Title

Robust Decision Making and Mechanism Design for Algorithmic Markets and Platforms

ISBN

9798189270932

Author

Kudva, Sukanya

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Robust Decision Making and Mechanism Design for Algorithmic Markets and Platforms

By

Sukanya Kudva

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Anil Aswani, Chair
Professor Shankar Sastry
Professor Rajan Udwani

Spring 2026

Robust Decision Making and Mechanism Design for Algorithmic Markets and Platforms

Abstract

Robust Decision Making and Mechanism Design for Algorithmic Markets and Platforms

by

Sukanya Kudva

Doctor of Philosophy in Engineering - Industrial Engineering and Operations Research

University of California, Berkeley

Professor Anil Aswani, Chair

As artificial intelligence and autonomous agents become primary actors in modern economic systems, the dynamics of market interactions have shifted from human-led deliberation to high-frequency, data-driven algorithmic competition. This dissertation investigates the theoretical foundations and practical design of robust algorithmic ecosystems, organized around three core pillars: robust mechanism design, reliable decision-making algorithms, and strategic data governance.

Part I addresses the vulnerability of classical auctions to strategic manipulation. We introduce the VCG-Posted Price (V-PoP) mechanism, which uses a collusion detection oracle to maintain incentive compatibility and efficiency in the presence of bid manipulation through coordination. We establish welfare and revenue guarantees, demonstrating how side information can be leveraged to improve on classical mechanisms.

Part II focuses on the challenge of building computationally efficient and reliable decision-making algorithms in complex environments. We first propose an exact iterative algorithm for continuous trilevel optimization, a framework essential for modeling hierarchical interactions like defender-attacker-defender games. By exploiting the parametric structure of KKT conditions, we observe finite-time convergence to the optima in our numerical experiments. Additionally, we develop a tensor completion algorithm using a gauge norm over the polytope of rank-1 tensors. This approach achieves the information-theoretic sampling rate while maintaining computational scalability, which we validate using numerical experiments on large-scale energy storage data.

Part III explores the role of strategic players in data sharing and governance. We analyze the economic viability of data dividends, identifying the regulatory conditions under which platforms are incentivized to compensate users for their data and invest in privacy protection. Finally, we examine partial coordination among electric vehicle (EV) charging stations, providing analytical conditions where coalition formation may paradoxically harm participants

or social welfare.

Together, these contributions build on techniques from mechanism design and game theory, optimization, and statistics to provide a framework for designing algorithmic systems that are efficient, equitable, and resilient to the strategic complexities of the modern digital economy.

To Amma, for her strength, love and sacrifice.

Contents

List of Figures

Acknowledgments

This thesis would not have been possible without the support of so many people; it truly took a village. I am grateful to Anil, whose mentorship has shaped my research and the way I approach ideas. Through hundreds of hours of brainstorming and discussion, he provided clarity and structure to my thoughts. It has been an honor to work with him. Shankar adopted me as an unofficial mentee, and the weekly group meetings have been a constant source of inspiration for new ideas. I am thankful for every discussion with him and his group that led to new directions and collaborations in my research. Rajan has always provided thoughtful feedback, and I appreciate the time and energy he has dedicated as a committee member. My collaborators, whose contributions are reflected in this thesis, have also made the journey more enjoyable. I could not have begun my PhD without my undergraduate mentors, Mallikarjuna Rao, Jayendran Venkateswaran, and Veeraruna Kavitha, who first introduced me to this field and sparked my curiosity.

Over the years, I have been fortunate to build a strong support system at Berkeley. My friends in the department, who shared similar experiences, have been my go-to for advice and comfort. The Indian community helped make Berkeley feel like home during my initial years. Friends from my apartment complex have been like family throughout my PhD, and my girlfriends, who have been my emotional anchors and co-work partners, have been invaluable in helping me make the final push over the past year.

I owe a special debt to my family, who were very supportive throughout this process. Amma has always been my role model, teaching me early on to be independent and accountable for my decisions. Aanu has always believed in the power of learning and the pursuit of knowledge. My little brother, Raghu, who is everything I am not, yet somehow contributes to half of who I am. My partner, Aman, has shared my curiosity and love for scientific research, and has been a constant source of strength and a reminder of all the beauty life holds.

Chapter 1

Introduction

1.1 Motivation

Automation and artificial intelligence (AI) agents are increasingly embedded in the operational fabric of modern markets and platforms. These systems are qualitatively different from those of a few decades ago, when decisions unfolded slowly and human judgment dominated system behavior. Today's environments are fast-paced, heavily data-driven, and continuously adaptive. Technology has evolved from a passive computational instrument to an *active participant*, that is, an autonomous decision-maker with its own learning dynamics, strategic objectives, and capacity to influence system-level outcomes.

AI agents now observe information, update beliefs, and choose actions to advance their own goals, often in competition or coordination with other agents. The resulting equilibria may diverge significantly from socially optimal outcomes, yielding inefficiencies or welfare-reducing externalities. In this context, mechanism design serves not only as a tool for aligning incentives but also as a framework for shaping the collective dynamics of these algorithmic ecosystems. As these systems adapt in real time, *robust mechanism design* aims to preserve desirable outcomes even when strategic interactions give rise to unforeseen or undesirable equilibria.

A defining feature of this paradigm shift is that failures or inefficiencies may arise *endogenously*, stemming from interactions among rational, adaptive agents rather than from exogenous shocks or human error. Understanding equilibrium behavior is therefore critical for developing regulatory and algorithmic interventions that improve welfare without impeding innovation or efficiency. At the same time, the growing autonomy and complexity of these systems demand *robust decision-making algorithms* that can operate effectively in uncertain, dynamic, and strategic environments. Such algorithms must reason about their informational limitations, anticipate the responses of other learning agents, and maintain performance guarantees even when faced with adversarial or misspecified conditions.

Another central dimension of these algorithmic markets is *data*. Data now functions as a key economic asset that determines the extent of personalization, innovation, and competitive advantage. The control and accessibility of data shape both the incentives and outcomes of AI-driven decision processes. Hence, questions surrounding data governance, sharing, and privacy must be analyzed through a strategic lens: who possesses what information, how it can be used, and how data externalities, such as privacy breaches or market concentration, affect the broader system.

Viewed through this lens, this dissertation explores three overarching research themes:

1. *Robust Mechanism Design:* How can we design mechanisms that remain robust and guarantee desirable outcomes under adaptive strategic interactions?
2. *Reliable Decision-making Algorithms:* What decision-making algorithms can perform reliably in environments populated by autonomous, strategic agents?
3. *Strategic Data Governance and Sharing:* How do data access, sharing, and governance affect incentives, competition, and welfare in algorithmic markets?

Through exploring questions under these themes, this work both borrows from and contributes to the methodological foundations of mechanism design and game theory, statistics, and optimization, aiming to bridge theoretical insights with the design of practically robust algorithmic systems.

1.2 Contributions and Outline

This dissertation is organized into three parts, each addressing one of the overarching research themes identified above.

Part I: Robust Mechanism Design. Chapter 2 studies collusion-proof auction design in the setting of a single-item, multi-unit auction where a subset of bidders may collude to manipulate their bids. Classical mechanisms such as the Vickrey–Clarke–Groves (VCG) mechanism achieve efficiency and dominant strategy incentive compatibility (DSIC) under truthful behavior, but are known to be vulnerable to bid-shading collusion. We first establish a Bulow–Klemperer-type result showing that recruiting additional honest bidders improves welfare even in the presence of colluders. We then propose the VCG-Posted Price (V-PoP) mechanism, which uses a black-box collusion detection oracle as side information to partition bidders into non-colluding and colluding subgroups and applies VCG and posted-price mechanisms, respectively, to each group. We prove that V-PoP is ex-post DSIC for all bidders including colluders, design an Item Split Oracle to optimally allocate items between the two subgroups, and establish probabilistic welfare and revenue guarantees. We further analyze the robustness of V-PoP to bidder misclassification and validate our results through numerical experiments across multiple valuation distributions. This chapter is based on joint work with Edward Dowling and Anil Aswani [103].

Part II: Reliable Decision-making Algorithms. Chapter 3 addresses the problem of solving a continuous trilevel optimization program, which is a hierarchical decision-making problem in which three agents make sequential decisions with coupled objectives and constraints. Such formulations capture Stackelberg-like games with three levels and arise naturally in defender–attacker–defender models, supply chains, and machine learning pipelines. We focus on trilevel programs with strongly convex quadratic objectives and linear constraints at all levels. Existing methods either focus on mixed-integer or robust linear models via branch-and-cut and decomposition, or use gradient-based methods for nonlinear cases. In contrast, we propose an exact iterative algorithm within a master–subproblem framework that exploits the parametric structure of the KKT conditions at the lower levels to derive local affine response mappings and their polytopic validity regions. These are progressively incorporated into a relaxed master problem, yielding a non-decreasing sequence of lower bounds; an upper bound is maintained by evaluating the followers' reactions at the current leader decision. We postulate finite time convergence to an optimal solution, establishing that the total number of distinct active-set regions (and hence the number of constraints generated) is finite. This chapter is based on joint work with Anil Aswani.

Chapter 4 develops a novel algorithm for tensor completion, which is the recovery of a low-rank tensor from partial, possibly noisy observations. As noted in prior works, there is an inherent trade-off between the computational complexity and the information-theoretic sample rate. We propose an algorithm that achieves the information-theoretic rate and, through numerical experiments, show that it scales well computationally as well. The key idea is to formulate the tensor completion problem as a convex optimization using a gauge norm defined over the polytope of rank-1 tensors with ± 1 entries. We prove that the gauge norm serves as a convex surrogate for tensor rank and derive generalization bounds matching the information-theoretic rate. The resulting convex program is solved via Blended Conditional Gradients (BCG) with an integer programming oracle for the weak separation subproblem, augmented by a heuristic to accelerate oracle calls. We demonstrate the algorithm's performance on synthetic tensors of varying order and size, and on a large-scale energy storage dataset from Ireland, where tensor completion is used to construct nonparametric predictive models scalable to grid-level deployments. This chapter is based on joint work with Xin Chen, Yongzheng Dai, Anil Aswani, and Chen Chen [37].

Part III: Strategic Data Governance and Sharing. Chapter 5 investigates the economic incentives governing data dividend payments by online platforms. We model the interaction between a platform and its users as a Stackelberg game, where the platform first sets dividend levels and data protection investment, and users subsequently choose between high and low levels of data sharing. We characterize the platform's optimal dividend policy as a function of users' privacy valuations, breach risk and loss, and the platform's revenue. Our analysis identifies when voluntary data dividends arise and when they do not. Major platforms operating under typical market conditions are unlikely to pay dividends voluntarily, but mandatory dividend requirements combined with breach liability can jointly incentivize platforms to invest in data protection and provide equitable compensation to users. This chapter is based on joint work with Anil Aswani [102].

Chapter 6 examines the game-theoretic consequences of partial coordination among electric vehicle (EV) charging stations. Stations are modeled as strategic agents in a linear, demand-sensitive price market, and we study a coalition–Nash equilibrium in which a coalition of stations minimizes its aggregate cost while the remaining stations act as individual Nash players. Contrary to the common intuition that coordination is universally beneficial, we establish analytical conditions under which the formation of a coalition is harmful either to coalition members, to non-coalition stations, or to society at large. We show that these outcomes depend critically on the composition of the coalition, not merely its size, and provide a characterization of harm in terms of the stations' demand and sensitivity. Our results are illustrated through numerical experiments, offering design guidance for regulators and operators of EV charging infrastructure. This chapter is based on joint work with Kshitij Kulkarni, Chinmay Maheshwari, Anil Aswani, and Shankar Sastry [104].

Part I

Robust Mechanism Design

Chapter 2

Collusion-Proof Auction Design Using Side Information

2.1 Introduction

We consider a single-item, multiple-unit auction, where r identical copies of the same item are being sold. There are m single-minded bidders, each desiring at most one item, and bidder i has a private valuation v_i for the item and is charged payment p_i on winning it. Each bidder i aims to maximize the quasi-linear utility $v_i \mathbb{1}\,(\text{bidder } i \text{ won}) - p_i$. The goal of the auctioneer is to design an *efficient* auction, with a payment scheme and allocation rule that maximizes the social welfare $\sum_i v_i \mathbb{1}\,(\text{bidder } i \text{ won})$. If some of the bidders *collude* to exchange information and monetary payments, they can collectively manipulate their bids and harm social welfare. The colluders can manipulate their bids in two ways: (1) *bid shade* to underquote their value and reduce the overall auction price, (2) *bid load* to overquote their value and win more items. There is limited prior work on collusion-proof auction design [77, 128, 154], which is the main question addressed in this chapter. Unlike prior work, we propose using side information about bidders' collusion to design better auctions.

