(n - k - 1)$ possible regions that λ may lie in. The inner "for loop" restricts λ in one region such that the set of buyers is fixed as $[s, n]$ (Type D), and the set of prosumers who neither buy nor have the intention to sell (Type C) is $[k + 1, s - 1]$.

After fixing the set of prosumers of each type, we solve the quadratic program (2.24). The objective (2.24a) follows from (2.23) by specifying the t number of Type A prosumers from whom the aggregator collects P, as well as the amount of selling $\left(C_i - \frac{p - b_i}{2a_i}\right)$ from each of the t prosumers. Constraints (2.24b), (2.24c), and (2.24d) correspond to the restrictions from each of the "for loops". Constraint (2.24e) follows from (2.22). Constraint (2.24f) ensures that those Type A prosumers may only sell an amount not exceeding their capacities. Constraint (2.24g) specifies that the total consumption of any Type D prosumers is still in the increasing part of the quadratic utility function. Constraint (2.24h) ensures that the total energy net consumption by all prosumers is nonnegative.

Each time we solve (2.24) (using any quadratic programming solver), there may or may not exist a solution. If a solution exists, we add it to the Solution-Set. On a system with n prosumers, Algorithm 2.1 would solve the quadratic program (2.24) for $\frac{1}{6}n(n^2 - 1)$ times. In the end, the aggregator will impose (P, p) that correspond to the solution (in the Solution-

Set) with maximum profit.

The above argument readily leads to the following result.

Proposition 2.4. *Under Assumption 2.5, Algorithm 2.1 returns the optimal solution leading to the maximum profit for the aggregator.*

Once we know the aggregator's optimal profit, optimal decision (P, p) and the corresponding market price λ, the prosumers' surplus and generator surplus can also be computed, and we can thus obtain the corresponding social welfare.

2.4.4 Illustrative Example

To gain deeper insights, we consider a power system with one node and 4 prosumers. We pick the parameters such that one prosumer always sells its DER supply (Prosumer 1 is always Type A), one always buys from the system operator (Prosumer 4 is always Type D), and one neither sells nor buys (Prosumer 3 is always Type C). We vary the capacity of Prosumer 2 and observe the changes of its behavior (selling vs. buying) under the two aggregation models. With that, we sort the prosumers in an increasing order of $2a_1C_1 + b_1, \ldots, 2a_4C_4 + b_4$, and pick the following parameters: $b_1 = 100,\ b_2 = 125,\ b_3 = 150,\ b_4 = 200,\ C_1 = 150,\ C_2 \in [80, 160],\ C_3 = 100,\ C_4 = 1,\ a_i = -b_i/400,\ \forall i$. With these values, we vary C_2 from 80 to 160, and plot the social welfares and z_2 in Figure 2.4. As we vary the capacity of Prosumer 2, we observe that its type changes twice, creating three regions of interest, shaded in light grey, light red, and light green, respectively.

- For $C_2 \in [80, 115)$, Prosumer 2 does not have the intention to sell, but under uniform pricing, it does not buy either. In the efficient model, it buys only when C_2 is small enough.

- For $C_2 \in [115, 150)$, Prosumer 2 has the intention to sell, but under uniform pricing, it does not sell. On the other hand, in the efficient model, it always sells.

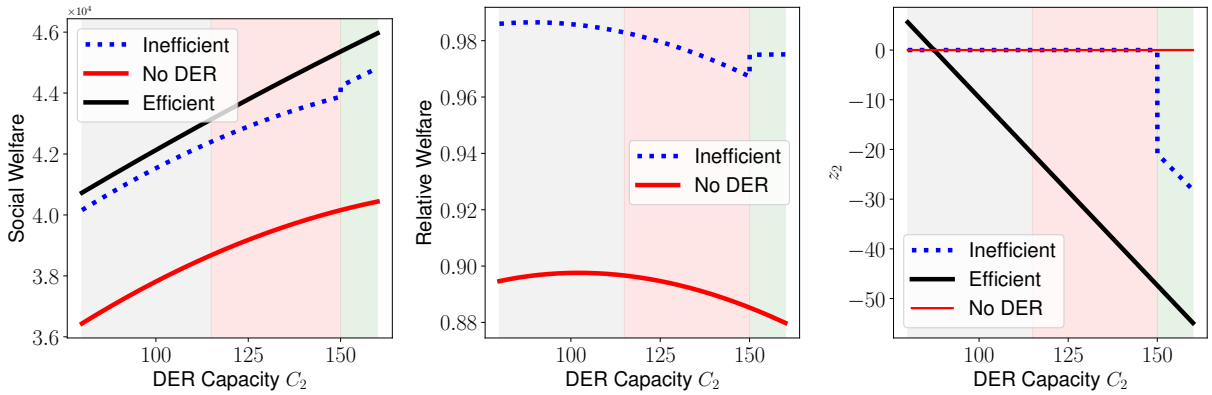

Figure 2.4: **Left:** Social welfare for each model vs. C_2. Uniform pricing is inefficient, but still yields welfare improvements. **Middle:** Relative social welfare of the inefficient model and the case in which there is no DER participation, to the efficient aggregation model. **Right:** z_2 as the capacity varies for the three models.

- For $C_2 \in [150, 160]$, Prosumer 2 has the intention to sell, and in fact sells. However, it does not sell as much DER as in the efficient model.

The above cases elaborate on the efficiency loss in view of Prosumer 2's behavior. With our parameter selection, other prosumers exhibit the same behavior (buying/selling) between the two models (efficient vs. inefficient). Figure 2.4 also illustrates the efficiency gap between the aggregation models. We note that uniform pricing achieves noticeable welfare improvement, compared to no DER participation. In fact, the inefficient model appears reasonably close to the efficient model. While we cannot ensure that this is generally the case, the inefficient model will always perform better than the no DER case. Finally, for $C_2 \geq 150$, as Prosumer 2 sells under both aggregation models, it does not sell as much DER supply as in the efficient model, causing an efficiency loss.

2.5 Efficient Aggregation Model with a Regulated Aggregator

In Section 2.3, we have shown that our unregulated aggregation model achieves full market efficiency if we allow the aggregator to charge differential participation fees P_i^k to each prosumer. When a profit-seeking aggregator is only allowed to impose uniform pricing policies,

we have shown in Section 2.4 that the aggregator's profit-maximization problem would be difficult to solve analytically. However, using our proposed Algorithm 2.1, we have illustrated through a numerical example that there will be some efficiency loss, though the degree of loss seems to be mild. It is thus interesting to study the following question: is there a mechanism that imposes a uniform pricing policy on the prosumers, while preserving full market efficiency? To address this question, we propose in this section a *regulated aggregation model* where the aggregator $\mathcal{A}$ is fully regulated, i.e., $\mathcal{A}$ does not make any pricing decisions, but is guaranteed some positive profit. We show that, with a regulated aggregator, we can have a uniform two-part pricing policy, i.e., all prosumers from the same location receive the same (P^k, p^k), while still achieving full market efficiency. We also remark that electric power utilities in most countries are in fact regulated, so the regulated aggregator's role can be potentially fulfilled by a utility company, which is consistent with the fact that they operate the distribution lines and directly interact with end-consumers.

Before proceeding to the model, we first introduce some notations from the direct participation model and the no participation model. In the competitive equilibrium of the direct participation model, as described in Proposition 2.1, we denote by λ^{k*} the market price at location k. Let d_i^{k*} (respectively, x_i^{k*}) be the equilibrium amount of energy bought (respectively, sold) by prosumer i at location k. Let z_i^k be such that $\frac{\partial u_i^k(z)}{\partial z}\Big|_{z=z_i^k} = \lambda^{k*}$. Then from Lemma 2.1, we have that $z_i^k = C_i^k - x_i^{k*} + d_i^{k*}$. In the competitive equilibrium of the no participation model, as noted in Remark 2.1, the market price at location k is denoted by $\hat{\lambda}^k$, and the prosumer i at location k buys $\hat{d}_i^k$ in the equilibrium. Let $\hat{z}_i^k$ be such that $\frac{\partial u_i^k(z)}{\partial z}\Big|_{z=\hat{z}_i^k} = \hat{\lambda}^k$. Then it follows that $\hat{z}_i^k = C_i^k + \hat{d}_i^k$ if $\hat{d}_i^k > 0$ and $\hat{z}_i^k \leq C_i^k + \hat{d}_i^k$ if $\hat{d}_i^k = 0$. We also note that $\lambda^{k*} \leq \hat{\lambda}^k$, and thus $\hat{d}_i^k \leq d_i^{k*}$ and $\hat{z}_i^k \leq z_i^k$.

Now, we proceed to the regulated aggregation model. The aggregator $\mathcal{A}$ sets a uniform price pair (P^k, p^k) for all prosumers from location k. The prices are exogenously given to the aggregator and do not necessarily maximize the aggregator's profit. The system operator

will impose an extra charge F_i^k on prosumer i at location k, where

$$F_i^k := \hat{d}_i^k \left(\hat{\lambda}^k - \lambda^{k*}\right) = \begin{cases} \left(\hat{z}_i^k - C_i^k\right)\left(\hat{\lambda}^k - \lambda^{k*}\right), & \text{if } \hat{z}_i^k > C_i^k, \\ 0, & \text{otherwise.} \end{cases} \tag{2.25}$$

The vector of these charges for all prosumers is denoted by $\mathbf{F}$. We note that, although this extra charge by the system operator F_i^k is different for different prosumers, it is a uniform policy for all prosumers from the same location: the charge is a linear function, with the linear factor being fixed as $\left(\hat{\lambda}^k - \lambda^{k*}\right)$, on $\hat{d}_i^k$, the amount of purchase made by the prosumer under the no participation model. This fee can be charged to prosumers in their utility bills. The system operator will forward this charge to the aggregator, so the aggregator earns profits from both the participation fee paid by those selling prosumers and the charges (forwarded by the system operator) paid by the buying prosumers. We will show that, with carefully selected prices and charges, the behaviors of the prosumers and the generators are the same as those in the direct participation model; thus, the same social welfare is achieved. In the end of this section, we will also discuss that this extra charge $\mathbf{F}$ is not necessary to achieve full market efficiency, but is imposed to offer the aggregator a higher profit. We now describe each party in the model separately.

2.5.1 Regulated Aggregator

In the regulated aggregation model, the aggregator $\mathcal{A}$ is not allowed to make the pricing decisions to maximize profit. Instead, $\mathcal{A}$ is regulated to set the following uniform prices (P^k, p^k):

$$p^k = \lambda^{k*}, \qquad P^k = \min_{i \in [n_k] | x_i^{k*} > 0} u_i^k(z_i^k) + \lambda^{k*}(C_i^k - z_i^k) - u_i^k(C_i^k). \tag{2.26}$$

The vectors of these prices at all locations are denoted by $(\mathbf{P}, \mathbf{p})$. The marginal price p^k is set to λ^{k*}, which is the market price in the direct participation model. The participation

fee P^i is set to be the minimum additional prosumer surplus gained from selling among those prosumers who would sell a positive amount of energy at the market price λ^{k*}, which ensures the same number of prosumers selling DER supply as in the efficient unregulated aggregation model. In addition, the aggregator also receives the extra charges F_i^k, forwarded by the system operator, where $F_i^k > 0$ if and only if $\hat{d}_i^k > 0$. Let λ^k be the market price and x_i^k be the prosumer's amount of energy to sell under the regulated aggregation model. The regulated aggregator earns the profit from location k: $\Pi^k = \sum_{i\in[n_k]|x_i^k>0} \left[\left(\lambda^k - p^k\right) x_i^k + P^k\right] + \sum_{i\in[n_k]} F_i^k$. As we will show later, in equilibrium, $\lambda^k = \lambda^{k*}$ and $x_i^k = x_i^{k*}$, which lead to $\Pi^k = \sum_{i\in[n_k]|x_i^{k*}>0} P^k + \sum_{i\in[n_k]} \hat{d}_i^k \left(\hat{\lambda}^k - \lambda^{k*}\right)$.

2.5.2 Prosumer's Problem

Consider prosumer i at location k. The prosumer sees the price pair (P^k, p^k) as well as the fee F_i^k charged by the system operator, and then decides the amount of energy it buys and sells to maximize its overall payoff. The prosumer's problem may be written as:

$$\max_{\substack{x_i^k\in[0,C_i^k]\\ d_i^k\in[0,Z-C_i^k+x_i^k]}} \pi_i^k(x_i^k, d_i^k) := \begin{cases} p^k x_i^k - P^k + u_i^k\left(d_i^k + C_i^k - x_i^k\right) - \lambda^k d_i^k - F_i^k, & \text{if } x_i^k > 0, \\ u_i^k\left(d_i^k + C_i^k\right) - \lambda^k d_i^k - F_i^k, & \text{if } x_i^k = 0. \end{cases} \tag{2.27}$$

We note that the prosumer's objective is the same as that in the unregulated aggregation model, except that the additional F_i^k is now being charged by the system operator.

2.5.3 Generator's Problem

For a given wholesale market price λ^{k*}, the conventional generators solve the same problem as described in Section 2.2.2, and the result of Lemma 2.2 still applies under the current model.

2.5.4 The Economic Dispatch Problem

The system operator now solves an optimization problem similar to that in the unregulated aggregation model as described in Section 2.3.5. The network constraints (2.3) and the constraints on the decisions of prosumers and generators (2.13) remain valid. The objective of the system operator is to maximize the social welfare, which includes the prosumer surplus (PS), aggregator surplus (AS), generator surplus (GS), and merchandizing surplus (MS):

$$
\begin{aligned}
\mathsf{PS} &:= \sum_{k\in[n]} \sum_{i\in[n_k]} \left(u_i^k(d_i^k - x_i^k + C_i^k) - \lambda^k d_i^k + p^k x_i^k - P^k \mathbb{1}\left\{x_i^k > 0\right\} - F_i^k \right), \\
\mathsf{AS} &:= \sum_{k\in[n]} \sum_{i\in[n_k]} \left(P^k \mathbb{1}\left\{x_i^k > 0\right\} + \lambda^k x_i^k - p^k x_i^k + F_i^k \right), \\
\mathsf{GS} &:= \sum_{k\in[n]} \sum_{j\in[N_k]} \left(\lambda^k y_j^k - c_j^k(y_j^k) \right), \\
\mathsf{MS} &:= \sum_{k\in[n]} \lambda^k h^k.
\end{aligned}
$$

The social welfare is the sum of the above four terms. By the supply-demand balance $\mathbf{h} = \mathbf{D} - \mathbf{Y} - \mathbf{X}$, and after canceling terms, we write the social welfare as

$$
\mathcal{W}_R := \mathsf{PS} + \mathsf{AS} + \mathsf{GS} + \mathsf{MS} = \sum_{k\in[n]} \left(\sum_{i\in[n_k]} u_i^k(d_i^k + C_i^k - x_i^k) - \sum_{j\in[N_k]} c_j^k(y_j^k) \right),
$$

which is again the same as $\mathcal{W}_B$. The system operator's economic dispatch problem is then:

$$
\begin{aligned}
&\max \qquad \mathcal{W}_R(\mathbf{h}, \mathbf{x}, \mathbf{d}, \mathbf{y}) \\
&\text{subject to} \qquad (2.3),\ (2.13).
\end{aligned}
\tag{2.29}
$$

Note that (2.29) is exactly the same problem as (2.15). As a result, the equilibrium

$(\mathbf{h}^*, \mathbf{x}^*, \mathbf{d}^*, \mathbf{y}^*)$ in Proposition 2.3 optimally solves (2.29), and the corresponding wholesale market prices $\boldsymbol{\lambda}$ are the same as in the unregulated aggregation model. We then have the following proposition.

Proposition 2.5 (Competitive Equilibrium)**.** *Under Assumptions 2.1, 2.2, 2.4, the tuple* $(\mathbf{h}^*, \mathbf{x}^*, \mathbf{d}^*, \mathbf{y}^*)$ *from Proposition 2.3 is an optimal solution to* (2.15)*. Let* $\boldsymbol{\lambda}$ *denote the corresponding optimal Lagrange multipliers of constraints* (2.3)*. Then, given* $(\mathbf{P}, \mathbf{p})$ *as in* (2.26) *and* $\mathbf{F}$ *as in* (2.25)*, we have that* $(\mathbf{x}^*, \mathbf{d}^*)$ *are optimal solutions to* (2.27)*, given* $(\mathbf{P}, \mathbf{p})$ *and* $\boldsymbol{\lambda}$*;* $\mathbf{y}^*$ *is consistent with Lemma 2.2, given* $\boldsymbol{\lambda}$*.*

The above argument, together with Theorem 2.1, imply that the regulated aggregation model again achieves the same market efficiency as the benchmark direct participation model.

Theorem 2.2. *Let* $\mathcal{W}_R^*$ *be the optimal social welfare of* (2.29)*, and let* $\mathcal{W}_B^*$ *be the optimal social welfare of* (2.7)*. Then, we have that* $\mathcal{W}_R^* = \mathcal{W}_B^*$*. Further, we have that the optimal* $\mathbf{x}^*, \mathbf{d}^*, \mathbf{y}^*$ *solving* (2.29) *are the same as those solving* (2.7)*.*

2.5.5 Discussions

In these discussions, we compare the prosumer surplus and aggregator surplus under two efficient aggregation models (leaving out the inefficient aggregation model since there are no analytical results from there). We also discuss the implications of the price pair $(\mathbf{P}, \mathbf{p})$ and the extra charge $\mathbf{F}$.

2.5.5.1 Comparison of Prosumer Surplus and Aggregator Surplus.

In this subsection, we explain that, comparing the no participation model with both efficient aggregation models (introduced in Section 2.3 and this section), each prosumer's surplus is not reduced, and sometimes increased. Furthermore, in the unregulated aggregation model and the regulated aggregation model, the profits of the aggregator could come from different

sets of prosumers, and the prosumer's surplus would thus be different. To further elaborate our discussions, we may divide the prosumers into three *types*, which are classified based on their utility functions and corresponding DER capacities, and are ordered from I to III in an increasing order according to their willingness to sell. Type I prosumers are those with with $d_i^{k*} \geq \hat{d}_i^k > 0$, i.e., $\frac{\partial u_i^k}{\partial z}\Big|_{z=C_i^k} > \hat{\lambda}^k \geq \lambda^{k*}$, and are always buying. These could be prosumers with small capacities (e.g., a prosumer with only one panel) and/or a high preference to consume energy, as captured by their utility functions. On the other extreme, Type III prosumers are those with $\hat{d}_i^k = d_i^{k*} = 0$, i.e., $\hat{\lambda}^k \geq \lambda^{k*} \geq \frac{\partial u_i^k}{\partial z}\Big|_{z=C_i^k}$, and they never buy but sell after aggregation. They can be those prosumers with many solar panels and batteries that could be discharged to cover their energy needs. Type II prosumers are those with $d_i^{k*} > \hat{d}_i^k = 0$, i.e., $\hat{\lambda}^k \geq \frac{\partial u_i^k}{\partial z}\Big|_{z=C_i^k} > \lambda^{k*}$, and they represent the middle ground, as they are buying after aggregation. They can be those prosumers with an intermediate number solar panels and limited storage.

- **Type I.** Prosumers with $d_i^{k*} \geq \hat{d}_i^k > 0$, i.e., $\frac{\partial u_i^k}{\partial z}\Big|_{z=C_i^k} > \hat{\lambda}^k \geq \lambda^{k*}$.
  - In the no participation model, a Type I prosumer purchases $\hat{d}_i^k = \hat{z}_i^k - C_i^k$ amount of energy, and its prosumer surplus is

$$u_i^k(\hat{z}_i^k) - \hat{\lambda}^k \left(\hat{z}_i^k - C_i^k\right). \tag{2.30}$$

  - In the unregulated aggregation model, the prosumer instead purchases $d_i^{k*} = z_i^k - C_i^k$ amount of energy, and its prosumer surplus is

$$u_i^k(z_i^k) - \lambda^{k*} \left(z_i^k - C_i^k\right). \tag{2.31}$$

  - In the regulated aggregation model, the prosumer still purchases $d_i^{k*} = z_i^k - C_i^k$

amount of energy, but is charged F_i^k. Its prosumer surplus becomes

$$\begin{aligned} & u_i^k(z_i^k) - \lambda^{k*}\left(z_i^k - C_i^k\right) - F_i^k \\ &= u_i^k(z_i^k) - \lambda^k\left(z_i^k - C_i^k\right) - \left(\hat{z}_i^k - C_i^k\right)\left(\hat{\lambda}^k - \lambda^{k*}\right). \end{aligned} \tag{2.32}$$

We note that the Type I prosumers' surplus is reduced in the regulated aggregation model, comparing to the unregulated aggregation model. However, these prosumers would still prefer the regulated aggregation model over no participation model. To see this, note that

$$\begin{aligned} & (2.32) - (2.30) \\ &= u_i^k(z_i^k) - \lambda^{k*}\left(z_i^k - C_i^k\right) - \left(\hat{z}_i^k - C_i^k\right)\left(\hat{\lambda}^k - \lambda^{k*}\right) - \left(u_i^k(\hat{z}_i^k) - \hat{\lambda}^k\left(\hat{z}_i^k - C_i^k\right)\right) \\ &= u_i^k(z_i^k) - u_i^k(\hat{z}_i^k) - \lambda^{k*}\left(z_i^k - \hat{z}_i^k\right) \geq 0, \end{aligned}$$

where the inequality follows since u_i^k is concave and $\lambda^{k*} = \frac{\partial u_i^k}{\partial z}\Big|_{z=z_i^k}$.

Additionally, we note that the system operator could actually set any $F_i^k \leq (2.31) - (2.30)$ in the regulated aggregation model, and the prosumers would still prefer the regulated model over the no participation model. However, we defined F_i^k as in (2.25) so that it is a uniform policy for all prosumers: everyone is charged a fee that is linear in its amount of purchase in the no participation model $\hat{d}_i^k$, where the linear factor is the difference of the market price $\hat{\lambda}^k - \lambda^{k*}$.

- **Type II.** Prosumers with $d_i^{k*} > \hat{d}_i^k = 0$, i.e., $\hat{\lambda}^k \geq \frac{\partial u_i^k}{\partial z}\Big|_{z=C_i^k} > \lambda^{k*}$.
  - In the no participation model, a Type II prosumer does not purchase from the grid, and its prosumer surplus is $u_i^k(C_i^k)$.
  - In the unregulated aggregation model, the prosumer purchases $d_i^{k*} = z_i^k - C_i^k$

amount of energy, and its prosumer surplus is $u_i^k(z_i^k) - \lambda^{k^*}\left(z_i^k - C_i^k\right)$.

– In the regulated aggregation model, the prosumer still purchases $d_i^{k^*} = z_i^k - C_i^k$ amount of energy, and is charged $F_i^k = \hat{d}_i^k\left(\hat{\lambda}^k - \lambda^{k^*}\right) = 0$. Its prosumer surplus is thus again $u_i^k(z_i^k) - \lambda^{k^*}\left(z_i^k - C_i^k\right)$.

We note that $u_i^k(z_i^k) - \lambda^{k^*}\left(z_i^k - C_i^k\right) - u_i^k(C_i^k) > 0$; thus, Type II prosumers' surplus is strictly smaller in the no participation model. Since $F_i^k = 0$, they are indifferent between the unregulated aggregation model and the regulated aggregation model. In theory, the system operator could set any charge $F_i^k \leq u_i^k(z_i^k) - \lambda^{k^*}\left(z_i^k - C_i^k\right) - u_i^k(C_i^k)$ to extract part of these prosumers' additional surplus and forward those to the aggregator. However, there does not seem to be a simple uniform policy for all prosumers. We therefore set $F_i^k = 0$ for Type II prosumers since $\hat{d}_i^k = 0$, and these prosumers are strictly better off in both aggregation models.

- **Type III.** Prosumers with $\hat{d}_i^k = d_i^{k^*} = 0$, i.e., $\hat{\lambda}^k \geq \lambda^{k^*} \geq \frac{\partial u_i^k}{\partial z}\Big|_{z=C_i^k}$.

  – In the no participation model, a Type III prosumer does not purchase from the grid, and its prosumer surplus is $u_i^k(C_i^k)$.

  – In the unregulated aggregation model, the prosumer sells $x_i^{k^*} = C_i^k - z_i^k$ amount of energy for a unit price λ^{k^*}, but also pays the participation fee $P_i^k = u_i^k\left(C_i^k - x_i^{k^*}\right) + \lambda^{k^*}\left(C_i^k - z_i^k\right) - u_i^k(C_i^k)$. Thus, its prosumer surplus is $u_i^k(z_i^k) + \lambda^{k^*}x_i^{k^*} - P_i^k = u_i^k(C_i^k)$.

  – In the regulated aggregation model, the prosumer still sells $x_i^{k^*} = C_i^k - z_i^k$ amount of energy for a unit price λ^{k^*}, but the participation fee is reduced to $P^k = \min_{i\in[n_k]|x_i^{k^*}>0} u_i^k(z_i^k) + \lambda^{k^*}(C_i^k - z_i^k) - u_i^k(C_i^k)$. Thus, its prosumer surplus becomes

$$u_i^k(z_i^k) + \lambda^{k^*}x_i^{k^*} - P^k$$

$$= u_i^k(z_i^k) + \lambda^{k^*} x_i^{k^*} - \left[\min_{i \in [n_k] | x_i^{k^*} > 0} u_i^k(z_i^k) + \lambda^{k^*}(C_i^k - z_i^k) - u_i^k(C_i^k) \right]. \quad (2.33)$$

We note that Type III prosumers earn the same surplus in both the no participation model and the unregulated aggregation model, due to the fact that the aggregator in the unregulated aggregation model extracts all additional prosumer surplus through participation fess P_i^k. In the regulated aggregation model, the participation fee is set to the minimum additional surplus among all Type III prosumers in the same location. Note that (2.33) $\geq u_i^k(C_i^k)$, where the equality holds if and only if

$$i = \arg\min_{i \in [n_k] | x_i^{k^*} > 0} u_i^k(z_i^k) + \lambda^{k^*}(C_i^k - z_i^k) - u_i^k(C_i^k). \quad (2.34)$$

In other words, those prosumers who have minimum additional surplus from selling at price λ^{k^*} (compared to no participation) are indifferent between the unregulated aggregation model and the regulated aggregation model. For other Type III prosumers, such that (2.34) does not hold, their prosumers' surplus (2.33) $> u_i^k(C_i^k)$, i.e., they earn strictly higher surplus from the regulated aggregation model, and thus would strictly prefer the regulated aggregation model over the unregulated aggregation model.

Table 2.1 summarizes the prosumer surplus under the three models. We remark that no prosumer is worse off after either aggregation model. However, prosumers of different types may prefer different aggregation models.

Prosumer Surplus				
Prosumer Types	No part.	Unreg. agg.	Reg. agg.	Note
Type I	Low	Highest	High	always buying
Type II	Low	High	High	buying after aggregation
Type III	Low	Low	High	selling after aggregation

Table 2.1: Comparison of prosumer surplus under different models

We then move our eyes from prosumers to the aggregator. In the unregulated aggregation

model, the aggregator earns profit

$$\mathsf{AS}^{\mathsf{U}} := \sum_{k\in[n]} \sum_{i\in[n_k]|i\in\textbf{TYPE III}} u_i^k(z_i^k) + \lambda^{k*}\left(C_i^k - z_i^k\right) - u_i^k(C_i^k).$$

In the regulated aggregation model, the aggregator earns profit

$$\mathsf{AS}^{\mathsf{R}} := \sum_{k\in[n]} \sum_{i\in\textbf{TYPE I}} \hat{d}_i^k\left(\hat{\lambda}^k - \lambda^{k*}\right) + \sum_{k\in[n]} |\{i\in[n_k] \mid i\in\textbf{TYPE III}\}| \cdot P^k,$$

where we recall that $P^k = \min_{i\in[n_k]|i\in\textbf{TYPE III}} u_i^k(z_i^k) + \lambda^{k*}(C_i^k - z_i^k) - u_i^k(C_i^k)$. We note that in the unregulated aggregation model, all of the aggregator's profit comes from Type III prosumers, and all Type III prosumers earn zero additional surplus compared with the no participation model. Moreover, Type I prosumers benefit the most in the unregulated aggregation model by retaining all additional surplus from buying energy at a lower market price (compared with the no participation model). In the regulated aggregation model, however, the aggregator's profit comes from both Type I prosumers and Type III prosumers. Type I prosumers now earn less surplus compared with the unregulated aggregation model (still better off compared with the no participation model), and the lost surplus is transferred to the aggregator's profit, which is the first term of AS^{R}. Type III prosumers get more surplus (except those marginal prosumers that satisfy $i = \arg\min_{i\in[n_k]|x_i^{k*}>0} u_i^k(z_i^k)+\lambda^{k*}(C_i^k-z_i^k)-u_i^k(C_i^k)$, who are indifferent) in the regulated model, compared with the unregulated model, and thus the profit earned by the aggregator from Type III prosumers becomes the second term of AS^{R}, instead of AS^{U}.

2.5.5.2 Discussion on the Prices and Extra Charge.

While we have provided some intuitions as we introduce the results, we now further discuss the implications of the prices and the extra charge. In both efficient aggregation models,

the wholesale market price is the same as that in the direct participation model. This fact by itself ensures that all prosumers who make a net purchase under the direct participation model (Type I and Type II prosumers) still purchase the same amount under both efficient aggregation models. Also in both efficient aggregation models, the equilibrium per-unit price offered by the aggregator is the same as the wholesale market price under the direct participation model, i.e., $p_i^k = \lambda^{k*}, \forall i \in [n_k], \forall k \in [n]$, and there is no price discrimination in $\mathbf{p}$ for prosumers from the same location. As a result, the aggregator earns zero profit from reselling the procured electricity.

In the unregulated aggregation model, the connection fee P_i^k is set to be each prosumer's additional surplus from selling: if prosumer i does not sell or buy, it retains $u_i^k(C_i^k)$; under the unregulated aggregation model, it sells $(C_i^k - z_i^k)$ at the price λ^{k*}, and has z_i^k to consume, which leads to a prosumer surplus of $u_i^k(z_i^k) + \lambda^{k*}(C_i^k - z_i^k)$ before the connection fee. This prosumer's selling behavior would not change, as long as the connection fee is charged such that $P_i^k \leq u_i^k(z_i^k) + \lambda^{k*}(C_i^k - z_i^k) - u_i^k(C_i^k)$. Thus, the maximum connection fee the aggregator can impose, without changing the prosumer's selling amount, is exactly that prosumer's additional surplus from selling.

When the aggregator is regulated, the connection fee is charged as the minimum additional prosumer surplus from selling, i.e., $P^k = \min_{i \in [n_k] | i \in \textbf{TYPE III}} u_i^k(z_i^k) + \lambda^{k*}(C_i^k - z_i^k) - u_i^k(C_i^k)$, so that for those sellers (Type III prosumers), it always holds that $P^k \leq u_i^k(z_i^k) + \lambda^{k*}(C_i^k - z_i^k) - u_i^k(C_i^k), \forall i \in [n_k] \cap \text{TYPE III}$. This guarantees that all sellers continue to sell the same amount as they would do in the direct participation model. With this set of (P^k, p^k), and without any extra charges (let $F_i^k = 0$ for all prosumers in all locations), all prosumers from location k have the exact same behavior as in the direct participation model, and we achieve full market efficiency without any type of price discrimination within the same location. The only potential drawback is that, compared with the unregulated model where the aggregator extracts all additional prosumer surplus from selling, most of

the additional prosumer surplus might be retained by the prosumers. In those locations with high population, the participation fee P^k could be small or even negligible, and the aggregator's profit would be small. Consider the case when there exists some $i' \in [n_k]$ such that $u^k_{i'}(z^k_{i'}) + \lambda^{k*}(C^k_{i'} - z^k_{i'}) - u^k_{i'}(C^k_{i'}) = 0$, then at this location, $P^k = 0$, which means the aggregator would earn zero profit from Type III prosumers in this location. If $F^k_i = 0$ for all prosumers, then the aggregator would earn zero profit from location k. This is the reason we impose this extra charge F^k_i, which is strictly positive for Type I prosumers, so that the aggregator always earn positive profit from this location even with the existence of a marginal Type III prosumer. On the other hand, if it is acceptable that the aggregator earns very little profit from certain locations, then the extra charge $\mathbf{F}$ is not necessary, and the market is still fully efficient.

As a summary, the regulated aggregation model (with or without the extra charge $\mathbf{F}$) is a good candidate for implementation as it avoids the differential pricing policy and achieves all of the following desirable outcomes:

1. Full market efficiency as in the direct participation model.
2. No prosumers are worse off compared with the no participation model.
3. The regulated aggregator earns positive profit.
4. A uniform (pricing and charging) policy for prosumers in the same location.

2.5.6 Numerical Example

We consider an example with $n = 3$ prosumers, one from each type. For ease of exposition, we consider a power system with one node and one generator with cost $c(y) = 0.1y^2 + 0.1y$. Prosumer i is equipped with an isoelastic utility function: $u_i(z) = \begin{cases} \frac{z^{1-\eta_i}-1}{1-\eta_i}, & \text{if } \eta_i \geq 0, \eta_i \neq 1, \\ \ln(z), & \text{if } \eta_i = 1. \end{cases}$

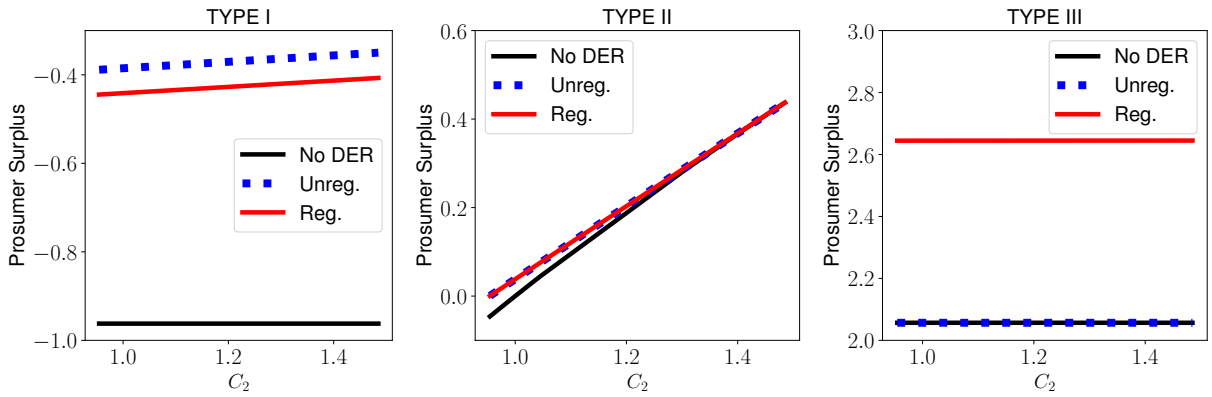

Figure 2.5: **Left:** TYPE I prosumer is always buying and achieves the maximum surplus for the unregulated model, but its surplus after regulation is still better than no participation. **Middle:** TYPE II prosumer achieves the maximum surplus for both the unregulated and regulated models. **Right:** TYPE III prosumer is always selling and achieves the maximum surplus for the regulated model; its surplus for the unregulated model is almost equal to the no participation model.

For our example, we distinguish among prosumer types by their capacities and risk-aversion parameters, and let $C_1 = 0.1,\ C_2 \in [0.95, 1.48],\ C_3 = 10,\ \eta_1 = 0.1,\ \eta_2 = 0.5,\ \eta_3 = 1.1$. With these parameters, prosumer $i = 1$ is TYPE I, prosumer $i = 2$ is TYPE II, and prosumer $i = 3$ is TYPE III. We also note that prosumer $i = 2$'s type changes for any $C_2 \notin [0.95, 1.48]$. Figure 2.5 illustrates the prosumer surplus for the no participation model, the unregulated model, and the regulated model, as C_2 varies from 0.95 to 1.48. The outcomes here are consistent with Table 2.1, and no prosumer is worse-off after DER aggregation, for both the regulated and the unregulated models. However, dependent on the prosumer type, some benefit further from regulation, and others do not. We pick three specific values of C_2. Starting with $C_2 = 0.5$, prosumer $i = 2$ is now TYPE I, and the aggregator's surpluses are given by $\mathsf{AS}(\text{No DER}) = 0,\ \mathsf{AS}(\text{Unreg.}) = 5.41,\ \mathsf{AS}(\text{Reg.}) = 5.53$, and the maximum surplus is attained for the regulated model. Next, when $C_2 = 1$, prosumer $i = 2$ is TYPE II, and the surpluses are given by $\mathsf{AS}(\text{No DER}) = 0,\ \mathsf{AS}(\text{Unreg.}) = 5.36,\ \mathsf{AS}(\text{Reg.}) = 5.41$, and the outcomes are similar. However, when $C_2 = 1.5$, prosumer $i = 2$ is TYPE III, and the surpluses are given by $\mathsf{AS}(\text{No DER}) = 0,\ \mathsf{AS}(\text{Unreg.}) = 5.32,\ \mathsf{AS}(\text{Reg.}) = 0.06$, and

hence, after regulation, because two prosumers are selling their DER supply, the aggregator only extracts the minimum P_i, so its surplus was reduced. Nevertheless, with the regulated model, the aggregator's surplus is always ensured to be positive. As a final remark, the regulated model might make the aggregator or a certain prosumer better-off or worse-off, compared to the unregulated model, but is always better than the no participation model, and it always attains an efficient market outcome.

2.6 Reducing Market Power of the Generators through Aggregation

In the previous sections, we proposed two efficient aggregation models: the unregulated aggregation model and the regulated aggregation model. There are two underlying assumptions in these models: all prosumers bid truthfully about their utility of consumption, and all generators bid truthfully about their cost of production. The former assumption is more justifiable as each prosumer represents a small fraction in the energy market; thus, it does not believe its bidding could make a difference in the market price. The later assumption, however, may not hold in general as the generators' bidding might affect the market price, and they may benefit from non-truthful (strategic) bidding. Such strategic bidding may result in higher market price (thus benefiting the generators) and reduced social welfare. The ability of the generators to influence the market price is referred to as the *market power* of the generators. In this section, we study the market power of the generators under the no participation model and the direct participation (or unregulated/regulated aggregation) model. We will show that, compared to no prosumer participation, the market power of the generators is reduced with prosumer participation (either direct or through aggregation), and the reduction in social welfare is also mitigated.

Suppose there are n prosumers, indexed by i, and N generators, indexed by j. For ease of exposition, in this section, we restrict the attention to one node (thus there is no network

constraints and we can drop the superscript k), and we also make the following simplifying assumptions on prosumers' utility of consumption and generators' cost functions.

Assumption 2.6. *For each prosumer i, the utility of consumption is quadratic and it is given by $u_i(z) = a_i z^2 + b_i z$, where $a_i < 0$ and $b_i \gg |a_i|$. So, the market price always falls into the range $(0, b_i)$. Each generator j's (true) cost function is linear in its production, and all generators have the same cost function, i.e., $c_j(y_j) = \alpha y_j$ for some $\alpha > 0$, and the optimal total supply always satisfies $y > 0$. Furthermore, there exists at least one prosumer i such that $2a_i C_i + b_i > \alpha$.*

As we will see in the rest of this section, Assumption 2.6 enables us to obtain explicit expressions for the generators' supply and the social welfare. In the following, we denote by $\mathcal{W}^T$ (resp., $\mathcal{W}^S$) the optimal social welfare of the model with prosumer participation (aggregation) when generators bid truthfully (resp., strategically). When no prosumer participation is allowed, the optimal social welfare under truthful bidding and strategic bidding of the generators are denoted by $\mathcal{W}^{TN}$ and $\mathcal{W}^{SN}$, respectively. Our main result in this section is summarized as the following theorem.

Theorem 2.3. *Under Assumption 2.6, the following inequalities hold:*

$$\mathcal{W}^T \geq \mathcal{W}^{TN}, \tag{2.35a}$$

$$\mathcal{W}^S \geq \mathcal{W}^{SN}, \tag{2.35b}$$

$$\mathcal{W}^T \geq \mathcal{W}^S, \tag{2.35c}$$

$$\mathcal{W}^{TN} \geq \mathcal{W}^{SN}, \tag{2.35d}$$

$$\mathcal{W}^{TN} - \mathcal{W}^{SN} \geq \mathcal{W}^T - \mathcal{W}^S. \tag{2.35e}$$

In Theorem 2.3, (2.35a) and (2.35b) state that the optimal social welfare with prosumer participation is always greater than that without prosumer participation, for both truthful

bidding and strategic bidding of the generators. Further, (2.35c) and (2.35d) state that the optimal social welfare under truthful bidding is always greater than that under strategic bidding of generators, for both the case with prosumer participation and the case without prosumer participation. Finally, (2.35e) implies that the loss of social welfare due to strategic bidding of the generators is reduced when there is prosumer participation, compared to the case when there is no prosumer participation.

For the rest of this section, we will provide a complete analysis of Theorem 2.3, and we also provide the proof in the appendix. We will first consider the case when there is prosumer participation (aggregation), and obtain the equilibria when generators bid truthfully and strategically. We will then move to the case when no prosumer participation is allowed, and obtain again the equilibria for both truthful bidding and strategic bidding of the generators. An illustrative example is provided towards the end of this section.

2.6.1 Full Prosumer Participation

We first consider the case when there is prosumer participation. As shown in the previous sections, given a market price, the prosumers' decisions under the unregulated aggregation model and the regulated aggregation model are the same as those under the direct participation model. The social welfare under these three models are also the same. Therefore, in this section, we will without loss of generality ignore the role of the aggregator and assume the direct participation of prosumers. We analyze the decisions of each party for the cases when generators bid truthfully and strategically.

2.6.1.1 Truthful bidding of the generators

Prosumers. Let λ^T be the market price when generators bid truthfully. Each prosumer i solves its payoff maximization problem:

$$\max_{C_i+d_i-x_i>0} u_i(C_i + d_i - x_i) - \lambda^T \cdot (d_i - x_i), \tag{2.36}$$

where we recall that d_i is the amount of purchase and x_i is the amount of energy sold. Let $z_i := C_i + d_i - x_i$ be the actual amount of energy consumed by prosumer i. Similar to Lemma 2.1, we can conclude that under Assumption 2.6, prosumer i's optimal response z_i^T satisfies

$$2a_i z_i^T + b_i = \lambda^T. \tag{2.37}$$

System operator. The system operator solves the economic dispatch problem to maximize the social welfare $\mathcal{W}^T$:

$$\begin{aligned} \mathcal{W}^T = \max_{z_i>0, y_j\geq 0} & \sum_i u_i(z_i) - \sum_j c_j(y_j) \\ \text{s.t.} \quad & \sum_j y_j = \sum_i (z_i - C_i). \end{aligned} \tag{2.38}$$

Since the generators are identical, we may without loss of generality restrict to $y_j = y_{j'}, \forall j \neq j'$, i.e., each generator supplies the same amount of energy y_j. Let $y := \sum_j y_j$ and $C := \sum_i C_i$. Consider the prosumers' and the system operator's problems, we have the following result.

Proposition 2.6. *Under Assumption 2.6, an optimal solution to* (2.38) *is given by*

$$z_i^T = \frac{\alpha - b_i}{2a_i}, \quad \forall i, \qquad y_j^T = \frac{-C + \sum_i \frac{\alpha - b_i}{2a_i}}{N}, \quad \forall j. \tag{2.39}$$

Furthermore, the equilibrium market price (the optimal Lagrange multiplier of the constraint) is given by $\lambda^T = \alpha$, and the optimal social welfare is then

$$\mathcal{W}^T = \alpha C - \sum_i \frac{(b_i - \alpha)^2}{4a_i}. \tag{2.40}$$

2.6.1.2 Strategic bidding of the generators

Prosumers. Let λ^S be the market price when generators bid strategically. Similar to (2.37), each prosumer i's optimal response z_i^S (under Assumption 2.6) now becomes

$$2a_i z_i^S + b_i = \lambda^S. \tag{2.41}$$

System operator. The system operator solves the economic dispatch problem to maximize the *apparent social welfare*, which is the "social welfare" when the system operator assumes the generators' bids are true, but is actually based on the (nontruthful) bidding $\tilde{c}_j$ of the generators:

$$\begin{aligned} \max_{z_i>0, y_j \geq 0} & \sum_i u_i(z_i) - \sum_j \tilde{c}_j(y_j) \\ \text{s.t.} \quad & \sum_j y_j = \sum_i (z_i - C_i). \end{aligned} \tag{2.42}$$

For each possible total supply y, we define the overall utility of consumption as

$$u(C+y) = \left\{ \max_{z_i>0} \sum_i u_i(z_i) \quad \text{s.t.} \quad \sum_i (z_i - C_i) = y \right\}. \tag{2.43}$$

In other words, $u(C+y)$ computes the total utility of consumption of all prosumers when the total energy supply from the generators is y. We can therefore rewrite the system operator's

problem equivalently as

$$\max_{y} u(C+y) - \sum_{j} \tilde{c}_j(y/N). \tag{2.44}$$

Generators. Each generator sets an optimal y_j to supply, and bids a tilted cost function $\tilde{c}_j$ instead of the true cost function c_j. The generator j considers the market price λ^S as a function of the total supply y, and aims to solve the profit maximization problem:

$$\max_{y_j} \lambda^S \left(y_j + \sum_{j' \neq j} y_{j'} \right) \cdot y_j - c_j(y_j). \tag{2.45}$$

To this end, the generator needs to compute the market price as a function of total supply. Since the total supply and the total (net) demand must be matched by the system operator, we have that

$$\sum_{i} (z_i^S - C_i) = y. \tag{2.46}$$

Combining (2.41) and (2.46), we obtain that

$$\lambda^S(y) = \frac{y + C + \sum_i \frac{b_i}{2a_i}}{\sum_i \frac{1}{2a_i}}. \tag{2.47}$$

Therefore, the generator's profit maximization problem (2.45) becomes:

$$\max_{y_j \geq 0} \frac{y_j + \sum_{j' \neq j} y_{j'} + C + \sum_i \frac{b_i}{2a_i}}{\sum_i \frac{1}{2a_i}} \cdot y_j - \alpha y_j. \tag{2.48}$$

Each generator solves (2.48), and since they are all identical, we only look at the equilibrium where $y_j = y_{j'}, \forall j, j' \in [N]$. Thus, from the first-order condition of (2.48), we obtain the

optimal supply amount for generator j:

$$y_j^S = \frac{-C + \sum_i \frac{\alpha - b_i}{2a_i}}{N+1}, \quad \forall j. \tag{2.49}$$

In order to sell y_j^S as given in (2.49), generator j will bid $\tilde{c}_j$ such that the system operator will allocate y_j^S amount of energy to j. The generator thus considers the economic dispatch problem that the system operator solves. Therefore, the generator will bid $\tilde{c}_j$ such that $y^S := \sum_j y_j^S = N y_j^S$ solves (2.44) optimally, i.e.,

$$\left.\frac{\partial u(C+y)}{\partial y}\right|_{y=Ny_j^S} = \left.\frac{\partial \tilde{c}_j(y_j)}{\partial y_j}\right|_{y_j=y_j^S}. \tag{2.50}$$

While there are many possible choices of $\tilde{c}_j$ that satisfies (2.50), one of the optimal bidding for the generator is to bid a linear cost function.

Lemma 2.5. *The following linear cost function is an optimal bidding strategy for generator j:*

$$\tilde{c}_j(y_j) = \frac{N\alpha \sum_i \frac{1}{2a_i} + C + \sum_i \frac{b_i}{2a_i}}{(N+1)\sum_i \frac{1}{2a_i}} \cdot y_j \tag{2.51}$$

Considering the prosumers' and the system operator's problems, we have the following result.

Proposition 2.7 (Competitive Equilibrium)**.** *Under Assumption 2.6, if all generators bid as in* (2.51)*, then, an optimal solution to* (2.42) *is given by*

$$z_i^S = \frac{\lambda^S - b_i}{2a_i}, \quad \forall i, \qquad y_j^S = \frac{-C + \sum_i \frac{\alpha - b_i}{2a_i}}{N+1}, \quad \forall j. \tag{2.52}$$

where the equilibrium market price (the optimal Lagrange multiplier of the equality constraint)

is

$$\lambda^S = \frac{N\alpha}{N+1} + \frac{C + \sum_i \frac{b_i}{2a_i}}{(N+1)\sum_i \frac{1}{2a_i}}. \tag{2.53}$$

Furthermore, the solutions in (2.52) *optimally solve the prosumer's/generator's problem. The social welfare in this case is*

$$\mathcal{W}^S = \left(\frac{\sum_i \left[\frac{\alpha N + b_i}{2a_i} + C_i\right]}{(N+1)\sum_i \frac{1}{2a_i}}\right)^2 \sum_i \frac{1}{4a_i} - \sum_i \frac{b_i^2}{4a_i} + \alpha \frac{N\sum_i \left(C_i - \frac{\alpha - b_i}{2a_i}\right)}{N+1}. \tag{2.54}$$

2.6.2 No Prosumer Participation

We next consider the case when the prosumers cannot sell back to the grid, i.e., each prosumer i can only purchase some $d_i \geq 0$ amount of energy. Since the amount of energy sold by each prosumer has to be $x_i = 0$, we will have $z_i = C_i + d_i$ for all prosumers. We first look at the prosumers' problem.

Prosumers. Let λ be the market price. Each prosumer i solves its payoff maximization problem:

$$\max_{d_i \geq 0} u_i(C_i + d_i) - \lambda d_i = \max_{d_i \geq 0} a_i(C_i + d_i)^2 + b_i(C_i + d_i) - \lambda d_i. \tag{2.55}$$

The optimal decision of prosumer i is thus

$$d_i = \left[\frac{\lambda - b_i}{2a_i} - C_i\right]^+. \tag{2.56}$$

Therefore, given any market price λ, the set of prosumers who make a strictly positive amount of purchase is

$$\mathcal{S}(\lambda) := \{i \mid 2a_i C_i + b_i > \lambda\}. \tag{2.57}$$

We will without loss of generality sort the prosumers in decreasing order of $2a_iC_i + b_i$, i.e., for any two prosumers i and i', we have that $2a_iC_i + b_i \geq 2a_{i'}C_{i'} + b_{i'}$ if $i \leq i'$. Under such ordering, for any $i \leq i'$, if prosumer i' makes positive purchase of electricity, then prosumer i must also make positive purchase. This sorting of prosumers will be helpful when we write the social welfare.

2.6.2.1 Truthful bidding of the generators

System operator. The system operator solves the economic dispatch problem that maximizes the social welfare:

$$\begin{aligned} \mathcal{W}^{TN} = \max_{d_i \geq 0, y_j \geq 0} & \sum_i u_i(C_i + d_i) - \sum_j c_j(y_j) \\ \text{s.t.} \quad & \sum_j y_j = \sum_i d_i. \end{aligned} \tag{2.58}$$

Since the generators are identical, we again restrict solutions to $y_j = y_{j'}, \forall j \neq j'$, i.e., each generator supplies the same amount of energy y_j. Considering the prosumers' and the system operator's problems, we have the following result.

Proposition 2.8. *Under Assumption 2.6, an optimal solution to* (2.38) *is given by*

$$d_i^{TN} = \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+, \quad \forall i, \qquad y_j^{TN} = \frac{\sum_i \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+}{N}, \quad \forall j. \tag{2.59}$$

Furthermore, the equilibrium market price (the optimal Lagrange multiplier of the constraint)

is given by $\lambda^{TN} = \alpha$. *The optimal social welfare is given by*

$$\mathcal{W}^{TN} = \sum_{\{i|2a_iC_i+b_i>\alpha\}} \left[\alpha C_i - \frac{(b_i-\alpha)^2}{4a_i}\right] + \sum_{\{i|2a_iC_i+b_i\leq\alpha\}} \left(a_iC_i^2 + b_iC_i\right). \tag{2.60}$$

2.6.2.2 Strategic bidding of the generators

System operator. The system operator solves the economic dispatch problem to maximize the *apparent social welfare*, which is the "social welfare" when the system operator assumes the generators' bids are true, but is actually based on the (nontruthful) bidding $\tilde{c}_j$ of the generators:

$$\begin{aligned} &\max_{d_i\geq 0, y_j\geq 0} \sum_i u_i(C_i+d_i) - \sum_j \tilde{c}_j(y_j) \\ &\text{s.t.} \quad \sum_j y_j = \sum_i d_i \end{aligned} \tag{2.61}$$

For each possible total supply y, the overall utility of consumption now becomes

$$u(C+y) = \left\{ \max_{d_i\geq 0} \sum_i u_i(C_i+d_i) \quad \text{s.t.} \quad \sum_i d_i = y \right\}. \tag{2.62}$$

As y increases, the number of prosumers with $d_i > 0$ will change in a discrete manner. However, we have the following useful result.

Lemma 2.6. $u(C+y)$ *is continuous and differentiable in* y.

Similar to (2.44), we can therefore rewrite the system operator's problem equivalently as

$$\max_{y\geq 0} u(C+y) - \sum_j \tilde{c}_j(y/N). \tag{2.63}$$

Generators. Each generator sets an optimal y_j to supply, and bids a tilted cost function $\tilde{c}_j$ instead of the true cost function c_j. The generator j considers the market price λ^{SN} as a

function of the total supply y, and aims to solve the profit maximization problem:

$$\max_{y_j} \lambda^{SN} \left(y_j + \sum_{j' \neq j} y_{j'} \right) \cdot y_j - c_j(y_j). \tag{2.64}$$

To this end, the generator needs to compute the market price as a function of total supply. Since the total supply and the total (net) demand must be matched by the system operator, we have that

$$\sum_i d_i^{SN} = y. \tag{2.65}$$

At any given y, we need to consider which prosumers have $d_i > 0$ and which prosumers have $d_i = 0$. Define

$$y^i := (2a_i C_i + b_i) \sum_{i'=1}^{i-1} \frac{1}{2a_{i'}} - \sum_{i'=1}^{i-1} \left(C_{i'} + \frac{b_{i'}}{2a_{i'}} \right), \quad \forall i = \{1, 2, \ldots, n\}. \tag{2.66}$$

Then, we have the following lemma.

Lemma 2.7. *Prosumer i's optimal decision $d_i > 0$ if and only if the total supply satisfies $y > y^i$.*

In other words, the set of prosumers with $d_i > 0$ may be written as

$$\mathcal{S}(y) = \left\{ i \mid y > y^i \right\}. \tag{2.67}$$

Combining (2.56) with $\lambda = \lambda^{SN}$, Lemma 2.7, and (2.65), we obtain that

$$\lambda^{SN}(y) = \frac{y + \sum_i \left(C_i + \frac{b_i}{2a_i} \right) \cdot \mathbb{1} \left\{ y > y^i \right\}}{\sum_i \frac{1}{2a_i} \cdot \mathbb{1} \left\{ y > y^i \right\}}. \tag{2.68}$$

Therefore, the generator's profit maximization problem (2.64) becomes:

$$\max_{y_j \geq 0} \frac{y_j + \sum_{j' \neq j} y_{j'} + \sum_i \left(C_i + \frac{b_i}{2a_i}\right) \cdot \mathbb{1}\left\{y_j + \sum_{j' \neq j} y_{j'} > y^i\right\}}{\sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{y_j + \sum_{j' \neq j} y_{j'} > y^i\right\}} \cdot y_j - \alpha y_j. \tag{2.69}$$

Each generator solves (2.69), and since they are all identical, we only look at the equilibrium where $y_j = y_{j'}, \forall j, j' \in [N]$. Thus, from the first-order condition of (2.69), we obtain the condition on optimal supply amount for generator j:

$$y_j^{SN} = \frac{\sum_i \left(\frac{\alpha - b_i}{2a_i} - C_i\right) \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\}}{N+1}, \quad \forall j. \tag{2.70}$$

In order to sell y_j^{SN} as given in (2.70), generator j will bid $\tilde{c}_j$ such that the system operator will allocate y_j^{SN} amount of energy to j. The generator thus considers the economic dispatch problem that the system operator solves. Therefore, the generator will bid $\tilde{c}_j$ such that $y^{SN} := \sum_j y_j^{SN} = N y_j^{SN}$ solves (2.63) optimally, i.e.,

$$\left.\frac{\partial u(C+y)}{\partial y}\right|_{y = N y_j^{SN}} = \left.\frac{\partial \tilde{c}_j(y_j)}{\partial y_j}\right|_{y_j = y_j^{SN}}. \tag{2.71}$$

While there are many possible choices of $\tilde{c}_j$ that satisfy (2.71), again one of the optimal bidding strategies for the generator is to bid a linear cost function.

Lemma 2.8. *The following linear cost function is an optimal bidding strategy for generator j:*

$$\tilde{c}_j(y_j) = \frac{N\alpha \sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\} + \sum_i \left(C_i + \frac{b_i}{2a_i}\right) \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\}}{(N+1) \sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\}} \cdot y_j \tag{2.72}$$

Considering the prosumers' and the system operator's problems, we have the following result.

Proposition 2.9 (Competitive Equilibrium)**.** *Under Assumption 2.6, if all generators bid*

as in (2.72)*, an optimal solution to* (2.61) *satisfies the following:*

$$d_i^{SN} = \left[-C_i + \frac{\lambda^{SN} - b_i}{2a_i}\right]^+, \quad \forall i, \qquad y_j^{SN} = \frac{\sum_i \left(-C_i + \frac{\alpha - b_i}{2a_i}\right) \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\}}{N+1}, \quad \forall j, \tag{2.73}$$

where the equilibrium market price (the optimal Lagrange multiplier of the constraint) is

$$\lambda^{SN} = \frac{N\alpha}{N+1} + \frac{\sum_i \left(C_i + \frac{b_i}{2a_i}\right) \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}}{(N+1)\sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}}. \tag{2.74}$$

Furthermore, (2.73) *optimally solves the prosumer's/generator's problem. The social welfare in this case is*

$$\begin{aligned}\mathcal{W}^{SN} = &\sum_i \left[\frac{1}{4a_i}\left(\frac{\sum_i \left[\frac{\alpha N + b_i}{2a_i} + C_i\right] \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}}{(N+1)\sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}}\right)^2 - \frac{b_i^2}{4a_i}\right] \cdot \mathbb{1}\left\{y^{SN} > y^i\right\} \\ &+ \sum_i \left(a_i C_i^2 + b_i C_i\right) \cdot \mathbb{1}\left\{y^{SN} \leq y^i\right\} + \alpha \frac{N\sum_i \left(C_i - \frac{\alpha - b_i}{2a_i}\right) \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}}{N+1}.\end{aligned} \tag{2.75}$$

2.6.3 Discussion

In the previous subsections, we have analyzed and derived the equilibrium social welfare for all four models, depending on if we allow prosumer participation (selling part of their capacity) and if the generators bid strategically. Table 2.2 summarizes the equilibrium social welfare under the four models.

	Full prosumer participation	No prosumer participation
Generators bidding truthfully	$\mathcal{W}^T$ (Proposition 2.6)	$\mathcal{W}^{TN}$ (Proposition 2.8)
Generators bidding strategically	$\mathcal{W}^S$ (Proposition 2.7)	$\mathcal{W}^{SN}$ (Proposition 2.9)

Table 2.2: Comparison of social welfare under different models

While we have obtained the valuable explicit expressions for the social welfare under all

four models, to complete the proof of Theorem 2.3, we need to compare the four expressions of $\mathcal{W}^T, \mathcal{W}^S, \mathcal{W}^{TN}$ and $\mathcal{W}^{SN}$, which is not obvious. In this subsection, we provide alternative expressions for these social welfares. To proceed, we first define the social welfare when there is no generator and no electricity market, i.e., each prosumer consumes the amount of energy that its capacity allows:

$$\mathcal{W}_0 := \sum_i u_i(C_i). \tag{2.76}$$

We also define the amount of market price rise due to strategic bidding of the generators as $\delta := \lambda^S - \lambda^T$, and $\delta^N := \lambda^{SN} - \lambda^{TN}$. With these definitions, we have the following expressions for the social welfare under different models, which use $\mathcal{W}_0$ as a reference.

Proposition 2.10. *We have the following expressions for the four social welfare.*

$$\mathcal{W}^T = \mathcal{W}_0 - \sum_i a_i \Big(z_i^T - C_i\Big)^2, \tag{2.77}$$

$$\mathcal{W}^S = \mathcal{W}_0 + \sum_i \left(-a_i \Big(z_i^S - C_i\Big)^2 + \delta \left(z_i^S - C_i\right)\right), \tag{2.78}$$

$$\mathcal{W}^{TN} = \mathcal{W}_0 - \sum_i a_i \left(d_i^{TN}\right)^2 = \mathcal{W}_0 - \sum_{\{i|z_i^T>0\}} a_i \left(z_i^T - C_i\right)^2, \tag{2.79}$$

$$\mathcal{W}^{SN} = \mathcal{W}_0 - \sum_i a_i \left(d_i^{SN}\right)^2 + \delta^N \sum_i d_i^{SN}. \tag{2.80}$$

We next obtain the expressions for δ and δ^N. Under full prosumer participation and strategic bidding of the generators, we have that $u'(z_i^S) = \lambda^S = \lambda^T + \delta = \alpha + \delta$. From the truthful bidding case, we also have that $\alpha = u_i'(z_i^T)$. Thus, we can write δ as

$$\delta = u'(z_i^S) - u'(z_i^T) = 2a_i(z_i^S - z_i^T), \quad \forall i, \tag{2.81}$$

or equivalently, by noting that $\sum_i (z_i^S - C_i) = y^S = y^T \cdot \frac{N}{N+1} = \frac{N}{N+1} \sum_i (z_i^T - C_i)$, we have

that

$$\delta = \frac{\sum_i \left(z_i^S - z_i^T\right)}{\sum_i \frac{1}{2a_i}} = \frac{-\left(\frac{1}{N+1}\right)\sum_i (z_i^T - C_i)}{\sum_i \frac{1}{2a_i}}. \tag{2.82}$$

Under the models that prosumers cannot sell, with generators bidding strategically, we have that $u_i'(C_i + d_i^{SN}) = \lambda^{SN} = \lambda^{TN} + \delta^N = \alpha + \delta^N$ for those prosumers with $d_i^{SN} > 0$. From the truthful bidding case, we also have that $\alpha = u_i'(C_i + d_i^{TN})$ for those prosumers with $d_i^{TN} > 0$. Thus, we can write δ^N as

$$\begin{aligned} \delta^N &= u_i'(C_i + d_i^{SN}) - u_i'(C_i + d_i^{TN}) = 2a_i(d_i^{SN} - d_i^{TN}) \\ &= 2a_i(d_i^{SN} - z_i^T), \quad \forall i \ \text{ s.t. } \ d_i^{SN} > 0. \end{aligned} \tag{2.83}$$

By noting that $\sum_{\{i|d_i^{SN}>0\}} d_i^{SN} = y^{SN} = \frac{N}{N+1}\sum_{\{i|d_i^{SN}>0\}}(z_i^T - C_i)$, we have that

$$\delta^N = \frac{\sum_{\{i|d_i^{SN}>0\}}(-d_i^{SN} + z_i^T - C_i)}{-\sum_{\{i|d_i^{SN}>0\}} \frac{1}{2a_i}} = \frac{\frac{1}{N+1}\sum_{\{i|d_i^{SN}>0\}}(z_i^T - C_i)}{-\sum_{\{i|d_i^{SN}>0\}} \frac{1}{2a_i}}. \tag{2.84}$$

With Proposition 2.10, (2.82), and (2.84), we are ready to show the relations in (2.35). Specifically, we have the following result.

Proposition 2.11 (Restating Theorem 2.3)**.** *The following relations hold:*

$$\mathcal{W}^T - \mathcal{W}^{TN} = -\sum_{\{i|z_i^T \le C_i\}} a_i \left(z_i^T - C_i\right)^2 \ge 0; \tag{2.85a}$$

$$\mathcal{W}^S - \mathcal{W}^{SN} = \sum_i \left(-a_i\left(z_i^S - C_i\right)^2 + \delta(z_i^S - C_i)\right) + \sum_i a_i \left(d_i^{SN}\right)^2 - \delta^N \sum_i d_i^{SN} \ge 0; \tag{2.85b}$$

$$\mathcal{W}^T - \mathcal{W}^S = \left(\frac{\sum_i (z_i^T - C_i)}{N+1}\right)^2 \cdot \frac{1}{\sum_i \frac{1}{-a_i}} \ge 0; \tag{2.85c}$$

$$\mathcal{W}^{TN} - \mathcal{W}^{SN} = \sum_{\{i|d_i^{SN}=0, z_i^T > C_i\}} (-a_i)(z_i^T - C_i)^2 + \left(\frac{\sum_{\{i|d_i^{SN}>0\}} (z_i^T - C_i)}{N+1} \right)^2 \cdot \frac{1}{\sum_{\{i|d_i^{SN}>0\}} \frac{1}{-a_i}} \geq 0; \tag{2.85d}$$

$$\mathcal{W}^{TN} - \mathcal{W}^{SN} \geq \mathcal{W}^T - \mathcal{W}^S. \tag{2.85e}$$

We have thus finished the proof of Theorem 2.3, and verified the benefit of aggregating distributed energy resources (DERs), i.e., by allowing the aggregation of DERs, the optimal social welfare is improved with either truthful bidding generators or strategic bidding generators, and the loss of social welfare due to strategic bidding of the generators is reduced in the full participation model compared to that in the no participation model. Thus, we can state that *DER aggregation mitigates market power of generators*. Next, we provide an illustrative example.

2.6.4 Illustrative Example

We consider an example with $n = 2$ and provide illustrations of our results. We let the true cost of each generator be $c(y) = \alpha y = 5y$. For each prosumer i, we let $a_i = -0.1$ and $b_i = 10$, and we distinguish between them via the capacities, with $C_1 = 10$ and $C_2 = 30$. The parameters are picked such that $2a_1C_1 + b_1 > \alpha$ and $2a_2C_2 + b_2 < \alpha$, i.e., prosumer $i = 1$ will always have a positive demand. To make our example more interesting, we vary the number of generators from $N = 1$ to $N = 20$ (outcomes saturate for $N > 20$). In Figure 2.6, we plot the social welfare for each market setup, efficiency loss due to strategic behavior of the generators, and prices. The key outcomes can be summarized as follows:

1. $\mathcal{W}^{SN} \to \mathcal{W}^{TN}$ and $\mathcal{W}^S \to \mathcal{W}^T$ as $N \to \infty$.

2. $\lambda^{SN} \to \alpha$ and $\lambda^S \to \alpha$ as $N \to \infty$.

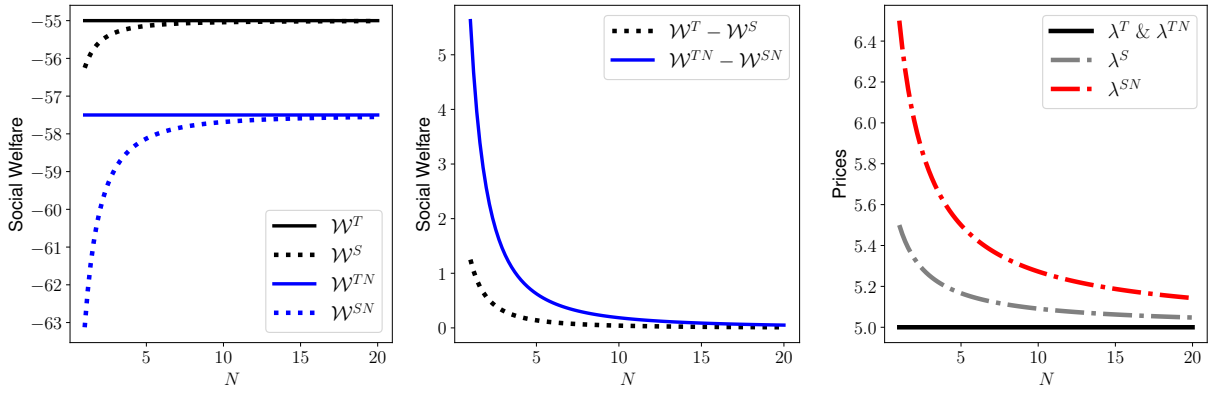

Figure 2.6: **Left:** Social welfare for each market setup vs. N. DER participation improves the social welfare. As N increases, strategic bidding converges to truthful bidding, and all inequalities provided in Theorem 2.3 hold. **Middle:** Quantification of market power in terms of social welfare loss. When DERs are integrated, market power is mitigated. **Right:** Price for each market setup vs. N. Highest price corresponds to strategic bidding without DER participation, but when DERs are present, the price becomes lower. All prices converge to $\alpha = 5$.

3. For $N < \infty$, $\alpha < \lambda^{SN} < \lambda^{S}$.

4. Efficient DER aggregation mitigates the market power of generators.

2.7 Conclusions

In this work, we have addressed questions surrounding the debate on whether or not it is possible to achieve full market efficiency in the presence of a monopolistic DER aggregator by studying three aggregation models. For all models, we utilized two-part pricing, where each prosumer pays a connection charge and sells its DER supply at a per-unit price offered by the aggregator. The first model has an unregulated profit-seeking aggregator, who adopts a discriminatory pricing policy in which each prosumer has a specific connection charge, and achieves full market efficiency. The second model has a profit-seeking aggregator who must not do price discrimination to prosumers in the same location, and we show numerically that there is a mild efficiency loss. In the third model, we avoid discriminatory pricing by

regulating the aggregator while ensuring that its profit is always positive. Both efficient models were shown to mitigate the market power of conventional generators; the welfare gap between truthful and strategic bidding was reduced.

There are several directions for future work.

1. We focused on DER capacities that have been already installed, and it would be interesting to see how efficient aggregation models would impact investments and rebates for the installations of such capacities (Kök et al. [130], Hu et al. [109], Aflaki and Netessine [6], Babich et al. [25]).

2. In our setup, the aggregator is monopolistic, and two-part pricing was necessary to attain full market efficiency, but if one-part pricing was adopted, it would not be possible to attain full market efficiency (Alshehri et al. [10, 11]). So, another research direction is to study the affect of competition among aggregators under different pricing policies, and perhaps, also study their strategic bidding in networked wholesale markets (Bimpikis et al. [38], Anupindi and Jiang [17], Nguyen and Kannan [167], Ruhi et al. [184]) or in zonal electricity markets (Aravena et al. [20]).

3. To achieve the full efficiency results, we need the true information about prosumers' utility functions. In practice, this could be observed from their bids/behavior via iterative interactions with the aggregator/utility or by empirical studies (see Khezeli et al. [129], Ata et al. [22]). If the utility parameters are mis-specified, then, the aggregator would set the wrong prices to the prosumers, and there will be efficiency loss due to that. It would be interesting to explore the consequences of not having a good prosumer utility parameter estimation.

4. While we considered a single period problem, there are multiple ways to extend our work to a multi-period setup. One possibility is to consider an open-loop game where all decisions are made at the beginning where each period t has a price pair (P_t, p_t).

In the absence of temporal constraints (such as ramping), the analysis for each period could be decoupled, and our efficiency results will still hold. In the presence of temporal constraints, the analysis becomes more complicated. While this still could be analyzed as an open-loop game, one might need to use tools from dynamic game theory, which might be difficult. Moreover, it is worth noting that if the aggregator has to offer the same prices for all periods, then the aggregator would need to find the optimal single (P, p) that is applied to all periods (then it will not be fully efficient), which could be explored in a future work.

5. We remark that while in our analysis we have assumed deterministic DER capacities, they are in fact often intermittent and uncertain. Some of this variability might be hedged against via the forward (often, day-ahead) markets. In view of recent developments regarding DER integration, stochastic models are more realistic and have been adopted by many authors (Alshehri et al. [11], Secomandi and Kekre [189], Alshehri et al. [12], Sunar and Birge [203], Han et al. [100], Wu and Kapuscinski [223], Zhou et al. [235], Alessio Trivella et al. [8], Peura and Bunn [174]). It would be interesting to study whether our results generalize to stochastic DER capacities.

2.8 Appendix

Proof of Proposition 2.1. First note that (2.7) can be equivalently written as:

$$\begin{aligned} \max \quad & \mathcal{W}_B(\mathbf{x}-\mathbf{d}, \mathbf{y}) \\ \text{subject to} \quad & \mathbf{1}^\top (\mathbf{X}-\mathbf{D}+\mathbf{Y}) = \mathbf{0},\ -\mathbf{B}(\mathbf{X}-\mathbf{D}+\mathbf{Y}) \leq \mathbf{f}, \\ & \mathbf{C}-\mathbf{Z} \leq \mathbf{x}-\mathbf{d} \leq \mathbf{C},\ \underline{\mathbf{y}} \leq \mathbf{y} \leq \overline{\mathbf{y}}. \end{aligned} \tag{2.86}$$

By Assumption 2.1 and Assumption 2.2, each u_i^k is strictly concave, and each c_j^k is strictly convex. Thus, $\mathcal{W}_B$ given by (2.6) is strictly concave. It can be seen from (2.86) that

the feasible set is compact, and is nonempty by Assumption 2.3. Therefore, there exists a unique optimal solution $((\mathbf{x}-\mathbf{d})^*, \mathbf{y}^*)$ to (2.86). By letting $\mathbf{h}^* = -\mathbf{Y}^* - (\mathbf{X}-\mathbf{D})^*$, we have a unique optimal solution $(\mathbf{h}^*, (\mathbf{x}-\mathbf{d})^*, \mathbf{y}^*)$ to (2.7).