In the absence of collusion, when the bidders' valuations are private, the well-known Vickrey-Clarke-Groves (VCG) mechanism [162, 78] can be used to ensure *dominant strategy incentive compatibility* (DSIC) for all bidders, that is truth-telling is the best strategy for the bidders regardless of what others report. This mechanism also maximizes the auctioneer's revenue among all welfare-maximizing auctions [101]. However, the VCG mechanism is known to be vulnerable to collusion unless the outcome is in the core [136, 46]. Moreover, it is also known that truthful, collusion-proof mechanisms have to be posted-price mechanisms, that is give bidders a fixed take-it-or-leave-it price [77]. Unfortunately, posted price mechanisms cannot approximate expected efficiency to even a constant factor and it is impossible to design an efficient, truthful, and collusion-proof mechanism [77], which is another reason why collusion is challenging to deal with. Given this, we attempt to address the following question:

> **Q:** How can we design an auction that achieves reasonably good performance in both welfare and revenue, even in the presence of bidder collusion, when compared against VCG outcomes in the no-collusion setting?

We model the colluding bidders to maximize the sum of their net utilities, and answer the above question in two parts: (1) we characterize colluding bids in a VCG auction and prove a Bulow-Klemperer[32] type result – that additional bidders improve welfare, even if colluding, (2) we propose a combination of collusion detection and auction design to achieve truthful, collusion-proof mechanisms that achieve probabilistic guarantees on expected welfare and revenue. The main idea is that once the colluding bidders are identified, we can implement VCG mechanism on the non-colluding bidders followed by posted price mechanism on the colluding bidders to achieve truthfulness and optimize welfare. We also discuss how to split the items between the two subgroups of bidders, and how to ensure truthfulness is preserved.

Our work falls under a line of literature referred to as mechanism design with side information or learning-augmented mechanism design. The idea is to incorporate external or expert information about strategic agents to improve the performance of classical mechanisms. In our setting, knowledge of which bidders are colluding serves as side information and enables the design of truthful mechanisms that outperform posted-price mechanisms, which are the only truthful and collusion-proof mechanisms in the absence of side information [77].

Currently, conclusions drawn from collusion detection algorithms are insufficient for legal incrimination, and instead some evidence of explicit communication between colluding parties is required [81]. Given this, we argue that our approach is a step towards developing collusion-resistant mechanisms, which can be useful as a preventative measure.

Main Technical Results

To get an intuition for the impact of collusion, consider the following example.

Example 2.1.1. *Consider 5 bidders with private valuations* $\{v_1 = 1, v_2 = 70, v_3 = 101, v_4 = 102, v_5 = 103\}$*, and* $r = 3$ *items. The VCG mechanism allocates items to bidders 3,4,5 at a price of 70 each, and social welfare is* $\sum_i v_i \mathbb{1}\,(\textit{bidder } i \textit{ won}) = (101 + 102 + 103) = 306$*. If bidders 3 and 5 collude, they could bid shade and quote 0, 103, respectively. In this case, naively implementing VCG allocates items to bidders 2,4,5 at a lower price of 1, and social welfare decreases to* $(70 + 102 + 103) = 275$*. The colluders benefit from bid shading as their net utility increases from* $(v_5 - p_5) + (v_3 - p_3) = (103 - 70) + (101 - 70) = 64$ *to* $(103 - 1) + 0 = 102$*.*

In the above example, the non-colluding bidders are allotted an extra item and also benefit from the price drop. The auctioneer, however, suffers a huge loss in revenue. In fact, this is always the case with the VCG mechanism, and the negative impact of collusion is limited to the auctioneer[143]. The observations can be summarized in the following lemma.

Lemma 2.1.2 (Informal)**.** *Consider the VCG mechanism for the single-item, multiple-unit auction with single-minded bidders. If the colluders aim to maximize their net additive utility, the following holds:*

1. *The colluding bidders never benefit from bid loading, and always bid shade. As a consequence, they never take more items but they do lead to a drop in the VCG auction price.*

2. *Neither the utility of each non-colluding bidder nor the net utility of colluding bidders decreases. However, the auctioneer's utility (i.e., their revenue) is non-increasing.*

3. *Due to collusion, the social welfare and revenue are also non-increasing, and the drop in welfare is entirely attributed to a decrease in the auctioneer's utility.*

It may seem counterintutitive that the colluding bidders do not bid load to take more items. The reason for this is that bid loading leads to a hike in the VCG price, and since the colluders do not have complementary effects in their net utility, they cannot make up for the loss from price hike. Instead, they bid shade, give up items if necessary, and cut down on the price to maximize their utility. This creates a win-win situation for both the colluding and non-colluding bidders while only hurting the auctioneer. These results extend to an auction with different types of items and where each bidders' valuation of a subset of items is complement-free [143].

Using the above characterization for colluding bidders' strategy, we prove a Bulow-Klemperer type result for VCG auctions with colluding bidders, as stated below. Similar to the original Bulow-Klemperer result [32], the main implication in our result is that recruiting more honest bidders improves welfare when compared to the best, welfare-maximizing, and collusion-proof auction mechanism.

Theorem 2.1.3 (Informal)**.** *Assume the bidders' true valuations come from an i.i.d distribution. In a multi-unit auction of a single good, when there are colluding bidders maximizing their net additive utility, the following holds:*

$$\mathbb{E}\big[\,\mathit{Welf}_{VCG}(\textit{Non-colluding} \cup \textit{Colluding} \cup \textit{New})\big] \geqslant \mathbb{E}\big[\,\mathit{Welf}_{OPT}(\textit{Non-colluding} \cup \textit{Colluding})\big]$$

where New denotes the set of newly introduced honest bidders, whose cardinality is at least that of the colluding bidders.

Despite the guarantees provided by the above theorem, colluding bidders may still deviate from truthful bidding in the VCG mechanism. To address this, we propose a hybrid mechanism that partitions bidders into non-colluding and colluding bidders, and applies two mechanisms, VCG and posted price, respectively. We refer to this mechanism as VCG-Posted Price (V-PoP). We design an Item Split Oracle to optimize the expected welfare (or a good proxy thereof) and calculate the optimal split of items berween the two sub-groups of bidders. We show counterexamples to illustrate how naive approaches do not preserve truthfulness and emphasise a careful design of the oracle. We show that V-PoP is ex-post dominant strategy incentive compatible (DSIC) for all bidders, even in the presence of collusion.

We further establish guarantees on the expected welfare and revenue under the following assumptions: (1) bidders' valuations are independent and identically distributed (i.i.d.), and (2) a black-box collusion detection algorithm is available to identify colluding bidders. Our main result is stated below.

Theorem 2.1.4 (Informal). *Let the bidders' valuations be i.i.d. random variables with cumulative distribution function $F(\cdot)$. When $F(\cdot)$ is known, the expected welfare of the proposed mechanism can be optimized such that*

$$\mathbb{E}[\mathit{Welf}_{\text{V-PoP}}(\textit{Non-colluding} \cup \textit{Colluding})] \geqslant \mathbb{E}[\mathit{Welf}_{\text{VCG}}(\textit{Non-colluding})]\,.$$

A similar guarantee holds for the expected revenue, i.e.,

$$\mathbb{E}[\mathit{Rev}_{\text{V-PoP}}(\textit{Non-colluding} \cup \textit{Colluding})] \geqslant \mathbb{E}[\mathit{Rev}_{\text{VCG}}(\textit{Non-colluding})]\,,$$

with a probability that depends on how the items are divided between the non-colluding and colluding bidders.

If, instead, $F(\cdot)$ is unknown but invertible, and its inverse (the quantile function) $Q(\cdot)$ satisfies

$$Lx \leqslant Q(x) \quad \text{for some known constant } L > 0,$$

we design two mechanisms that respectively optimize minorants of the expected welfare and expected revenue. We further show that the corresponding welfare and revenue guarantees hold probabilistically. Numerical simulations demonstrate that these mechanisms perform well empirically. Note that since the expected revenue serves as a lower bound on the expected welfare, it can be interpreted as a minorant of the welfare objective.

Finally, we conduct numerical experiments using different distributions for bidders' valuations to compare four mechanisms: (1) VCG assuming both non-colluding and colluding bidders are truthful, (2) VCG applied only to the non-colluding bidders, (3) VCG in the presence of collusion and bid shading (i.e., untruthful VCG), and (4) our proposed V-PoP mechanism. In our setup, we observe that V-PoP consistently outperforms VCG applied only to the non-colluding bidders, both in terms of average welfare and average revenue. As the number of non-colluding bidders increases, the performance of V-PoP approaches that of the best baseline, the VCG mechanism with all bidders assumed to be truthful.

Since V-PoP enforces truthfulness among colluding bidders through a posted-price mechanism and does not guarantee that all items are sold, some welfare and revenue loss is inevitable. However, when the group of colluding bidders is sufficiently large, no items remain unsold under V-PoP, and it outperforms all other mechanisms except the idealized VCG with fully truthful bidders, a theoretical upper bound that no mechanism can surpass.

Related Work

Mechanism design with side information. In contrast to our work that considers the knowledge of colluding bidders as side information, most prior work considers side information about bidder types, i.e., parameters related to preferences or item valuations [50, 14, 167]. For instance, information about correlated bidder types has been used to identify bidders generating the least welfare (i.e., the weakest types) and to improve the performance of the VCG mechanism [14, 167]. Devanur et al. [50] study revenue-maximizing auctions where bidders lie on a continuum and can be distinguished using side information. Learning-augmented algorithms have also been used to design strategy-proof mechanisms in application-specific domains such as job scheduling [15] and facility location [2].

Prior-independent mechanism design. The VCG mechanism, widely adopted for welfare maximization, does not require prior information about the distribution of bidders' valuations [78, 41]. In contrast, the Myerson auction which is optimal for revenue maximization, does rely on such priors. Subsequent work on prior-independent auctions has achieved constant-factor approximation guarantees to the optimal expected revenue and analyzed sample complexity requirements for heterogeneous bidders and specific distribution classes [51, 42]. The truthful auctions proposed in this chapter offer similar constant-approximation guarantees for welfare and revenue. While we do not assume complete knowledge of the valuation distributions, we require certain parameters that bound their quantiles.

Another foundational result in this domain is the Bulow–Klemperer theorem, which shows that welfare maximization via a prior-independent VCG with additional bidders can achieve nearly optimal revenue [32, 83, 55]. In a similar spirit, we establish a Bulow–Klemperer type result in the presence of colluding bidders, showing that in a (non-truthful) VCG mechanism, adding more bidders improves total welfare, even when some bidders collude.

Collusion-resistant mechanisms. Most auction mechanisms, including VCG, are vulnerable to collusion [117, 11]. The notion of collusion-proofness, where no subset of bidders can benefit through communication or side payments, is quite strong and restricts mechanisms to posted-price formats, which cannot optimize non-trivial objectives without prior information about the bids [77]. As a result, weaker notions of collusion-resistance, such as t-truthfulness and group strategy-proofness, are often considered. Under t-truthfulness, where any coalition of size at most t maximizes its total expected utility by bidding truthfully, approximately efficient and optimal mechanisms have been proposed using consensus techniques [77]. An even weaker notion, group strategy-proofness, disallows monetary transfers among colluding bidders and requires that any deviation benefiting one bidder must harm another [92, 21]. Mechanisms satisfying this property have been characterized and studied in various settings, including network design games [128, 38, 154] and queueing models [119, 95]. In contrast, this chapter introduces a new notion of collusion-resistant mechanism design: Given knowledge of which bidders are colluding, how can we guarantee that they bid truthfully?

Collusion detection. We assume access to a black-box collusion detection algorithm that classifies bidders into colluding and non-colluding groups. Collusion detection, particularly in public procurement auctions, has been extensively studied using statistical techniques [36, 44, 115]. A comparative study of machine learning approaches for collusion detection is presented in [71], while a game-theoretic approach based on mutual information is proposed in [28]. In the more recent context of tacit algorithmic collusion [7, 34], a variety of methods have been developed, including dynamic testing [81] and statistical testing [82].

2.2 Preliminaries

We consider a set of all bidders M with $|\mathsf{M}| = m$. We assume M is partitioned into two disjoint subsets: the non-colluding bidders N and the colluding bidders C, such that $\mathsf{M} = \mathsf{N} \cup \mathsf{C}$ and $m = |\mathsf{M}| = |\mathsf{N}| + |\mathsf{C}| = n + c$.

Let $b = (b_1, \cdots, b_{\mathrm{M}})$ and $v = (v_1, \cdots, v_{\mathrm{M}})$ be the bid and true valuation vectors, respectively, where b_i is the bid received from bidder i. Designing an auction mechanism comprises two components: (1) an allocation scheme $x_i(b)$, and (2) a payment scheme $p_i(b)$ for each bidder i. These are determined before the bids are collected.

Definition 2.2.1 (Utility of Bidders and Auctioneer)**.** *We assume that each bidder i is rational and submits bid b_i to maximize their utility $u_i(b_i, b_{i-}) = v_i x_i(b_i, b_{i-}) - p_i(b_i, b_{i-})$, where b_{i-} are the bids of all other bidders except i. The auctioneer receives utility from the total payments collected, defined as $u_a(b) = \sum_{i \in \mathsf{M}} p_i(b)$.*

Given this framework, we define welfare and revenue. In the literature, auctions that maximize welfare and revenue are referred to as *efficient* and *optimal* auctions, respectively. We focus on efficiency while maximizing revenue to the extent possible.

Definition 2.2.2 (Welfare and Revenue)**.** *Welfare is the total utility of all participants (all bidders and the auctioneer). Hence, welfare is $Welf(b) = u_a(b) + \sum_{i \in \mathsf{M}} u_i(b) = \sum_{i \in \mathsf{M}} v_i x_i(b)$. Revenue is the total payment made to the auctioneer: $Rev(b) = u_a(b) = \sum_{i \in \mathsf{M}} p_i(b)$.*

We now introduce notation based on order statistics to allow for clear reference to ranked bids and valuations. For any set of bidders S, let b_k^{S} and v_k^{S} be the k^{th}-largest bid and true valuations within S, respectively. Define $\mathsf{B}_r^{\mathrm{S}} = \{b_1^{\mathrm{S}}, \cdots, b_r^{\mathrm{S}}\}$ and $\mathsf{V}_r^{\mathrm{S}} = \{v_1^{\mathrm{S}}, \cdots, v_r^{\mathrm{S}}\}$ to be the sets of the top r bids and true valuations from S, respectively. The complements of these sets are defined as $\overline{\mathsf{B}_r^{\mathrm{S}}} = \mathsf{B}_s^{\mathrm{S}} \backslash \mathsf{B}_r^{\mathrm{S}} = \{b_{r+1}^{\mathrm{S}}, \cdots, b_n^{\mathrm{S}}\}$ and similarly, $\overline{\mathsf{V}_r^{\mathrm{S}}} = \mathsf{V}_s^{\mathrm{S}} \backslash \mathsf{V}_r^{\mathrm{S}} = \{v_{r+1}^{\mathrm{S}}, \cdots, v_n^{\mathrm{S}}\}$.

To differentiate observed bidders' valuations v from their underlying random variables, we use capital letters V to denote the latter. Also, $\mathsf{V}_r^{\mathrm{S}}$ and $\mathcal{V}_r^{\mathrm{S}}$ denote sets of observed and random bidders' valuations. Similarly, notations for observed and realized bids and their sets are b, $\mathsf{B}_r^{\mathrm{S}}$ and B, $\mathcal{B}_r^{\mathrm{S}}$, respectively.

Definition 2.2.3 (VCG Mechanism)**.** *In an auction with r identical items, the VCG mechanism prescribes: (1) allocation $x_i(b) = \mathbf{1}\left(b_i \in \mathsf{B}_r^{\mathrm{M}}\right)$, (2) payment $p_i(b) = b_{r+1}^{\mathrm{M}} \mathbf{1}\left(b_i \in \mathsf{B}_r^{\mathrm{M}}\right)$ for each bidder i. In the absence of collusion, the mechanism ensures (ex-post) DSIC by guaranteeing each winning bidder i an information rent of $v_i - p_i(b)$, which incentivizes truth-telling*

$(b = v)$. This property, in turn, guarantees efficiency since the auction maximizes the welfare $\sum_i v_i x_i(b)$ and the items are given to bidders who value them the most.

Model for Colluding Bidders

Let b^{N} and b^{C} be the bids submitted by the non-colluding and colluding bidders, respectively. True valuations v^{N} and v^{C} are defined similarly.

Definition 2.2.4 (Utility of Colluding Bidders)**.** *We assume that the colluding bidders act as a unified entity to maximize the sum of their individual utilities, allowing for both information exchange and monetary transfers amongst them. The joint utility of the colluding bidders is: $u_{\mathrm{C}}(b^{\mathrm{C}}, b^{\mathrm{N}}) = \sum_{i \in \mathsf{C}} v_i x_i(b^{\mathrm{C}}, b^{\mathrm{N}}) - p_i(b^{\mathrm{C}}, b^{\mathrm{N}})$.*

2.3 VCG in the Presence of Collusion

We characterize the effects of colluding bidders on welfare and revenue when the VCG auction is conducted without accounting for collusion. Define r_{N} and r_{C} as the number of items allotted to non-colluding (N) and colluding (C) bidders, respectively, such that $r_{\mathrm{N}} + r_{\mathrm{C}} = r$ (total items).

The phenomena of bid shading (bidding a value lower than the true item valuation) and shill bidding (a single bidder entering the auction as multiple bidders) have been studied in literature [143]. Their main observation is that substitute items incentivize integration and bid shading (as in our setup), while complement items incentivize disintegration (shill bidding) and possibly higher bids. In our setting without complements, the losing colluding bidders optimally shade bids to zero, effectively behaving as a single non-competitive bidder. In the next lemma, we state these results for our setting and give a simpler proof tailored to this special case for completeness.

Lemma 2.3.1 (VCG Equilibrium in Presence of Collusion)**.** *Let $r_{\mathrm{C}}^{Col}(r_{\mathrm{C}}^*)$, $r_{\mathrm{N}}^{Col}(r_{\mathrm{N}}^*)$ and $b^{Col}(b^*)$ be the number of items allocated to colluding bidders, items allocated to non-colluding bidders, and the optimal bids in the presence (absence) of collusion, respectively, using the VCG mechanism. Then, the following holds:*

1. *Colluding bidders do not obtain more items through collusion, i.e., $r_{\mathrm{C}}^{Col} \leqslant r_{\mathrm{C}}^*$. Furthermore, colluding bidders always shade their bids and never bid load, i.e., $b_i^{\mathrm{C},Col} \leqslant v_i^{\mathrm{C}}$, $\forall i \in C$. In fact, they either bid 0 or their true valuation v_i^{C}.*

2. *The utility of each non-colluding bidder and of the collective of colluding bidders does not decrease due to collusion. That is, $u_i(b^*) \leqslant u_i(b^{Col})$ $\forall i \in N$ and $u_{\mathrm{C}}(b^*) \leqslant u_{\mathrm{C}}(b^{Col})$. However, the revenue of the auctioneer does not increase, and could instead decrease. That is, $u_a(b^*) \geqslant u_a(b^{Col})$.*

3. *Welfare and revenue are both non-increasing as a consequence of collusion. That is,* $Welf(b^*) \geqslant Welf(b^{Col})$ *and* $Rev(b^*) \geqslant Rev(b^{Col})$.

Corollary 2.3.2. *The fact that colluding bidders only bid shade and do not bid load leads to an interesting observation that additional bidders improve welfare and revenue, even if colluding. That is,* $Welf_{VCG}(\mathsf{N} \cup \mathsf{C}) \geqslant Welf_{VCG}(\mathsf{N})$ *and* $Rev_{VCG}(\mathsf{N} \cup \mathsf{C}) \geqslant Rev_{VCG}(\mathsf{N})$.

We use the above observations to prove a Bulow-Klemperer type result.

Theorem 2.3.3 (BK Theorem for Collusion)**.** *Say, the true valuations of the bidders come from an i.i.d distribution. In a multi-unit auction of identical goods, when there are colluding bidders maximizing their net additive utility, the following holds:*

$$\mathbb{E}\big[Welf_{VCG}(\mathsf{N} \cup \mathsf{C} \cup \widetilde{\mathsf{N}})\big] \geqslant \mathbb{E}\big[Welf_{OPT}(\mathsf{N} \cup \mathsf{C})\big]$$

where $Welf_{VCG}(\mathsf{S})$ *and* $Welf_{OPT}(\mathsf{S})$ *denote the welfare achieved using a VCG mechanism and a collusion-proof efficient mechanism on bidders in set* S*, some of who could be colluding and reporting untruthful bids. Importantly,* $\widetilde{\mathsf{N}}$ *denote a set of new non-colluding bidders such that* $|\widetilde{\mathsf{N}}| = |\mathsf{C}|$.

The key idea behind the Bulow-Klemperer theorem in literature [32] is that the effort of finding optimal auctions is better spent on recruiting an additional bidder. Our theorem has a similar interpretation: recruiting $|\mathsf{C}|$ more honest bidders improves both welfare and revenue when compared to the best, welfare-maximizing, and collusion-proof auction mechanism.

2.4 Truthful Auction in the Presence of Collusion

Since colluding bidders are untruthful in the VCG mechanism, we henceforth focus on designing mechanisms that remain truthful in the presence of collusion. To this end, we introduce a mechanism, which we refer to as VCG- Posted Price (V-PoP) mechanism. This mechanism partitions the bidders into two sub-groups- of non-colluding and colluding bidders- and implements the VCG mechanism and the posted price on these subgroups, respectively. The choice of posted price for colluding bidders is because it is the only collusion-proof and truthful mechanism [77]. To ensure truthfulness among all bidders, it is crucial to carefully design how allocated items are split between the two subgroups; the choice of VCG or posted price mechanisms alone is insufficient. A naive implementation may incentivize both colluding and non-colluding bidders to overbid to increase the likelihood that more items are allocated to their subgroup, thereby improving their own chances of winning (see Remark 2.4.1 for concrete example). The mechanism is outlined in Algorithm 1.

Notice that our posted price is set using the VCG price from non-colluding bidders, and hence all winning bidders are charged the same price. Next, we design the Item Split Oracle to ensure truthfulness.

Algorithm 1 VCG-Posted Price (V-PoP) Mechanism

Input: Total number of items r; non-colluding bidders N and colluding bidders C with sealed/unrevealed bids b^{N} and b^{C}; Item Split Oracle (b^{N}, n, c)
Output: Allocation x_i and payment p_i for each bidder $i \in \mathsf{N} \cup \mathsf{C}$.
Calculate $k^* \leftarrow$Item Split Oracle (b^{N}, n, c)
Phase 1: Run VCG on N with k^* items
Winning non-colluding bidders: $\mathsf{W}_{\mathrm{N}} \leftarrow \{i \in \mathsf{N} \mid b_i \in \mathsf{B}^{\mathrm{N}}_{k^*}\}$.
Phase 2: Run posted price on C with $r - k^*$ items
Qualifying colluding bidders: $\hat{\mathsf{W}}_{\mathrm{C}} \leftarrow \{i \in \mathsf{C} \mid b^{\mathrm{C}}_i > b^{\mathrm{N}}_{k^*+1}\}$.
if $|\hat{\mathsf{W}}_{\mathrm{C}}| > r - k^*$ **then**
    Winning colluding bidders: randomly select a subset $\mathsf{W}_{\mathrm{C}} \subseteq \hat{\mathsf{W}}_{\mathrm{C}}$ of size $r - k^*$.
else
    $\mathsf{W}_{\mathrm{C}} \leftarrow \hat{\mathsf{W}}_{\mathrm{C}}$.
end if
Winners: $\mathsf{W} \leftarrow \mathsf{W}_{\mathrm{N}} \cup \mathsf{W}_{\mathrm{C}}$
Allocate items: $x_i \leftarrow 1\ \forall i \in \mathsf{W}$, 0 otherwise.
Charge price: $p_i \leftarrow b^{\mathrm{N}}_{k^*+1}\ \forall i \in \mathsf{W}$.

Algorithm 2 Item Split Oracle

Input: Total number of items r; number of non-colluding bidders and colluding bidders n, c respectively; non-colluding bids b^{N}; function $M(k, \overline{\mathsf{B}^{\mathrm{N}}_k})$.
Output: Optimal split of items k^*, where k^* and $r - k^*$ items are to be allocated to non-colluding and colluding bidders respectively.
$k^* \leftarrow \arg\max_{k \in \{0, \cdots, r\}} \mathbb{E}\left[M(k, \overline{\mathcal{B}^{\mathrm{N}}_k})\right]$

Remark 2.4.1. *With an example, we show that a naive implementation of either the Item Split Oracle(b^{N}, n, c) or the final allocation of items in Algorithm 1 incentivizes untruthful bidding. Say k^* is computed using the Item Split Oracle, which in turn does not use the bids b^{N}, so the oracle itself cannot be manipulated through untruthful bids. If, after Phases 1 and 2, the r items are allocated by randomly selecting winners from $\mathsf{W}_{\mathrm{N}} \cup \hat{\mathsf{W}}_{\mathrm{C}}$ (instead of W as in Algorithm 1), the mechanism is no longer truthful. Colluders could strategically bid the maximum allowed value to increase their expected allocation without affecting the price. For example, let $r = 2$, non-colluding bids be $\{1, 2\}$, and colluding valuations be $\{1, 2, 3\}$. If the price is $p = 1$, then colluders bid $\{3, 3, 3\}$ to strictly increase their expected number of allocated items when items are randomly distributed among those with bids above p.*

In the below theorem, we claim that our mechanism is truthful for all bidders when the Item Split Oracle in Algorithm 2 is used. The intuition is that not using the realized non-colluding bids in the oracle is an obvious way to ensure truthfulness.