We write the Lagrangian of (2.7) as

$$\begin{aligned}\mathcal{L} = \sum_{k\in[n]} \Bigg(& \sum_{i\in[n_k]} u_i^k \left(C_i^k - (x_i^k - d_i^k)\right) - \sum_{j\in[N_k]} c_j^k(y_j^k) + \lambda^k \left(h^k + X^k - D^k + Y^k\right) \\ & + \sum_{i\in[n_k]} \left(\overline{\mu}_i^k \left(C_i^k - (x_i^k - d_i^k)\right) + \underline{\mu}_i^k \left(x_i^k - d_i^k - C_i^k + Z\right)\right) \\ & + \sum_{j\in[N_k]} \left(\overline{\nu}_j^k (\overline{y}_j^k - y_j^k) + \underline{\nu}_j^k (y_j^k - \underline{y}_j^k)\right)\Bigg) \\ & + \text{extra terms}, \end{aligned} \tag{2.87}$$

where $\lambda^k, \overline{\mu}_i^k, \underline{\mu}_i^k, \overline{\nu}_j^k, \underline{\nu}_j^k$ are the Lagrange multipliers corresponding to the first constraint in (2.3) and (2.4), and the "extra terms" correspond to the other constraints in (2.3). Under Assumptions 2.1 and 2.2, it should be clear that the optimal solution $(x_i^k - d_i^k)^* \neq C_i^k$, $(x_i^k - d_i^k)^* \neq C_i^k - Z$, and $y_j^{k*} \neq \overline{y}_j^k$, $y_j^{k*} \neq \underline{y}_j^k$. Then, from the KKT optimality conditions, we have that $\overline{\mu}_i^k = \underline{\mu}_i^k = \overline{\nu}_j^k = \underline{\nu}_j^k = 0$. Further, we have that

$$\nabla_{(x_i^k - d_i^k)} \mathcal{L} = \frac{\partial u_i^k}{\partial (x_i^k - d_i^k)}\Big|_{(x_i^k - d_i^k)^*} + \lambda^k = 0, \tag{2.88a}$$

$$\nabla_{y_j^k} \mathcal{L} = -\frac{\partial c_j^k}{\partial y_j^k}\Big|_{y_j^{k*}} + \lambda^k = 0, \tag{2.88b}$$

where the first equality is equivalent to Lemma 2.1 and the second equality is equivalent to Lemma 2.2. □

Proof of Lemma 2.3. Consider prosumer i at location k. By Assumption 2.1, $\frac{\partial u_i^k(z)}{\partial z}$ is continuous and ranges from ∞ to 0. By Intermediate Value Theorem, for any given λ^k and p_i^k, there exist a (z_λ, z_p) such that (2.9) holds. By strict concavity of u_i^k, $\frac{\partial u_i^k(z)}{\partial z}$ is strictly

decreasing, so the (z_λ, z_p) is unique.

We next find the optimal solution to (2.8). We look at the two cases when $x_i^k > 0$ and $x_i^k = 0$ separately. Recall that $p_i^k \leq \lambda^k$. Then for any $x_i^k > 0$ and $d_i^k \geq x_i^k$, we have that

$$\pi_i^k \left(x_i^k, d_i^k\right) \leq \pi_i^k \left(0, d_i^k - x_i^k\right).$$

For any $x_i^k > 0$ and $0 < d_i^k < x_i^k$, we have that

$$\pi_i^k \left(x_i^k, d_i^k\right) \leq \pi_i^k \left(x_i^k - d_i^k, 0\right).$$

Therefore, we may without loss of generality restrict to the case when x_i^k and d_i^k are not both strictly positive. We can rewrite (2.8) as

$$\pi_i^k(x_i^k, d_i^k) = \begin{cases} p_i^k x_i^k - P_i^k + u_i^k \left(C_i^k - x_i^k\right), & \text{if } x_i^k > 0, \\ u_i^k \left(d_i^k + C_i^k\right) - \lambda^k d_i^k, & \text{if } x_i^k = 0. \end{cases} \tag{2.89}$$

- If $C_i^k \leq z_p$, then $u_i^k \left(C_i^k - x_i^k\right) + p_i^k x_i^k \leq u_i^k \left(C_i^k\right)$ for any $x_i^k > 0$. Thus we have the optimal $x_i^{k*} = 0$. If we further have that $C_i^k < z_\lambda$, then the first order condition of $u_i^k \left(d_i^k + C_i^k\right) - \lambda^k d_i^k$ leads to $d_i^{k*} = z_\lambda - C_i^k$. If we instead have $C_i^k \geq z_\lambda$, then $u_i^k \left(d_i^k + C_i^k\right) - \lambda^k d_i^k \leq u_i^k \left(C_i^k\right)$ for any $d_i^k > 0$.
- If $C_i^k > z_p$, then $u_i^k \left(d_i^k + C_i^k\right) - \lambda^k d_i^k < u_i^k \left(C_i^k\right)$ for any $d_i^k > 0$. We thus have $d_i^{k*} = 0$. When $x_i^k > 0$, the first order condition of $p_i^k x_i^k - P_i^k + u_i^k \left(C_j^i - x_i^k\right)$ leads to $x_i^k = C_i^k - z_p$. Moreover, it is only optimal to have $x_i^{k*} = C_i^k - z_p > 0$ if $p_i^k x_i^k - P_i^k + u_i^k \left(C_j^i - x_i^k\right) \geq u_i^k \left(C_i^k\right)$, or equivalently, $P_i^k \leq p_i^k \left(C_i^k - z_p\right) + u_i^k \left(z_p\right) - u_i^k \left(C_i^k\right)$. Therefore, we have that $x_i^{k*} = \left(C_i^k - z_p\right) \cdot \mathbb{1}\left\{\mathcal{X}\right\}$, where $\mathcal{X} = \left\{P_i^k \leq p_i^k \left(C_i^k - z_p\right) + u_i^k \left(z_p\right) - u_i^k \left(C_i^k\right)\right\}$.

□

Proof of Lemma 2.4. Consider prosumer i at location k. Let (z_λ, z_p) be as in (2.9). Then $z_\lambda \leq z_p$ since $p_i^k \leq \lambda^k$. First note that when $C_i^k \leq z_\lambda \leq z_p$, $x_i^{k*} = 0$ by Lemma 2.3, and $\Pi_i^k\left(P_i^k, p_i^k\right) = 0$ for any $(P_i^k, p_i^k) \in \mathbb{R}_+^2$, i.e., the aggregator $\mathcal{A}$ earns zero profit from the prosumer regardless of its pricing decisions.

For those prosumers with $C_i^k > z_\lambda$, we have that $d_i^{k*} = 0$ from Lemma 2.3. Since the aggregator earns zero profit if the prosumer does not sell its energy, the aggregator would choose a p_i^k such that $C_i^k > z_p$ (because otherwise $x_i^{k*} = 0$) and thus $x_i^{k*} = C_i^k - z_p$. We can add in the aggregator's problem the constraint that the prosumers would benefit from selling:

$$\begin{aligned} &\max_{P_i^k, p_i^k \geq 0} \; P_i^k + \left(\lambda^k - p_i^k\right) x_i^{k*} \\ &\text{s.t. } p_i^k x_i^{k*} + u_i^k\left(C_i^k - x_i^{k*}\right) - P_i^k \geq u_i^k\left(C_i^k\right), \end{aligned} \tag{2.90}$$

where x_i^{k*} is the optimal response of the prosumer, as a function of (P_i^k, p_i^k). We observe from (2.90) that, the optimal P_i^k should satisfy $P_i^k = p_i^k x_i^{k*} + u_i^k\left(C_i^k - x_i^{k*}\right) - u_i^k\left(C_i^k\right)$ in the maximization problem. Thus, (2.90) can be rewritten as

$$\begin{aligned} &\max_{p_i^k \geq 0} p_i^k x_i^{k*} + u_i^k\left(C_i^k - x_i^{k*}\right) - u_i^k\left(C_i^k\right) + \left(\lambda^k - p_i^k\right) x_i^{k*} \\ =&\max_{z_p \geq 0} u_i^k(z_p) - u_i^k\left(C_i^k\right) + \lambda^k\left(C_i^k - z_p\right), \end{aligned} \tag{2.91}$$

and the first order condition of (2.91) leads to

$$\left.\frac{\partial u_i^k(z_p)}{\partial z_p}\right|_{z_p = z_p^*} = \lambda^k. \tag{2.92}$$

By (2.9), we also have that $\left.\frac{\partial u_i^k(z)}{\partial z}\right|_{z=z_p^*} = p_i^{k*}$. We thus conclude that $p_i^{k*} = \lambda^k$. This further leads to $z_p^* = z_\lambda$, and thus

$$P_i^{k*} = p_i^{k*}\left(C_i^k - z_p^*\right) + u_i^k\left(z_p^*\right) - u_i^k\left(C_i^k\right) = \lambda^k\left(C_i^k - z_\lambda\right) + u_i^k\left(C_i^k - z_\lambda\right) - u_i^k\left(C_i^k\right).$$

□

Proof of Proposition 2.3. In this proof, we show that the optimal solution to (2.7) can be used to construct an optimal solution to (2.15).

First note that the objectives of both problems have the same expression, i.e.,

$$\mathcal{W}_A(\mathbf{h},\mathbf{x},\mathbf{d},\mathbf{y}) = \mathcal{W}_B(\mathbf{h},\mathbf{x}-\mathbf{d},\mathbf{y}) = \sum_{k\in[n]}\left(\sum_{i\in[n_k]} u_i^k(d_i^k + C_j^i - x_i^k) - \sum_{j\in[N_i]} c_j^i(y_j^i)\right).$$

While we took $(\mathbf{x}-\mathbf{d})$ as a single vector of decision variables in solving (2.7) and obtained a unique optimal solution $(\mathbf{h}^*,(\mathbf{x}-\mathbf{d})^*,\mathbf{y}^*)$ in Proposition 2.1, we can equivalently consider $\mathbf{x}$ and $\mathbf{d}$ as two separate vectors of decision variables. Consider the constraints (2.4) and constraints (2.13). For any $\mathbf{x},\mathbf{d},\mathbf{y}$ satisfying constraints (2.13), we have that $\mathbf{x}-\mathbf{d}\le\mathbf{C}$ and $\mathbf{x}-\mathbf{d}\ge\mathbf{C}-\mathbf{Z}$, which implies constraints (2.4). With the other constraints being the same, the feasible region of (2.15) is a subset of the feasible region of (2.7), which implies that $\mathcal{W}_B^*\ge\mathcal{W}_A^*$.

Let $(\mathbf{h}^*,(\mathbf{x}-\mathbf{d})^*,\mathbf{y}^*)$ be the optimal solution to (2.7). Let $\mathbf{x}^*=[(\mathbf{x}-\mathbf{d})^*]^+$ and $\mathbf{d}^*=[-(\mathbf{x}-\mathbf{d})^*]^+$, then $\mathbf{x}^*-\mathbf{d}^*=(\mathbf{x}-\mathbf{d})^*$. We show that $(\mathbf{h}^*,\mathbf{x}^*,\mathbf{d}^*,\mathbf{y}^*)$ is feasible to (2.15). It suffices to show that $\mathbf{0}\le\mathbf{x}^*\le\mathbf{C}$ and $\mathbf{0}\le\mathbf{d}^*\le\mathbf{Z}-\mathbf{C}+\mathbf{x}^*$. By definition, $\mathbf{x}^*\ge\mathbf{0}$ and $\mathbf{d}^*\ge\mathbf{0}$. If $(\mathbf{x}-\mathbf{d})^*<\mathbf{0}$, then $\mathbf{x}^*=\mathbf{0}$ and $\mathbf{d}^*=-(\mathbf{x}-\mathbf{d})^*\le\mathbf{Z}-\mathbf{C}=\mathbf{Z}-\mathbf{C}+\mathbf{x}^*$. If $(\mathbf{x}-\mathbf{d})^*\ge\mathbf{0}$, then $\mathbf{x}^*=(\mathbf{x}-\mathbf{d})^*\le\mathbf{C}$ and $\mathbf{d}^*=\mathbf{0}$. In either case, we have that $\mathbf{x}^*,\mathbf{d}^*$ are feasible to (2.13), and thus $(\mathbf{h}^*,\mathbf{x}^*,\mathbf{d}^*,\mathbf{y}^*)$ is feasible to (2.15). Therefore, this set of $(\mathbf{h}^*,\mathbf{x}^*,\mathbf{d}^*,\mathbf{y}^*)$ is an optimal solution to (2.15), and thus $\mathcal{W}_A^*=\mathcal{W}_B^*$. Further, the corresponding optimal Lagrange multipliers of constraints (2.3) are the same in problem (2.7) and problem (2.15), i.e., the wholesale market prices $\boldsymbol{\lambda}$ are the same under both models.

From Proposition 2.1, we know that y_j^{k*} and λ^k are consistent with Lemma 2.2, $\forall k\in[n],\forall j\in[N_k]$. Moreover, $\left(x_i^k-d_i^k\right)^*$ and λ^k are consistent with Lemma 2.1, $\forall k\in[n],\forall i\in[n_k]$. Consider an arbitrary prosumer i at an arbitrary location k. With $p_i^k=\lambda^k$, (z_λ,z_p)

as defined in (2.9), and $P_i^k = \left[\lambda^k(C_i^k - z_\lambda) + u_i^k(z_\lambda) - u_i^k(C_i^k)\right]^+$, we have that $z_\lambda = z_p = C_i^k - (x_i^k - d_i^k)^*$. If $C_i^k \leq z_p = C_i^k - (x_i^k - d_i^k)^*$, then $(x_i^k - d_i^k)^* < 0$, which implies that $d_i^{k*} = \left[-(x_i^k - d_i^k)^*\right]^+ = \left[z_\lambda - C_i^k\right]^+$ and $x_i^{k*} = \left[(x_i^k - d_i^k)^*\right]^+ = 0$. If $C_i^k > z_p = C_i^k - (x_i^k - d_i^k)^*$, then $(x_i^k - d_i^k)^* > 0$, which implies that $d_i^{k*} = \left[-(x_i^k - d_i^k)^*\right]^+ = 0$ and $x_i^{k*} = \left[(x_i^k - d_i^k)^*\right]^+ = C_i^k - z_p$. In either case, the optimal response x_i^{k*} and d_i^{k*} matches those described in Lemma 2.3. Further, by Lemma 2.4, the choice of $p_i^k = \lambda^k$ and $P_i^k = \left[\lambda^k(C_i^k - z_\lambda) + u_i^k(z_\lambda) - u_i^k(C_i^k)\right]^+$ are optimal for the aggregator to maximize its profit from the prosumer j. We have thus verified all statements listed in Proposition 2.3. □

Proof of Theorem 2.1. Immediately follows from the proof of Proposition 2.3. □

Proof of Proposition 2.5. Since $(\mathbf{h}^*, \mathbf{x}^*, \mathbf{d}^*, \mathbf{y}^*)$ and $\boldsymbol{\lambda}$ are the same as those in Proposition 2.3, it follows from Proposition 2.3 that $\mathbf{y}^*$ is consistent with Lemma 2.2. We next show that $(\mathbf{x}^*, \mathbf{d}^*)$ are optimal solutions to (2.27), given $(\mathbf{P}, \mathbf{p})$ as in (2.26), $\mathbf{F}$ as in (2.25), and $\boldsymbol{\lambda}$. Note that the marginal price is again $p^k = \lambda^{k*}$. Consider prosumer i at location k. Since (x_i^{k*}, d_i^{k*}) is optimal for (2.8), we have that

$$(x_i^{k*}, d_i^{k*}) = \underset{x_i^k \in [0, C_i^k],\ d_i^k \in [0, Z - C_i^k + x_i^k]}{\arg\max} \begin{cases} p_i^k x_i^k - P_i^k + u_i^k\left(d_i^k + C_i^k - x_i^k\right) - \lambda^k d_i^k, & \\ & \text{if } x_i^k > 0, \\ u_i^k\left(d_i^k + C_i^k\right) - \lambda^k d_i^k, & \text{if } x_i^k = 0. \end{cases}$$

Offsetting both cases by the same constant F_i^k will not change the optimality of (x_i^{k*}, d_i^{k*}),

i.e.,

$$(x_i^{k*}, d_i^{k*}) = \underset{x_i^k \in [0, C_i^k],\ d_i^k \in [0, Z - C_i^k + x_i^k]}{\arg\max} \begin{cases} p_i^k x_i^k - P_i^k + u_i^k \left(d_i^k + C_i^k - x_i^k \right) - \lambda^k d_i^k - F_i^k, & \\ & \text{if } x_i^k > 0, \\ u_i^k \left(d_i^k + C_i^k \right) - \lambda^k d_i^k - F_i^k, & \text{if } x_i^k = 0. \end{cases} \tag{2.93}$$

There are two cases:

- $x_i^{k*} > 0$ and $d_i^{k*} = 0$. This implies that the optimal decision is to sell x_i^{k*}, and earn the upper profit of (2.93): $p_i^k x_i^{k*} - P_i^k + u_i^k \left(d_i^k + C_i^k - x_i^{k*} \right) - \lambda^k d_i^k - F_i^k$. Since $P^k \leq P_i^k$, by changing P_i^k to P^k, we are adding a nonnegative constant to the objective, and thus does not change the optimal decision. Therefore, we have that

  $$(x_i^{k*}, d_i^{k*})$$
  $$= \underset{x_i^k \in [0, C_i^k],\ d_i^k \in [0, Z - C_i^k + x_i^k]}{\arg\max} \begin{cases} p_i^k x_i^k - P^k + u_i^k \left(d_i^k + C_i^k - x_i^k \right) - \lambda^k d_i^k - F_i^k, & \\ & \text{if } x_i^k > 0, \\ u_i^k \left(d_i^k + C_i^k \right) - \lambda^k d_i^k - F_i^k, & \text{if } x_i^k = 0, \end{cases}$$

  which says that (x_i^{k*}, d_i^{k*}) is also an optimal solution to (2.27).

- $x_i^{k*} = 0$ and $d_i^{k*} \geq 0$. This implies that the optimal decision is to buy d_i^{k*} and earn the lower profit of (2.93). For these prosumers, by Proposition 2.2 and Proposition 2.3, we know that $u_i^k \left(d_i^{k*} + C_i^k \right) - \lambda^k d_i^{k*} - F_i^k \geq \max_{x_i^k > 0, d_i^k = 0} p_i^k x_i^k - P_i^k + u_i^k \left(d_i^k + C_i^k - x_i^k \right) - \lambda^k d_i^k - F_j^i$ even with $P_i^k = 0$. Therefore, we again have that

  $$(x_i^{k*}, d_i^{k*})$$

$$= \underset{x_i^k \in [0, C_i^k],\ d_i^k \in [0, Z - C_i^k + x_i^k]}{\arg\max} \begin{cases} p_i^k x_i^k - P^k + u_i^k \left(d_i^k + C_i^k - x_i^k \right) - \lambda^k d_i^k - F_i^k, & \\ & \text{if } x_i^k > 0, \\ u_i^k \left(d_i^k + C_i^k \right) - \lambda^k d_i^k - F_i^k, & \text{if } x_i^k = 0, \end{cases}$$

which also says that (x_i^{k*}, d_i^{k*}) is also an optimal solution to (2.27).

□

Proof of Proposition 2.6. We write the Lagrangian of (2.38) as

$$\mathcal{L} = \sum_i u_i(z_i) - \sum_j c_j(y_j) + \lambda \left(\sum_j y_j - \sum_i (z_i - C_i) \right) + \mu_i z_i + \nu_j y_j, \tag{2.94}$$

where λ, μ_i, ν_j are the Lagrange multipliers of the constraints. The KKT optimality conditions are

$$\frac{\partial \mathcal{L}}{\partial z_i} = 2a_i z_i + b_i - \lambda + \mu_i = 0, \quad \forall i, \tag{2.95a}$$

$$\frac{\partial \mathcal{L}}{\partial y_j} = -\alpha + \lambda + \nu_j = 0, \quad \forall j, \tag{2.95b}$$

$$\lambda \left(\sum_j y_j - \sum_i (z_i - C_i) \right) = 0, \tag{2.95c}$$

$$\mu_i z_i = 0, \quad \forall i, \tag{2.95d}$$

$$\nu_j y_j = 0, \quad \forall j, \tag{2.95e}$$

$$\sum_j y_j - \sum_i (z_i - C_i) = 0, \tag{2.95f}$$

$$z_i > 0, \quad \forall i, \tag{2.95g}$$

$$y_j \geq 0, \quad \forall j, \tag{2.95h}$$

$$\lambda, \mu_i, \nu_j \geq 0, \quad \forall i, j. \tag{2.95i}$$

By Assumption 2.6, we have that $y_j > 0$ and thus $\nu_j = 0$ for all $j \in [N]$. With some algebra, we can conclude from (2.95) that

$$\lambda^T = \alpha, \quad z_i^T = \frac{\alpha - b_i}{2a_i}, \quad \sum_j y_j^T = \sum_i (z_i^T - C_i) = -C + \sum_i \frac{\alpha - b_i}{2a_i}.$$

We note that the optimal z_i^T we derived from the above also satisfies (2.37). The optimal social welfare is given by

$$\mathcal{W}^T = \sum_i u_i(z_i^T) - \sum_j c_j(y_j^T) = \alpha C - \sum_i \frac{(b_i - \alpha)^2}{4a_i}.$$

This completes the proof of Proposition 2.6. □

Proof of Lemma 2.5. Recall from (2.43) that

$$u(C + y) = \left\{ \max_{z_i > 0} \sum_i u_i(z_i) \quad \text{s.t.} \quad \sum_i (z_i - C_i) = y \right\},$$

which is itself an optimization problem. We write its Lagrangian:

$$\mathcal{L} = \sum_i u_i(z_i) + \lambda \left(y - \sum_i (z_i - C_i) \right) + \mu_i z_i,$$

where λ and μ_i are the Lagrange multipliers of the constraints. The KKT optimality conditions are

$$\frac{\partial \mathcal{L}}{\partial z_i} = 2a_i(z_i) + b_i - \lambda + \mu_i = 0, \quad \forall i \tag{2.96a}$$

$$\lambda \left(y - \sum_i (z_i - C_i) \right) = 0 \tag{2.96b}$$

$$\mu_i z_i = 0, \quad \forall i \tag{2.96c}$$

$$\sum_i (z_i - C_i) = y \tag{2.96d}$$

$$z_i > 0, \quad \forall i \tag{2.96e}$$

$$\lambda, \mu_i \geq 0, \quad \forall i. \tag{2.96f}$$

Thus, the optimal $z_i = \frac{\lambda - b_i}{2a_i}$, and thus

$$\sum_i (z_i - C_i) = \sum_i \frac{\lambda - b_i}{2a_i} - \sum_i C_i = y,$$

which implies that

$$\lambda = \frac{y + \sum_i C_i + \sum_i \frac{b_i}{2a_i}}{\sum_i \frac{1}{2a_i}}.$$

Therefore, we have that

$$\begin{aligned} u(C+y) &= \sum_i a_i \left(\frac{\lambda - b_i}{2a_i} \right)^2 + \sum_i b_i \left(\frac{\lambda - b_i}{2a_i} \right) \\ &= \sum_i \frac{1}{4a_i} \left(\frac{y + C + \sum_i \frac{b_i}{2a_i}}{\sum_i \frac{1}{2a_i}} - b_i \right)^2 + \sum_i \frac{b_i}{2a_i} \left(\frac{y + C + \sum_i \frac{b_i}{2a_i}}{\sum_i \frac{1}{2a_i}} - b_i \right). \end{aligned} \tag{2.97}$$

From (2.49) and (2.97), it follows that

$$\left. \frac{\partial u(C+y)}{\partial y} \right|_{y = N y_j^S} = \frac{\frac{-C + \sum_i \frac{\alpha - b_i}{2a_i}}{N+1} \cdot N + C + \sum_i \frac{b_i}{2a_i}}{\sum_i \frac{1}{2a_i}} = \frac{N\alpha \sum_i \frac{1}{2a_i} + C + \sum_i \frac{b_i}{2a_i}}{(N+1) \sum_i \frac{1}{2a_i}}.$$

Therefore, bidding (2.51) ensures that the condition (2.50) is satisfied, which ensures that the system operator will optimally assign y_j^S to generator j. □

Proof of Proposition 2.7. We write the Lagrangian of (2.42) as

$$\mathcal{L} = \sum_i u_i(z_i) - \sum_j \tilde{c}_j(y_j) + \lambda \left(\sum_j y_j - \sum_i (z_i - C_i) \right) + \mu_i z_i + \nu_j y_j, \tag{2.98}$$

where λ, μ_i, ν_j are the Lagrange multipliers of the constraints, and $\tilde{c}_j$ is given in (2.51). The KKT optimality conditions are

$$\frac{\partial \mathcal{L}}{\partial z_i} = 2a_i z_i + b_i - \lambda + \mu_i = 0, \quad \forall i, \tag{2.99a}$$

$$\frac{\partial \mathcal{L}}{\partial y_j} = \frac{N\alpha \sum_i \frac{1}{2a_i} + C + \sum_i \frac{b_i}{2a_i}}{(N+1)\sum_i \frac{1}{2a_i}} + \lambda + \nu_j = 0, \quad \forall j, \tag{2.99b}$$

$$\lambda \left(\sum_j y_j - \sum_i (z_i - C_i) \right) = 0, \tag{2.99c}$$

$$\mu_i z_i = 0, \quad \forall i, \tag{2.99d}$$

$$\nu_j y_j = 0, \quad \forall j, \tag{2.99e}$$

$$\sum_j y_j - \sum_i (z_i - C_i) = 0, \tag{2.99f}$$

$$z_i > 0, \quad \forall i, \tag{2.99g}$$

$$y_j \geq 0, \quad \forall j, \tag{2.99h}$$

$$\lambda, \mu_i, \nu_j \geq 0, \quad \forall i, j. \tag{2.99i}$$

By Assumption 2.6, we have that $y_j > 0$ and thus $\nu_j = 0$ for all $j \in [N]$. With some algebra, we can conclude from (2.99) that

$$\lambda^S = \frac{N\alpha \sum_i \frac{1}{2a_i} + C + \sum_i \frac{b_i}{2a_i}}{(N+1)\sum_i \frac{1}{2a_i}},$$

$$z_i^S = \frac{\lambda^S - b_i}{2a_i},$$

$$y_j^S = \frac{\sum_j y_j^S}{N} = \frac{\sum_i (z_i^S - C_i)}{N} = \frac{-C + \sum_i \frac{\lambda^S - b_i}{2a_i}}{N} = \frac{-C + \sum_i \frac{\alpha - b_i}{2a_i}}{N+1}.$$

We note that the optimal z_i^S we derived from the above also satisfies (2.41), and the optimal y_j^S from the above also satisfies (2.49), i.e., z_i^S and y_j^S are optimal to prosumer i and generator j, respectively. The optimal social welfare is given by

$$\begin{aligned}\mathcal{W}^S &= \sum_i u_i(z_i^S) - \sum_j c_j(y_j^S) \\ &= \left(\frac{\sum_i \left[\frac{\alpha N + b_i}{2a_i} + C_i \right]}{(N+1)\sum_i \frac{1}{2a_i}} \right)^2 \sum_i \frac{1}{4a_i} - \sum_i \frac{b_i^2}{4a_i} + \alpha \frac{N \sum_i \left(C_i - \frac{\alpha - b_i}{2a_i} \right)}{N+1}.\end{aligned}$$

This completes the proof of Proposition 2.7. □

Proof of Proposition 2.8. We write the Lagrangian of (2.58) as

$$\mathcal{L} = \sum_i u_i(C_i + d_i) - \sum_j c_j(y_j) + \lambda \left(\sum_j y_j - \sum_i d_i \right) + \mu_i d_i + \nu_j y_j, \tag{2.100}$$

where λ, μ_i, ν_j are the Lagrange multipliers of the constraints. The KKT optimality conditions are

$$\frac{\partial \mathcal{L}}{\partial d_i} = 2a_i(C_i + d_i) + b_i - \lambda + \mu_i = 0, \quad \forall i, \tag{2.101a}$$

$$\frac{\partial \mathcal{L}}{\partial y_j} = -\alpha + \lambda + \nu_j = 0, \quad \forall j, \tag{2.101b}$$

$$\lambda \left(\sum_j y_j - \sum_i d_i \right) = 0, \tag{2.101c}$$

$$\mu_i d_i = 0, \quad \forall i, \tag{2.101d}$$

$$\nu_j y_j = 0, \quad \forall j, \tag{2.101e}$$

$$\sum_j y_j - \sum_i d_i = 0, \tag{2.101f}$$

$$d_i \geq 0, \quad \forall i, \tag{2.101g}$$

$$y_j \geq 0, \quad \forall j, \tag{2.101h}$$

$$\lambda, \mu_i, \nu_j \geq 0, \quad \forall i, j. \tag{2.101i}$$

By Assumption (2.6), we have that $y_j > 0$ and thus $\nu_j = 0$ for all $j \in [N]$ by (2.101e). From (2.101b), we have that $\lambda^{TN} = \alpha$. Also (2.101a), (2.101d), and (2.101g) together imply that

$$d_i^{TN} = \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+, \quad \forall i.$$

From (2.101c), we then have that

$$\sum_i d_i^{TN} = \sum_i \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+ = \sum_j y_j^{TN} = N \cdot y_j^{TN},$$

which implies that

$$y_j^{TN} = \frac{\sum_i \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+}{N}, \quad \forall j.$$

Therefore, we may write the social welfare as

$$\begin{aligned}
\mathcal{W}^{TN} &= \sum_i u_i(C_i + d_i^{TN}) - \sum_j c_j(y_j^{TN}) \\
&= \sum_i \left[a_i \left(C_i + \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+ \right)^2 + b_i \left(C_i + \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+ \right) \right] \\
&\quad - \alpha \sum_i \left[-C_i + \frac{\alpha - b_i}{2a_i}\right]^+
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\{i|2a_iC_i+b_i>\alpha\}} \left[\frac{(\alpha-b_i)^2}{4a_i} + \frac{b_i(\alpha-b_i)}{2a_i} + \alpha C_i - \frac{\alpha(\alpha-b_i)}{2a_i} \right] \\
&\quad + \sum_{\{i|2a_iC_i+b_i\leq\alpha\}} \left(a_iC_i^2 + b_iC_i \right) \\
&= \sum_{\{i|2a_iC_i+b_i>\alpha\}} \left[\alpha C_i - \frac{(b_i-\alpha)^2}{4a_i} \right] + \sum_{\{i|2a_iC_i+b_i\leq\alpha\}} \left(a_iC_i^2 + b_iC_i \right).
\end{aligned}
$$

□

Proof of Lemma 2.6. Recall from (2.62) that

$$
u(C+y) = \left\{ \max_{d_i\geq 0} \sum_i u_i(C_i+d_i) \quad \text{s.t.} \quad \sum_i d_i = y \right\},
$$

which is itself an optimization problem. We write its Lagrangian:

$$
\mathcal{L} = \sum_i u_i(C_i+d_i) + \lambda \left(y - \sum_i d_i \right) + \mu_i d_i,
$$

where λ and μ_i are the Lagrange multipliers of the constraints. The KKT optimality conditions are

$$
\frac{\partial \mathcal{L}}{\partial d_i} = 2a_i(C_i+d_i) + b_i - \lambda + \mu_i = 0, \quad \forall i \tag{2.102a}
$$

$$
\lambda \left(y - \sum_i d_i \right) = 0 \tag{2.102b}
$$

$$
\mu_i d_i = 0, \quad \forall i \tag{2.102c}
$$

$$
\sum_i d_i = y \tag{2.102d}
$$

$$
d_i \geq 0, \quad \forall i \tag{2.102e}
$$

$$
\lambda, \mu_i \geq 0, \quad \forall i. \tag{2.102f}
$$

The optimal d_i is given by

$$d_i = \frac{\lambda - \mu_i - b_i}{2a_i} - C_i.$$

Since $\mu_i d_i = 0$, for any given λ, we have the set of prosumers with $d_i > 0$, i.e.,

$$\mathcal{S}(\lambda) = \{i \mid d_i > 0\} = \left\{ i \;\middle|\; \frac{\lambda - b_i}{2a_i} - C_i > 0 \right\} = \{i \mid \lambda < 2a_i C_i + b_i\}. \tag{2.103}$$

Similarly, any prosumers in $\mathcal{S}^c(\lambda) := \{i \mid \lambda \geq 2a_i C_i + b_i\}$ will have $d_i = 0$. Thus, we have that

$$\sum_i d_i = \sum_{i \in \mathcal{S}(\lambda)} \left[\frac{\lambda - b_i}{2a_i} - C_i \right] = y,$$

which implies that

$$\lambda = \frac{y + \sum_{i \in \mathcal{S}(\lambda)} C_i + \sum_{i \in \mathcal{S}(\lambda)} \frac{b_i}{2a_i}}{\sum_{i \in \mathcal{S}(\lambda)} \frac{1}{2a_i}}. \tag{2.104}$$

We will figure out an expression of the set $\mathcal{S}$ as a function of y. Recall that prosumers are sorted in decreasing order of $2a_i C_i + b_i$. If $\lambda \geq 2a_1 C_1 + b_1$, then all prosumers have $d_i = 0$, and $\mathcal{S}(\lambda)$ is empty. As λ decreases to $2a_1 C_1 + b_1$, the first prosumer is included in the set.

When the set $\mathcal{S}(\lambda)$ does not change, as y increases, λ will decrease according to (2.104). When y increases to some critical point y^i that the prosumer $i > 1$ is just about to be included in the set $\mathcal{S}$, we look at the corresponding λ right before i is included:

$$\lambda_- = 2a_i C_i + b_i = \frac{y^i + \sum_{i'=1}^{i-1} \left(C_{i'} + \frac{b_{i'}}{2a_{i'}} \right)}{\sum_{i'=1}^{i-1} \frac{1}{2a_{i'}}}, \tag{2.105}$$

which implies that

$$y^i = (2a_iC_i + b_i)\sum_{i'=1}^{i-1}\frac{1}{2a_{i'}} - \sum_{i'=1}^{i-1}\left(C_{i'} + \frac{b_{i'}}{2a_{i'}}\right). \tag{2.106}$$

When prosumer i is just included in the set, we have that

$$\lambda_+ = \frac{y^i + \sum_{i'=1}^{i}\left(C_{i'} + \frac{b_{i'}}{2a_{i'}}\right)}{\sum_{i'=1}^{i}\frac{1}{2a_{i'}}}. \tag{2.107}$$

One can verify that $\lambda_+ = \lambda_-$, which implies that as y increases, λ continuously decreases, even at those critical points when more prosumers are being added to the set $\mathcal{S}$. Therefore, the set of prosumers with $d_i > 0$ can be expressed as

$$\mathcal{S}(y) = \left\{i \mid y > y^i\right\}. \tag{2.108}$$

We may rewrite $u(C+y)$ as

$$\begin{aligned}
u(C+y) &= \sum_{i\in\mathcal{S}^c(y)}\left(a_iC_i^2 + b_iC_i\right) + \sum_{i\in\mathcal{S}(y)}\left[a_i\left(\frac{\lambda - b_i}{2a_i}\right)^2 + b_i\left(\frac{\lambda - b_i}{2a_i}\right)\right] \\
&= \sum_{i\in\mathcal{S}}\left[\frac{1}{4a_i}\left(\frac{y + \sum_{i\in\mathcal{S}}C_i + \sum_{i\in\mathcal{S}}\frac{b_i}{2a_i}}{\sum_{i\in\mathcal{S}}\frac{1}{2a_i}} - b_i\right)^2\right. \\
&\quad \left. + \frac{b_i}{2a_i}\left(\frac{y + \sum_{i\in\mathcal{S}}C_i + \sum_{i\in\mathcal{S}}\frac{b_i}{2a_i}}{\sum_{i\in\mathcal{S}}\frac{1}{2a_i}} - b_i\right)\right] + \sum_{i\in\mathcal{S}^c}\left(a_iC_i^2 + b_iC_i\right)
\end{aligned}$$

When y is within the range that $\mathcal{S}$ does not change, we have that

$$\frac{\partial u(C+y)}{\partial y} = \sum_{i\in\mathcal{S}}\left[\frac{1}{2a_i}\left(\frac{y + \sum_{i\in\mathcal{S}}C_i + \sum_{i\in\mathcal{S}}\frac{b_i}{2a_i}}{\sum_{i\in\mathcal{S}}\frac{1}{2a_i}} - b_i\right)\frac{1}{\sum_{i\in\mathcal{S}}\frac{1}{2a_i}} + \frac{b_i}{2a_i}\frac{1}{\sum_{i\in\mathcal{S}}\frac{1}{2a_i}}\right]$$

$$
\begin{aligned}
&= \sum_{i\in\mathcal{S}} \frac{1}{2a_i} \frac{1}{\sum_{i\in\mathcal{S}} \frac{1}{2a_i}} \left(\frac{y + \sum_{i\in\mathcal{S}} C_i + \sum_{i\in\mathcal{S}} \frac{b_i}{2a_i}}{\sum_{i\in\mathcal{S}} \frac{1}{2a_i}} \right) \\
&= \frac{y + \sum_{i\in\mathcal{S}} C_i + \sum_{i\in\mathcal{S}} \frac{b_i}{2a_i}}{\sum_{i\in\mathcal{S}} \frac{1}{2a_i}} \\
&= \lambda.
\end{aligned} \tag{2.109}
$$

Thus, we can conclude that, as y increases, $\frac{\partial u(C+y)}{\partial y}$ continuously decreases, and the overall utility of consumption $u(C+y)$ is continuous and differentiable in y. □

Proof of Lemma 2.7. Follows directly from the proof of Lemma 2.6. □

Proof of Lemma 2.8. From (2.70), and (2.109), we have that

$$
\begin{aligned}
\left.\frac{\partial u(C+y)}{\partial y}\right|_{y=Ny_j^{SN}} &= \frac{N\frac{\sum_i \left(\frac{\alpha-b_i}{2a_i} - C_i\right)\cdot \mathbb{1}\left\{Ny_j^{SN} > y^i\right\}}{N+1} + \sum_i \left(C_i + \frac{b_i}{2a_i}\right)\cdot \mathbb{1}\left\{Ny_j^{SN} > y^i\right\}}{\sum_{i\in\mathcal{S}} \frac{1}{2a_i}} \\
&= \frac{N\alpha \sum_i \frac{1}{2a_i}\cdot \mathbb{1}\left\{Ny_j^{SN} > y^i\right\} + \sum_i \left(C_i + \frac{b_i}{2a_i}\right)\cdot \mathbb{1}\left\{Ny_j^{SN} > y^i\right\}}{(N+1)\sum_i \frac{1}{2a_i}\cdot \mathbb{1}\left\{Ny_j^{SN} > y^i\right\}}.
\end{aligned}
$$

Therefore, bidding (2.72) ensures that the condition (2.71) is satisfied, which ensures that the system operator will optimally assign y_j^{SN} to generator j. □

Proof of Proposition 2.9. First consider the equivalent problem (2.62), given the bids $\tilde{c}_j(y_j)$, from (2.71), the system operator chooses the optimal $y^{SN} = Ny_j^{SN}$ where y_j^{SN} satisfies (2.70).

Next, we write the Lagrangian of (2.61) as

$$
\mathcal{L} = \sum_i u_i(C_i + d_i) - \sum_j \tilde{c}_j(y_j) + \lambda \left(\sum_j y_j - \sum_i d_i \right) + \mu_i d_i + \nu_j y_j, \tag{2.110}
$$

where λ, μ_i, ν_j are the Lagrange multipliers of the constraints, and $\tilde{c}_j$ is given in (2.51). The

KKT optimality conditions are

$$\frac{\partial \mathcal{L}}{\partial d_i} = 2a_i(C_i + d_i) + b_i - \lambda + \mu_i = 0, \quad \forall i, \tag{2.111a}$$

$$\frac{\partial \mathcal{L}}{\partial y_j} = \frac{N\alpha \sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\} + \sum_i \left(C_i + \frac{b_i}{2a_i}\right) \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\}}{(N+1)\sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{N y_j^{SN} > y^i\right\}} + \lambda + \nu_j = 0, \quad \forall j, \tag{2.111b}$$

$$\lambda \left(\sum_j y_j - \sum_i d_i \right) = 0, \tag{2.111c}$$

$$\mu_i d_i = 0, \quad \forall i, \tag{2.111d}$$

$$\nu_j y_j = 0, \quad \forall j, \tag{2.111e}$$

$$\sum_j y_j - \sum_i d_i = 0, \tag{2.111f}$$

$$d_i \geq 0, \quad \forall i, \tag{2.111g}$$

$$y_j \geq 0, \quad \forall j, \tag{2.111h}$$

$$\lambda, \mu_i, \nu_j \geq 0, \quad \forall i, j. \tag{2.111i}$$

By Assumption 2.6, we have that $y_j > 0$ and thus $\nu_j = 0$ for all $j \in [N]$. Using (2.111) and y_j^{SN} from (2.70), we can conclude that

$$\lambda^{SN} = \frac{N\alpha}{N+1} + \frac{\sum_i \left(C_i + \frac{b_i}{2a_i}\right) \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}}{(N+1)\sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{y^{SN} > y^i\right\}},$$

$$d_i^{SN} = \left[-C_i + \frac{\lambda^{SN} - b_i}{2a_i} \right]^+$$

We note that the optimal d_i^{SN} and λ^{SN} in the above also satisfies (2.56), and is thus optimal for the prosumer i. The y_j^{SN} from the above also satisfies (2.70), and is thus optimal for the

generator j. The optimal social welfare is given by

$$\begin{aligned}
\mathcal{W}^{SN} &= \sum_i u_i(C_i + d_i^{SN}) - \sum_j c_j(y_j^{SN}) \\
&= \sum_i \left[\frac{1}{4a_i} \left(\frac{\sum_i \left[\frac{\alpha N + b_i}{2a_i} + C_i \right] \cdot \mathbb{1}\left\{ y^{SN} > y^i \right\}}{(N+1)\sum_i \frac{1}{2a_i} \cdot \mathbb{1}\left\{ y^{SN} > y^i \right\}} \right)^2 - \frac{b_i^2}{4a_i} \right] \cdot \mathbb{1}\left\{ y^{SN} > y^i \right\} \\
&\quad + \sum_i \left(a_i C_i^2 + b_i C_i \right) \cdot \mathbb{1}\left\{ y^{SN} \leq y^i \right\} + \alpha \frac{N \sum_i \left(C_i - \frac{\alpha - b_i}{2a_i} \right) \cdot \mathbb{1}\left\{ y^{SN} > y^i \right\}}{N+1}.
\end{aligned}$$

This completes the proof of Proposition 2.9. □

Proof of Proposition 2.10. First, from the Taylor expansion, we have that

$$u_i(C_i) = u_i(z_i) - u_i'(z_i)(z_i - C_i) + \frac{1}{2} u_i''(z_i)(z_i - C_i)^2. \tag{2.112}$$

Then, we can write the social welfare $\mathcal{W}^T$ as

$$\begin{aligned}
\mathcal{W}^T &= \sum_i u_i(z_i^T) - \sum_j c_j(y_j^T) \\
&= \sum_i u_i(z_i^T) - \alpha y^T \\
&= \sum_i u_i(z_i^T) - \alpha \sum_i z_i^T \\
&= \sum_i \left[u_i(C_i) + u_i'(z_i^T) \cdot (z_i^T - C_i) - \frac{1}{2} u_i''(z_i^T) \cdot \left(z_i^T - C_i \right)^2 \right] - \alpha \sum_i \left(z_i^T - C_i \right) \\
&= \sum_i \left[u_i(C_i) + \alpha \left(z_i^T - C_i \right) - \frac{1}{2} u_i''(z_i^T) \cdot \left(z_i^T - C_i \right)^2 - \alpha \left(z_i^T - C_i \right) \right] \\
&= \mathcal{W}_0 - \sum_i a_i \left(z_i^T - C_i \right)^2,
\end{aligned}$$

where we have used the fact that $u_i'(z_i^T) = \alpha$.

When the generators bid strategically, we have that

$$\begin{aligned}\mathcal{W}^S &= \sum_i u_i(z_i^S) - \sum_j c_j(y_j^S) \\ &= \sum_i u_i(z_i^S) - \alpha \sum_i \left(z_i^S - C_i\right) \\ &= \sum_i \left[u_i(C_i) + u_i'(z_i^S) \cdot \left(z_i^S - C_i\right) - \frac{1}{2} u_i''(z_i^S) \cdot \left(z_i^S - C_i\right)^2\right] - \alpha \sum_i \left(z_i^S - C_i\right).\end{aligned}$$

In this case, we also have that $u'(z_i^S) = \lambda^S = \lambda^T + \delta = \alpha + \delta$. Therefore,

$$\begin{aligned}\mathcal{W}^S &= \sum_i \left[u_i(C_i) + (\alpha + \delta) \cdot \left(z_i^S - C_i\right) - a_i \left(z_i^S - C_i\right)^2 - \alpha \left(z_i^S - C_i\right)\right] \\ &= \mathcal{W}_0 + \sum_i \left[-a_i \left(z_i^S - C_i\right)^2 + \delta \left(z_i^S - C_i\right)\right].\end{aligned}$$

We next move to the case of no prosumer participation. Under truthful bidding, we have that

$$\begin{aligned}\mathcal{W}^{TN} &= \sum_i u_i(C_i + d_i^{TN}) - \sum_j c_j(y_j^{TN}) \\ &= \sum_i u_i(C_i + d_i^{TN}) - \alpha y^{TN} \\ &= \sum_i u_i(C_i + d_i^{TN}) - \alpha \sum_i d_i^{TN}\end{aligned}$$

For those prosumers with $d_i^{TN} > 0$, we have that $d_i^{TN} = z_i^T - C_i$ and $u'(C_i + d_i^{TN}) = \lambda^{TN} = \alpha$. Thus,

$$\begin{aligned}\mathcal{W}^{TN} &= \sum_{\{i|d_i^{TN}>0\}} u_i(C_i + d_i^{TN}) + \sum_{\{i|d_i^{TN}=0\}} u_i(C_i) - \alpha \sum_{\{i|d_i^{TN}>0\}} d_i^{TN} \\ &= \sum_{\{i|d_i^{TN}>0\}} \left[u_i(C_i) + u_i'(C_i + d_i^{TN}) \cdot d_i^{TN} - \frac{1}{2} u_i''(C_i + d_i^{TN}) \cdot \left(d_i^{TN}\right)^2 - \alpha d_i^{TN}\right]\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\{i|d_i^{TN}=0\}} u_i(C_i) \\
&= \mathcal{W}_0 - \sum_i a_i \left(d_i^{TN}\right)^2 = \mathcal{W}_0 - \sum_{\{i|z_i^T > C_i\}} \left(z_i^T - C_i\right)^2.
\end{aligned}
$$

Under strategic bidding, we have that

$$
\begin{aligned}
\mathcal{W}^{SN} &= \sum_i u_i(C_i + d_i^{SN}) - \sum_j c_j(y_j^{SN}) \\
&= \sum_i u_i(C_i + d_i^{SN}) - \alpha \sum_i d_i^{SN} \\
&= \sum_{\{i|d_i^{SN}>0\}} u_i(C_i + d_i^{SN}) + \sum_{\{i|d_i^{SN}=0\}} u_i(C_i) - \alpha \sum_{\{i|d_i^{SN}>0\}} d_i^{SN} \\
&= \sum_{\{i|d_i^{SN}>0\}} \left[u_i(C_i) + u_i'(C_i + d_i^{SN}) \cdot d_i^{SN} - \frac{1}{2} u_i''(C_i + d_i^{SN}) \cdot \left(d_i^{SN}\right)^2 - \alpha d_i^{SN}\right] \\
&\quad + \sum_{\{i|d_i^{SN}=0\}} u_i(C_i)
\end{aligned}
$$

For those i such that $d_i^{SN} > 0$, we also have that $u'(C_i + d_i^{SN}) = \lambda^{SN} = \lambda^{TN} + \delta^N = \alpha + \delta^N$. Thus,

$$
\begin{aligned}
\mathcal{W}^{SN} &= \sum_{\{i|d_i^{SN}>0\}} \left[(\alpha + \delta^N) \cdot d_i^{SN} - a_i \left(d_i^{SN}\right)^2 - \alpha d_i^{SN}\right] + \sum_i u_i(C_i) \\
&= \mathcal{W}_0 - \sum_i a_i d_i^{SN} + \delta^N \sum_i d_i^{SN}.
\end{aligned}
$$

This completes the proof of Proposition 2.10. □

Proof of Proposition 2.11. In the following, we prove each relation separately.

- (2.85a) follows directly from (2.77) and (2.79):

$$\begin{aligned}\mathcal{W}^T - \mathcal{W}^{TN} &= -\sum_i a_i\Big(z_i^T - C_i\Big)^2 + \sum_{\{i|z_i^T > C_i\}} a_i\left(z_i^T - C_i\right)^2 \\ &= -\sum_{\{i|z_i^T \le C_i\}} a_i\left(z_i^T - C_i\right)^2 \ge 0.\end{aligned}$$

- From (2.78) and (2.80), we have that

$$\begin{aligned}&\mathcal{W}^S - \mathcal{W}^{SN} \\ &= \sum_i \left(-a_i\Big(z_i^S - C_i\Big)^2 + \delta\left(z_i^S - C_i\right)\right) + \sum_i a_i\left(d_i^{SN}\right)^2 - \delta^N \sum_i d_i^{SN} \\ &= \sum_i \left(-a_i\left(\frac{\delta}{2a_i} + z_i^T - C_i\right)^2 + \delta\left(\frac{\delta}{2a_i} + z_i^T - C_i\right)\right) \\ &\quad + \sum_{\{i|d_i^{SN}>0\}} a_i\left(\frac{\delta^N}{2a_i} + z_i^T - C_i\right)^2 - \delta^N \sum_i d_i^{SN} \\ &= -\sum_i \left(-\frac{\delta^2}{4a_i} + a_i\left(z_i^T - C_i\right)^2\right) \\ &\quad + \sum_{\{i|d_i^{SN}>0\}} \left(\frac{\delta^N{}^2}{4a_i} + a_i\left(z_i^T - C_i\right)^2 + \delta^N\left(z_i^T - C_i\right)\right) \\ &\quad - \delta^N \sum_{\{i|d_i^{SN}>0\}} \left(\frac{\delta^N}{2a_i} + z_i^T - C_i\right) \\ &= -\sum_i \left(-\frac{\delta^2}{4a_i} + a_i\left(z_i^T - C_i\right)^2\right) + \sum_{\{i|d_i^{SN}>0\}} \left(-\frac{\delta^N{}^2}{4a_i} + a_i\left(z_i^T - C_i\right)^2\right) \\ &= -\sum_{\{i|d_i^{SN}=0\}} a_i\left(z_i^T - C_i\right)^2 + \frac{\left(\sum_i\left(z_i^T - C_i\right)\right)^2}{(N+1)^2\sum_i \frac{1}{a_i}} - \frac{\left(\sum_{\{i|d_i^{SN}>0\}}\left(z_i^T - C_i\right)\right)^2}{(N+1)^2\sum_{\{i|d_i^{SN}>0\}}\frac{1}{a_i}},\end{aligned}$$

where the first term is nonnegative. To see (2.85b), it suffices to show that

$$\frac{\left(\sum_i \left(z_i^T - C_i\right)\right)^2}{(N+1)^2 \sum_i \frac{1}{a_i}} \geq \frac{\left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)\right)^2}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{a_i}}. \tag{2.113}$$

Note that in the no participation case, under strategic bidding, the generators earn total profit $\delta^N \sum_{\{i|d_i^{SN}>0\}} d_i^{SN}$, which must be at least the profit earned under the truthful bidding, i.e.,

$$\delta^N \sum_{\{i|d_i^{SN}>0\}} d_i^{SN} \geq \delta \sum_{\{i|z_i^S>0\}} z_i^S,$$

which implies that

$$\frac{\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)}{(N+1)\sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} \sum_{\{i|d_i^{SN}>0\}} d_i^{SN} \geq \frac{\sum_i \left(z_i^T - C_i\right)}{(N+1)\sum_i \frac{1}{-2a_i}} \sum_{\{i|z_i^S>C_i\}} \left(z_i^S - C_i\right).$$

Thus, we have that

$$\begin{aligned}\frac{N\left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)\right)^2}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} &\geq \frac{\sum_i \left(z_i^T - C_i\right)}{(N+1)\sum_i \frac{1}{-2a_i}} \sum_{\{i|z_i^S>C_i\}} \left(z_i^S - C_i\right) \\ &\geq \frac{\sum_i \left(z_i^T - C_i\right)}{(N+1)\sum_i \frac{1}{-2a_i}} \sum_i \left(z_i^S - C_i\right),\end{aligned}$$

which implies that

$$\frac{N\left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)\right)^2}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} \geq \frac{N\left(\sum_i \left(z_i^T - C_i\right)\right)^2}{(N+1)^2 \sum_i \frac{1}{-2a_i}},$$

which implies (2.113).

- To see (2.85c), from (2.77), (2.78), and (2.82), we have that

$$
\begin{aligned}
&\mathcal{W}^T - \mathcal{W}^S \\
&= -\sum_i a_i \left(\left(z_i^T - C_i \right)^2 - \left(z_i^S - C_i \right)^2 \right) - \left(\frac{N}{(N+1)^2 \sum_i \frac{1}{-2a_i}} \right) \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \\
&= -\sum_i a_i \left(\left(z_i^T - C_i \right)^2 - \left(\frac{\delta}{2a_i} + z_i^T - C_i \right)^2 \right) \\
&\quad - \left(\frac{N}{(N+1)^2 \sum_i \frac{1}{-2a_i}} \right) \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \\
&= \sum_i a_i \left(\frac{\delta^2}{4a_i^2} + \frac{\delta \left(z_i^T - C_i \right)}{a_i} \right) - \left(\frac{N}{(N+1)^2 \sum_i \frac{1}{-2a_i}} \right) \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \\
&= \sum_i \left(\frac{\delta^2}{4a_i} + \delta \left(z_i^T - C_i \right) \right) - \left(\frac{N}{(N+1)^2 \sum_i \frac{1}{-2a_i}} \right) \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \\
&= \sum_i \left(\frac{\frac{\left(\sum_i \left(z_i^T - C_i \right) \right)^2}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2}}{4a_i} + \delta \left(z_i^T - C_i \right) \right) + \left(\frac{N}{(N+1)^2 \sum_i \frac{1}{2a_i}} \right) \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \\
&= \left(\frac{\left(\sum_i \left(z_i^T - C_i \right) \right)^2}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2} \right) \sum_i \frac{1}{4a_i} + \sum_i \delta \left(z_i^T - C_i \right) \\
&\quad + \left(\frac{N}{(N+1)^2 \sum_i \frac{1}{2a_i}} \right) \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \\
&= \left(\frac{\left(\sum_i \left(z_i^T - C_i \right) \right)^2 \left(\sum_i \frac{1}{4a_i} \right)}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2} \right) + \left(\frac{\left(\sum_i \left(z_i^T - C_i \right) \right)^2}{-(N+1) \sum_i \frac{1}{2a_i}} \right) \\
&\quad + \left(\frac{N \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \left(\sum_i \frac{1}{2a_i} \right)}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2} \right) \\
&= \left(\frac{\left(\sum_i \left(z_i^T - C_i \right) \right)^2 \left(\sum_i \frac{1}{4a_i} \right)}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2} \right) + \left(\frac{\left(\sum_i \left(z_i^T - C_i \right) \right)^2 \left(-(N+1) \sum_i \frac{1}{2a_i} \right)}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2} \right)
\end{aligned}
$$

$$
\begin{aligned}
&+ \left(\frac{N \left(\sum_i \left(z_i^T - C_i \right) \right)^2 \left(\sum_i \frac{1}{2a_i} \right)}{(N+1)^2 \left(\sum_i \frac{1}{2a_i} \right)^2} \right) \\
&= \left(\frac{\sum_i \left(z_i^T - C_i \right)}{(N+1) \sum_i \frac{1}{2a_i}} \right)^2 \left(- \sum_i \frac{1}{4a_i} \right) \qquad (2.114) \\
&\geq 0.
\end{aligned}
$$

- To see (2.85d), from (2.79), (2.80), and (2.84), we have that

$$
\begin{aligned}
&\mathcal{W}^{TN} - \mathcal{W}^{SN} \\
&= - \sum_{\{i | z_i^T > C_i\}} a_i \left(z_i^T - C_i \right)^2 + \sum_i a_i (d_i^{SN})^2 - \delta^N \sum_i d_i^{SN} \\
&= - \sum_{\{i | z_i^T > C_i\}} a_i \left(z_i^T - C_i \right)^2 + \sum_{\{i | d_i^{SN} > 0\}} a_i (d_i^{SN})^2 - \delta^N \sum_i d_i^{SN} \\
&= - \sum_{\{i | z_i^T > C_i\}} a_i \left(z_i^T - C_i \right)^2 + \sum_{\{i | d_i^{SN} > 0\}} a_i \left(\frac{\delta^N}{2a_i} + z_i^T - C_i \right)^2 - \delta^N \sum_i d_i^{SN} \\
&= - \sum_{\{i | z_i^T > C_i\}} a_i \left(z_i^T - C_i \right)^2 + \sum_{\{i | d_i^{SN} > 0\}} a_i \left(\frac{\delta^N}{2a_i} + z_i^T - C_i \right)^2 \\
&\quad - \left(\frac{N}{(N+1)^2 \sum_{\{i | d_i^{SN} > 0\}} \frac{1}{-2a_i}} \right) \left(\sum_{\{i | d_i^{SN} > 0\}} \left(z_i^T - C_i \right) \right)^2 \\
&= - \sum_{\{i | d_i^{SN} = 0,\ z_i^T > C_i\}} a_i \left(z_i^T - C_i \right)^2 - \sum_{\{i | d_i^{SN} > 0\}} a_i \left(z_i^T - C_i \right)^2 \\
&\quad + \sum_{\{i | d_i^{SN} > 0\}} a_i \left(\frac{\delta^N}{2a_i} + z_i^T - C_i \right)^2
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \left(\frac{N}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} \right) \left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i \right) \right)^2 \\
&= - \sum_{\{i|d_i^{SN}=0,\ z_i^T>C_i\}} a_i \left(z_i^T - C_i \right)^2 - \sum_{\{i|d_i^{SN}>0\}} a_i \left(z_i^T - C_i \right)^2 \\
&\quad + \sum_{\{i|d_i^{SN}>0\}} \left(\frac{{\delta^N}^2}{4a_i} + a_i \left(z_i^T - C_i \right)^2 + \delta^N \left(z_i^T - C_i \right) \right) \\
&\quad - \left(\frac{N}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} \right) \left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i \right) \right)^2 \\
&= - \sum_{\{i|d_i^{SN}=0,\ z_i^T>C_i\}} a_i \left(z_i^T - C_i \right)^2 + \sum_{\{i|d_i^{SN}>0\}} \left(\frac{(\delta^N)^2}{4a_i} + \delta^N \left(z_i^T - C_i \right) \right) \\
&\quad - \left(\frac{N}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} \right) \left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i \right) \right)^2 \\
&= - \sum_{\{i|d_i^{SN}=0,\ z_i^T>C_i\}} a_i \left(z_i^T - C_i \right)^2 + \sum_{\{i|d_i^{SN}>0\}} \frac{\frac{\left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i \right) \right)^2}{(N+1)^2 \left(\sum_{\{i|d_i^{SN}>0\}} \frac{1}{2a_i} \right)^2}}{4a_i} \\
&\quad + \sum_{\{i|d_i^{SN}>0\}} \delta^N \left(z_i^T - C_i \right) \\
&\quad - \left(\frac{N}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{-2a_i}} \right) \left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i \right) \right)^2 \\
&= - \sum_{\{i|d_i^{SN}=0,\ z_i^T>C_i\}} a_i \left(z_i^T - C_i \right)^2 \\
&\quad + \frac{\left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i \right) \right)^2 \left(\sum_{\{i|d_i^{SN}>0\}} \frac{1}{4a_i} \right)}{(N+1)^2 \left(\sum_{\{i|d_i^{SN}>0\}} \frac{1}{2a_i} \right)^2}
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{\{i|d_i^{SN}>0\}} \delta^N \left(z_i^T - C_i\right) + \left(\frac{N \left(\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)\right)^2}{(N+1)^2 \sum_{\{i|d_i^{SN}>0\}} \frac{1}{2a_i}} \right) \\
&= - \sum_{\{i|d_i^{SN}=0,\ z_i^T>C_i\}} a_i \left(z_i^T - C_i\right)^2 \\
&\quad + \left(\frac{\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)}{(N+1) \sum_{\{i|d_i^{SN}>0\}} \frac{1}{2a_i}} \right)^2 \left(- \sum_{\{i|d_i^{SN}>0\}} \frac{1}{4a_i} \right) \\
&= \sum_{\{i|d_i^{SN}=0,\ z_i^T>C_i\}} (-a_i) \left(z_i^T - C_i\right)^2 \\
&\quad + \left(\frac{\sum_{\{i|d_i^{SN}>0\}} \left(z_i^T - C_i\right)}{N+1} \right)^2 \left(\frac{1}{\sum_{\{i|d_i^{SN}>0\}} \frac{1}{-a_i}} \right) \qquad (2.115) \\
&\geq 0.
\end{aligned}
$$

- From (2.113), we can conclude that (2.114) $\geq$ (2.115), which implies (2.85e).

□

CHAPTER 3

FINITE-SAMPLE ANALYSIS OF DECENTRALIZED Q-LEARNING FOR STOCHASTIC GAMES

3.1 Introduction

Multi-agent learning has received extensive attention in recent years, and has achieved considerable success in application areas including traffic control, network routing, energy distribution, robotic systems, and social economic problems, where multiple agents learn concurrently how to solve a task by interacting with the same environment (Stone and Veloso [201]). The canonical model for dynamic multi-agent interactions is *stochastic games*, also known as Markov games (Littman [147]), which were first introduced by Shapley [191]; we refer the interested reader to Busoniu et al. [45] and Zhang et al. [232] for comprehensive surveys. Compared to repeated games, i.e., repeated play of static games (Wu et al. [224]), stochastic games are more general in the sense that each stage game is affected by the previous joint actions of all agents, who may or may not be in a network (Correa et al. [59], Kempe et al. [128], Qu et al. [177], Lin et al. [146]) through the system state evolution, and thus are applicable to a broader set of problem settings (Stern and Birge [200], Birge et al. [41]).

There are different attributes that may be associated with a stochastic game. Depending on the number of agents and their reward functions, a stochastic game can be a two-agent zero-sum game and/or a multi-agent general-sum game, where the former games have two agents, and the reward of one agent is always the negation of the reward of the other agent, representing a fully competitive relationship, and the latter games may have N agents for any $N \geq 2$, with no restrictions on their reward functions (Başar and Olsder [26]). Depending on the length of the game, a stochastic game can have either finite horizon or infinite horizon, where the objective of each agent is to choose a policy (that maps from a state to an action) to maximize its total reward over the length of the horizon (if finite), or to maximize its

total discounted reward or time-averaged reward over an infinite horizon.

In all of these different types of stochastic games, the most commonly studied notion on the agents' joint policy is the so-called *Markov perfect equilibrium* (Maskin and Tirole [156]), or simply *Nash equilibrium* (Nash [166]) in some literature (Das et al. [60]). Roughly speaking, in a Markov perfect equilibrium, each agent's policy is a *best reply* (i.e., maximizes its own total (discounted) reward) to all other agents' joint policy. Mainstreams of research include analyzing the hardness to compute the equilibria (e.g., Daskalakis et al. [62], Daskalakis [61], Garg et al. [95]), approximating and analyzing the equilibria (e.g., Brânzei et al. [43], Adsul et al. [5], Boodaghians et al. [42]), designing algorithms to find the equilibria with the knowledge of the transitions and rewards (e.g., Hu and Wellman [108], Hansen et al. [101]) or without such knowledge (e.g. Arslan and Yüksel [21]).

With the recent boom of reinforcement learning (RL), there is a growing interest in applying the methods of RL to stochastic games; see Shoham et al. [193] and the references therein. The adaptive decision-making framework of RL, together with the context of multiple interacting learners, lead to multi-agent RL (MARL). MARL corresponds to the learning problem in a multi-agent system in which multiple RL agents learn simultaneously by interacting with the stochastic environment via a trial-and-error approach, from which they receive rewards for their actions. These MARL algorithms can be either centralized (meaning that there is a central controller who has full access to the game setup as well as each agent's actions and rewards, and who provides coordination among these agents) or decentralized (meaning that each agent makes decisions based on local information without a coordinator). Developing convergent decentralized MARL algorithms, however, is well known to be challenging. In contrast to the single-agent scenario, such as bandit problems (Frazier et al. [84], Emamjomeh-Zadeh et al. [68]), the state evolution and the rewards earned by each agent depend on not only the current state and this agent's action, but also the actions taken by other agents. As a consequence, the existing single-agent learning algorithms can-

not be directly extended to multi-agent settings, due to the fact that the environment is now non-stationary from each agent's perspective, resulting in potential non-convergence, as shown in Tan [210], Claus and Boutilier [55]. Such nonstationarity issue is the key challenge to address in order to develop a converging multi-agent learning algorithm.

To address the nonstationarity issue, Arslan and Yüksel [21] proposed a decentralized Q-learning algorithm and proved the asymptotic convergence of the algorithm for a subclass of stochastic games – weakly acyclic games (Young [229]). Roughly speaking, a weakly acyclic game is a stochastic game such that best-reply dynamics cannot enter inescapable oscillations (see Definition 3.4 later for a precise definition). The standard Q-learning, a widely used model-free, value-based RL algorithm, has been applied to specific multiple-agent systems (Tan [210], Sen et al. [190]), but no analytical results exist regarding the convergence properties of Q-learning in a stochastic game setting. In the decentralized Q-learning algorithm by Arslan and Yüksel [21], agents do not update their policies for an extended period of time, which is called an *exploration phase*, during which the environment becomes stationary from each agent's perspective. During the exploration phases, the Q-functions of each agent are still being updated. The policy of each agent is then updated at the end of each exploration phase, according to the current values of the Q-functions. This "explore-then-update" procedure is repeated, and Arslan and Yüksel [21] have shown that the joint policy asymptotically converges to a Markov perfect equilibrium as the length of each exploration phase and the number of exploration phases go to infinity.

3.1.1 Contributions

In this chapter, we use the algorithm from Arslan and Yüksel [21] as a starting point, and build upon it with several new developments. We summarize our main contributions as follows.

- We study the non-asymptotic convergence guarantee, namely, sample complexity, of

the decentralized Q-learning algorithm in Arslan and Yüksel [21] (Theorem 3.1 of this chapter). We note that many of the generalizations from asymptotic convergence to finite-sample analysis turn out to be nontrivial. A brief overview of this is provided in Section 3.5. To the best of our knowledge, this is the first sample complexity result for convergence to Markov perfect equilibrium on multi-player general-sum stochastic games with infinite horizon. Moreover, the sample complexity result is expressed explicitly in the game parameters (Corollary 3.1), which is made possible by developing upper and lower bounds (Proposition 3.2) on some implicit parameters (e.g., the minimum probability of stationary distribution $\mu_{\min}$ and the mixing time t_{mix}).

- We apply linear function approximation to approximate the Q-functions in this general-sum stochastic game. Instead of maintaining a Q-function on all state-action pairs, each agent now restricts its attention to a linear space with smaller dimensions, compared to the large state/action space. Under the restriction of a smaller-dimensional linear space, the original Q-functions will not be fully recovered. As a result, the original Markov perfect equilibria might not be reachable. To this end, we define a new notion of equilibrium – *linear approximated equilibrium* (Definition 3.5). Roughly speaking, in a linear approximated equilibrium, each agent's policy is a best reply, according to the linearly approximated Q-functions (instead of the original ones) to all other agents' joint policy. The algorithm with linear function approximation (Algorithm 3.2) is shown in Section 3.4. We again prove the sample complexity result for the algorithm, i.e., we provide finite lower bounds on the length of exploration phases and the number of exploration phases needed for the joint policy of all agents to converge to a linear approximated equilibrium with high probability. This work also appears to be the first to apply function approximation to general-sum stochastic games.

- We provide numerical experiments of both algorithms (with and without linear function approximation) on the classical Grid World game (Sutton and Barto [207]) with minor

modifications. Specifically, we have two agents, one of whom moves in the vertical direction and the other one moves in the horizontal direction. At each time step, the state moves in the direction determined by the joint actions of both agents, and both agents receive a negative reward of -1, except when the current state is at a "terminal" state, where both agents receive a reward of 0 and the state does not change for all joint actions. This cooperative game belongs to a common subclass of stochastic games with identical reward functions for all agents, namely, the Markov team problems (Ho [105], Yüksel and Başar [230]). The experimental results confirm that, as the length and the number of exploration phases increase, the joint policy adopted by both agents converges to a Markov perfect equilibrium (in the tabular setting) or a linear approximated equilibrium (in the function approximation setting) with higher probability.

3.1.2 Related Work

This chapter is related to several sets of previous work. We mention below some of the most relevant ones.

3.1.2.1 Stochastic Games (SGs)

Stochastic games were proposed by Shapley [191] and can be viewed as a generalization of the Markov Decision Process (MDP) to the multi-agent setting. Since then, this framework has become a classical model for multi-agent learning, and there is a long line of work in finding the Markov perfect equilibrium (Nash equilibrium) of different types of stochastic games under various assumptions. The works of Littman [147, 148], Hu and Wellman [108], Hansen et al. [101], Wei et al. [220] assumed full knowledge of the transition kernel and the reward functions, while Wei et al. [219], Jia et al. [120], Sidford et al. [194], Zhang et al. [231], Wei et al. [221] assumed certain reachability conditions, e.g., access to some simulators that allow

each agent to directly sample transitions and rewards for each state-action pair. Most of these works were aimed at designing algorithms that asymptotically converge to Markov perfect equilibria.

Another set of recent works focused on the non-asymptotic sample complexity/regret guarantees for learning in SGs. For two-player zero-sum games, Bai and Jin [27], Xie et al. [225] developed the first provably-efficient learning algorithms based on optimistic value iteration. Liu et al. [150] improved upon these works with a model-based "Optimistic Nash Value Iteration" algorithm and achieves the best-known sample complexity for finding an ϵ-Nash equilibrium. For multi-player general-sum games, Liu et al. [150] provided the first sample complexity guarantees for finding Markov perfect (Nash) equilibria, correlated equilibria (CE), or coarse correlated equilibria (CCE), where CE and CCE are some other notions of equilibria which can be viewed as relaxations of Markov perfect equilibria.

It is worth noting that the algorithms that appeared in all of the aforementioned works are centralized algorithms. In contrast to those, Daskalakis et al. [63] established the sample complexity of independent policy gradient methods in zero-sum SGs. Tian et al. [212] proved sublinear regret in finite-horizon SGs, under the name of *online agnostic learning*. Sayin et al. [186] provided a decentralized Q-learning algorithm for two-player zero-sum stochastic games and showed its asymptotic convergence. Bai et al. [28] proposed a (decentralized) V-learning algorithm and proved sample complexity results for two-player zero-sum games. Concurrent works of Jin et al. [122], Song et al. [198], Mao and Başar [154] developed finite-sample convergence results (to CE and CCE) of the V-learning algorithm for multi-player general-sum stochastic games with finite horizons (in episodic setting), while in this work, we study the finite-sample convergence to Markov perfect equilibrium for multi-player general-sum stochastic games with infinite horizon, along with the use of function approximation.

3.1.2.2 Best Reply Process

Best reply (BR) processes, also known as best response dynamics (Hopkins [107]) or best response schemes (Lei and Shanbhag [137]), describe a policy update scheme for the agents, where each agent selects the policy that maximizes its payoff given other agents' joint policy (Fudenberg et al. [86], Başar and Olsder [26]). Work on BR processes mostly falls into one of two directions: The first one studies whether (under certain assumptions or no assumption) the BR processes converge to a Nash Equilibrium (NE) if NE exists (e.g., Harks and Klimm [102], Milchtaich [160]). The second one considers how fast it takes for the BR processes to converge to a NE (e.g., Even-Dar and Mansour [71], Fabrikant et al. [74], Ieong et al. [114], Syrgkanis [209], Aydın and Eksin [24]). It is well known that BR processes do not necessarily always converge to a NE, even if one exists. However, for the class of *weakly acyclic games* (Fabrikant et al. [75], Apt and Simon [19]), which includes all *potential games* as special cases, BR processes are guaranteed to converge to one of the equilibria of the game (Monderer and Shapley [162], Rosenthal [183]).

Recently, there has been a growing interest in developing different variants of the BR process that may be applied to different classes of games, e.g., the proximal BR processes (Facchinei and Pang [76], Pang et al. [170]), the Gauss–Seidel BR processes (Facchinei et al. [77]), and the BR processes with a deviator (Feldman et al. [80]). As for stochastic games, Lei et al. [139] proposed several generalizations of the proximal BR processes and showed their convergence. The asynchronous BR processes and their connections to block-coordinate descent (BCD) schemes were further investigated in Lei and Shanbhag [138]. Swenson et al. [208] and Leslie et al. [141] studied the convergence of *continuous-time* BR processes in potential games and zero-sum stochastic games, respectively. Chen et al. [53] analyzed the sample complexity of a discrete and doubly smoothed variant of the BR process with temporal-difference (TD)-learning and minimax value iteration on two-player zero-sum games.