Theorem 2.4.2 (V-PoP Truthfulness)**.** *The mechanism in Algorithm 1 is (ex-post) DSIC for all bidders, including colluding bidders* C*.*

Let $\mathrm{Welf}_{\text{V-PoP}}(\mathsf{N} \cup \mathsf{C}; k)$ and $\mathrm{Rev}_{\text{V-PoP}}(\mathsf{N} \cup \mathsf{C}; k)$ be the welfare and revenue, respectively, from the V-PoP mechanism when k items are allocated to non-colluding bidders. In the rest of the chapter, we focus on optimizing V-PoP to achieve good welfare and revenue guarantees.

Exact Expected Welfare Optimization

The key idea of our mechanism is to set the value of k^* using expected welfare (or an appropriate proxy thereof) as $M(k, \overline{\mathsf{B}^{\mathrm{N}}_k})$ in the Item Split Oracle to achieve strong welfare guarantees. We use prior information about the distribution of bidders' valuations and some of the realized bids b^{N} to calculate this function. In the next lemma, we derive a useful expression for the expected welfare conditioned on the price. We use the following assumption, which is required to compute the expected welfare.

Assumption 2.4.1. *Let the true valuations of all the bidders be i.i.d. and come from a distribution with cumulative density function (CDF)* $\mathsf{F}(\cdot)$*, which is invertible, with the inverse as the quantile function* $\mathsf{Q}(\cdot)$*. Sufficient conditions for invertibility of* $F(\cdot)$ *are monotonicity and continuous differentiability.*

Lemma 2.4.3 ((Conditional) Expected Welfare)**.** *Let Assumption 2.4.1 hold. The expected welfare conditioned on the price from the V-PoP mechanism can be calculated as follows:*

$$\mathbb{E}\Big[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k) \mid \overline{\mathsf{V}^{\mathrm{N}}_k}\Big] = \mathbb{E}\Big[Q(U) \;\Big|\; U \geqslant F(v^{\mathrm{N}}_{k+1})\Big] \cdot \Big(k + \mathbb{E}\Big[G\big(|\hat{\mathsf{W}}_{\mathrm{C}}|, v^{\mathrm{N}}_{k+1}; k\big) \;\Big|\; v^{\mathrm{N}}_{k+1}\Big]\Big),$$

where $G\big(|\hat{\mathsf{W}}_{\mathrm{C}}|, v^{\mathrm{N}}_{k+1}; k\big) = |\hat{\mathsf{W}}_{\mathrm{C}}| 1\left\{|\hat{\mathsf{W}}_{\mathrm{C}}| \leqslant r-k\right\} + \frac{\mathbb{E}\left[\sum_{i=1}^{r-k} Q(U_i^{|\hat{\mathsf{W}}_{\mathrm{C}}|}) | U_{|\hat{\mathsf{W}}_{\mathrm{C}}|}^{|\hat{\mathsf{W}}_{\mathrm{C}}|} \geqslant F(v^{\mathrm{N}}_{k+1})\right]}{\mathbb{E}\left[Q(U) | U \geqslant F(v^{\mathrm{N}}_{k+1})\right]} 1\big\{|\hat{\mathsf{W}}_{\mathrm{C}}| > r-k\big\}$*,* $U \sim \mathrm{Uniform}[0,1]$ *and* $|\hat{\mathsf{W}}_{\mathrm{C}}| \mid v^{\mathrm{N}}_{k+1} \sim \mathrm{Binomial}\big(C,\, 1 - F(v^{\mathrm{N}}_{k+1})\big)$ *is the number of colluding bids above price cut-off* v^{N}_{k+1}*.*

When the CDF $F(\cdot)$ and quantile $Q(\cdot)$ are known, the above lemma gives a way to easily compute the conditional expected welfare numerically. In the next theorem, we derive welfare guarantees when $M(k, \overline{\mathsf{B}^{\mathrm{N}}_k})$ is set to this expected welfare.

Theorem 2.4.4 (Expected Welfare Optimization)**.** *Let Assumption 2.4.1 hold. In the Item Split Oracle, set* $M\big(k, \overline{\mathsf{B}^{\mathrm{N}}_k}\big) = \mathbb{E}\Big[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k) \mid \overline{\mathsf{V}^{\mathrm{N}}_k} = \overline{\mathsf{B}^{\mathrm{N}}_k}\Big]$ *as in Lemma 2.4.3. Then, the VCG-Posted Price mechanism (*V-PoP*) has the following welfare guarantees:*

$$\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)\right] \;\geqslant\; \mathbb{E}\left[\mathrm{Welf}_{VCG}(\mathsf{N})\right],$$

where $\mathrm{Welf}_{VCG}(\mathsf{N})$ *is the welfare achieved using the VCG mechanism on non-colluding bidders. Moreover, the following revenue guarantee holds for all the approaches with probability* *at least $P(k^*)$:*

$$\mathbb{E}\left[\mathrm{Rev}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)\right] \geqslant \mathbb{E}\left[\mathrm{Rev}_{VCG}(\mathsf{N})\right],$$

where $P(k^*) = \sum_{q=r-k^*}^{\mathrm{C}} \binom{\mathrm{C}}{q} \frac{B(C+N-k^*-q, q+k^*+1)}{B(N-k^*, k^*+1)}$ with $B(\cdot,\cdot)$ is the beta function and $\mathrm{Rev}_{VCG}(\mathsf{N})$ is the revenue achieved using the VCG mechanism on non-colluding bidders.

By optimally splitting items between the two groups, V-PoP achieves at least the welfare (and often revenue) of running VCG on only non-colluders, leveraging side information for improved efficiency.

Corollary 2.4.5. *Since $\mathbb{E}\left[\mathit{Welf}_{VCG}(\mathsf{N})\right] \leqslant \mathbb{E}\left[\mathit{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)\right] \leqslant \mathbb{E}\left[\mathit{Welf}_{VCG}(\mathsf{N} \cup \mathsf{C})\right]$. For finite number of items r and colluding bidders $|\mathsf{C}|$, $\lim_{|\mathsf{N}| \to \infty} \mathbb{E}[\mathit{Welf}_{VCG}(\mathsf{N})] = \mathbb{E}[\mathit{Welf}_{VCG}(\mathsf{N} \cup \mathsf{C})]$ should hold. Therefore,*

$$\lim_{|\mathsf{N}| \to \infty} \mathbb{E}\left[\mathit{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)\right] = \mathbb{E}\left[\mathit{Welf}_{VCG}(\mathsf{N} \cup \mathsf{C})\right].$$

As the number of non-colluding bidders grow large, V-PoP's performance converges to that of the ideal VCG mechanism with all bidders, showing that the hybrid design asymptotically closes the welfare gap due to collusion.

Minorant-based Expected Welfare Optimization

If $F(\cdot)$ and $Q(\cdot)$ are unknown, we could calculate a minorant of the expected welfare with some minimal information about $Q(\cdot)$ and optimize that instead in our mechanism.

Assumption 2.4.2. *[Quantile Function] Let $\exists\ L > 0$ s.t. $Lx \leqslant \mathsf{Q}(x) \quad \forall x \in [0, 1]$.*

The above assumption is satisfied by many distributions. For example, distributions with bounded support starting from $\beta \geqslant 0$ and bounded from above as $f(V) \leqslant \frac{1}{L}\ \forall V$, where $f(\cdot)$ is the probability density function, satisfy it. The next lemma links order statistics of bidders' valuations to those of samples from a uniform distribution, establishing a simple way to bound valuations using the quantile function's upper and lower slopes.

Lemma 2.4.6. *Given that Assumptions 2.4.1 and 2.4.2 hold, an order statistic V_k^{S} satisfies $LU_k^{\mathrm{S}} \leqslant V_k^{\mathrm{S}}$, where U_k^{S} is the k-th largest order statistic among S i.i.d. standard uniform samples. Furthermore, $F(V_k^{\mathrm{N}}) = U_k^{\mathrm{N}}$.*

Next, we construct two different minorants for expected welfare.

Lemma 2.4.7 (Minorant of Welfare)**.** *Let Assumptions 2.4.1 and 2.4.2 hold. Then,*

$$\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k)\right] \geqslant \mathbb{E}\left[\frac{L\left(1+U_{k+1}^{\mathrm{N}}\right)}{2} \cdot r(k, U_{k+1}^{\mathrm{N}})\right],$$

where $r(k, U_{k+1}^{\mathrm{N}}) = k + \sum_{q=0}^{\mathrm{C}} \hat{G}(q, U_{k+1}^{\mathrm{N}}; k) \binom{\mathrm{C}}{q} \left(1 - U_{k+1}^{\mathrm{N}}\right)^q \left(U_{k+1}^{\mathrm{N}}\right)^{\mathrm{C}-q}$ can be interpreted as the effective number of items sold, in terms of welfare, in the mechanism. and $\hat{G}(q, U_{k+1}^{\mathrm{N}}; k) =$

$q\mathbb{1}\{q \leqslant r-k\} + \frac{2(r-k)}{(1+U^{\mathrm{N}}_{k+1})}\big(1 - \frac{\left(1-U^{\mathrm{N}}_{k+1}\right)(r-k+1)}{2(q+1)}\big)\mathbb{1}\{q > r-k\}$*, and the expectation is over* $U^{\mathrm{N}}_{k+1} \sim$ *Beta* $(N-k, k+1)$.

Since the utility of all bidders and the auctioneer is non-negative, $\mathrm{Welf}_{\text{V-PoP}}(\mathsf{N} \cup \mathsf{C}; k) \geqslant \mathrm{Rev}_{\text{V-PoP}}(\mathsf{N} \cup \mathsf{C}; k)$. Hence, a minorant of the revenue is also a valid minorant of the welfare.

Lemma 2.4.8 (Minorant of Revenue)**.** *Let Assumptions 2.4.1 and 2.4.2 hold. When all bidders are truthful,*

$$\mathbb{E}\left[\mathrm{Rev}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k)\right] \geqslant \mathbb{E}\left[LU^{\mathrm{N}}_{k+1} \cdot \tilde{r}(k, U^{\mathrm{N}}_{k+1})\right],$$

where $\tilde{r}(k, U^{\mathrm{N}}_{k+1}) = k + \sum_{q=0}^{\mathrm{C}} \min\{r-k, q\}\binom{\mathrm{C}}{q}\left(1 - U^{\mathrm{N}}_{k+1}\right)^q \left(U^{\mathrm{N}}_{k+1}\right)^{\mathrm{C}-q}$ *is expected number of items sold in the mechanism and the expectation is over* $U^{\mathrm{N}}_{k+1} \sim$ *Beta* $(N-k, k+1)$.

In the next theorem, we discuss welfare and revenue guarantees for when the welfare and revenue-based minorants are used in the V-PoP mechanism. We show that optimizing over the welfare minorants ensures that the mechanism maintains truthfulness and still achieves a quantifiable welfare and revenue guarantees, even with incomplete distributional knowledge.

Theorem 2.4.9 (Minorant-based Welfare Optimization)**.** *Let Assumptions 2.4.1 and 2.4.2 hold. In the Item Split Oracle, set*

$$\mathbb{E}\big[M(k, \overline{\mathcal{B}^{\mathrm{N}}_k})\big] = \mathbb{E}\big[\tfrac{L\left(1+U^{\mathrm{N}}_{k+1}\right)}{2} \cdot r(k, U^{\mathrm{N}}_{k+1})\big] \quad or \quad \mathbb{E}\big[M(k, \overline{\mathcal{B}^{\mathrm{N}}_k})\big] = \mathbb{E}\big[LU^{\mathrm{N}}_{k+1} \cdot \tilde{r}(k, U^{\mathrm{N}}_{k+1})\big],$$

where $r(k, U^{\mathrm{N}}_{k+1}) = k + \sum_{q=0}^{\mathrm{C}} \hat{G}(q, U^{\mathrm{N}}_{k+1}; k)\binom{\mathrm{C}}{q}\left(1 - U^{\mathrm{N}}_{k+1}\right)^q\left(U^{\mathrm{N}}_{k+1}\right)^{\mathrm{C}-q}$ *with* $\hat{G}(q, U^{\mathrm{N}}_{k+1}; k) = q\mathbb{1}\{q \leqslant r-k\} + \frac{2(r-k)}{(1+U^{\mathrm{N}}_{k+1})}\big(1 - \frac{(1-U^{\mathrm{N}}_{k+1})(r-k+1)}{2(q+1)}\big)\mathbb{1}\{q > r-k\}$ *and* $\tilde{r}(k, U^{\mathrm{N}}_{k+1}) = k + \sum_{q=0}^{\mathrm{C}} \min\{r-k, q\}\binom{\mathrm{C}}{q}\left(1 - U^{\mathrm{N}}_{k+1}\right)^q\left(U^{\mathrm{N}}_{k+1}\right)^{\mathrm{C}-q}$ *is expected number of items sold in the mechanism, and the expectation is over* $U^{\mathrm{N}}_{k+1} \sim$ *Beta* $(N-k, k+1)$*. As shown in Lemmas 2.4.7 and 2.4.8, these choices of* $M(\cdot,\cdot)$ *serve as minorant of expected welfare and revenue, respectively. Under this setting, the VCG-Posted Price mechanism has the following welfare and revenue guarantees with probability at least* $P(k^*)$*:*

$$\begin{aligned}\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)\right] \;&\geqslant\; \mathbb{E}\left[\mathrm{Welf}_{VCG}(\mathsf{N})\right],\\ \mathbb{E}\left[\mathrm{Rev}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)\right] \;&\geqslant\; \mathbb{E}\left[\mathrm{Rev}_{VCG}(\mathsf{N})\right]\end{aligned}$$

where $P(k^*) = \sum_{q=r-k^*}^{\mathrm{C}} \binom{\mathrm{C}}{q}\frac{B(C+N-k^*-q, q+k^*+1)}{B(N-k^*, k^*+1)}$ *with* $B(\cdot,\cdot)$ *is the beta function and* $\mathrm{Welf}_{VCG}(\mathsf{N})$ *and* $\mathrm{Rev}_{VCG}(\mathsf{N})$ *is the welfare and revenue respectively achieved using the VCG mechanism on non-colluding bidders.*

2.5 Numerical Experiments

To evaluate the performance of our proposed mechanism, we conduct numerical simulations and compare it against the standard VCG auction, both with and without bidder collusion. We conduct multi-unit auctions with a total number of items $r = 10$, the number of colluding bidders $C = 10$ and 100, and the number of non-colluding bidders (N) varies from 1 to 50 to observe performance under different percentages of collusion in the market. The bidders' private valuations for an item are drawn independently from a common distribution. We perform tests using different distributions and corresponding quantile function $Q(\cdot)$ and parameter L as listed in the Table 2.1.

Distribution	**Quantile Function,** $Q(x)$	**L**
Uniform	$Q(x) = x$	1
Trapezoidal	$Q(x) = 2 - \sqrt{4 - 3x}$	3/4
Quadratic	$Q(x) = 1 - (1 - x)^{1/3}$	1/3
Sinusoid	$Q(x) = \sin(\pi x/2)$	1

Table 2.1: Bidder Valuation Distributions and Quantile Function Properties

We simulate and compare the outcomes of four different mechanisms:

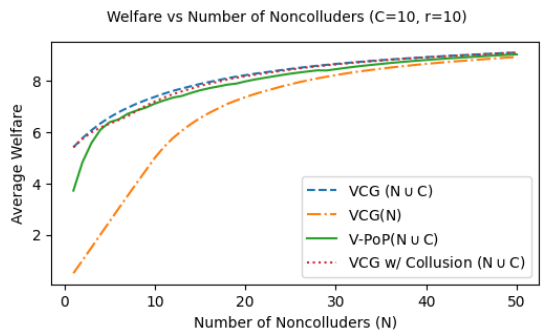
1. *Truthful VCG with All Bidders* (VCG($\mathsf{N} \cup \mathsf{C}$)): It is a standard VCG auction where all $|\mathsf{N} \cup \mathsf{C}|$ bidders are assumed to bid their true valuations.

2. *Truthful VCG with Non-colluders* (VCG(N)): It measures the VCG outcome in an auction with only the N non-colluding bidders.

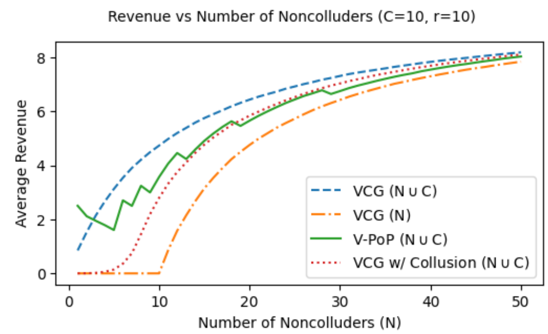
3. *VCG with Collusive Bid-Shaving* (VCG w/ Collusion($\mathsf{N} \cup \mathsf{C}$)): This scenario models a VCG auction where the N non-colluders bid truthfully, but the C colluders strategically "shave" their bids. The colluders calculate and submit bids that maximize their collective utility, based on the non-colluders' bids as derived in Lemma A.0.1.

4. *VCG-Posted Price* (V-PoP ($\mathsf{N} \cup \mathsf{C}$)): This is our proposed mechanism, with a minorant-based optimization using $\mathbb{E}\left[M(k, \overline{\mathsf{B}}_k^{\mathsf{N}})\right] = \mathbb{E}\left[\frac{L\left(1+U_{k+1}^{\mathsf{N}}\right)}{2} \cdot r(k, U_{k+1}^{\mathsf{N}})\right]$. The expectation is evaluated numerically using a sum-based integration method over $U_{k+1}^{\mathsf{N}} \sim$Beta$(N - k, k + 1)$ distribution.

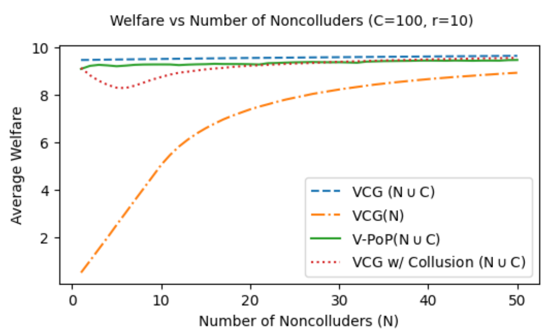
The experiments follow a Monte Carlo simulation approach. For each value of N (the number of non-colluders), we run 1,000 independent repetitions of the mechanisms. In each repetition, new valuations are drawn for all bidders, and the outcomes of the four mechanisms are recorded. The performance of each mechanism is evaluated based on two key metrics, which are averaged over the 1,000 repetitions: (a) total revenue: the total payment received by the seller, and (b) social welfare: the sum of the valuations of the bidders who receive an



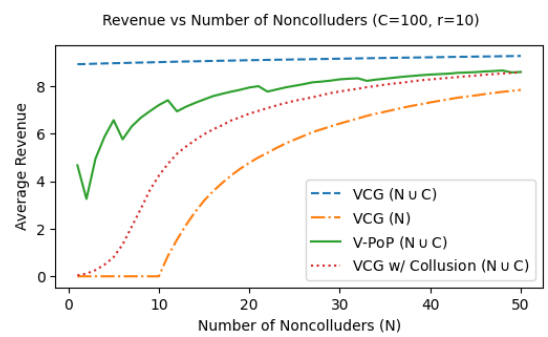
(a)



(b)



(c)



(d)

Figure 2.1: Numerical results for Uniform distribution.

item. We plot how these metrics evolve for each mechanism as the number of non-colluding bidders increases from 1 to 50.

In Figure 2.1, we compare the performance of the mechanisms under the uniform distribution. As established in Corollary 2.4.5, the expected welfare satisfies

$$\mathbb{E}[\mathrm{Welf}_{VCG}(\mathsf{N})] \leqslant \mathbb{E}[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)] \leqslant \mathbb{E}[\mathrm{Welf}_{VCG}(\mathsf{N} \cup \mathsf{C})]\,,$$

and we observe in the average welfare plots that

$$\lim_{|\mathsf{N}| \to \infty} \mathbb{E}[\mathrm{Welf}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k^*)] = \mathbb{E}[\mathrm{Welf}_{VCG}(\mathsf{N} \cup \mathsf{C})]\,.$$

The proposed V-PoP($\mathsf{N} \cup \mathsf{C}$) mechanism consistently outperforms VCG(N) in both average welfare and revenue. Empirically, V-PoP achieves performance well beyond its theoretical guarantees established in Theorem 2.4.9.

As the number of colluding bidders increases from $C = 10$ to $C = 100$, both welfare and revenue improve. Although this may seem counterintuitive, the effect can be explained by the higher probability that all items reserved for colluders are sold, as also supported by Theorem 2.4.9. Similar qualitative behavior is observed under other value distributions–sinusoidal, trapezoidal, and quadratic–the corresponding plots of which are included in the appendix.

2.6 Conclusion

This chapter introduces the VCG-Posted Price (V-PoP) mechanism, a new approach to auction design that remains truthful and efficient even in the presence of bidder collusion. Our results demonstrate that side information, specifically the identification of colluding bidders, can be leveraged to design mechanisms that are simultaneously truthful and collusion-proof. By combining a VCG mechanism for non-colluding bidders with a posted-price mechanism for colluding bidders, V-PoP achieves provable guarantees on expected welfare and revenue.

We also show that colluding bidders optimally shade their bids, never overbidding, and that adding colluding bidders can only improve welfare and revenue relative to the non-collusive baseline, establishing a Bulow–Klemperer type result for collusive environments.

Empirical evaluations further confirm that V-PoP consistently outperforms VCG restricted to non-colluding bidders and approaches the performance of fully truthful VCG outcomes. In follow up work, we generalize the item split oracle to use some of the realized non-colluding bids while still preserving truthfulness [103]. An important direction for future work is to extend these results to settings with heterogeneous items and online settings.

Part II

Reliable Decision-making Algorithms

Chapter 3

Trilevel Optimization Problem

3.1 Introduction

Multilevel optimization problems arise naturally in settings where decisions are taken sequentially over time or are organized in a hierarchy of decision makers [114]. Such models capture strategic interactions between agents with different levels of authority or information, and have been widely adopted in engineering, operations, and economics.

An important application is the defender–attacker–defender (DAD) models for critical infrastructure protection, in which a defender first allocates limited security resources, an attacker subsequently disrupts components, and the defender finally deploys remedial actions to restore service [164, 75, 40]. DAD models have been used to study the resilience of power grids, water distribution systems, and gas networks under targeted attacks. Multilevel structures also appear in decentralized industrial environments, such as supply chains and electricity markets, where upstream and downstream agents optimize coupled decisions. Examples include the coordinated planning of manufacturers, distribution centers, and retailers in supply chains, and hierarchies involving regulators, transmission system operators, and generation companies in electricity markets [97, 63, 113]. More recently, trilevel formulations have emerged in models that rely on upstream prediction for decision making, where one seeks to learn ML models whose outputs are subsequently used in downstream optimization problems [56].

In this chapter, we consider a continuous trilevel problem with strongly convex, quadratic objectives and linear constraints at all three levels. Most trilevel algorithms, which focus on mixed integer or robust linear models and rely on branch-and-cut and decomposition methods [17, 164, 75, 40, 62], or use gradient-based methods for non-linear trilevel problems [142, 140, 110, 91]. In contrast, our method treats a fully continuous problem and develops an exact, iterative method to provide an optimal solution in finite time. The key difficulty lies in handling the nested bilevel structure at levels 2 and 3, and we exploit the parametric structure of the lower levels to derive optimal responses given the corresponding active sets of Karush-Kuhn-Tucker (KKT) conditions. We postulate that the finiteness of active set regions and corresponding distinct constraints to tighten our relaxation problem leads to finite time convergence of our algorithm.

Related Work

Bilevel optimization. The common solution strategy for bilevel problems with convex lower level is KKT reformulation yielding a single level mathematical program with complementarity constraints (MPCC) [48]. While KKT reformulations are exact when the lower level is strongly convex (for instance, with a strictly convex objective and independent linear constraints), they suffer from degenerate constraint structure and find local optimal solutions at lower levels instead of global optima [47, 26]. For mixed integer bilevel programs, various cutting plane approaches have been studied, where the master problem is a high point relaxation that progressively adds valid inequalities derived from the lower level subproblem, yielding monotone lower and upper bound sequences [98, 120]. Our work extends this paradigm from two to three levels, using KKT-derived response mappings for optimal values at lower levels.

Multilevel optimization. Most trilevel optimization algorithms studied in literature consider linear or mixed integer linear models, motivated by sequential decision problems in energy systems, network security, and supply chains [17, 164, 75, 40, 62]. These methods typically employ nested decomposition or generalized Benders decomposition, in which the trilevel problem is successively reduced to bilevel and single level subproblems and cuts are passed back through the hierarchy [27, 13, 74]. Finite convergence in such settings usually rests on the finiteness of discrete first-stage decisions or extreme points in polyhedral feasible regions, and complexity results show that trilevel linear programming is Σ_2^p-hard [90]. In contrast, our work addresses continuous trilevel problems with strongly convex quadratic objectives, where the finiteness of cuts follows from the finite number of active sets arising from the KKT conditions.

Gradient-based methods have also been used for trilevel problems with non-linearity, and these propagate gradients through two levels of nested optimality, and usually provide asymptotic convergence to stationary points [142, 140, 110]. Zeroth-order distributed methods extend this to settings where gradients are unavailable, using polynomial approximations and zeroth-order cuts [91]. A common limitation of these methods is that they target stationary points. Our algorithm, by contrast, is an exact method, albeit for trilevel problems with nested quadratic programs, and that could provide finite-time convergence to an optimal solution.