In the standard (discrete-time) BR process, at each time when the joint policy is not an equilibrium, *one* arbitrary agent is chosen to improve its best policy given the policies of others, i.e., most of the aforementioned works consider the *asynchronous* BR process. In the weakly acyclic game that we study in this chapter, all agents may update their policies synchronously. Moreover, it is sometimes unrealistic for each agent to compute the best reply policy given all other agents' joint policies. To accommodate the synchronous policy updates and to alleviate the computation of the best reply policies, Young [229] proposed the BR process with *inertia*, which lets each agent keep its current policy if it is a best reply. If its current policy is not a best reply, the agent still keeps the current policy with a certain probability (inertia), but updates to another random policy in other cases. In Arslan and Yüksel [21] as well as in this work, an agent cannot even assess *whether* its current policy is a best reply, which is the reason why we adopt Q-learning to mimic the BR process with inertia.

3.1.2.3 Finite-sample Analysis for Q-learning

Q-learning (Watkins and Dayan [218]) has been recognized as one of the workhorses of RL. Beyond asymptotic convergence analysis, a considerable amount of prior work has studied its finite-sample performance (Even-Dar et al. [72], Beck and Srikant [33], Wainwright [217], Qu and Wierman [176], Li et al. [142]), with the sharpest results so far by Li et al. [142]. To address the setting with massive state-action spaces, Q-learning has also been blended with linear function approximation (Melo et al. [159]), with its finite-sample analysis being investigated in Chen et al. [51, 52]. It is not clear yet whether similar results can be established for decentralized Q-learning in infinite horizon multi-agent general-sum SGs, which is the focus of our work.

3.1.3 Organization

The rest of this chapter is organized as follows. In Section 3.2, we formally introduce stochastic games, weakly acyclic games, and Best Reply Process with Inertia (Young [229]). Some useful properties of the Best Reply Process with Inertia are also developed. Then, in Section 3.3, we introduce the algorithm from Arslan and Yüksel [21], called Algorithm 3.1, and provide a finite-sample analysis, through which some bounds on the convergent measures are also established. We then move to Section 3.4, and introduce the algorithm with linear function approximation (Algorithm 3.2), define the linear approximated equilibrium, and derive sample complexity results on convergence of Algorithm 3.2 (either to a linear approximated equilibrium or to a Markov perfect equilibrium, under different assumptions). Full numerical studies on the classical Grid World game are provided at the end of Section 3.3 as well as in Section 3.4. Finally, we conclude the chapter in Section 3.5, where we also highlight some of the technical difficulties in establishing our results. Proofs of two of the main results are provided in Appendix 3.6.

3.2 Preliminaries

3.2.1 Stochastic Games

A (finite) discounted stochastic game has the following ingredients (Fink [82]).

- A finite set of agents, with the i-th agent referred to as agent i for $i \in \{1, \ldots, N\} =: [N]$;
- a finite set $\mathcal{S}$ of states;
- a finite set $\mathcal{A}^i$ of actions for each agent i;
- a nonnegative deterministic reward function r^i for each agent i, which determines agent i's reward, i.e., $r^i\left(s, a^1, \ldots, a^N\right) \in [0, r^i_{\max}]$ at each state $s \in \mathcal{S}$ and for each joint action $\left(a^1, \ldots, a^N\right) \in \mathcal{A} := \mathcal{A}^1 \times \cdots \times \mathcal{A}^N$;

- a discount factor $\gamma^i \in (0,1)$ for each agent i;
- a random initial state $s_0 \in \mathcal{S}$;
- a transition kernel for the probability $P[s'|s, a^1, \ldots, a^N]$ of each state transition from $s \in \mathcal{S}$ to $s' \in \mathcal{S}$ for each joint N-tuple of actions $(a^1, \ldots, a^N) \in \mathcal{A}^1 \times \cdots \times \mathcal{A}^N$.

The dynamic evolution of the game can be described as follows. At each time $t \geq 0$, each agent i observes the state s_t, and chooses an action $a_t^i \in \mathcal{A}^i$. The agent then receives a reward $r^i(s_t, a_t) \in [0, r^i_{\max}]$ where $a_t := \left(a_t^1, \ldots, a_t^N\right)$, i.e., the reward of each agent, is determined by the state as well as the joint action selected by all agents. The system then transits to the next state s_{t+1} according to the transition kernel $P[\cdot|s_t, a^1, \ldots, a^N]$. We note that the information structure we consider here is *fully decentralized*, in the sense that each agent, when choosing its action, has access to only the current (and past) states, as well as its own history of actions and rewards, while the rewards and the state transitions are determined by the joint actions of all agents. Each agent does not have access to other agents' actions and rewards. Although the reward r^i that an agent receives depends on the state and the joint actions of all agents, the agents do *not* have full knowledge of their own reward functions, but only observe the reward they receive. In fact, an agent can be completely oblivious to the presence of other agents.

A policy for an agent is a rule of choosing an action at any time, based on the agent's history of observations. While an agent may use any function of the available information as its policy, without loss of optimality, we focus on stationary (i.e., time-invariant) policies where an agent's action at time t is based solely on the current state s_t, i.e., a stationary policy of agent i, denoted by π^i, is a mapping from the state space $\mathcal{S}$ to $\mathcal{P}\left(\mathcal{A}^i\right)$, the set of probability distributions on $\mathcal{A}^i$. The set of such stationary policies of agent i is denoted by $\Delta^i := \left\{\pi^i : \mathcal{S} \to \mathcal{P}\left(\mathcal{A}^i\right)\right\}$. The set of *deterministic* stationary policies of agent i is denoted by $\Pi^i := \left\{\pi^i : \mathcal{S} \to \mathcal{A}^i\right\} \subset \Delta^i$. We let $\Delta := \times_{i=1}^N \Delta^i$, $\Delta^{-i} := \times_{j \neq i} \Delta^j$, and $\Pi := \times_{i=1}^N \Pi^i$, $\Pi^{-i} :=$

$\times_{j\neq i}\Pi^j$. Further, we denote the joint actions and joint policies by $a := \left(a^1, \ldots, a^N\right)$ and $\pi := \left(\pi^1, \ldots, \pi^N\right)$, respectively. The joint actions and policies of all agents except agent i are denoted by $a^{-i} := \left(a^1, \ldots, a^{i-1}, a^{i+1}, \ldots, a^N\right)$ and $\pi^{-i} := \left(\pi^1, \ldots, \pi^{i-1}, \pi^{i+1}, \ldots, \pi^N\right)$, respectively. The joint actions and policies may also be written as $a = \left(a^i, a^{-i}\right)$ and $\pi = \left(\pi^i, \pi^{-i}\right)$, respectively.

The objective of each agent i is to find a policy $\pi^i \in \Delta^i$ that maximizes the total expected discounted reward, or the *value function*:

$$V_\pi^i(s) = \mathbb{E}\left[\sum_{t=0}^{\infty}\left(\gamma^i\right)^t r^i\left(s_t, a_t\right) \;\middle|\; s_0 = s\right], \quad \forall s \in \mathcal{S}, \tag{3.1}$$

where the expectation is taken over the joint distribution of a given by $\pi(s)$, as well as the random state s given by the transition kernel at each step. The *Q-function* (or *action-value function*) of agent i, $Q_\pi^i : \mathcal{S} \times \mathcal{A}^i \to \mathbb{R}$ of a joint policy π is defined by

$$Q_\pi^i(s, a^i) = \mathbb{E}\left[\sum_{t=0}^{\infty}\left(\gamma^i\right)^t r^i\left(s_t, a_t^i, a_t^{-i}\right) \;\middle|\; s_0 = s, a_0^i = a^i\right], \quad \forall s \in \mathcal{S}, \tag{3.2}$$

where the actions of i are taken according to the policy π^i except the initial action $a_0^i = a^i$, and the joint actions a^{-i} are taken according to the joint policy π^{-i}. In addition, for any $\pi^{-i} \in \Delta^{-i}$, we define the *Bellman operator* of agent i as a self-mapping of $\mathcal{S} \times \mathcal{A}^i$: $\forall (s, a^i)$,

$$\mathcal{T}_{\pi^{-i}}^i(Q^i)(s, a^i) := \mathbb{E}_{\pi^{-i}(s)}\left[r^i\left(s, a^i, a^{-i}\right) + \gamma^i \sum_{s' \in \mathcal{S}} P\left[s' \mid s, a^i, a^{-i}\right] \max_{\hat{a}^i \in \mathcal{A}^i} Q^i(s', \hat{a}^i)\right], \tag{3.3}$$

where the expectation is taken over the joint distribution of a^{-i} given by $\pi^{-i}(s)$.

We next define the *Markov perfect equilibrium* of a stochastic game (Maskin and Tirole [157]).

Definition 3.1. *A joint policy* $\pi^* = \left(\pi^{*1}, \ldots, \pi^{*N}\right) \in \Delta$ *is a (Markov perfect) equilibrium if*

$$V^i_{\left(\pi^{*i}, \pi^{*-i}\right)}(s) = \max_{\pi^i \in \Delta^i} V^i_{\left(\pi^i, \pi^{*-i}\right)}(s), \quad \forall s \in \mathcal{S},\ i \in \{1, \ldots, N\}\,.$$

We denote by Π_{eq} the set of all equilibrium joint policies. It is well known that any finite discounted stochastic game admits at least one equilibrium joint policy (Fudenberg and Tirole [85]). However, a deterministic equilibrium joint policy may not exist in general. In the following, we revisit a set of games, termed *weakly acyclic games*, for which a deterministic joint equilibrium policy always exists.

3.2.2 Weakly Acyclic Games

Definition 3.2. *A policy* $\pi^{*i} \in \Delta^i$ *is called a best reply to* $\pi^{-i} \in \Delta^{-i}$ *(for agent* i*) if*

$$V^i_{\left(\pi^{*i}, \pi^{-i}\right)}(s) = \max_{\pi^i \in \Delta^i} V^i_{\left(\pi^i, \pi^{-i}\right)}(s), \quad \forall s \in \mathcal{S}.$$

A best reply $\pi^{*i} \in \Delta^i$ *to* $\pi^{-i} \in \Delta^{-i}$ *is called a strict best reply to* $\left(\pi^i, \pi^{-i}\right)$ *if*

$$V^i_{\left(\pi^{*i}, \pi^{-i}\right)}(s) > V^i_{\left(\pi^i, \pi^{-i}\right)}(s), \quad \textit{for some } s \in \mathcal{S}.$$

We denote by $\Pi^i_{\pi^{-i}}$ the agent i's set of deterministic best replies to any $\pi^{-i} \in \Delta^{-i}$, i.e.,

$$\Pi^i_{\pi^{-i}} := \left\{\pi^{*i} \in \Pi^i :\ V^i_{\left(\pi^{*i}, \pi^{-i}\right)}(s) = \max_{\pi^i \in \Delta^i} V^i_{\left(\pi^i, \pi^{-i}\right)}(s), \quad \forall s \in \mathcal{S}\right\}.$$

Agent i's best replies to any $\pi^{-i} \in \Delta^{-i}$ can be characterized by the optimal Q-functions $Q^i_{\left(\pi^{*i}, \pi^{-i}\right)}$. With some abuse of notation, we simply write $Q^i_{\pi^{-i}}$ in place of $Q^i_{\left(\pi^{*i}, \pi^{-i}\right)}$. This

optimal Q-function satisfies the fixed point equation of the Bellman operator: $\forall \left(s, a^i\right)$,

$$Q^i_{\pi^{-i}}(s, a^i) = \mathbb{E}_{\pi^{-i}(s)} \left[r^i \left(s, a^i, a^{-i}\right) + \gamma^i \sum_{s' \in \mathcal{S}} P \left[s' \mid s, a^i, a^{-i} \right] \max_{\hat{a}^i \in \mathcal{A}^i} Q^i_{\pi^{-i}}(s', \hat{a}^i) \right]. \quad (3.4)$$

The optimal Q-function $Q^i_{\pi^{-i}}(s, a^i)$ is agent i's expected discounted value-to-go from the initial state s, assuming that i initially chooses action a^i and uses an optimal policy thereafter while all other agents use the joint policy π^{-i}. Then, we can rewrite agent i's set of deterministic best replies to π^{-i} as

$$\Pi^i_{\pi^{-i}} = \left\{ \pi^{*i} \in \Pi^i : \; Q^i_{\pi^{-i}} \left(s, \pi^{*i}(s)\right) = \max_{a^i \in \mathcal{A}^i} Q^i_{\pi^{-i}}(s, a^i), \quad \forall s \in \mathcal{S} \right\}. \quad (3.5)$$

From (3.5) and Definition 3.1, we have that a deterministic joint policy $\pi^* \in \Pi_{\text{eq}}$ if $\pi^{*i} \in \Pi^i_{(\pi^*)^{-i}}$ for all $i \in [N]$.

We next define the *best reply graph* on the set of deterministic joint policies Π. Specifically, each node (vertex) in the graph is a deterministic joint policy $\pi \in \Pi$, and there is a directed edge from π_k to π_l if for some $i \in [N]$, $\pi^i_l \neq \pi^i_k$, $\pi^j_l = \pi^j_k, \forall j \neq i$, and $\pi^i_l \in \Pi^i_{\pi^{-i}_k}$. When there is a directed edge from π_k to π_l, we also say that π_l is an *out-neighbor* of π_k in the strict best reply graph. We then define the *strict best reply path* and the *weakly acyclic game*.

Definition 3.3. *A sequence of deterministic joint policies $\pi_0, \pi_1, \ldots$ is called a strict best reply path if for each k, π_k and π_{k+1} differ for exactly one agent, say agent i, and π^i_{k+1} is a strict best reply with respect to π_k.*

Definition 3.4. *A discounted stochastic game is called weakly acyclic under strict best replies if there is a strict best reply path starting from each deterministic joint policy and ending at a deterministic equilibrium policy.*

Figure 3.1 shows the strict best reply graph of a weakly acyclic game. A node with no outgoing edges is an equilibrium policy (π_5 and π_6 in this graph). The game illustrated in

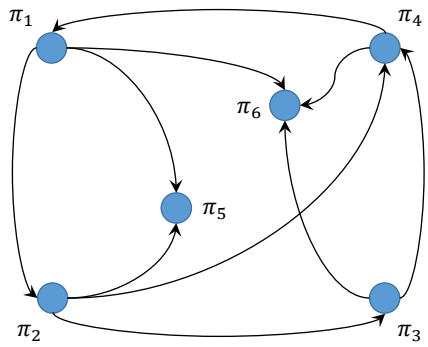

Figure 3.1: The strict best reply graph of a stochastic game.

Figure 3.1 is weakly acyclic under strict best replies since there is a path from every node to an equilibrium. Also note that a weakly acyclic game may have cycles in its strict best reply graph, for example, $\pi_1 \to \pi_2 \to \pi_3 \to \pi_4 \to \pi_1$ in Figure 3.1.

If the game has no cycles in its strict best reply graph, we may consider the process of letting only one agent switch to one of its best replies at each step, and such a process will continue until no agent has strict best replies, at which time the joint policy of all agents is a deterministic equilibrium joint policy. However, as described above, the strict best reply graph of a weakly acyclic game may contain cycles. We next introduce the Best Reply Process with Inertia (Young [229]) as Algorithm 3.0, which assigns to each agent a strict positive probability of choosing each of its strict best replies.

Algorithm 3.0 Best Reply Process with Inertia (for agent i)

Set parameters
$\lambda^i \in (0,1)$: inertia

Initialize $\pi_0^i \in \Pi^i$ (arbitrary)
Iterate $k \geq 0$ **do**
  if $\pi_k^i \in \Pi^i_{\pi_k^{-i}}$ **then**
    $\pi_{k+1}^i = \pi_k^i$
  else
    $$\pi_{k+1}^i = \begin{cases} \pi_k^i & \text{w.p. } \lambda^i \\ \text{any } \pi^i \in \Pi^i_{\pi_k^{-i}} & \text{w.p. } (1-\lambda^i)/|\Pi^i_{\pi_k^{-i}}| \end{cases}$$
  end if
end

Under the Best Reply Process with Inertia (BRPI), if the joint policy π_k is an equilibrium policy at step k, the policy will never change in the following steps; otherwise, the joint policy at step $k+1$, denoted by π_{k+1}, can be any joint policy that is an out-neighbor of π_k in the strict best reply graph with strict positive probability. For each $\pi \in \Pi$, there exists a strict best reply path of minimum length L_π, and letting $L := \max_{\pi\in\Pi} L_\pi$ be the maximum length of the shortest strict best reply path from any policy to an equilibrium policy. Then, starting from an arbitrary joint policy π_0 and letting all agents update their policies following the BRPI, the joint policy π_L in L steps later will be an equilibrium policy with some positive probability $p_{\min}$, i.e., $p_{\min} := \min_\pi P\left[\pi_L \in \Pi_{\text{eq}} \mid \pi_0 = \pi\right]$. We then have the following lemma, which provides a lower bound on $p_{\min}$.

Lemma 3.1. *Let all agents update their policies by the BRPI. We have that*

$$p_{\min} \geq \left(\min_{j\in\{1,\dots,N\}} \left\{ \frac{1-\lambda^j}{\left|\Pi^j\right|} \cdot \prod_{i\neq j} \lambda^i \right\} \right)^L =: \hat{p}. \tag{3.6}$$

Proof of Lemma 3.1. Let $\pi_0 = \hat{\pi}_0 \in \Pi$ be an arbitrary initial joint policy. If $\hat{\pi}_0 \in \Pi_{\text{eq}}$, then it holds that $P\left[\pi_L = \hat{\pi}_0 \in \Pi_{\text{eq}} \mid \pi_0 = \hat{\pi}_0 \in \Pi_{\text{eq}}\right] = 1$; otherwise, let l be the length of the shortest strict best reply path from $\hat{\pi}_0$ to an equilibrium policy, where $l \leq L$. Let the sequence of policies along the path be $\hat{\pi}_0, \hat{\pi}_1, \dots, \hat{\pi}_l$, with $\hat{\pi}_l \in \Pi_{\text{eq}}$. Further, let $i_1, \dots, i_l$ be the agent that changes its policy at each update, i.e., $\hat{\pi}_{n-1}$ and $\hat{\pi}_n$ differ only at agent i_n, for all $n = 1, \dots, l$. Then, by the policy updates in the BRPI, we have that

$$\begin{aligned}
&P\left[\pi_L \in \Pi_{\text{eq}} \mid \pi_0 = \hat{\pi}_0\right] \\
&\geq P\left[\pi_1 = \hat{\pi}_1, \dots, \pi_l = \hat{\pi}_l, \pi_{l+1} = \hat{\pi}_l, \dots, \pi_L = \hat{\pi}_l \mid \pi_0 = \hat{\pi}_0\right] \\
&= P\left[\pi_1 = \hat{\pi}_1 \mid \pi_0 = \hat{\pi}_0\right] P\left[\pi_2 = \hat{\pi}_2 \mid \pi_0 = \hat{\pi}_0, \pi_1 = \hat{\pi}_1\right] \cdot \cdots \\
&\qquad \cdot P\left[\pi_l = \hat{\pi}_l \mid \pi_0 = \hat{\pi}_0, \dots \pi_{l-1} = \hat{\pi}_{l-1}\right] \cdot P\left[\pi_{l+1} = \cdots = \pi_L = \hat{\pi}_l \mid \pi_0 = \hat{\pi}_0, \dots \pi_l = \hat{\pi}_l\right]
\end{aligned}$$

$$
\begin{aligned}
&\geq \left(\frac{1-\lambda^{i_1}}{\left|\Pi^{i_1}\right|}\cdot\prod_{i\neq i_1}\lambda^i\right)\left(\frac{1-\lambda^{i_2}}{\left|\Pi^{i_2}\right|}\cdot\prod_{i\neq i_2}\lambda^i\right)\cdots\left(\frac{1-\lambda^{i_l}}{\left|\Pi^{i_l}\right|}\cdot\prod_{i\neq i_l}\lambda^i\right)\cdot 1\\
&=\prod_{j\in\{i_1,\ldots,i_l\}}\left(\frac{1-\lambda^{j}}{\left|\Pi^{j}\right|}\cdot\prod_{i\neq j}\lambda^i\right)\geq\left(\min_{j\in\{1,\ldots,N\}}\left\{\frac{1-\lambda^{j}}{\left|\Pi^{j}\right|}\cdot\prod_{i\neq j}\lambda^i\right\}\right)^{l}\\
&\geq\left(\min_{j\in\{1,\ldots,N\}}\left\{\frac{1-\lambda^{j}}{\left|\Pi^{j}\right|}\cdot\prod_{i\neq j}\lambda^i\right\}\right)^{L}.
\end{aligned}
$$

Note that since the above holds for any arbitrary initial joint policy $\hat{\pi}_0$, we conclude that it is a lower bound for $p_{\min}$. □

This implies that the BRPI will reach an equilibrium policy in a finite number of steps w.p. 1. We further have the following result.

Proposition 3.1. *Let all agents update their deterministic policies according to the BRPI. We have that*

$$
P\left[\pi_k\in\Pi_{\text{eq}}\right]\geq 1-\delta,
$$

provided that

$$
k\geq\frac{L\cdot\log\delta}{\log\left(1-\left(\min_{j\in\{1,\ldots,N\}}\left\{\frac{1-\lambda^{j}}{\left|\Pi^{j}\right|}\cdot\prod_{i\neq j}\lambda^i\right\}\right)^{L}\right)}+L.
$$

Proof of Proposition 3.1. For any initial joint policy $\pi_0=\hat{\pi}_0$, we have that

$$
P\left[\pi_L\in\Pi_{\text{eq}}\mid\pi_0=\hat{\pi}_0\right]\geq p_{\min}.
$$

We first show that

$$P\left[\pi_{nL} \in \Pi_{\text{eq}}\right] \geq p_{\min}\left[1 + (1-p_{\min}) + (1-p_{\min})^2 + \cdots + (1-p_{\min})^{n-1}\right]$$
$$= 1 - (1-p_{\min})^n. \tag{3.7}$$

We show (3.7) by induction on n. The base case holds since $P\left[\pi_L \in \Pi_{\text{eq}}\right] \geq p_{\min}$. As for the induction step, assuming that (3.7) holds for $\pi_{(n-1)L}$, we have that

$$\begin{aligned}
&P\left[\pi_{nL} \in \Pi_{\text{eq}}\right] \\
&= P\left[\pi_{nL} \in \Pi_{\text{eq}} \mid \pi_{(n-1)L} \in \Pi_{\text{eq}}\right] P\left[\pi_{(n-1)L} \in \Pi_{\text{eq}}\right] \\
&\quad + P\left[\pi_{nL} \in \Pi_{\text{eq}} \mid \pi_{(n-1)L} \notin \Pi_{\text{eq}}\right] P\left[\pi_{(n-1)L} \notin \Pi_{\text{eq}}\right] \\
&\geq P\left[\pi_{(n-1)L} \in \Pi_{\text{eq}}\right] + p_{\min} P\left[\pi_{(n-1)L} \notin \Pi_{\text{eq}}\right] \\
&= P\left[\pi_{(n-1)L} \in \Pi_{\text{eq}}\right] + p_{\min}\left(1 - P\left[\pi_{(n-1)L} \in \Pi_{\text{eq}}\right]\right) \\
&= (1-p_{\min}) P\left[\pi_{(n-1)L} \in \Pi_{\text{eq}}\right] + p_{\min} \geq (1-p_{\min})\left(1 - (1-p_{\min})^{n-1}\right) + p_{\min} \\
&= 1 - (1-p_{\min})^n,
\end{aligned}$$

which completes the induction step. Therefore, if k satisfies

$$k \geq \frac{L \log \delta}{\log(1-p_{\min})} + L, \tag{3.8}$$

then, we have that

$$P\left[\pi_k \in \Pi_{\text{eq}}\right] \geq P\left[\pi_{\left\lfloor \frac{k}{L} \right\rfloor L} \in \Pi_{\text{eq}}\right] \geq 1 - (1-p_{\min})^{\left\lfloor \frac{k}{L} \right\rfloor} \geq 1 - (1-p_{\min})^{\left\lfloor \frac{\log \delta}{\log(1-p_{\min})} + 1 \right\rfloor}$$
$$\geq 1 - (1-p_{\min})^{\frac{\log \delta}{\log(1-p_{\min})}} = 1 - (1-p_{\min})^{\log_{1-p_{\min}} \delta} = 1 - \delta.$$

The proof is completed by taking the lower bound of $p_{\min}$ from Lemma 3.1 to (3.8). □

We note that in applying the BRPI, each agent i needs to construct $\Pi^i_{\pi_k^{-i}}$ at step k, which can be done according to (3.5) by first computing $Q^i_{\pi_k^{-i}}$ by solving the fixed point equation (3.4). However, since we assume that agents do not have access to the state transition probabilities P, neither do they know the joint policy π_k^{-i} of other agents, they would not be able to compute (3.4) directly. In the next section, we introduce and analyze the sample complexity of the Q-learning algorithm for stochastic games, where agents would be able to approximate their best replies and adjust their policies accordingly.

3.3 Decentralized Q-learning in Tabular Setting

Recall that in the decentralized setting, at any time t, each agent has access to the history of state realizations up to time t, its own set of actions $\mathcal{A}^i$ and discount factor γ^i, as well as its own history of actions. Since each agent is completely oblivious to the existence of other agents, agent i may view the decision making problem as a stationary Markov decision process, and use the standard Q-learning algorithm:

$$Q^i_{t+1}(s_t, a^i_t) = (1-\eta^i_t)Q^i_t(s_t, a^i_t) + \eta^i_t \left[r^i(s_t, a^i_t, a^{-i}_t) + \gamma^i \max_{a^i \in \mathcal{A}^i} Q^i_t(s_{t+1}, a^i) \right], \tag{3.9a}$$

$$Q^i_{t+1}(s, a^i) = Q_t(s, a^i), \quad \forall (s, a^i) \neq (s_t, a^i_t), \tag{3.9b}$$

where η^i_t is agent i's step size at time t. A common approach for agent i to select its actions is the so-called ϵ-greedy method, i.e., by exploiting the learned Q-functions with high probability and randomly exploring any action with some small probability. If agent i uses Q-learning (3.9) with the ϵ-greedy method while all other agents use a fixed joint policy π^{-i}, then, agent i solves a stationary MDP and $P\left[Q^i_t \to Q^i_{\pi^{-i}}\right] = 1$ following the convergence result of Q-learning on stationary MDPs (Tsitsiklis [214]). However, when all agents use Q-learning (3.9) with the ϵ-greedy method, then the MDP becomes nonstationary and convergence of the Q functions is not guaranteed (Leslie and Collins [140]). To overcome

this difficulty, Arslan and Yüksel [21] proposed a fully decentralized Q-learning algorithm where all agents use constant policies for extended periods of time, termed *exploration phases*. The kth exploration phase runs through time $t = t_k, \ldots, t_{k+1} - 1$, where $t_{k+1} = t_k + T_k$ for some positive integer T_k (with $t_0 = 0$). During the kth exploration phase, each agent i has some deterministic *baseline policy* π_k^i, but uses the same randomized policy $\bar{\pi}_k^i$ throughout the phase, where

$$\bar{\pi}_k^i(s_t) := \begin{cases} \pi_k^i(s_t), & \text{w.p. } 1 - \rho^i \\ \text{any } a^i \in \mathcal{A}^i, & \text{w.p. } \rho^i / |\mathcal{A}^i|, \end{cases}$$

for some $\rho^i \in (0, 1)$. Equivalently, we can write

$$\bar{\pi}_k^i = (1 - \rho^i)\pi_k^i + \rho^i \nu^i, \tag{3.10}$$

where ν^i is the random policy that assigns the uniform distribution on $\mathcal{A}^i$ to each s. In words, agent i plays the baseline policy with probability $1 - \rho^i$, and plays all actions uniformly with probability $\rho^i / |\mathcal{A}^i|$. We denote by $\bar{\Pi}$ the set of joint policies in the form of (3.10) for each agent, i.e., $\bar{\Pi} := \left\{ \bar{\pi} \mid \bar{\pi}^i = (1 - \rho^i)\pi^i + \rho^i \nu^i, \pi^i \in \Pi^i, \forall i \in [N] \right\}$. Each agent updates its Q function after each step according to (3.9), but updates its baseline policy only at the end of every exploration phase, by using the BRPI with some estimated $\Pi^i_{\pi_k^{-i}}$. The complete algorithm is presented as Algorithm 3.1.

Arslan and Yüksel [21] proved that the joint policy π_k obtained from Algorithm 3.1 *asymptotically* converges to some equilibrium policy. We will show in this chapter the non-asymptotic convergence guarantees of the algorithm. To proceed, we first impose the following two assumptions.

Assumption 3.1. *There exist some $\kappa > 0$, and a finite integer $H \geq 1$, such that for any*

Algorithm 3.1 Q-learning for agent i

Set parameters

$\mathbb{Q}^i$: some compact subset of the Euclidian space $\mathbb{R}^{|\mathcal{S}\times\mathcal{A}^i|}$

$\{T_k\}_{k\geq 0}$: sequence of integers in $[1,\infty)$, the length of the kth exploration phase

$K \in \mathbb{Z}_+$: number of exploration phases

$\rho^i \in (0,1)$: experimentation probability

$\lambda^i \in (0,1)$: inertia

$\zeta^i \in (0,\infty)$: tolerance level for sub-optimality

$\{\eta_t^i\}_{t\geq 0}$: sequence of step sizes

Initialize $\pi_0^i \in \Pi^i$ (arbitrary), $Q_0^i \in \mathbb{Q}^i$ (arbitrary)

Receive s_0

for $k = 1, 2 \ldots$ **do**

  for $t = t_k, \ldots, t_{k+1} - 1$ **do**

    $a_t^i = \bar{\pi}_k^i(s_t) := \begin{cases} \pi_k^i(s_t), & \text{w.p. } 1-\rho^i \\ \text{any } a^i \in \mathcal{A}^i, & \text{w.p. } \rho^i/|\mathcal{A}^i| \end{cases}$

    Receive $r^i(s_t, a_t^i, a_t^{-i})$

    Receive s_{t+1} (selected according to $P[\,\cdot \mid s_t, a_t^i, a_t^{-i}]$)

    $Q_{t+1}^i(s_t, a_t^i) = (1-\eta_t^i) Q_t^i(s_t, a_t^i) + \eta_t^i \left[r^i(s_t, a_t^i, a_t^{-i}) + \gamma^i \max_{a^i \in \mathcal{A}^i} Q_t^i(s_{t+1}, a^i) \right]$

    $Q_{t+1}^i(s, a^i) = Q_t^i(s, a^i)$, for all $(s, a^i) \neq (s_t, a_t^i)$

  end for

  $\Pi_{k+1}^i = \left\{ \hat{\pi}^i \in \Pi^i : Q_{t_{k+1}}^i(s, \hat{\pi}^i(s)) \geq \max_{a^i \in \mathcal{A}^i} Q_{t_{k+1}}^i(s, a^i) - \frac{1}{2}\zeta^i, \text{ for all } s \right\}$

  if $\pi_k^i \in \Pi_{k+1}^i$ **then**

    $\pi_{k+1}^i = \pi_k^i$

  else

    $\pi_{k+1}^i = \begin{cases} \pi_k^i, & \text{w.p. } \lambda^i \\ \text{any } \pi^i \in \Pi_{k+1}^i, & \text{w.p. } (1-\lambda^i)/|\Pi_{k+1}^i| \end{cases}$

  end if

  $Q_{t_{k+1}}^i \leftarrow$ projection of $Q_{t_{k+1}}^i$ onto $\mathbb{Q}^i$

end for

pair of states (s', s), there exists a sequence of joint actions $\tilde{a}_0, \ldots, \tilde{a}_{H-1} \in \mathcal{A}$ such that

$$P[s_H = s' \mid (s_0, a_0, \ldots, a_{H-1}) = (s, \tilde{a}_0, \ldots, \tilde{a}_{H-1})] \geq \kappa.$$

Recall from the definition of $\bar{\pi}_k$ that each agent has positive probability of choosing any action $a^i \in \mathcal{A}^i$, which implies that the joint actions taken by all agents can be any $a \in \mathcal{A}$ with positive probability. This, together with Assumption 3.1, implies that all states

communicate with each other in the Markov chain induced by the joint policy $\bar{\pi}_k$, i.e., the Markov chain is irreducible. This is the same assumption as made in Arslan and Yüksel [21] except that we denote by κ the lower bound on the probabilities.

Assumption 3.2. *For any joint policy $\bar{\pi}_k$, the induced Markov chain is aperiodic.*

It is common to assume that the Markov chain induced by the behavior policy is ergodic in analyzing the sample complexity of single-agent Q-learing (Li et al. [142]). Assumption 3.2, together with Assumption 3.1, ensures that the Markov chain is finite, irreducible, and aperiodic, which implies that the chain is uniformly ergodic (Paulin et al. [172]) and admits a unique stationary distribution.

Let $\mu_{\bar{\pi}_k}$ be the stationary distribution over all states of the Markov chain induced by $\bar{\pi}_k$, and let $\mu^i_{\bar{\pi}_k}$ be the stationary distribution over all $(s, a^i) \in \mathcal{S} \times \mathcal{A}^i$ pairs. We further define

$$\mu_{\min,k} := \min_{i\in[N]} \min_{(s,a^i)\in\mathcal{S}\times\mathcal{A}^i} \mu^i_{\bar{\pi}_k}\left(s, a^i\right). \tag{3.11}$$

Here, $\min_{(s,a^i)\in\mathcal{S}\times\mathcal{A}^i} \mu^i_{\bar{\pi}_k}\left(s, a^i\right) := \mu^i_{\min,k}$ is the minimum probability of the stationary distribution over all state-action pairs from the perspective of agent i, and $\mu_{\min,k}$ is obtained by taking the minimum over all agents. Intuitively, the smaller $\mu_{\min,k}$ is, the more samples are needed to ensure that all state-action pairs (from the perspective of each agent) are visited sufficiently many times during the kth exploration phase. Moreover, we define the mixing time of agent i at the kth exploration phase as:

$$t^i_{\text{mix},k}(\alpha) := \min\left\{t \;\middle|\; \max_{(s_0,a^i_0)\in\mathcal{S}\times\mathcal{A}^i} d_{\text{TV}}\left(P^t(\cdot \mid s_0, a^i_0), \mu^i_{\bar{\pi}_k}\right) \le \alpha\right\}, \tag{3.12}$$

where $\alpha \in (0,1)$, $P^t(\cdot \mid s_0, a^i_0)$ is the distribution of (s_t, a^i_t) conditioned on the initial state-action pair (s_0, a^i_0), and d_{TV} measures the total variation between two distributions. Intuitively, $t^i_{\text{mix},k}$ describes how fast sample trajectory of the Markov chain converges to

the stationary distribution of state-action pairs from the perspective of agent i. Further, let $t_{\text{mix},k}(\alpha) := \max_{i\in[N]} t^i_{\text{mix},k}(\alpha)$. Note that the convergence rate of a uniformly ergodic Markov chain to its stationary distribution is exponential (Häggström et al. [99]). We therefore do not expect $t_{\text{mix},k}$ to be excessively large.

We next define the minimum separation between the agents' optimal Q-functions (with respect to deterministic policies), which is regarded as an upper bound of the tolerance level ζ^i for all agents.

$$\bar{\zeta} := \min_{\substack{i,s,a^i,\tilde{a}^i,\pi^{-i}\in\Pi^{-i}:\\ Q^i_{\pi^{-i}}(s,a^i)\neq Q^i_{\pi^{-i}}(s,\tilde{a}^i)}} \left| Q^i_{\pi^{-i}}(s,a^i) - Q^i_{\pi^{-i}}(s,\tilde{a}^i) \right|. \tag{3.13}$$

For notational convenience, we let $A := \max_{i\in[N]} |\mathcal{A}^i|$, $\bar{\gamma} := \max_{i\in[N]} \gamma^i$, and $\underline{\gamma} = \min_{i\in[N]} \gamma^i$.We now present our main theorem on the sample complexity of Algorithm 3.1.

Theorem 3.1. *Consider a discounted stochastic game that is weakly acyclic under strict best replies* (3.5)*. Suppose that each agent updates its policies by Algorithm 3.1. Let Assumptions 3.1 and 3.2 hold. Then, there exist some constants c_0 and c_1 such that, for any $0 < \delta < 1$, one has that for all $k \geq K$,*

$$P\left[\pi_k \in \Pi_{\text{eq}}\right] \geq 1 - \delta,$$

provided that for all $i \in [N]$ and $k \in [K]$,

$$T_k \geq \frac{c_0}{\mu_{\min,k}} \left\{ \frac{1}{(1-\bar{\gamma})^5\epsilon^2} + \frac{t_{\text{mix},k}\left(\frac{1}{4}\right)}{1-\bar{\gamma}} \right\} \log\left(\frac{NL|\mathcal{S}|AT_k}{\tilde{\delta}}\right) \log\left(\frac{1}{(1-\bar{\gamma})^2\epsilon}\right), \tag{3.14a}$$

$$K \geq \frac{\left[\left(1-\tilde{\delta}\right)^2 \hat{p} - \tilde{\delta}^2\right] L}{\left[\tilde{\delta} + \left(1-\tilde{\delta}\right)\hat{p}\right]^2 \tilde{\delta}}, \tag{3.14b}$$

$$\eta_t^i = \frac{c_1}{\log\left(\frac{NL|\mathcal{S}||\mathcal{A}^i|T_k}{\tilde{\delta}}\right)} \min\left\{\frac{(1-\bar{\gamma})^4\epsilon^2}{\bar{\gamma}^2}, \frac{1}{t_{\text{mix},k}\left(\frac{1}{4}\right)}\right\}, \quad \forall t = t_k, \ldots, t_{k+1}-1, \tag{3.14c}$$

$$\rho^i = 1 - \left(1 - \frac{(\bar{\zeta}/8 - \epsilon)(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}} := \rho, \tag{3.14d}$$

$$\zeta^i = \frac{\bar{\zeta}}{2}, \tag{3.14e}$$

where $\bar{\zeta}$ *and* $\hat{p}$ *are as defined in* (3.13) *and* (3.6)*, respectively, and* Γ *is some absolute constant (formally defined in* (3.22)*) which depends only on the game parameters,* $\epsilon := \min\left\{\frac{\bar{\zeta}}{16}, \frac{1}{2(1-\underline{\gamma})}\right\}$*, and* $\tilde{\delta}$ *is a unique element in* $(0, \delta)$ *such that*

$$\delta = 1 - \left(\frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta} + \left(1-\tilde{\delta}\right)\hat{p}} - \tilde{\delta}\right)\left(1-\tilde{\delta}\right).$$

Theorem 3.1 provides a finite-sample result for Algorithm 3.1. To be more explicit on the results, we further obtain the following bounds for $t_{\text{mix},k}(\alpha)$ and $\mu_{\min,k}$.

Proposition 3.2. *For all* $k \in [K]$ *and* $i \in [N]$*, we have that*

$$\mu_{\min,k}^i \leq \left[1 - (|\mathcal{S}|-1)\,\kappa\left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H\right] \cdot \frac{\rho^i}{|\mathcal{A}^i|}, \tag{3.15a}$$

$$\mu_{\min,k}^i \geq \kappa \frac{\rho^i}{|\mathcal{A}^i|}\left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H, \tag{3.15b}$$

$$t_{\text{mix},k}(\alpha) \leq (H+1)\left(\frac{\log\alpha}{\log\left[1 - |\mathcal{S}|\kappa\left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H \cdot \min_{i\in[N]} \rho^i\right]} + 1\right). \tag{3.15c}$$

With $\rho^i = \rho$ for all $i \in [N]$ as in Theorem 3.1, we deduce that

$$\mu_{\min,k} \leq \left[1 - (|\mathcal{S}| - 1)\,\kappa \frac{\rho^{NH}}{A^{NH}}\right] \cdot \frac{\rho}{A}, \tag{3.16a}$$

$$\mu_{\min,k} \geq \kappa \frac{\rho^{NH+1}}{A^{NH+1}}, \tag{3.16b}$$

$$t_{\text{mix},k}(\alpha) \leq (H+1)\left((-\log\alpha)\frac{A^{NH}}{\kappa\rho^{NH}} \cdot \min\left\{1, |\mathcal{S}|^{-1}\rho^{-1}\right\} + 1\right). \tag{3.16c}$$

Corollary 3.1. *By applying* (3.16b) *and* (3.16c) *to Theorem 3.1, we may express the sample complexity of each exploration phase* (3.14a) *as*

$$T_k \geq \frac{c_0 A^{NH+1}}{\kappa\rho^{NH+1}} \left\{ \frac{1}{(1-\bar{\gamma})^5\epsilon^2} + \frac{(H+1)\left((\log 4)\frac{A^{NH}}{\kappa\rho^{NH}} \cdot \min\left\{1, |\mathcal{S}|^{-1}\rho^{-1}\right\} + 1\right)}{1-\bar{\gamma}} \right\}$$
$$\cdot \log\left(\frac{NL|\mathcal{S}|AT_k}{\tilde{\delta}}\right) \log\left(\frac{1}{(1-\bar{\gamma})^2\hat{\epsilon}_k}\right). \tag{3.17}$$

Note that the joint action space of all agents has size $\mathcal{O}(A^N)$. In Proposition 3.2, we have eliminated the dependence on $\mu_{\min,k}$ and $t_{\text{mix,k}}$, so that the sample complexity of T_k is explicitly represented by the parameters of the game. In the following two subsections, we provide complete analyses and proofs for Theorem 3.1 and Proposition 3.2.

3.3.1 Proof of Theorem 3.1.

We first introduce the following lemma, which is an application of the sample complexity result on single agent Q-learning (Li et al. [142]).

Lemma 3.2. *Fix any arbitrary $\pi_k \in \Pi$. For any $0 < \hat{\delta} < 1$ and $0 < \epsilon \leq \frac{1}{1-\underline{\gamma}}$, there exist some constants $c_{0,k}$ and $c_{1,k}^1, \ldots, c_{1,k}^N$ such that*

$$P\left[\left|Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}}\right|_\infty \leq \epsilon,\ \forall i \in [N]\right] \geq 1 - \hat{\delta},$$

provided that the iteration number T_k and the learning rates η_t^i obey

$$T_k \geq \frac{c_{0,k}}{\mu_{\min,k}} \left\{ \frac{1}{(1-\bar{\gamma})^5 \epsilon^2} + \frac{t_{\text{mix},k}\left(\frac{1}{4}\right)}{1-\bar{\gamma}} \right\} \log\left(\frac{N|\mathcal{S}|AT_k}{\hat{\delta}}\right) \log\left(\frac{1}{(1-\bar{\gamma})^2\epsilon}\right), \tag{3.18a}$$

$$\eta_t^i = \frac{c_{1,k}^i}{\log\left(\frac{N|\mathcal{S}||\mathcal{A}^i|T_k}{\hat{\delta}}\right)} \min\left\{ \frac{(1-\gamma^i)^4\epsilon^2}{\left(\gamma^i\right)^2}, \frac{1}{t_{\text{mix},k}^i\left(\frac{1}{4}\right)} \right\}, \quad \forall t = t_k, \ldots, t_{k+1}-1,\ i \in [N], \tag{3.18b}$$

where $\bar{\gamma} = \max_i \gamma^i$ and $A = \max_i |\mathcal{A}^i|$.

Proof of Lemma 3.2. Note that in the kth exploration phase, agents adopt the joint policy $\bar{\pi}_k$ as defined in (3.10). Also by Assumptions 3.1 and 3.2, the (finite) Markov chain induced by the joint policy is irreducible and aperiodic. Theorem 1 of Li et al. [142] implies that for any agent i, there exist some constants $c_{0,k}^i$ and $c_{1,k}^i$ such that for any $0 < \delta_0 < 1$ and $0 < \epsilon \leq \frac{1}{1-\gamma^i}$,

$$P\left[\left|Q_{t_{k+1}}^i - Q_{\bar{\pi}_k^{-i}}^i\right|_\infty \leq \epsilon\right] \geq 1 - \delta_0$$

provided that the iteration number T_k and the learning rates η_t^i obey

$$T_k \geq \frac{c_{0,k}^i}{\mu_{\min,k}^i} \left\{ \frac{1}{(1-\gamma^i)^5 \epsilon^2} + \frac{t_{\text{mix},k}^i\left(\frac{1}{4}\right)}{1-\gamma^i} \right\} \log\left(\frac{|\mathcal{S}||\mathcal{A}^i|T_k}{\delta_0}\right) \log\left(\frac{1}{(1-\gamma^i)^2\epsilon}\right),$$

$$\eta_t^i = \frac{c_{1,k}^i}{\log\left(\frac{|\mathcal{S}||\mathcal{A}^i|T_k}{\delta_0}\right)} \min\left\{ \frac{(1-\gamma^i)^4\epsilon^2}{\left(\gamma^i\right)^2}, \frac{1}{t_{\text{mix},k}^i\left(\frac{1}{4}\right)} \right\}, \quad \forall t = t_k, \ldots, t_{k+1}-1,$$

where $\mu_{\min,k}^i := \min_{(s,a^i)\in\mathcal{S}\times\mathcal{A}^i} \mu_{\pi_k}^i\left(s, a^i\right)$, and $\mu_{\min,k}$ and $t_{\text{mix},k}^i$ are as defined in (3.11)

and (3.12), respectively. Let $c_{0,k} := \max_{i\in[N]} c^i_{0,k}$. Then, with

$$T_k \geq \frac{c_{0,k}}{\mu_{\min,k}} \left\{ \frac{1}{(1-\bar{\gamma})^5\epsilon^2} + \frac{t_{\text{mix},k}\left(\frac{1}{4}\right)}{1-\bar{\gamma}} \right\} \log\left(\frac{|\mathcal{S}|AT_k}{\delta_0}\right) \log\left(\frac{1}{(1-\bar{\gamma})^2\epsilon}\right),$$

$$\eta^i_t = \frac{c^i_{1,k}}{\log\left(\frac{|\mathcal{S}||\mathcal{A}^i|T_k}{\delta_0}\right)} \min\left\{ \frac{(1-\gamma^i)^4\epsilon^2}{\left(\gamma^i\right)^2}, \frac{1}{t^i_{\text{mix},k}\left(\frac{1}{4}\right)} \right\}, \quad \forall t = t_k, \ldots, t_{k+1}-1,\ i \in [N],$$

we have that

$$P\left[\left| Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty \leq \epsilon \right] \geq 1-\delta_0, \quad \forall i \in [N],$$

which implies that

$$P\left[\left| Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty > \epsilon \right] \leq \delta_0, \quad \forall i \in [N].$$

From the union bound,

$$P\left[\left| Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty > \epsilon,\ \exists i \in [N] \right] \leq \sum_{i\in[N]} P\left[\left| Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty > \epsilon \right] \leq N\delta_0.$$

Therefore,

$$P\left[\left| Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty \leq \epsilon,\ \forall i \in [N] \right] = 1 - P\left[\left| Q^i_{t_{k+1}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty > \epsilon,\ \exists i \in [N] \right] \geq 1 - N\delta_0.$$

The proof is completed by taking $\hat{\delta} = N\delta_0$. □

Lemma 3.2 bounds the approximation error of Q-learning for each agent, i.e., the difference of the Q-function obtained at the end of the kth exploration phase and the optimal Q-function in the best reply to $\bar{\pi}^{-i}$. Our next goal is to bound the approximation error of policy perturbation. Recall the definition of the randomized policy in (3.10), and consider

the joint policies of all agents except i. With probability $\prod_{j\neq i}(1-\rho^j)$, all agents $j \neq i$ end up playing their baseline policies, which results in $\left|Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}}\right| = 0$, i.e. the approximation error of policy perturbation becomes zero in this case. When not all agents play their baseline policies, let $\varphi^{-i} \in \Delta^{-i}$ be some convex combination of the policies in Δ^{-i} of the form where each agent $j \neq i$ either uses a baseline policy $\pi^j \in \Pi^j$ or the uniform distribution. More precisely, let J denote the subset of agents choosing the baseline policies, and let

$$\varphi^{-i} = \sum_{J\subset\{1,\dots,N\}\setminus\{i\}} a_J \varphi_J^{-i}, \tag{3.21}$$

where $a_J := \frac{\prod_{j\in J}(1-\rho^j)\prod_{j\notin J\cup\{i\}}\rho^j}{1-\prod_{j\neq i}(1-\rho^j)}$ and $\varphi_J \in \Delta^{-i}$ is such that $\varphi_J^j = \pi^j$ for $j \in J$ and $\varphi_J^j = \nu^j$ for $j \notin J \cup \{i\}$. Denote by $\bar{\Delta}^{-i} \subset \Delta^{-i}$ the set of all policies in the form of (3.21). Note that $\bar{\Delta}^{-i}$ is a finite set. Recall the definition of the Bellman operator from (3.3). We then define

$$\Gamma := \max_{(\pi^{-i},\varphi^{-i})\in\Pi^{-i}\times\bar{\Delta}^{-i}} \left|\mathcal{T}^i_{\pi^{-i}}(Q^i_{\pi^{-i}}) - \mathcal{T}^i_{\varphi^{-i}}(Q^i_{\pi^{-i}})\right|_\infty. \tag{3.22}$$

We next have the following lemma on the approximation error due to policy perturbation.

Lemma 3.3. *Fix any arbitrary $\pi_k \in \Pi$. For any $\tilde{\epsilon} > 0$, if ρ^i satisfies*

$$\rho^i \leq 1 - \left(1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}}, \quad \forall i \in [N], \tag{3.23}$$

then, we have that

$$\left|Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}}\right|_\infty \leq \tilde{\epsilon}, \quad \forall i \in [N], k \in [K].$$

Proof of Lemma 3.3. First note that, for all $i \in [N]$ and $k \in [K]$,

$$\left|Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}}\right|_\infty = \left|\mathcal{T}^i_{\pi_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\bar{\pi}_k^{-i}})\right|_\infty$$

$$\leq \left| \mathcal{T}^i_{\pi_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\pi_k^{-i}}) \right|_\infty + \left| \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\bar{\pi}_k^{-i}}) \right|_\infty . \quad (3.24)$$

By definition of $\bar{\pi}_k^{-i}$, we have that $P\left[\bar{\pi}_k^{-i} = \pi_k^{-i}\right] = \prod_{j\neq i}(1-\rho^j)$. With probability $1 - \prod_{j\neq i}(1-\rho^j)$, $\bar{\pi}_k^{-i} \neq \pi_k^{-i}$ and $\bar{\pi}_k^{-i} \in \bar{\Delta}^{-i}$. Thus, the first term of (3.24) can be bounded by

$$\left| \mathcal{T}^i_{\pi_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\pi_k^{-i}}) \right|_\infty \leq \left(1 - \prod_{j\neq i}(1-\rho^j)\right) \times \left| \mathcal{T}^i_{\pi_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\varphi_k^{-i}}(Q^i_{\pi_k^{-i}}) \right|_\infty , \quad (3.25)$$

for some $\varphi_k^{-i} \in \bar{\Delta}^{-i}$. On the other hand, by the contraction mapping of the Bellman operator, we have that

$$\left| \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}(Q^i_{\bar{\pi}_k^{-i}}) \right|_\infty \leq \gamma^i \left| Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty . \quad (3.26)$$

Substituting (3.25) and (3.26) back into (3.24), we have that

$$\begin{aligned} \left| Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty &\leq \left(1 - \prod_{j\neq i}(1-\rho^j)\right) \times \left| \mathcal{T}^i_{\pi_k^{-i}}(Q^i_{\pi_k^{-i}}) - \mathcal{T}^i_{\varphi_k^{-i}}(Q^i_{\pi_k^{-i}}) \right|_\infty \\ &\quad + \gamma^i \left| Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty \\ &\leq \left(1 - \prod_{j\neq i}(1-\rho^j)\right) \Gamma + \gamma^i \left| Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty , \end{aligned}$$

which implies that

$$\left| Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty \leq \frac{\left(1 - \prod_{j\neq i}(1-\rho^j)\right)\Gamma}{1-\gamma^i} \leq \frac{\left(1 - \prod_{j\neq i}(1-\rho^j)\right)\Gamma}{1-\bar{\gamma}}.$$

If for all $i \in [N]$, $\rho^i \leq 1 - \left(1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}}$, then, we have that $1 - \rho^j \geq \left(1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}}$,

which implies that $\prod_{j\neq i}(1-\rho^j) \geq 1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\Gamma}$, and thus

$$\left| Q^i_{\pi_k^{-i}} - Q^i_{\bar{\pi}_k^{-i}} \right|_\infty \leq \frac{\left(1 - \prod_{j\neq i}(1-\rho^j)\right)\Gamma}{1-\bar{\gamma}} \leq \tilde{\epsilon}.$$

The above holds for all $i \in [N]$ and $k \in [K]$, which completes the proof. □

Recall from (3.13) that $\bar{\zeta}$ is the minimum separation between the entries of agents' optimal Q-functions (with respect to the deterministic policies):

$$\bar{\zeta} := \min_{\substack{i,s,a^i,\tilde{a}^i,\pi^{-i}\in\Pi^{-i}: \\ Q^i_{\pi^{-i}}(s,a^i)\neq Q^i_{\pi^{-i}}(s,\tilde{a}^i)}} \left| Q^i_{\pi^{-i}}(s,a^i) - Q^i_{\pi^{-i}}(s,\tilde{a}^i) \right|.$$

We assume that $\bar{\zeta} > 0$ to avoid trivial cases, and consider $\bar{\zeta}$ as an upper bound on ζ^i for all i. We next define the following random event for any arbitrary $\pi_k \in \Pi$:

$$E_k := \left\{ \omega \in \Omega : \left| Q^i_{t_{k+1}} - Q^i_{\pi_k^{-i}} \right|_\infty < \frac{1}{4}\min\{\zeta^i, \bar{\zeta} - \zeta^i\}, \forall i \right\}.$$

With this definition of E_k, we show that, if E_k is not empty and $\pi_k \in \Pi_{\text{eq}}$, then $\pi_{k+1} = \pi_k$ with probability 1.

Lemma 3.4. *Given any $\pi_k \in \Pi$ and the corresponding E_k, for all k, we have that*

$$P\left[\pi_{k+1} = \pi_k \mid E_k,\ \pi_k \in \Pi_{\text{eq}}\right] = 1.$$

Proof of Lemma 3.4. Let $\hat{a}^{i*} := \arg\max_{\hat{a}^i} Q^i_{t_{k+1}}\left(s, \hat{a}^i\right)$. Then, conditioned on E_k and $\pi_k \in \Pi_{\text{eq}}$, we have that

$$\max_{\hat{a}^i} Q^i_{t_{k+1}}\left(s, \hat{a}^i\right) - Q^i_{t_{k+1}}\left(s, \pi^i_k(s)\right)$$

$$
\begin{aligned}
&= Q^i_{t_{k+1}}\left(s, \hat{a}^{i*}\right) - Q^i_{t_{k+1}}\left(s, \pi^i_k(s)\right) \\
&= \left[Q^i_{t_{k+1}}\left(s, \hat{a}^{i*}\right) - Q^i_{\pi^{-i}_k}\left(s, \pi^i_k(s)\right)\right] + \left[Q^i_{\pi^{-i}_k}\left(s, \pi^i_k(s)\right) - Q^i_{t_{k+1}}\left(s, \pi^i_k(s)\right)\right] \\
&< Q^i_{t_{k+1}}\left(s, \hat{a}^{i*}\right) - Q^i_{\pi^{-i}_k}\left(s, \pi^i_k(s)\right) + \frac{1}{2}\min\left\{\zeta^i, \bar{\zeta} - \zeta^i\right\} \\
&< \left[Q^i_{t_{k+1}}\left(s, \hat{a}^{i*}\right) - Q^i_{\pi^{-i}_k}\left(s, \hat{a}^{i*}\right)\right] + \left[Q^i_{\pi^{-i}_k}\left(s, \hat{a}^{i*}\right) - Q^i_{\pi^{-i}_k}\left(s, \pi^i_k(s)\right)\right] \\
&\quad + \frac{1}{4}\min\left\{\zeta^i, \bar{\zeta} - \zeta^i\right\} \\
&< \frac{1}{4}\min\left\{\zeta^i, \bar{\zeta} - \zeta^i\right\} + \frac{1}{4}\min\left\{\zeta^i, \bar{\zeta} - \zeta^i\right\} \leq \frac{1}{2}\min\left\{\zeta^i, \bar{\zeta} - \zeta^i\right\},
\end{aligned}
$$

where the second-to-last inequality follows since $Q^i_{\pi^{-i}_k}\left(s, \hat{a}^i\right) - Q^i_{\pi^{-i}_k}\left(s, \pi^i_k(s)\right) < 0$, which follows from $\pi_k \in \Pi_{\text{eq}}$. It follows that $Q^i_{t_{k+1}}\left(s, \pi^i_k(s)\right) \geq \max_{\hat{a}^i} Q^i_{t_{k+1}}\left(s, \hat{a}^i\right) - \frac{1}{2}\zeta^i$ for all i. Then, by Algorithm 3.1 (lines 11-13), we have that $\pi_{k+1} = \pi_k$ with probability 1. □

Recall that L is the maximum length of the shortest strict best reply path from any policy to an equilibrium policy. Our next lemma lower bounds the conditional probability of π_{k+L} being an equilibrium policy, given that π_k is not an equilibrium policy and $E_k, \ldots, E_{k+L-1}$.

Lemma 3.5. *Let*

$$
\hat{p} := \left(\min_{j \in \{1,\ldots,N\}} \left\{\frac{1-\lambda^j}{\left|\Pi^j\right|} \cdot \prod_{i \neq j} \lambda^i\right\}\right)^L, \tag{3.27}
$$

which is the same as that in (3.6)*. We then have that*

$$
P\left[\pi_{k+L} \in \Pi_{\text{eq}} \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k \notin \Pi_{\text{eq}}\right] \geq \hat{p}. \tag{3.28}
$$

Proof of Lemma 3.5. We begin with an important observation. Consider some $\pi_k \notin \Pi_{\text{eq}}$; then, there must exist at least one agent, say agent i, whose policy π^i_k is not the best reply to π^{-i}_k, i.e., $\pi^i_k \notin \Pi^i_{\pi^{-i}_k}$. In this case, we claim that $\pi^i_k \notin \Pi^i_{k+1}$, where Π^i_{k+1} is as defined in Algorithm 3.1 (line 11). In other words, the "else" statement in Algorithm 3.1 (line 15) will be

executed. To see this, it suffices to show that $Q^i_{t_{k+1}}(s, \pi^i_k(s)) < \max_{a^i \in \mathcal{A}^i} Q^i_{t_{k+1}}(s, a^i) - \frac{1}{2}\zeta^i$ for some $s \in \mathcal{S}$. Conditioned on E_k, we have that

$$Q^i_{\pi^{-i}_k}(s, a^i) - \frac{1}{4}\min\{\zeta^i, \bar{\zeta} - \zeta^i\} < Q^i_{t_{k+1}}(s, a^i) < Q^i_{\pi^{-i}_k}(s, a^i) + \frac{1}{4}\min\{\zeta^i, \bar{\zeta} - \zeta^i\},$$

i.e., $Q^i_{t_{k+1}}(s, a^i)$ lies within a distance of $\frac{1}{4}\min\{\zeta^i, \bar{\zeta} - \zeta^i\}$ to $Q^i_{\pi^{-i}_k}(s, a^i)$. Moreover, we note that $\frac{1}{4}\min\{\zeta^i, \bar{\zeta} - \zeta^i\} \leq \frac{1}{8}\bar{\zeta}$. Recall that $\left\{Q^i_{\pi^{-i}_k}(s, a^i) : a^i \in \mathcal{A}^i\right\}$ are dispersed with spacing being at least $\bar{\zeta}$, where $\bar{\zeta}$ is as defined in (3.13) as the minimum separation between the optimal Q-functions. Thus, it follows that the possible range of $Q^i_{t_{k+1}}(s, a^i)$ for all $a^i \in \mathcal{A}^i$ are mutually exclusive, which implies that the τ-th best action under $Q^i_{\pi^{-i}_k}$ is identical to that under $Q^i_{t_{k+1}}$, i.e.,

$$\arg\max_{a^i \in \mathcal{A}^i} \left(Q^i_{\pi^{-i}_k}(s, a^i)\right)_{(\tau)} = \arg\max_{a^i \in \mathcal{A}^i} \left(Q^i_{t_{k+1}}(s, a^i)\right)_{(\tau)},$$

where $(\cdot)_{(\tau)}$ represents the τ-th largest value. For instance, when $\tau = 1$, we have that $\arg\max_{a^i \in \mathcal{A}^i} Q^i_{\pi^{-i}_k}(s, a^i) = \arg\max_{a^i \in \mathcal{A}^i} Q^i_{t_{k+1}}(s, a^i)$, which are denoted by $a^{i*}_{\pi^{-i}_k}(s)$ and $a^{i*}_{t_{k+1}}(s)$, respectively.

Since $\pi^i_k \notin \Pi^i_{\pi^{-i}_k}$, it follows that $\pi^i_k(s) \neq \arg\max_{a^i \in \mathcal{A}^i} Q^i_{\pi^{-i}_k}(s, a^i) =: a^{i*}_{\pi^{-1}_k}(s)$ for some $s \in \mathcal{S}$. Then, we have that

$$\begin{aligned}
\max_{a^i \in \mathcal{A}^i} Q^i_{t_{k+1}}(s, a^i) - Q^i_{t_{k+1}}(s, \pi^i_k(s)) &> \left(\max_{a^i \in \mathcal{A}^i} Q^i_{\pi^{-i}_k}(s, a^i) - \frac{1}{8}\bar{\zeta}\right) - \left(Q^i_{\pi^{-i}_k}(s, \pi^i_k(s)) + \frac{1}{8}\bar{\zeta}\right) \\
&= \left(Q^i_{\pi^{-i}_k}\left(s, a^{i*}_{\pi^{-i}_k}(s)\right) - Q^i_{\pi^{-i}_k}(s, \pi^i_k(s))\right) - \frac{1}{4}\bar{\zeta} \\
&\geq \bar{\zeta} - \frac{1}{4}\bar{\zeta} = \frac{3}{4}\bar{\zeta} \geq \frac{3}{4}\zeta^i > \frac{1}{2}\zeta^i
\end{aligned}$$

as desired. Now, we are ready to prove the statement.

Let l be the length of the shortest strict best reply path from π_k to an equilibrium

policy. Then $l \leq L$. Let the sequence of policies along the path be $\pi_0, \pi_1, \ldots, \pi_l$, with $\pi_0 = \pi_k \notin \Pi_{\text{eq}}$ and $\pi_l \in \Pi_{\text{eq}}$. Further, let $i_1, \ldots, i_l$ be the agent that changes its policy at each update, i.e., π_{n-1} and π_n differ only at agent i_n, for all $n = 1, \ldots, l$. Then, based on the aforementioned observation, we can use the two probabilities in the policy update rule in Algorithm 3.1 (line 15) to yield

$$
\begin{aligned}
&P\left[\pi_{k+L} \in \Pi_{\text{eq}} \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k \notin \Pi_{\text{eq}}\right] \\
&\geq P\left[\pi_{k+L} = \pi_l \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k \notin \Pi_{\text{eq}}\right] \\
&\geq P\big[\pi_{k+1} = \pi_1, \pi_{k+2} = \pi_2, \ldots, \pi_{k+l} = \pi_l, \\
&\qquad \pi_{k+l+1} = \cdots = \pi_{k+L} = \pi_l \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k \notin \Pi_{\text{eq}}\big] \\
&\geq P\left[\pi_{k+1} = \pi_1 \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k = \pi_0\right] \\
&\quad \cdot P\left[\pi_{k+2} = \pi_2 \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1\right] \\
&\quad \cdot P\left[\pi_{k+3} = \pi_3 \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1, \pi_{k+2} = \pi_2\right] \cdot \cdots \\
&\quad \cdot P\left[\pi_{k+l} = \pi_l \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1, \ldots, \pi_{k+l-1} = \pi_{l-1}\right] \\
&\quad \cdot P\left[\pi_{k+l+1} = \pi_l \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1, \ldots, \pi_{k+l} = \pi_l\right] \cdot \cdots \\
&\quad \cdot P\left[\pi_{k+L} = \pi_l \;\middle|\; E_k, \ldots, E_{k+L-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1, \ldots, \pi_{k+l} = \pi_l, \ldots, \pi_{k+L-1} = \pi_l\right] \\
&\geq \prod_{j \in \{i_1, \ldots, i_l\}} \left(\frac{1 - \lambda^j}{\left|\Pi^j\right|} \cdot \prod_{i \neq j} \lambda^i \right) \geq \left(\min_{j \in \{1, \ldots, N\}} \left\{ \frac{1 - \lambda^j}{\left|\Pi^j\right|} \cdot \prod_{i \neq j} \lambda^i \right\} \right)^l \\
&\geq \left(\min_{j \in \{1, \ldots, N\}} \left\{ \frac{1 - \lambda^j}{\left|\Pi^j\right|} \cdot \prod_{i \neq j} \lambda^i \right\} \right)^L,
\end{aligned}
$$

where we have used the fact from Lemma 3.4: given $\pi_l \in \Pi_{\text{eq}}$ and the events $E_k, \ldots, E_{k+L-1}$, the conditional probability that $\pi_s \in \Pi_{\text{eq}}$ is 1 for all $s \geq l$. □

We will then bound $P\left[E_k, \ldots, E_{k+L-1}\right]$. Before that, we first look at $P[E_k]$. We would like $P[E_k]$ to be as large as possible. Note that $\frac{1}{4} \min\{\zeta^i, \bar{\zeta} - \zeta^i\} \leq \frac{1}{8}\bar{\zeta}$, with equality holding

when $\zeta^i = \frac{1}{2}\bar{\zeta}$. We next have the following lemma.

Lemma 3.6. *Let $\zeta^i = \frac{\bar{\zeta}}{2}$ for all $i \in [N]$. Fix an arbitrary $\pi_k \in \Pi$. For any $0 < \hat{\delta} < 1$, we have that*

$$P\left[E_k\right] \geq 1 - \hat{\delta},$$

provided that $\rho^i \leq 1 - \left(1 - \frac{(\bar{\zeta}/8 - \epsilon)(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}}$, and T_k and η_t^i satisfy (3.18), where ϵ can take any value in $0 < \epsilon < \min\left\{\frac{\bar{\zeta}}{8}, \frac{1}{1-\gamma_{\min}}\right\}$.

Proof of Lemma 3.6. A direct implication of Lemma 3.2 and Lemma 3.3 is that when T_k and η_t^i satisfy (3.18), and ρ^i satisfies (3.23), then, by triangle inequality, we have that

$$P\left[\left|Q^i_{t_{k+1}} - Q^i_{\pi_k^{-i}}\right|_\infty \leq \epsilon + \tilde{\epsilon}, \quad \forall i \in [N]\right] \geq 1 - \hat{\delta}. \tag{3.29}$$

The lemma then follows by taking $\tilde{\epsilon} = \frac{1}{8}\bar{\zeta} - \epsilon$. □

We then have the following lemma which bounds $P\left[E_k, \ldots, E_{k+L-1}\right]$.

Lemma 3.7. *For any arbitrary sequence of joint policies $\pi_k, \ldots, \pi_{k+L-1} \in \Pi$, and for any $0 < \tilde{\delta} < 1$, we have that*

$$P\left[E_k, \ldots, E_{k+L-1}\right] \geq 1 - \tilde{\delta},$$

provided that for all $i \in [N]$ and for all $\hat{k} \in \{k, \ldots, k+L-1\}$,

$$T_{\hat{k}} \geq \frac{c_{0,\hat{k}}}{\mu_{\min,\hat{k}}} \left\{\frac{1}{(1-\bar{\gamma})^5\epsilon^2} + \frac{t_{\text{mix},\hat{k}}\left(\frac{1}{4}\right)}{1-\bar{\gamma}}\right\} \log\left(\frac{NL|\mathcal{S}|AT_{\hat{k}}}{\tilde{\delta}}\right) \log\left(\frac{1}{(1-\bar{\gamma})^2\epsilon}\right), \tag{3.30a}$$

$$\eta_t^i = \frac{c^i_{1,\hat{k}}}{\log\left(\frac{NL|\mathcal{S}||\mathcal{A}^i|T_{\hat{k}}}{\tilde{\delta}}\right)} \min\left\{\frac{(1-\gamma^i)^4\epsilon^2}{\left(\gamma^i\right)^2}, \frac{1}{t^i_{\text{mix},\hat{k}}\left(\frac{1}{4}\right)}\right\}, \quad \forall t = t_{\hat{k}}, \ldots, t_{\hat{k}+1} - 1, \tag{3.30b}$$

$$\rho^i \leq 1 - \left(1 - \frac{\left(\bar{\zeta}/8 - \epsilon\right)(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}} \tag{3.30c}$$

$$\zeta^i = \frac{\bar{\zeta}}{2} \tag{3.30d}$$

where ϵ *can take any value in* $0 < \epsilon < \min\left\{\frac{\bar{\zeta}}{8}, \frac{1}{1-\gamma_{\min}}\right\}$.

Proof of Lemma 3.7. When the conditions of Lemma 3.6 are satisfied, we have $P\left[E_k^c\right] < \hat{\delta}$, where E_k^c is the complement of E_k. Then,

$$\begin{aligned} P\left[E_k, \ldots, E_{k+L-1}\right] &= 1 - P\left[(E_k, \ldots, E_{k+L-1})^c\right] = 1 - P\left[E_k^c \cup \cdots \cup E_{k+L-1}^c\right] \\ &\geq 1 - \left(P\left[E_k^c\right] + P\left[E_{k+1}^c\right] + \cdots + P\left[E_{k+L-1}^c\right]\right) = 1 - L\hat{\delta}. \end{aligned}$$

By taking $\tilde{\delta} = L\hat{\delta}$, it follows that the conditions (3.18) now become (3.30), and the lemma is thus proved. □

As for the choice of ϵ, we would like to make $T_{\hat{k}}$ as small as possible. As ϵ increases, the term $\frac{1}{(1-\bar{\gamma}^5\epsilon^2)}$ decreases, while ρ^i also decreases, which leads to a smaller $\mu_{\min,\hat{k}}$ and a larger $t_{\text{mix},\hat{k}}$. Therefore, we would like to choose an optimal $\hat{\epsilon}_{\hat{k}}$ for $T_{\hat{k}}$, such that (ignoring the logarithmic factors)

$$\hat{\epsilon}_{\hat{k}} := \underset{0<\epsilon<\min\left\{\frac{\bar{\zeta}}{8}, \frac{1}{1-\gamma_{\min}}\right\}}{\arg\min} \frac{1}{\mu_{\min,\hat{k}}}\left(\frac{1}{(1-\bar{\gamma})^4\epsilon^2} + t_{\text{mix},\hat{k}}\left(\frac{1}{4}\right)\right), \tag{3.31}$$

or, using the result of Proposition 3.2 for the bounds of $t_{\text{mix},\hat{k}}\left(\frac{1}{4}\right)$ and $\mu_{\min,\hat{k}}$,

$$\hat{\epsilon}_{\hat{k}} = \underset{\epsilon\in\left(0,\min\left\{\frac{\bar{\zeta}}{8}, \frac{1}{1-\gamma_{\min}}\right\}\right)}{\arg\min} \frac{A^{(N+1)H}}{\kappa\rho^{(N+1)H}}$$

$$\cdot \left(\frac{1}{(1-\bar{\gamma})^4 \epsilon^2} + (H+1) \left(\frac{\log(1/4)}{\log\left[\frac{A^{NH} - |\mathcal{S}|\kappa\rho^{NH}}{A^{NH} - (|\mathcal{S}|-1)\kappa\rho^{NH}} \right]} + 1 \right) \right), \tag{3.32}$$

and we can define $\hat{\epsilon} := \max_{\hat{k} \in \{k, \ldots, k+L-1\}} \hat{\epsilon}_k$, and let $\rho^i = 1 - \left(1 - \frac{(\bar{\zeta}/8 - \hat{\epsilon})(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}}$, which leads to the optimal sample complexity for $T_{\hat{k}}$. For simplicity, we choose that $\epsilon = \frac{1}{2}\min\left\{\frac{\bar{\zeta}}{8}, \frac{1}{1-\gamma_{\min}}\right\}$ in Theorem 3.1 and do not worry about optimizing over ϵ.

Note also that the result of Lemma 3.7 holds for any realization of $\pi_k \in \Pi$. Therefore, under the same conditions, we in fact have that

$$P\left[E_k, \ldots, E_{k+L-1} \;\middle|\; \pi_k \in \Pi_{\text{eq}}\right] \geq 1 - \tilde{\delta}, \tag{3.33a}$$

$$P\left[E_k, \ldots, E_{k+L-1} \;\middle|\; \pi_k \notin \Pi_{\text{eq}}\right] \geq 1 - \tilde{\delta}. \tag{3.33b}$$

By Lemma 3.4 and (3.33a), under conditions (3.30), we have that for all k,

$$P\left[\pi_k = \pi_{k+1} = \cdots = \pi_{k+L} \;\middle|\; \pi_k \in \Pi_{\text{eq}}\right] \geq 1 - \tilde{\delta}. \tag{3.34}$$

By Lemma 3.5 and (3.33b), under conditions (3.30), we have that for all k,

$$P\left[\pi_{k+L} \in \Pi_{\text{eq}} \;\middle|\; \pi_k \notin \Pi_{\text{eq}}\right] \geq \hat{p}\left(1 - \tilde{\delta}\right). \tag{3.35}$$

As a notation, let $p_k := P\left[\pi_k \in \Pi_{\text{eq}}\right]$. Then, (3.34) and (3.35) together imply that

$$p_{(n+1)L} \geq p_{nL}\left(1 - \tilde{\delta}\right) + (1 - p_{nL})\hat{p}\left(1 - \tilde{\delta}\right). \tag{3.36}$$

Rearranging the above, we obtain that

$$\begin{aligned} p_{(n+1)L} - p_{nL} &\geq \left(1-\tilde{\delta}\right)\hat{p} - \tilde{\delta}p_{nL} - \left(1-\tilde{\delta}\right)\hat{p}p_{nL} \\ &= \left[\tilde{\delta} + \left(1-\tilde{\delta}\right)\hat{p}\right]\left[\frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}} - p_{nL}\right] \qquad (3.37) \\ &\geq -\tilde{\delta} \qquad (3.38) \end{aligned}$$

Note that $p_{(n+1)L} - p_{nL} \geq 0$ as long as $p_{nL} \leq \frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}}$. Further, if $p_{nL} \leq \frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}} - \tilde{\delta}$, then from (3.37), we have that $p_{(n+1)L} - p_{nL} \geq \left[\tilde{\delta} + \left(1-\tilde{\delta}\right)\hat{p}\right]\tilde{\delta}$; if $p_{nL} > \frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}}$, then $p_{(n+1)L} - p_{nL} \geq -\tilde{\delta}$ from (3.38). Therefore, we have that

$$p_{nL} \geq \frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}} - \tilde{\delta}, \quad \forall n \geq \tilde{n}, \qquad (3.39)$$

where

$$\tilde{n} := \frac{\frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}} - \tilde{\delta}}{\left[\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}\right]\tilde{\delta}} = \frac{\left(1-\tilde{\delta}\right)^2\hat{p} - \tilde{\delta}^2}{\left[\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}\right]^2\tilde{\delta}}. \qquad (3.40)$$

This, together with (3.34), implies that for all $n \geq \tilde{n}$,

$$P\left[\pi_{nL} = \pi_{nL+1} = \cdots = \pi_{nL+L} \in \Pi_{\text{eq}}\right] \geq \left(\frac{\left(1-\tilde{\delta}\right)\hat{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\hat{p}} - \tilde{\delta}\right)\left(1-\tilde{\delta}\right) := f(\tilde{\delta}). \qquad (3.41)$$

Therefore, if the number of exploration phases $k \geq K := \tilde{n}L$, then $P\left[\pi_k \in \Pi_{\text{eq}}\right] \geq f(\tilde{\delta})$. Note that $f(\tilde{\delta})$ is continuous, decreasing in $\tilde{\delta}$, and $f(0) = 1$, $f(\delta) < 1 - \delta$ for any $0 < \delta < 1$.

Thus, we can take $\tilde{\delta} \in (0, \delta)$ such that

$$\left(\frac{\left(1 - \tilde{\delta}\right) \hat{p}}{\tilde{\delta} + \left(1 - \tilde{\delta}\right) \hat{p}} - \tilde{\delta} \right) \left(1 - \tilde{\delta}\right) = 1 - \delta, \tag{3.42}$$

which leads to $P\left[\pi_k \in \Pi_{\text{eq}}\right] \geq 1 - \delta$ for all $k \geq K$, and completes the proof of Theorem 3.1.

3.3.2 Proof of Proposition 3.2.

We first show the bounds for $\mu^i_{\min,k}$ and $\mu_{\min,k}$. By Assumption 3.1, for any s_1 and s_{H+1}, there exists a sequence of joint actions $\tilde{a}_1, \ldots, \tilde{a}_H$ such that $P[s_{H+1} = s \mid (s_1, a_1, \ldots, a_H) = (s, \tilde{a}_1, \ldots, \tilde{a}_H)] \geq \kappa$. Thus, we have that

$$\begin{aligned} P\left(s_{H+1} = s \mid s_1\right) &= \sum_{a_1, \ldots, a_H} P\left(s_{H+1} = s \mid (s_1, a_1, \ldots, a_H)\right) P\left((s_1, a_1, \ldots, a_H) \mid s_1\right) \\ &\geq P\left(s_{H+1} = s \mid (s_1, \tilde{a}_1, \ldots, \tilde{a}_H)\right) P\left((s_1, \tilde{a}_1, \ldots, \tilde{a}_H) \mid s_1\right) \\ &\geq \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H, \quad \forall s_1, s \in \mathcal{S}, \end{aligned} \tag{3.43}$$

where the last inequality follows from Assumption 3.1 and the action selection (Line 5) of Algorithm 3.1. Then, we can write the lower bound for the stationary distribution $\mu_{\bar{\pi}_k}$ over all states:

$$\mu_{\bar{\pi}_k}(s) = \sum_{s_1 \in \mathcal{S}} \mu_{\bar{\pi}_k}(s_1) P\left(s_{H+1} = s \mid s_1\right) \geq \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H, \quad \forall s \in \mathcal{S}, \tag{3.44}$$

where the inequality follows since (3.43) is a uniform lower bound for all (s_1, s), and also $\sum_{s_1 \in \mathcal{S}} \mu_{\pi_k}(s_1) = 1$. Note that (3.44) also implies the following upper bound:

$$\mu_{\bar{\pi}_k}(s) = 1 - \sum_{\bar{s} \in \mathcal{S} \setminus \{s\}} \mu_{\bar{\pi}_k}(\bar{s}) \leq 1 - (|\mathcal{S}| - 1)\, \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H, \quad \forall s \in \mathcal{S}. \tag{3.45}$$

From (3.44) and (3.45), and by noting that $\min_{(s,a^i) \in \mathcal{S} \times \mathcal{A}^i} P(a^i \mid s) = \frac{\rho^i}{|\mathcal{A}^i|}$, we can also write the bounds for the minimum probability of the stationary distribution over state-action pairs (from the perspective of agent i):

$$\mu^i_{\min,k} = \min_{(s,a^i) \in \mathcal{S} \times \mathcal{A}^i} \mu^i_{\bar{\pi}_k}\left(s, a^i\right) = \min_{(s,a^i) \in \mathcal{S} \times \mathcal{A}^i} P(a^i \mid s) \mu_{\bar{\pi}_k}(s) \geq \kappa \frac{\rho^i}{|\mathcal{A}^i|} \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H, \tag{3.46a}$$

$$\begin{aligned} \mu^i_{\min,k} &= \min_{(s,a^i) \in \mathcal{S} \times \mathcal{A}^i} \mu^i_{\bar{\pi}_k}\left(s, a^i\right) = \min_{(s,a^i) \in \mathcal{S} \times \mathcal{A}^i} \mu_{\bar{\pi}_k}(s) \cdot P(a^i \mid s) \\ &\leq \left[1 - (|\mathcal{S}| - 1)\, \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \right] \cdot \frac{\rho^i}{|\mathcal{A}^i|}. \end{aligned} \tag{3.46b}$$

With $\rho^i = \rho$ for all $i \in [N]$ and recalling that $A := \max_{i \in [N]} \mathcal{A}^i$, (3.46) leads to

$$\mu_{\min,k} = \min_{i \in [N]} \mu^i_{\min,k} \geq \kappa \frac{\rho^{NH+1}}{A^{NH+1}}, \tag{3.47a}$$

$$\mu_{\min,k} = \min_{i \in [N]} \mu^i_{\min,k} \leq \left[1 - (|\mathcal{S}| - 1)\, \kappa \frac{\rho^{NH}}{A^{NH}} \right] \cdot \frac{\rho}{A}. \tag{3.47b}$$

We next show the upper bound for $t_{\text{mix},k}$. To proceed, we first introduce the following two lemmas. Lemma 3.8 follows directly from the coupling inequality under the *Doeblin condition* (see Diaconis [65] for a detailed explanation). Lemma 3.9 follows directly from Dobrushin's theorem (see Dobrushin [67] for details).

Lemma 3.8. *If* $P\left(s_{H+1}=\bar{s}, a^i_{H+1}=\bar{a}^i \mid s_0, a^i_0\right) \geq c_1 \mu^i_{\bar{\pi}_k}\left(\bar{s}, \bar{a}^i\right)$ *for all* s_0, a^i_0 *and* $\bar{s}, \bar{a}^i$*, then,*

$$\begin{aligned} d_{\mathrm{TV}}\left(P^t(\cdot \mid s_0, a^i_0), \mu^i_{\bar{\pi}_k}\right) &= \max_{\bar{s}, \bar{a}^i}\left|P\left(s_t=\bar{s}, a^i_t=\bar{a}^i \mid s_0, a^i_0\right)-\mu^i_{\bar{\pi}_k}\left(\bar{s}, \bar{a}^i\right)\right| \\ &\leq (1-c_1)^{\lfloor t/(H+1)\rfloor}. \end{aligned}$$

Lemma 3.9. *Let* $s, \bar{s}, \hat{s}$ *be any state and* $a^i, \bar{a}^i, \hat{a}^i$ *be any action of agent* i*. Define*

$$c_2 := \min_{\bar{s}, \bar{a}^i, \hat{s}, \hat{a}^i} \sum_{s, a^i} \min\left(P^{H+1}\left(s, a^i \mid \bar{s}, \bar{a}^i\right), P^{H+1}\left(s, a^i \mid \hat{s}, \hat{a}^i\right)\right). \tag{3.48}$$

Then, we have that

$$\begin{aligned} d_{\mathrm{TV}}\left(P^t(\cdot \mid s_0, a^i_0), \mu^i_{\bar{\pi}_k}\right) &= \max_{\bar{s}, \bar{a}^i}\left|P\left(s_t=\bar{s}, a^i_t=\bar{a}^i \mid s_0, a^i_0\right)-\mu^i_{\bar{\pi}_k}\left(\bar{s}, \bar{a}^i\right)\right| \\ &\leq (1-c_2)^{\lfloor t/(H+1)\rfloor}. \end{aligned}$$

We will use both Lemma 3.8 and Lemma 3.9 to prove the upper bound of $t_{\mathrm{mix},k}$. To apply Lemma 3.8, we need to find the parameter c_1 such that $P\left(s_{H+1}=\bar{s}, a^i_{H+1}=\bar{a}^i \mid s_0, a^i_0\right) \geq c_1 \mu^i_{\bar{\pi}_k}\left(\bar{s}, \bar{a}^i\right)$ for all s_0, a^i_0 and $\bar{s}, \bar{a}^i$. To apply Lemma 3.9, we need to find the c_2 (or a lower bound of it) that satisfies (3.48). Here, $(1-c_2)$ is also known as the *Dobrushin ergodicity coefficient* (Gaubert and Qu [96]). These lead to the following lemma.

Lemma 3.10. *For all* $i \in [N]$ *and for any* $(s_0, a^i_0), (\bar{s}, \bar{a}^i) \in \mathcal{S} \times \mathcal{A}^i$*, we have that*

$$P^{H+1}\left(\bar{s}, \bar{a}^i \mid s_0, a^i_0\right) = P\left(s_{H+1}=\bar{s}, a^i_{H+1}=\bar{a}^i \mid s_0, a^i_0\right) \geq \kappa\left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H \cdot \frac{\rho^i}{|\mathcal{A}^i|}, \tag{3.49a}$$

$$P\left(s_{H+1}=\bar{s}, a^i_{H+1}=\bar{a}^i \mid s_0, a^i_0\right) \geq c_1 \mu^i_{\bar{\pi}_k}\left(\bar{s}, \bar{a}^i\right), \tag{3.49b}$$

where

$$c_1 = \frac{\kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H}{1 - (|\mathcal{S}| - 1)\, \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H}. \tag{3.49c}$$

Proof of Lemma 3.10. By Assumption 3.1, we know that for all s, s', there exist $\tilde{a}_0, \ldots, \tilde{a}_{H-1}$ such that

$$P[s_H = s' \mid (s_0, a_0, \ldots, a_{H-1}) = (s, \tilde{a}_0, \ldots, \tilde{a}_{H-1})] \geq \kappa.$$

Then, we have that

$$\begin{aligned}
& P\left(s_{H+1} = \bar{s}, a^i_{H+1} = \bar{a}^i \mid s_0, a^i_0\right) \\
&= P\left(s_{H+1} = \bar{s} \mid s_0, a^i_0\right) \cdot P\left(a^i_{H+1} = \bar{a} \mid s_{H+1} = \bar{s}, s_0, a^i_0\right) \\
&= P\left(s_{H+1} = \bar{s} \mid s_0, a^i_0\right) \cdot P\left(a^i_{H+1} = \bar{a} \mid s_{H+1} = \bar{s}\right) \\
&\geq P\left(s_{H+1} = \bar{s} \mid s_0, a^i_0\right) \cdot \frac{\rho^i}{|\mathcal{A}^i|}.
\end{aligned} \tag{3.50}$$

Then, it suffices to find a lower bound for $P\left(s_{H+1} = \bar{s} \mid s_0, a^i_0\right)$. Note that

$$\begin{aligned}
P\left(s_{H+1} = \bar{s} \mid s_0, a^i_0\right) &= \sum_{s_1} P\left(s_{H+1} = \bar{s} \mid s_1, s_0, a^i_0\right) P\left(s_1 \mid s_0, a^i_0\right) \\
&= \sum_{s_1} P\left(s_{H+1} = \bar{s} \mid s_1\right) P\left(s_1 \mid s_0, a^i_0\right).
\end{aligned}$$

We make the following observations.

- For any s_1, by the law of total probability, we have that

$$P\left(s_{H+1} = \bar{s} \mid s_1\right)$$

$$
\begin{aligned}
&= \sum_{\{a_1,\ldots,a_H\}} P\left(s_{H+1} = \bar{s} \mid s_1, a_1, \ldots, a_H\right) P\left(a_1, \ldots, a_H \mid s_1\right) \\
&= P\left(s_{H+1} = \bar{s} \mid s_1, \tilde{a}_1, \ldots, \tilde{a}_H\right) P\left(\tilde{a}_1, \ldots, \tilde{a}_H \mid s_1\right) \\
&\quad + \sum_{\{a_1,\ldots,a_H\} \neq \{\tilde{a}_1,\ldots,\tilde{a}_H\}} P\left(s_{H+1} = \bar{s} \mid s_1, a_1, \ldots, a_H\right) P\left(a_1, \ldots, a_H \mid s_1\right) \\
&\geq \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H .
\end{aligned}
$$

- Since the above is true for any $s_1 \in \mathcal{S}$, and $\sum_{s_1} P\left(s_1 \mid s_0, a_0^i\right) = 1$, we have that the convex combination

$$
P\left(s_{H+1} = \bar{s} \mid s_0, a_0^i\right) = \sum_{s_1} P\left(s_{H+1} = \bar{s} \mid s_1\right) P\left(s_1 \mid s_0, a_0^i\right) \geq \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H .
$$

Thus, (3.50) now becomes

$$
\begin{aligned}
P\left(s_{H+1} = \bar{s}, a_{H+1}^i = \bar{a}^i \mid s_0, a_0^i\right) &\geq P\left(s_{H+1} = \bar{s} \mid s_0, a_0^i\right) \cdot \frac{\rho^i}{|\mathcal{A}^i|} \\
&\geq \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \cdot \frac{\rho^i}{|\mathcal{A}^i|}.
\end{aligned}
$$

This completes the proof of (3.49a).

With (3.49a), we can see that any c_1 satisfying $\kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \cdot \frac{\rho^i}{|\mathcal{A}^i|} \geq c_1 \mu_{\bar{\pi}_k}^i \left(\bar{s}, \bar{a}^i\right)$ for all $\bar{s}, \bar{a}^i$ and for all i will lead to (3.49b). Equivalently, we need

$$
c_1 \leq \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \cdot \min_{i \in [N]} \frac{\frac{\rho^i}{|\mathcal{A}^i|}}{\mu_{\min,k}^i}.
$$

One option is to choose $c_1 = \kappa \left(\prod_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \cdot \min_{i \in [N]} \frac{\rho^i}{|\mathcal{A}^i|}$, which essentially uses $\mu_{\min,k}^i \leq$

1. However, note that we can use the upper bound of $\mu^i_{\min,k}$ from (3.46), and choose that

$$\begin{aligned} c_1 &= \kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \cdot \min_{i\in[N]} \frac{\frac{\rho^i}{|\mathcal{A}^i|}}{\left[1 - (|\mathcal{S}|-1)\,\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H\right] \cdot \frac{\rho^i}{|\mathcal{A}^i|}} \\ &= \frac{\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H}{1 - (|\mathcal{S}|-1)\,\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H}. \end{aligned}$$

This completes the proof of the lemma. □

It follows from (3.48) and Lemma 3.10 that

$$c_2 \geq |\mathcal{S}||\mathcal{A}^i| \cdot \kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \cdot \frac{\rho^i}{|\mathcal{A}^i|} = |\mathcal{S}|\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|} \right)^H \rho^i. \tag{3.51}$$

The rest of the proof is divided into two parts. We first show that

$$t_{\text{mix},k}(\alpha) \leq (H+1) \left((-\log \alpha) \frac{A^{NH}}{\kappa \rho^{NH}} + 1 \right). \tag{3.52}$$

Combining Lemma 3.8 and Lemma 3.10, we have that for all $i \in [N]$ and $(s_0, a_0^i) \in \mathcal{S}\times\mathcal{A}^i$,

$$\begin{aligned} & d_{\text{TV}} \left(P^t(\cdot \mid s_0, a_0^i), \mu^i_{\bar{\pi}_k} \right) \\ &\leq (1 - c_1)^{\lfloor t/(H+1) \rfloor} \\ &= \left[\frac{1 - (|\mathcal{S}|-1)\,\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H - \kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H}{1 - (|\mathcal{S}|-1)\,\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H} \right]^{\lfloor t/(H+1) \rfloor} \\ &= \left[\frac{1 - |\mathcal{S}|\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H}{1 - (|\mathcal{S}|-1)\,\kappa \left(\prod_{i\in[N]} \frac{\rho^i}{|\mathcal{A}^i|}\right)^H} \right]^{\lfloor t/(H+1) \rfloor}. \end{aligned}$$

Let t be such that

$$\left[\frac{1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H}{1-(|\mathcal{S}|-1)\,\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H}\right]^{\lfloor t/(H+1)\rfloor} \leq \left[\frac{1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H}{1-(|\mathcal{S}|-1)\,\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H}\right]^{t/(H+1)-1}$$
$$= \alpha. \tag{3.53}$$

Then, we have that $\max_{(s_0,a_0^i)\in\mathcal{S}\times\mathcal{A}^i} d_{\mathrm{TV}}\left(P^t(\cdot \mid s_0,a_0^i),\mu^i_{\pi_k}\right) \leq \alpha, \forall i \in [N]$, which implies that $t_{\mathrm{mix},k}(\alpha) = \max_{i\in[N]} t^i_{\mathrm{mix},k} \leq t$. Therefore, by solving (3.53) for t, we obtain an upper bound for $t_{\mathrm{mix},k}$:

$$t_{\mathrm{mix},k}(\alpha) \leq t = (H+1)\left(\frac{-\log\alpha}{\log\left[\frac{1-(|\mathcal{S}|-1)\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H}{1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H}\right]}+1\right). \tag{3.54}$$

With $\rho^i = \rho$ for all $i \in [N]$, (3.54) becomes

$$t_{\mathrm{mix},k}(\alpha) \leq (H+1)\left(\frac{-\log\alpha}{\log\left[\frac{1-(|\mathcal{S}|-1)\kappa\left(\frac{\rho}{\prod_{i\in[N]}|\mathcal{A}^i|}\right)^H}{1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho}{\prod_{i\in[N]}|\mathcal{A}^i|}\right)^H}\right]}+1\right)$$
$$\leq (H+1)\left(\frac{(-\log\alpha)\left[1-(|\mathcal{S}|-1)\kappa\left(\frac{\rho}{\prod_{i\in[N]}|\mathcal{A}^i|}\right)^H\right]}{\kappa\left(\frac{\rho}{\prod_{i\in[N]}|\mathcal{A}^i|}\right)^H}+1\right)$$

$$\leq (H+1)\left((-\log\alpha)\frac{\prod_{i\in[N]}|\mathcal{A}^i|^H}{\kappa\rho^H}+1\right)\leq (H+1)\left((-\log\alpha)\frac{A^{NH}}{\kappa\rho^{NH}}+1\right).$$

This completes the proof of (3.52).

Next, we show that

$$t_{\text{mix},k}(\alpha)\leq (H+1)\left((-\log\alpha)\frac{A^{NH}}{|\mathcal{S}|\kappa\rho^{NH+1}}+1\right). \tag{3.55}$$

With (3.51) and Lemma 3.9, we have that for all $i\in[N]$ and $(s_0,a_0^i)\in\mathcal{S}\times\mathcal{A}^i$,

$$d_{\text{TV}}\left(P^t(\cdot\mid s_0,a_0^i),\mu^i_{\bar{\pi}_k}\right)\leq(1-c_2)^{\lfloor t/(H+1)\rfloor}\leq\left[1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H\rho^i\right]^{\lfloor t/(H+1)\rfloor}.$$

Let t be such that

$$\begin{aligned}\left[1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H\rho^i\right]^{\lfloor t/(H+1)\rfloor}&\leq\left[1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H\rho^i\right]^{t/(H+1)-1}\\&=\alpha,\quad\forall i\in[N].\end{aligned} \tag{3.56}$$

Then, we have that $\max_{(s_0,a_0^i)\in\mathcal{S}\times\mathcal{A}^i} d_{\text{TV}}\left(P^t(\cdot\mid s_0,a_0^i),\mu^i_{\bar{\pi}_k}\right)\leq\alpha,\forall i\in[N]$, which implies that $t_{\text{mix},k}(\alpha)=\max_{i\in[N]}t^i_{\text{mix},k}\leq t$. Therefore, by solving (3.56) for t, we obtain an upper bound for $t_{\text{mix},k}$:

$$t_{\text{mix},k}(\alpha)\leq t=\max_{i\in[N]}(H+1)\left(\frac{-\log\alpha}{-\log\left[1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H\rho^i\right]}+1\right)$$

$$= (H+1)\left(\frac{-\log\alpha}{-\log\left[1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho^i}{|\mathcal{A}^i|}\right)^H\cdot\min_{i\in[N]}\rho^i\right]}+1\right). \quad (3.57)$$

With $\rho^i=\rho$ for all $i\in[N]$, (3.57) becomes

$$\begin{aligned} t_{\mathrm{mix},k}(\alpha) &\le (H+1)\left(\frac{-\log\alpha}{-\log\left[1-|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho}{|\mathcal{A}^i|}\right)^H\cdot\rho\right]}+1\right)\\ &\approx (H+1)\left(\frac{-\log\alpha}{|\mathcal{S}|\kappa\left(\prod_{i\in[N]}\frac{\rho}{|\mathcal{A}^i|}\right)^H\cdot\rho}+1\right)\\ &= (H+1)\left((-\log\alpha)\frac{A^{NH}}{|\mathcal{S}|\kappa\rho^{NH+1}}+1\right). \end{aligned}$$

This completes the proof of (3.55).

Combining (3.52) and (3.55), the proof of (3.16c) is completed.

We have thus finished the proof of Proposition 3.2.

3.3.3 Numerical Experiments: A Grid-World Game

To demonstrate the effectiveness of Algorithm 3.1, we test it on the classical grid-world experiment (Sutton and Barto [207]). We consider a 3×3 grid. Each state is represented by a pair $(s(1),s(2))$, where $s(1)$ is the "x"-coordinate and $s(2)$ is the "y"-coordinate on the grid. The set of states are $\mathcal{S}=\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$ (as shown in Figure 3.2). The upper-left state $(1,1)$ is called the *terminal state*: once the system reaches state $(1,1)$, it will stay there and will not transit to other states. There are two agents, where agent 1 decides its actions in the vertical direction, and agent 2 decides its actions in the horizontal direction, i.e., $\mathcal{A}^1=\{\text{up},\text{stay},\text{down}\}$, $\mathcal{A}^2=\{\text{left},\text{stay},\text{right}\}$. The joint actions of agent 1 and agent 2 will determine the next state, which could be the

same as the current state (if both agents stay) or move one cell in the respective direction on the grid (except when the system is at the terminal state, it will not transit to other states for any joint actions taken by the agents). To ensure the next state stays within the grid, the agents can only choose actions which do not point to a direction outside of the grid. For instance, at state $(1,1)$, $(1,2)$ and $(1,3)$, agent 1 can only choose "down" or "stay", but not "up". Thus, for each agent, there are 6 states where the agent can take two actions, and 3 states where the agent can take three actions, which implies that we have 1728×1728 possible joint policies in total (for each of agent, $|\Pi^i| = 2^6 \times 3^3 = 1728$). When the system is at the terminal state, all actions yield a reward of 0. When the system is at any non-terminal state, all the actions yield a reward of -1.

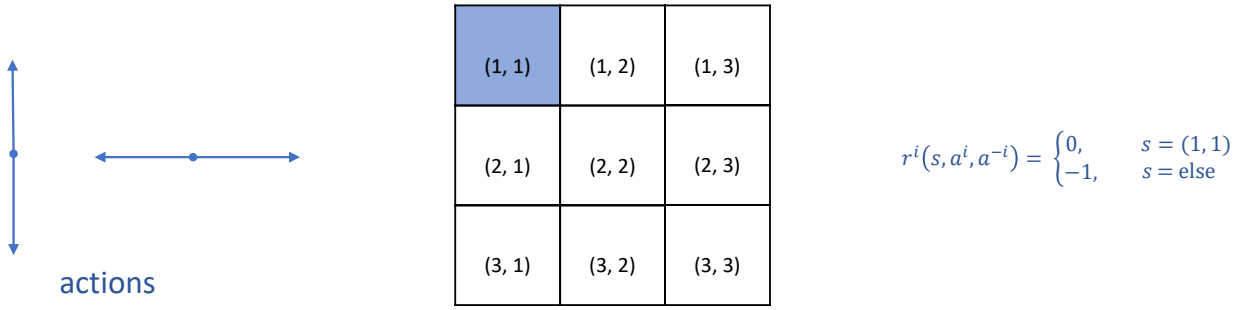

Figure 3.2: Illustration for the grid-world experiment.

For this grid-world stochastic game, if not considering actions taken on the terminal state, there are 16 *optimal* equilibrium joint policies (as shown in Figure 3.3). These optimal equilibria will lead the agents to reach the terminal state as soon as possible and produce the maximum value of reward. Besides these optimal equilibria, there are also some *suboptimal* equilibria. For example, one suboptimal equilibrium is when agent 1 chooses "down" for all of the non-terminal states while agent 2 chooses "right" for all of these states. This joint policy is still a Markov perfect equilibrium, but results in both agents earning less rewards (relative to those of the optimal equilibria). To obtain the set of all equilibrium joint policies, for each $\pi^{-i} \in \Pi^{-i}$, we use the standard Q-learning algorithm (3.9) to get the set of its best reply policies $\Pi^i_{\pi^{-i}}$. Then, for each best reply policy $\pi^{*i} \in \Pi^i_{\pi^{-i}}$, we check if π^{-i} also turns

out to be a best reply policy of π^{*i}. If so, then the joint policy $(\pi^{-i},\ \pi^{*i})$ constitutes a Markov perfect equilibrium. Moreover, if we start from any joint policy in $\Pi^1 \times \Pi^2$, there is a strict best reply path to one of these equilibria, which means the grid-world stochastic game is weakly acyclic under strict best replies.

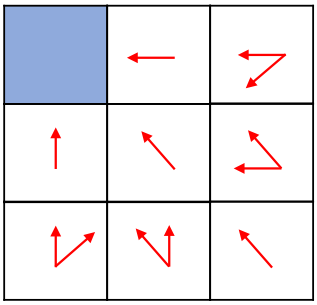

Figure 3.3: The optimal equilibria for the grid-world experiment.

In our numerical experiments, we let $\rho^i = 0.4$, $\lambda^i = 0.3$, $\gamma^i = 0.75$, $\eta_k^i = 1/k^{0.5}$ for all i and k. We fix the length of the exploration phases (the inner "for" loop of Algorithm 3.1), i.e., $T_k = T$, $\forall k = 1, \ldots, K$. Since both the length of the exploration phases T and the number of policy updates K are critical for the learning process, we run our experiments with different parameters of T and K. For each set of K and T, we start from a random joint policy in $\Pi^1 \times \Pi^2$ and run Algorithm 3.1, with the initial values of the Q-table set to all 0's. Then, we have a set of joint policies $\{\pi_k \mid k = 1, \ldots, K\}$, and record the fraction of these policies that belong to the set of equilibrium joint policies obtained before. This process is repeated 50 times, and the experimental results are shown in Figure 3.4. The solid lines represent the fraction of π_k's which are equilibrium joint policies (the number of π_k's which are equilibrium joint policies divided by K), averaged over 50 repeated runs, and the shaded region represents the min-max interval.

It can be observed from Figure 3.4a that the minimum of the fraction $\frac{1}{K}\sum_{k=1}^{K} \mathbf{I}_{\{\pi_k \in \Pi_{eq}\}}$ over the 50 repeated runs is greater than 0 when $K \geq 20$. This implies that the joint policy converges to an equilibrium at the end of the algorithm for all given initial policies. A similar

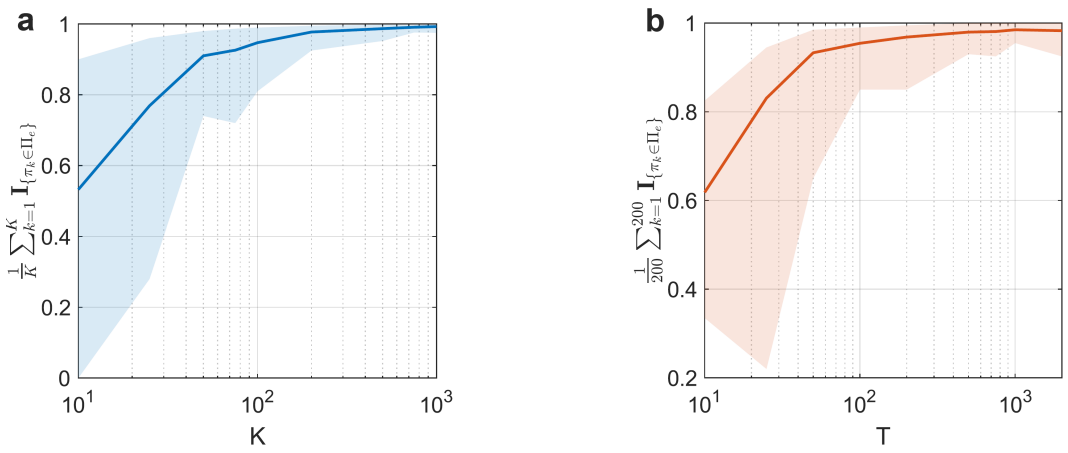

Figure 3.4: Experimental results of grid-world when **Algorithm 3.1** is applied. **a**, The fraction of times at which π_k visits an equilibrium when K ranges in the interval [10, 1000]. **b**, The fraction of times at which π_k visits an equilibrium when T ranges in the interval [10, 2000]. In **a**, we fix the value of T as 200. In **b**, we fix the value of K as 200. The solid lines are the average of 50 repeated runs, and the shaded regions represent the min-max intervals.

phenomenon can also be observed from Figure 3.4b. Meanwhile, it can be seen that π_k visits an equilibrium more often as K and T increase. This is consistent with Theorem 3.1, as π_K is expected to be at equilibrium with higher probability as K and T increase.

3.4 Decentralized Q-learning with Linear Function Approximation

Algorithm 3.1 in the previous section extends the single agent Q-learning algorithm. Specifically, in each exploration phase (an inner "for" loop of Algorithm 3.1), each agent updates and keeps track of its Q function which has dimension $|\mathcal{S}||\mathcal{A}^i|$. One major challenge of such an algorithm is the curse of dimensionality – when the number of state-action pairs is large, it becomes intractable. One popular approach to overcome this obstacle is to approximate the optimal Q functions with functions from a much smaller space. We next describe the linear function approximation method where each agent's Q function is approximated by a linear combination of d basis functions. To the best of our knowledge, there is no existing result on either the convergence or the sample complexity of decentralized Q-learning with

linear function approximation for general-sum stochastic games.

Let $\left\{\phi_j^i : \mathcal{S} \times \mathcal{A}^i \mapsto \mathbb{R} \mid 1 \leq j \leq d\right\}$ be the set of basis functions (features) of agent i. We denote by $\phi^i := \left[\phi_1^i, \phi_2^i, \ldots, \phi_d^i\right] \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}^i| \times d}$ the feature matrix of agent i. The linear subspace $\mathcal{W}^i$ spanned by the features $\left\{\phi_j^i\right\}$ is $\mathcal{W}^i = \left\{Q_{\theta^i}^i := {\phi^i}^\top \theta^i \mid \theta^i \in \Theta^i \subset \mathbb{R}^d\right\}$ where Θ^i is some compact subset of $\mathbb{R}^d$ which contains the zero point and has diameter D^i, i.e., $D^i = \sup\left\{\left\|\theta_j^i - \theta_{j'}^i\right\|_2 \;\middle|\; \theta_j^i, \theta_{j'}^i \in \Theta^i\right\}$. We use $\mathcal{W}^i$ as the approximation function space for agent i.

With the linear function class, we start with approximating agent i's optimal Q-function satisfying the fixed point equation of the Bellman operator, as given in (3.4). Specifically, for any joint policy played by all other agents, $\pi^{-i} \in \Delta^{-i}$, we define

$$\theta_{\pi^{-i}}^i := \underset{\theta^i \in \Theta^i}{\arg\min} \left\| Q_{\pi^{-i}}^i - {\phi^i}^\top \theta^i \right\|_2^2, \tag{3.58}$$

where we recall that $Q_{\pi^{-i}}^i$ satisfies (3.4). Agent i's set of deterministic best replies to π^{-i} under linear function approximation is then given by

$$\widetilde{\Pi}_{\pi^{-i}}^i = \left\{\tilde{\pi}^i \in \Pi^i : \; {\phi^i}\left(s, \tilde{\pi}^i(s)\right)^\top \theta_{\pi^{-i}}^i = \max_{a^i \in \mathcal{A}^i} \phi^i(s, a^i)^\top \theta_{\pi^{-i}}^i, \quad \forall s \in \mathcal{S}\right\}. \tag{3.59}$$

With (3.59), we now define the (deterministic) linear approximated equilibrium.

Definition 3.5. *A deterministic joint policy $\pi^* \in \Pi$ is a linear approximated equilibrium if*

$$\pi^{*i} \in \widetilde{\Pi}_{(\pi^*)^{-i}}^i, \quad \forall i \in [N].$$

We denote by $\widetilde{\Pi}_{\text{eq}}$ the set of linear approximated equilibria. Our goal is to find a set of $\left\{\theta_{(\pi^*)^{-i}}^i \mid \forall i \in [N]\right\}$ for some $\pi^* \in \widetilde{\Pi}_{\text{eq}}$ such that ${\phi^i}^\top \theta_{(\pi^*)^{-i}}^i$ best represents $Q_{(\pi^*)^{-i}}^i$ in the sense of (3.58) for all $i \in [N]$. For this goal to be feasible, we assume that $\widetilde{\Pi}_{\text{eq}}$ is nonempty.

Assumption 3.3. *There exists at least one deterministic joint policy that is a linear ap-*

proximated equilibrium.

Similar to the tabular case, we can define the *best reply graph* on the set of deterministic joint policies, where each vertex is a deterministic joint policy and there is a direct edge from π_k to π_l if for some $i \in [N]$, $\pi_l^i \neq \pi_k^i$, $\pi_l^i = \pi_k^i, \forall j \neq i$, and $\pi_l^i \in \widetilde{\Pi}^i_{\pi_k^{-i}}$. The *strict best reply path* and the *weakly acyclic game* are defined analogously as in Definition 3.3 and Definition 3.4. We again consider the weakly acyclic game, with the newly defined best replies as in (3.59). There exists a strict best reply path from any $\pi \in \Pi$ to some $\pi^* \in \widetilde{\Pi}_{\text{eq}}$. Let $\widetilde{L}_\pi$ be the minimum length of the strict best reply paths from π to a linear approximated equilibrium policy, and let $\widetilde{L} := \max_{\pi \in \Pi} \widetilde{L}_\pi$. We will again apply the BRPI, where the deterministic best reply sets $\widetilde{\Pi}^i_{\pi^{-i}}$ are approximated using Q-learning with linear function approximations.

In the fully decentralized setting, each agent is completely oblivious to other agents. Agent i may use the standard Q-learning algorithm under linear function approximation (see Bertsekas and Tsitsiklis [34], Melo et al. [159]), i.e.,

$$\theta_{t+1}^i = \theta_t^i + \eta_t^i \phi^i(s_t, a_t^i) \left[r^i(s_t, a_t^i, a_t^{-i}) + \gamma^i \max_{a^i \in \mathcal{A}^i} \phi^i(s_{t+1}, a^i)^\top \theta_t^i - \phi^i(s_t, a_t^i)^\top \theta_t^i \right], \quad (3.60)$$

and selects its actions by taking $a_t^i = \arg\max_{a^i \in \mathcal{A}^i} \phi^i(s, a^i)^\top \theta^i$ with high probability and randomly exploring any actions with some small probability. The same problem as in the tabular setting arises here: if all agents use (3.60) and select their actions with the ϵ-greedy method, the environment becomes nonstationary and the convergence of the θ^i's is not guaranteed. In the same spirit of Algorithm 3.1, we let each agent play the behavior policy $\bar{\pi}_k^i$ as defined in (3.10) during the kth exploration phase, so that the environment is stationary within each exploration phase. Instead of maintaining and updating a $|\mathcal{S}||\mathcal{A}^i|$-dimensional Q function, agent i updates a d-dimensional vector θ^i according to (3.60). Note that (3.60) may also be viewed as the stochastic approximation algorithm for solving the following

equation:

$$\mathbb{E}_{\mu^i_{\bar{\pi}_k}}\left[\phi^i(s,a^i)\left(r^i(s,a^i,a^{-i})+\gamma^i\max_{\hat{a}^i\in\mathcal{A}^i}\phi^i(s',\hat{a}^i)^\top\theta^i-\phi^i(s,a^i)^\top\theta^i\right)\right]=0. \tag{3.61}$$

The complete decentralized Q-learning algorithm with linear function approximation is presented as Algorithm 3.2. In essence, we replace the update of Q-functions in Algorithm 3.1 with (3.60) as in line 8 of Algorithm 3.2. Furthermore, in determining whether to update the current baseline policy π^i_k, we use the best reply set $\widetilde{\Pi}^i_{k+1}$, which is defined with $\phi^i_{t_{k+1}}(s,a^i)^\top\theta^i_{t_{k+1}}$, in replacement of the set Π^i_{k+1}.

We will show in this section the finite-sample convergence guarantee of Algorithm 3.2. To proceed, we first impose the following assumptions for all agents, in addition to Assumptions 3.1, 3.2, and 3.3.

Assumption 3.4. *The features $\{\phi^i_j\}_{1\le j\le d}$ are linearly independent and are normalized so that $\|\phi^i(s,a^i)\|\le 1$ for all state-action pairs (s,a^i).*

Assumption 3.4 is imposed without loss of generality (Chen et al. [51]): we can always scale the basis functions to ensure that $\max_{(s,a^i)\in\mathcal{S}\times\mathcal{A}^i}\|\phi^i(s,a^i)\|\le 1$, and any dependent features can be discarded.

Assumption 3.5. (3.61) *has a unique solution, and there exists $\xi^i>0$ such that the following inequality holds for all $\theta^i\in\Theta^i$:*

$$(\gamma^i)^2\mathbb{E}_\mu\left[\max_{a^i\in\mathcal{A}^i}(\phi^i(s,a^i)^\top\theta^i)^2\right]-\mathbb{E}_\mu\left[\left(\phi^i(s,a^i)^\top\theta^i\right)^2\right]\le\xi^i\|\theta^i\|_2^2. \tag{3.62}$$

We note that (3.61) may not admit a solution in general, and the iteration for θ^i_t in (3.60) might diverge. Assumption 3.5, which is exactly the same as Assumption 3.3 in Chen et al. [51], ensures the convergence of (3.60). See Chen et al. [51], Lee and He [136], Melo et al. [159] for detailed discussions on this point.

Algorithm 3.2 Q-learning for agent i with linear function approximation

Set parameters

Θ^i: some compact subset of the Euclidian space $\mathbb{R}^d$ with diameter D^i

$\{T_k\}_{k\geq 0}$: sequence of integers in $[1,\infty)$, the length of the kth exploration phase

$K \in \mathbb{Z}_+$: number of exploration phases

$\rho^i \in (0,1)$: experimentation probability

$\lambda^i \in (0,1)$: inertia

$\zeta^i \in (0,\infty)$: tolerance level for sub-optimality

$\{\eta_t^i\}_{t\geq 0}$: sequence of step sizes

Initialize $\pi_0^i \in \Pi^i$ (arbitrary), $\theta_0^i \in \mathbb{R}^d$ (arbitrary)

Receive s_0

for $k = 1, 2 \ldots$ **do**

  for $t = t_k, \ldots, t_{k+1} - 1$ **do**

    $a_t^i = \bar{\pi}_k^i(s_t) := \begin{cases} \pi_k^i(s_t), & \text{w.p. } 1-\rho^i \\ \text{any } a^i \in \mathcal{A}^i, & \text{w.p. } \rho^i/|\mathcal{A}^i| \end{cases}$

    Receive $r^i(s_t, a_t^i, a_t^{-i})$

    Receive s_{t+1} (selected according to $P[\,\cdot \mid s_t, a_t^i, a_t^{-i}]$)

    $\theta_{t+1}^i = \theta_t^i + \eta_t^i \phi(s_t, a_t^i) \left[r^i(s_t, a_t^i, a_t^{-i}) + \gamma^i \max_{a^i \in \mathcal{A}^i} \phi^i(s_{t+1}, a^i)^\top \theta_t^i - \phi(s_t, a_t^i)^\top \theta_t^i \right]$

  end for

  $\widetilde{\Pi}_{k+1}^i = \left\{ \tilde{\pi}^i \in \Pi^i : \phi_{t_{k+1}}^i(s, \tilde{\pi}^i(s))^\top \theta_{t_{k+1}}^i \geq \max_{a^i \in \mathcal{A}^i} \phi_{t_{k+1}}^i(s, a^i)^\top \theta_{t_{k+1}}^i - \frac{1}{2}\zeta_\theta^i,\ \forall s \right\}$

  if $\pi_k^i \in \widetilde{\Pi}_{k+1}^i$, **then**

    $\pi_{k+1}^i = \pi_k^i$

  else

    $\pi_{k+1}^i = \begin{cases} \pi_k^i, & \text{w.p. } \lambda^i \\ \text{any } \pi^i \in \widetilde{\Pi}_{k+1}^i, & \text{w.p. } (1-\lambda^i)/|\widetilde{\Pi}_{k+1}^i| \end{cases}$

  end if

  $\theta_{t_{k+1}}^i \leftarrow$ projection of $\theta_{t_{k+1}}^i$ onto Θ^i

end for

Let b be an upper bound on the minimum Bellman error under joint policy $\pi \in \Pi \cup \bar{\Pi}$, i.e.,

$$\min_{\theta^i \in \Theta^i} \left\| Q_{\theta^i}^i - \mathcal{T}_{\pi^{-i}}^i \left(Q_{\theta^i}^i \right) \right\|_2 \leq b, \quad \forall i \in [N], \pi^{-i} \in \Pi^{-i} \cup \bar{\Pi}^{-i}. \tag{3.63}$$

Similar to (3.13) where $\bar{\zeta}$ is defined, we define the minimum separation between the linear

approximated Q-functions:

$$\bar{\zeta}_\theta := \min_{\substack{i,s,a^i,\tilde{a}^i,\pi^{-i}\in\Pi^{-i}:\\ Q^i_{\theta^i_{\pi^{-i}}}(s,a^i)\neq Q^i_{\theta^i_{\pi^{-i}}}(s,\tilde{a}^i)}} \left| Q^i_{\theta^i_{\pi^{-i}}}(s,a^i) - Q^i_{\theta^i_{\pi^{-i}}}(s,\tilde{a}^i) \right|. \tag{3.64}$$

We next have the following assumption on the bound of minimum Bellman error.

Assumption 3.6. *The upper bound b on the minimum Bellman error as given in* (3.63) *satisfies* $b < \frac{(1-\bar{\gamma})\bar{\zeta}_\theta}{8}$.

Assumption 3.6 bounds the minimum Bellman error from a constant factor of the minimum separation between the linear approximated Q-factors. This assumption is necessary to achieve a good approximation of $\widetilde{\Pi}^i_{\pi_k^{-i}}$ by $\widetilde{\Pi}^i_{k+1}$ in line 10 of Algorithm 3.2, so that the algorithm closely mimics the BRPI. Without Assumption 3.6, there is no guarantee that a policy that is not a linear approximated equilibrium will be updated, i.e., an agent may not update its current policy even if it is not a best reply.

For notational convenience, we let $D := \max_{i\in[N]} D^i$, $\xi_{\min} := \min_{i\in[N]} \xi^i$, and $\eta_{\min} = \min_{i,t} \eta^i_t$. With the above definitions and assumptions, we now present our second main theorem on the sample complexity of Algorithm 3.2, whose proof is given in Appendix 3.6.1.

Theorem 3.2. *Consider a discounted stochastic game that is weakly acyclic under strict best replies* (3.59). *Suppose that each agent updates its policies by Algorithm 3.2. Let Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6 hold. Then, for any $0 < \delta < 1$, one has that for all $k \geq K$,*

$$P\left[\pi_k \in \widetilde{\Pi}_{\text{eq}}\right] \geq 1 - \delta$$

provided that for all $i \in [N]$ *and* $k \in [K]$,

$$\eta_t^i = \eta^i \leq \frac{\epsilon^2 \tilde{\delta} \xi^i}{456 N \widetilde{L} \left(1 + \gamma^i + r_{\max}^i\right)^2 (D^i + 1)^2 t_{\mathrm{mix},k}^i(\eta^i)}, \quad \forall t = t_k, \ldots, t_{k+1} - 1, \tag{3.65a}$$

$$T_k \geq t_{\mathrm{mix},k}(\eta_{\min}) + \frac{\log \frac{\epsilon^2 \tilde{\delta}}{2N\widetilde{L}(2D+1)^2}}{\log\left(1 - \xi_{\min}\eta_{\min}/2\right)}, \tag{3.65b}$$

$$K \geq \frac{\left[\left(1 - \tilde{\delta}\right)^2 \tilde{p} - \tilde{\delta}^2\right] \widetilde{L}}{\left[\tilde{\delta} + \left(1 - \tilde{\delta}\right)\tilde{p}\right]^2 \tilde{\delta}}, \tag{3.65c}$$

$$\rho^i \leq 1 - \left(1 - \frac{\left(\bar{\zeta}_\theta/8 - \epsilon\right)(1 - \bar{\gamma}) - 2b}{\widetilde{\Gamma}}\right)^{\frac{1}{N-1}} := \rho, \tag{3.65d}$$

$$\zeta_\theta^i = \frac{\bar{\zeta}_\theta}{2}, \tag{3.65e}$$

where $\widetilde{\Gamma}$ *is some absolute constant which depends only on the game parameters (formally defined in* (3.72)*),* $\tilde{p} := \hat{p}^{\widetilde{L}/L}$, $\epsilon := \min\left\{\frac{\bar{\zeta}_\theta}{16} - \frac{b}{1-\bar{\gamma}}, \frac{1}{2(1-\underline{\gamma})}\right\}$, *and* $\tilde{\delta}$ *is such that*

$$\delta = 1 - \left(\frac{\left(1 - \tilde{\delta}\right)\tilde{p}}{\tilde{\delta} + \left(1 - \tilde{\delta}\right)\tilde{p}} - \tilde{\delta}\right)\left(1 - \tilde{\delta}\right).$$

Corollary 3.2. *Recall from Proposition 3.2 that*

$$t_{\mathrm{mix},k}(\alpha) \leq (H+1)\left((-\log\alpha)\frac{A^{NH}}{\kappa\rho^{NH}} \cdot \min\left\{1, |\mathcal{S}|^{-1}\rho^{-1}\right\} + 1\right).$$

By applying this upper bound on the mixing time to Theorem 3.2, we may express the η_t^i *(*$\forall t = t_k, \ldots, t_{k+1} - 1$*) and* T_k *in Theorem 3.2 as*

$$\eta_t^i \leq \frac{\epsilon^2 \tilde{\delta} \xi^i \kappa \rho^{NH}}{456 N \widetilde{L}\left(1 + \gamma^i + r_{\max}^i\right)^2 (D^i+1)^2 (H+1)\left(\kappa\rho^{NH} - A^{NH} \cdot \min\left\{1, |\mathcal{S}|^{-1}\rho^{-1}\right\} \log \eta^i\right)}, \tag{3.66a}$$

$$T_k \geq (H+1)\left((-\log \eta_{\min})\frac{A^{NH}}{\kappa \rho^{NH}} \cdot \min\left\{1, |\mathcal{S}|^{-1}\rho^{-1}\right\} + 1\right) + \frac{\log \frac{\epsilon^2 \tilde{\delta}}{2N\widetilde{L}(2D+1)^2}}{\log\left(1 - \xi_{\min}\eta_{\min}/2\right)}. \tag{3.66b}$$

We note that Theorem 3.2 provides the sample complexity for the joint deterministic baseline policy π_k to converge to a *linear approximated equilibrium* in $\widetilde{\Pi}_{\text{eq}}$, as defined in Definition 3.5, which may or may not be the equilibrium as defined in Definition 3.1 due to linear approximation. However, for the special case when each agent's optimal Q-function $Q^i_{\bar{\pi}_k^{-i}}$ is realizable in $\mathcal{W}^i$ for all joint behavior policies $\bar{\pi}_k \in \bar{\Pi}$, we will be able to show the convergence of π_k to an equilibrium in Π_{eq}. Formally, we have the following realizability assumption.

Assumption 3.7. *For any joint policy $\bar{\pi}_k$ as in* (3.10)*, agent i's optimal Q-function $Q^i_{\bar{\pi}_k^{-i}}$ is realizable in $\mathcal{W}^i$, i.e., there exists $\theta^i_{\bar{\pi}_k^{-i}} \in \Theta^i \subset \mathbb{R}^d$ such that*

$$Q^i_{\bar{\pi}_k^{-i}}(s, a^i) = \phi^i(s, a^i)^\top \theta^i_{\bar{\pi}_k^{-i}}, \forall (s, a^i) \in \mathcal{S} \times \mathcal{A}^i.$$

With this additional assumption, we arrive at the following result, whose proof can be found in Appendix 3.6.2.

Theorem 3.3. *Consider a discounted stochastic game that is weakly acyclic under strict best replies* (3.5)*. Suppose that each agent updates its policies by Algorithm 3.2. Let Assumptions 3.1, 3.2, 3.4, 3.5, and 3.7 hold. Then, for any $0 < \delta < 1$, one has for all $k \geq K$,*

$$P\left[\pi_k \in \Pi_{\text{eq}}\right] \geq 1 - \delta,$$

provided that for all $i \in [N]$ and $k \in [K]$,

$$\eta^i_t = \eta^i \leq \frac{\epsilon^2 \tilde{\delta} \xi^i}{456NL\left(1 + \gamma^i + r^i_{\max}\right)^2 (D^i+1)^2 t^i_{\text{mix},k}(\eta^i)}, \quad \forall t = t_k, \ldots, t_{k+1} - 1, \tag{3.67a}$$

$$T_k \geq t_{\text{mix},k}(\eta_{\min}) + \frac{\log \frac{\epsilon^2 \tilde{\delta}}{2NL(2D+1)^2}}{\log\left(1 - \xi_{\min}\eta_{\min}/2\right)}, \tag{3.67b}$$

$$K \geq \frac{\left[\left(1-\tilde{\delta}\right)^2 \hat{p} - \tilde{\delta}^2\right] L}{\left[\tilde{\delta} + \left(1-\tilde{\delta}\right)\hat{p}\right]^2 \tilde{\delta}} \tag{3.67c}$$

$$\rho^i = 1 - \left(1 - \frac{(\bar{\zeta}/2 - \epsilon)(1-\bar{\gamma})}{\Gamma}\right)^{\frac{1}{N-1}} \tag{3.67d}$$

$$\zeta^i = \frac{\bar{\zeta}}{2} \tag{3.67e}$$

where Γ, $\bar{\zeta}$ *and* $\hat{p}$ *are absolute constants as defined in* (3.22), (3.13) *and* (3.27), *respectively (which depend only on the game parameters),* $\epsilon := \min\left\{\frac{\bar{\zeta}}{16}, \frac{1}{2(1-\underline{\gamma})}\right\}$, *and* $\tilde{\delta}$ *is such that*

$$\delta = 1 - \left(\frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta} + \left(1-\tilde{\delta}\right)\tilde{p}} - \tilde{\delta}\right)\left(1-\tilde{\delta}\right).$$

3.4.1 Numerical Experiments

Similar to the tabular cases, we use again the grid-world stochastic game to demonstrate the effectiveness of Algorithm 3.2. All of the details about the agents, states, actions, reward, and parameters in this setting are the same as those in Section 3.3.3. We construct the feature vectors using a polynomial basis (Sutton and Barto [207]). Specifically, we use order-3 polynomial-basis features, where each feature can be written as

$$\phi_j^i(s,a) = s(1)^{c_{1,j}} s(2)^{c_{2,j}} a^{c_{3,j}},$$

where $j = 1, \ldots, d$, each $c_{k,j}$ is an integer in the set $\{0,1,2,3\}$. Since $|\mathcal{S} \times \mathcal{A}^i| = 21$ (for both agents) in this stochastic game, we choose $d = 18$. To obtain the set of linear approximated equilibria, for each $\pi^{-i} \in \Pi^{-i}$ $(i = 1, 2)$, we first compute $\theta^i_{\pi^{-i}}$ by solving the quadratic programming problem (3.58), and deduce the set of best reply policies under linear function

approximation $\widetilde{\Pi}^i_{\pi^{-i}}$ from (3.59). Then, for each best reply policy $\tilde{\pi}^i \in \widetilde{\Pi}^i_{\pi^{-i}}$, we check if $\tilde{\pi}^i$ is also a best reply policy of π^{-i} under linear function approximation. If so, then the joint policy $(\pi^{-i}, \tilde{\pi}^i)$ is a linear approximated equilibrium.

As in Section 3.3.3, we also run experiments with different parameters of K and T, and the process is repeated 50 times. The experimental results are presented in Figure 3.5. We can see a similar phenomenon as in the tabular case. It can be seen from Figure 3.5a that the minimum of the fraction $\frac{1}{K}\sum_{k=1}^{K} \mathbf{I}_{\{\pi_k \in \Pi_{eq}\}}$ of the 50 repeated runs is greater than 0 when $K \geq 100$. It can also be seen from Figure 3.5b that the minimum of $\frac{1}{K}\sum_{k=1}^{K} \mathbf{I}_{\{\pi_k \in \Pi_{eq}\}}$ is greater than 0 when $T \geq 200$. This implies that the joint policy converges to a linear approximated equilibrium at the end of the algorithm for all given initial policies. Moreover, it can be seen that π_k visits a linear approximated equilibrium more often as K and T increase. This is consistent with Theorem 3.2, as π_K is expected to be at a linear approximated equilibrium with higher probability as K and T increase.

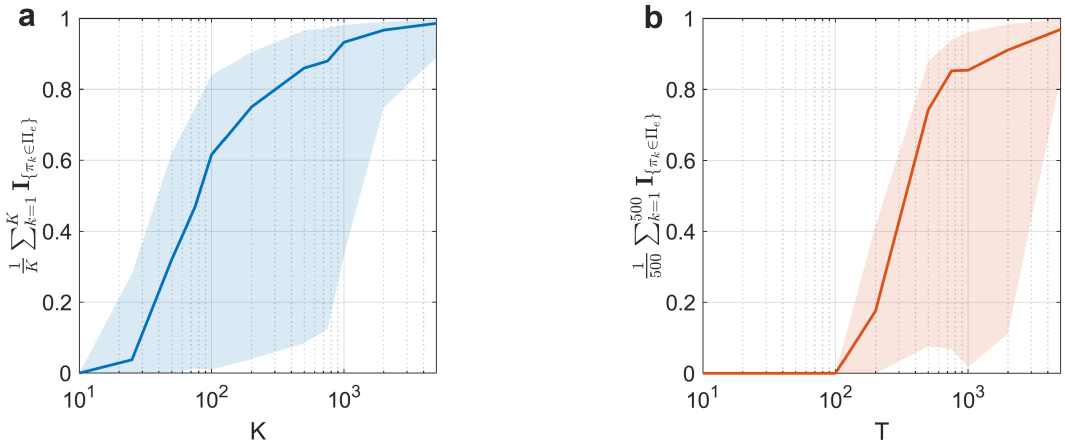

Figure 3.5: Experimental results of grid-world when **Algorithm 3.2** is applied. **a**, The fraction of times at which π_k visits an equilibrium when K ranges in the interval [10, 5000]. **b**, The fraction of times at which π_k visits an equilibrium when T ranges in the interval [10, 5000]. In **a**, we fix the value of T as 1000. In **b**, we fix the value of K as 500. The solid lines are the average of 50 runs, and the shaded regions represent the min-max interval.

3.5 Conclusions and Future Work

This chapter was aimed at deriving sample complexity results on decentralized Q-learning algorithms for a class of general-sum stochastic games – weakly acyclic games. The main takeaways of this chapter can be summarized as follows. First, we have established finite-sample guarantees for Algorithm 3.1, whose asymptotic convergence was shown earlier in Arslan and Yüksel [21]. Second, we have imposed linear function approximation to the algorithm (as Algorithm 3.2), and provided finite-sample analysis for its convergence to a linear approximated equilibrium – a new notion of equilibrium that we have introduced.

Regarding the first point, we note that there are some nontrivial generalizations from the asymptotic convergence to finite-sample analysis of Algorithm 3.1. One example is in the proof of Lemma 3.5, where we have to ensure that Π^i_{k+1} well approximates $\Pi^i_{\pi_k^{-i}}$, so that any non-equilibrium policy would not stop updating before converging to equilibrium (the "if-else" statement in Algorithm 1). Under the linear approximation setting, the behavior that $Q^i_{\theta^i_{k+1}}$ and $Q^i_{\theta^i_{\pi_k^{-i}}}$ are separated by a distance related to the minimum Bellman error (Lemma 11) adds considerable complications to the analysis, as it may hinder the algorithm from updating a linear approximated non-equilibrium policy. We address this issue in the proof of Lemma 3.14. In addition, Lemmas 3.5 and 3.14 provide closed-form expressions for the lower bounds $\hat{p}$ and $\tilde{p}$, which did not exist in Arslan and Yüksel [21] but are crucial for developing the finite-sample guarantees. Another example of new development lies in Proposition 3.2. The notions of the minimum probability of stationary distribution $\mu_{\min}$ and the mixing time t_{mix} that appeared in the sample complexity results seem implicit, while Proposition 3.2 bounds $\mu_{\min}$ and t_{mix} under the current game set up, and thus the results in both theorems can be expressed explicitly in terms of the game parameters. We further note that, even with these (and other) developments, there is room for improvement on the results. For instance, in the chapter we picked ϵ as the middle point of its possible range, while it remains open to optimize over ϵ in (3.31) while keeping it simple to obtain a

tighter bound on T_k.

Regarding the second point, we note that in the linear approximated equilibrium, each agent's policy is a best reply (to other agents' joint policy) within the linear space spanned by the set of features (basis functions). When the dimension of the feature set is large enough so that the original Q-functions can be fully recovered for all state-action pairs, each linear approximated equilibrium is naturally also a Markov perfect equilibrium and vice versa. When we have a smaller feature set, the relationship between linear approximated equilibria and Markov perfect equilibria can be general: a joint policy can be both a linear approximated equilibrium and a Markov perfect equilibrium, or it can only be one of these equilibria but not the other one. Fixing a certain (small) number of features, it would be interesting to investigate the question of how to select features so that the set of linear approximated equilibria overlaps the most with the set of Markov perfect equilibria.

We further note that, while we have derived the sample complexities of Algorithm 3.1 and Algorithm 3.2 for converging to a Markov perfect equilibrium and a linear approximated equilibrium, respectively, it is unknown which equilibrium these algorithms converge to. In other words, a weakly acyclic stochastic game may possess multiple (linear approximated) equilibria and some equilibria could be strictly better than others for all agents. Most recently, Yongacoglu et al. [227] proposed learning algorithms that converge to optimal equilibria for stochastic teams and common interest games. Moreover, Yongacoglu et al. [228] eliminated the weakly acyclic assumption (existence of a best-reply path from any policy to an equilibrium policy) but introduced the ϵ-satisficing property (which is shown to hold for two-player games and N-player symmetric games), and proposed an algorithm that guarantees convergence of the joint baseline policy to an ϵ-equilibrium. Sayin and Unlu [187] proved the convergence of logit-Q learning in infinite-horizon discounted identical-interest Markov games. Building upon these recent works, designing algorithms (and analyzing their sample complexities) that converge to different types of euilibira for stochastic games with

various assumptions could be promising future research directions.

3.6 Appendix

3.6.1 Proof of Theorem 3.2

We first introduce the following lemma, which is an application of the sample complexity result on the convergence of θ for single agent Q-learning (Chen et al. [51]).

Lemma 3.11. *Fix any arbitrary $\pi_k \in \Pi$. For any $\epsilon \geq 0$ and $0 < \hat{\delta} < 1$, we have that*

$$P\left[\left|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right|_\infty \leq \epsilon,\ \forall i \in [N]\right] \geq 1 - \hat{\delta},$$

provided that the iteration number T_k and the learning rates η^i_t obey

$$\eta^i_t = \eta^i \leq \frac{\epsilon^2 \hat{\delta} \xi^i}{456N\left(1+\gamma^i+r^i_{\max}\right)^2 (D^i+1)^2 t^i_{\mathrm{mix},k}(\eta^i)}, \quad \forall t = t_k, \ldots, t_{k+1}-1, \forall i \in [N], \tag{3.68a}$$

$$T_k \geq t_{\mathrm{mix},k}(\eta_{\min}) + \frac{\log \frac{\epsilon^2 \hat{\delta}}{2N(2D+1)^2}}{\log\left(1-\xi_{\min}\eta_{\min}/2\right)}, \tag{3.68b}$$

where $D = \max_i D^i$, $\eta_{\min} = \min_i \eta^i$, and $\xi_{\min} = \min_i \xi^i$.

Proof of Lemma 3.11. Note that in the kth exploration phase, agents adopt the joint policy $\bar{\pi}_k$ as defined in (3.10). Under Assumptions 3.1, 3.2, 3.4, and 3.5, Theorem 3.1 of Chen et al. [51] implies that for any agent i,

$$\mathbb{E}\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2^2\right] \leq \beta_1\left(1-\xi^i\eta^i_t/2\right)^{T_k - t^i_{\mathrm{mix},k}(\eta^i_t)} + 2\beta_2\eta^i_t t^i_{\mathrm{mix},k}(\eta^i_t)/\xi^i,$$

where $\beta_1 = (2D^i+1)^2$, $\beta_2 = 114\left(1+\gamma^i+r^i_{\max}\right)^2(D^i+1)^2$, and $\eta^i_t = \eta^i$ are such that

$\eta^i \leq \frac{\xi^i}{228\left(1+\gamma^i+r^i_{\max}\right)^2 t^i_{\text{mix},k}(\eta^i)}$.

For any given $\epsilon > 0$ and $0 < \delta_0 < 1$, let η^i_t and T_k be such that

$$\eta^i_t = \eta^i \leq \frac{\epsilon^2 \delta_0 \xi^i}{456\left(1+\gamma^i+r^i_{\max}\right)^2 (D^i+1)^2 t^i_{\text{mix},k}(\eta^i)}, \quad \forall t = t_k, \ldots, t_{k+1}-1 \tag{3.69a}$$

$$T_k \geq t^i_{\text{mix},k}(\eta^i) + \frac{\log \frac{\epsilon^2\delta_0}{2(2D^i+1)^2}}{\log\left(1-\xi^i\eta^i/2\right)}, \quad \forall i \in [N]. \tag{3.69b}$$

Then, we have that $\forall i \in [N]$,

$$\begin{aligned}\mathbb{E}\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2^2\right] &\leq \beta_1\left(1-\xi^i\eta^i_t/2\right)^{T_k - t^i_{\text{mix},k}(\eta^i_t)} + 2\beta_2\eta^i_t t^i_{\text{mix},k}(\eta^i_t)/\xi^i \\ &\leq \frac{\epsilon^2\delta_0}{2} + \frac{\epsilon^2\delta_0}{2} = \epsilon^2\delta_0.\end{aligned}$$

By Markov inequality, this further implies that

$$P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2^2 \geq \epsilon^2\right] \leq \frac{\mathbb{E}\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2^2\right]}{\epsilon^2} \leq \delta_0, \quad \forall i \in [N],$$

which is equivalent to

$$P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2 \geq \epsilon\right] \leq \frac{\mathbb{E}\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2^2\right]}{\epsilon^2} \leq \delta_0, \quad \forall i \in [N].$$

From the union bound,

$$P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2 \geq \epsilon,\ \exists i \in [N]\right] \leq \sum_{i\in[N]} P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2 \geq \epsilon\right] \leq N\delta_0.$$

Therefore,

$$P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2 \le \epsilon,\ \forall i \in [N]\right] \ge 1 - P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2 \ge \epsilon,\ \exists i \in [N]\right] \ge 1 - N\delta_0.$$

Furthermore, since $\left|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right|_\infty \le \left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2$, we have that

$$P\left[\left|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right|_\infty \le \epsilon,\ \forall i \in [N]\right] \ge P\left[\left\|\theta^i_{t_{k+1}} - \theta^i_{\bar{\pi}_k^{-i}}\right\|_2 \le \epsilon,\ \forall i \in [N]\right] \ge 1 - N\delta_0.$$

By taking $\hat{\delta} = N\delta_0$, η^i_t as in (3.69a), and

$$T_k \ge t_{\text{mix},k}(\eta_{\min}) + \frac{\log \frac{\epsilon^2 \hat{\delta}}{2N(2D+1)^2}}{\log\left(1 - \xi_{\min}\eta_{\min}/2\right)} \ge t^i_{\text{mix},k}(\eta^i) + \frac{\log \frac{\epsilon^2 \delta_0}{2(2D^i+1)^2}}{\log\left(1 - \xi^i\eta^i/2\right)}, \quad \forall i \in [N],$$

the proof is completed. □

Lemma 3.11 bounds the approximation error of $\theta^i_{t_{k+1}}$ for each agent. By noting that $|\phi^i|_\infty \le \|\phi^i\|_2 \le 1$, Lemma 3.11 implies that for an arbitrary $\bar{\pi}_k$ as in (3.10) and for any $\epsilon > 0$ and $0 < \hat{\delta} < 1$,

$$\begin{aligned} P\left[\left|Q^i_{\theta^i_{t_{k+1}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right|_\infty \le \epsilon,\ \forall i \in [N]\right] &= P\left[\left|{\phi^i}^\top \theta^i_{t_{k+1}} - {\phi^i}^\top \theta^i_{\bar{\pi}_k^{-i}}\right|_\infty \le \epsilon,\ \forall i \in [N]\right] \\ &\ge 1 - \hat{\delta} \end{aligned} \tag{3.70}$$

when the same conditions (3.68) are satisfied.

Our next goal is to bound the approximation error of policy perturbation. Recall the definition of the randomized policy in (3.10), and consider the joint policies of all agents except i. With probability $\prod_{j\neq i}(1-\rho^j)$, all agents $j \neq i$ end up playing their baseline policies, which results in $\left|Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right|_\infty = 0$, i.e. the approximation error of policy

perturbation becomes zero in this case. When not all agents play their baseline policies, let $\varphi^{-i} \in \Delta^{-i}$ be some convex combination of the policies in Δ^{-i} of the form where each agent $j \neq i$ either uses a baseline policy $\pi^j \in \Pi^j$ or the uniform distribution. More precisely, let J denote the subset of agents choosing the baseline policies, and let

$$\varphi^{-i} = \sum_{J \subset \{1,\dots,N\} \setminus \{i\}} a_J \varphi_J^{-i}, \tag{3.71}$$

where $a_J := \frac{\prod_{j \in J}(1-\rho^j) \prod_{j \notin J \cup \{i\}} \rho^j}{1 - \prod_{j \neq i}(1-\rho^j)}$ and $\varphi_J \in \Delta^{-i}$ is such that $\varphi_J^j = \pi^j$ for $j \in J$ and $\varphi_J^j = \nu^j$ for $j \notin J \cup \{i\}$. Denote by $\bar{\Delta}^{-i} \subset \Delta^{-i}$ the set of all policies in the form of (3.71). Note that $\bar{\Delta}^{-i}$ is a finite set. We then define

$$\widetilde{\Gamma} := \max_{(\pi^{-i}, \varphi^{-i}) \in \Pi^{-i} \times \bar{\Delta}^{-i}} \left| \mathcal{T}^i_{\pi^{-i}} \left(Q^i_{\theta^i \, \pi^{-i}} \right) - \mathcal{T}^i_{\varphi^{-i}} \left(Q^i_{\theta^i \, \pi^{-i}} \right) \right|_\infty . \tag{3.72}$$

We next have the following lemma on the approximation error due to policy perturbation.

Lemma 3.12. *Fix any arbitrary $\pi_k \in \Pi$. For any $\tilde{\epsilon} > 0$, if ρ^i satisfies*

$$\rho^i \leq 1 - \left(1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\widetilde{\Gamma}} \right)^{\frac{1}{N-1}}, \quad \forall i \in [N], \tag{3.73}$$

then, we have that

$$\left| Q^i_{\theta^i \, \pi_k^{-i}} - Q^i_{\theta^i \, \bar{\pi}_k^{-i}} \right|_\infty = \left| {\phi^i}^\top \theta^i_{\pi_k^{-i}} - {\phi^i}^\top \theta^i_{\bar{\pi}_k^{-i}} \right|_\infty \leq \tilde{\epsilon} + \frac{2b}{1-\bar{\gamma}}, \quad \forall i \in [N], k \in [K].$$

Proof of Lemma 3.12. First note that, for all $i \in [N]$ and $k \in [K]$,

$$\begin{aligned}
&\left| Q^i_{\theta^i \, \pi_k^{-i}} - Q^i_{\theta^i \, \bar{\pi}_k^{-i}} \right|_\infty \\
&= \left| Q^i_{\theta^i \, \pi_k^{-i}} - \mathcal{T}^i_{\pi_k^{-i}} \left(Q^i_{\theta^i \, \pi_k^{-i}} \right) + \mathcal{T}^i_{\pi_k^{-i}} \left(Q^i_{\theta^i \, \pi_k^{-i}} \right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}} \left(Q^i_{\theta^i \, \bar{\pi}_k^{-i}} \right) + \mathcal{T}^i_{\bar{\pi}_k^{-i}} \left(Q^i_{\theta^i \, \bar{\pi}_k^{-i}} \right) - Q^i_{\theta^i \, \bar{\pi}_k^{-i}} \right|_\infty
\end{aligned}$$

$$
\begin{aligned}
&= \left| Q^i_{\theta^i_{\pi_k^{-i}}} - \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) \right|_\infty + \left| \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right) \right|_\infty \\
&\quad + \left| \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right) - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty \\
&\leq b + \left| \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right) \right|_\infty + b \\
&\leq \left| \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) \right|_\infty + \left| \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right) \right|_\infty + 2b. \qquad (3.74)
\end{aligned}
$$

By definition of $\bar{\pi}_k^{-i}$, we have that $P\left[\bar{\pi}_k^{-i} = \pi_k^{-i}\right] = \prod_{j\neq i}(1-\rho^j)$. With probability $1 - \prod_{j\neq i}(1-\rho^j)$, $\bar{\pi}_k^{-i} \neq \pi_k^{-i}$ and $\bar{\pi}_k^{-i} \in \bar{\Delta}^{-i}$. Thus, the first term of (3.74) can be bounded by

$$
\begin{aligned}
&\left| \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) \right|_\infty \\
&\leq \left(1 - \prod_{j\neq i}(1-\rho^j)\right) \times \left| \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\varphi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) \right|_\infty, \qquad (3.75)
\end{aligned}
$$

for some $\phi_k^{-i} \in \bar{\Delta}^{-i}$. On the other hand, by the contraction mapping of the Bellman operator, we have that

$$
\left| \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\bar{\pi}_k^{-i}}\left(Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right) \right|_\infty \leq \gamma^i \left| Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty. \qquad (3.76)
$$

Substituting (3.75) and (3.76) back into (3.74), we have that

$$
\begin{aligned}
\left| Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty \leq &\left(1 - \prod_{j\neq i}(1-\rho^j)\right) \times \left| \mathcal{T}^i_{\pi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) - \mathcal{T}^i_{\varphi_k^{-i}}\left(Q^i_{\theta^i_{\pi_k^{-i}}}\right) \right|_\infty \\
&+ \gamma^i \left| Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty + 2b
\end{aligned}
$$

$$\leq \left(1 - \prod_{j \neq i}(1-\rho^j)\right) \widetilde{\Gamma} + \gamma^i \left| Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty + 2b,$$

which implies that

$$\left| Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty \leq \frac{\left(1 - \prod_{j \neq i}(1-\rho^j)\right) \widetilde{\Gamma} + 2b}{1-\gamma^i} \leq \frac{\left(1 - \prod_{j \neq i}(1-\rho^j)\right) \widetilde{\Gamma} + 2b}{1-\bar{\gamma}}.$$

If for all $i \in [N]$, $\rho^i \leq 1 - \left(1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\widetilde{\Gamma}}\right)^{\frac{1}{N-1}}$, then, we have that $1 - \rho^j \geq \left(1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\widetilde{\Gamma}}\right)^{\frac{1}{N-1}}$, which implies that $\prod_{j \neq i}(1-\rho^j) \geq 1 - \frac{\tilde{\epsilon}(1-\bar{\gamma})}{\widetilde{\Gamma}}$, and thus

$$\left| Q^i_{\theta^i_{\pi_k^{-i}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} \right|_\infty \leq \frac{\left(1 - \prod_{j \neq i}(1-\rho^j)\right) \widetilde{\Gamma} + 2b}{1-\bar{\gamma}} \leq \tilde{\epsilon} + \frac{2b}{1-\bar{\gamma}}.$$

The above holds for all $i \in [N]$ and $k \in [K]$, which completes the proof. □

Recall from (3.64) that $\bar{\zeta}_\theta$ is the minimum separation of agents' optimal linear approximated Q-factors. By Assumption 3.6, we have that $\bar{\zeta}_\theta > \frac{8b}{1-\bar{\gamma}}$. We consider $\bar{\zeta}_\theta$ as an upper bound on ζ^i_θ for all i. We next define the following random event for any arbitrary $\pi_k \in \Pi$:

$$\widetilde{E}_k := \left\{ \omega \in \Omega : \left| Q^i_{\theta^i_{t_{k+1}}} - Q^i_{\theta^i_{\pi_k^{-i}}} \right|_\infty < \frac{1}{4} \min\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\}, \forall i \right\}.$$

With this definition of $\widetilde{E}_k$, we show that, if $\widetilde{E}_k$ is not empty and $\pi_k \in \widetilde{\Pi}_{\text{eq}}$, then $\pi_{k+1} = \pi_k$ with probability 1.