Parametric programming. A key technical element in our algorithm is the observation that for a strongly convex quadratic program with linear constraints, the optimal primal solution is an affine function of the parameters on each active set region of the KKT constraints, and the number of such regions is finite [158, 131]. Multi-parametric quadratic programming exploits this structure to precompute the piecewise affine solution map over all parameter regions, and has found application in model predictive control [3]. Our algorithm leverages this structure and we solve level 3 subproblem and an upper bound problem of level 2 to identify the active set region at the current iterate, derive the local affine response mapping and its validity polytope, and add these as constraints to a relaxed master problem. A brute force approach that enumerates all possible active set regions using multi-parametric programming has been studied previously [61, 94]; in contrast, we propose an iterative algorithm that tightens lower and upper bounds for convergence.

3.2 Problem Description

We examine a hierarchical scenario involving three utility maximizing players. Our objective is to determine the optimal strategy for the leader, accounting for the rational responses of the subordinates to those initial decisions. The optimal decisions of all participants constitute a generalized equilibrium, extending the standard Stackelberg model, which typically involves two levels, to more levels. We model this scenario using a trilevel optimization problem with quadratic objectives and linear constraints as follows.

$$\begin{aligned}
\min_{x^{(1)},x^{(2)},x^{(3)}} \quad & C_1^\top x + x^\top H_1 x \\
\text{s.t.} \quad & A_1 x \leqslant B_1, \quad l_1 \leqslant x^{(1)} \leqslant u_1 \\
& (x^{(2)}, x^{(3)}) \in \arg\min_{x^{(2)},x^{(3)}} \begin{cases} & C_2^\top x + x^\top H_2 x \\ \text{s.t.} & A_2 x \leqslant B_2, \quad l_2 \leqslant x^{(2)} \leqslant u_2 \\ & x^{(3)} \in \arg\min_{x^{(3)}} \begin{cases} & C_3^\top x + x^\top H_3 x \\ \text{s.t.} & A_3 x \leqslant B_3, \quad l_3 \leqslant x^{(3)} \leqslant u_3 \end{cases} \end{cases}
\end{aligned} \tag{3.1}$$

In the above problem, $x^{(i)} \in \mathbb{R}^{d_i}$ is the decision variable of the player at level i and $x = (x^{(1)}, x^{(2)}, x^{(3)})$. Let $d = \sum_{i=1}^{3} d_i$. Then, the problem parameters are $C_i \in \mathbb{R}^d$, $H_i \in \mathbb{R}^{d\times d}$, $A_i \in \mathbb{R}^{m_i\times d}$, $B_i \in \mathbb{R}^{m_i}$ and $l_i, u_i \in \mathbb{R}^{d_i}$.

3.3 Iterative Algorithm for Trilevel Optimization

We propose an iterative algorithm within a master-subproblem framework to solve the trilevel optimization problem. The master problem provides a sequence of non-decreasing lower bounds by accumulating optimality constraints, while the subproblems yield a sequence of upper bounds by evaluating the followers' reactions. The algorithm terminates when the optimality gap between the two bounds falls below a specified threshold. We use the following assumption to generate useful optimality constraints for the master problem.

Assumption 3.3.1 (Strong Convexity)**.** *The quadratic objectives are strongly convex, i.e.,* $H_i \succ 0$ *(positive definite matrix).*

Our iterative algorithm is summarized in Algorithm 3, and the different components of the algorithm is explained in the subsequent subsections.

Algorithm 3 Iterative Algorithm for Trilevel Optimization

Input: Convergence tolerance $\epsilon > 0$.
Initialize: LBD $\leftarrow -\infty$, UBD $\leftarrow +\infty$, iteration $k \leftarrow 1$.
Initial Constraints: $\mathcal{P} \leftarrow \varnothing$.
while (UBD $-$ LBD) $> \epsilon$ **do**
    Step 1: Master Problem
    Solve Master Problem (3.2)
    $(x_k, f_{1,k}) \leftarrow$ (Optimal solution, Objective value).
    LBD $\leftarrow f_{1,k}$
    Step 2: Generate Constraints
    Solve Level 2 and Level 3 subproblems (3.6) and (3.3) at fixed upper level decisions
    $(x_k^{(1)})$ and $(x_k^{(1)}, x_k^{(2)})$, respectively.
    if Level 2 subproblem is infeasible **then**
        Generate feasibility constraint using (3.10), (3.5).
    else
        Generate optimality constraint using (3.7), (3.8).
    end if
    Generate Level 3 optimality constraint using (3.4), (3.5).
    Update constraint set: $\mathcal{P} \leftarrow \mathcal{P} \cup \{\text{new constraints}\}$.
    Step 3: Upper Bound Update
    if Level 2 subproblem is feasible **then**
        Solve Upper Bound problem (3.11).
        $(x_u, f_{1,u}) \leftarrow$ (Optimal solution, Objective value).
        if $f_{1,u} <$ UBD **then**
            $UBD \leftarrow f_{1,u}$
            $\mathbf{x}^* \leftarrow \mathbf{x}_u$
        end if
    end if
    $k \leftarrow k + 1$
end while
return $\mathbf{x}^*$, UBD.

Master Problem

The master problem (3.2), which is formulated by relaxing the second and third level objectives, gives a valid lower bound. This is because it is a *high-point relaxation*, wherein the lower-level optimality conditions

$$x^{(2)}, x^{(3)} \in \arg\min\{f_2(x) : x \in \Omega_2\}$$
$$x^{(3)} \in \arg\min\{f_3(x) : x \in \Omega_3\}$$

are relaxed, where $f_i()$ and Ω_i are objective and constraint region of level i. The Master Problem minimizes the first-level objective $f_1(x)$ over the polyhedral set defined by the union of all physical constraints and the accumulated optimality and feasibility constraints. Since the feasible region of the Master Problem at iteration k, Ω^k_{MP}, is a superset of the true feasible region of the trilevel problem Ω_{true} (i.e., $\Omega_{true} \subseteq \Omega^k_{MP}$), it follows that:

$$\min_{x \in \Omega^k_{MP}} f_1(x) \leqslant \min_{x \in \Omega_{true}} f_1(x)$$

The Master Problem is defined below for iteration k.

$$\begin{aligned} x_k = \arg\min_{x} \quad & C_1^\top x + x^\top H_1 x \\ \text{s.t.} \quad & A_i x \leqslant B_i, \quad \forall i \in \{1,2,3\} \\ & l_i \leqslant x_i \leqslant u_i, \quad \forall i \in \{1,2,3\} \\ & \text{constraints from previous iterations } k' < k\text{:} \\ & \quad \text{(1) Level 2 Optimality: If (3.8) then (3.7)} \\ & \quad \text{(2) Level 2 Feasibility: If (3.5) then (3.10)} \\ & \quad \text{(3) Level 3 Optimality: If (3.5) then (3.4).} \end{aligned} \tag{3.2}$$

As the iteration count k increases, the addition of valid constraints systematically prunes Ω^k_{MP}, yielding a non-decreasing sequence of lower bounds that converges to the global optimum.

Level 3 Subproblem: Optimality constraints

In this step, we fix the decisions of the first and second levels to $x_k^{(1)}$ and $x_k^{(2)}$, the optimal values from the Master Problem, respectively. Let $x_k^{(12)} = (x_k^{(1)}, x_k^{(2)})$ represent these fixed parameters. The third-level agent solves the following quadratic program:

$$\hat{x}^{(3)} = \arg\min_{x^{(3)}} \begin{cases} x^{(3)\top} H_3^{(3,3)} x^{(3)} + (2x_k^{(12)\top} H_3^{(12,3)} + C_3^{(3)\top}) x^{(3)} \\ \text{s.t. } A_3^{(3)} x^{(3)} \leqslant B_3 - A_3^{(12)} x_k^{(12)}, \quad l_3 \leqslant x^{(3)}, \quad x^{(3)} \leqslant u_3. \end{cases} \tag{3.3}$$

where $H_3^{(3,3)}$ is the sub-block of matrix H_3 corresponding to $x^{(3)}$, and $H_3^{(12,3)}$ contains the bilinear terms between the upper levels and $x^{(3)}$. Similar notations are used for A_3. Let λ_3 denote the dual variables.

Since $H_3^{(3,3)}$ is positive definite, it is a convex optimization problem, and the optimality of $\hat{x}^{(3)}$ is determined by the Karush-Kuhn-Tucker (KKT) conditions. The KKT conditions, given the active constraint set A_3, are expressed as the linear system:

$$\text{Optimal Response: } \hat{x}^{(3)} = K_3^{(\mathsf{A}_3)\dagger(x)}(N_3^{(\mathsf{A}_3)} + M_3^{(\mathsf{A}_3)} x^{(12)}) \tag{3.4}$$

$$\begin{aligned} \text{Dual Validity: } & \hat{\lambda}_3 = K_3^{(\mathsf{A}_3)\dagger(\lambda)}(N_3^{(\mathsf{A}_3)} + M_3^{(\mathsf{A}_3)} x^{(12)}) \\ & \hat{\lambda}_3 + (\mathbf{I}_{d_3+|\mathsf{A}_3|} - K_3^{(\mathsf{A}_3)\dagger} K_3^{(\mathsf{A}_3)})^{(\lambda)} y_{null} \geqslant \mathbf{0} \\ \text{Primal Validity: } & G_3^{(\mathsf{I}_3)} \hat{x}^{(3)} \leqslant W_3^{(\mathsf{I}_3)} + S_3^{(\mathsf{I}_3)} x^{(12)} \end{aligned} \tag{3.5}$$

In this notation, the superscript (A_3) denotes indices corresponding to the active constraint set, while the superscript (I_3) denotes the inactive set. The constituent matrices are defined as follows (where $\mathbf{I}_n$ is the identity matrix of size n and $\mathbf{0}$ is zero matrix of appropriate dimension):

$$G_3 = \begin{pmatrix} A_3^{(3)} \\ -\mathbf{I}_{d_3} \\ \mathbf{I}_{d_3} \end{pmatrix}, \quad W_3 = \begin{pmatrix} B_3 \\ -l_3 \\ u_3 \end{pmatrix}, \quad S_3 = \begin{pmatrix} -A_3^{(12)} \\ \mathbf{0} \end{pmatrix}, \quad K_3^{(\mathsf{A}_3)} = \begin{pmatrix} H_3^{(3,3)} & \frac{1}{2} G_3^{(\mathsf{A}_3)\top} \\ G_3^{(\mathsf{A}_3)} & \mathbf{0} \end{pmatrix},$$
$$N_3^{(\mathsf{A}_3)} = \begin{pmatrix} -\frac{1}{2} C_3^{(3)} \\ W_3^{(\mathsf{A}_3)} \end{pmatrix}, \quad M_3^{(\mathsf{A}_3)} = \begin{pmatrix} -H_3^{(12,3)\top} \\ S_3^{(\mathsf{A}_3)} \end{pmatrix}$$

Furthermore, $K_3^{(\mathsf{A}_3)\dagger}$ is the Moore-Penrose pseudoinverse. The pseudoinverse and its null space projector are (row-wise) partitioned as $K_3^{(\mathsf{A}_3)\dagger} = [K_3^{(\mathsf{A}_3)\dagger(x)}; K_3^{(\mathsf{A}_3)\dagger(\lambda)}]$ and $\mathbf{I}_{d_3+|\mathsf{A}_3|} - K_3^{(\mathsf{A}_3)\dagger} K_3^{(\mathsf{A}_3)} = [(\mathbf{I}_{d_3+|\mathsf{A}_3|} - K_3^{(\mathsf{A}_3)\dagger} K_3^{(\mathsf{A}_3)})^{(x)}; (\mathbf{I}_{d_3+|\mathsf{A}_3|} - K_3^{(\mathsf{A}_3)\dagger} K_3^{(\mathsf{A}_3)})^{(\lambda)}]$ to indicate the indices corresponding to the primal and dual variables respectively. $y_{null} \in \mathbb{R}^{d_3+|\mathsf{A}_3|}$ is an auxiliary decision variable in the dual null space.

The way to interpret these optimality conditions is that the optimal response (3.4) is valid within the polytopic trust region defined by dual and primal validity constraints (3.5).

Remark 3.3.1 (Uniqueness of primal variables)**.** *Since the level subproblem is a convex optimization problem with a strictly convex objective, the optimal values of the primal variables are unique and given in the closed-form solution in equation (3.4).*

Remark 3.3.2 (Non-uniqueness of dual variables)**.** *When the gradients of active constraints are linearly dependent, the dual variables λ may not be unique. In constraints (3.5), we use the null space projector $(\mathbf{I}_{d_3+|\mathsf{A}_3|} - K_3^{(\mathsf{A}_3)\dagger} K_3^{(\mathsf{A}_3)})^{(\lambda)}$ to account for the non-unique component of the dual variables so that $\lambda_3 = \hat{\lambda}_3 + (\mathbf{I}_{d_3+|\mathsf{A}_3|} - K_3^{(\mathsf{A}_3)\dagger} K_3^{(\mathsf{A}_3)})^{(\lambda)} y_{null}$.*

Remark 3.3.3 (Feasibility of level 3 subproblem)**.** *For the given $(x_k^{(1)}, x_k^{(2)})$ from the Master Problem, the level 3 subproblem is always feasible. This is because the point $(x_k^{(1)}, x_k^{(2)}, x_k^{(3)})$ from the Master Problem is a feasible point for the level 3 subproblem by construction.*

The Master Problem (3.2) incorporates the optimality of the third level by adding the optimal linear mapping in equation (3.4) subject to the polytopic validity constraints in (3.5). This combination of a linear mapping and its validity polytope effectively replaces the Level 3 optimization within the Master Problem for the current region of validity.