Lemma 3.13. *Given any $\pi_k \in \Pi$ and the corresponding $\widetilde{E}_k$, for all k, we have that*

$$P\left[\pi_{k+1} = \pi_k \mid \widetilde{E}_k,\ \pi_k \in \widetilde{\Pi}_{\text{eq}}\right] = 1.$$

Proof of Lemma 3.13. Let $\hat{a}^{i*} := \arg\max_{\hat{a}^i} Q^i_{t_{k+1}}\left(s, \hat{a}^i\right)$. Then, conditioned on E_k and

$\pi_k \in \widetilde{\Pi}_{\text{eq}}$, we have that

$$
\begin{aligned}
&\max_{\hat{a}^i} Q^i_{\theta^i_{t_{k+1}}}\left(s, \hat{a}^i\right) - Q^i_{\theta^i_{t_{k+1}}}\left(s, \pi^i_k(s)\right) \\
&= Q^i_{\theta^i_{t_{k+1}}}\left(s, \hat{a}^{i*}\right) - Q^i_{\theta^i_{t_{k+1}}}\left(s, \pi^i_k(s)\right) \\
&= \left[Q^i_{\theta^i_{t_{k+1}}}\left(s, \hat{a}^{i*}\right) - Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \pi^i_k(s)\right)\right] + \left[Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \pi^i_k(s)\right) - Q^i_{\theta^i_{t_{k+1}}}\left(s, \pi^i_k(s)\right)\right] \\
&< Q^i_{\theta^i_{t_{k+1}}}\left(s, \hat{a}^{i*}\right) - Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \pi^i_k(s)\right) + \frac{1}{4}\min\left\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\right\} \\
&< \left[Q^i_{\theta^i_{t_{k+1}}}\left(s, \hat{a}^{i*}\right) - Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \hat{a}^{i*}\right)\right] + \left[Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \hat{a}^{i*}\right) - Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \pi^i_k(s)\right)\right] \\
&\quad + \frac{1}{4}\min\left\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\right\} \\
&< \frac{1}{4}\min\left\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\right\} + \frac{1}{4}\min\left\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\right\} \\
&\leq \frac{1}{2}\min\left\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\right\},
\end{aligned}
$$

where the second-to-last inequality follows since $Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \hat{a}^{i*}\right) - Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, \pi^i_k(s)\right) < 0$, which follows from $\pi_k \in \widetilde{\Pi}_{\text{eq}}$. It follows that $Q^i_{\theta^i_{t_{k+1}}}\left(s, \pi^i_k(s)\right) \geq \max_{\hat{a}^i} Q^i_{\theta^i_{t_{k+1}}}\left(s, \hat{a}^i\right) - \frac{1}{2}\zeta^i_\theta$ for all i. Then, by Algorithm 3.2 (lines 10-12), we have that $\pi_{k+1} = \pi_k$ with probability 1. □

Recall that $\widetilde{L}$ is the maximum length of the shortest strict best reply path from any policy to a linear approximated equilibrium policy. Our next lemma lower bounds the conditional probability of π_{k+L} being a linear approximated equilibrium policy, given that π_k is not a linear approximated equilibrium policy and given $\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1}$.

Lemma 3.14. *Let*

$$
\tilde{p} := \left(\min_{j \in \{1,\ldots,N\}} \left\{\frac{1-\lambda^j}{\left|\Pi^j\right|} \cdot \prod_{i \neq j} \lambda^i\right\}\right)^{\widetilde{L}} = \hat{p}^{\widetilde{L}/L}. \tag{3.77}
$$

We have that

$$P\left[\pi_{k+\widetilde{L}} \in \widetilde{\Pi}_{\text{eq}} \mid \widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1}, \pi_k \notin \widetilde{\Pi}_{\text{eq}}\right] \geq \tilde{p}. \tag{3.78}$$

Proof of Lemma 3.14. The proof is similar to that of Lemma 3.5. Consider some $\pi_k \notin \widetilde{\Pi}_{\text{eq}}$; there must exist at least one agent, say i, whose policy π_k^i is not the best reply to π_k^{-i}, i.e., $\pi_k^i \notin \widetilde{\Pi}^i_{\pi_k^{-i}}$. In this case, we claim that $\pi_k^i \notin \widetilde{\Pi}^i_{k+1}$, where $\widetilde{\Pi}^i_{k+1}$ is as defined in Algorithm 3.2 (line 10). In other words, the "else" statement in Algorithm 3.2 (line 14) will be executed. To see this, it suffices to show that $\phi^i_{t_{k+1}}(s, \pi_k^i(s))^\top \theta^i_{t_{k+1}} < \max_{a^i \in \mathcal{A}^i} \phi^i_{t_{k+1}}(s, a^i)^\top \theta^i_{t_{k+1}} - \frac{1}{2}\zeta^i_\theta$, for some $s \in \mathcal{S}$. Conditioned on $\widetilde{E}_k$, we have that

$$Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i) - \frac{1}{4}\min\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\} < Q^i_{\theta^i_{t_{k+1}}}(s, a^i) < Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i) + \frac{1}{4}\min\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\},$$

i.e., $Q^i_{\theta^i_{t_{k+1}}}(s, a^i)$ lies within a distance of $\frac{1}{4}\min\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\}$ to $Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i)$. Moreover, we note that $\frac{1}{4}\min\{\zeta^i_\theta, \bar{\zeta}_\theta - \zeta^i_\theta\} \leq \frac{1}{8}\bar{\zeta}_\theta$. Recall that $\{Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i) : a^i \in \mathcal{A}^i\}$ are dispersed with spacing being at least $\bar{\zeta}_\theta$, where $\bar{\zeta}_\theta$ is as defined in (3.64) as the minimum separation between the approximated Q-factors. Thus, it follows that the possible ranges of $Q^i_{\theta^i_{t_{k+1}}}(s, a^i)$ for all $a^i \in \mathcal{A}^i$ are mutually exclusive, which implies that the τ-th best action under $Q^i_{\theta^i_{\pi_k^{-i}}}$ is identical to that under $Q^i_{\theta^i_{t_{k+1}}}$, i.e.,

$$\arg\max_{a^i \in \mathcal{A}^i} \left(Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i)\right)_{(\tau)} = \arg\max_{a^i \in \mathcal{A}^i} \left(Q^i_{\theta^i_{t_{k+1}}}(s, a^i)\right)_{(\tau)},$$

where $(\cdot)_{(\tau)}$ represents the τ-th largest value. For instance, when $\tau = 1$, we have that $\arg\max_{a^i \in \mathcal{A}^i} Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i) = \arg\max_{a^i \in \mathcal{A}^i} Q^i_{\theta^i_{t_{k+1}}}(s, a^i)$, which are denoted by $a^{i*}_{\theta^i_{\pi_k^{-i}}}(s)$ and $a^{i*}_{\theta^i_{t_{k+1}}}(s)$, respectively.

Since $\pi_k^i \notin \widetilde{\Pi}^i_{\pi_k^{-i}}$, it follows that $\pi_k^i(s) \neq \arg\max_{a^i \in \mathcal{A}^i} Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i) =: a^{i*}_{\theta^i_{\pi_k^{-1}}}(s)$ for some $s \in \mathcal{S}$. Then, we have that

$$
\begin{aligned}
&\max_{a^i \in \mathcal{A}^i} Q^i_{\theta^i_{t_{k+1}}}(s, a^i) - Q^i_{\theta^i_{t_{k+1}}}(s, \pi_k^i(s)) \\
&> \left(\max_{a^i \in \mathcal{A}^i} Q^i_{\theta^i_{\pi_k^{-i}}}(s, a^i) - \frac{1}{8}\bar{\zeta}_\theta \right) - \left(Q^i_{\theta^i_{\pi_k^{-i}}}(s, \pi_k^i(s)) + \frac{1}{8}\bar{\zeta}_\theta \right) \\
&= \left(Q^i_{\theta^i_{\pi_k^{-i}}}\left(s, a^{i*}_{\theta^i_{\pi_k^{-i}}}(s)\right) - Q^i_{\theta^i_{\pi_k^{-i}}}(s, \pi_k^i(s)) \right) - \frac{1}{4}\bar{\zeta}_\theta \\
&\geq \bar{\zeta}_\theta - \frac{1}{4}\bar{\zeta}_\theta = \frac{3}{4}\bar{\zeta}_\theta \geq \frac{3}{4}\zeta^i_\theta > \frac{1}{2}\zeta^i_\theta
\end{aligned}
$$

as desired. Now, we are ready to prove the statement.

Let l be the length of the shortest strict best reply path from π_k to a linear approximated equilibrium policy. Then, $l \leq \widetilde{L}$. Let the sequence of policies along the path be $\pi_0, \pi_1, \ldots, \pi_l$, with $\pi_0 = \pi_k \notin \widetilde{\Pi}_{\text{eq}}$ and $\pi_l \in \widetilde{\Pi}_{\text{eq}}$. Further, let $i_1, \ldots, i_l$ be the agent that changes its policy at each update, i.e., π_{n-1} and π_n differ only at agent i_n, for all $n = 1, \ldots, l$. Then, we use the two probabilities in the policy update rule in Algorithm 3.2 (line 14) to yield

$$
\begin{aligned}
&P\left[\pi_{k+\widetilde{L}} \in \widetilde{\Pi}\text{eq} \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k \notin \widetilde{\Pi}_{\text{eq}}\right] \\
&\geq P\left[\pi_{k+\widetilde{L}} = \pi_l \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k \notin \widetilde{\Pi}_{\text{eq}}\right] \\
&\geq P\big[\pi_{k+1} = \pi_1, \pi_{k+2} = \pi_2, \ldots, \pi_{k+l} = \pi_l, \\
&\qquad \pi_{k+l+1} = \cdots = \pi_{k+\widetilde{L}} = \pi_l \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k \notin \widetilde{\Pi}_{\text{eq}}\big] \\
&\geq P\left[\pi_{k+1} = \pi_1 \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k = \pi_0\right] \\
&\quad \cdot P\left[\pi_{k+2} = \pi_2 \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1\right] \\
&\quad \cdot P\left[\pi_{k+3} = \pi_3 \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1, \pi_{k+2} = \pi_2\right] \cdot \cdots \\
&\quad \cdot P\left[\pi_{k+l} = \pi_l \mid E_k, \ldots, E_{k+\widetilde{L}-1}, \pi_k = \pi_0, \pi_{k+1} = \pi_1, \ldots, \pi_{k+l-1} = \pi_{l-1}\right]
\end{aligned}
$$

$$\cdot P\left[\pi_{k+l+1}=\pi_l \mid E_k,\dots,E_{k+\widetilde{L}-1},\pi_k=\pi_0,\pi_{k+1}=\pi_1,\dots,\pi_{k+l}=\pi_l\right]\cdot\cdots$$
$$\cdot P\left[\pi_{k+\widetilde{L}}=\pi_l \mid E_k,\dots,E_{k+\widetilde{L}-1},\pi_k=\pi_0,\pi_{k+1}=\pi_1,\dots,\pi_{k+l}=\pi_l,\dots,\pi_{k+\widetilde{L}-1}=\pi_l\right]$$
$$\geq \prod_{j\in\{i_1,\dots,i_l\}}\left(\frac{1-\lambda^j}{\left|\Pi^j\right|}\cdot\prod_{i\neq j}\lambda^i\right)\geq\left(\min_{j\in\{1,\dots,N\}}\left\{\frac{1-\lambda^j}{\left|\Pi^j\right|}\cdot\prod_{i\neq j}\lambda^i\right\}\right)^l$$
$$\geq\left(\min_{j\in\{1,\dots,N\}}\left\{\frac{1-\lambda^j}{\left|\Pi^j\right|}\cdot\prod_{i\neq j}\lambda^i\right\}\right)^{\widetilde{L}},$$

where we have used the fact from Lemma 3.13: given $\pi_l\in\widetilde{\Pi}_{\text{eq}}$ and the events $E_k,\dots,E_{k+\widetilde{L}-1}$, the conditional probability that $\pi_s\in\widetilde{\Pi}_{\text{eq}}$ is 1 for all $s\geq l$. □

We will then bound $P\left[\widetilde{E}_k,\dots,\widetilde{E}_{k+\widetilde{L}-1}\right]$. Before doing that, we first look at $P[\widetilde{E}_k]$. We would like $P[\widetilde{E}_k]$ to be as large as possible. Note that $\frac{1}{4}\min\{\zeta_\theta^i,\bar{\zeta}_\theta-\zeta_\theta^i\}\leq\frac{1}{8}\bar{\zeta}_\theta$, with equality holding when $\zeta_\theta^i=\frac{1}{2}\bar{\zeta}_\theta$. We next have the following lemma.

Lemma 3.15. *Let $\zeta_\theta^i=\frac{\bar{\zeta}_\theta}{2}$ for all $i\in[N]$. Fix an arbitrary $\pi_k\in\Pi$. For any $0<\hat{\delta}<1$, we have that*

$$P\left[\widetilde{E}_k\right]\geq 1-\hat{\delta},$$

provided that $\rho^i\leq 1-\left(1-\frac{(\bar{\zeta}_\theta/8-\epsilon)(1-\bar{\gamma})-2b}{\widetilde{\Gamma}}\right)^{\frac{1}{N-1}}$, and T_k and η_t^i satisfy (3.68), where ϵ can take any value in $0<\epsilon<\min\left\{\frac{\bar{\zeta}_\theta}{8}-\frac{2b}{1-\bar{\gamma}},\frac{1}{1-\gamma_{\min}}\right\}$.

Proof of Lemma 3.15. A direct implication of Lemma 3.11 with (3.70) and Lemma 3.12 is that when T_k and η_t^i satisfy (3.68), and ρ^i satisfies (3.73), then, by triangle inequality, we have that

$$P\left[\left|Q^i_{\theta^i_{t_{k+1}}}-Q^i_{\theta^i_{\pi_k^{-i}}}\right|_\infty\leq\epsilon+\tilde{\epsilon}+\frac{2b}{1-\bar{\gamma}},\quad\forall i\in[N]\right]\geq 1-\hat{\delta}. \tag{3.79}$$

The lemma then follows by taking $\tilde{\epsilon}=\frac{1}{8}\bar{\zeta}_\theta-\epsilon-\frac{2b}{1-\bar{\gamma}}$. □

We then have the following lemma which bounds $P\left[\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1}\right]$.

Lemma 3.16. *For any arbitrary sequence of joint policies $\pi_k, \ldots, \pi_{k+\widetilde{L}-1} \in \Pi$, and for any $0 < \tilde{\delta} < 1$, we have that*

$$P\left[\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1}\right] \geq 1 - \tilde{\delta},$$

provided that for all $i \in [N]$ and for all $\hat{k} \in \{k, \ldots, k + \widetilde{L} - 1\}$,

$$\eta_t^i = \eta^i \leq \frac{\epsilon^2 \tilde{\delta} \xi^i}{456 N \widetilde{L} \left(1 + \gamma^i + r_{\max}^i\right)^2 (D^i + 1)^2 t_{\text{mix},k}^i(\eta^i)}, \quad \forall t = t_{\hat{k}}, \ldots, t_{\hat{k}+1} - 1, \forall i \in [N], \tag{3.80a}$$

$$T_{\hat{k}} \geq t_{\text{mix},k}(\eta_{\min}) + \frac{\log \frac{\epsilon^2 \tilde{\delta}}{2N\widetilde{L}(2D+1)^2}}{\log\left(1 - \xi_{\min}\eta_{\min}/2\right)}, \tag{3.80b}$$

$$\rho^i \leq 1 - \left(1 - \frac{\left(\bar{\zeta}_\theta/8 - \epsilon\right)(1 - \bar{\gamma}) - 2b}{\widetilde{\Gamma}}\right)^{\frac{1}{N-1}}, \tag{3.80c}$$

$$\zeta^i = \frac{\bar{\zeta}}{2}, \tag{3.80d}$$

where ϵ can take any value in $0 < \epsilon < \min\left\{\frac{\bar{\zeta}_\theta}{8} - \frac{2b}{1-\bar{\gamma}}, \frac{1}{1-\gamma_{\min}}\right\}$.

Proof of Lemma 3.16. When the conditions of Lemma 3.15 are satisfied, we have that $P\left[\widetilde{E}_k^c\right] < \hat{\delta}$, where $\widetilde{E}_k^c$ is the complement of $\widetilde{E}_k$. Then,

$$\begin{aligned} P\left[\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1}\right] &= 1 - P\left[\left(\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1}\right)^c\right] = 1 - P\left[\widetilde{E}_k^c \cup \cdots \cup \widetilde{E}_{k+\widetilde{L}-1}^c\right] \\ &\geq 1 - \left(P\left[\widetilde{E}_k^c\right] + P\left[\widetilde{E}_{k+1}^c\right] + \cdots + P\left[\widetilde{E}_{k+\widetilde{L}-1}^c\right]\right) = 1 - \widetilde{L}\hat{\delta}. \end{aligned}$$

By taking $\tilde{\delta} = \widetilde{L}\hat{\delta}$, it follows that the conditions (3.68) now become (3.80), and the lemma is proved. □

For simplicity, we choose $\epsilon = \frac{1}{2}\min\left\{\frac{\bar{\zeta}_\theta}{8} - \frac{2b}{1-\bar{\gamma}}, \frac{1}{1-\gamma_{\min}}\right\}$ in Theorem 3.2. Note that

the result of Lemma 3.16 holds for any realization of $\pi_k \in \Pi$. Therefore, under the same conditions, we in fact have that

$$P\left[\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1} \mid \pi_k \in \widetilde{\Pi}_{\text{eq}}\right] \geq 1 - \tilde{\delta}, \tag{3.81a}$$

$$P\left[\widetilde{E}_k, \ldots, \widetilde{E}_{k+\widetilde{L}-1} \mid \pi_k \notin \widetilde{\Pi}_{\text{eq}}\right] \geq 1 - \tilde{\delta}. \tag{3.81b}$$

By Lemma 3.13 and (3.81a), under conditions (3.80), we have that for all k,

$$P\left[\pi_k = \pi_{k+1} = \cdots = \pi_{k+\widetilde{L}} \mid \pi_k \in \widetilde{\Pi}_{\text{eq}}\right] \geq 1 - \tilde{\delta}. \tag{3.82}$$

By Lemma 3.14 and (3.81b), under conditions (3.80), we have that for all k,

$$P\left[\pi_{k+\widetilde{L}} \in \widetilde{\Pi}_{\text{eq}} \mid \pi_k \notin \widetilde{\Pi}_{\text{eq}}\right] \geq \tilde{p}\left(1 - \tilde{\delta}\right). \tag{3.83}$$

As a notation, let $p_k := P\left[\pi_k \in \widetilde{\Pi}_{\text{eq}}\right]$. Then, (3.82) and (3.83) together imply that

$$p_{(n+1)L} \geq p_{nL}\left(1 - \tilde{\delta}\right) + (1 - p_{nL})\tilde{p}\left(1 - \tilde{\delta}\right). \tag{3.84}$$

Rearranging the above, we obtain that

$$\begin{aligned} p_{(n+1)L} - p_{nL} &\geq \left(1 - \tilde{\delta}\right)\tilde{p} - \tilde{\delta}p_{nL} - \left(1 - \tilde{\delta}\right)\tilde{p}p_{nL} \\ &= \left[\tilde{\delta} + \left(1 - \tilde{\delta}\right)\tilde{p}\right]\left[\frac{\left(1 - \tilde{\delta}\right)\tilde{p}}{\tilde{\delta} + \left(1 - \tilde{\delta}\right)\tilde{p}} - p_{nL}\right] \qquad (3.85) \\ &\geq -\tilde{\delta} \qquad (3.86) \end{aligned}$$

Note that $p_{(n+1)L} - p_{nL} \geq 0$ as long as $p_{nL} \leq \frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}}$. Further, if $p_{nL} \leq \frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}} - \tilde{\delta}$,

then from (3.85), we have that $p_{(n+1)L} - p_{nL} \geq \left[\tilde{\delta} + \left(1-\tilde{\delta}\right)\tilde{p}\right]\tilde{\delta}$; if $p_{nL} > \frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}}$, then $p_{(n+1)L} - p_{nL} \geq -\tilde{\delta}$ from (3.86). Therefore, we have that

$$p_{nL} \geq \frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}} - \tilde{\delta}, \quad \forall n \geq \tilde{n}, \tag{3.87}$$

where

$$\tilde{n} := \frac{\frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}} - \tilde{\delta}}{\left[\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}\right]\tilde{\delta}} = \frac{\left(1-\tilde{\delta}\right)^2\tilde{p} - \tilde{\delta}^2}{\left[\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}\right]^2\tilde{\delta}}. \tag{3.88}$$

This, together with (3.82), implies that for all $n \geq \tilde{n}$,

$$P\left[\pi_{nL} = \pi_{nL+1} = \cdots = \pi_{nL+L} \in \widetilde{\Pi}_{\text{eq}}\right] \geq \left(\frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}} - \tilde{\delta}\right)\left(1-\tilde{\delta}\right) := f(\tilde{\delta}). \tag{3.89}$$

Therefore, if the number of exploration phases $k \geq K := \tilde{n}\widetilde{L}$, then $P\left[\pi_k \in \widetilde{\Pi}_{\text{eq}}\right] \geq f(\tilde{\delta})$. Note that $f(\tilde{\delta})$ is continuous, decreasing in $\tilde{\delta}$, and $f(0) = 1$, $f(\delta) < 1 - \delta$ for any $0 < \delta < 1$. Thus, we can take $\tilde{\delta} \in (0, \delta)$ such that

$$\left(\frac{\left(1-\tilde{\delta}\right)\tilde{p}}{\tilde{\delta}+\left(1-\tilde{\delta}\right)\tilde{p}} - \tilde{\delta}\right)\left(1-\tilde{\delta}\right) = 1 - \delta, \tag{3.90}$$

which leads to $P\left[\pi_k \in \widetilde{\Pi}_{\text{eq}}\right] \geq 1 - \delta$ for all $k \geq K$, and this completes the proof of Theorem 3.2.

3.6.2 Proof of Theorem 3.3

Following Lemma 3.11 and by noting that $|\phi^i|_\infty \leq \|\phi^i\|_2 \leq 1$, Lemma 3.11 implies that for an arbitrary $\bar{\pi}_k$ as in (3.10) and for any $\epsilon > 0$ and $0 < \hat{\delta} < 1$,

$$P\left[\left|Q^i_{\theta^i_{t_{k+1}}} - Q^i_{\theta^i_{\bar{\pi}_k^{-i}}}\right|_\infty \leq \epsilon,\ \forall i \in [N]\right] = P\left[\left|{\phi^i}^\top \theta^i_{t_{k+1}} - {\phi^i}^\top \theta^i_{\bar{\pi}_k^{-i}}\right|_\infty \leq \epsilon,\ \forall i \in [N]\right]$$
$$\geq 1 - \hat{\delta}$$

when the conditions (3.68) are satisfied. Under Assumption 3.7, we have that $Q^i_{\theta^i_{\bar{\pi}_k^{-i}}} = Q^i_{\bar{\pi}_k^{-i}}$, which leads to

$$P\left[\left|Q^i_{\theta^i_{t_{k+1}}} - Q^i_{\bar{\pi}_k^{-i}}\right|_\infty \leq \epsilon,\ \forall i \in [N]\right] \geq 1 - \hat{\delta}.$$

The above is exactly the same as the result in Lemma 3.2, and the rest of the proof follows the proof of Theorem 3.1.

CHAPTER 4

DESIGN OF POWER PURCHASE AGREEMENTS WITH RENEWABLE ENERGY GENERATORS

4.1 Introduction

It is with little doubt that addressing the energy and environmental challenges, including climate change, is one of the top priorities of our time (United Nations [215]). Among all economic sectors, electricity production is the second largest contributor to U.S. greenhouse gas (GHG) emissions, and the largest emitting sector in the world (Environmental Protection Agency [69], International Energy Agency [115]). Renewable energy sources, such as solar and wind, emit little to no greenhouse gases, reducing the negative impact of energy production on the planet. Thus, they are not only crucial in meeting energy needs but also in reducing greenhouse gas emissions and mitigating climate change. Renewable energy investments have recently soared across the globe. According to International Energy Agency [116], global energy investment is set to increase by 8% in 2022 to reach USD 2.4 trillion, with the anticipated rise coming mainly in clean energy. Moreover, the pace of growth of clean energy has accelerated significantly to 12% since 2020, comparing to only 2% a year in the five years after the Paris Agreement was signed in 2015. These investments in renewable energy provide significant environmental, economic, and societal benefits. From an environmental perspective, renewable energy sources such as solar and wind emit little to no greenhouse gases or air pollutants, reducing the negative impact of energy production on the planet. By transitioning to renewable energy, we can reduce our reliance on fossil fuels, which are finite resources and contribute to climate change. Economically, renewable energy has the potential to create new jobs and boost economic growth. According to International Renewable Energy Agency [117], the renewable energy sector employed over 12.7 million people globally in 2021, and this number is expected to continue to grow. Additionally,

investing in renewable energy can lead to cost savings over time. On societal impacts, access to renewable energy can provide energy security, particularly in remote or underserved areas, and improve the health and well-being of communities by reducing air pollution. Additionally, renewable energy projects often involve collaboration with local communities and can provide opportunities for community ownership and participation.

Despite the numerous benefits of renewable energy, there are also challenges that need to be addressed. One of the main challenges is the intermittency of renewable energy sources such as solar and wind power. Unlike traditional energy sources, renewable energy sources are dependent on weather conditions and may not always be available when needed. With the increasing integration of renewable energy, the energy supply becomes more volatile, which in turn also leads to financial instability to both the generators and the consumers of renewable energy.

One solution to address these uncertainties brought by the renewable energy is the power purchase agreements. A power purchase agreement (PPA) is a contractual agreement between a firm (buyer) and a renewable energy generator (seller) (Wu and Babich [222]). These agreements are also called *renewable PPAs*. PPAs provide more financial certainty to both the buyer and the seller, thus removing a significant roadblock to building new renewable facilities (Bruck et al. [44]). Many Fortune 500 companies, such as Amazon and Meta, have made significant investments in renewable energy (Holter [106]). In 2021 alone, Amazon signed 6.2GW of PPAs, accounting for over 20% of a record 31.1GW of clean-power purchase deals inked by private companies around the world (Chediak [47]). In 2022, global renewable PPA volume was 36.7GW, which is 18% higher than the 2021 figure, and the volume of total renewable PPAs signed by corporations between 2008 and 2022 exceeded the entire energy generation capacity of France (PV Tech [175]). The current designs of Power Purchase Agreements (PPAs) can differ in their structures, with varying specifications for the quantity of electricity to be delivered, which can be expressed in kWh or as a total gener-

ation from specific facilities, such as solar farms. The pricing of the electricity can be either fixed or indexed to market rates, and the delivery point can either be on-site or off-site. Furthermore, payment schedules may vary depending on the specific terms agreed upon by the buyer and seller.

As the number of new renewable energy deals continue to grow, it is becoming more important than ever to have a better design of PPAs that, while financially benefiting the firm, provides incentives for more investment on renewable energy facilities, and makes the firm's electricity consumption more eco-friendly. Toward this goal, in this chapter, we design a PPA where the firm agrees to make a certain transfer payment to the renewable generator, and the generator invests that payment to build new renewable energy facilities, such as solar photovoltaics (PVs) and/or wind turbines. The firm will then have access to all electricity generation from the new facilities for a long-term period, e.g., 20 years (Christophers [54]). The firm may dynamically decide when to start the PPA on an ongoing basis, based on the evolving market conditions, and the transfer payment (amount of investment) is also specified by the firm. The firm's objective is to maximize its long-term discounted benefit (total savings) from signing the PPA. For ease of presentation, we consider solar PPAs in this chapter, while our model can be easily modified to suit PPAs with other types of renewable generations.

4.1.1 Contributions

The contribution of this chapter is three-fold. First, we mathematically formulate the firm's decision problem as an optimal stopping problem and provide analytical solutions. Specifically, we provide the optimal policy for the firm, where the firm signs the PPA once the total electricity demand in the market reaches some constant threshold, and the amount of transfer payment is set such that the capacity of newly added renewable facilities optimizes the firm's expected long-term total discounted savings. In this work, we also characterize

the effect of the renewable energy facilities' production level (for solar, this is determined by the weather and the conversion efficiency of PVs), the length of the PPA, and the investment cost of renewable facilities. We conclude that with an increased PV efficiency or with an increased length of the PPA, the firm will optimally sign a PPA earlier, with a smaller capacity of new renewable facilities, and the firm attains higher expected value. The expected total new generation from the PPA may increase or decrease with the PV efficiency, depending on the variation and efficiency of PV generations. Furthermore, under certain conditions, a higher investment cost will make the firm wait longer to sign the PPA, but the firm would also invest in a higher capacity of renewable facilities, which leads to a higher expected total new generation from the PPA.

Second, we consider the same PPA model with technology price discount, i.e., the investment cost exponentially decreases in time. This additional discount on investment cost leads to a more complicated optimal policy for the firm: the firm should start the PPA as soon as the total demand in the market hits some time-dependent threshold. We obtain the optimal threshold as a function of time, and provide explicit expressions for the firm's optimal investment capacity. We also derive the distribution of the firm's waiting time (before signing the PPA). The characterization of the effect of the renewable energy facilities' production level, the length of the PPA, and the investment cost are similar to those in the original PPA model, with slight differences. The optimal capacity is no longer monotonically decreasing with the PV efficiency and the length of the PPA. Instead, the change of optimal capacity might increase or decrease with respect to the PV efficiency, depending on the ratio of the variation and efficiency of PV generations. The optimal capacity first decreases and then slightly increases with the length of the PPA, where the increase reflects the additional benefit of discount on investment costs at a later time.

Third, we consider an extension model where the firm no longer dynamically decides when to start the PPA, but needs to commit at the very beginning if it would sign a PPA,

and if yes, when the PPA would start and how much to invest. We again characterize the effect of the production level, the length of the PPA, and the investment cost on the firm's optimal capacity, savings, and the total new generation due to the PPA.

4.1.2 Literature Review

Integration and operation of renewable energy has been studied extensively in literature from different aspects. We list below a few of them.

Competition and equilibrium: Al-Gwaiz et al. [7] studied the competition of conventional power generators with different levels of flexibility and the impact of intermittent renewable generators on the competition. Sunar and Birge [204] analyzed equilibrium in the day-ahead electricity market with both renewable generators and conventional generators, and characterized the equilibrium bidding strategy of renewable generators in the day-ahead market.

Pricing policies: Alizamir et al. [9] studied the dynamic control of prices of feed-in tariff policies for promoting renewable energy. Kök et al. [131] considered the impact of pricing policies, i.e., flat pricing versus peak pricing, on the investment levels of a utility firm in both renewable and conventional energy sources. Mamaghani and Çakanyıldırım [152] studied the interplay between a higher solar adoption level and a higher electricity price. Singh and Scheller-Wolf [196] analyzed the effect of tariff structures in the regulator's social welfare maximization problem in a market with a regulated utility; an unregulated, price-setting, profit-maximizing solar system installer; and customers who endogenously determine whether to adopt solar or not.

Investment problem: Kök et al. [132] considered the capacity investment problem for a utility firm that invests in both renewable and conventional energy. Angelus [14] studied the

investment decisions on distributed energy resources by the end consumers. Hu et al. [110] focused on the organization's one-time capacity investment in a renewable energy-producing technology with supply intermittency and net metering compensation and emphasized the importance of data granularity. Kaps et al. [124] considered the joint optimization of investment in renewable generation capacity and storage. An empirical study by Huang et al. [111] considered the effect of noisy customer reviews on solar marketplace.

Distributed energy resources (DERs): Sunar and Swaminathan [205] analyzed how net-metered distributed renewable energy technologies, such as rooftop solar panels adopted by end-users, impact utility profits. Chen et al. [49] studied the competitive aggregation of DERs. Chen et al. [50] designed a coordination mechanism between the distribution system operator and the DER aggregators to ensure system reliability while providing open access to the aggregators. Gao et al. [88, 89, 92] proposed efficient aggregations of DERs through a two-part pricing policy, and analyzed the effect of DERs on the market power of conventional generators. Mamaghani and Çakanyıldırım [152] studied the interplay between a higher solar adoption level and a higher electricity price. Singh and Scheller-Wolf [196] analyzed the effect of tariff structures in the regulator's social welfare maximization problem in a market with a regulated utility; an unregulated, price-setting, profit-maximizing solar system installer; and customers who endogenously determine whether to adopt solar or not.

Power purchase agreements (PPAs): Wu and Babich [222] analyzed different types of power plants' misreporting in a unit-contingent power purchase agreement with an electricity distributor. Trivella et al. [213] employed approximate dynamic programming to study power purchase agreements. However, there is no prior study that analyzes the truly optimal design of PPAs with renewable energy generators, which is the topic of our study.

Apart from these, Peng et al. [173] studied the optimization on the joint operations of different types of energy resources, i.e., renewable (wind and solar), flexible (natural gas), and

storage capacities. Birge [39] proposed methods to uncover the network structure (including electric power networks) across commodity markets and provided implications that these hidden connections may have for predicting risk propagation and cascading failures. For more papers on sustainable operations and renewable energy, see Sunar and Swaminathan [206] for a comprehensive review.

4.1.3 Organization

The rest of this chapter is organized as follows. In Section 4.2, we introduce our first PPA model by describing the renewable generator and the firm's problems. We also characterize the electricity price dynamics. In Section 4.3, we formulate the firm's dynamic decisions as an optimal stopping problem, and derive the optimal solution to the PPA model. We also provide a complete analysis on the properties of the optimal investment capacity, the firm's expected savings, and the total new generation from the PPA. Then, a more complicated model, with the additional discount on investment cost, is proposed and solved in Section 4.4. The effect of discount on investment cost is also discussed. In Section 4.5, we study another extension of the PPA model where the firm needs to make all the investment decisions upfront. The chapter concludes with Section 4.6. For the ease of presentation, all proofs are left to Appendix 4.7.

4.2 Power Purchase Agreement Model

In this section, we introduce our first power purchase agreement (PPA) model, which includes the problems faced by the firm (or utility) and the renewable generator, as well as the electricity price dynamics in the spot market. In a nutshell, at time $t = 0$, the renewable generator first announces $\{Q_t, t \geq 0\}$, the production amount of one unit of the renewable facility, which is a general stochastic process, as well as the cost function for the investment of renewable facilities. As time evolves, the firm observes its own demand as well as the

total demand for electricity in the spot market, and dynamically (at each time instant t) decides if it would sign a PPA of length T (that starts immediately) with the renewable generator to minimize its expected total cost. A transfer amount C is also specified by the firm, to be paid at the start of the PPA. If a PPA is signed at time τ, then the renewable generator invests all C to build new renewable energy facilities, and the firm has access to all electricity production from these facilities from τ to $\tau + T$. However, the newly built renewable facilities have a lifespan $\hat{T}$ (which is greater than T) and can last till time $\tau + \hat{T}$.

4.2.1 Renewable Generator's Problem

The renewable generator is assumed to be a passive decision maker who always invests all C to maximize the size (capacity) of the new renewable energy facilities. Let $I(k)$ be the cost function of investing k units of renewable facilities. Assume that $I(k)$ is given by the following linear function

$$I(k) = bk, \quad k \geq 0, \tag{4.1}$$

where b is a positive constant coefficient. If a PPA is signed, the renewable generator earns the transfer C. The renewable generator builds renewable facilities of capacity K such that

$$I(K) = C \Longrightarrow K = \frac{C}{b}. \tag{4.2}$$

4.2.2 Electricity Spot Price Dynamics

Let the total demand for electricity in the spot market be $\{D_t, t \geq 0\}$ that follows a geometric Brownian motion (GBM), i.e., $dD_t = \mu_D D_t dt + \sigma_D D_t dW_t$, with the assumption that $\mu_D > \sigma_D^2/2$. The GBM process of energy demand has been commonly assumed in literature, e.g. Djauhari et al. [66], Marathe and Ryan [155].

The production from one unit of the renewable facility is a general stationary stochastic process $\{Q_t, t \geq 0\}$, with mean μ_Q and standard deviation σ_Q, which is independent of $\{D_t\}$. Let $\hat{Q}_t$ be the total generation from the new renewable facility from the PPA. With K capacity of facility, the total electricity production from the renewable facility is $\hat{Q}_t := KQ_t$. Then, the net electricity demand is given by the stochastic process

$$N_t = D_t - KQ_t, \quad t \geq 0. \tag{4.3}$$

Assume that the electricity price in spot market has the form

$$p_t = \theta N_t, \tag{4.4}$$

where θ is some positive constant. Let p_t^N denote the spot market price at time t, assuming no PPA is signed, and let p_t^Y denote the spot market price at time t if a PPA is signed. Then, from (4.4) we have that $p_t^Y = \theta(D_t - KQ_t)$ and $p_t^N = \theta D_t$.

4.2.3 Firm's Problem

Consider a firm that needs to satisfy an uncertain residual electricity demand, i.e., the excess demand that is not met by existing generation sources or power contracts. The firm's residual demand at each time instant t is given by

$$\{U_t, t \geq 0\}. \tag{4.5}$$

We assume that $U_t = \alpha D_t$ (α being a constant in $[0, 1]$), i.e., the firm's residual demand is always a constant fraction of the total demand. The residual demand can be satisfied by procuring electricity from the spot market at the spot market price. The firm also has the option to make a long term capacity contract with a new renewable generator, which gives

the firm access to all electricity produced by the renewable generator for T time length. Such a contract may fulfill all or part of the residual demand, while the remaining residual demand may still be satisfied from the spot market.

Let λ_d be the discount rate of uninvested cash (real interest rate). Throughout the rest of this chapter, we make the following assumption: $\lambda_d > 2\mu_D + \sigma_D^2$.

Without a PPA, the firm's total cost, starting from τ and discounted to time τ, for procuring electricity from the spot market is given by

$$\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^N U_s ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^N U_s ds + \int_{\tau+\hat{T}}^{\infty} e^{-\lambda_d(s-\tau)} p_s^N U_s ds. \tag{4.6}$$

With a PPA starting from τ and lasts till $\tau+T$, the firm's total cost for procuring electricity from the spot market becomes

$$\int_{\tau}^{\tau+T} e^{-\lambda_d(s-t)} p_s^Y \left[U_s - \hat{Q}_s\right]^+ ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^Y U_s ds + \int_{\tau+\hat{T}}^{\infty} e^{-\lambda_d(s-\tau)} p_s^Y U_s ds. \tag{4.7}$$

We assume that, with the PPA, the firm always own all electricity produced by the renewable generator during $(\tau, \tau+T)$, and when $\hat{Q}_t > U_t$, the firm can sell $\hat{Q}_t - U_t$ back, again at the spot market price. Therefore, when the PPA is signed at time τ, the firm will earn the following revenue from selling:

$$\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^Y \left[\hat{Q}_s - U_s\right]^+ ds. \tag{4.8}$$

Thus, the firm's overall cost, with a PPA signed at τ, is given by

$$\begin{aligned}(4.7) - (4.8) + C = &\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^Y \left[U_s - \hat{Q}_s\right] ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^Y U_s ds \\ &+ \int_{\tau+\hat{T}}^{\infty} e^{-\lambda_d(s-\tau)} p_s^Y U_s ds + C.\end{aligned} \tag{4.9}$$

Note that $\hat{T}$ is the lifespan of the renewable facilities, and thus $N_t = D_t$ for $t > \tau + \hat{T}$ with and without the PPA, which results in $p_t^Y = p_t^N$ for $t \in (\tau + \hat{T}, \infty)$. The firm's savings (discounted to time τ) from signing a PPA with transfer amount C at time τ is then

$$\begin{aligned}(4.6) - (4.9) = &\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^N U_s ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^N U_s ds \\ &- \int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^Y \left[U_s - \hat{Q}_s\right] ds - \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^Y U_s ds - C.\end{aligned} \tag{4.10}$$

Note that p_s^Y can be viewed as a function of K, and thus (4.10) is a function of C and K. The firm's decision include dynamically choosing a time τ to start the PPA, as well as choosing a transfer C, or equivalently, choosing an investment capacity K (since $C = bK$ from (4.2)). The firm's objective is to maximize its discounted expected saving from a PPA:

$$\begin{aligned}&\max_{\tau, C} \mathbb{E}\left[e^{-\lambda_d \tau} \cdot (4.10)\right] \\ &\text{s.t. } I(K) = C\end{aligned} \tag{4.11}$$

Note that from (4.11), the firm has to decide both C, the investment amount, and τ, the starting time of the PPA. On one hand, the firm's optimal decision on C, given any starting time τ, is such that (4.10) is maximized. On the other hand, the firm dynamically decides on τ, based on the evolving process of D_t, to maximize its discounted expected saving, assuming that C is chosen optimally.

4.3 Analysis and Solution to the Power Purchase Agreement Model

In this section, we derive the optimal solution to the PPA model and provide a complete analysis on the properties of the optimal investment capacity, the firm's expected savings,

and the total new generation from the PPA.

4.3.1 Signing PPA at Time τ

To solve the PPA model, we need to find the firm's optimal decision on whether/when to sign a PPA and on the transfer amount C. We will formulate the firm's dynamic decision on whether to sign a PPA (at each time instance) as an optimal stopping problem. Before that, however, we first study in this subsection how much the firm could save if it signs the PPA at an arbitrary given time t, which builds the foundation for the optimal stopping problem that we present in the next subsection.

When the firm signs a PPA at time τ, its saving (4.10) is optimized by choosing an optimal transfer C. Correspondingly, according to (4.2), C/b units of renewable facilities will be built. In the following lemma, we derive the firm's expected saving if a PPA is signed at time τ.

Lemma 4.1. *If a PPA is signed at time τ, and the firm sets the transfer $C = bK$ where K is the amount of new renewable facilities built, then, the firm's expected saving from this PPA, discounted to time τ, is given by*

$$K\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)T}\right]+K\frac{\alpha\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right] - K^2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]-bK. \tag{4.12}$$

Lemma 4.1 provides the expected saving of the firm in terms of D_τ, the total demand of the spot market at time τ, and K, the capacity of new renewable energy facilities. Since the firm chooses $C = bK$, this saving can be further optimized over K. In the following proposition, we present the optimal capacity and saving of the firm.

Proposition 4.1. *Suppose that the firm signs the PPA at time $\tau > 0$. Then, the optimal*

renewable energy capacity added because of the PPA is

$$K(D_\tau) := \max\left\{ \frac{\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}, 0 \right\}, \tag{4.13}$$

and the firm's optimal expected saving from the PPA is

$$S(D_\tau) := \begin{cases} \dfrac{\left[\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}, & \\ & \text{if } D_\tau \geq \dfrac{b(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}, \\ 0, & \text{if } D_\tau < \dfrac{b(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}. \end{cases} \tag{4.14}$$

From Proposition 4.1, we see that the firm would choose a capacity $K > 0$, resulting in a positive expected saving S, if and only if $D_\tau \geq \frac{b(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}$. In other words, when the investment cost b is high enough: $b > \frac{D_\tau\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{\lambda_d-\mu_D}$, then the firm's optimal investment capacity would be zero, meaning that the cost is too high for the firm to make any investment.

Next, we are interested in how the optimal capacity and the saving change with respect to the production process (described by μ_Q and σ_Q) and T, the length of the PPA. The results are summarized as Proposition 4.2.

Proposition 4.2. *The optimal renewable capacity and the firm's optimal expected saving change with respect to different problem parameters as follows.*

1. *There exists a unique threshold $\bar{\mu}$ such that $\frac{\partial K(D_\tau)}{\partial \mu_Q} < 0$ if and only if $\mu_Q > \bar{\mu}$.*
2. *There exists a unique threshold $\hat{\mu}$ such that $\frac{\partial S(D_\tau)}{\partial \mu_Q} > 0$ if and only if $\mu_Q > \hat{\mu}$.*

3. *If* $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] > b$*, then, there exists a threshold* T_1 *such that when* $T < T_1$*, we have that* $\frac{\partial K(D_\tau)}{\partial T} < 0$*. If* $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] \leq b$*, then, there exists a threshold* T_2 *such that when* $T < T_2$*, we have that* $\frac{\partial K(D_\tau)}{\partial T} > 0$*.*

4. *If* $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] \neq b$*, then, there exists a threshold* T_3 *such that when* $T < T_3$*, we have that* $\frac{\partial S(D_\tau)}{\partial T} < 0$*. If* $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] \neq b$*, then, there exists a threshold* T_4 *such that when* $T < T_4$*, we have that* $\frac{\partial S(D_\tau)}{\partial T} > 0$*.*

5. $\frac{\partial K(D_\tau)}{\partial \sigma_Q} < 0$, $\frac{\partial S(D_\tau)}{\partial \sigma_Q} < 0$.

From the first item of Proposition 4.2, we see that $K(D_\tau)$ first increases with μ_Q, then after $\mu_Q > \bar{\mu}$, the investment capacity starts to decrease. On the other hand, $S(D_\tau)$ first decreases with μ_Q, then after $\mu_Q > \hat{\mu}$, the firm's saving starts to increase. The change of $K(D_\tau)$ and $S(D_\tau)$ with respect to T are less tractable, but we obtain their limiting behaviors as $T \to 0$. Finally, as the variance of production σ_Q increases, both $K(D_\tau)$ and $S(D_\tau)$ decrease.

4.3.2 Optimal Time to Sign PPA - Dynamic Decision

We now come back to the original model where the firm needs to dynamically decide $\tau \geq 0$, the time to sign a PPA, and the transfer amount C to the renewable energy generator. We will use x as a generic notation to represent the realization of the initial total demand for electricity in the spot market. Let $V(x)$ be the firm's expected saving if it optimally chooses the time to sign the PPA and the transfer payment C given the initial demand realization is x. Recall from (4.2) that optimizing with respect to C is equivalent to optimizing with respect to K, and the optimal saving at the stopping time is already given by (4.14). Therefore, the firm's value function can be written as the following optimal stopping problem:

$$V(x) = \max_{\tau \geq 0} \mathbb{E}\left[e^{-\lambda_d \tau} S(D_\tau) \mid D_0 = x\right], \tag{4.15}$$

where we recall that τ is the time to sign a PPA (or the starting time of the PPA, or the time the firm stops waiting).

We assume that $V(x)$ is twice continuously differentiable and nonnegative. From (4.14), we know that $S(x)$ is continuously differentiable, nonnegative, and monotone. The decision to start or not to start a PPA at any time t when the realization $D_t = x$ depends on the comparison of $V(x)$ and $S(x)$. If $V(x) > S(x)$, the optimal τ^* that solves (4.15) is strictly positive, and it is more beneficial for the firm to wait, since the expected value of waiting is higher than the value of starting a PPA immediately. The set $\{x \mid V(x) > S(x)\}$ is called the *continuation region*. If $V(x) = S(x)$, then one optimal $\tau^* = 0$, and it is optimal for the firm to start the PPA immediately with the expected saving $S(x)$. The set $\{x \mid V(x) \leq S(x)\}$ is called the *stopping region*. We next have the following lemma on the characterization of $V(x)$.

Lemma 4.2. *The value function satisfies the following Hamilton–Jacobi–Bellman (HJB) equation:*

$$V(x) = \max\left\{S(x),\ \frac{1}{\lambda_d}\mu_D x V'(x) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V''(x)\right\}. \tag{4.16}$$

Lemma 4.2 implies that in the stopping region, $V(x) = S(x)$, and in the continuation region, we have

$$V(x) = \frac{1}{\lambda_d}\mu_D x V'(x) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V''(x). \tag{4.17}$$

The firm's decision to continue waiting or to stop waiting (start a PPA) only depends on the realized total market demand. As the firm waits, we have that $S(x) < \frac{1}{\lambda_d}\mu_D x V'(x) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V''(x)$. The market demand continues to evolve while the firm is waiting, until x reaches some x_* such that $S(x_*) = \frac{1}{\lambda_d}\mu_D x_* V'(x_*) + \frac{1}{2\lambda_d}\sigma_D^2 x_*^2 V''(x_*)$, at which point the firm would stop waiting. The set $\left\{x \mid V(x) = S(x) = \frac{1}{\lambda_d}\mu_D x V'(x) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V''(x)\right\}$ is called

the *optimal stopping boundary*. Since both $V(x)$ and $S(x)$ are continuously differentiable, at x_* in the optimal stopping boundary, we have that

$$V(x_*) = S(x_*), \tag{4.18a}$$

$$V'(x_*) = S'(x_*), \tag{4.18b}$$

where (4.18a) is the *value matching condition* and (4.18b) is the *smooth pasting condition*.

It remains to find the x_* in the optimal stopping boundary, such that the firm would optimally start a PPA when x first reaches x_*. In the following proposition, we formalize the optimal policy and specify x_* by solving the differential equation (4.17) subject to the boundary conditions (4.18).

Proposition 4.3. *Suppose that the firm dynamically decides when to sign a PPA with a renewable energy generator. For the firm, it is optimal to sign the PPA at*

$$\tau^* = \inf\{t \geq 0 \mid V(D_t) = S(D_t)\} = \inf\{t \geq 0 \mid D_t \geq x_*\}. \tag{4.19}$$

where x_* is the demand threshold and is given by

$$x_* = \frac{b\omega_+(\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q\left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}}\right]}, \tag{4.20}$$

where

$$\omega_+ = \frac{\sigma_D^2 - 2\mu_D + \sqrt{(2\mu_D - \sigma_D^2)^2 + 8\sigma_D^2\lambda_d}}{2\sigma_D^2} > 2. \tag{4.21}$$

Proposition 4.3 states that the firm's optimal policy is to start the PPA when the total market demand first reaches x_*, which is a given explicitly by (4.20). We also note that the obtained x_* is positive. This can be seen by noting that $\mu_D < \lambda_d$, $e^{(-\lambda_d + \mu_D)T} < 1$, and

$\omega_+ > 2$.

Next, we have the following corollary, which provides explicit expressions for the optimal invested capacity and the firm's optimal expected saving, as well as distribution of the waiting time before starting the PPA (for a given initial demand).

Corollary 4.1. *From Proposition 4.3 and Proposition 4.1, we can obtain the optimal additional capacity and the optimal expected saving:*

$$K^* = \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]}, \tag{4.22}$$

$$S(x_*) = \frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]}, \tag{4.23}$$

with ω_+ as given in (4.21)*. Moreover, the value function in the continuation region is given by*

$$V(x) = \frac{\frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}} \cdot x^{\omega_+}. \tag{4.24}$$

Furthermore, let the initial demand be some $D_0 < x_*$, then, the optimal time to sign a PPA follows the inverse Gaussian distribution $\mathrm{IG}\left(\frac{\ln\left(\frac{x_*}{D_0}\right)}{\mu_D-\sigma_D^2/2}, \left(\frac{\ln\left(\frac{x_*}{D_0}\right)}{\sigma_D}\right)^2\right)$*, with a mean* $\frac{\ln\left(\frac{x_*}{D_0}\right)}{\mu_D-\sigma_D^2/2}$.

Following Corollary 4.1, we are interested in how K^*, $V(x)$, and $\mathbb{E}[\tau^*]$ change with respect to the production process (described by μ_Q and σ_Q) and T, the length of the PPA. The results are summarized as Proposition 4.4.

Proposition 4.4. *Under the optimal policy, the newly added renewable capacity K^*, the firm's optimal expected saving $S(x_*)$, the value function $V(x)$, and the expected stopping time $\mathbb{E}[\tau^*]$ change as the following with respect to different problem parameters.*

$$\frac{\partial K^*}{\partial \mu_Q} < 0, \qquad \frac{\partial K^*}{\partial \sigma_Q} < 0, \qquad \frac{\partial K^*}{\partial T} < 0, \tag{4.25a}$$

$$\frac{\partial V(x)}{\partial \mu_Q} > 0, \qquad \frac{\partial V(x)}{\partial \sigma_Q} < 0, \tag{4.25b}$$

$$\frac{\partial \mathbb{E}[\tau^*]}{\partial \mu_Q} < 0, \qquad \frac{\partial \mathbb{E}[\tau^*]}{\partial \sigma_Q} = 0, \qquad \frac{\partial \mathbb{E}[\tau^*]}{\partial T} < 0. \tag{4.25c}$$

To demonstrate the results of Proposition 4.4, we also numerically show the changes of K^* and $V(x)$ with respect to μ_Q, σ_Q, and T. In all numerical studies in this chapter, we choose the following "default" parameters (i.e., the non-varying parameters are set to these values when making the plots): $\mu_D = 0.001$, $\sigma_D = 0.015$, $\lambda_d = 0.015$, $\alpha = 0.004$, $b = 300$, $\theta = 4 \times 10^{-14}$, $\mu_Q = 2000$, $\sigma_Q = 80$, $T = 20$, $\hat{T} = 50$, $D_0 = 4 \times 10^{12}$. With these numbers, Figure 4.1 shows how the optimal capacity K^* changes with respect to μ_Q, σ_Q, and T; Figure 4.2 shows how the value function $V(x)$ changes with respect to μ_Q, σ_Q, and T, when $x = D_0$.

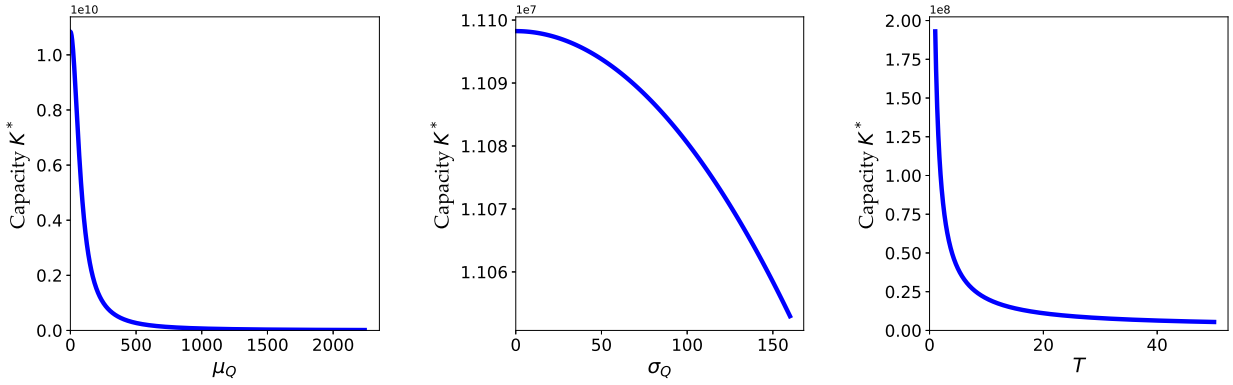

Figure 4.1: Numerical illustration of how the optimal capacity K^* changes with respect to μ_Q, σ_Q, and T.

Proposition 4.4 conveys several messages. First, as shown in (4.25a) and illustrated in

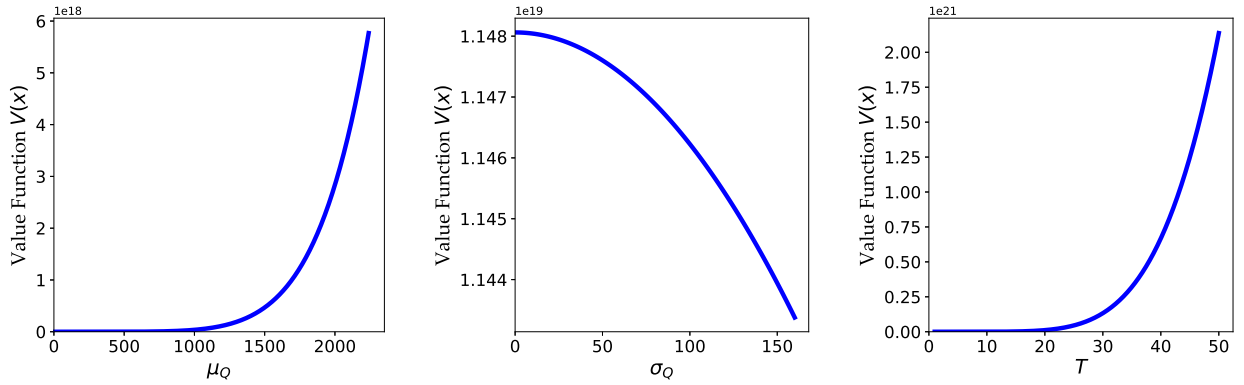

Figure 4.2: Numerical illustration of how the value function $V(x)$ changes with respect to μ_Q, σ_Q, and T.

Figure 4.1, the firm's optimal investment capacity decreases with respect to μ_Q, σ_Q, and T. As the mean production per unit of renewable facility increases, the capacity of newly built renewable facilities is smaller, since the firm may now reach the optimal generation from a smaller amount of facilities. This optimal amount of capacity also decreases with more variance on the generation, since the instability of the generation would likely make the firm benefit less from the renewable facilities. When the length of the PPA is longer, the firm would also have fewer new facilities, as the firm is benefiting for a longer term from each unit of renewable facility.

Second, the value function is the expected discounted saving to go, assuming the firm makes the decisions optimally, and given the current total market demand is x. As shown in (4.25b) and illustrated in Figure 4.2, the value function increases with the production level μ_Q, and decreases with the variance of production σ_Q. While the change of the value function with respect to the length of the PPA is not analytically tractable, it is intuitive that, with all other conditions fixed, the longer the length of the PPA, the more benefit the firm gets. Thus, the value function is higher with a longer PPA, which is also consistent with the numerical studies.

Moreover, as shown in (4.25c), the expected waiting time before the firm starts a PPA decreases with a higher mean production level μ_Q, or with a longer length T, but the expected

waiting time does not change with respect to the variance σ_Q.

We next consider the effect of varying the investment cost, i.e., changing the parameter b in (4.1). The following proposition summarizes how the optimal capacity K^* and the total new generation due to the PPA, $\mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]$, change with respect to b.

Proposition 4.5. *When $D_0 < x_*$, under the firm's optimal policy, increasing the investment cost parameter b results in*

- *a larger capacity for the new renewable facility and*
- *more total new renewable energy output with probability* 1.

Otherwise when $D_0 \geq x_*$, such an increase in b reduces the added renewable energy capacity and production with probability 1.

Numerical illustration of Proposition 4.5 when $D_0 < x_*$ is given in Figure 4.3. When $D_0 \geq x_*$, the firm does not wait and signs the PPA immediately, and the optimal capacity is then given by $K(D_0)$ from (4.13), which decreases as b increases. Effectively, the total new renewable generation due to the PPA also decreases. When $D_0 < x_*$, the firm waits to sign a PPA. As the investment cost increases, the firm delays the PPA to sign it at a larger x_*. Since the total demand is now higher, the wholesale market price is also higher, which gives the firm more motivation to invest for a larger renewable energy capacity. In summary, reducing the investment cost for renewable energy is effective in shortening the firm's time to sign a renewable PPA, as long as the current total demand is not higher than the threshold x_*. If the current demand is already higher than x_*, however, further reducing the investment cost will reduce the capacity of new renewable facilities.

Lastly, we look at the total expected generation from the new capacities due to the PPA. The results are summarized as Proposition 4.6.

Proposition 4.6. *Under the optimal policy, the total expected generation from newly added capacities, over the lifespan of these facilities, is $\mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]$, which changes as follows*

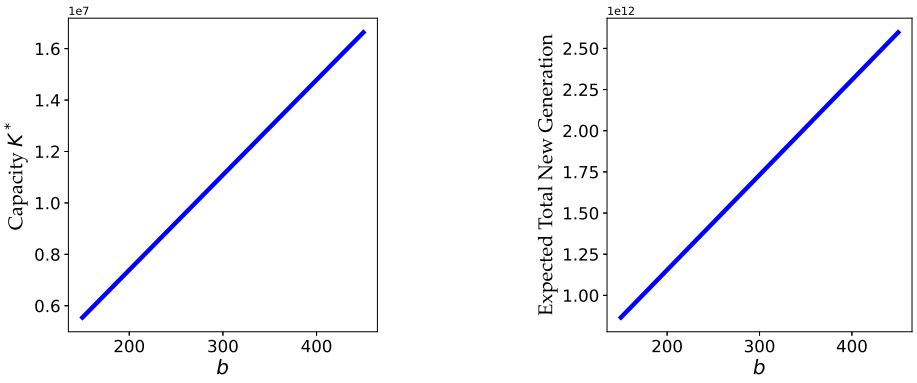

Figure 4.3: Numerical illustration of how the optimal capacity and the total new generation change with respect to b when $D_0 < x_*$.

with respect to different problem parameters.

$$\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} \geq 0 \text{ if } \mu_Q \leq \sigma_Q, \qquad \frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} < 0 \text{ if } \mu_Q > \sigma_Q, \tag{4.26a}$$

$$\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \sigma_Q} < 0, \qquad \frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial T} < 0. \tag{4.26b}$$

We also show (4.26a) numerically in Figure 4.4. Proposition 4.6 implies that when μ_Q is relatively small compared with σ_Q, i.e., the coefficient of variation σ_Q/μ_Q is greater than 1, the total generation from new facilities is increasing with respect to μ_Q; when μ_Q is relatively large compared with σ_Q, i.e., the coefficient of variation σ_Q/μ_Q is smaller than 1, the total generation from new facilities is decreasing with respect to μ_Q. The total expected generation decreases with σ_Q and T, which follows directly from (4.25a).

4.4 Power Purchase Agreement Model with Technology Price Discount

In our previous model, the cost function of investing k units of renewable facilities is assumed to be $I(k) = bk$. This simple price structure of the renewable facilities ensured the tractability

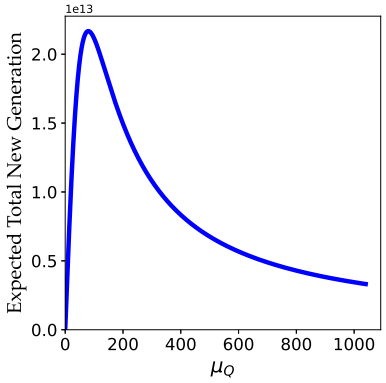

Figure 4.4: Numerical illustration of the change of expected total new generation with respect to μ_Q.

of the model. However, given that the firm's decision to sign a PPA is dynamically evolving, and the firm's waiting time before signing the PPA may span a longer period (in years), it is more convincing that, as the technology advances, the investment cost will decrease over an extended period of time. In this section, we impose an additional discount rate λ_c on the investment cost, i.e., the investment cost function, original given as (4.1), now becomes

$$I_t(k) = e^{-\lambda_c t} bk, \tag{4.27}$$

where b is again some positive constant coefficient. If a PPA is signed at time t, the renewable generator earns the transfer amount C, and builds renewable facilities of capacity K such that

$$I_t(K) = C \Longrightarrow K = \frac{C}{be^{-\lambda_c t}}. \tag{4.28}$$

We will see that this additional discount on investment cost will change the structure of the firm's optimal policy, and the analysis will become more involved. In this section, we derive the optimal solution of the PPA model with technology price discount, and provide a complete analysis on the properties of the optimal investment capacity and the firm's

expected savings.

4.4.1 Signing PPA at Time τ

Similar to Section 4.3, we start with the questions of how much capacity the firm should invest, and how much the firm could save, if it signs the PPA at an arbitrary given time τ. While our focus later will be the firm's dynamic decision on whether to sign a PPA (at each time instance), which is again formulated as an optimal stopping problem, this subsection builds the foundation for our analysis in the following subsections.

When the firm signs a PPA at time τ, its saving (4.10) is optimized by choosing an optimal transfer C. Correspondingly, according to (4.28), $\frac{C}{be^{-\lambda_c \tau}}$ unites of renewable facilities will be built. The following lemma provides the firm's expected saving if a PPA is signed at time τ.

Lemma 4.3. *If a PPA is signed at time τ, and the firm sets the transfer $C = bKe^{-\lambda_c \tau}$ where K is the amount of new renewable facilities built, then, the firm's expected saving from this PPA, discounted to time τ, is given by*

$$K\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)T}\right]+K\frac{\alpha\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right] - K^2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]-bKe^{-\lambda_c\tau}. \tag{4.29}$$

Lemma 4.3 provides the expected saving of the firm in terms of D_τ, the total demand of the spot market at time τ, and K, the capacity of new renewable energy facilities. Since the firm chooses $C = bKe^{-\lambda_c \tau}$, this saving can be further optimized over K. The following proposition presents the optimal capacity and saving of the firm.

Proposition 4.7. *Suppose that the firm signs the PPA at time $\tau > 0$. Then, the newly*

added capacity of renewable energy facilities because of the PPA is

$$K(D_\tau,\tau) := \max\left\{\frac{\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{-\lambda_c\tau}}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]},0\right\}, \tag{4.30}$$

and the firm's optimal expected saving from the PPA is

$$S(D_\tau,\tau) := \begin{cases} \dfrac{\left[\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{-\lambda_c\tau}\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}, \\ \qquad\qquad \text{if } D_\tau \geq \dfrac{be^{-\lambda_c\tau}(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}, \\ 0, \qquad\qquad \text{if } D_\tau < \dfrac{be^{-\lambda_c\tau}(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}. \end{cases} \tag{4.31}$$

From Proposition 4.7, we see that the firm would choose a capacity $K > 0$, and resulting in a positive expected saving S, if and only if $D_\tau \geq \frac{be^{-\lambda_c\tau}(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}$. In other words, when the discounted investment cost $be^{-\lambda_c\tau}$ is high enough, i.e., $be^{-\lambda_c\tau} > \frac{D_\tau\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{\lambda_d-\mu_D}$, then the firm's optimal investment capacity would be zero, meaning that the discounted cost is too high for the firm to make any investment. Note that comparing to the case without technology discount (Proposition 4.1), both the capacity and the saving now have an additional argument τ, i.e., they depend on both the total spot market demand and the time, where the time shows up in the term $be^{-\lambda_c\tau}$. This fact will make the analysis more complicated as we consider the dynamic decision of the firm, which we will show in the next subsection.

Next, we are interested in how the optimal capacity and the saving changes with respect to the production process (described by μ_Q and σ_Q) and T, the length of the PPA. The results are summarized as Proposition 4.2.

Proposition 4.8. *The optimal renewable capacity and the firm's optimal expected saving change with respect to different problem parameters as follows.*

1. *There exists a unique threshold* $\bar{\mu}^d$ *such that* $\frac{\partial K(D_\tau,\tau)}{\partial \mu_Q} < 0$ *if and only if* $\mu_Q > \bar{\mu}^d$.
2. *There exists a unique threshold* $\hat{\mu}^d$ *such that* $\frac{\partial S(D_\tau,\tau)}{\partial \mu_Q} > 0$ *if and only if* $\mu_Q > \hat{\mu}^d$.
3. *If* $\frac{\theta \mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] > be^{-\lambda_c \tau}$, *then, there exists a threshold* T_1 *such that when* $T < T_1$, *we have that* $\frac{\partial K(D_\tau,\tau)}{\partial T} < 0$. *If* $\frac{\theta \mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] \leq be^{-\lambda_c \tau}$, *then, there exists a threshold* T_2 *such that when* $T < T_2$, *we have that* $\frac{\partial K(D_\tau,\tau)}{\partial T} > 0$.
4. *If* $\frac{\theta \mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] \neq be^{-\lambda_c \tau}$, *then, there exists a threshold* T_3 *such that when* $T < T_3$, *we have that* $\frac{\partial S(D_\tau,\tau)}{\partial T} < 0$. *If* $\frac{\theta \mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] \neq be^{-\lambda_c \tau}$, *then, there exists a threshold* T_4 *such that when* $T < T_4$, *we have that* $\frac{\partial S(D_\tau,\tau)}{\partial T} > 0$.
5. $\frac{\partial K(D_\tau,\tau)}{\partial \sigma_Q} < 0$, $\frac{\partial S(D_\tau,\tau)}{\partial \sigma_Q} < 0$.

From the first item of Proposition 4.8, we see that $K(D_\tau, \tau)$ first increases with μ_Q, then after $\mu_Q > \bar{\mu}^d$, the investment capacity starts to decrease. On the other hand, $S(D_\tau, \tau)$ first decreases with μ_Q, then after $\mu_Q > \hat{\mu}^d$, the firm's saving starts to increase. Note that due to the technology price discount, the thresholds $\bar{\mu}^d < \bar{\mu}$ and $\hat{\mu} < \hat{\mu}^d$, comparing to the case without the discount λ_c. The change of $K(D_\tau, \tau)$ and $S(D_\tau, \tau)$ with respect to T are less tractable, but we obtain their limiting behaviors as $T \to 0$. Finally, as the variance of production σ_Q increases, both $K(D_\tau, \tau)$ and $S(D_\tau, \tau)$ decrease.

4.4.2 Optimal Time to Sign PPA - Dynamic Decision

We now come back to the model where the firm needs to dynamically decide $\tau \geq 0$, the time to sign a PPA, and the transfer amount C to the renewable energy generator. Recall that x is a generic notation to represent the realization of the initial total demand for electricity in the spot market. Different from the previous section, the value function now not only depends

on the initial total demand x, but also the current time t, due to the fact that the investment cost coefficient is discounted (with a rate λ_c) over time. We let $V(x,t)$ be the firm's expected saving if it optimally chooses the time to sign the PPA and the transfer payment C given the initial demand realization is x and the current time is t. Recall from (4.28) that optimizing with respect to C is equivalent to optimizing with respect to K, and the optimal saving at the stopping time is already given by (4.31). Therefore, the firm's value function can be written as the following optimal stopping problem:

$$V(x,t) = \max_{\tau \geq 0} \mathbb{E}\left[e^{-\lambda_d \tau} S(D_\tau, \tau) \mid D_t = x\right], \tag{4.32}$$

where we recall that τ is the time to sign a PPA (or the starting time of the PPA, or the time the firm stops waiting).

We assume that $V(x,t)$ is twice continuously differentiable along x (the directional derivative exists and is continuous), continuously differentiable along t, and nonnegative. From (4.31), we know that $S(x,t)$ is continuously differentiable along x and t, nonnegative, and monotone. The decision to start or not to start a PPA at any time t when the realization $D_t = x$ depends on the comparison of $V(x,t)$ and $S(x,t)$. If $V(x,t) > S(x,t)$, the optimal τ^* that solves (4.32) is strictly positive, and it is more beneficial for the firm to wait, since the expected value of waiting is higher than the value of starting a PPA immediately. The set $\{(x,t) \mid V(x,t) > S(x,t)\}$ is called the *continuation region*. If $V(x,t) = S(x,t)$, then one optimal $\tau^* = 0$, and it is optimal for the firm to start the PPA immediately with the expected saving $S(x,t)$. The set $\{(x,t) \mid V(x,t) \leq S(x,t)\}$ is called the *stopping region*. We next have the following lemma on the characterization of $V(x,t)$.

Lemma 4.4. *The value function satisfies the following Hamilton–Jacobi–Bellman (HJB)*

equation:

$$V(x,t) = \max\left\{S(x,t),\ \frac{1}{\lambda_d}\mu_D x V_x(x,t) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V_{xx}(x,t) + \frac{1}{\lambda_d}V_t(x,t)\right\}. \tag{4.33}$$

Lemma 4.4 implies that in the stopping region, $V(x,t) = S(x,t)$, and in the continuation region, we have

$$V(x,t) = \frac{1}{\lambda_d}\mu_D x V_x(x,t) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V_{xx}(x,t) + \frac{1}{\lambda_d}V_t(x,t). \tag{4.34}$$

Contrast to the previous section, the firm's decision to continue waiting or to stop waiting (start a PPA) now depends on both the realized total market demand and the current time. Specifically, at each time instance t, the firm checks the realization of total market demand x, and see whether the pair (x,t) is in the continuation region or the stopping region. As the firm waits, we have that $S(x,t) < \frac{1}{\lambda_d}\mu_D x V_x(x,t) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V_{xx}(x,t) + V_t(x,t)$. The market demand and the time continue to evolve while the firm is waiting, until at some time t, the demand is $x_*(t)$ (or equivalently, the pair (x,t) reaches some $(x_*(t),t)$) such that $S(x_*(t),t) = \frac{1}{\lambda_d}\mu_D x_*(t) V_x(x_*(t),t) + \frac{1}{2\lambda_d}\sigma_D^2 x_*^2(t) V_{xx}(x_*(t),t) + V_t(x_*(t),t)$, at which point the firm would stop waiting. The set

$$\left\{(x,t) \mid V(x,t) = S(x,t) = \frac{1}{\lambda_d}\mu_D x V_x(x,t) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V_{xx}(x,t) + V_t(x,t)\right\}$$

is called the *optimal stopping boundary*. Since both $V(x,t)$ and $S(x,t)$ are continuously differentiable, at $(x_*(t),t)$ in the optimal stopping boundary, we have that

$$V(x_*(t),t) = S(x_*(t),t), \quad \forall(t, x_*(t)) \tag{4.35a}$$

$$V_x(x_*(t),t) = S_x(x_*(t),t), \quad \forall(t, x_*(t)) \tag{4.35b}$$

$$V_t(x_*(t),t) = S_t(x_*(t),t), \quad \forall(t, x_*(t)) \tag{4.35c}$$

where (4.35a) is the *value matching condition*, (4.35b) and (4.35c) are the *smooth pasting conditions*.

It remains to find, for each time t, the critical market demand $x_*(t)$, such that the firm would optimally start a PPA at the first time t when the market demand is $x_*(t)$. In the following proposition, we formalize the optimal policy and specify $(x_*(t), t)$ in the optimal stopping boundary by solving the differential equation (4.34) subject to the boundary conditions (4.35).

Proposition 4.9. *Suppose that the firm dynamically chooses when to sign a PPA with a renewable energy generator. For the firm, it is optimal to sign the PPA at*

$$\tau^* = \inf \{t \geq 0 \mid D_t \geq x_*(t)\}. \tag{4.36}$$

where $x_*(t)$ is the demand threshold function of t and is given by

$$x_*(t) = \frac{b\omega_+ e^{-\lambda_c t}(\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}}\right]}, \tag{4.37}$$

where

$$\omega_+ = \frac{\sigma_D^2 - 2\mu_D - 2\lambda_c + \sqrt{\sigma_D^4 + 4\mu_D^2 + 4\lambda_c^2 + 12\sigma_D^2\lambda_c + 8\mu_D\lambda_c - 4\sigma_D^2\mu_D + 8\sigma_D^2\lambda_d}}{2\sigma_D^2} > 2. \tag{4.38}$$

Proposition 4.9 states that the firm's optimal policy is to start the PPA at the first time t when the total market demand matches $x_*(t)$, which is a given explicitly by (4.37). We also note that the obtained $x_*(t)$ is positive. This can be seen by noting that $\mu_D < \lambda_d$, $e^{(-\lambda_d + \mu_D)T} < 1$, and $\omega_+ > 2$. To avoid trivial cases, we assume that $D_0 < x_*(0)$ for the rest of this section, since otherwise the firm would start the PPA immediately at time $t = 0$ without waiting.

Next, we have the following corollary, which provides explicit expressions for the optimal invested capacity and the firm's optimal expected saving, as well as distribution of the waiting time before starting the PPA (for a given initial demand).

Corollary 4.2. *From Proposition 4.9 and Proposition 4.7, we can obtain the optimal additional capacity and the optimal expected saving:*

$$K^* = \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}, \tag{4.39}$$

$$S^* = \frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}, \tag{4.40}$$

with ω_+ *as given in* (4.38)*, and* $D_* := x_*(0) = \frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}$.

Moreover, the value function in the continuation region is given by

$$V(x,t) = \frac{\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}} \cdot x^{\omega_+} e^{\hat{\lambda}t}, \tag{4.41}$$

where $\hat{\lambda} = \frac{-\sigma_D^2(3\sigma_D^2+2\mu_D)\lambda_c - 2\sigma_D^2\lambda_c^2 + \sqrt{\sigma_D^4\lambda_c^2\left[\sigma_D^4+4\mu_D^2+4\lambda_c^2+12\sigma_D^2\lambda_c+8\mu_D\lambda_c-4\sigma_D^2\mu_D+8\sigma_D^2\lambda_d\right]}}{2\sigma_D^4}$.

Furthermore, let the initial demand be some $D_0 < D_*$*, then, the optimal time to sign a PPA follows the inverse Gaussian distribution* $\mathrm{IG}\left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c+\mu_D-\sigma_D^2/2}, \left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\sigma_D}\right)^2\right)$*, with a mean* $\frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c+\mu_D-\sigma_D^2/2}$.

Following Corollary 4.2, we are interested in how K^*, S^*, $V(x,t)$, and $\mathbb{E}[\tau^*]$ change with respect to the production process (described by μ_Q and σ_Q) and T, the length of the PPA. The results are summarized as Proposition 4.10.

Proposition 4.10. *Under the optimal policy, the newly added renewable capacity K^*, the firm's optimal expected saving S^*, the value function $V(x,t)$, and the expected stopping time $\mathbb{E}[\tau^*]$ change as the following with respect to different problem parameters.*

$$\begin{cases} \frac{\partial K^*}{\partial \mu_Q} \geq 0, & \text{if } \frac{\sigma_Q^2}{\mu_Q^2} \geq \xi \\ \frac{\partial K^*}{\partial \mu_Q} < 0, & \text{if } \frac{\sigma_Q^2}{\mu_Q^2} < \xi \end{cases}, \qquad \frac{\partial K^*}{\partial \sigma_Q} < 0, \tag{4.42a}$$

$$\frac{\partial V(x,t)}{\partial \mu_Q} > 0, \qquad \frac{\partial V(x,t)}{\partial \sigma_Q} < 0, \tag{4.42b}$$

$$\frac{\partial \mathbb{E}[\tau^*]}{\partial \mu_Q} < 0, \qquad \frac{\partial \mathbb{E}[\tau^*]}{\partial \sigma_Q} = 0, \qquad \frac{\partial \mathbb{E}[\tau^*]}{\partial T} < 0, \tag{4.42c}$$

where ξ is some constant that only depends on μ_D, σ_D, and λ_c.

To demonstrate the results of Proposition 4.10, we also numerically show the changes of K^* and $V(x)$ with respect to μ_Q, σ_Q, and T. The "default" parameters (i.e., the non-varying parameters are set to these value when making the plots) are set to the same as in the previous section: $\mu_D = 0.001$, $\sigma_D = 0.015$, $\lambda_d = 0.015$, $\alpha = 0.004$, $b = 300$, $\theta = 4 \times 10^{-14}$, $\mu_Q = 2000$, $\sigma_Q = 80$, $T = 20$, $\hat{T} = 50$, $D_0 = 4 \times 10^{12}$. In addition, we set $\lambda_c = 0.2$. With these numbers, Figure 4.5 shows how the optimal capacity K^* changes with respect to μ_Q, σ_Q, and T; Figure 4.6 shows how the value function $V(x,t)$ changes with respect to μ_Q, σ_Q, and T, when $t = 5$ and $x = D_0 e^{5\mu_D}$.

Proposition 4.10 conveys several messages. First, as shown in (4.42a) and illustrated in Figure 4.5, the firm's optimal investment capacity decreases with respect to μ_Q when σ_Q^2/μ_Q^2, square of the coefficient of variation of the production, is smaller than some constant ξ; on the other hand when it is greater than ξ, we have K^* increasing with respect to μ_Q. This is different from Proposition 4.4 (when there is no discount λ_c on the investment cost) where the firm's optimal investment capacity is always decreasing with respect to μ_Q. This might be due to the fact that with additional discount on the investment cost, the firm faces a

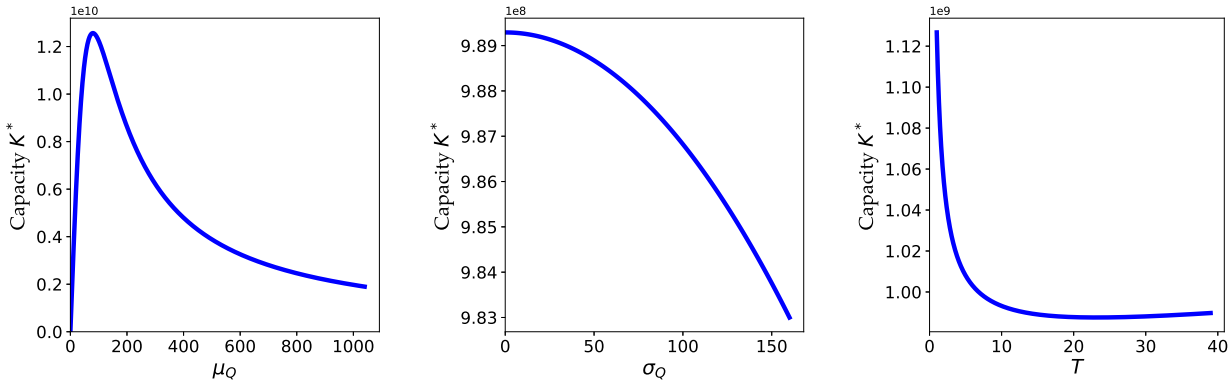

Figure 4.5: Numerical illustration of how the optimal capacity K^* changes with respect to μ_Q, σ_Q, and T.

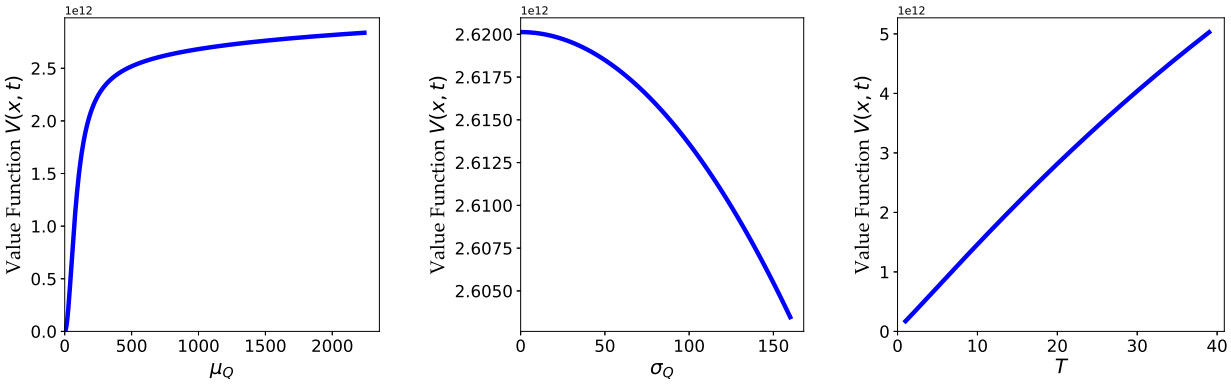

Figure 4.6: Numerical illustration of how the value function $V(x,t)$ changes with respect to μ_Q, σ_Q, and T.

lower investment cost, so that when the coefficient of variation is high, the firm would invest more capacity (at the discounted cost) with a higher mean production. When the coefficient of variation is low, however, the firm does not need more capacity to hedge the production variance, and with a higher μ_Q, the firm can reach the optimal generation from less capacity. This optimal amount of capacity also decreases with more variance on the generation, since the instability of the generation would likely make the firm benefit less from the renewable facilities. With the additional discount on investment cost, it is not analytically tractable to analyze the change of K^* with respect to the length of the PPA. In the right of Figure 4.5, we see that the optimal capacity first decreases and then increases with T. Intuitively, when T

is within the smaller range, a higher T will let the firm decrease the capacity since the firm can benefit each unit of capacity for a longer term, and this behavior is similar to how K^* changes with respect to T in previous section. When T is in the larger range, however, the optimal capacity starts to increase, since the additional discount on the investment cost has made it cheaper to invest in renewable facilities, and a higher capacity provides more benefit to the firm when T is longer.

Second, the value function is the expected discounted saving to go, assuming the firm makes the decisions optimally, given the current time t and the current total market demand x. As shown in (4.42b) and illustrated in Figure 4.6, the value function increases with the production level μ_Q, and decreases with the variance of production σ_Q. While the change of the value function with respect to the length of the PPA is not analytically tractable, it is intuitive that, with all other conditions fixed, the longer the length of the PPA, the more benefit the firm gets. Thus, the value function is higher with a longer PPA, which is also consistent with the numerical studies.

The reasoning for how the value function and the expected stopping time changes with respect to different problem parameters is the similar to those described after Proposition 4.4.

We next consider the effect of varying the investment cost, i.e., changing the parameter b in (4.1). The following proposition summarizes how the optimal capacity K^* and the total new generation due to the PPA, $\mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]$, change with respect to b.

Proposition 4.11. *At any time t, when $D_t < x_*(t)$, under the firm's optimal policy, increasing the investment cost parameter b results in*

- *a larger capacity for the new renewable facility and*
- *more total new renewable energy output with probability* 1.

Otherwise when $D_t \geq x_*(t)$, such an increase in b reduces the added renewable energy capacity and production with probability 1.

Numerical illustration of Proposition 4.11 when $D_t < x_*(t)$ is given in Figure 4.7. When $D_t \geq x_*(t)$, the firm does not wait anymore and signs the PPA immediately at time t, and the optimal capacity is then given by $K(D_t, t)$ from (4.30), which decreases as b increases. Effectively, the total new renewable generation due to the PPA also decreases. When $D_t < x_*(t)$, the firm waits to sign a PPA. As the investment cost increases, the firm delays the PPA to sign it at a larger $x_*(t)$. Since the total demand is now higher, the wholesale market price is also higher, which gives the firm more motivation to invest for a larger renewable energy capacity. In summary, reducing the investment cost for renewable energy is effective in shortening the firm's time to sign a renewable PPA, as long as the current total demand is not higher than the threshold $x_*(t)$. If the current demand is already higher than $x_*(t)$, however, further reducing the investment cost will reduce the capacity of new renewable facilities.

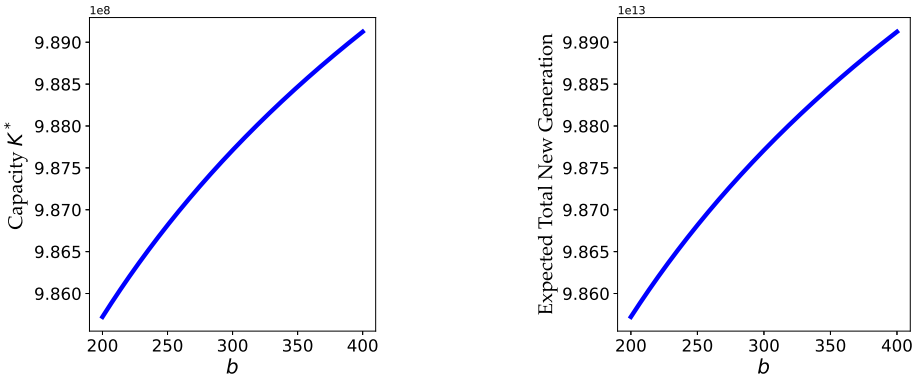

Figure 4.7: Numerical illustration of how the optimal capacity and the expected total new generation change with respect to b when $D_t < x_*(t)$.

Lastly, we look at the total expected generation from the new capacities due to the PPA. The results are summarized as Proposition 4.12.

Proposition 4.12. *Under the optimal policy, the total expected generation from newly added capacities, over the lifespan of these facilities, is* $\mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]$, *which changes as follows*

with respect to different problem parameters.

$$\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} \geq 0 \text{ if } \frac{\sigma_Q^2}{\mu_Q^2} \geq \zeta, \quad \frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} < 0 \text{ if } \frac{\sigma_Q^2}{\mu_Q^2} < \zeta, \tag{4.43a}$$

$$\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \sigma_Q} < 0, \tag{4.43b}$$

where ζ is some constant that only depends on μ_D, σ_D, and λ_c.

In Figure 4.8, we also numerically show how $\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]$ changes with respect to μ_Q. Proposition 4.12 implies that when the coefficient of variation σ_Q/μ_Q is large (greater than $\sqrt{\zeta}$), the total generation from new facilities is increasing with respect to μ_Q; when σ_Q/μ_Q is small (smaller than $\sqrt{\zeta}$), the total generation from new facilities is decreasing with respect to μ_Q. This result is similar to Proposition 4.6 when there is no discount λ_c on investment cost, except that in Proposition 4.6, we essentially have the threshold $\zeta = 1$. In Proposition 4.12, however, one can easily verify that the threshold $\zeta > 1$. In other words, when the investment cost becomes cheaper over time, the threshold on the coefficient of variance becomes higher, e.g., if $1 < \sigma_Q^2/\mu_Q^2 < \zeta$, the total generation from new facilities increases with respect to the mean production when there is no discount on investment cost, but decreases with respect to the mean production when there is discount on investment cost.

The total expected generation from new facilities decreases with σ_Q, which follows directly from (4.42a).

4.5 Power Purchase Agreement: Advanced Planning

In this section, we study an extension of the power purchase agreement (PPA) model with technology price discount, where the firm needs to commit at the very beginning if it would

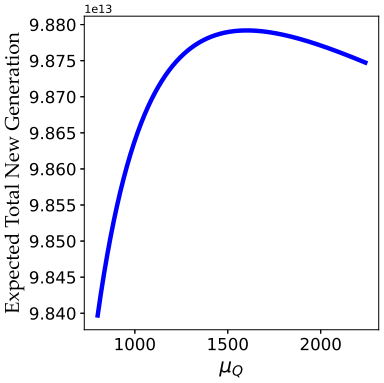

Figure 4.8: Numerical illustration of the change of expected total new generation with respect to μ_Q.

sign a PPA, and if yes, when the PPA would start and how much to invest. In other words, the firm is no longer dynamically observing realized the market conditions (demands) and making decisions, but is instead making a one-time decision at time $t = 0$, based on the expectations future electricity demand and renewable generations.

Specifically, at time $t = 0$, the firm sees $\{Q_t, t \geq 0\}$, the production amount of one unit of the renewable facility, announced by the renewable energy generator, as well as the cost function for the investment of renewable facilities. Then, the firm decides if it would sign a PPA of length T with the renewable generator. In a PPA, the firm chooses a starting time t_s to minimize its total expected cost. It also specifies a transfer payment C to be paid at time $t = 0$, and has access to all electricity production from the renewable generator from t_s to $t_s + T$. All these specifications of the PPA are also determined at time $t = 0$ immediately following the renewable generator's announcement, and therefore we refer to this type of agreement as *advanced planning* PPA. The renewable generator is assumed to be a passive decision maker who will sign the PPA as long as C is nonnegative. If a PPA is signed, the renewable generator invests all C to maximize the size (capacity) of the new renewable energy facilities.

If the PPA starts at time t_s, the cost function of investing k units of renewable facilities

is assumed to be

$$I_{t_s}(k) = e^{-\lambda_c t_s} bk, \tag{4.44}$$

where b is some positive constant and λ_c is the technology discount rate. If the transfer payment is C and the PPA starts at time t_s, then, the newly built facilities of capacity K is given by

$$I_{t_s}(K) = C \Longrightarrow K = \frac{C}{be^{-\lambda_c t_s}}. \tag{4.45}$$

The firm needs to decide at time $t = 0$ if it would sign a PPA, and if so, how much the transfer payment would be and when the PPA would start, with the objective of maximizing the total discounted saving (throughout the whole time horizon) from the PPA. Recall that the firm's demand at each time instant is given by the stochastic process $\{U_t, t \geq 0\}$, which can be satisfied by the electricity procured from the spot market at the spot market price, or by the generation from the PPA. We again let p_t^N denote the spot market price at time t, assuming no PPA is signed, and let p_t^Y denote the spot market price at time t if a PPA is signed. Recall that Q_t is the random generation from each unit of renewable facility at time t, and $\hat{Q}_t := KQ_t$ is the total generation from the new renewable facility by the PPA. Let λ_d be the discount rate of cash dollar (real interest rate). Then, without a PPA, the firm's total cost for procuring electricity from the spot market, discounted to time $t = 0$, is given by

$$\int_0^\infty e^{-\lambda_d s} p_s^N U_s ds. \tag{4.46}$$

With a PPA starting from t_s and lasting till $t_s + T$, the firm's total cost for procuring

electricity from the spot market becomes

$$\int_0^{t_s} e^{-\lambda_d s} p_s^Y U_s ds + \int_{t_s}^{t_s+T} e^{-\lambda_d s} p_s^Y \left[U_s - \hat{Q}_s\right] ds + \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} p_s^Y U_s ds + \int_{t_s+\hat{T}}^{\infty} e^{-\lambda_d s} p_s^Y U_s ds, \tag{4.47}$$

where we recall that $\hat{T}$ is the lifespan of the renewable facilities.

Note that the net electricity demand N_t is the same with or without the PPA before t_s or after $t_s + \hat{T}$. Therefore, $p_t^Y = p_t^N$ for $t < t_s$ and $t > t_s + \hat{T}$. The firm's saving from signing a PPA is given by

$$(4.46) - (4.47) - e^{-\lambda_d t_s} C = \int_{t_s}^{t_s+T} e^{-\lambda_d s} p_s^N U_s ds - \int_{t_s}^{t_s+T} e^{-\lambda_d s} p_s^Y \left[U_s - \hat{Q}_s\right] ds + \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} p_s^N U_s ds - \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} p_s^Y U_s ds - e^{-\lambda_d t_s} C. \tag{4.48}$$

The firm aims to maximize its expected saving from signing a PPA, i.e.,

$$\max_{t_s, C \geq 0} \mathbb{E}\left[(4.48)\right]. \tag{4.49}$$

The optimization of (4.49) can be divided into two steps: first, for any given t_s, choose an optimal C; then, optimize over t_s. In the following lemma, we first show the firm's expected saving when the PPA starts at t_s and the firm chooses a transfer amount C, or equivalently, an investment capacity $K = \frac{C}{be^{-\lambda_c t_s}}$.

Lemma 4.5. *If a PPA is signed at time $t = 0$ and scheduled to start at time t_s, and the firm sets the transfer $C = e^{-\lambda_c t_s} bK$, where K is the amount of new renewable facilities to*

be built, then, the firm's expected saving from this PPA, discounted to time $t = 0$, is given by

$$\frac{\theta K^2\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[e^{-\lambda_d(t_s+T)}-e^{-\lambda_d t_s}\right] + \frac{\theta\mu_Q D_0 K}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+T)}+\alpha e^{(\mu_D-\lambda_d)(t_s+\hat{T})}-(1+\alpha)e^{(\mu_D-\lambda_d)t_s}\right]-be^{-(\lambda_d+\lambda_c)t_s}K. \tag{4.50}$$

Lemma 4.5 provides the expected saving of the firm in terms of D_0, t_s, and K. This saving can be further optimized over K, which leads to the following proposition.

Proposition 4.13. *Suppose that the firm signs the PPA at time $t = 0$, and the PPA is scheduled to start at time t_s. Then, the newly added renewable facilities because of the PPA is*

$$K(t_s)=\max\left\{\frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]e^{\mu_D t_s}}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]},0\right\}, \tag{4.51}$$

and if $K(t_s) > 0$, the firm's optimal expected saving from the PPA is

$$S(t_s)=\frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]^2 e^{(2\mu_D-\lambda_d)t_s}}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}. \tag{4.52}$$

From Proposition 4.13, we see that the firm would choose a capacity $K > 0$, resulting in a positive expected saving S, if and only if

$$D_0 \geq \frac{(\lambda_d-\mu_D)be^{-(\lambda_d+\lambda_c)t_s}}{\theta\mu_Q\left[(1+\alpha)e^{(\mu_d-\lambda_d)t_s}-e^{(\mu_d-\lambda_d)(t_s+T)}-\alpha e^{(\mu_d-\lambda_d)(t_s+\hat{T})}\right]}.$$

In other words, when the discounted investment cost $be^{-(\lambda_d+\lambda_c)t}$ is high enough, i.e.,

$$be^{-(\lambda_d+\lambda_c)t_s} > \frac{D_0\theta\mu_Q\left[(1+\alpha)e^{(\mu_d-\lambda_d)t_s} - e^{(\mu_d-\lambda_d)(t_s+T)} - \alpha e^{(\mu_d-\lambda_d)(t_s+\hat{T})}\right]}{\lambda_d-\mu_D},$$

then the firm's optimal investment capacity would be zero, meaning that the cost is too high for the firm to make any investment.

Next, the firm also needs to choose t_s, the starting time of the PPA. In other words, the firm needs to optimize $S(t_s)$ over t_s, subject to the constraint that $be^{-(\lambda_d+\lambda_c)t_s} > \frac{D_0\theta\mu_Q\left[(1+\alpha)e^{(\mu_d-\lambda_d)t_s}-e^{(\mu_d-\lambda_d)(t_s+T)}-\alpha e^{(\mu_d-\lambda_d)(t_s+\hat{T})}\right]}{\lambda_d-\mu_D}$. This leads to the following proposition.

Proposition 4.14. *Suppose that the firm signs a PPA at time $t = 0$, with the objective of maximizing the long-term expected saving. Then, the optimal scheduled starting time of the PPA is*

$$t_s^* = \max\left\{\frac{\log\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)}{\lambda_c+\mu_D}, 0\right\}. \tag{4.53}$$

When $t_s^* > 0$, the newly added renewable facilities because of the PPA is

$$K^* = \frac{\left[\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]}{\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}, \tag{4.54}$$

and the firm's optimal expected saving from the PPA is

$$S^* = \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]^2}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot\left(\frac{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{b(2\lambda_c+\lambda_d)}\right)^{\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}}. \tag{4.55}$$

Following Proposition 4.14, we are interested in how K^*, S^*, t_s^*, and the total expected additional generation $\mathbb{E}\left[K^*\int_0^{\hat{T}} Q_t dt\right]$ change with respect to the production process (μ_Q and σ_Q) and the investment cost b. The results are summarized as Proposition 4.15.

Proposition 4.15. *When the firm signs the PPA at time $t=0$, with an optimal scheduled starting time $t_s^*>0$, the newly added renewable capacity K^*, the firm's optimal expected saving S^*, and the optimal starting time t_s^* change as the following with respect to different problem parameters.*

$$\begin{cases}\frac{\partial K^*}{\partial\mu_Q}\geq 0, & \text{if } \frac{\sigma_Q^2}{\mu_Q^2}\geq\frac{\lambda_c+2\mu_D}{\lambda_c}\\ \frac{\partial K^*}{\partial\mu_Q}<0, & \text{if } \frac{\sigma_Q^2}{\mu_Q^2}<\frac{\lambda_c+2\mu_D}{\lambda_c}\end{cases}, \qquad \frac{\partial K^*}{\partial\sigma_Q}<0, \qquad \frac{\partial K^*}{\partial b}>0, \tag{4.56a}$$

$$\frac{\partial S^*}{\partial\mu_Q}\geq 0, \qquad \frac{\partial S^*}{\partial\sigma_Q}<0, \qquad \frac{\partial S^*}{\partial b}<0, \tag{4.56b}$$

$$\frac{\partial t_s^*}{\partial\mu_Q}<0, \qquad \frac{\partial t_s^*}{\partial\sigma_Q}=0, \qquad \frac{\partial t_s^*}{\partial T}<0, \qquad \frac{\partial t_s^*}{\partial b}\geq 0. \tag{4.56c}$$

Furthermore, the total expected generation from newly added capacities, over the lifespan of these facilities, is $\mathbb{E}\left[K^*\int_0^{\hat{T}} Q_t dt\right]$, which changes as follows with respect to different problem

parameters.

$$\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} \geq 0 \text{ if } \frac{\sigma_Q^2}{\mu_Q^2} \geq \frac{\mu_D}{2\lambda_c + \mu_D}, \quad \frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} < 0 \text{ if } \frac{\sigma_Q^2}{\mu_Q^2} < \frac{\mu_D}{2\lambda_c + \mu_D}, \tag{4.57a}$$

$$\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \sigma_Q} < 0, \qquad \frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial b} > 0. \tag{4.57b}$$

To demonstrate the results of Proposition 4.15, we also numerically show the changes of K^*, S^*, and $\mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]$ with respect to μ_Q, σ_Q, T, and b. The "default" parameters (i.e., the non-varying parameters are set to default when making the plots) are set to the same as in the previous section: $\mu_D = 0.001$, $\sigma_D = 0.015$, $\lambda_d = 0.015$, $\alpha = 0.004$, $b = 300$, $\theta = 4 \times 10^{-14}$, $\mu_Q = 2000$, $\sigma_Q = 80$, $T = 20$, $\hat{T} = 50$, $D_0 = 4 \times 10^{12}$, and $\lambda_c = 0.2$. With these numbers, Figure 4.9 and Figure 4.10 show how the optimal capacity K^* and the optimal saving S^*, respectively, change with respect to $mu-Q$, σ_Q, and T; Figure 4.11 shows how the optimal capacity K^*, optimal saving S^*, and the expected total new generation due to the PPA change with respect to the investment cost b; Figure 4.12 shows the change of expected total new generation with respect to μ_Q.

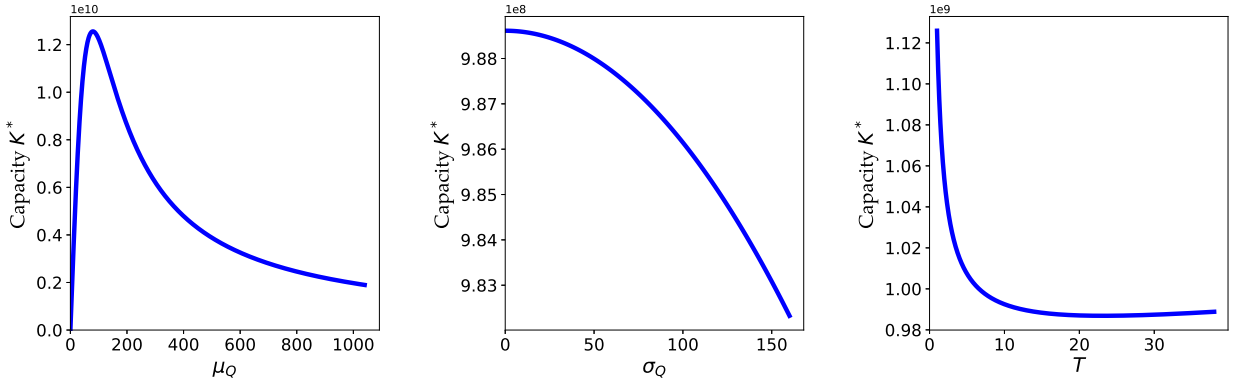

Figure 4.9: Numerical illustration of how the optimal capacity K^* changes with respect to μ_Q, σ_Q, and T.

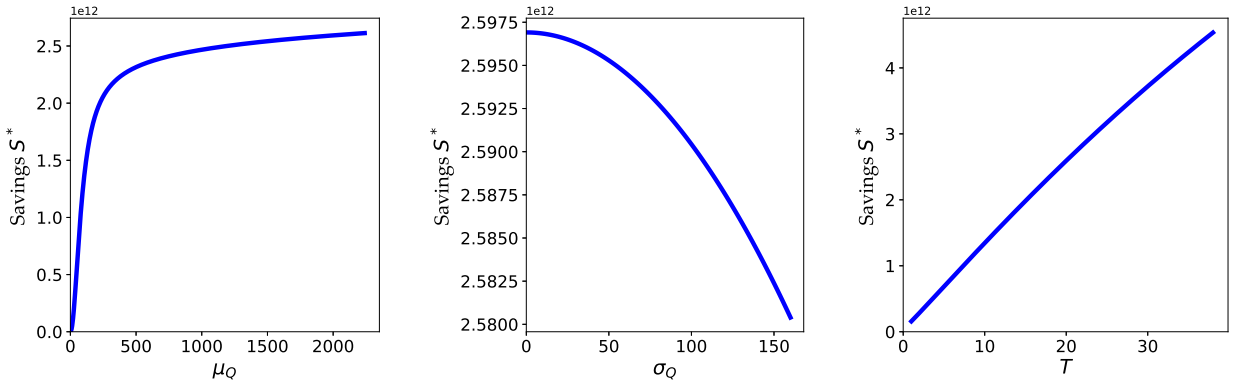

Figure 4.10: Numerical illustration of how the firm's optimal saving S^* changes with respect to μ_Q, σ_Q, and T.

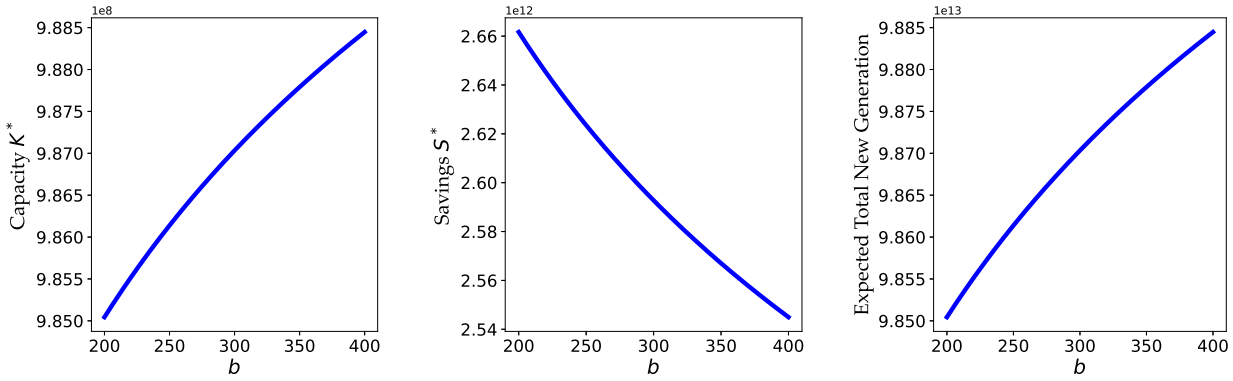

Figure 4.11: Numerical illustration of how the optimal capacity, the optimal saving, and the expected total new generation change with respect to b when $t_s^* > 0$.

Proposition 4.15 conveys several messages. First, as shown in (4.56a) and illustrated in Figure 4.9, the firm's optimal investment capacity K^* decreases with respect to μ_Q when σ_Q^2/μ_Q^2, square of the coefficient of variation of the production, is smaller than $\frac{\lambda_c+2\mu_D}{\lambda_c}$; on the other hand when it is greater than $\frac{\lambda_c+2\mu_D}{\lambda_c}$, we have K^* decreasing with respect to μ_Q. This is part similar to Proposition 4.10 in the dynamic decision PPA, though the constant threshold for σ_Q^2/μ_Q^2 is different. The optimal capacity changes in a similar way with respect to σ_Q and T, comparing to the previous section. Next, as shown in (4.56b) and illustrated in Figure 4.10, the firm's optimal saving is monotonically increasing with respect to μ_Q, decreasing with respect to σ_Q, and increasing with respect to T. These results are similar to

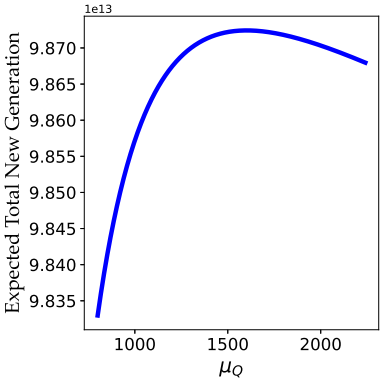

Figure 4.12: Numerical illustration of the change of expected total new generation with respect to μ_Q.

how the value function $V(x,t)$ changes in the previous section, since the firm now makes all decisions at time 0, the firm's optimal saving is the same as the value function with $t = 0$. We also note that, as shown in (4.56c), the starting time t_s^* is earlier when μ_Q is larger, or when the length of the PPA is longer, similar to the results in the previous section.

As shown in (4.15), (4.57b), and illustrated in Figure 4.11, the optimal capacity K^* and the expected total new generation increase with respect to the investment cost b. These results and the reasoning are similar to those in the previous section. The optimal saving S^* decreases with respect to b.

The expected generation from newly added capacities $\mathbb{E}\left[K^* Q_t \hat{T}\right]$ decreases with μ_Q when the square of the coefficient of variation of the production is smaller than $\frac{\mu_D}{2\lambda_c+\mu_D}$, but increases with μ_Q when $\sigma_Q^2/\mu_Q^2 \geq \frac{\mu_D}{2\lambda_c+\mu_D}$, as shown in (4.57a) and illustrated in Figure 4.12. The expected amount of new generation also decreases with σ_Q, and its behavior with respect to T would also be similar to K^* with respect to T.

4.6 Conclusions and Future Directions

In this chapter, we have proposed a power purchase agreement (PPA) model between the firm and the renewable generator. We have formulated the firm's dynamic decision problem

on when to start the PPA of certain length and how much to invest (transfer) as an infinite horizon optimal stopping problem. We have defined the value function, derived the HJB equation, and solved the optimal policy for the firm, i.e., the firm should sign the PPA as soon as the total electricity demand in the market reaches some constant threshold. We have concluded that with an increased PV efficiency μ_Q or with an increased T, the length of the PPA, the firm will optimally sign a PPA earlier, with a smaller capacity of new renewable facilities, and the firm attains higher expected value. The optimal capacity and the total new renewable generation may increase or decrease with the investment cost b, depending on the initial total demand. Moreover, the expected total new generation from the PPA increases (resp. decreases) with μ_Q, if the coefficient of variation σ_Q/μ_Q is greater (resp. smaller) than 1.

Following this model, we have also considered the same formulation but with a decreasing investment cost over time. This additional discount on the investment cost makes the model more complicated to analyze, but we have again obtained the firm's optimal policy: the firm should start the PPA as soon as the total demand in the market hits some time-dependent threshold. The effect of the PV efficiency, the length of the PPA, and the investment cost are similar to those in the previous model, with slight differences. Specifically, the optimal capacity might increase or decrease with respect to the PV efficiency μ_Q, depending on the coefficient of variation σ_Q/μ_Q. The optimal capacity also first decreases and then slightly increases with the length of the PPA.

Finally, we have studied the model where the firm needs to commit at the beginning if it would sign a PPA, and if yes, when to start it and how much to invest. In this case, the firm no longer makes dynamic decisions. We have also characterized the effect of the production level, the length of the PPA, and the investment cost on the firm's optimal capacity, savings, and the total new generation due to the PPA.

There are several follow-up research questions that we may consider. First, many govern-

ments are starting to offer rebates or tax credits for building new renewable energy facilities. These rebates or tax credits might further change the formulation of the PPA and may result in a different optimal investment policy for the firm. Second, we may also consider the periodic behavior of the energy demand and renewable energy generation, i.e., at different seasons, the demand process D_t and the production process Q_t may no longer be stationary, but instead become periodically time-varying. While these non-stationary processes will significantly complicate the analysis, it would be interesting to explore if the firm's optimal policy also exhibits some type of periodic behavior. Lastly, it would also be meaningful to test our model with real-world renewable energy generation data and demand data, which may hopefully provide more insights to the business as well as to policy makers.

4.7 Appendix

4.7.1 Proofs for Section 4.3

Proof of Lemma 4.1. Recall that $p_t^N = \theta N_t = \theta D_t$ and $p_t^Y = \theta N_t = \theta(D_t - KQ_t)$. We also have that $U_t = \alpha D_t$ and $C = bK$. Before taking the expectation, the firm's random saving if signing a PPA at time τ is

$$
\begin{aligned}
(4.10) = &\int_\tau^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^N U_s ds + \int_{t+T}^{t+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^N U_s ds \\
&- \int_\tau^{\tau+T} e^{-\lambda_d(s-\tau)} p_s^Y \left[U_s - \hat{Q}_s\right] ds - \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} p_s^Y U_s ds - C \\
= &\int_\tau^{\tau+T} e^{-\lambda_d(s-\tau)} \theta D_s U_s ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} \theta D_s U_s ds \\
&- \int_\tau^{\tau+T} e^{-\lambda_d(s-t)} \theta (D_s - KQ_s) \left[U_s - KQ_s\right] ds \\
&- \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} \theta (D_s - KQ_s) U_s ds - bK \\
= &\int_\tau^{\tau+T} e^{-\lambda_d(s-\tau)} \theta D_s U_s ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)} \theta D_s U_s ds
\end{aligned}
$$

$$
\begin{aligned}
&- \int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}\theta\left[D_sU_s - KD_sQ_s - KQ_sU_s + K^2Q_s^2\right]ds \\
&- \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)}\theta(D_s - KQ_s)U_s ds - bK \\
&= \int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}\theta\left[KD_sQ_s + KQ_sU_s - K^2Q_s^2\right]ds + \int_{\tau+T}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)}\theta KQ_sU_s ds \\
&\quad - bK \\
&= \theta K\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}D_sQ_s ds + \theta K\int_{\tau}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)}Q_sU_s ds \\
&\quad - \theta K^2\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}Q_s^2 ds - bK \\
&= \theta K\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}D_sQ_s ds + \alpha\theta K\int_{\tau}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)}Q_sD_s ds \\
&\quad - \theta K^2\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}Q_s^2 ds - bK.
\end{aligned}
$$

We further note that, from our assumptions on Q_t and D_t, $\mathbb{E}[Q_t] = \mu_Q$ and $\mathbb{E}[D_{t+t'}] = D_te^{\mu_D t'}$. Thus, the expectation of this saving is

$$
\begin{aligned}
\mathbb{E}\left[(4.10)\right] &= \theta K\mu_Q\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}D_\tau e^{\mu_D(s-\tau)}ds + \alpha\theta K\mu_Q\int_{\tau}^{\tau+\hat{T}} e^{-\lambda_d(s-\tau)}D_\tau e^{\mu_D(s-\tau)}ds \\
&\quad - \theta K^2\left(\sigma_Q^2 + \mu_Q^2\right)\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}ds - bK \\
&= K\theta\mu_QD_\tau\int_{\tau}^{\tau+T} e^{(\mu_D-\lambda_d)(s-\tau)}ds + K\alpha\theta\mu_QD_\tau\int_{\tau}^{\tau+\hat{T}} e^{(\mu_D-\lambda_d)(s-\tau)}ds \\
&\quad - K^2\theta\left(\sigma_Q^2 + \mu_Q^2\right)\int_{\tau}^{\tau+T} e^{-\lambda_d(s-\tau)}ds - bK \\
&= K\frac{\theta\mu_QD_\tau}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D-\lambda_d)T}\right] + K\frac{\alpha\theta\mu_QD_\tau}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D-\lambda_d)\hat{T}}\right] \\
&\quad - K^2\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_dT}\right] - bK.
\end{aligned}
$$

This completes the proof of Lemma 4.1. □

Proof of Proposition 4.1. From Lemma 4.1, the firm's expected discounted saving (4.12) is

a quadratic function of K:

$$K\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)T}\right]+K\frac{\alpha\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right]$$
$$-K^2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]-bK.$$

Since second order coefficient $-\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]$ is negative, we conclude that the expected discounted saving is concave in K. Moreover, note that the capacity can only be nonnegative. From this and the first-order condition, we conclude that the optimal capacity is

$$K(D_\tau)=\max\left\{\frac{\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)T}\right]+\frac{\alpha\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right]-b}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]},0\right\}$$
$$=\max\left\{\frac{\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]},0\right\},$$

and the corresponding saving is

$$S(D_\tau):=\begin{cases}\dfrac{\left[\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}, & \\ \qquad\qquad\qquad\qquad \text{if } \frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\geq b, & \\ 0, \qquad\qquad\qquad\qquad \text{if } \frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]<b. & \end{cases}$$

□

Proof of Proposition 4.2. We prove each item separately.

1. We take the derivative of $K(D_\tau)$ with respect to μ_Q.

$$\frac{\partial K(D_\tau)}{\partial \mu_Q} = \frac{\partial \frac{\frac{\theta \mu_Q D_\tau}{\mu_D - \lambda_d}\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right] - b}{2\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right]}}{\partial \mu_Q} = \frac{num}{4\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right]\right)^2},$$

where

$$\begin{aligned} num = &\frac{\theta D_\tau}{\mu_D - \lambda_d}\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right] \cdot 2\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right]\right) \\ &- \left(\frac{\theta \mu_Q D_\tau}{\mu_D - \lambda_d}\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right] - b\right) \cdot \frac{4\theta\mu_Q}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right], \end{aligned}$$

which is positive if and only if

$$\begin{aligned} &\frac{\theta D_\tau}{\mu_D - \lambda_d}\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right] \cdot 2\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right]\right) \\ &> \left(\frac{\theta \mu_Q D_\tau}{\mu_D - \lambda_d}\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right] - b\right) \cdot \frac{4\theta\mu_Q}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right] \\ \iff &\frac{\theta D_\tau\left[1 + \alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{\lambda_d - \mu_D} \cdot \frac{2\theta(\sigma_Q^2+\mu_Q^2)\left[1 - e^{-\lambda_d T}\right]}{\lambda_d} \\ &> \frac{\theta \mu_Q D_\tau\left[1 + \alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right] - b(\lambda_d - \mu_D)}{\lambda_d - \mu_D} \cdot \frac{4\theta\mu_Q\left[1 - e^{-\lambda_d T}\right]}{\lambda_d}. \end{aligned}$$

With the assumption that $\lambda_d > 2\mu_D + \sigma_D^2$, this is equivalent to

$$\begin{aligned} &\theta D_\tau\left[1 + \alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right] \cdot (\sigma_Q^2 + \mu_Q^2) \\ &> \left[\theta \mu_Q D_\tau\left[1 + \alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right] - b(\lambda_d - \mu_D)\right] \cdot 2\mu_Q \\ \iff &\theta D_\tau\left[1 + \alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\mu_Q^2 \\ &+ \theta D_\tau\left[1 + \alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\sigma_Q^2 \end{aligned}$$

$$
\begin{aligned}
&> 2\theta D_\tau \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\mu_Q^2 - 2b(\lambda_d-\mu_D)\mu_Q \\
\iff \quad &\theta D_\tau \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\mu_Q^2 - 2b(\lambda_d-\mu_D)\mu_Q \\
&- \theta D_\tau \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\sigma_Q^2 < 0,
\end{aligned}
$$

which is quadratic in μ_Q. The quadratic equation

$$
\begin{aligned}
&\theta D_\tau \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\mu_Q^2 - 2b(\lambda_d-\mu_D)\mu_Q \\
&- \theta D_\tau \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\sigma_Q^2 = 0
\end{aligned}
$$

has two roots, one of which is negative and the other is positive. Therefore, we have that $\frac{\partial K(D_\tau)}{\partial \mu_Q} > 0$ if and only if

$$
\begin{aligned}
&\mu_Q < \text{positive root} \\
&= \frac{b(\lambda_d-\mu_D) + \sqrt{b^2(\lambda_d-\mu_D)^2 + \theta^2 D_\tau^2 \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]^2}}{\theta D_\tau \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \\
&:= \bar{\mu}. \qquad (4.58)
\end{aligned}
$$

2. We take the derivative of $S(D_\tau)$ with respect to μ_Q. If the total demand satisfies $D_\tau \geq \frac{b(\mu_D-\lambda_d)}{\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}$, then, we have that

$$
\begin{aligned}
\frac{\partial S(D_\tau)}{\partial \mu_Q} &= \frac{\partial \frac{\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\partial \mu_Q} \\
&= \frac{num}{\left(4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]\right)^2},
\end{aligned}
$$

where

$$
\begin{aligned}
num = 2 &\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right] \\
&\cdot \frac{\theta D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right] \\
&\cdot \frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)\left[e^{-\lambda_d T}-1\right]}{-\lambda_d} \\
&-\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]^2 \cdot \frac{8\theta\mu_Q\left[e^{-\lambda_d T}-1\right]}{-\lambda_d},
\end{aligned}
$$

which is positive if and only if

$$
\begin{aligned}
&\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right] \\
&\cdot \frac{\theta D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]\cdot\left(\sigma_Q^2+\mu_Q^2\right) \\
&> \left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]^2\cdot\mu_Q. \qquad (4.59)
\end{aligned}
$$

We next discuss on the sign of $\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]$.

- If $\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b>0$, then (4.59) is equivalent to

$$
\begin{aligned}
&\frac{\theta D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]\cdot\left(\sigma_Q^2+\mu_Q^2\right) \\
&> \left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]\cdot\mu_Q \\
\iff\quad &\frac{\theta D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]\cdot\sigma_Q^2+b\mu_Q>0,
\end{aligned}
$$

  which always holds since both terms are positive.

- If $\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b=0$, then both sides of (4.59) are zero, and $num=0$.

- If $\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b<0$, then $S(D_\tau)=0$ according to (4.14).

  Therefore, we conclude that $num > 0$ and thus $\frac{\partial S(D_\tau)}{\partial \mu_Q} > 0$ if and only if $\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b>0$, which is

$$\mu_Q > \frac{b\left(\lambda_d-\mu_D\right)}{\theta D_\tau\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} := \hat{\mu}. \tag{4.60}$$

  Otherwise, $S(D_\tau)=0$ and thus $\frac{\partial S(D_\tau)}{\partial \mu_Q}=0$.

3. We take the derivative of $K(D_\tau)$ with respect to T.

$$\frac{\partial K(D_\tau)}{\partial T}=\frac{\partial\frac{\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b}{2\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\partial T}=\frac{num}{4\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]\right)^2},$$

  where

$$\begin{aligned} num=&\theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}\cdot 2\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]\\ &-\left(\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right)\cdot 2\theta(\sigma_Q^2+\mu_Q^2)e^{-\lambda_d T}. \end{aligned} \tag{4.61}$$

  While it is intractable to solve the equation (4.61) $=0$ for T, we take the limit on T. We first study the case when $T\to 0^+$.

- If $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] > b$, then

$$\lim_{T\to 0^+} \frac{\partial K(D_\tau)}{\partial T} = \frac{-\left(\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] - b\right) \cdot 2\theta(\sigma_Q^2 + \mu_Q^2)}{\lim_{T\to 0^+} 4\left(\frac{\theta(\sigma_Q^2 + \mu_Q^2)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]\right)^2} = -\infty. \tag{4.62}$$

- If $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] < b$, then

$$\lim_{T\to 0^+} \frac{\partial K(D_\tau)}{\partial T} = \frac{-\left(\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] - b\right) \cdot 2\theta(\sigma_Q^2 + \mu_Q^2)}{\lim_{T\to 0^+} 4\left(\frac{\theta(\sigma_Q^2 + \mu_Q^2)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]\right)^2} = +\infty. \tag{4.63}$$

- If $\frac{\theta\mu_Q D_\tau \alpha}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] = b$, then

$$\lim_{T\to 0^+} \frac{\partial K(D_\tau)}{\partial T} = \frac{num'}{\lim_{T\to 0^+} 4\frac{\theta(\sigma_Q^2 + \mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T} - 1\right] \cdot e^{-\lambda_d T}},$$

where

$$\begin{aligned} num' = \lim_{T\to 0^+} & \theta\mu_Q D_\tau(\mu_Q - \lambda_d)e^{(\mu_Q - \lambda_d)T} \cdot \frac{e^{-\lambda_d T} - 1}{-\lambda_d} \\ & + \theta\mu_Q D_\tau e^{(\mu_D - \lambda_d)T} \cdot e^{-\lambda_d T} - \theta\mu_Q D_\tau e^{(\mu_D - \lambda_d)T} e^{-\lambda_d T} \\ & + \left(\frac{\theta\mu_Q D_\tau}{\mu_D - \lambda_d}\left[e^{(\mu_D - \lambda_d)T} + \alpha e^{(\mu_D - \lambda_d)\hat{T}} - 1 - \alpha\right] - b\right)\lambda_d e^{-\lambda_d T}. \end{aligned}$$

Thus,

$$\lim_{T\to 0^+} \frac{\partial K(D_\tau)}{\partial T}$$

$$= \frac{\lim_{T\to 0^+} \frac{\partial num'}{\partial T}}{\lim_{T\to 0^+} 4\theta(\sigma_Q^2+\mu_Q^2)e^{-\lambda_d T}e^{-\lambda_d T} + 4\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right](-\lambda_d)e^{-\lambda_d T}}$$
$$= \frac{num''}{4\theta(\sigma_Q^2+\mu_Q^2)},$$

where

$$\begin{aligned} num'' = \lim_{T\to 0^+} & \theta\mu_Q D_\tau(\mu_Q-\lambda_d)e^{(\mu_Q-\lambda_d)T}\cdot e^{-\lambda_d T} + \theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}\lambda_d e^{-\lambda_d T} \\ & - \left(\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right)\lambda_d^2 e^{-\lambda_d T} \\ = & \theta\mu_Q^2 D_\tau. \end{aligned}$$

This leads to

$$\lim_{T\to 0^+}\frac{\partial K(D_\tau)}{\partial T} = \frac{\mu_Q^2 D_\tau}{4(\sigma_Q^2+\mu_Q^2)}. \tag{4.64}$$

Therefore, when $\frac{\theta\mu_Q D_\tau\alpha}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right] > b$, we have that $\lim_{T\to 0^+}\frac{\partial K(D_\tau)}{\partial T} < 0$; otherwise, $\lim_{T\to 0^+}\frac{\partial K(D_\tau)}{\partial T} > 0$.

4. We take the derivative of $S(D_\tau)$ with respect to T.

$$\begin{aligned} \frac{\partial S(D_\tau)}{\partial T} &= \frac{\partial\frac{\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\partial T} \\ &= \frac{num}{\left(4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]\right)^2}, \end{aligned}$$

where

$$num = 2\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]\theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}$$
$$\cdot 4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]$$
$$-\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]^2\cdot 4\theta(\sigma_Q^2+\mu_Q^2)e^{-\lambda_d T}. \tag{4.65}$$

While it is intractable to solve the equation (4.65) $= 0$ for T, we take the limit on T. We first study the case when $T \to 0^+$.

- If $\frac{\theta\mu_Q D_\tau\alpha}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right] \neq b$, then

$$\lim_{T\to 0^+}\frac{\partial S(D_\tau)}{\partial T} = \frac{-\left[\frac{\theta\mu_Q D_\tau\alpha}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right]-b\right]^2\cdot 4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lim_{T\to 0^+}\left(4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]\right)^2} = -\infty. \tag{4.66}$$

- If $\frac{\theta\mu_Q D_\tau\alpha}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right] = b$, then

$$\lim_{T\to 0^+}\frac{\partial S(D_\tau)}{\partial T} = \frac{num'}{\lim_{T\to 0^+}\left(8\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]\cdot e^{-\lambda_d T}\right)},$$

where

$$num' = \lim_{T\to 0^+}\left[2\theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}\theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}\right.$$
$$\left.+2\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]\right.$$

$$
\begin{aligned}
&\cdot \theta\mu_Q D_\tau(\mu_D-\lambda_d)e^{(\mu_D-\lambda_d)T}\Big] \cdot \frac{1}{-\lambda_d}\left[e^{-\lambda_d T}-1\right] \\
&+2\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right] \\
&\cdot \theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}\cdot e^{-\lambda_d T} \\
&-2\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right] \\
&\cdot \theta\mu_Q D_\tau e^{(\mu_D-\lambda_d)T}\cdot e^{-\lambda_d T} \\
&-\left[\frac{\theta\mu_Q D_\tau}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\right]^2 \\
&\cdot (-\lambda_d)e^{-\lambda_d T}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\lim_{T\to 0^+}\frac{\partial S(D_\tau)}{\partial T} \\
&=\frac{\lim_{T\to 0^+}\frac{\partial num'}{\partial T}}{\lim_{T\to 0^+} 8\theta\left(\sigma_Q^2+\mu_Q^2\right)e^{-\lambda_d T}e^{-\lambda_d T}+8\theta\left(\sigma_Q^2+\mu_Q^2\right)\left[e^{-\lambda_d T}-1\right]e^{-\lambda_d T}} \\
&=\frac{num''}{8\theta\left(\sigma_Q^2+\mu_Q^2\right)},
\end{aligned}
$$

where $num''=\lim_{T\to 0^+}\frac{\partial num'}{\partial T}=2\theta^2\mu_Q^2 D_\tau^2$. This leads to

$$
\lim_{T\to 0^+}\frac{\partial S(D_\tau)}{\partial T}=\frac{\theta\mu_Q^2 D_\tau^2}{4\left(\sigma_Q^2+\mu_Q^2\right)}. \tag{4.67}
$$

Therefore, when $\frac{\theta\mu_Q D_\tau\alpha}{\lambda_d-\mu_D}\left[1-e^{(\mu_D-\lambda_d)\hat{T}}\right]=b$, we have that $\lim_{T\to 0^+}\frac{\partial S(D_\tau)}{\partial T}>0$; otherwise, $\lim_{T\to 0^+}\frac{\partial S(D_\tau)}{\partial T}<0$.

5. From (4.13) and (4.14), we see that σ_Q only exists in the denominators of $K(D_\tau)$ and $S(D_\tau)$ as σ_Q^2. Therefore, we have that $\frac{\partial K(D_\tau)}{\partial\sigma_Q}<0$ and $\frac{\partial S(D_\tau)}{\partial\sigma_Q}<0$.

□

Proof of Lemma 4.2. We first derive the HJB equation (4.16). Recall the value function $V(x) = \max_{\tau \geq 0} \mathbb{E}\left[e^{-\lambda_d \tau} S(D_\tau) \mid D_0 = x\right]$, where D_t follows $dD_t = \mu_D D_t dt + \sigma_D D_t dW_t$. The value function also follows $V(x) = \mathbb{E}\left[e^{-\lambda_d h} V(D_h) \mid D_0 = x\right]$. By Ito's Lemma, we have

$$V(D_h) = V(D_0) + \int_0^h V'(D_0) dD_s + \int_0^h \frac{1}{2} V''(D_0)(dD_s)^2.$$

Then, we have that

$$\begin{aligned} V(D_h) &= V(x) + \int_0^h V'(x)\left[\mu_D x ds + \sigma_D x dW_s\right] + \int_0^h \frac{1}{2} V''(x)\left[\mu_D x ds + \sigma_D x dW_s\right]^2 \\ &= V(x) + V'(x)\mu_D x h + \int_0^h V'(x)\sigma_D x dW_s + \frac{1}{2} V''(x) \int_0^h \left[\mu_D x ds + \sigma_D x dW_s\right]^2 \\ &= V(x) + V(x)\mu_D x h + \frac{1}{2} V''(x) \int_0^h \sigma_D^2 x^2 ds + \int_0^h V'(x)\sigma_D x dW_s + \frac{1}{2} V''(x) o(h). \end{aligned}$$

Then,

$$\begin{aligned} &V(x) = \\ &\mathbb{E}\left[(1 - \lambda_d h + o(h))x \left\{ V(x) + V'(x)\mu_D x h + \frac{1}{2} V''(x)\sigma_D^2 x^2 h + \int_0^h V'(x)\sigma_D x dW_s + o(h) \right\}\right] \\ &= V(x) + V'(x)\mu_D x h + \frac{1}{2} V''(x)\sigma_D^2 x^2 h + o(h) - \lambda_d h V(x), \end{aligned}$$

which is

$$0 = -\lambda_d V(x) + V''(x)\mu_D x + \frac{1}{2} V''(x)\sigma_D^2 x^2 + \frac{o(h)}{h}.$$

Since $\lim_{h \to 0} \frac{o(h)}{h} = 0$, we have

$$0 = -\lambda_d V(x) + V'(x)\mu_D x + \frac{1}{2} V''(x)\sigma_D^2 x^2,$$

which is

$$\lambda_d V(x) = \mu_D x V'(x) + \frac{1}{2}\sigma_D^2 x^2 V''(x).$$

Therefore, in the continuation region, we have that $V(x) = \frac{1}{\lambda_d}\mu_D x V'(x) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V''(x)$. In the stopping region, we have that $V(x) = S(x)$. The value function takes the maximum value from stopping and continuing. Thus, it satisfies the HJB equation as given in (4.16). □

Proof of Proposition 4.3. It follows from the arguments before and after Lemma 4.2 that the firm would optimally start a PPA at the first time when $V(D_t) = S(D_t)$, since when $V(D_t) < S(D_t)$, the expected saving of the firm from waiting is higher, and when $V(D_t) = S(D_t)$, the expected saving from waiting is at most as much as the expected saving from starting a PPA. Since both $V(x)$ and $S(x)$ are continuously differentiable, there exists some threshold x_* such that $V(D_t) = S(D_t)$ when the realization of D_t first reaches x_*. Thus, it is optimal for the firm to sign the PPA at τ^* as given in (4.19).

To identify the demand threshold x_*, we solve the differential equation (4.17) along with the boundary conditions (4.18). We start by taking a guess. Let $V(x) = kx^\omega$. Then $xV'(x) = \omega k x^\omega$ and $x^2V''(x) = \omega(\omega-1)kx^\omega$. Substituting these into (4.17) leads to

$$\sigma_D^2\omega(\omega-1)kx^\omega + 2\mu_D\omega kx^\omega - 2\lambda_d kx^\omega = 0,$$

which is

$$kx^\omega\left[\sigma_D^2\omega^2 + (2\mu_D - \sigma_D^2)\omega - 2\lambda_d\right] = 0. \tag{4.68}$$

Note that k, μ_D, σ_D are constants and $x > 0$. The roots of (4.68) are

$$\omega_\pm = \frac{\sigma_D^2 - 2\mu_D \pm \sqrt{(2\mu_D - \sigma_D^2)^2 + 8\sigma_D^2\lambda_d}}{2\sigma_D^2}. \tag{4.69}$$

It follows that $V(x) = kx^{\omega_\pm}$ are solutions to the differential equation (4.17). This implies that the general solution is a linear combination of the respective solutions, i.e.,

$$V(x) = k_1 x^{\omega_+} + k_2 x^{\omega_-}. \tag{4.70}$$

From the properties of geometric Brownian motion, when $x = 0$, it will remain zero forever and the optimal threshold x_* will never be reached. Thus, we have that $V(0) = 0$. From (4.69), it is clear that $\omega_+ > 0$ and $\omega_- < 0$. Therefore, as $x \to 0$, we have $k_2 x^{\omega_-} \to +\infty$ for $k_2 > 0$ and $k_2 x^{\omega_-} \to -\infty$ for $k_2 < 0$. This implies that $V(0) = 0$ only when $k_2 = 0$. We thus have

$$V(x) = k_1 x^{\omega_+}. \tag{4.71}$$

Combining (4.71) with (4.18), we obtain that

$$k_1 x_*^{\omega_+} = S(x_*), \tag{4.72a}$$

$$\omega_+ k_1 x_*^{\omega_+ - 1} = S'(x_*). \tag{4.72b}$$

Substituting (4.72a) in (4.72b), we have that

$$\frac{\omega_+ S(x_*)}{x_*} = S'(x_*). \tag{4.73}$$

Combining (4.73) with (4.14), we have that

$$\omega_+ \frac{\left[\frac{\theta\mu_Q x_*}{\lambda_d - \mu_D}\left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}}\right] - b\right]^2}{4\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]}$$

$$
\begin{aligned}
=& x_* \frac{2\left[\frac{\theta\mu_Q x_*}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b\right]}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\cdot \frac{\theta\mu_Q}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right],
\end{aligned}
$$

which, after simplification, becomes

$$
\begin{aligned}
&\omega_+\left[\frac{\theta\mu_Q x_*}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b\right] \\
&= 2\frac{\theta\mu_Q x_*}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right].
\end{aligned}
$$

From this, we can solve for x_*, and obtain (4.20) as in the proposition, and ω_+ is given as (4.21). Finally, we show that $\omega_+ > 2$. Since μ_D and σ_D are both positive, we have the following:

$$
\begin{aligned}
& \omega_+ = \frac{\sigma_D^2 - 2\mu_D + \sqrt{(2\mu_D-\sigma_D^2)^2+8\sigma_D^2\lambda_d}}{2\sigma_D^2} > 2 \\
\iff \quad & \sigma_D^2 - 2\mu_D + \sqrt{(2\mu_D-\sigma_D^2)^2+8\sigma_D^2\lambda_d} > 4\sigma_D^2 \\
\iff \quad & \sqrt{(2\mu_D-\sigma_D^2)^2+8\sigma_D^2\lambda_d} > 2\mu_D + 3\sigma_D^2 \\
\iff \quad & (2\mu_D-\sigma_D^2)^2+8\sigma_D^2\lambda_d > (2\mu_D+3\sigma_D^2)^2 \\
\iff \quad & \sigma_D^4 - 4\mu_D\sigma_D^2 + 8\sigma_D^2\lambda_d > 9\sigma_D^4 + 12\mu_D\sigma_D^2 \\
\iff \quad & 8\sigma_D^2\lambda_d > 8\sigma_D^4 + 16\mu_D\sigma_D^2 \\
\iff \quad & \lambda_d > 2\mu_D + \sigma_D^2,
\end{aligned}
$$

where we recall that the last inequality is exactly our assumption on λ_d. □

Proof of Corollary 4.1. First, we check if the optimal capacity and the saving is positive with the given x_*, i.e., which region x_* lies in (4.13) and (4.14). With x_* as in (4.20), we

have that

$$
\begin{aligned}
&\frac{\theta\mu_Q x_*}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]-b\\
&=\frac{\theta\mu_Q}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]\\
&\quad\cdot\frac{b\omega_+\left(\mu_D-\lambda_d\right)}{(\omega_+-2)\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}-b\\
&=\frac{b\omega_+}{\omega_+-2}-b>0,
\end{aligned}
$$

which means that the optimal $K(x_*)$ and $S(x_*)$ are always positive.

Plugging (4.20) into (4.13), we obtain that

$$
\begin{aligned}
&K^*=K(x_*)\\
&=\frac{\frac{\theta\mu_Q}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]\frac{b\omega_+(\mu_D-\lambda_d)}{(\omega_+-2)\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}-b}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}\\
&=\frac{\frac{b\omega_+}{(\omega_+-2)}-b}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}=\frac{\frac{b}{\omega_+-2}}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}.
\end{aligned}
$$

Plugging (4.20) into (4.14), and obtain that

$$
\begin{aligned}
&S(x_*)=\\
&\frac{\left[\frac{\theta\mu_Q}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]\frac{b\omega_+(\mu_D-\lambda_d)}{(\omega_+-2)\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}-b\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}\\
&=\frac{\left[\frac{b\omega_+}{(\omega_+-2)}-b\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}=\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}.
\end{aligned}
$$

Next, we obtain $V(x)$ in the continuation region, i.e., the solution of (4.17). From the proof of Proposition 4.3, we have that $V(x) = k_1 x^{\omega_+}$, where k_1 can be obtained through (4.72a):

$$k_1 = \frac{S(x_*)}{x_*^{\omega_+}} = \frac{\frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\left(\frac{b\omega_+(\mu_D-\lambda_d)}{(\omega_+-2)\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}\right)^{\omega_+}}.$$

Thus, we have that

$$V(x) = k_1 x^{\omega_+} = \frac{\frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\left(\frac{b\omega_+(\mu_D-\lambda_d)}{(\omega_+-2)\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}\right)^{\omega_+}} \cdot x^{\omega_+}.$$

Finally, we look at the distribution of the firm's waiting time before signing a PPA. Since D_t is a geometric Brownian motion, i.e., $dD_t = \mu_D D_t dt + \sigma_D D_t dW_t$, we have that $D_t = D_0 \exp\left(\left(\mu_D - \frac{\sigma_D^2}{2}\right)t + dW_t\right)$, which implies that

$$\ln\left(\frac{D_t}{D_0}\right) = \left(\mu_D - \frac{\sigma_D^2}{2}\right)t + dW_t.$$

Thus, $\ln\left(\frac{D_t}{D_0}\right)$ is a Brownian motion with drift term $\left(\mu_D - \frac{\sigma_D^2}{2}\right)$. The firm's waiting time would be the first passage time of $\ln\left(\frac{D_t}{D_0}\right)$ to a fixed level $\ln\left(\frac{x_*}{D_0}\right)$. It is well known that the first passage time of a Brownian motion with drift follows an inverse Gaussian distribution (Schrödinger [188], Smoluchowski [197], Folks and Chhikara [83]). In the current context, we have that $\tau^* \sim \text{IG}\left(\frac{\ln\left(\frac{x_*}{D_0}\right)}{\mu_D - \sigma_D^2/2}, \left(\frac{\ln\left(\frac{x_*}{D_0}\right)}{\sigma_D}\right)^2\right)$ and $\mathbb{E}[\tau^*] = \frac{\ln\left(\frac{x_*}{D_0}\right)}{\mu_D - \sigma_D^2/2}$. □

Proof of Proposition 4.4. First, both the numerator and the denominator of (4.22) are pos-

itive, and μ_Q, σ_Q, and T only exist in the denominators. It is clear that as μ_Q increases, the denominator of (4.22) increases, thus decreasing K^*. The same also holds for σ_Q and T, thus we have (4.25a).

We next look at the value function $V(x)$. We take the derivatives of (4.24) with respect to μ_Q and σ_Q. Let $k_1 := \frac{\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}}$. Then $V(x) = k_1 x^{\omega_+}$, and we have that

$$\frac{\partial V(x)}{\partial \mu_Q} = \frac{\partial k_1}{\partial \mu_Q} = \frac{num}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{2\omega_+}},$$

where

$$\begin{aligned}
num = & \frac{-\left[\frac{b}{\omega_+-2}\right]^2 \cdot 2\theta\mu_Q \frac{[1-e^{-\lambda_d T}]}{\lambda_d}}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
& \cdot \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q(1+\alpha)\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+} \\
& - \frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
& \cdot \omega_+ \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+-1} \\
& \cdot \left(\frac{-b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q^2\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-\left[\frac{b}{\omega_+ - 2}\right]^2 \cdot 2\theta\mu_Q \frac{[1-e^{-\lambda_d T}]}{\lambda_d}}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad \cdot \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q(1+\alpha)\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+} \\
&\quad - \frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot \omega_+ \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+-1} \\
&\quad \cdot \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right) \cdot \frac{-1}{\mu_Q} \\
&= \frac{-\left[\frac{b}{\omega_+ - 2}\right]^2 \cdot 2\theta\mu_Q \frac{[1-e^{-\lambda_d T}]}{\lambda_d}}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad \cdot \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q(1+\alpha)\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+} \\
&\quad - \frac{\left[\frac{b}{\omega_+ - 2}\right]^2 \cdot \frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad \cdot \omega_+ \left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+} \frac{-1}{\mu_Q}.
\end{aligned}
$$

Note that $num > 0$ if and only if $\frac{-2\mu_Q}{(\sigma_Q^2+\mu_Q^2)} > \omega_+ \cdot \left(-\frac{1}{\mu_Q}\right)$, which is $\omega_+(\sigma_Q^2+\mu_Q^2) > 2\mu_Q^2$, or $(\omega_+-2)\mu_Q^2 + \omega_+\sigma_Q^2 > 0$, which always holds. Therefore, we have that $\frac{\partial V(x)}{\partial \mu_Q} > 0$.

On the other side, we have that $\frac{\partial V(x)}{\partial \sigma_Q} = \frac{\partial k_1}{\partial \sigma_Q} < 0$ since we only have σ_Q^2 in the denominator of k_1.