Lemma 3.3.4 (Correctness of Level 3 Optimality Constraints)**.** *Level 3 optimality constraints are added to the Master Problem as*

$$\text{If } (x^{(1)}, x^{(2)}) \in \text{Polytope (3.5) then } x^{(3)} \text{ satisfies (3.4)},$$

holds for all $x \in \Omega_{\text{true}}$.

Level 2 Subproblem: Feasibility and Optimality constraints

Optimality constraints

To compute the optimality constraints for the second level, we first solve the corresponding bilevel subproblem with the first-level decision fixed at $x_k^{(1)}$, obtained from the optimal solution of the Master Problem. Let $x^{(23)} = (x^{(2)}, x^{(3)})$. This problem is defined as:

$$\begin{aligned} \hat{x}^{(23)} = \arg\min_{x^{(23)}} \quad & x^{(23)\top} H_2^{(23,23)} x^{(23)} + (2x_k^{(1)\top} H_2^{(1,23)} + C_2^{(23)\top}) x^{(23)} \\ \text{s.t.} \quad & A_2^{(23)} x^{(23)} \leqslant B_2 - A_2^{(1)} x_k^{(1)}, \quad l_2 \leqslant x^{(2)}, \quad x^{(2)} \leqslant u_2 \\ & x^{(3)} \in \arg\min_{x^{(3)}} \begin{cases} x^{(23)\top} H_3^{(23,23)} x^{(23)} + (2x_k^{(1)\top} H_3^{(1,23)} + C_3^{(23)\top}) x^{(23)} \\ \text{s.t.} \ A_3^{(23)} x^{(23)} \leqslant B_3 - A_3^{(1)} x_k^{(1)}, \quad l_3 \leqslant x^{(3)} \leqslant u_3. \end{cases} \end{aligned} \tag{3.6}$$

To render this problem tractable for constraint generation, we replace the lower-level optimization (Level 3) with its explicit optimal response mapping. This mapping is constructed analogously to (3.4) but is derived using the active constraint set associated with the bilevel problem above and remains valid within its corresponding polytopic region, as in (3.5). This reformulation transforms the bilevel problem into a single level quadratic program, denoted as (P_2^{UBD}), defined below, which also provides an upper bound due to Lemma (3.3.5).

$$\begin{aligned} P_2^{UBD}(x_k^{(1)}) = \min_{x^{(23)}} \quad & x^{(23)\top} H_2^{(23,23)} x^{(23)} + (2x_k^{(1)\top} H_2^{(1,23)} + C_2^{(23)\top}) x^{(23)} \\ \text{s.t.} \quad & A_2^{(23)} x^{(23)} \leqslant B_2 - A_2^{(1)} x_k^{(1)} \\ & l_2 \leqslant x^{(2)} \leqslant u_2 \\ & \text{(3.4) and (3.5) using level 3 active constraints } \mathsf{A}_3' \text{ from bilevel problem.} \end{aligned}$$

Lemma 3.3.5 (Level 2 Upper Bound Problem)**.** *Fix* $x^{(1)}$*. Let* $\Omega_2(x^{(1)})$ *denote the feasible set of the level 2 subproblem* (3.6)*, and let* $\bar{\Omega}_2(x^{(1)})$ *denote the feasible set of* $P_2^{\text{UBD}}(x^{(1)})$*, where the level 3 response is represented by* (3.4) *together with the validity polytope* (3.5) *for a fixed active set* A_3'*. Then the following hold:*

1. $\bar{\Omega}_2(x^{(1)}) \subseteq \Omega_2(x^{(1)})$.

2. The bilevel optimal point $\hat{x}^{(23)}$ *lies in the restricted feasible set, i.e.,* $\hat{x}^{(23)} \in \bar{\Omega}_2(x^{(1)})$.

In particular, $P_2^{\mathrm{UBD}}(x^{(1)})$ *is an exact restriction of the level-2 bilevel feasible set on the active-set cell associated with* A_3'*, and it contains the bilevel optimal solution.*

(P_2^{UBD}) is a convex optimization problem, and KKT conditions are necessary and sufficient for optimality. Let λ_2 represent its dual variables. Also, A_3' and A_2 denote the active constraint sets corresponding to third level constraints in bilevel problem (3.6) and upper-bound problem (P_2^{UBD}), respectively. Using KKT conditions, given the active constraint sets A_3' and A_2, the following optimality constraints can be derived.

$$\begin{aligned}
&\text{Upper Bound on Optimal Response:}\\
&\hat{x}_{UB}^{(23)} = K_2^{(\mathsf{A}_2)\dagger(x)}(N_2^{(\mathsf{A}_2)} + M_2^{(\mathsf{A}_2)}x^{(1)})\\
&x^{(23)\top}H_2^{(23,23)}x^{(23)} + (2x^{(1)\top}H_2^{(1,23)} + C_2^{(23)\top})x^{(23)} \leqslant \hat{x}_{UB}^{(23)\top}H_2^{(23,23)}\hat{x}_{UB}^{(23)} + \\
&\qquad\qquad (2x^{(1)\top}H_2^{(1,23)} + C_2^{(23)\top})\hat{x}_{UB}^{(23)}
\end{aligned} \tag{3.7}$$

$$\begin{aligned}
\text{Dual Validity: } & \hat{\lambda}_2 = K_2^{(\mathsf{A}_2)\dagger(\lambda)}(N_2^{(\mathsf{A}_2)} + M_2^{(\mathsf{A}_2)}x^{(1)})\\
& \hat{\lambda}_2 + (\mathbf{I}_{d_2+d_3+|\mathsf{A}_2|} - K_2^{(\mathsf{A}_2)\dagger}K_2^{(\mathsf{A}_2)})^{(\lambda)}\tilde{y}_{null} \geqslant \mathbf{0}\\
\text{Primal Validity: } & G_2^{(\mathsf{I}_2)}\hat{x}_{UB}^{(23)} \leqslant W_2^{(\mathsf{I}_2)} + S_2^{(\mathsf{I}_2)}x^{(1)}
\end{aligned} \tag{3.8}$$

Remark 3.3.6 (Upper bound on Optimal Response)**.** *Unlike in (3.4), where the optimal response admits a closed-form linear expression, only a quadratic inequality can be obtained for the optimal objective of the second-level subproblem. This is because the constraints are derived from the upper-bounding problem* (P_2^{UBD})*. Also, using this inequality in the Master Problem gives it the flexibility to handle degeneracies in the optimal decisions at Level 2.*

We use similar notations as before and the constituent matrices are defined as follows:

$$G_2 = \begin{pmatrix} A_2^{(2)} & A_2^{(3)} \\ -\mathbf{I}_{d_2} & \mathbf{0} \\ \mathbf{I}_{d_2} & \mathbf{0} \\ Q^{(2)} & \mathbf{0} \\ -(K_3^{(\mathsf{A}_3')\dagger(x)}M_3^{(\mathsf{A}_3')})^{(2)} & \mathbf{I}_{d_3} \end{pmatrix}, \quad W_2 = \begin{pmatrix} B_2 \\ -l_2 \\ u_2 \\ -Q^{(0)} \\ K_3^{(\mathsf{A}_3')\dagger(x)}N_3^{(\mathsf{A}_3')} \end{pmatrix}, \quad S_2 = \begin{pmatrix} -A_2^{(1)} \\ \mathbf{0} \\ \mathbf{0} \\ -Q^{(1)} \\ (K_3^{(\mathsf{A}_3')\dagger(x)}M_3^{(\mathsf{A}_3')})^{(1)} \end{pmatrix},$$

$$Q^{(0)} = \begin{pmatrix} -K_3^{(\mathsf{A}_3')\dagger(\lambda)}N_3^{(\mathsf{A}_3')} - (\mathbf{I}_{d_3+|\mathsf{A}_3'|} - K_3^{(\mathsf{A}_3')\dagger}K_3^{(\mathsf{A}_3')})^{(\lambda)}y_{null} \\ G_3^{(\mathsf{I}_3')}K_3^{(\mathsf{A}_3')\dagger(x)}N_3^{(\mathsf{A}_3')} - W_3^{(\mathsf{I}_3')} \end{pmatrix}, \quad K_2^{(\mathsf{A}_2)} = \begin{pmatrix} H_2^{(23,23)} & \frac{1}{2}G_2^{(\mathsf{A}_2)\top} \\ G_2^{(\mathsf{A}_2)} & \mathbf{0} \end{pmatrix},$$

$$N_2^{(\mathsf{A}_2)} = \begin{pmatrix} -\frac{1}{2}C_2^{(23)} \\ W_2^{(\mathsf{A}_2)} \end{pmatrix}, \quad M_2^{(\mathsf{A}_2)} = \begin{pmatrix} -H_2^{(1,23)\top} \\ S_2^{(\mathsf{A}_2)} \end{pmatrix}, \quad (Q^{(1)}\ Q^{(2)}) = \begin{pmatrix} -K_3^{(\mathsf{A}_3')\dagger(\lambda)}M_3^{(\mathsf{A}_3')} \\ G_3^{(\mathsf{I}_3')}K_3^{(\mathsf{A}_3')\dagger(x)}M_3^{(\mathsf{A}_3')} - S_3^{(\mathsf{I}_3')} \end{pmatrix}.$$

Note that there are two auxillary decision variables $y_{null} \in \mathbb{R}^{d_3+|\mathsf{A}_3'|}$ and $\tilde{y}_{null} \in \mathbb{R}^{d_3+d_2+|\mathsf{A}|}$ corresponding to level 3 and level 2 dual variable space, respectively. As for the interpretation of the optimality constraints, the upper bound on optimal response (3.7) is valid within the polytopic trust region defined by dual and primal validity constraints (3.8).

Remark 3.3.7 (Active Constraint Sets)**.** *The active constraint sets used to calculate the optimality constraints for level 2 and level 3 subproblems are different. These are calculated by solving their corresponding subproblems (3.6) and (3.3), respectively. This ensures that these optimality constraints are tight in some neighborhood of the upper decision variables* $(x_k^{(1)})$ *and* $(x_k^{(1)}, x_k^{(2)})$*, respectively, for the subproblems.*

Lemma 3.3.8 (Correctness of Level 2 Optimality Constraints)**.** *Assume that the level 2 subproblem is feasible. Then, the corresponding optimality constraints are added to the Master Problem as*

$$\textit{If } x^{(1)} \in \textit{Polytope (3.8) then } (x^{(2)}, x^{(3)}) \textit{ satisfies (3.7)},$$

holds for all $x \in \Omega_{\text{true}}$.

Feasibility constraints

If the level 2 subproblem (3.6) is infeasible for a given level 1 decision $x_k^{(1)}$, we generate a feasibility constraint to prune the level 1 search space and ensure that the Master Problem (3.2) does not pick $x_k^{(1)}$ again. Consider the active constraints set A_3 of level 3 subproblem (3.3) using $(x_k^{(1)}, x_k^{(2)})$ from the Master Problem. Then, the consolidated constraint system for the second level is:

$$G_2 x^{(23)} \leqslant W_2 + S_2 x^{(1)} \tag{3.9}$$

where G_2, W_2, S_2 are as defined earlier constructed using the active constraint set A_3. These matrices incorporate Level 2 constraints and Level 3 optimal response and validity constraints. Following Farkas' Lemma, infeasibility is certified by a vector $\omega \geqslant \mathbf{0}$ satisfying $\omega^\top G_2 = \mathbf{0}$ and $\omega^\top (W_2 + S_2 x^{(1)}) < \mathbf{0}$. Hence, the Master Problem is augmented with the following feasibility constraint inequality:

$$(\omega^\top S_2) x^{(1)} \geqslant -\omega^\top W_2 \tag{3.10}$$

The domain of validity of the feasibility constraint is as defined by the polytopic constraints in (3.5). The Farkas multiplier ω is obtained using the following problem :

$$\omega = \arg\max\{-\omega^\top (W_2 + S_2 x^{(1)}) : \omega^\top G_2 = 0,\ \mathbf{0} \leqslant \omega \leqslant \mathbf{1}\}.$$

Lemma 3.3.9 (Correctness of Level 2 Feasibility Constraints)**.** *Assume that the level 2 subproblem is infeasible. Then, the following feasibility constraints are added to the Master Problem*

$$\textit{If } (x^{(1)}, x^{(2)}) \in \textit{Polytope (3.5) then } x^{(1)} \textit{ satisfies (3.10)},$$

holds for all $x \in \Omega_{\text{true}}$.