Finally, we look at $\mathbb{E}[\tau^*]$. From Corollary 4.1, we have that $\mathbb{E}[\tau^*] = \frac{\ln\left(\frac{x_*}{D_0}\right)}{\mu_D - \sigma_D^2/2}$. With the

assumption that $\mu_D > \sigma_D^2/2$, we have that

$$\begin{aligned}
\frac{\partial \mathbb{E}[\tau^*]}{\partial \mu_Q} &= \frac{1}{\mu_D - \sigma_D^2/2} \cdot \frac{D_0}{x_*} \cdot \frac{\partial x^*}{\partial \mu_Q} = \frac{D_0}{\mu_D - \sigma_D^2/2} \cdot \left(-\frac{1}{\mu_Q}\right) < 0, \\
\frac{\partial \mathbb{E}[\tau^*]}{\partial \sigma_Q} &= \frac{1}{\mu_D - \sigma_D^2/2} \cdot \frac{D_0}{x_*} \cdot \frac{\partial x^*}{\partial \sigma_Q} = 0, \\
\frac{\partial \mathbb{E}[\tau^*]}{\partial T} &= \frac{1}{\mu_D - \sigma_D^2/2} \cdot \frac{D_0}{x_*} \cdot \frac{\partial x^*}{\partial T} \\
&= \frac{D_0}{\mu_D - \sigma_D^2/2} \cdot \frac{(\omega_+ - 2)\theta\mu_Q(1+\alpha)\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right]}{b\omega_+(\mu_D - \lambda_d)} \\
&\quad \cdot \frac{b\omega_+(\mu_D - \lambda_d) \cdot (\omega_+ - 2)\theta\mu_Q(1+\alpha)(\lambda_d - \mu_D)e^{(\mu_D-\lambda_d)T}}{(\omega_+ - 2)^2\theta^2\mu_Q^2(1+\alpha)^2\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right]^2} \\
&= \frac{D_0}{\mu_D - \sigma_D^2/2} \cdot \frac{(\lambda_d - \mu_D)e^{(\mu_D-\lambda_d)T}}{\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right]} < 0.
\end{aligned}$$

□

Proof of Proposition 4.5. When $D_0 < x_*$, the firm wait to sign the PPA, and the optimal expected capacity is given by (4.22), where it follows that a higher b will lead to a higher K^*, which in turn increases the total renewable energy output $K^* \int_0^{\hat{T}} Q_t dt$ with probability 1. On the other side when $D_0 \geq x_*$, the firm does not wait and signs the PPA immediately, in which case the optimal capacity for the firm is given by (4.13), where it follows that a higher b will lead to a lower $K(D_0)$, which in turn decreases the total renewable energy output $K(D_0) \int_0^{\hat{T}} Q_t dt$ with probability 1. □

Proof of Proposition 4.6. We first note that

$$\begin{aligned}
&\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} \\
&= \frac{\partial \mathbb{E}\left[K^* Q_t \hat{T}\right]}{\partial \mu_Q} = \frac{\partial K^* \mathbb{E}\left[Q_t\right] \hat{T}}{\partial \mu_Q} = \frac{\partial K^*}{\partial \mu_Q} \cdot \mu_Q \hat{T} + K^* \hat{T}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-\frac{b}{\omega_+ - 2} \cdot \frac{\theta}{\lambda_d}\left[1 - e^{-\lambda_d T}\right] \cdot 2\mu_Q^2}{\left(\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]\right)^2}\hat{T} + \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]}\hat{T} \\
&= \frac{-\frac{b}{\omega_+ - 2} \cdot \frac{\theta}{\lambda_d}\left[1 - e^{-\lambda_d T}\right] \cdot 2\mu_Q^2 + \frac{b}{\omega_+ - 2} \cdot \frac{\theta}{\lambda_d}\left[1 - e^{-\lambda_d T}\right] \cdot \left(\sigma_Q^2 + \mu_Q^2\right)}{\left(\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]\right)^2}\hat{T} \\
&= \frac{\frac{b}{\omega_+ - 2} \cdot \frac{\theta}{\lambda_d}\left[1 - e^{-\lambda_d T}\right] \cdot \left(\sigma_Q^2 - \mu_Q^2\right)}{\left(\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]\right)^2}\hat{T},
\end{aligned}
$$

which is positive if and only if $\mu_Q < \sigma_Q$.

On the other side, from (4.25a), it should be clear that $\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \sigma_Q} = \mu_Q \hat{T} \frac{\partial K^*}{\partial \sigma_Q} < 0$, and $\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial T} = \mu_Q \hat{T} \frac{\partial K^*}{\partial T} < 0$. □

4.7.2 Proofs for Section 4.4

Proof of Lemma 4.3. The proof follows directly from the proof of Lemma 4.1, with the difference being that the transfer C is replaced with $bKe^{-\lambda_c \tau}$ here, instead of bK in Lemma 4.1. □

Proof of Proposition 4.7. From Lemma 4.3, the firm's expected discounted saving (4.29) is a quadratic function of K. Moreover, note that the capacity can only be nonnegative. It follows that the optimal capacity is

$$
K(D_\tau, \tau) = \max\left\{ \frac{\frac{\theta \mu_Q D_\tau}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)T}\right] + \frac{\alpha \theta \mu_Q D_\tau}{\lambda_d - \mu_D}\left[1 - e^{(\mu_D - \lambda_d)\hat{T}}\right] - be^{-\lambda_c \tau}}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]}, 0 \right\}
$$

$$= \max\left\{ \frac{\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{-\lambda_c\tau}}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}, 0 \right\},$$

and the corresponding saving is

$$S(D_\tau,\tau) := \begin{cases} \dfrac{\left[\frac{\theta\mu_Q D_\tau}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{-\lambda_c\tau}\right]^2}{4\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}, \\ \qquad\qquad \text{if } D_\tau \geq \dfrac{be^{-\lambda_c\tau}(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}, \\ 0, \qquad\qquad \text{if } D_\tau < \dfrac{be^{-\lambda_c\tau}(\lambda_d-\mu_D)}{\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}. \end{cases}$$

□

Proof of Proposition 4.8. The proof of this proposition follows the proof of Proposition 4.2, with the only change being the cost of investment coefficient b is now replaced with $be^{-\lambda_c\tau}$. Here, we have that

$$\bar{\mu}^d = \frac{be^{-\lambda_c\tau}(\lambda_d-\mu_D)+\sqrt{b^2e^{-2\lambda_c\tau}(\lambda_d-\mu_D)^2+\theta^2D_\tau^2\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]^2}}{\theta D_\tau\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}, \tag{4.74}$$

$$\hat{\mu}^d = \frac{be^{-\lambda_c\tau}\left(\lambda_d-\mu_D\right)}{\theta D_\tau\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}. \tag{4.75}$$

The rest of the proof is omitted as they are similar to the proof of Proposition 4.2. □

Proof of Lemma 4.4. We first derive the HJB equation (4.33). At any time t, suppose the realization of the total demand is x. The firm can choose to start a PPA or continue to wait. If the firm optimally chooses to start the PPA, its saving is $S(x,t)$, and it must hold

that $V(x,t) = S(x,t)$; if the firm optimally chooses to wait, and suppose the firm waits for the next δt time and the change in total demand is δx, then, we have that

$$V(x,t) = e^{-\lambda_d \delta t}\mathbb{E}\left[V(x+\delta x, t+\delta t)\right] \tag{4.76}$$

Expanding (4.76) using Taylor series expansion, we obtain

$$\begin{aligned} V(x+\delta x, t+\delta t) \approx & V(x,t) + V_x(x,t)\delta x + V_t(x,t)\delta t + \frac{V_{xx}(x,t)}{2}(\delta x)^2 + \frac{V_{tt}(x,t)}{2}(\delta t)^2 \\ & + V_{xt}(x,t)\delta x \delta t + o\left((\delta x)^3\right) + o\left((\delta t)^3\right) + o\left((\delta x)^2 \delta t\right) + o\left((\delta t)^2 \delta x\right) \end{aligned} \tag{4.77}$$

Neglecting higher order terms, and substituting (4.77) in (4.76), we have that

$$\begin{aligned} V(x,t) &\approx e^{-\lambda_d \delta t}\mathbb{E}\Bigg[V(x,t) + V_x(x,t)\delta x + V_t(x,t)\delta t + \frac{V_{xx}(x,t)}{2}(\delta x)^2 \\ &\quad + \frac{V_{tt}(x,t)}{2}(\delta t)^2 + V_{xt}(x,t)\delta x \delta t\Bigg] \\ &= e^{-\lambda_d \delta t}\Bigg[V(x,t) + V_x(x,t)\mathbb{E}[\delta x] + V_t(x,t)\delta t + \frac{V_{xx}(x,t)}{2}\mathbb{E}(\delta x)^2 \\ &\quad + \frac{V_{tt}(x,t)}{2}(\delta t)^2 + V_{xt}(x,t)\delta t \mathbb{E}[\delta x]\Bigg]. \end{aligned} \tag{4.78}$$

Subtracting $e^{-\lambda_d \delta t}V(x,t)$ from both sides of (4.78), then dividing by δt, we obtain

$$\begin{aligned} &\frac{\left(1 - e^{-\lambda_d \delta t}\right)V(x,t)}{\delta t} \\ &= \frac{e^{-\lambda_d \delta t}}{\delta t}\left[V_x(x,t)\mathbb{E}[\delta x] + V_t(x,t)\delta t + \frac{V_{xx}(x,t)}{2}\mathbb{E}(\delta x)^2 + \frac{V_{tt}(x,t)}{2}(\delta t)^2 + V_{xt}(x,t)\delta t \mathbb{E}[\delta x]\right]. \end{aligned} \tag{4.79}$$

Given the geometric Brownian motion assumption of the demand process, we have that $\delta x = \mu_D x \delta t + \sigma_D x \delta W_t$, and thus $\mathbb{E}[\delta x] = \mu_D x \delta t$, $\mathbb{E}[(\delta x)^2] = (\sigma_D x)^2 \delta t$. Substituting these

expectations in (4.79), and take the limit $\delta t \to 0$, we arrive at

$$\lambda_d V(x,t) = \mu_D x V_x(x,t) + \frac{1}{2}\sigma_D^2 x^2 V_{xx}(x,t) + V_t(x,t). \tag{4.80}$$

Therefore, in the continuation region, we have $V(x,t) = \frac{1}{\lambda_d}\mu_D x V_x(x,t) + \frac{1}{2\lambda_d}\sigma_D^2 x^2 V_{xx}(x,t) + \frac{1}{\lambda_d}V_t(x,t)$. In the stopping region, we have that $V(x,t) = S(x,t)$. The value function takes the maximum value from stopping and continuing. Thus, the value function satisfies the HJB equation as given in (4.33). □

Proof of Proposition 4.9. It follows from the arguments before and after Lemma 4.4 that the firm would optimally start a PPA at the first time when $V(D_t,t) = S(D_t,t)$, since when $V(D_t,t) < S(D_t,t)$, the expected saving of the firm from waiting is higher, and when $V(D_t,t) = S(D_t,t)$, the expected saving from waiting is at most as much as the expected saving from starting a PPA. Since both $V(x,t)$ and $S(x,t)$ are continuously differentiable, there exists some threshold function $x_*(t)$ such that $V(D_t,t) = S(D_t,t)$ at the first time when the realization of D_t reaches $x_*(t)$. Thus, it is optimal for the firm to sign the PPA at τ^* as given in (4.36).

To identify the demand threshold function $x_*(t)$, we solve the differential equation (4.34) along with the boundary conditions (4.35). We start by taking a guess. Let $V(x,t) = kx^\omega e^{\hat{\lambda}t}$. Then $xV_x(x,t) = \omega k x^\omega e^{\hat{\lambda}t}$, $x^2 V_{xx}(x,t) = \omega(\omega-1)kx^\omega e^{\hat{\lambda}t}$, and $V_t(x,t) = \hat{\lambda}kx^\omega e^{\hat{\lambda}t}$. Substituting these into (4.34) leads to

$$\lambda_d k x^\omega e^{\hat{\lambda}t} = \mu_D \omega k x^\omega e^{\hat{\lambda}t} + \frac{1}{2}\sigma_D^2 \omega(\omega-1)kx^\omega e^{\hat{\lambda}t} + \hat{\lambda}kx^\omega e^{\hat{\lambda}t},$$

which is

$$kx^\omega e^{\hat{\lambda}t}\left[\sigma_D^2\omega(\omega-1) + 2\mu_D\omega + 2\left(\hat{\lambda} - \lambda_d\right)\right] = 0.$$

Since $kx^{\omega}e^{\hat{\lambda}t} > 0$ for any $x > 0$ (assuming $k > 0$), this implies that

$$\sigma_D^2\omega^2 + (2\mu_D - \sigma_D^2)\omega + 2\left(\hat{\lambda} - \lambda_d\right) = 0 \tag{4.81}$$

Note that μ_D, σ_D are constants. The roots of (4.81) are

$$\omega_{\pm} = \frac{\sigma_D^2 - 2\mu_D \pm \sqrt{(2\mu_D - \sigma_D^2)^2 - 8\sigma_D^2 2\left(\hat{\lambda} - \lambda_d\right)}}{2\sigma_D^2}. \tag{4.82}$$

It follows that $V(x) = kx^{\omega_{\pm}}e^{\hat{\lambda}t}$ are solutions to the differential equation (4.34). This implies that the general solution is a linear combination of the respective solutions, i.e.,

$$V(x,t) = \left(k_1 x^{\omega_+} + k_2 x^{\omega_-}\right) e^{\hat{\lambda}t}. \tag{4.83}$$

From the properties of geometric Brownian motion, when $x = 0$, it will remain zero forever and the optimal threshold $x_*(t)$ will never be reached. Thus, we have that $V(0,t) = 0$. For $\hat{\lambda}$, we restrict ourselves to the condition $0 < \hat{\lambda} < \lambda_d$, which will be verified after we derive the explicit expression for $\hat{\lambda}$. For now, with the condition that $0 < \hat{\lambda} < \lambda_d$, and also from (4.82), it is clear that $\omega_+ > 0$ and $\omega_- < 0$. Therefore, as $x \to 0$, we have $k_1 x^{\omega_+}e^{\hat{\lambda}t} \to 0$, $k_2 x^{\omega_-}e^{\hat{\lambda}t} \to +\infty$ for $k_2 > 0$ and $k_2 x^{\omega_-}e^{\hat{\lambda}t} \to -\infty$ for $k_2 < 0$. This implies that $V(0,t) = 0$ only when $k_2 = 0$. We thus have

$$V(x,t) = k_1 x^{\omega_+}e^{\hat{\lambda}t}. \tag{4.84}$$

Combining (4.84) with (4.35), we obtain that

$$k_1 x_*^{\omega_+}e^{\hat{\lambda}t} = S(x_*,t), \tag{4.85a}$$

$$\omega_+ k_1 x_*^{\omega_+ - 1}e^{\hat{\lambda}t} = S_x(x_*,t), \tag{4.85b}$$

$$\hat{\lambda} k_1 x_*^{\omega_+} e^{\hat{\lambda} t} = S_t(x_*, t). \tag{4.85c}$$

Substituting (4.85a) in (4.85b) and (4.85c), we have that

$$\frac{\omega_+ S(x_*, t)}{x_*} = S_x(x_*, t), \tag{4.86a}$$

$$\hat{\lambda} S(x_*, t) = S_t(x_*, t) \tag{4.86b}$$

Combining (4.86) with (4.31), and solve for x_*, we obtain that

$$x_*(t) = \frac{\omega_+ S(x_*, t)}{S_x(x_*, t)} = \frac{b\omega_+ e^{-\lambda_c t}\left(\lambda_d - \mu_D\right)}{(\omega_+ - 2)\theta\mu_Q\left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}}\right]}, \tag{4.87}$$

$$\hat{\lambda} = \frac{S_t(x_*, t)}{S(x_*, t)} = \frac{2b\lambda_c e^{-\lambda_c t}}{\left[\frac{\theta\mu_Q x_*(t)}{\lambda_d - \mu_D}\left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}}\right] - b e^{-\lambda_c t}\right]} = (\omega_+ - 2)\lambda_c, \tag{4.88}$$

where $x_*(t)$ is as (4.37) in the Proposition. It remains to solve for ω_+ and $\hat{\lambda}$ using (4.88) and (4.82). Plugging (4.82) in (4.88), we have that

$$\hat{\lambda} = \frac{(-3\sigma_D^2 - 2\mu_D) + \sqrt{(2\mu_D - \sigma_D^2)^2 - 8\sigma_D^2(\hat{\lambda} - \lambda_d)}}{2\sigma_D^2}\lambda_c.$$

With some algebra, we obtain that $\hat{\lambda}$ needs to satisfy

$$\sigma_D^4\hat{\lambda}^2 + \sigma_D^2(3\sigma_D^2 + 2\mu_D)\lambda_c\hat{\lambda} + 2\sigma_D^2\lambda_c^2\hat{\lambda} = \lambda_c^2(4\mu_D + 2\sigma_D^2)(-\sigma_D^2) + 2\sigma_D^2\lambda_c^2\lambda_d.$$

Solving the above for $\hat{\lambda}$, we have that

$$\hat{\lambda} = \frac{-\sigma_D^2(3\sigma_D^2 + 2\mu_D)\lambda_c - 2\sigma_D^2\lambda_c^2 + \sqrt{\Delta}}{2\sigma_D^4}, \tag{4.89}$$

where

$$\begin{aligned}\Delta &= \sigma_D^4\lambda_c^2(3\sigma_D^2 + 2\mu_D + 2\lambda_c)^2 - 4\sigma_D^4\left[\sigma_D^2\lambda_c^2(4\mu_D + 2\sigma_D^2) - 2\sigma_D^2\lambda_c^2\lambda_d\right] \\ &= \sigma_D^4\lambda_c^2\left[\sigma_D^4 + 4\mu_D^2 + 4\lambda_c^2 + 12\sigma_D^2\lambda_c + 8\mu_D\lambda_c - 4\sigma_D^2\mu_D + 8\sigma_D^2\lambda_d\right]. \qquad (4.90)\end{aligned}$$

Thus, from $\omega_+ - 2 = \frac{\hat{\lambda}}{\lambda_c}$, we have that

$$\omega_+ = \frac{\sigma_D^2 - 2\mu_D - 2\lambda_c + \sqrt{\sigma_D^4 + 4\mu_D^2 + 4\lambda_c^2 + 12\sigma_D^2\lambda_c + 8\mu_D\lambda_c - 4\sigma_D^2\mu_D + 8\sigma_D^2\lambda_d}}{2\sigma_D^2},$$

which is (4.38) in the proposition. Finally, we verify that $0 < \hat{\lambda} < \lambda_c$ (note that $\hat{\lambda} > 0$ also implies that $\omega_+ > 2$). To verify that $\hat{\lambda} > 0$, we need that

$$\Delta > \sigma_D^4(3\sigma_D^2 + 2\mu_D)^2\lambda_c^2 + 4\sigma_D^4\lambda_c^4 + 4\sigma_D^4\lambda_c^3(3\sigma_D^2 + 2\mu_D),$$

which is

$$\begin{aligned}&\sigma_D^4\lambda_c^2\left[\sigma_D^4 + 4\mu_D^2 + 4\lambda_c^2 + 12\sigma_D^2\lambda_c + 8\mu_D\lambda_c - 4\sigma_D^2\mu_D + 8\sigma_D^2\lambda_d\right] \\ &> \sigma_D^4(3\sigma_D^2 + 2\mu_D)^2\lambda_c^2 + 4\sigma_D^4\lambda_c^4 + 4\sigma_D^4\lambda_c^3(3\sigma_D^2 + 2\mu_D).\end{aligned}$$

After some simplification, this is equivalent to

$$\sigma_D^6\lambda_c^2\left[\lambda_d - \sigma_D^2 - 2\mu_D\right] > 0 \Longleftrightarrow \lambda_d > \sigma_D^2 + 2\mu_D.$$

Note that $\lambda_d > \sigma_D^2 + 2\mu_D$ is exactly our assumption on λ_d.

To verify that $\hat{\lambda} < \lambda_d$, it suffices to show that

$$\frac{-\sigma_D^2(3\sigma_D^2 + 2\mu_D)\lambda_c - 2\sigma_D^2\lambda_c^2 + \sqrt{\Delta}}{2\sigma_D^4} < \lambda_d,$$

which, after some algebra, becomes

$$\sigma_D^4\lambda_c^2\left[-8\sigma_D^4-16\sigma_D^2\mu_D\right]<4\sigma_D^8\lambda_d^2+4\sigma_D^4\lambda_d\sigma_D^2(3\sigma_D^2+2\mu_D)\lambda_c,$$

which always holds since the left-hand-side is negative and the right-hand-side is positive. □

Proof of Corollary 4.2. First, we check if the optimal capacity and the saving is positive with the given $x_*(t)$, i.e., which region $x_*(t)$ lies in (4.30) and (4.31). With $x_*(t)$ as in (4.37), we have that

$$\begin{aligned}
&\frac{\theta\mu_Q x_*(t)}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{-\lambda_c t}\\
&=\frac{\theta\mu_Q}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\\
&\quad\cdot\frac{b\omega_+e^{-\lambda_c t}\left(\lambda_d-\mu_D\right)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}-be^{-\lambda_c t}\\
&=\frac{be^{-\lambda_c t}\omega_+}{\omega_+-2}-be^{-\lambda_c t}>0,
\end{aligned}$$

which means that $K(x_*(t),t)$ and $S(x_*(t),t)$ are always positive.

Plugging (4.37) into (4.30), we obtain the optimal capacity as a function of the PPA starting time, i.e., the optimal capacity if the firm starts the PPA at time τ:

$$\begin{aligned}
K^*(\tau)=K(x_*(\tau))&=\frac{\frac{\theta\mu_Q x_*(\tau)}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-b}{2\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\\
&=\frac{\frac{b\omega_+e^{-\lambda_c\tau}}{\omega_+-2}-be^{-\lambda_c\tau}}{2\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)}=\frac{\frac{be^{-\lambda_c\tau}}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}.
\end{aligned}\tag{4.91}$$

Plugging (4.20) into (4.14), we obtain the firm's saving as a function of the PPA starting

time, i.e., the firm's saving if it starts the PPA at time τ:

$$S^*(\tau) = S(x_*(\tau)) = \frac{\left[\frac{b\omega_+ e^{-\lambda_c \tau}}{(\omega_+ - 2)} - be^{-\lambda_c \tau}\right]^2}{4\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)} = \frac{\left[\frac{be^{-\lambda_c \tau}}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}. \tag{4.92}$$

To obtain the expected optimal capacity and saving, we need to take the expectation of (4.91) and (4.92) over the firm's optimal stopping time (time to start the PPA). Therefore, we need to have the distribution of the firm's waiting time.

Since D_t is a geometric Brownian motion, i.e., $dD_t = \mu_D D_t dt + \sigma_D D_t dW_t$, we have that $D_t = D_0 \exp\left(\left(\mu_D - \frac{\sigma_D^2}{2}\right)t + dW_t\right)$. Let $D_* := \frac{b\omega_+(\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}$. Then, we can write that

$$\begin{aligned}
P[\tau^* > s] &= P\left[\forall t \in [0,s], D_t < x_*(t)\right] \\
&= P\left[\forall t \in [0,s], D_0 \exp\left(\left(\mu_D - \frac{\sigma_D^2}{2}\right)t + \sigma_D W_t\right) < D_* \exp(-\lambda_c t)\right] \\
&= P\left[\forall t \in [0,s], \ln(D_0) + \left(\left(\mu_D - \frac{\sigma_D^2}{2}\right)\nu + \sigma_D W_t\right) < \ln(D_*) + (-\lambda_c t)\right] \\
&= P\left[\max_{t\in[0,s]} \left(\lambda_c + \mu_D - \sigma_D^2/2\right)t + \sigma_D W_t < \ln\left(\frac{D_*}{D_0}\right)\right].
\end{aligned}$$

We need the expected first hitting time to $\ln\left(\frac{D_*}{D_0}\right)$ for the Brownian motion with drift $\left(\lambda_c + \mu_D - \sigma_D^2/2\right)t + \sigma_D W_t$. This first hitting time follows the inverse Gaussian distribution $\text{IG}\left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c + \mu_D - \sigma_D^2/2}, \left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\sigma_D}\right)^2\right)$, thus the expected first hitting time is given by

$$\mathbb{E}[\tau^*] = \frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c + \mu_D - \sigma_D^2/2}. \tag{4.93}$$

Now, we come back to the expected optimal capacity and saving. By taking the expec-

tation of (4.91) and (4.92) over τ^*, we have that

$$
\begin{aligned}
K^* &= \mathbb{E}_{\tau^*}[K^*(\tau^*)] = \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\mathbb{E}\left[e^{-\lambda_c \tau^*}\right] \\
&= \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \exp\left[\frac{\left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\sigma_D}\right)^2}{\frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c+\mu_D-\sigma_D^2/2}}\left(1-\sqrt{1-\frac{2\left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c+\mu_D-\sigma_D^2/2}\right)^2(-\lambda_c)}{\left(\frac{\ln\left(\frac{D_*}{D_0}\right)}{\sigma_D}\right)^2}}\right)\right] \\
&= \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot \exp\left[\frac{(\lambda_c+\mu_D-\sigma_D^2/2)\ln\left(\frac{D_*}{D_0}\right)}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)\right] \\
&= \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)},
\end{aligned}
$$

$$
\begin{aligned}
S^* &= \mathbb{E}_{\tau^*}\left[S(x_*(\tau^*),\tau^*)\right] \\
&= \frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot \exp\left[\frac{(\lambda_c+\mu_D-\sigma_D^2/2)\ln\left(\frac{D_*}{D_0}\right)}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-2\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)\right] \\
&= \frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}.
\end{aligned}
$$

Finally, we obtain $V(x,t)$ in the continuation region, i.e., the solution of (4.34). From the proof of Proposition 4.9, we have that $V(x,t) = k_1 x^{\omega_+} e^{\hat{\lambda} t}$, where k_1 can be obtained

through (4.85a):

$$k_1 = \frac{S(x_*,t)}{x_*^{\omega_+}e^{\hat{\lambda}t}} = \frac{\frac{\left[\frac{be^{-\lambda_c t}}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}}{\left(\frac{b\omega_+e^{-\lambda_c t}(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}e^{\hat{\lambda}t}}$$

$$= \frac{\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}}\cdot e^{-2\lambda_c t}\cdot e^{\omega_+\lambda_c t}\cdot e^{-\hat{\lambda}t}$$

$$= \frac{\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}},$$

where we recall from (4.88) that $\hat{\lambda} = (\omega_+ - 2)\lambda_c$. Thus, we have that

$$V(x,t) = k_1 x^{\omega_+}e^{\hat{\lambda}t} = \frac{\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}}{\left(\frac{b\omega_+(\lambda_d-\mu_D)}{(\omega_+-2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\omega_+}}\cdot x^{\omega_+}e^{\hat{\lambda}t},$$

where $\hat{\lambda}$ is derived from (4.89) and (4.90) as

$$\hat{\lambda} = \frac{-\sigma_D^2(3\sigma_D^2+2\mu_D)\lambda_c - 2\sigma_D^2\lambda_c^2}{2\sigma_D^4} + \frac{+\sqrt{\sigma_D^4\lambda_c^2\left[\sigma_D^4+4\mu_D^2+4\lambda_c^2+12\sigma_D^2\lambda_c+8\mu_D\lambda_c-4\sigma_D^2\mu_D+8\sigma_D^2\lambda_d\right]}}{2\sigma_D^4}.$$

□

Proof of Proposition 4.10. We first look at how the newly added renewable capacity K^* changes with respect to μ_Q. Recall that

$$K^* = \frac{\frac{b}{\omega_+ - 2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}.$$

We take partial derivative with respect to μ_Q:

$$\begin{aligned}
\frac{\partial K^*}{\partial \mu_Q} &= \frac{-\frac{b}{\omega_+-2}\cdot\frac{2\theta\mu_Q}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)} \\
&\quad + \frac{\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right) \\
&\quad \cdot \frac{1}{D_0}\cdot\left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)-1} \\
&\quad \cdot \frac{-b\omega_+\left(\mu_D-\lambda_d\right)}{(\omega_+-2)\theta\mu_Q^2\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]} \\
&= \frac{-\frac{b}{\omega_+-2}\cdot\frac{2\theta\mu_Q}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)} \\
&\quad + \frac{\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right) \\
&\quad \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)} \cdot \frac{1}{-\mu_Q},
\end{aligned} \tag{4.94}$$

where we note that (4.94) is positive if and only if

$$\frac{\frac{b}{\omega_+-2}\cdot\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)$$
$$\cdot\frac{1}{-\mu_Q}+\frac{-\frac{b}{\omega_+-2}\cdot\frac{2\theta\mu_Q}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}>0,$$

or equivalently,

$$-2\mu_Q^2-(\sigma_Q^2+\mu_Q^2)\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)>0. \tag{4.95}$$

After rearranging, (4.95) becomes

$$\begin{aligned}&-\mu_Q^2\cdot\left[\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2\right]\\ &>\sigma_Q^2\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)\end{aligned} \tag{4.96}$$

From (4.96), we then check if $\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2$ is always positive:

$$\begin{aligned}&\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2>0\\ \iff\quad&\left(\lambda_c+\mu_D-\sigma_D^2/2\right)\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2\sigma_D^2>0\\ \iff\quad&\left(\lambda_c+\mu_D+3\sigma_D^2/2\right)>\left(\lambda_c+\mu_D-\sigma_D^2/2\right)\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\end{aligned}$$

$$
\begin{aligned}
&\iff \quad \left(\lambda_c+\mu_D+3\sigma_D^2/2\right)^2 > \left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2\left(1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}\right)\\
&\iff \quad \left(\lambda_c+\mu_D+3\sigma_D^2/2\right)^2 > \left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2+2\lambda_c\sigma_D^2\\
&\iff \quad 9\sigma_D^4/4+3\lambda_c\sigma_D^2+3\mu_D\sigma_D^2 > \sigma_D^4/4-\lambda_c\sigma_D^2-\mu_D\sigma_D^2+2\lambda_c\sigma_D^2\\
&\iff \quad 2\sigma_D^2+2\lambda_c\sigma_D^2+4\mu_D\sigma_D^2>0\\
&\iff \quad 1+\lambda_c+2\mu_D>0,
\end{aligned}
$$

which always holds. This implies that $\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2$ is always positive. Thus, $\frac{\partial K^*}{\partial \mu_Q}>0$ if and only if

$$
-\mu_Q^2/\sigma_Q^2 > \frac{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2},
$$

equivalently,

$$
\frac{\sigma_Q^2}{\mu_Q^2} > \frac{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+2}{-\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)} := \xi.
$$

Next, we look at how the newly added renewable capacity K^* changes with respect to σ_Q. Note that σ_Q exists only in the denominator of K^*. We thus have that $\frac{\partial K^*}{\partial \sigma_Q}<0$.

We next look at the value function $V(x,t)$. We take the derivatives of (4.41) with respect to μ_Q and σ_Q. Let $k_1 := \frac{\frac{\left[\frac{b}{\omega_+-2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{-\lambda_d}\left[e^{-\lambda_d T}-1\right]}}{\left(\frac{b\omega_+(\mu_D-\lambda_d)}{(\omega_+-2)\theta\mu_Q\left[e^{(\mu_D-\lambda_d)T}+\alpha e^{(\mu_D-\lambda_d)\hat{T}}-1-\alpha\right]}\right)^{\omega_+}}$. Then $V(x,t)=$

$k_1 x^{\omega_+} e^{\hat{\lambda} t}$, and we have that

$$\frac{\partial V(x,t)}{\partial \mu_Q} = \frac{\partial k_1}{\partial \mu_Q} = \frac{num}{\left(\frac{b\omega_+(\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q \left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{2\omega_+}},$$

where

$$\begin{aligned} num = & \frac{-\left[\frac{b}{\omega_+ - 2}\right]^2 \cdot 2\theta\mu_Q \frac{[1-e^{-\lambda_d T}]}{\lambda_d}}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\ & \cdot \left(\frac{b\omega_+ (\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q(1+\alpha)\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\omega_+} \\ & - \frac{\left[\frac{b}{\omega_+ - 2}\right]^2}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\ & \cdot \omega_+ \left(\frac{b\omega_+ (\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\omega_+ - 1} \\ & \cdot \left(\frac{-b\omega_+ (\lambda_d - \mu_D)}{(\omega_+ - 2)\theta\mu_Q^2\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right). \end{aligned}$$

Note that $num > 0$ if and only if $\frac{-2\mu_Q}{(\sigma_Q^2+\mu_Q^2)} > \omega_+ \cdot \left(-\frac{1}{\mu_Q}\right)$, which is $\omega_+(\sigma_Q^2+\mu_Q^2) > 2\mu_Q^2$, or $(\omega_+ - 2)\mu_Q^2 + \omega_+\sigma_Q^2 > 0$, which always holds. Therefore, we have that $\frac{\partial V(x)}{\partial \mu_Q} > 0$.

Finally, we look at $\mathbb{E}[\tau^*]$. From Corollary 4.2, we have that $\mathbb{E}[\tau^*] = \frac{\ln\left(\frac{D_*}{D_0}\right)}{\lambda_c+\mu_D-\sigma_D^2/2}$. With the assumption that $\mu_D > \sigma_D^2/2$, we have that

$$\begin{aligned} \frac{\partial \mathbb{E}[\tau^*]}{\partial \mu_Q} &= \frac{1}{\lambda_c+\mu_D-\sigma_D^2/2} \cdot \frac{D_0}{D_*} \cdot \frac{\partial D^*}{\partial \mu_Q} = \frac{D_0}{\lambda_c+\mu_D-\sigma_D^2/2} \cdot \left(-\frac{1}{\mu_Q}\right) < 0, \\ \frac{\partial \mathbb{E}[\tau^*]}{\partial \sigma_Q} &= \frac{1}{\lambda_c+\mu_D-\sigma_D^2/2} \cdot \frac{D_0}{D_*} \cdot \frac{\partial D^*}{\partial \sigma_Q} = 0, \end{aligned}$$

$$
\begin{aligned}
\frac{\partial \mathbb{E}[\tau^*]}{\partial T} &= \frac{1}{\lambda_c + \mu_D - \sigma_D^2/2} \cdot \frac{D_0}{D_*} \cdot \frac{\partial D^*}{\partial T} \\
&= \frac{D_0}{\lambda_c + \mu_D - \sigma_D^2/2} \cdot \frac{(\omega_+ - 2)\theta\mu_Q(1+\alpha)\left[1+\alpha - e^{(\mu_D-\lambda_d)T} - \alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{b\omega_+(\lambda_d - \mu_D)} \\
&\quad \cdot \frac{b\omega_+(\mu_D - \lambda_d)\cdot(\omega_+ - 2)\theta\mu_Q(1+\alpha)(\lambda_d - \mu_D)e^{(\mu_D-\lambda_d)T}}{(\omega_+ - 2)^2\theta^2\mu_Q^2(1+\alpha)^2\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right]^2} \\
&= \frac{D_0}{\lambda_c + \mu_D - \sigma_D^2/2} \cdot \frac{(\lambda_d - \mu_D)e^{(\mu_D-\lambda_d)T}}{\left[e^{(\mu_D-\lambda_d)T} + \alpha e^{(\mu_D-\lambda_d)\hat{T}} - 1 - \alpha\right]} < 0.
\end{aligned}
$$

□

Proof of Proposition 4.11. When $D_t < x_*(t)$, the firm waits to sign the PPA, and the optimal expected capacity is given by (4.39), where it follows that a higher b will lead to a higher K^*, which in turn increases the total renewable energy output $K^* \int_0^{\hat{T}} Q_t dt$ with probability 1. On the other side when $D_t \geq x_*(t)$, the firm does not wait and signs the PPA immediately, in which case the optimal capacity for the firm is given by (4.30), where it follows that a higher b will lead to a lower $K(D_t, t)$, which in turn decreases the total renewable energy output $K(D_t, t) \int_0^{\hat{T}} Q_t dt$ with probability 1. □

Proof of Proposition 4.12. We first show $\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q}$.

$$
\begin{aligned}
&\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} \\
&= \frac{\partial \mathbb{E}\left[K^* Q_t \hat{T}\right]}{\partial \mu_Q} = \frac{\partial K^*}{\partial \mu_Q} \cdot \mu_Q \hat{T} + K^* \hat{T} \\
&= \frac{-\frac{b}{\omega_+ - 2} \cdot \frac{2\theta\mu_Q^2}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2 + \mu_Q^2)}{\lambda_d}\left[1 - e^{-\lambda_d T}\right]\right)^2} \cdot \left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c + \mu_D - \sigma_D^2/2}{\sigma_D^2}\left(1 - \sqrt{1 - \frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c + \mu_D - \sigma_D^2/2\right)^2}}\right)} \cdot \hat{T}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{-\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)\\
&\quad\cdot\left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}\cdot\hat{T}\\
&+\frac{\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\cdot\left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}\cdot\hat{T}\\
&=terms\cdot\left(\frac{D_*}{D_0}\right)^{\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}\cdot\hat{T},
\end{aligned}
$$

where

$$
\begin{aligned}
terms=&\frac{-\frac{b}{\omega_+-2}\cdot\frac{2\theta\mu_Q^2}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}\\
&+\frac{-\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)\\
&+\frac{\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}.
\end{aligned}
$$

Then, $\frac{\partial\mathbb{E}\left[K^*\int_0^{\hat{T}}Q_t dt\right]}{\partial\mu_Q}$ is positive if and only if $terms>0$, which is

$$
\begin{aligned}
&\frac{-\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)\\
&+\frac{-\frac{b}{\omega_+-2}\cdot\frac{2\theta\mu_Q^2}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}+\frac{\frac{b}{\omega_+-2}}{\frac{\theta(\sigma_Q^2+\mu_Q^2)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}>0,
\end{aligned}
$$

or equivalently,

$$-\mu_Q^2-(\sigma_Q^2+\mu_Q^2)\cdot\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+\sigma_Q^2>0,$$

which is

$$\begin{aligned}&-\mu_Q^2\left[\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)+1\right]\\&>\sigma_Q^2\left[\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)-1\right],\end{aligned}$$

which, after rearranging, becomes

$$\frac{\sigma_Q^2}{\mu_Q^2}>\frac{1+\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}{1-\frac{\lambda_c+\mu_D-\sigma_D^2/2}{\sigma_D^2}\left(1-\sqrt{1-\frac{2\sigma_D^2(-\lambda_c)}{\left(\lambda_c+\mu_D-\sigma_D^2/2\right)^2}}\right)}:=\zeta.$$

Finally, we note that $\frac{\partial\mathbb{E}\left[K^*\int_0^{\hat{T}}Q_t dt\right]}{\partial\sigma_Q}=\mu_Q\hat{T}\frac{\partial K^*}{\partial\sigma_Q}<0.$ □

4.7.3 Proofs for Section 4.5

Proof of Lemma 4.5. The firm's random saving is given by (4.48). We further note that $p_s^Y=\theta(D_s-KQ_s)$, $p_s^N=\theta D_s$, and $C=e^{-\lambda_c t_s}bK$. Thus, the firm's random saving can be written as

$$\begin{aligned}(4.48)=&\int_{t_s}^{t_s+T}e^{-\lambda_d s}p_s^N U_s ds-\int_{t_s}^{t_s+T}e^{-\lambda_d s}p_s^Y\left[U_s-\hat{Q}_s\right]ds\\&+\int_{t_s+T}^{t_s+\hat{T}}e^{-\lambda_d s}p_s^N U_s ds-\int_{t_s+T}^{t_s+\hat{T}}e^{-\lambda_d s}p_s^Y U_s ds-e^{-\lambda_d t_s}C\end{aligned}$$

$$
\begin{aligned}
&= \theta \int_{t_s}^{t_s+T} e^{-\lambda_d s} D_s U_s ds - \theta \int_{t_s}^{t_s+T} e^{-\lambda_d s} (D_s - KQ_s) \left[U_s - KQ_s\right] ds \\
&\quad + \theta \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} D_s U_s ds - \theta \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} (D_s - KQ_s) U_s ds - e^{-\lambda_d t_s} b e^{-\lambda_c t_s} K \\
&= \theta \int_{t_s}^{t_s+T} e^{-\lambda_d s} D_s U_s ds - \theta \int_{t_s}^{t_s+T} e^{-\lambda_d s} \left[D_s U_s - D_s K Q_s - K Q_s U_s + K^2 Q_s^2\right] ds \\
&\quad + \theta \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} D_s U_s ds - \theta \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} (D_s - KQ_s) U_s ds - e^{-\lambda_d t_s} b e^{-\lambda_c t_s} K \\
&= \theta \int_{t_s}^{t_s+T} e^{-\lambda_d s} \left[KQ_s(U_s + D_s) - K^2 Q_s^2\right] ds + \theta \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} K Q_s U_s ds \\
&\quad - b e^{-\lambda_d t_s} e^{-\lambda_c t_s} K \\
&= \theta(1+\alpha) K \int_{t_s}^{t_s+T} e^{-\lambda_d s} \left[Q_s D_s\right] ds - \theta K^2 \int_{t_s}^{t_s+T} e^{-\lambda_d s} Q_s^2 ds \\
&\quad + \theta \alpha K \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} Q_s D_s ds - b e^{-(\lambda_d+\lambda_c) t_s} K.
\end{aligned}
$$

From our assumptions on Q_t and D_t, we have that $\mathbb{E}[Q_t] = \mu_Q$, $\mathbb{E}[D_t] = D_0 e^{\mu_D t}$, and $\mathbb{E}[Q_t^2] = \text{Var}[Q_t] + \mathbb{E}[Q_t]^2 = \sigma_Q^2 + \mu_Q^2$. Thus, the firm's expected saving is

$$
\begin{aligned}
\mathbb{E}[(4.48)] &= \theta(1+\alpha)\mu_Q K \int_{t_s}^{t_s+T} e^{-\lambda_d s} D_0 e^{\mu_D s} ds - \theta K^2 \int_{t_s}^{t_s+T} e^{-\lambda_d s} \left(\sigma_Q^2 + \mu_Q^2\right) ds \\
&\quad + \theta \alpha \mu_Q K \int_{t_s+T}^{t_s+\hat{T}} e^{-\lambda_d s} D_0 e^{\mu_D s} ds - b e^{-(\lambda_d+\lambda_c) t_s} K \\
&= \frac{\theta(1+\alpha)\mu_Q D_0 K}{\mu_D - \lambda_d} \left[e^{(\mu_D - \lambda_d)(t_s+T)} - e^{(\mu_D - \lambda_d) t_s}\right] \\
&\quad - \frac{\theta K^2 \left(\sigma_Q^2 + \mu_Q^2\right)}{-\lambda_d} \left[e^{-\lambda_d (t_s+T)} - e^{-\lambda_d t_s}\right] \\
&\quad + \frac{\theta \alpha \mu_Q D_0 K}{\mu_D - \lambda_d} \left[e^{(\mu_D - \lambda_d)(t_s+\hat{T})} - e^{(\mu_D - \lambda_d)(t_s+T)}\right] - b e^{-(\lambda_d+\lambda_c) t_s} K \\
&= -\frac{\theta K^2 \left(\sigma_Q^2 + \mu_Q^2\right)}{-\lambda_d} \left[e^{-\lambda_d (t_s+T)} - e^{-\lambda_d t_s}\right] \\
&\quad + \frac{\theta \mu_Q D_0 K}{\mu_D - \lambda_d} \left[e^{(\mu_D - \lambda_d)(t_s+T)} - e^{(\mu_D - \lambda_d) t_s}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{\theta\alpha\mu_Q D_0 K}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+\hat{T})}-e^{(\mu_D-\lambda_d)t_s}\right]-be^{-(\lambda_d+\lambda_c)t_s}K\\
&=\frac{\theta\mu_Q D_0 K}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+T)}+\alpha e^{(\mu_D-\lambda_d)(t_s+\hat{T})}-(1+\alpha)e^{(\mu_D-\lambda_d)t_s}\right]\\
&\quad+\frac{\theta K^2\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[e^{-\lambda_d(t_s+T)}-e^{-\lambda_d t_s}\right]-be^{-(\lambda_d+\lambda_c)t_s}K.
\end{aligned}
$$

This completes the proof of Lemma 4.5. □

Proof of Proposition 4.13. From Lemma 4.5, the firm's expected discounted saving (4.50) is a quadratic function of K. Moreover, note that the capacity can only be nonnegative. It follows that the optimal capacity is

$$
\begin{aligned}
K(t_s)&=\frac{\frac{\theta\mu_Q D_0}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+T)}-e^{(\mu_D-\lambda_D)t_s}\right]+\frac{\theta\alpha\mu_Q D_0}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+\hat{T})}-e^{(\mu_D-\lambda_d)t_s}\right]}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d(t_s+T)}-e^{-\lambda_d t_s}\right]}\\
&\quad-\frac{be^{-(\lambda_d+\lambda_c)t_s}}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d(t_s+T)}-e^{-\lambda_d t_s}\right]}\\
&=\frac{\frac{\theta\mu_Q D_0}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+T)}+\alpha e^{(\mu_D-\lambda_d)(t_s+\hat{T})}-(1+\alpha)e^{(\mu_D-\lambda_d)t_s}\right]-be^{-(\lambda_d+\lambda_c)t_s}}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d(t_s+T)}-e^{-\lambda_d t_s}\right]}\\
&=\frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]e^{\mu_D t_s}}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]},
\end{aligned}
$$

and the corresponding saving (if $K(t_s)>0$) is

$$
\begin{aligned}
S(t_s)&=\frac{\left[\frac{\theta\mu_Q D_0}{\mu_D-\lambda_d}\left[e^{(\mu_D-\lambda_d)(t_s+T)}+\alpha e^{(\mu_D-\lambda_d)(t_s+\hat{T})}-(1+\alpha)e^{(\mu_D-\lambda_d)t_s}\right]-be^{-(\lambda_d+\lambda_c)t_s}\right]^2}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{-\lambda_d}\left[e^{-\lambda_d(t_s+T)}-e^{-\lambda_d t_s}\right]}\\
&=\frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]^2 e^{(2\mu_D-\lambda_d)t_s}}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}.
\end{aligned}
$$

□

Proof of Proposition 4.14. We first find t_s^*, which maximizes $S(t_s)$. Recall that

$$S(t_s) = \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]^2 e^{(2\mu_D-\lambda_d)t_s}}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}.$$

Then,

$$\begin{aligned}
&\frac{\partial S(t_s)}{\partial t_s} \\
&= \frac{2\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot b(\lambda_c+\mu_D)e^{(-\lambda_c-\mu_D)t_s}e^{(2\mu_D-\lambda_d)t_s} \\
&\quad + \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]^2}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot (2\mu_D-\lambda_d)e^{(2\mu_D-\lambda_d)t_s} \\
&= \frac{e^{(2\mu_D-\lambda_d)t_s}\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot \Bigg\{2b(\lambda_c+\mu_D)e^{(-\lambda_c-\mu_D)t_s} \\
&\quad - (\lambda_d-2\mu_D)\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right]\Bigg\}.
\end{aligned}$$

Since $K(t_s^*) > 0$, we have that $\frac{\partial S(t_s)}{\partial t_s} = 0$ if and only if t_s is positive and satisfies

$$\begin{aligned}
&2b(\lambda_c+\mu_D)e^{(-\lambda_c-\mu_D)t_s} \\
&= (\lambda_d-2\mu_D)\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s}\right],
\end{aligned}$$

which is

$$
\begin{aligned}
& b\left[2(\lambda_c+\mu_D)+(\lambda_d-2\mu_D)\right]e^{(-\lambda_c-\mu_D)t_s} \\
&= (\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right],
\end{aligned}
$$

or, equivalently,

$$
b\left(2\lambda_c+\lambda_d\right)e^{(-\lambda_c-\mu_D)t_s} = (\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right].
$$

Therefore, we have that

$$
t_s^* = \left[\frac{\log\left(b(2\lambda_c+\lambda_d)\right)-\log\left((\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\right)}{\lambda_c+\mu_D}\right]^+.
$$

When $t_s^* > 0$, we plug in t_s^* to (4.51), and obtain that

$$
\begin{aligned}
K^* = K(t_s^*) &= \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s^*}\right]e^{\mu_D t_s^*}}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&= \left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\right. \\
&\quad \left.-\frac{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{(2\lambda_c+\lambda_d)}\right] \\
&\quad \cdot \frac{1}{\frac{2\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad \cdot \left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}
\end{aligned}
$$

$$= \frac{\left[\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]}{\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}.$$

When $t_s^* > 0$, we plug in t_s^* to (4.51), and obtain that

$$S^* = S(t_s^*) = \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]-be^{(-\lambda_c-\mu_D)t_s^*}\right]^2 e^{(2\mu_D-\lambda_d)t_s^*}}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$= \left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\right.$$
$$\left.-\frac{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{(2\lambda_c+\lambda_d)}\right]^2$$
$$\cdot\frac{1}{\frac{4\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$\cdot\left(\frac{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{b(2\lambda_c+\lambda_d)}\right)^{\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}}$$
$$= \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]^2}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$\cdot\left(\frac{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{b(2\lambda_c+\lambda_d)}\right)^{\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}}.$$

□

Proof of Proposition 4.15. We first look at how the newly added renewable capacity K^*

changes with respect to μ_Q. Recall that when $t_s^* > 0$, we have that

$$K^* = \frac{\left[\frac{\mu_Q D_0}{\lambda_d - \mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]}{\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}.$$

We take the partial derivative with respect to μ_Q:

$$\frac{\partial K^*}{\partial \mu_Q} = \left\{\frac{\frac{D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}\right.$$
$$\left.-\frac{\left[\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]\cdot\frac{2\mu_Q}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}\right\}$$
$$\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}$$
$$+\frac{\left[\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]}{\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}\cdot\frac{-\mu_D}{\lambda_c+\mu_D}$$
$$\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}\cdot\mu_Q^{\frac{-\lambda_c-2\mu_D}{\lambda_c+\mu_D}}$$
$$=\frac{\frac{D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{\left(\sigma_Q^2-\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}$$

$$\cdot \left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \cdot \mu_Q^{\frac{-\mu_D}{\lambda_c+\mu_D}}$$

$$+ \frac{\left[\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]}{\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \cdot \frac{-\mu_D}{\lambda_c+\mu_D}$$

$$\cdot \left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \cdot \mu_Q^{\frac{-\lambda_c-2\mu_D}{\lambda_c+\mu_D}}$$

$$= \frac{\frac{D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{\left(\sigma_Q^2-\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}$$

$$\cdot \left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \cdot \mu_Q^{\frac{-\mu_D}{\lambda_c+\mu_D}}$$

$$+ \frac{\frac{D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}$$

$$\cdot \left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\frac{\mu_D}{\lambda_c+\mu_D}}$$

$$\cdot \mu_Q \cdot \frac{-\mu_D}{\lambda_c+\mu_D} \cdot \mu_Q^{\frac{-\lambda_c-2\mu_D}{\lambda_c+\mu_D}}$$

$$= \frac{\frac{D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{1}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}$$

$$\cdot \left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]} \right)^{\frac{\mu_D}{\lambda_c+\mu_D}}$$

$$\cdot \left((\sigma_Q^2-\mu_Q^2)\mu_Q^{\frac{-\mu_D}{\lambda_c+\mu_D}} - (\sigma_Q^2+\mu_Q^2)\frac{\mu_D}{\lambda_c+\mu_D}\mu_Q^{\frac{-\mu_D}{\lambda_c+\mu_D}} \right)$$

$$= \frac{\frac{D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{1}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}$$
$$\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}}$$
$$\cdot\left((\sigma_Q^2-\mu_Q^2)-(\sigma_Q^2+\mu_Q^2)\frac{\mu_D}{\lambda_c+\mu_D}\right),$$

which is nonnegative if and only if

$$\frac{\lambda_c}{\lambda_c+\mu_D}\sigma_Q^2 \geq \frac{\lambda_c+2\mu_D}{\lambda_c+\mu_D}\mu_Q^2,$$

or equivalently, $\frac{\sigma_Q^2}{\mu_Q^2} \geq \frac{\lambda_c+2\mu_D}{\lambda_c}$.

Also note that σ_Q exists only in the denominator of K^*, and b exists only in the numerator of K^*. We thus have that $\frac{\partial K^*}{\partial \sigma_Q} < 0$ and $\frac{\partial K^*}{\partial b} > 0$.

We then look at how the firm's optimal expected saving S^* changes with respect to μ_Q. Recall that when $t_s^* > 0$, we have that

$$S^* = \frac{\left[\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]^2}{\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}$$
$$\cdot\left(\frac{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{b(2\lambda_c+\lambda_d)}\right)^{\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}}.$$

We take the partial derivative of S^* with respect to μ_Q:

$$\frac{\partial S^*}{\partial \mu_Q}$$

$$
\begin{aligned}
&= \Bigg\{ \frac{2\mu_Q \left(\frac{\theta D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right] \cdot \frac{\lambda_c + \mu_D}{(2\lambda_c + \lambda_d)} \right)^2 \cdot \frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d}}{\left(\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d} \left[1 - e^{-\lambda_d T} \right] \right)^2} \\
&\quad - \frac{\left[\frac{\theta \mu_Q D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right] \cdot \frac{\lambda_c + \mu_D}{(2\lambda_c + \lambda_d)} \right]^2 \cdot \frac{2\theta\mu_Q}{\lambda_d}}{\left(\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d} \left[1 - e^{-\lambda_d T} \right] \right)^2} \Bigg\} \\
&\quad \cdot \left[1 - e^{-\lambda_d T} \right] \cdot \left(\frac{(\lambda_d - 2\mu_D) \frac{\theta \mu_Q D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right]}{b(2\lambda_c + \lambda_d)} \right)^{\frac{\lambda_d - 2\mu_D}{\lambda_c + \mu_D}} \\
&\quad + \frac{\left[\frac{\theta \mu_Q D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right] \cdot \frac{\lambda_c + \mu_D}{(2\lambda_c + \lambda_d)} \right]^2}{\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d} \left[1 - e^{-\lambda_d T} \right]} \cdot \frac{\lambda_d - 2\mu_D}{\lambda_c + \mu_D} \\
&\quad \cdot \left(\frac{(\lambda_d - 2\mu_D) \frac{\theta D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right]}{b(2\lambda_c + \lambda_d)} \right)^{\frac{\lambda_d - 2\mu_D}{\lambda_c + \mu_D}} \cdot \mu_Q^{\frac{\lambda_d - \lambda_c - 3\mu_D}{\lambda_c + \mu_D}} \\
&= \frac{\left[\frac{\theta D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right] \cdot \frac{\lambda_c + \mu_D}{(2\lambda_c + \lambda_d)} \right]^2 \cdot \frac{2\theta\mu_Q\sigma_Q^2}{\lambda_d} \left[1 - e^{-\lambda_d T} \right]}{\left(\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d} \left[1 - e^{-\lambda_d T} \right] \right)^2} \\
&\quad \cdot \left(\frac{(\lambda_d - 2\mu_D) \frac{\theta \mu_Q D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right]}{b(2\lambda_c + \lambda_d)} \right)^{\frac{\lambda_d - 2\mu_D}{\lambda_c + \mu_D}} \\
&\quad + \frac{\left[\frac{\theta D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right] \cdot \frac{\lambda_c + \mu_D}{(2\lambda_c + \lambda_d)} \right]^2 \cdot \frac{\theta\left(\mu_Q^2\sigma_Q^2 + \mu_Q^4\right)}{\lambda_d} \left[1 - e^{-\lambda_d T} \right]}{\left(\frac{\theta\left(\sigma_Q^2 + \mu_Q^2\right)}{\lambda_d} \right)^2} \\
&\quad \cdot \left[1 - e^{-\lambda_d T} \right] \cdot \left(\frac{(\lambda_d - 2\mu_D) \frac{\theta D_0}{\lambda_d - \mu_D} \left[1 + \alpha - e^{(\mu_D - \lambda_d)T} - \alpha e^{(\mu_D - \lambda_d)\hat{T}} \right]}{b(2\lambda_c + \lambda_d)} \right)^{\frac{\lambda_d - 2\mu_D}{\lambda_c + \mu_D}} \\
&\quad \cdot \frac{\lambda_d - 2\mu_D}{\lambda_c + \mu_D} \cdot \mu_Q^{\frac{\lambda_d - \lambda_c - 3\mu_D}{\lambda_c + \mu_D}}
\end{aligned}
$$

$$= \frac{\left[\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]^2}{\left(\frac{\theta\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2}\cdot\frac{\theta\left[1-e^{-\lambda_d T}\right]}{\lambda_d}$$

$$\cdot\left(\frac{(\lambda_d-2\mu_D)\frac{\theta D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}{b(2\lambda_c+\lambda_d)}\right)^{\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}}$$

$$\cdot\left(2\mu_Q\sigma_Q^2\cdot\mu_Q^{\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}}+(\mu_Q^2\sigma_Q^2+\mu_Q^4)\cdot\frac{\lambda_d-2\mu_D}{\lambda_c+\mu_D}\cdot\mu_Q^{\frac{\lambda_d-\lambda_c-3\mu_D}{\lambda_c+\mu_D}}\right)$$

$$\geq 0.$$

Also note that σ_Q and b exist only in the denominator of S^*. We thus have that $\frac{\partial S^*}{\partial \sigma_Q} < 0$ and $\frac{\partial S^*}{\partial b} < 0$.

We next look at the optimal scheduled starting time t_s^*. Recall that when $t_s^* > 0$, it is given by

$$t_s^* = \frac{\log\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)}{\lambda_c+\mu_D}.$$

Note that increasing either μ_Q or T would increase the denominator within the natural log in the numerator, which then decreases t_s^* (assuming the denominator within the natural log is still smaller than the numerator, so that t_s^* is still positive). Changing σ_Q does not affect t_s^*. Increasing b will increase t_s^* if $t_s^* > 0$. If $t_s^* = 0$, then t_s^* may stay at 0 or become positive with a higher b.

Finally, we look at how the expected additional production due to PPA changes with respect to μ_Q, σ_Q, and b. Note that

$$\frac{\partial \mathbb{E}\left[K^*\int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q} = \frac{\partial \mathbb{E}\left[K^* Q_t \hat{T}\right]}{\partial \mu_Q} = \hat{T}\frac{\partial K^*}{\partial \mu_Q}\cdot\mu_Q + \hat{T}K^*.$$

Thus,

$$
\begin{aligned}
&\frac{\frac{\partial \mathbb{E}\left[K^* \int_0^{\hat{T}} Q_t dt\right]}{\partial \mu_Q}}{\hat{T}} \\
&= \frac{\left[\frac{\mu_Q D_0}{\lambda_d - \mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]}{\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]} \\
&\quad\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \\
&\quad+\frac{\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{1}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \\
&\quad\cdot\left((\sigma_Q^2-\mu_Q^2)-(\sigma_Q^2+\mu_Q^2)\frac{\mu_D}{\lambda_c+\mu_D}\right) \\
&= \frac{\left[\frac{\mu_Q D_0}{\lambda_d - \mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\right]\cdot\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \\
&\quad+\frac{\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{1}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \\
&\quad\cdot\left((\sigma_Q^2-\mu_Q^2)-(\sigma_Q^2+\mu_Q^2)\frac{\mu_D}{\lambda_c+\mu_D}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{1}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \\
&\quad\cdot\left(\sigma_Q^2+\mu_Q^2+\sigma_Q^2-\mu_Q^2-\left(\sigma_Q^2+\mu_Q^2\right)\frac{\mu_D}{\lambda_c+\mu_D}\right) \\
&= \frac{\frac{\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]\cdot\frac{\lambda_c+\mu_D}{(2\lambda_c+\lambda_d)}\cdot\frac{1}{\lambda_d}\left[1-e^{-\lambda_d T}\right]}{\left(\frac{\left(\sigma_Q^2+\mu_Q^2\right)}{\lambda_d}\left[1-e^{-\lambda_d T}\right]\right)^2} \\
&\quad\cdot\left(\frac{b(2\lambda_c+\lambda_d)}{(\lambda_d-2\mu_D)\frac{\theta\mu_Q D_0}{\lambda_d-\mu_D}\left[1+\alpha-e^{(\mu_D-\lambda_d)T}-\alpha e^{(\mu_D-\lambda_d)\hat{T}}\right]}\right)^{\frac{\mu_D}{\lambda_c+\mu_D}} \\
&\quad\cdot\left(2\sigma_Q^2-\left(\sigma_Q^2+\mu_Q^2\right)\frac{\mu_D}{\lambda_c+\mu_D}\right),
\end{aligned}
$$

which is nonnegative if and only if

$$
2\sigma_Q^2 \geq \sigma_Q^2\frac{\mu_D}{\lambda_c+\mu_D}+\mu_Q^2\frac{\mu_D}{\lambda_c+\mu_D},
$$

or, equivalently,

$$
\sigma_Q^2\frac{2\lambda_c+\mu_D}{\mu_D} \geq \mu_Q^2.
$$

As for σ_Q and b, we have that $\frac{\partial\mathbb{E}\left[K^*Q_t\hat{T}\right]}{\partial\sigma_Q}<0$ and $\frac{\partial\mathbb{E}\left[K^*Q_t\hat{T}\right]}{\partial b}>0$, which follow directly from $\frac{\partial K^*}{\partial\sigma_Q}<0$ and $\frac{\partial K^*}{\partial b}>0$. □

CHAPTER 5

GREEDY ALGORITHMS FOR THE FREIGHT CONSOLIDATION PROBLEM

5.1 Introduction

The spiking high container prices since the COVID-19 pandemic have caused significant issues in global supply chains. In this chapter, we consider the (ocean) *freight consolidation problem* (FCP) - a combinatorial optimization problem that is being solved every day and every hour by some of the world's leading freight forwarders. In a nutshell, the freight consolidation problem aims to optimize the assignments of shipments to containers at the origin ports, such as Yantian Port (Shenzhen) and Port of Shanghai. In the FCP, there are a set of shipments and a set of candidate containers that can be used. The origin/destinations of each shipment and each container, as well as the estimated departure/arrival dates of each container, are predetermined as the shipment/container becomes available at the port. There are two major costs: cost of assigning a shipment to a container (shipment cost), and cost of procuring a container (container cost). We further explain these costs in slightly more detail:

- The shipment cost takes into account everything related to sending the shipment boxes to their final destinations. Starting from the origin port, the remaining cycle of a shipment includes arriving at a destination port, being sorted and loaded to rail or truck, and delivering to their destinations. If a shipment is assigned to two containers that arrive at different ports, the remaining rail and/or trucking costs will be different. Furthermore, many shipments also have time window requirements, and based on the arrival time of different containers, there may be different lateness costs. Therefore, we have a shipment cost associated with assigning each shipment to each container. If a container is not feasible for a shipment due to time window or destination ports, the

corresponding shipment cost (of assigning that shipment to that container) is assigned as ∞.

- The container cost is the cost of using a container. There is a set of containers available at the origin port, each with its own destination, departure time, and cost of procurement. If we decide to assign any shipment to a container, then we have to pay the procurement cost for that container.

Moreover, if we find there is no proper container to assign a shipment, there is always an option to "coload" that shipment, i.e., use a third-party shipper, e.g., Shipco, to fulfill that shipment. The cost associated with assigning the shipment to a third-party shipper is called the "coloading" cost. In our formulation, the "coloading" option can be viewed as a container with unlimited capacity, and the coloading costs are equivalently viewed as the shipment cost of assigning a shipment to this "coloading" container.

The freight forwarder aims to fulfill all shipments at hand while minimizing the total cost, which includes both shipment costs and container costs, subject to certain constraints. Specifically, each container has its own size in three-dimensions, as does each shipment. A container also has a maximum weight limit. In reality, we need to ensure that the total weight of all shipments assigned to a container does not exceed the weight limit of that container, and the center of mass (of a loading plan of these shipments) is not too far away from the center of the container. Moreover, these shipments should be able to fit into the container in three dimensions. Assuming a shipment is packed in a three-dimensional box, there are six possible rotations (orientations) of a box when being loaded to the container. Some boxes do not allow all six rotations, and some boxes are not stackable (which means they have to be put on the top). Given all these practical constraints, the problem of loading any given set of shipments to a container is a separate NP-hard problem, which is called the *container loading problem* in literature (see Aydemir and Yigit [23] for a comprehensive review). It would be too complicated to consider all container-loading constraints in our freight consolidation

model. Therefore, we simplify the constraints by just having a weight capacity constraint and a volume capacity constraint for each container, ignoring the actual three-dimensional packing feasibility constraint. Despite that FCP does not reflect all practical constraints, we believe it is the simplest model to capture the most important features of the problem.

Up till now, a keen reader would recognize that our FCP can be viewed as a combination of the generalized assignment problem (GAP) and the bin packing problem (BPP), in a more complicated version. The shipment costs mimic the costs of assigning jobs in GAP, while in FCP we have two sets of capacity constraints (both weight and volume). The container cost is the cost of using each container (bin), while we have different costs for each container (bin). Therefore, FCP is already complicated in its nature and is expected to be difficult to solve. In this chapter, we prove the non-approximability result of FCP, i.e., there is no constant factor approximation to FCP in polynomial time, unless $P = NP$. As a remedy, we propose a series of heuristics. With simulated data that aims to reflect the actual practice, we show that our heuristics return solutions with small optimality gaps.

The remaining of the chapter is organized as follows. In Section 5.2, we provide a comprehensive literature review on the Bin Packing and related problems. In Section 5.3 we formally introduce the FCP and provide the non-approximability result. In Section 5.4, we provide main greedy heuristics for solving the FCP. Section 5.5 provides some numerical experiments on these heuristics. The chapter concludes with Section 5.6.

5.2 Literature Review

5.2.1 Classical Bin Packing Problem

We first review the classical (one-dimensional) bin packing problem (BPP). In the classical bin packing problem, we are given a set of items, each with a one-dimensional size, and an unlimited number of containers (bins) with the same sizes. The BPP asks to minimize the

total number of bins used, subject to the constraints that the total size of items added to each bin does not exceed the size of the bin. BPP is strongly NP-hard (Hartmanis [104]), meaning that no full polynomial time approximation scheme (FPTAS) exists. Over the years, many heuristics have been developed to provide high-quality solutions for practical purposes. The traditional heuristics are all for the "online" version of BPP, meaning that the list of items are shown one by one, and a decision for each item is made final as soon as the item is shown. Classical heuristics include the following.

- First Fit (FF) (Johnson et al. [123]): Upon seeing an item, it is inserted to the first bin (according to the indices of the bins) that has room for it. A new bin is opened if the item does not fit into any existing bin.

- Next Fit (NF) (Johnson et al. [123]): Upon seeing an item, it is inserted to the last existing bin (according to the indices of the bins) that has room for it. A new bin is opened if the item does not fit into any existing bin.

- Best Fit (BF) (Rhee and Talagrand [181]): Upon seeing an item, it is inserted to the fullest bin that has room for it. A new bin is opened if the item does not fit into any existing bin.

- Worst Fit (WF) (Coffman et al. [56]): Upon seeing an item, it is inserted to the emptiest bin (among those existing ones) that has room for it. A new bin is opened if the item does not fit into any existing bin.

- Almost Worst Fit (AWF) (Coffman et al. [56]): Upon seeing an item, it is inserted to the second emptiest bin that has room for it. A new bin is opened if the item does not fit into any existing bin.

For the "offline" problem, on the other hand, we are given access to the full list of items from the beginning (before making any decisions). The above heuristics may also be used, but

combined with some sorting of the items. For example, FF-Decreasing uses the First Fit heuristic on the presorted list of items, where the items are listed in decreasing order of their sizes. Other heuristics such as BF-Decreasing, NF-Decreasing, FF-Increasing are defined similarly. We refer to Coffman Jr et al. [57] for a survey on the worst-case analysis of the above algorithms.

There are also algorithms that have both online and offline flavor for BPP. One example is the Better-Fit heuristic algorithm (BFH) (Bhatia et al. [36]). In BFH, an existing item from a bin is removed and replaced with the current item if the current item better fills the bin. If the packing of the current item results in a smaller remaining space than the packing of the existing item, then the existing item is removed from the bin it is in. The replaced item is then packed again using BFH. Such procedure continues for all items until better-fit cannot pack a replaced item, in which case it is packed with BF heuristic.

In recent years, there are also developments of more complicated metaheuristic approaches for solving the BPP. Examples include the Whale Optimization Algorithm (WOA) (Mirjalili and Lewis [161]) (may be improved with Lévy Flights (Abdel-Basset et al. [1])), (Adaptive) Cuckoo Search (may also incorporate with Lévy Flights) (Yang and Deb [226]), Squirrel Search Algorithm (Jain et al. [118]), the Fitness-Dependent Optimizer (FDO) (Abdullah and Ahmed [3], Abdul-Minaam et al. [2]), and so on. Since BPP is still not so close to our FCP, we do not extend our discussions on these metaheuristics. We refer to (Munien and Ezugwu [165]) for a comprehensive survey of the aforementioned algorithms.

5.2.2 Variations of BPP

One major restriction of the classical BPP is that the objective is simply minimizing the number of bins used, and these bins are assumed to be identical. In our FCP, however, containers may differ in their size/dimensions, and the costs of containers are different from each other. Luckily, a number of variations of the classical BPP have also been studied in

the literature.

5.2.2.1 Bin Packing Problem with General Cost Structures (GCBP)

In GCBP, the cost of a bin is not one, but depends on the number of items actually inserted into this bin. Specifically, the cost of a bin is given by a function $f : \{0, 1, 2, \ldots, n\} \to \mathbb{R}^+$, where f is a monotonically non-decreasing concave function, and $f(0) = 0$. In words, if the bin has been inserted k items, the cost of that bin would be $f(k)$. GCBP was first proposed in Anily et al. [15], where the worst-case performance of some BPP heuristics was analyzed. Specifically, it was shown that many common heuristics for BPP, such as FF, BF, and NF as described in Section 5.2.1 do not have a finite asymptotic approximation ratio, while NF-Decreasing was shown to have an asymptotic approximation ratio of exactly 2. Moreover, the BF-Increasing, FF-Increasing and NF-Increasing achieve a better asymptotic approximation ratio of approximately 1.691. It was also shown in Anily et al. [15] that any heuristic that is independent of f has an asymptotic approximation ratio of at least $\frac{4}{3}$. Later, Epstein and Levin [70] developed an asymptotic fully polynomial time approximation scheme (AFPTAS) and proved the tight bound of 1.5 asymptotic approximation ratio.

5.2.2.2 Generalized Bin Packing Problem (GBPP)

GBPP was first introduced in Baldi et al. [30]. In GBPP, a set of items I with volume and profit has to be loaded into proper bins. Items can be either compulsory or non-compulsory, i.e., the item set is partitioned into two subsets: items in I^C are mandatory to load into any bin, and items in I^{NC} are optional. Bins are also classified in bin types, where bins belonging to the same type have the same capacity and cost. Moreover, for each bin type, there is a maximum number of bins that can be used. The objective is to accommodate all compulsory items and possibly non-compulsory items into appropriate bins in order to minimize the overall cost, which is the total cost of all used bins deducted by the total profit

earned from the items.

GBPP differs from FCP in two ways: first, only one set of capacity constraints are considered; second, in GBPP, each item has the same profit (or cost) if inserted into different bins, while in FCP, items would cost differently if inserted into different containers. Even though GBPP is a still a much simplified version of the FCP, it was shown in Baldi et al. [31] and Baldi and Bruglieri [29] that GBPP cannot be approximated by any constant, unless $P = NP$.

5.2.2.3 Generalized Bin Packing Problem with Bin-Dependent Item Profits (GBPPI)

GBPPI extends GBPP by allowing that when an item is inserted into different bins, the profit earned from that item may be different. In this sense, GBPPI is the closest model to FCP, with the only difference being the absence of an additional set of capacity constraints on containers. GBPPI was introduced in Baldi et al. [32], and to the best of our knowledge, there has been no further studies on the same problem since then. Since this is closely relevant to our problem, we discuss the algorithms in Baldi et al. [32] in more detail. The overall approach can be described in three steps.

1. Constructive Heuristics. Items are given in a presorted list, and are visited one by one. All containers are closed initially. Let p_{ij} be the profit of inserting i to bin j, and let $\Phi_{res}(j)$ be the remaining space of bin j after inserting i. Upon seeing an item i, compute a weighted profit of inserting item i to bin j for all bins that are opened and has enough capacity for item i. The weighted profit is calculated as

$$\alpha \cdot p_{ij} + (1 - \alpha) \cdot \Phi_{res}(j), \tag{5.1}$$

where α is some parameter that can be configured. We then insert i to the bin j that

results in a maximum weighted profit.

This insertion process may be generalized by looking at N items each time, where N is another parameter to be configured, rather than just one item. Specifically, we look at item i and the succeeding $N-1$ items in the list. For each item, we find the best bin according to (5.1), and then select the best item-bin pair that maximizes the weighted profit.

If no bin is feasible, there are two different heuristics to choose a new bin to open:

- Best Profitable (BP). BP heuristics considers item i and the remaining succeeding items in the item list, and selects the bin that maximizes the overall profit, which is the sum of profits of the items that can be inserted into the bin deducted by the cost of that bin. If the overall profit is negative and item i is non-compulsory, then item i is discarded.
- Best Assignment (BA). BA heuristics selects the bin that maximizes the profit for item i.

At the end when all items are inserted to some bins, a post-optimization procedure is performed, which consists of two parts. First, for each bin used in the solution, we try to perform (if possible) the best swap with a bin that has not been used. Second, we remove bins from the solution that are not profitable and do not contain compulsory items.

2. Greedy Adaptive Search Procedure (GASP). GASP, shown as Algorithm 5.1, is a metaheuristic that uses BA or BP as a subroutine. The multi-start initialization generates some initial solution and sets the initial parameters of α, N that will be used in the BP or BA constructive heuristics. Before reaching some preset time limit, the algorithm at each round first sorts the items uniformly randomly. The BP or BA heuristic is then performed, and if the resulting solution is better than the best

one found so far, we replace the best solution as the current one, and perform "1 to 1" swaps to search the neighborhood of the current solution. A swap consists of unloading one item to create sufficient room to insert another item that was not part of the solution. If the heuristic solution is not better than the best one, the counter *numConsecutive* is incremented. If no better solution is found after performing $MAXCONSECUTIVE$ number of constructive heuristics, we jump to the LONG-TERM INITIALIZATION PROCEDURE which will reset different parameters for α, N.

Algorithm 5.1 The GASP (Baldi et al. [32])

```
1: IS : Initial solution provided by the MULTI-START INITIALIZATION procedure
2: BS : best solution
3: BS := IS
4: numConsecutive : number of consecutive non-improving solutions
5: numConsecutive := 0
6: while time limit has not been reached do
7:     sort the items
8:     perform either the BP or the BA constructive heuristic
9:     store the resulting solution as CS
10:    if CS < BS then
11:        BS := CS
12:        perform "1 to 1" swaps
13:        numConsecutive := 0
14:    else
15:        numConsecutive := numConsecutive + 1
16:    end if
17:    SCORE UPDATE procedure
18:    if numConsecutive = MAXCONSECUTIVE then
19:        LONG-TERM REINITIALIZATION procedure
20:        numConsecutive := 0
21:    end if
22: end while
```

3. Model-Based Matheuristic (MBM). MBM is a parallel matheuristic for the GBPPI. During each iteration we feed the MBM a solution from GASP. Then, the set of bins used in the solution is randomly partitioned into P subsets, where P is the total number of threads available for the parallel computing. Each thread then solves the GBPPI

problem using a solver with some time limit, e.g. 1 second, where the problem instance only uses a subset of bins, the items loaded to those bins, and the items not loaded in the solution. The partial solutions returned by the solver are then merged to create a new current solution, and if the current solution is better, we save it as the best solution. This process is repeated until some time limit is reached.

In Baldi et al. [32], the above algorithms were also tested using both artificial instances and some instances from the parcel delivery in last-mile logistics.

5.3 Problem Formulation and Non-Approximability Result

In this section, we first define what we call the Freight Consolidation Problem (FCP). Then, we present the non-approximability result of the FCP. An instance of the FCP is given by a set of shipments and a set of containers. Each shipment has a weight and a volume, and each container has its own weight limit (capacity) and volume limit. There is a cost associated with assigning each shipment to each container (shipment cost), and, if any container is used (been assigned any shipment), there will be a procurement cost of that container (container cost). The goal is to assign all shipments to some containers to minimize the overall cost (total of shipment costs and container costs), subject to the volume and weight capacity constraints of these containers. In the following, we formulate the FCP as an integer linear program (ILP).

Sets:

- $\mathcal{S} = \{1, 2, \ldots, |\mathcal{S}|\}$ - set of shipments (indexed by s)
- $\mathcal{C} = \{1, 2, \ldots, |\mathcal{C}|\}$ - set of containers (indexed by c)

Parameters:

- ξ_{sc} - cost of packing shipment piece s into container c, assigned ∞ if cannot ship s with c

- p_c - procurement cost of container c
- ϕ_s - weight of shipment s
- Φ_c - weight limit of container c
- v_s - volume of shipment s
- V_c - volume limit of container c

Binary decision variables:

- $\mu_{sc} = 1$ if s is assigned to container c
- $\mu_c = 1$ if container c is used

The optimization problem (FCP):

$$\min_{\mu_{sc},\mu_c} \sum_{c\in\mathcal{C}} \sum_{s\in\mathcal{S}} \xi_{sc}\mu_{sc} + \sum_{c\in\mathcal{C}} p_c\mu_c \tag{5.2a}$$

$$\text{s.t.} \sum_{c\in\mathcal{C}} \mu_{sc} = 1, \quad \forall s \in \mathcal{S}, \tag{5.2b}$$

$$\sum_{s\in\mathcal{S}} \phi_s\mu_{sc} \leq \Phi_c, \quad \forall c \in \mathcal{C}, \tag{5.2c}$$

$$\sum_{s\in\mathcal{S}} v_s\mu_{sc} \leq V_c, \quad \forall c \in \mathcal{C}, \tag{5.2d}$$

$$\mu_c \geq \mu_{sc}, \quad \forall s \in \mathcal{S},\ \forall c \in \mathcal{C}, \tag{5.2e}$$

$$\mu_{sc}, \mu_c \in \{0,1\}, \quad \forall s \in \mathcal{S}, c \in \mathcal{C}.$$

The objective (5.2a) is to minimize the total cost, which includes both the cost of shipping and the cost of containers. (5.2b) implies that each shipment must be assigned to one of the containers. (5.2c) and (5.2d) ensure that the total weight (resp. volume) of shipments assigned to each container does not exceed the weight (resp. volume) limit of that container.

Lastly, (5.2e) forces us to pay the cost of a container as long as at least one of the shipments is assigned to that container.

The *approximation ratio* of any algorithm that solves FCP is defined as follows.

Definition 5.1. *Given the minimization problem* (5.2)*, an instance* π *of the problem, an algorithm* ALG*, the optimum* $\text{OPT}(\pi) \geq 0$*, and value* $\text{ALG}(\pi)$ *of the solution computed by the algorithm, the* approximation ratio *of the algorithm* ALG *is the infimum* $\alpha \geq 1$ *such that*

$$\text{ALG}(\pi) \leq \alpha \cdot \text{OPT}(\pi), \quad \forall \pi, \tag{5.3}$$

i.e., for all instances, the output of the algorithm incurs a total cost that is at most α *times the optimal value.*

We next have the following non-approximability result for FCP.

Proposition 5.1. *For any constant* α*, there is no polynomial-time algorithm for the Freight Consolidation Problem (FCP)* (5.2) *with approximation ratio* α*, unless* $P = NP$*.*

Proof. We prove by reduction from the decision version of the Bin Packing Problem (BPP). Consider an instance $\hat{\pi}$ of the BPP, which consists of n items, each with a volume v_i for $i = 1, \ldots, n$, and unlimited number of bins, each with a capacity V, where $V \geq v_i$ for all $i = 1, \ldots, n$. The decision version of the BPP asks if it is feasible to assign all items to the bins such that at most k bins are used. This instance $\hat{\pi}$ of BPP can be transformed into an instance π of the FCP as follows. The instance π of the FCP would include n shipments, each with volume v_i for $i = 1, \ldots, n$. The weight of these shipments are all 0. There are also $k+n$ containers with volume capacity V and weight capacity one. The cost of procuring each of the containers $1, \ldots, k$ is one, and the cost of procuring each of the containers $k+1, \ldots, k+n$ are $k\alpha$. All shipment costs ξ are zero. We note that, if $\hat{\pi}$ for BPP has a solution, then the optimal value of the FCP is at most k; otherwise if $\hat{\pi}$ does not have a solution, then the

optimal value of the FCP must be greater than $k\alpha$ since at least one container with cost $k\alpha$ must be used.

Suppose that to the contrary a polynomial time algorithm approximating the FCP with a constant $\alpha > 1$ exists, then through such an algorithm we would be able to determine if an instance of the BPP has a solution: the algorithm would return value $\leq k\alpha$ for the instances of the FCP corresponding to the instances of the BPP which have a solution, and the algorithm would return value $> k\alpha$ for those corresponding to the instances of BPP without a solution. Unless $P = NP$, this is impossible since the decision version of the BPP is NP-complete. □

Since there is no constant factor approximation for the FCP (assuming $P \neq NP$), we propose in the next section some intuitive greedy heuristics for the problem.

5.4 Proposed Heuristics

In this section, we propose a series of greedy-type heuristics that find solutions that are (hopefully) close to optimal.

5.4.1 Greedy Cost-Feasibility Algorithm (GR)

5.4.1.1 Overview

In this subsection, we propose a greedy heuristic for the FCP, which we call the Greedy Cost-Feasibility algorithm. In this algorithm, we first assign all shipments to the containers such that the shipping cost is the lowest, i.e., for each shipment s, we find one container c' such that $\xi_{sc'} = \min_c \xi_{sc}$, and assign shipment s to container c'. This assignment provides a lower bound on the total shipping costs. The assignment, however, may not be feasible as some of the capacity constraints of the containers may be violated. In each of the following steps, the algorithm moves one shipment at a time, from one container to another, to make

the assignment move towards feasibility, while keeping the increment of the shipping cost at a minimum.

5.4.1.2 Overflow Score

We define an "overflow score" on each container for any given assignment, and use this overflow score together with the shipping costs to determine which shipment to be moved to which container. For any assignment μ, the overflow score for container c is defined as

$$O_c(\mu) := \beta_1 \cdot \frac{\left[\sum_{\{s|\mu_{sc}=1\}} v_s - V_c\right]^+}{V_c} + \beta_2 \cdot \frac{\left[\sum_{\{s|\mu_{sc}=1\}} \phi_s - \Phi_c\right]^+}{\Phi_c}, \tag{5.4}$$

where β_1, β_2 are some adjustable parameters that satisfy $\beta_1, \beta_2, \beta_1 + \beta_2 \in [0, 1]$. The first term of the overflow score measures the percentage volume overflow of container c, and the second term measures the percentage weight overflow of container c. These two terms are summed together with weights β_1, β_2 to obtain the overflow score of container c.

The total overflow score of an assignment is then defined as

$$O(\mu) := \sum_c O_c(\mu). \tag{5.5}$$

5.4.1.3 Moving Towards Feasibility

After computing the overflow score of each container given the initial assignment, we find those containers with $O_c(\mu) > 0$, i.e., containers that are not feasible. For each shipment in these containers, we try to move the shipment out of its current container to another container, and compute the new overflow score O'. Let μ denote the current assignment, and $\mu^{sc'}$ denote the new assignment that moves shipment s from its current container to container c'. If we move the shipment s from its current container c to container c', we will

have the following cost-feasibility ratio:

$$\mathcal{R}(s, c') := \frac{\xi_{sc'} - \xi_{sc}}{O(\mu) - O(\mu^{sc'})}. \tag{5.6}$$

The algorithm decides to move the shipment s from c to c' that minimizes the above ratio. In other words, in deciding which move to take, we choose the move that incurs least incremental shipping cost per unit reduction of the overflow score.

Since there are always coloading options for those shipments in the overflowed containers, at each round after the move, the overflow score is guaranteed to decrease. We repeat this process until the overflow score decreases to zero, at which time we have a feasible solution.

In the end, we also perform a post-adjustment procedure by looking at each used container (containers with $\mu_c = 1$)[1] and the shipments assigned to it. We will remove that container and coload all shipments assigned to it if it is more profitable to do so.

5.4.1.4 Algorithm Summary

The complete Greedy Cost-Feasibility (GR) algorithm is given as Algorithm 5.2.

5.4.2 Greedy + Local Search (GRL)

The next heuristic we introduce is Greedy with Local Search (GRL).

5.4.2.1 Overview

From the solution of GR, we perform local movements of shipments. Specifically, we search in two neighborhoods of a solution: the "shift" neighborhood, which consists of all solutions obtained by reassigning one shipment from the current solution, and the "swap" neighborhood, which consists of all solutions obtained by swapping the assignment of two shipments

1. In the rest of this chapter, we also say a container c is "opened" if $\mu_c = 1$, and "closed" if $\mu_c = 0$.

Algorithm 5.2 GREEDY COST-FEASIBILITY (GR)

Input: shipment info, container info, β_1, β_2 ▷ β_1, β_2 are adjustable parameters
Output: Assignment of each shipment to a container
GREEDY PROCEDURE
Assign each shipment to its shipment cost-minimizing container, i.e., assign s to a c' such that $\xi_{sc'} \leq \xi_{sc}, \forall c$. Denote the current assignment by μ.
Compute the overflow score of the current assignment $O(\mu)$.
while $O(\mu) > 0$ **do**
    For each shipment-container pair (s, c), compute the cost-feasibility ratio $\mathcal{R}(s, c)$ if s is reassigned to c.
    Find the pair (s, c) with the minimum $\mathcal{R}(s, c)$. Reassign s to c.
    Compute the new overflow score.
end while
POST-ADJUSTMENT PROCEDURE
for each container c with $\mu_c = 1$ **do**
    Find all shipments s that has been assigned to c.
    if $p_c + \sum_{s \text{ assigned to } c} \xi_{sc} > \sum_{s \text{ assigned to } c} \xi_{s1}$ **then**
        $\mu_c = 0$, coload all these shipments. ▷ Coload all shipments in c if more profitable
    end if
end for

from the current solution. In searching each neighborhood, there are two standard ways of performing movements: first-admissible (FA) and best-admissible (BA).

- In the first-admissible scheme, we randomly search the neighborhood and take the move as soon as we find a better solution.
- In the best-admissible scheme, we search all possible moves and thus all solutions in the neighborhood, and choose to take the move that leads to the most reduction in the shipment cost.

It has been shown in Osman [169] that for the generalized assignment problem (GAP), BA returns a slightly better solution, but takes much longer time to generate the solution. We therefore choose FA in our implementations for two reasons: first, the (potentially) slightly better solution from BA may not be worth the extra time; second, our problem size is much larger than those that have been experimented upon in the GAP literature.

5.4.2.2 Searching the "Shift" Neighborhood

The search of the "shift" neighborhood is performed in cycles. In each cycle, we first randomly sort the list of all shipments. Then, starting from the first shipment s in the list, we sort the set of opened containers (those with $\mu_c = 1$ in the GR solution) in increasing order of μ_{sc}, and try to reassign this shipment to each container in the container list. If the reassignment is feasible, the shipment is reassigned permanently, and a new cycle is started. Otherwise, we move to the next container in the sorted container list. If no container before the current assigned container is feasible, i.e., no reassignment of the current shipment can lead to reduction in cost while keeping feasibility, we skip this shipment and move to the next shipment. This process is repeated until we reach a cycle where no feasible improvement relocation can be made, at which time the solution is locally optimal in its "shift" neighborhood.

5.4.2.3 Searching the "Swap" Neighborhood

The search of the "swap" neighborhood is also performed in cycles. We first generate a list of all pairs of shipments. In each cycle, we sort this list randomly. Then, starting from the first shipment pair in the list, we try to swap the assignment of the two shipments. If the assignment after the swap is feasible for both containers, and the swap leads to a reduction in the total shipment cost, the swap is made permanent and a new cycle will start. Otherwise, we move to the next pair of shipments. This process is repeated until we reach a cycle where no swaps are made after visiting all shipment pairs, at which time the solution is locally optimal in its "swap" neighborhood.

5.4.2.4 Local Optimal Solution in Both Neighborhoods

Given any input solution, we first repeatedly search the "shift" neighborhood. We always keep the best solution found so far, and the search is repeated until no better solution is found

after $Max\_Nonimprove\_S$ consecutive number of searches. Next, we search the "swap" neighborhood of the best solution found so far (locally optimal within the "shift" neighborhood), after which we reach a locally optimal solution within the "swap" neighborhood. If the new solution is better than the solution before searching the "swap" neighborhood, we will again repeatedly search the "shift" neighborhood and then the "swap" neighborhood. The whole process is repeated until no better solution is found after $Max\_Nonimprove$ consecutive number repetitions, at which point the solution is locally optimal within both neighborhoods.

5.4.2.5 Algorithm Summary

The complete Greedy + Local Search (GRL) algorithm is given as Algorithm 5.3.

5.4.3 Greedy + Local Search + Varying Containers (GRLV)

We now introduce the heuristic that is based on GRL, but tries to vary the set of used (opened) containers.

5.4.3.1 Overview

This heuristic consists of two layers. In the first layer, we generate a set of "seed" solutions. In the second layer, we try to vary the set of used containers on each "seed" solution, and finally return the best solution found throughout the process.

There are several intuitions behind this heuristic. First, the local search can be combined with the post-adjustment: Every time after running local search and finding a locally optimal solution, we can check again if deleting some containers and coloading all shipments in those containers can be more profitable. If such containers exist, we proceed to delete these containers. Then we can redo the local search and the post-adjustment, and repeat this process till the post-adjustment does not delete any more containers. Second, every time

Algorithm 5.3 GREEDY + LOCAL SEARCH (GRL)

Input: shipment info, container info, β_1, β_2, $Max\_Nonimprove\_S$, $Max\_Nonimprove$
Output: Assignment of each shipment to a container

Run GREEDY PROCEDURE (as in Algorithm 5.2).
Run POST-ADJUSTMENT PROCEDURE (as in Algorithm 5.2), save as "initial solution".
LOCAL-SEARCH PROCEDURE
"best solution" = "initial solution"
$Outer\_counter = 0$
while $Outer\_counter < Max\_Nonimprove$ **do**
    $Inner\_counter = 0$
    "best shift solution" = "initial solution"
    while $Inner\_counter < Max\_Nonimprove\_S$ **do**
        Search the "shift" neighborhood of the "initial solution", save as "shift solution"
        if "shift solution" has lower total cost than "best shift solution" **then**
            "best shift solution" = "shift solution"
            $Inner\_counter = 0$
        else
            $Inner\_counter = Inner\_counter + 1$
        end if
    end while
    Search the "swap" neighborhood of the "best shift solution", save as "swap solution"
    while "swap solution" has lower cost than "best shift solution" **do**
        $Inner\_counter = 0$
        "best shift solution" = "swap solution"
        while $Inner\_counter < Max\_Nonimprove\_S$ **do**
            Search the "shift" neighborhood of the "swap solution", save as "shift solution"
            if "shift solution" has lower total cost than "best shift solution" **then**
                "best shift solution" = "shift solution"
                $Inner\_counter = 0$
            else
                $Inner\_counter = Inner\_counter + 1$
            end if
        end while
        Search the "swap" neighborhood of the "best shift solution", save as "swap solution"
    end while
    if "swap solution" has lower cost than the "best solution" **then**
        "best solution" = "swap solution"
        $Outer\_counter = 0$
    else
        $Outer\_counter = Outer\_counter + 1$
    end if
end while
Return "best solution"

we perform some procedure that might change the set of used (opened) containers, we might do further local search based on the current solution, or we can also build a new solution from scratch, again using the GREEDY PROCEDURE, but this time fixing the set of unopened containers , i.e., set $\xi_{sc} = \infty$ for all containers that are not open before applying

the Greedy Procedure. Third, every time we try to vary the set of containers, we can either add/delete one container at a time, or we can add/delete a number of containers altogether. In the following, we describe the procedures/subroutines that are used in this heuristic.

5.4.3.2 Adjusted Local Search

We may combine the Post-Adjustment Procedure with the Local-Search Procedure, then iterate both procedures repeatedly until the set of opened containers no longer changes so that we obtain a local optimum within both neighborhoods. We define the Adjust-Local Procedure as Algorithm 5.4.

Algorithm 5.4 Adjust-Local Procedure

Input: initial solution
Output: updated solution
"updated solution" = "initial solution"
$Num\_del\_master = 1$
while $Num\_del\_master > 0$ **do**
    Run Local-Search Procedure on "updated solution", save as "updated solution"
    Run Post-Adjustment Procedure on "updated solution", save as "updated solution"
    Save the number of deleted containers in the Post-Adjustment Procedure as Num_del_master
end while
Return "updated solution"

5.4.3.3 Adding One of the Deleted Containers Back

Since the Post-Adjustment Procedure deletes some containers, we try to add one of those deleted containers back to the solution and then perform Adjust-Local Procedure. In the end, we save the best solution found during this process. The Add-One Procedure is defined as Algorithm 5.5.

Algorithm 5.5 ADD-ONE PROCEDURE

Input: initial solution
Output: updated solution
"updated solution" = "initial solution"
Run ADJUST-LOCAL PROCEDURE on "initial solution", save as "cand solution", save the set of deleted containers as S
Run ADJUST-LOCAL PROCEDURE on "initial solution", save as "updated solution"
for each container in set S **do**
    Reopen the container in the "cand solution", and add the shipments what were assigned to this container in the "initial solution" to this container, save as "current solution"
    if "current solution" has lower total cost than "updated solution" **then**
        "updated solution" = "current solution"
    end if
    Close this container in the "cand solution"
end for
Return "updated solution"

5.4.3.4 Deleting a Chain of Containers

We observe that the GR solution, even after the POST-ADJUSTMENT PROCEDURE, uses more containers than the optimal solution returned by the solver. Based on an initial solution, we try to delete a chain of containers. Specifically, we sort the containers in increasing order of their profit, i.e., for each container c, we compute:

$$\text{Profit of using container } c := \sum_{s:\mu_{sc}=1} \xi_{s1} - \left(p_c + \sum_{s:\mu_{sc}=1} \xi_{sc} \right), \tag{5.7}$$

which is the total coloading cost of the shipments assigned to container c deducted by the total shipping cost of those shipments and the procurement cost of the container. This is the actual "saving" from using container c for these shipments, compared with the cost of coloading all these shipments.

We delete the top k containers in the list from the initial solution and perform the ADJUST-LOCAL PROCEDURE, where k ranges from 0 to num_cont_del (a preset parameter). In the end, we output the best solution among these $(k+1)$ solutions. The DEL-CHAIN PROCEDURE is defined as Algorithm 5.6.

Algorithm 5.6 DEL-CHAIN PROCEDURE

Input: initial solution, num_cont_del
Output: updated solution
"updated solution" = "initial solution"
"current solution" = "initial solution"
Sort the containers used in the "initial solution" in increasing order of their total profit (5.7). Save as "sorted list"
for $j \in \{0, 1, 2, \ldots, num\_cont\_del\}$ **do**
    Delete the jth container from the "current solution", coload all shipments previously assigned to that container, save as "current solution"
    Run ADJUST-LOCAL PROCEDURE on "current solution", save as "new solution"
    if "new solution" has lower total cost than "updated solution" **then**
        "updated solution" = "new solution"
    end if
end for
Return "updated solution"

5.4.3.5 Deleting One More Container

Given an initial solution, we may again sort the containers in increasing order of their profits (5.7), and try to delete one container from the top num_cont_del containers in the sorted list. The best solution is saved in the end. We define the DEL-ONE PROCEDURE as Algorithm 5.7.

Algorithm 5.7 DEL-ONE PROCEDURE

Input: initial solution, num_cont_del
Output: updated solution
"updated solution" = "initial solution"
Sort the containers used in the "initial solution" in increasing order of their total profit (5.7). Save as "sorted list"
for $j \in \{0, 1, 2, \ldots, num\_cont\_del\}$ **do**
    Delete the jth container from the "initial solution", coload all shipments previously assigned to that container, save as "current solution"
    Run ADJUST-LOCAL PROCEDURE on "current solution", save as "current solution"
    if "current solution" has lower total cost than "updated solution" **then**
        "updated solution" = "current solution"
    end if
end for
Return "updated solution"

5.4.3.6 Deleting Containers One by One

Starting from some initial solution, we can repeatedly perform DEL-ONE PROCEDURE, until further deleting any containers leads to no improvement in the solution. The DEL-OBO PROCEDURE is defined as Algorithm 5.8.

Algorithm 5.8 DEL-OBO PROCEDURE

Input: initial solution, *num_cont_del*
Output: updated solution
"updated solution" = "initial solution"
Run DEL-ONE PROCEDURE on "initial solution", save as "current solution"
if "current solution" has lower total cost than the "updated solution" **then**
    "updated solution" = "current solution"
    while "current solution" has lower total cost than the "updated solution" **do**
        "updated solution" = "current solution"
        Run DEL-ONE PROCEDURE on "current solution", save as "current solution"
    end while
end if
Return "updated solution"

5.4.3.7 Algorithm Summary

The complete Greedy + Local Search + Varying Containers (GRLV) algorithm is given as Algorithm 5.9.

5.5 Experiments

In this section, we provide experimental results on our proposed heuristics, including GR, GRL, and GRLV. We first generate a set of instances that hopefully reflects part of the reality. Each of these instances is generated as the following:

- **Containers:** We have 150 containers in an instance (not including the "coloading" container), each with a weight capacity $\Phi_c = 28,000$ (kg) and a volume capacity $V_c = 76$ (m^3), which reflects the capacities of the most used containers (40' high-cube container). The container cost p_c is sampled from a truncated Normal distribution

Algorithm 5.9 GREEDY + LOCAL SEARCH + VARYING CONTAINERS (GRLV)

Input: shipment info, container info, β_1, β_2, $Max\_Nonimprove\_S$, $Max\_Nonimprove$, num_cont_del

Output: Assignment of each shipment to a container

1: Run GREEDY PROCEDURE (as in Algorithm 5.2), save as "GR solution"
2: Run POST-ADJUSTMENT PROCEDURE (as in Algorithm 5.2) on "GR solution", save as "PA solution"
3: Run ADJUST-LOCAL PROCEDURE on "GR solution", save as "LC solution"
4: Run GREEDY PROCEDURE on "PA" solution, i.e., first set $\xi_{sc} = \infty$ for all containers that are not open (used) in the "PA solution", then run GREEDY PROCEDURE. Save the solution as "PA_GR solution"
5: Run GREEDY PROCEDURE on "LC" solution, i.e., first set $\xi_{sc} = \infty$ for all containers that are not open (used) in the "LC solution", then run GREEDY PROCEDURE. Save the solution as "LC_GR solution"
6: Run ADD-ONE PROCEDURE on "PA solution", save as "PA_one solution"
7: Run ADD-ONE PROCEDURE on "LC solution", save as "LC_one solution"
8: Run ADD-ONE PROCEDURE on "PA_GR solution", save as "PA_GR_one solution"
9: Run ADD-ONE PROCEDURE on "LC_GR solution", save as "LC_GR_one solution"
10: Run DEL-CHAIN PROCEDURE on "PA solution", save as "CHAIN_PA solution"
11: Run DEL-CHAIN PROCEDURE on "LC solution", save as "CHAIN_LC solution"
12: Run DEL-CHAIN PROCEDURE on "PA_GR solution", save as "CHAIN_PA_GR solution"
13: Run DEL-CHAIN PROCEDURE on "LC_GR solution", save as "CHAIN_LC_GR solution"
14: Run DEL-CHAIN PROCEDURE on "PA_one solution", save as "CHAIN_PA_one solution"
15: Run DEL-CHAIN PROCEDURE on "LC_one solution", save as "CHAIN_LC_one solution"
16: Run DEL-CHAIN PROCEDURE on "PA_GR_one solution", save as "CHAIN_PA_GR_one solution"
17: Run DEL-CHAIN PROCEDURE on "LC_GR_one solution", save as "CHAIN_LC_GR_one solution"
18: Run DEL-OBO PROCEDURE on "PA solution", save as "OBO_PA solution"
19: Run DEL-OBO PROCEDURE on "LC solution", save as "OBO_LC solution"
20: Run DEL-OBO PROCEDURE on "PA_GR solution", save as "OBO_PA_GR solution"
21: Run DEL-OBO PROCEDURE on "LC_GR solution", save as "OBO_LC_GR solution"
22: Run DEL-OBO PROCEDURE on "PA_one solution", save as "OBO_PA_one solution"
23: Run DEL-OBO PROCEDURE on "LC_one solution", save as "OBO_LC_one solution"
24: Run DEL-OBO PROCEDURE on "PA_GR_one solution", save as "OBO_PA_GR_one solution"
25: Run DEL-OBO PROCEDURE on "LC_GR_one solution", save as "OBO_LC_GR_one solution"
26: Return the best solution among {"CHAIN_PA solution", "CHAIN_LC solution", "CHAIN_PA_GR solution", "CHAIN_LC_GR solution", "CHAIN_PA_one solution", "CHAIN_LC_one solution", "CHAIN_PA_GR_one solution", "CHAIN_LC_GR_one solution", "OBO_PA solution", "OBO_LC solution", "OBO_PA_GR solution", "OBO_LC_GR solution", "OBO_PA_one solution", "OBO_LC_one solution", "OBO_PA_GR_one solution", "OBO_LC_GR_one solution"}.

(lower bounded at 0) with the mean 9000 and the standard deviation 4000. The "coloading" container, however, has a cost 0, and infinite weight and volume capacities.

- **Shipments:** We have 1000 shipments in an instance, each with its weight and volume sample from the truncated bivariate Normal distribution (lower bounded at 0) with the means $(2000, 10)$ and the covariance matrix $\begin{bmatrix} 250,000,000 & 1,000,000 \\ 1,000,000 & 4,500 \end{bmatrix}$.

- **Shipment costs:** Each shipment has a limited number of feasible non-coloading con-

tainers. For each shipment, the number of feasible containers is sampled from the truncated Normal distribution (lower bounded at 0) with the mean 10 and the standard deviation 10. Then, if shipment s has k number of feasible containers, we randomly select k containers from the container set, plus the "coloading" container. The shipment costs ξ_{sc} are sampled from a truncated Normal distribution (lower bounded at 0) with the mean 3500 and the standard deviation $10,000$.

The experiments were run on 20 simulated instances generated as above. These instances have much larger sizes than any of those tested in the Bin Packing or Generalized Assignment Problem literature. In GR, we set the parameters $\beta_1 = \beta_2 = 0.5$. In GRL, we further set the parameters $Max\_Nonimprove\_S = 1$ and $Max\_Nonimprove = 10$. In GRLV, we start with generating different GR solutions by setting different parameters of β_1, β_2 (β_1 ranging from 1 to 5 and β_2 ranging from 1 to 5). We then fix the set of β_1, β_2 that gives the best GR solution, and the parameter num_cont_del is set to 5. The benchmark is the solution of the integer linear program (5.2) returned by the Gurobi solver whose default optimality gap is 0.01%, and the solving time limit is set to 60 seconds. The setups of the experiments are described as follows.

- Program used for implementation: Julia Version 1.7.2.
- Solver used for solving the ILP: Gurobi Version 9.5.1 (academic license).
- Machine used for running: Surface Book 2 with Intel Core i7-8650 CPU @ (1.90 GHz 2.11 GHz) and 16 GB RAM.

The results of the experiments, including the optimality gaps (compared with the optimal solutions returned by the solver) and the runtimes (in seconds) of all heuristics, averaged over the 20 instances, are summarized as Table 5.1.

Finally, we remark that while the solver is able to solve these instances to a smaller optimality gap with shorter runtime, the problem size is expected to grow significantly in

Metric	Solver	GR	GRL	GRLV
Average Optimality Gap	0.01%	8.36%	4.56%	3.73%
Average Runtime (s)	26.18	7.99	72.43	3056.92

Table 5.1: Summary of experimental results

the near future. It is likely that the solver will not be able to solve the problem when its size grows larger in the next few years. Given this expectation, a freight forwarder should be prepared to not rely on the integer linear program solver for the FCP. Therefore, our proposed heuristics will still be practically relevant.

5.6 Conclusion and Future Direction

In this chapter, we have properly defined the freight consolidation problem (FCP) - a proven important and practically relevant problem faced by freight forwarders every day and every hour at the origin ports. We proved the non-approximability result of the FCP, and proposed a series of greedy based heuristics to solve the problem. Our solutions are shown to perform well in the numerical experiments with simulated data. For future improvement of this work, we may consider more generalized definitions of the neighborhood in the local search. We may also generate the set of used (opened) containers by some types of genetic algorithms. Furthermore, it might be helpful to use Tabu list and Tabu search to avoid repeated search of candidate solutions.

REFERENCES

[1] Mohamed Abdel-Basset, Gunasekaran Manogaran, Laila Abdel-Fatah, and Seyedali Mirjalili. An improved nature inspired meta-heuristic algorithm for 1-D bin packing problems. *Personal and Ubiquitous Computing*, 22(5):1117–1132, 2018.

[2] Diaa Salama Abdul-Minaam, Wadha Mohammed Edkheel Saqar Al-Mutairi, Mohamed A. Awad, and Walaa H. El-Ashmawi. An adaptive fitness-dependent optimizer for the one-dimensional bin packing problem. *IEEE Access*, 8:97959–97974, 2020.

[3] Jaza Mahmood Abdullah and Tarik Ahmed. Fitness dependent optimizer: inspired by the bee swarming reproductive process. *IEEE Access*, 7:43473–43486, 2019.

[4] Daniel Adelman and Canan Uçkun. Dynamic electricity pricing to smart homes. *Operations Research*, 67(6):1520–1542, 2019.

[5] Bharat Adsul, Jugal Garg, Ruta Mehta, Milind Sohoni, and Bernhard Von Stengel. Fast algorithms for rank-1 bimatrix games. *Operations Research*, 69(2):613–631, 2021.

[6] Sam Aflaki and Serguei Netessine. Strategic investment in renewable energy sources: The effect of supply intermittency. *Manufacturing & Service Operations Management*, 19(3):489–507, 2017.

[7] Majid Al-Gwaiz, Xiuli Chao, and Owen Q. Wu. Understanding how generation flexibility and renewable energy affect power market competition. *Manufacturing & Service Operations Management*, 19(1):114–131, 2017.

[8] Selvaprabu Nadarajah Alessio Trivella, Stein-Erik Fleten, Denis Mazieres, and David Pisinger. Managing shutdown decisions in merchant commodity and energy production: A social commerce perspective. *Manufacturing & Service Operations Management*, 23 (2):311–330, 2021.

[9] Saed Alizamir, Francis de Véricourt, and Peng Sun. Efficient feed-in-tariff policies for renewable energy technologies. *Operations Research*, 64(1):52–66, 2016.

[10] Khaled Alshehri, Mariola Ndrio, Subhonmesh Bose, and Tamer Başar. The impact of aggregating distributed energy resources on electricity market efficiency. In *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2019.

[11] Khaled Alshehri, Mariola Ndrio, Subhonmesh Bose, and Tamer Başar. Quantifying market efficiency impacts of aggregated distributed energy resources. *IEEE Transactions on Power Systems*, 35(5):4067–4077, 2020.

[12] Khaled Alshehri, Subhonmesh Bose, and Tamer Başar. Centralized volatility reduction for electricity markets. *International Journal of Electrical Power & Energy Systems*, 133:107101, 2021.

[13] Rumen Andonov, Vincent Poirriez, and Sanjay Rajopadhye. Unbounded Knapsack problem: Dynamic programming revisited. *European Journal of Operational Research*, 123(2):394–407, 2000.

[14] Alexandar Angelus. Distributed renewable power generation and implications for capacity investment and electricity prices. *Production and Operations Management*, 30 (12):4614–4634, 2021.

[15] Shoshana Anily, Julien Bramel, and David Simchi-Levi. Worst-case analysis of heuristics for the bin packing problem with general cost structures. *Operations Research*, 42 (2):287–298, 1994.

[16] Miguel F. Anjos and Juan A. Gómez. Operations research approaches for building demand response in a smart grid. In *Leading developments from INFORMS communities*, pages 131–152. INFORMS, 2017.

[17] Ravi Anupindi and Li Jiang. Capacity investment under postponement strategies, market competition, and demand uncertainty. *Management Science*, 54(11):1876–1890, 2008.

[18] Ali Aouad and Danny Segev. An approximate dynamic programming approach to the incremental Knapsack problem. *arXiv preprint arXiv:2010.07633*, 2020.

[19] Krzysztof R. Apt and Sunil Simon. A classification of weakly acyclic games. *Theory and Decision*, 78(4):501–524, 2015.

[20] Ignacio Aravena, Quentin Lété, Anthony Papavasiliou, and Yves Smeers. Transmission capacity allocation in zonal electricity markets. *Operations Research*, 69(4):1240–1255, 2021.

[21] Gürdal Arslan and Serdar Yüksel. Decentralized Q-learning for stochastic teams and games. *IEEE Transactions on Automatic Control*, 62(4):1545–1558, 2016.

[22] Baris Ata, A Serasu Duran, and Ozge Islegen. An analysis of time-based pricing in retail electricity markets. *Available at SSRN 2826055*, 2018.

[23] Merve Aydemir and Tuncay Yigit. A review of the solutions for the container loading problem, and the use of heuristics. In *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering*, pages 690–700. Springer, 2019.

[24] Sarper Aydın and Ceyhun Eksin. Decentralized inertial best-response with voluntary and limited communication in random communication networks. *Automatica*, 145: 110566, 2022.

[25] Volodymyr Babich, Ruben Lobel, and Şafak Yücel. Promoting solar panel investments: Feed-in-tariff vs. tax-rebate policies. *Manufacturing & Service Operations Management*, 22(6):1148–1164, 2020.

[26] Tamer Başar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory*, volume 23. SIAM, 1999.

[27] Yu Bai and Chi Jin. Provable self-play algorithms for competitive reinforcement learning. In *International Conference on Machine Learning*, pages 551–560. PMLR, 2020.

[28] Yu Bai, Chi Jin, and Tiancheng Yu. Near-optimal reinforcement learning with self-play. In *Advances in Neural Information Processing Systems*, volume 33, pages 2159–2170, 2020.

[29] Mauro Maria Baldi and Maurizio Bruglieri. On the generalized bin packing problem. *International Transactions in Operational Research*, 24(3):425–438, 2017.

[30] Mauro Maria Baldi, Teodor Gabriel Crainic, Guido Perboli, and Roberto Tadei. The generalized bin packing problem. *Transportation Research Part E: Logistics and Transportation Review*, 48(6):1205–1220, 2012.

[31] Mauro Maria Baldi, Teodor Gabriel Crainic, Guido Perboli, and Roberto Tadei. Asymptotic results for the generalized bin packing problem. *Procedia-Social and Behavioral Sciences*, 111:663–671, 2014.

[32] Mauro Maria Baldi, Daniele Manerba, Guido Perboli, and Roberto Tadei. A generalized bin packing problem for parcel delivery in last-mile logistics. *European Journal of Operational Research*, 274(3):990–999, 2019.

[33] Carolyn L. Beck and Rayadurgam Srikant. Error bounds for constant step-size Q-learning. *Systems & Control Letters*, 61(12):1203–1208, 2012.

[34] Dimitri P. Bertsekas and John N. Tsitsiklis. Neuro-dynamic programming: an overview. In *Proceedings of 1995 34th IEEE Conference on Decision and Control*, volume 1, pages 560–564. IEEE, 1995.

[35] Dimitris Bertsimas, Eugene Litvinov, Xu Andy Sun, Jinye Zhao, and Tongxin Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2013.

[36] Avnish K. Bhatia, M. Hazra, and S.K. Basu. Better-fit heuristic for one-dimensional bin-packing problem. In *2009 IEEE International Advance Computing Conference*, pages 193–196. IEEE, 2009.

[37] Daniel Bienstock, Jay Sethuraman, and Chun Ye. Approximation algorithms for the incremental Knapsack problem via disjunctive programming. *arXiv preprint arXiv:1311.4563*, 2013.

[38] Kostas Bimpikis, Shayan Ehsani, and Rahmi İlkılıç. Cournot competition in networked markets. *Management Science*, 65(6):2467–2481, 2019.

[39] John R. Birge. Network structure and its impact on commodity markets. *Production and Operations Management*, 30(12):4568–4574, 2021.

[40] John R. Birge, Ali Hortaçsu, and J. Michael Pavlin. Inverse optimization for the recovery of market structure from market outcomes: An application to the MISO electricity market. *Operations Research*, 65(4):837–855, 2017.

[41] John R. Birge, Hongfan Chen, N Bora Keskin, and Amy Ward. To interfere or not to interfere: Information revelation and price-setting incentives in a multiagent learning environment. *Available at SSRN 3864227*, 2021.

[42] Shant Boodaghians, Bhaskar Ray Chaudhury, and Ruta Mehta. Polynomial time algorithms to find an approximate competitive equilibrium for chores. *arXiv preprint arXiv:2107.06649*, 2021.

[43] Simina Brânzei, Vasilis Gkatzelis, and Ruta Mehta. Nash social welfare approximation for strategic agents. *Operations Research*, 2021.

[44] Maira Bruck, Peter Sandborn, and Navid Goudarzi. A levelized cost of energy (LCOE) model for wind farms that include power purchase agreements (PPAs). *Renewable Energy*, 122:131–139, 2018.

[45] Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172, 2008.

[46] Timothy M. Chan. Approximation Schemes for 0-1 Knapsack. In Raimund Seidel, editor, *1st Symposium on Simplicity in Algorithms (SOSA 2018)*, volume 61 of *OpenAccess Series in Informatics (OASIcs)*, pages 5:1–5:12, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-064-4. doi:10.4230/OASIcs.SOSA.2018.5. URL `http://drops.dagstuhl.de/opus/volltexte/2018/8299`.

[47] Mark Chediak. Amazon (AMZN) leads corporate clean energy purchases in record year, Jan 2022. URL `https://www.bloomberg.com/news/articles/2022-01-31/amazon-amzn-leads-corporate-clean-energy-purchases-in-record-year?leadSource=uverify+wall`.

[48] Chandra Chekuri and Sanjeev Khanna. A polynomial time approximation scheme for the multiple Knapsack problem. *SIAM Journal on Computing*, 35(3):713–728, 2005.

[49] Cong Chen, Ahmed S. Alahmed, Timothy D. Mount, and Lang Tong. Competitive DER aggregation for participation in wholesale markets. *arXiv preprint arXiv:2207.00290*, 2022.

[50] Cong Chen, Subhonmesh Bose, and Lang Tong. DSO-DERA coordination for the wholesale market participation of distributed energy resources. *arXiv preprint arXiv:2211.16585*, 2022.

[51] Zaiwei Chen, Sheng Zhang, Thinh T. Doan, John-Paul Clarke, and Siva Theja Maguluri. Finite-sample analysis of nonlinear stochastic approximation with applications in reinforcement learning. *arXiv e-prints*, pages arXiv–1905, 2019.

[52] Zaiwei Chen, John Paul Clarke, and Siva Theja Maguluri. Target network and truncation overcome the deadly triad in Q-learning. *arXiv preprint arXiv:2203.02628*, 2022.

[53] Zaiwei Chen, Kaiqing Zhang, Eric Mazumdar, Asuman Ozdaglar, and Adam Wierman. A finite-sample analysis of payoff-based independent learning in zero-sum stochastic games. *arXiv preprint arXiv:2303.03100*, 2023.

[54] Brett Christophers. Taking renewables to market: Prospects for the after-subsidy energy transition: The 2021 antipode rgs-ibg lecture. *Antipode*, 54(5):1519–1544, 2022.

[55] Caroline Claus and Craig Boutilier. The dynamics of reinforcement learning in cooperative multiagent systems. *AAAI Conference on Artificial Intelligence*, 1998(746-752): 2, 1998.

[56] Edward G. Coffman, Gabor Galambos, Silvano Martello, and Daniele Vigo. Bin packing approximation algorithms: Combinatorial analysis. In *Handbook of Combinatorial Optimization*, pages 151–207. Springer, 1999.

[57] Edward G. Coffman Jr, Michael R. Garey, and David S. Johnson. Approximation algorithms for bin packing: A survey. In Dorit S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 46–93. PWS Publishing Co., Boston, MA, USA, 1996.

[58] Jeffrey J. Cook, Kristen Ardani, Eric O'Shaughnessy, Brittany Smith, and Robert Margolis. Expanding PV value: Lessons learned from utility-led distributed energy resource aggregation in the United States. Technical report, National Renewable Energy Laboratory, 2018.

[59] José Correa, Cristóbal Guzmán, Thanasis Lianeas, Evdokia Nikolova, and Marc Schröder. Network pricing: How to induce optimal flows under strategic link operators. *Operations Research*, 2021.

[60] Ankush Das, Shankara Narayanan Krishna, Lakshmi Manasa, Ashutosh Trivedi, and Dominik Wojtczak. On pure Nash equilibria in stochastic games. In *International Conference on Theory and Applications of Models of Computation*, pages 359–371. Springer, 2015.

[61] Constantinos Daskalakis. On the complexity of approximating a Nash equilibrium. *ACM Transactions on Algorithms (TALG)*, 9(3):1–35, 2013.

[62] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1): 195–259, 2009.

[63] Constantinos Daskalakis, Dylan J. Foster, and Noah Golowich. Independent policy gradient methods for competitive reinforcement learning. In *Advances in Neural Information Processing Systems*, 2020.

[64] Federico Della Croce, Ulrich Pferschy, and Rosario Scatamacchia. On approximating the incremental Knapsack problem. *Discrete Applied Mathematics*, 2019.

[65] Persi Diaconis. The mathematics of mixing things up. *Journal of Statistical Physics*, 144(3):445–458, 2011.

[66] Maman Abdurachman Djauhari, Norhaidah Mohd Asrah, Lee Siaw Li, and Ismail Djakaria. Forecasting model of electricity consumption in malaysia: A geometric Brownian motion approach. *Solid State Technology*, 63(3):40–46, 2020.

[67] Roland L. Dobrushin. Central limit theorem for nonstationary Markov chains. I. *Theory of Probability & Its Applications*, 1(1):65–80, 1956.

[68] Ehsan Emamjomeh-Zadeh, Chen-Yu Wei, Haipeng Luo, and David Kempe. Adversarial online learning with changing action sets: Efficient algorithms with approximate regret bounds. In *Algorithmic Learning Theory*, pages 599–618. PMLR, 2021.

[69] Environmental Protection Agency. Sources of greenhouse gas emissions, April 28, 2023. URL `https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions`. Accessed on April 30, 2023.

[70] Leah Epstein and Asaf Levin. Bin packing with general cost structures. *Mathematical Programming*, 132(1):355–391, 2012.

[71] Eyal Even-Dar and Yishay Mansour. Fast convergence of selfish rerouting. In *SODA*, volume 5, pages 772–781. Citeseer, 2005.

[72] Eyal Even-Dar, Yishay Mansour, and Peter Bartlett. Learning rates for Q-learning. *Journal of Machine Learning Research*, 5(1), 2003.

[73] Bruce H. Faaland. The multiperiod Knapsack problem. *Operations Research*, 29(3): 612–616, 1981.

[74] Alex Fabrikant, Christos Papadimitriou, and Kunal Talwar. The complexity of pure Nash equilibria. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 604–612, 2004.

[75] Alex Fabrikant, Aaron D. Jaggard, and Michael Schapira. On the structure of weakly acyclic games. In *International Symposium on Algorithmic Game Theory*, pages 126–137. Springer, 2010.

[76] Francisco Facchinei and Jong-Shi Pang. 12 Nash equilibria: the variational approach. *Convex optimization in signal processing and communications*, page 443, 2010.

[77] Francisco Facchinei, Veronica Piccialli, and Marco Sciandrone. Decomposition algorithms for generalized potential games. *Computational Optimization and Applications*, 50(2):237–262, 2011.

[78] Yuri Faenza and Igor Malinovic. A ptas for the time-invariant incremental Knapsack problem. In *International Symposium on Combinatorial Optimization*, pages 157–169. Springer, 2018.

[79] Yuri Faenza, Danny Segev, and Lingyi Zhang. Approximation algorithms for the generalized incremental Knapsack problem. *arXiv preprint arXiv:2009.07248*, 2020.

[80] Michal Feldman, Yuval Snappir, and Tami Tamir. The efficiency of best-response dynamics. In *International Symposium on Algorithmic Game Theory*, pages 186–198. Springer, 2017.

[81] FERC. Participation of distributed energy resource aggregations in markets operated by Regional Transmission Organizations and Independent System Operators. Order No. 2222, 2020.

[82] Arlington M. Fink. Equilibrium in a stochastic n-person game. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)*, 28(1):89–93, 1964.

[83] J. Leroy Folks and Raj S. Chhikara. The inverse Gaussian distribution and its statistical application—a review. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3):263–275, 1978.

[84] Peter Frazier, David Kempe, Jon Kleinberg, and Robert Kleinberg. Incentivizing exploration. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 5–22, 2014.

[85] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[86] Drew Fudenberg, Fudenberg Drew, and David K. Levine. *The Theory of Learning in Games*, volume 2. MIT press, 1998.

[87] Deqiang Gan and Eugene Litvinov. Energy and reserve market designs with explicit consideration to lost opportunity costs. *IEEE Transactions on Power Systems*, 18(1): 53–59, 2003.

[88] Zuguang Gao, Khaled Alshehri, and John R. Birge. Aggregating distributed energy resources: efficiency and market power. *Available at SSRN 3931052*, 2021.

[89] Zuguang Gao, Khaled Alshehri, and John R. Birge. On efficient aggregation of distributed energy resources. In *Proc. 60th IEEE Conference on Decision and Control (CDC)*, pages 7064–7069. IEEE, 2021.

[90] Zuguang Gao, John R. Birge, and Varun Gupta. Approximation schemes for multiperiod binary Knapsack problems. In *Computer Science–Theory and Applications: 16th International Computer Science Symposium in Russia, CSR 2021, Sochi, Russia, June 28–July 2, 2021, Proceedings 16*, pages 131–146. Springer, 2021.

[91] Zuguang Gao, Qianqian Ma, Tamer Başar, and John R. Birge. Finite-sample analysis of decentralized Q-learning for stochastic games. *arXiv preprint arXiv:2112.07859*, 2021.

[92] Zuguang Gao, Khaled Alshehri, and John R. Birge. Quantification of market power mitigation via efficient aggregation of distributed energy resources. In *Proc. 61st IEEE Conference on Decision and Control (CDC)*, pages 4406–4411. IEEE, 2022.

[93] Zuguang Gao, John R. Birge, Richard Li-Yang Chen, and Maurice Cheung. Greedy algorithms for the freight consolidation problem. In *22nd Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[94] Zuguang Gao, Qianqian Ma, Tamer Başar, and John R. Birge. Sample complexity of decentralized tabular Q-learning for stochastic games. In *2023 American Control Conference (ACC)*. IEEE, 2023.

[95] Jugal Garg, Ruta Mehta, Vijay V. Vazirani, and Sadra Yazdanbod. ETR-completeness for decision versions of multi-player (symmetric) Nash equilibria. *ACM Transactions on Economics and Computation (TEAC)*, 6(1):1–23, 2018.

[96] Stéphane Gaubert and Zheng Qu. Dobrushin's ergodicity coefficient for Markov operators on cones. *Integral Equations and Operator Theory*, 81(1):127–150, 2015.

[97] Monica Greer. *Electricity marginal cost pricing: applications in eliciting demand responses*. Elsevier, 2012.

[98] Justin Gundlach and Romany Webb. Distributed energy resource participation in wholesale markets: Lessons from the California ISO. *Energy Law Journal*, 39(1):47–77, 2018.

[99] Olle Häggström et al. *Finite Markov Chains and Algorithmic Applications*, volume 52. Cambridge University Press, 2002.

[100] Liyang Han, Jalal Kazempour, and Pierre Pinson. Monetizing customer load data for an energy retailer: A cooperative game approach. In *2021 IEEE Madrid PowerTech*, pages 1–6. IEEE, 2021.

[101] Thomas Dueholm Hansen, Peter Bro Miltersen, and Uri Zwick. Strategy iteration is strongly polynomial for 2-player turn-based stochastic games with a constant discount factor. *Journal of the ACM*, 60(1):1, 2013.

[102] Tobias Harks and Max Klimm. On the existence of pure Nash equilibria in weighted congestion games. *Mathematics of Operations Research*, 37(3):419–436, 2012.

[103] Jeff Hartline and Alexa Sharp. An incremental model for combinatorial maximization problems. In Carme Àlvarez and María Serna, editors, *Experimental Algorithms*, pages 36–48, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-34598-5.

[104] Juris Hartmanis. Computers and intractability: A guide to the theory of NP-completeness (michael r. garey and david s. johnson). *SIAM Review*, 24(1):90–91, 1982.

[105] Yu-Chi Ho. Team decision theory and information structures. *Proceedings of the IEEE*, 68(6):644–654, 1980.

[106] Mikael Holter. Amazon surges ahead on renewable energy despite gap with rivals. *Time*, 2021. URL `https://time.com/6213666/amazon-renewable-energy/`. Accessed on April 30, 2023.

[107] Ed Hopkins. A note on best response dynamics. *Games and Economic Behavior*, 29 (1-2):138–150, 1999.

[108] Junling Hu and Michael P. Wellman. Nash Q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, 4(Nov):1039–1069, 2003.

[109] Shanshan Hu, Gilvan C. Souza, Mark E. Ferguson, and Wenbin Wang. Capacity investment in renewable energy technology with supply intermittency: Data granularity matters! *Manufacturing & Service Operations Management*, 17(4):480–494, 2015.

[110] Shanshan Hu, Gilvan C. Souza, Mark E. Ferguson, and Wenbin Wang. Capacity investment in renewable energy technology with supply intermittency: Data granularity matters! *Manufacturing & Service Operations Management*, 17(4):480–494, 2015.

[111] Herbie Huang, Nur Sunar, Jayashankar M. Swaminathan, and Rahul Roy. Do noisy customer reviews discourage platform sellers? Empirical analysis of an online solar marketplace. *Manufacturing & Service Operations Management* (to appear), 2020.

[112] Shaojun Huang, Qiuwei Wu, Shmuel S. Oren, Ruoyang Li, and Zhaoxi Liu. Distribution locational marginal pricing through quadratic programming for congestion management in distribution networks. *IEEE Transactions on Power Systems*, 30(4): 2170–2178, July 2015.

[113] Oscar H. Ibarra and Chul E. Kim. Fast approximation algorithms for the Knapsack and sum of subset problems. *Journal of the ACM (JACM)*, 22(4):463–468, 1975.

[114] Samuel Ieong, Robert McGrew, Eugene Nudelman, Yoav Shoham, and Qixiang Sun. Fast and compact: A simple class of congestion games. In *AAAI*, volume 5, pages 489–494, 2005.

[115] International Energy Agency. Greenhouse gas emissions by sector 2019. https://www.iea.org/data-and-statistics/charts/greenhouse-gas-emissions-by-sector-2019, 2021. Accessed on April 30, 2023.

[116] International Energy Agency. World energy investment 2022. Technical report, OECD/IEA, Paris, France, 2022. URL https://www.iea.org/reports/world-energy-investment-2022. Accessed on April 16, 2023.

[117] International Renewable Energy Agency. Renewable energy and jobs annual review 2022. Technical report, 2022. URL https://www.irena.org/publications/2022/Mar/Renewable-Energy-and-Jobs-Annual-Review-2022.

[118] Mohit Jain, Vijander Singh, and Asha Rani. A novel nature-inspired algorithm for optimization: Squirrel search algorithm. *Swarm and Evolutionary Computation*, 44: 148–175, 2019.

[119] Klaus Jansen. A fast approximation scheme for the multiple Knapsack problem. In Mária Bieliková, Gerhard Friedrich, Georg Gottlob, Stefan Katzenbeisser, and György Turán, editors, *SOFSEM 2012: Theory and Practice of Computer Science*, pages 313–324, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-27660-6.

[120] Zeyu Jia, Lin F. Yang, and Mengdi Wang. Feature-based Q-learning for two-player stochastic games. *arXiv preprint arXiv:1906.00423*, 2019.

[121] Ce Jin. An Improved FPTAS for 0-1 Knapsack. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 76:1–76:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-109-2. doi:10.4230/LIPIcs.ICALP.2019.76.

[122] Chi Jin, Qinghua Liu, Yuanhao Wang, and Tiancheng Yu. V-learning–a simple, efficient, decentralized algorithm for multiagent RL. *arXiv preprint arXiv:2110.14555*, 2021.

[123] David S. Johnson, Alan Demers, Jeffrey D. Ullman, Michael R. Garey, and Ronald L. Graham. Worst-case performance bounds for simple one-dimensional packing algorithms. *SIAM Journal on Computing*, 3(4):299–325, 1974.

[124] Christian Kaps, Simone Marinesi, and Serguei Netessine. When should the off-grid sun shine at night? optimum renewable generation and energy storage investments. *Optimum Renewable Generation and Energy Storage Investments.(January 5, 2022)*, 2022.

[125] Hans Kellerer and Ulrich Pferschy. A new fully polynomial time approximation scheme for the Knapsack problem. *Journal of Combinatorial Optimization*, 3(1):59–71, 1999.

[126] Hans Kellerer and Ulrich Pferschy. Improved dynamic programming in connection with an fptas for the Knapsack problem. *Journal of Combinatorial Optimization*, 8(1): 5–11, 2004.

[127] Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack Problems*. Springer, 2004.

[128] David Kempe, Sixie Yu, and Yevgeniy Vorobeychik. Inducing equilibria in networked public goods games through network structure modification. *arXiv preprint arXiv:2002.10627*, 2020.

[129] Kia Khezeli, Weixuan Lin, and Eilyan Bitar. Learning to buy (and sell) demand response. *IFAC-PapersOnLine*, 50(1):6761–6767, 2017. ISSN 2405-8963. doi:https://doi.org/10.1016/j.ifacol.2017.08.1193. URL `https://www.sciencedirect.com/science/article/pii/S2405896317316920`. 20th IFAC World Congress.

[130] A. Gürhan Kök, Kevin Shang, and Şafak Yücel. Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science*, 64(1): 131–148, 2018.

[131] A Gürhan Kök, Kevin Shang, and Şafak Yücel. Impact of electricity pricing policies on renewable energy investments and carbon emissions. *Management Science*, 64(1): 131–148, 2018.

[132] A Gürhan Kök, Kevin Shang, and Şafak Yücel. Investments in renewable and conventional energy: The role of operational flexibility. *Manufacturing & Service Operations Management*, 22(5):925–941, 2020.

[133] Hoong Chuin Lau and Min Kwang Lim. Multi-period multi-dimensional Knapsack problem and its application to available-to-promise. 2004.

[134] Michael Lavillotti. DER market design: Aggregations. Technical report, New York ISO, 2018.

[135] Eugene L. Lawler. Fast approximation algorithms for Knapsack problems. *Mathematics of Operations Research*, 4(4):339–356, 1979.

[136] Donghwan Lee and Niao He. A unified switching system perspective and ODE analysis of Q-learning algorithms. *arXiv preprint arXiv:1912.02270*, 2019.

[137] Jinlong Lei and Uday V. Shanbhag. Distributed variable sample-size gradient-response and best-response schemes for stochastic Nash equilibrium problems over graphs. *arXiv preprint arXiv:1811.11246*, 2018.

[138] Jinlong Lei and Uday V. Shanbhag. Asynchronous schemes for stochastic and misspecified potential games and nonconvex optimization. *Operations Research*, 68(6): 1742–1766, 2020.

[139] Jinlong Lei, Uday V. Shanbhag, Jong-Shi Pang, and Suvrajeet Sen. On synchronous, asynchronous, and randomized best-response schemes for stochastic Nash games. *Mathematics of Operations Research*, 45(1):157–190, 2020.

[140] David S. Leslie and Edmund J. Collins. Individual Q-learning in normal form games. *SIAM Journal on Control and Optimization*, 44(2):495–514, 2005.

[141] David S. Leslie, Steven Perkins, and Zibo Xu. Best-response dynamics in zero-sum stochastic games. *Journal of Economic Theory*, 189:105095, 2020.

[142] Gen Li, Yuting Wei, Yuejie Chi, Yuantao Gu, and Yuxin Chen. Sample complexity of asynchronous Q-learning: Sharper analysis and variance reduction. In *Advances in Neural Information Processing Systems*, volume 33, pages 7031–7043, 2020.

[143] Jianming Lian, Di Wu, Karanjit Kalsi, and Hua Chen. Theoretical framework for integrating distributed energy resources into distribution systems. In *2017 IEEE Power Energy Society General Meeting*, pages 1–5, July 2017.

[144] Edward Y.H. Lin and M.C. Chen. A dynamic programming approach to the multiple-choice multi-period Knapsack problem and the recursive APL2 code. *Journal of Information and Optimization Sciences*, 31(2):289–303, 2010.

[145] Edward Y.H. Lin and Chung-Min Wu. The multiple-choice multi-period Knapsack problem. *Journal of the Operational Research Society*, 55(2):187–197, 2004.

[146] Yiheng Lin, Guannan Qu, Longbo Huang, and Adam Wierman. Multi-agent reinforcement learning in stochastic networked systems. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[147] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine Learning Proceedings*, pages 157–163. Elsevier, 1994.

[148] Michael L. Littman. Friend-or-foe Q-learning in general-sum games. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 322–328. Morgan Kaufmann Publishers Inc., 2001.

[149] Eugene Litvinov, Tongxin Zheng, Gary Rosenwald, and Payman Shamsollahi. Marginal loss modeling in LMP calculation. *IEEE Transactions on Power Systems*, 19(2):880–888, 2004.

[150] Qinghua Liu, Tiancheng Yu, Yu Bai, and Chi Jin. A sharp analysis of model-based reinforcement learning with self-play. In *International Conference on Machine Learning*, pages 7001–7010. PMLR, 2021.

[151] Lars Ljungqvist and Thomas J. Sargent. *Recursive macroeconomic theory*. MIT press, 2018.

[152] Fariba Farajbakhsh Mamaghani and Metin Çakanyıldırım. Harvesting solar power foments prices in a vicious cycle: Breaking the cycle with price mechanisms. *Metin, Harvesting Solar Power Foments Prices in a Vicious Cycle: Breaking the Cycle with Price Mechanisms*, 2021.

[153] Saeed D. Manshadi and Mohammad E. Khodayar. A hierarchical electricity market structure for the smart grid paradigm. *IEEE Transactions on Smart Grid*, 7(4):1866–1875, July 2016.

[154] Weichao Mao and Tamer Başar. Provably efficient reinforcement learning in decentralized general-sum Markov games. *arXiv preprint arXiv:2110.05682*, 2021.

[155] Rahul R. Marathe and Sarah M. Ryan. On the validity of the geometric Brownian motion assumption. *The Engineering Economist*, 50(2):159–192, 2005.

[156] Eric Maskin and Jean Tirole. A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs. *Econometrica: Journal of the Econometric Society*, pages 549–569, 1988.

[157] Eric Maskin and Jean Tirole. Markov perfect equilibrium: I. Observable actions. *Journal of Economic Theory*, 100(2):191–219, 2001.

[158] George B. Mathews. On the partition of numbers. *Proceedings of the London Mathematical Society*, 1(1):486–490, 1896.

[159] Francisco S. Melo, Sean P. Meyn, and M. Isabel Ribeiro. An analysis of reinforcement learning with function approximation. In *Proceedings of the 25th International Conference on Machine Learning*, pages 664–671, 2008.

[160] Igal Milchtaich. Congestion games with player-specific payoff functions. *Games and Economic Behavior*, 13(1):111–124, 1996.

[161] Seyedali Mirjalili and Andrew Lewis. The whale optimization algorithm. *Advances in Engineering Software*, 95:51–67, 2016.

[162] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

[163] Eduardo Moreno, Daniel Espinoza, and Marcos Goycoolea. Large-scale multi-period precedence constrained Knapsack problem: a mining application. *Electronic notes in discrete mathematics*, 36:407–414, 2010.

[164] Fabio Moret and Pierre Pinson. Energy collectives: A community and fairness based approach to future electricity markets. *IEEE Transactions on Power Systems*, 34(5): 3994–4004, 2018.

[165] Chanaleä Munien and Absalom E. Ezugwu. Metaheuristic algorithms for one-dimensional bin-packing problems: A survey of recent advances and applications. *Journal of Intelligent Systems*, 30(1):636–663, 2021.

[166] John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[167] Thành Nguyen and Karthik Kannan. Welfare implications in intermediary networks. *Information Systems Research*, 32(2):378–393, 2021.

[168] Elli Ntakou and Michael Caramanis. Price discovery in dynamic power markets with low-voltage distribution-network participants. In *2014 IEEE PES T&D Conference and Exposition*, pages 1–5, April 2014.

[169] Ibrahim H. Osman. Heuristics for the generalised assignment problem: simulated annealing and tabu search approaches. *Operations-Research-Spektrum*, 17(4):211–225, 1995.

[170] Jong-Shi Pang, Suvrajeet Sen, and Uday V. Shanbhag. Two-stage non-cooperative games with risk-averse players. *Mathematical Programming*, 165(1):235–290, 2017.

[171] Yael Parag and Benjamin K. Sovacool. Electricity market design for the prosumer era. *Nature energy*, 1(4):1–6, 2016.

[172] Daniel Paulin et al. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20, 2015.

[173] Xiaoshan Peng, Owen Q. Wu, and Gilvan Souza. Renewable, flexible, and storage capacities: Friends or foes? *Available at SSRN 3983678*, 2021.

[174] Heikki Peura and Derek W. Bunn. Renewable power and electricity prices: The impact of forward markets. *Management Science*, 67(8):4772–4788, 2021.

[175] PV Tech. Corporations purchased record clean power in 2022: BloombergNEF. April 2023. URL `https://www.pv-tech.org/corporations-purchased-record-clean-power-in-2022-bloombergnef`. Accessed on April 30, 2023.

[176] Guannan Qu and Adam Wierman. Finite-time analysis of asynchronous stochastic approximation and Q-learning. In *Conference on Learning Theory*, pages 3185–3205. PMLR, 2020.

[177] Guannan Qu, Adam Wierman, and Na Li. Scalable reinforcement learning of localized policies for multi-agent networked systems. In *Learning for Dynamics and Control*, pages 256–266. PMLR, 2020.

[178] Farrokh A. Rahimi and Ali Ipakchi. Transactive energy techniques: Closing the gap between wholesale and retail markets. *The Electricity Journal*, 25(8):29 – 35, 2012. ISSN 1040-6190.

[179] R. Randeniya. *Multiple-choice Multi-period Knapsack Problem (MCMKP): Application and Solution Approach*. University of New Brunswick, Faculty of Administration, 1994.

[180] Donguk Rhee. *Faster fully polynomial approximation schemes for Knapsack problems*. PhD thesis, Massachusetts Institute of Technology, 2015.

[181] Wansoo T. Rhee and Michel Talagrand. Online bin packing with items of random size. *Mathematics of Operations Research*, 18(2):438–445, 1993.

[182] Ingmar Ritzenhofen, John R. Birge, and Stefan Spinler. The structural impact of renewable portfolio standards and feed-in tariffs on electricity markets. *European Journal of Operational Research*, 255(1):224–242, 2016.

[183] Robert W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[184] Navid Azizan Ruhi, Krishnamurthy Dvijotham, Niangjun Chen, and Adam Wierman. Opportunities for price manipulation by aggregators in electricity markets. *IEEE Trans. Smart Grid*, 9(6):5687–5698, 2018.

[185] Mehran Samavati, Daryl Essam, Micah Nehring, and Ruhul Sarker. A methodology for the large-scale multi-period precedence-constrained Knapsack problem: an application in the mining industry. *International Journal of Production Economics*, 193:12–20, 2017.

[186] Muhammed Sayin, Kaiqing Zhang, David Leslie, Tamer Başar, and Asuman Ozdaglar. Decentralized Q-learning in zero-sum markov games. *Advances in Neural Information Processing Systems*, 34:18320–18334, 2021.

[187] Muhammed O. Sayin and Onur Unlu. Logit-Q learning in Markov games. *arXiv preprint arXiv:2205.13266*, 2022.

[188] Erwin Schrödinger. Zur theorie der fall-und steigversuche an teilchen mit brownscher bewegung. *Physikalische Zeitschrift*, 16:289–295, 1915.

[189] Nicola Secomandi and Sunder Kekre. Optimal energy procurement in spot and forward markets. *Manufacturing & Service Operations Management*, 16(2):270–282, 2014.

[190] Sandip Sen, Mahendra Sekaran, John Hale, et al. Learning to coordinate without sharing information. In *AAAI*, volume 94, pages 426–431, 1994.

[191] Lloyd S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.

[192] David B. Shmoys and Éva Tardos. An approximation algorithm for the generalized assignment problem. *Mathematical programming*, 62(1-3):461–474, 1993.

[193] Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, 171(7):365–377, 2007.

[194] Aaron Sidford, Mengdi Wang, Lin Yang, and Yinyu Ye. Solving discounted stochastic two-player games with near-optimal time and sample complexity. In *International Conference on Artificial Intelligence and Statistics*, pages 2992–3002. PMLR, 2020.

[195] Siddharth Prakash Singh and Alan Scheller-Wolf. That's not fair: Tariff structures for electric utilities with rooftop solar. *Manufacturing & Service Operations Management*, 2021.

[196] Siddharth Prakash Singh and Alan Scheller-Wolf. That's not fair: Tariff structures for electric utilities with rooftop solar. *Manufacturing & Service Operations Management*, 24(1):40–58, 2022.

[197] Marian Smoluchowski. Notiz über die berechnung der brown'schen molekularbewegung bei der ehrenhaft-millikan'schen versuchsanordnung. *Pisma Mariana Smoluchowskiego*, 1(2):477–485, 1927.

[198] Ziang Song, Song Mei, and Yu Bai. When can we learn general-sum Markov games with a large number of players sample-efficiently? *arXiv preprint arXiv:2110.04184*, 2021.

[199] Paul M. Sotkiewicz and Jesus M. Vignolo. Nodal pricing for distribution networks: efficient pricing for efficiency enhancing DG. *IEEE Transactions on Power Systems*, 21(2):1013–1014, May 2006.

[200] Matt Stern and John R. Birge. Dynamic learning in strategic pricing games. *Available at SSRN 3579123*, 2020.

[201] Peter Stone and Manuela Veloso. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8(3):345–383, 2000.

[202] Tao Sun, Lang Tong, and Donghan Feng. On the dynamics of distributed energy adoption: equilibrium, stability, and limiting capacity. *IEEE Trans. Automatic Control*, 65 (1):102–114, 2020.

[203] Nur Sunar and John R. Birge. Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. *Management Science*, 65(2):714–734, 2019.

[204] Nur Sunar and John R. Birge. Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. *Management Science*, 65(2):714–734, 2019.

[205] Nur Sunar and Jayashankar M. Swaminathan. Net-metered distributed renewable energy: A peril for utilities? *Management Science*, 67(11):6716–6733, 2021.

[206] Nur Sunar and Jayashankar M. Swaminathan. Socially relevant and inclusive operations management. *Production and Operations Management*, 2022.

[207] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT press, 2018.

[208] Brian Swenson, Ryan Murray, and Soummya Kar. On best-response dynamics in potential games. *SIAM Journal on Control and Optimization*, 56(4):2734–2767, 2018.

[209] Vasilis Syrgkanis. The complexity of equilibria in cost sharing games. In *International Workshop on Internet and Network Economics*, pages 366–377. Springer, 2010.

[210] Ming Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *International Conference on Machine Learning*, pages 330–337, 1993.

[211] J. Terra, R. Ferreira, C. Borges, and M. Carvalho. Using distribution-level locational marginal pricing to value distributed generation: Impacts on revenues captured by generation agents. In *2017 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, pages 1–6, Sep. 2017.

[212] Yi Tian, Yuanhao Wang, Tiancheng Yu, and Suvrit Sra. Provably efficient online agnostic learning in Markov games. *arXiv preprint arXiv:2010.15020*, 2020.

[213] Alessio Trivella, Danial Mohseni-Taheri, and Selvaprabu Nadarajah. Meeting corporate renewable power targets. *Management science*, 69(1):491–512, 2023.

[214] John N. Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16(3):185–202, 1994.

[215] United Nations. Transforming our world: the 2030 agenda for sustainable development. `https://sdgs.un.org/2030agenda`, 2015. Accessed on April 30, 2023.

[216] Vijay V. Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013.

[217] Martin J. Wainwright. Stochastic approximation with cone-contractive operators: Sharp ℓ_∞-bounds for Q-learning. *arXiv preprint arXiv:1905.06265*, 2019.

[218] Christopher J.C.H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4): 279–292, 1992.

[219] Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. Online reinforcement learning in stochastic games. In *Advances in Neural Information Processing Systems*, pages 4987–4997, 2017.

[220] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. *arXiv preprint arXiv:2006.09517*, 2020.

[221] Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Last-iterate convergence of decentralized optimistic gradient descent/ascent in infinite-horizon competitive Markov games. *arXiv preprint arXiv:2102.04540*, 2021.

[222] Owen Q. Wu and Volodymyr Babich. Unit-contingent power purchase agreement and asymmetric information about plant outage. *Manufacturing & Service Operations Management*, 14(2):245–261, 2012.

[223] Owen Q. Wu and Roman Kapuscinski. Curtailing intermittent generation in electrical systems. *Manufacturing & Service Operations Management*, 15(4):578–595, 2013.

[224] Yutong Wu, Ali Khodabakhsh, Bo Li, Evdokia Nikolova, and Emmanouil Pountourakis. Eliciting information with partial signals in repeated games. *arXiv preprint arXiv:2109.04343*, 2021.

[225] Qiaomin Xie, Yudong Chen, Zhaoran Wang, and Zhuoran Yang. Learning zero-sum simultaneous-move Markov games using function approximation and correlated equilibrium. In *Conference on Learning Theory*, pages 3674–3682. PMLR, 2020.

[226] Xin-She Yang and Suash Deb. Cuckoo search via Lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pages 210–214. IEEE, 2009.

[227] Bora Yongacoglu, Gurdal Arslan, and Serdar Yüksel. Decentralized learning for optimality in stochastic dynamic teams and games with local control and global state information. *IEEE Transactions on Automatic Control*, pages 1–1, 2021. doi:10.1109/TAC.2021.3121228.

[228] Bora Yongacoglu, Gurdal Arslan, and Serdar Yüksel. Satisficing paths and independent multi-agent reinforcement learning in stochastic games. *arXiv preprint arXiv:2110.04638*, 2021.

[229] H. Peyton Young. *Strategic Learning and its Limits*. OUP Oxford, 2004.

[230] Serdar Yüksel and Tamer Başar. *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. Springer Science & Business Media, 2013.

[231] Kaiqing Zhang, Sham M Kakade, Tamer Başar, and Lin F. Yang. Model-based multi-agent RL in zero-sum Markov games with near-optimal sample complexity. *arXiv preprint arXiv:2007.07461*, 2020.

[232] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. In *Handbook of Reinforcement Learning and Control*, pages 321–384. Springer, 2021.

[233] Feng Zhao, Tongxin Zheng, and Eugene Litvinov. Constructing demand curves in forward capacity market. *IEEE Trans. Power Systems*, 33(1):525–535, 2018.

[234] Tongxin Zheng and Eugene Litvinov. Ex post pricing in the co-optimized energy and reserve market. *IEEE Trans. Power Systems*, 21(4):1528–1538, 2006.

[235] Yangfang (Helen) Zhou, Alan Scheller-Wolf, Nicola Secomandi, and Stephen Smith. Managing wind-based electricity generation in the presence of storage and transmission capacity. *Production and Operations Management*, 28(4):970–989, 2019.