Upper Bound Problem

The Upper Bound problem is formulated to identify a globally feasible candidate for the original trilevel optimization problem by fixing the first-level decision to $x_k^{(1)}$ obtained from the Master Problem. The resulting subproblem is a bilevel program:

$$
\begin{aligned}
(x_u^{(2)}, x_u^{(3)}) = \arg\min_{x^{(23)}} \quad & C_1^\top x + x^\top H_1 x \\
\text{s.t.} \quad & A_i^{(23)} x^{(23)} \leqslant B_i - A_i^{(1)} x_k^{(1)}, \quad \in \{1, 2\} \\
& l_2 \leqslant x^{(2)} \leqslant u_2 \\
& x^{(23)\top} H_2^{(23,23)} x^{(23)} + (2 x_k^{(1)\top} H_2^{(1,23)} + C_2^{(23)\top}) x^{(23)} \leqslant f_2^*(x_k^{(1)}) \\
& x^{(3)} \in \arg\min_{x^{(3)}} \begin{cases} x^{(23)\top} H_3^{(23,23)} x^{(23)} + (2 x_k^{(1)\top} H_3^{(1,23)} + C_3^{(23)\top}) x^{(23)} \\ \text{s.t. } A_3^{(23)} x^{(23)} \leqslant B_3 - A_3^{(1)} x_k^{(1)}, \quad l_3 \leqslant x^{(3)} \leqslant u_3. \end{cases}
\end{aligned}
\tag{3.11}
$$

where $f_2^*(x_k^{(1)})$ is the optimal objective value of the second-level subproblem (3.6). This problem ensures that $(x_k^{(1)}, x_u^{(2)}, x_u^{(3)})$ is feasible point of the trilevel problem (3.1), providing it a valid upper bound.

Lemma 3.3.10 (Correctness of Upper Bound Problem)**.** *Let $(x_u^{(2)}, x_u^{(3)})$ be an optimal solution of the upper-bound problem* (3.11) *at $x_k^{(1)}$. Then, the point $x_u := (x_k^{(1)}, x_u^{(2)}, x_u^{(3)})$ is feasible for the original trilevel problem* (3.1)*. Consequently, the corresponding objective value $C_1^\top x_u + x_u^\top H_1 x_u$ is a valid upper bound on the optimal value of the trilevel problem* (3.1)*.*

Remarks on Convergence

We now posit that the proposed iterative algorithm terminates in finitely many iterations and returns a globally optimal solution of the trilevel problem (3.1). The main intuition is that the master problem provides a monotonically nondecreasing sequence of lower bounds by iteratively refining a relaxation of the trilevel problem, whereas the upper bound problem generates a nonincreasing sequence of feasible upper bounds. The constraints in each iteration is derived from a specific active set combination of $(\mathsf{A}_2, \mathsf{A}_3, \mathsf{A}_3')$, and there are only a finite number of such combinations. We postulate that convergence is reached when an active set combination repeats, at which point the lower and upper bounds converge. The latter statement is key to the convergence proof and is proved in the next lemma.

Lemma 3.3.11 (Closing optimality gap)**.** *Suppose that at iteration k the algorithm generates level 2 and level 3 constraints using active sets $(\mathsf{A}_2, \mathsf{A}_3, \mathsf{A}_3')$, and adds the corresponding constraints* (3.4)*–*(3.5) *and* (3.7)*–*(3.8) *to the master problem. Assume that at a later iteration $k' > k$, the master solution $x_{k'} = (x_{k'}^{(1)}, x_{k'}^{(2)}, x_{k'}^{(3)})$ lies in the validity polytopes of these constraints and that the level 2 and level 3 subproblems at $x_{k'}$ exhibit the same active sets*

$(\mathsf{A}_2, \mathsf{A}_3, \mathsf{A}_3')$. *Then* $x_{k'}$ *is feasible for the original trilevel problem* (3.1)*, and the lower and upper bounds coincide at iteration* k'.

When the feasible region Ω_{true} of the trilevel problem (3.1) is nonempty and compact, the quadratic objectives are strongly convex, i.e., $H_i \succ 0$ for $i = 1, 2, 3$, and the constraints at each level are linearly independent, then we postulate that the iterative algorithm (3) generates a monotone sequence of lower bounds and upper bounds, terminates after finitely many oracle calls, and returns a globally optimal solution to the trilevel problem.

3.4 Numerical Experiments

Numerical experiments are conducted using a randomized trilevel optimization benchmark designed to evaluate the solver's performance across varying problem dimensions. For each test configuration, the generator constructs instances with positive definite Hessian matrices (Q) to ensure strong convexity in the objective functions. Objective and constraint coefficients are sampled from the disjoint set $(-5, -0.5) \cup (0.5, 5.0)$, and constraints are generated to guarantee full row rank and a non-empty feasible region. The experiments evaluate performance across six distinct configurations of decision variables (n_i) and constraints (m_i), and with 2 problem instances for each configuration within a 120 second time limit per instance. The below table reports average solve times, average lower and upper bounds, and total iteration counts.

Table 3.1: Numerical Results for Trilevel Optimization Benchmark

Variables (n_1, n_2, n_3)	**Constraints** (m_1, m_2, m_3)	**Avg. Time** (seconds)	**Avg. LBD**	**Avg. UBD**	**Ave. Iter.** (count)
(2, 2, 2)	(1, 1, 1)	1.945	-0.500	-0.499	2.0
(2, 2, 2)	(2, 2, 2)	0.705	-6.914	-6.915	0.5
(3, 2, 2)	(1, 1, 1)	1.304	-2.638	-2.637	0.5
(3, 2, 2)	(2, 2, 2)	1.079	-4.720	-4.720	1.5
(3, 2, 3)	(2, 2, 2)	1.418	-4.427	-4.425	1.5
(3, 2, 3)	(1, 1, 1)	3.963	0.206	0.205	2.0

3.5 Conclusion

In conclusion, we proposed an exact iterative method for continuous trilevel optimization problems with strongly convex quadratic objectives and linear constraints. By exploiting the parametric structure of follower problems, the algorithm constructs and refines a master problem over finitely many active set regions, and we showed finite time convergence to an optimal solution in the numerical experiments.

Chapter 4

Tensor Completion Via Integer Optimization

4.1 Introduction

A tensor is a multilinear operator, and it can be represented as an array of numbers referenced by multiple indices. For a tensor $\psi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$ of order p, we use $\psi_x := \psi_{x_1,\ldots,x_p}$ to refer to an entry corresponding to the indices $x = (x_1, \ldots, x_p)$, where $x_i \in [r_i]$ and $[s] := \{1, \ldots, s\}$. Furthermore, we define parameters: $\rho = \sum_i r_i$, $\pi = \prod_i r_i$, and $\mathcal{R} = [r_1] \times \cdots \times [r_p]$. Though matrices are special cases of tensors, problems like computing rank, singular values, and nuclear norm for tensors with $p \geqslant 3$ are NP-hard [85].

In this chapter, we address one such difficult problem called *tensor completion*. Here, a small subset of tensor entries are observed – possibly with noise. Under an assumption of low-rankness, the problem is to fill-in the remaining, unobserved entries and remove noise, if any. Some applications of tensor completion include recommendation systems [73, 93], information diffusion [172], nonparametric regression [8], computer vision [54, 144, 152], bioinformatics [24, 1] and control systems [153, 173].

Related Work

The information-theoretic rate for estimation error in tensor completion is $\sqrt{k \cdot \sum_i r_i/n}$, where: k is tensor rank, r_i is the i-th dimension of the tensor, and n is the number of samples [70, 121, 88]. Past work has studied tradeoffs between computational complexity and this rate [16].

Initial work on tensor completion used decomposition methods based on CP and Tucker decomposition [157]. Other approaches accounted for robustness to outliers and data corruptions [89, 88], or imposed fixed-rank constraints [100]. Alternative approaches used various tensor norms as a convex surrogate for tensor rank [121, 169, 86]. Our approach falls into this last category, where we give a new construction of the nuclear-∞ norm [66] using a gauge function to allow the use of integer optimization.

Past approaches to tensor completion are computationally efficient without achieving the information-theoretic rate and rely on heuristics to compute non-certifiably optimal solutions for NP-hard problems, where optimal solutions (if found) theoretically achieve the information-theoretic rate [100, 169, 88]. Some specialized tensor completion algorithms have successfully achieved the information-theoretic rate in practical computation time, such as rank-1 nonnegative tensor completion formulated as conic optimization with exponential constraints [8], symmetric tensor completion using a variant of the Frank-Wolfe algorithm [134] and two-stage non-convex optimization [33], and first-order optimization methods leveraging integer optimization for a weak separation oracle in nonnegative tensors [31, 129].

Main Contribution

Our algorithm does not assume tensor properties like non-negativity or symmetricity and works for general tensors. It is guaranteed to reach a global optimum in a linear (in the required accuracy as represented in bits) number of oracle calls, where the oracle solves a weak separation problem using integer optimization and a heuristic. More importantly, it attains the information-theoretic rate, i.e., data efficiency, while finding optimal solutions on instances of general tensors of sizes up to $10^{\times 7}$ within minutes.

The main idea behind our algorithm is to define the tensor completion problem using a gauge norm constructed using a convex polytope with its vertices as rank-1 tensors. We relate the gauge norm to tensor rank and analyze the norm's computational and statistical complexities in relation to NP-hardness and low Rademacher complexity, respectively. Consequently, the tensor completion problem using gauge norm is also NP-hard to solve to arbitrary accuracy. Nevertheless, since the formulation is a convex optimization problem, we design an algorithm using a Franke-Wolfe-like first order method called Blended Conditional Gradients (BCG) [29]. We construct a weak separation oracle for BCG using an integer linear optimization formulation and deploy an additional heuristic to accelerate the computation of oracle calls.

Going from Nonnegative to General Tensors

This chapter extends an approach for nonnegative tensors [31] to general tensors, which we explain below is a nontrivial generalization. The key idea in [31] was to define a gauge norm using a 0-1 polytope representing the convex hull of rank-1 nonnegative tensors with {0, 1} entries. The polytope was such that linear separation problems over it could be written using integer linear constraints. For illustration consider the set $\{\zeta = \prod_{k=1}^{p} \nu_k : \nu_k \in \{0,1\}\}$. Its linearization $\{\zeta : 0 \leqslant \zeta \leqslant \nu_k, \sum_k \nu_k + (1-p) \leqslant \zeta, \nu_k \in \{0,1\}\}$ is easily possible since $\zeta = 0$ if any single $\nu_k = 0$, and $\zeta = 1$ if all $\nu_k = 1$. The natural generalization of this idea, pursued here, is a gauge norm using a polytope that is the convex hull of all (general) rank-1 tensors with entries ± 1. However, for the new set $\{\zeta = \prod_{k=1}^{p} \nu_k : \nu_k \in \{-1,+1\}\}$ satisfies $\zeta = +1$ if an even number of $\nu_k = -1$, and $\zeta = -1$ if an odd number of $\nu_k = -1$. Hence, the linearization is fundamentally different, more challenging, and requires new computational design and theoretical analysis.

General tensors also have well-posedness issues that do not occur for nonnegative tensors. In particular, the best low-rank approximation problem is not well-posed for tensors [145]. In contrast, the best low-rank approximation problem *is* well-posed for nonnegative tensors [132]. Given these phenomena for general tensors, there is a technical challenge in determining whether a specific gauge norm, as defined above, can act as a convex surrogate for the rank of tensors.

Applications to Electromobility and Battery Storage

Increased use of renewable energy and the rapid growth of electric vehicles (EVs) has introduced new complexities in energy grid management, namely integration of consumers who store surplus energy to share with the grid [171]. This has resulted in a bidirectional and dynamic energy flow, making grid management more intricate. EVs serve as a prime example, as they both consume significant energy and can store and reinject surplus energy into the grid [155].

Making predictions of electricity demand, generation, and storage is beneficial to optimal operation of the grid, but scaling model estimation and predictions to billion-device-scale systems is challenging because of the high computational burden of typical nonparametric models (e.g., random forests, neural networks). Here, we propose the use of tensor completion to construct nonparametric models because they have extremely low computation requirements for prediction. Indeed, these use of tensor completion for regression has been previously considered for bioengineered systems [8]. Here, we use household energy storage data from Ireland [159] to demonstrate a proof of concept in this direction.

4.2 Preliminaries

By definition, a rank-1 tensor can be written as the tensor product of vectors, that is $\psi = \bigotimes_{k=1}^{p} \theta^{(k)}$, where $\theta^{(k)} \in \mathbb{R}^{r_k}$. Equivalently, each tensor entry is $\psi_x \coloneqq \psi_{x_1,\cdots,x_p} = \prod_{k=1}^{p} \theta^{(k)}_{x_k}$, where $\theta^{(k)}_{x_k}$ is the x_k-th element of vector $\theta^{(k)}$ for any index value $x_k \in [r_k]$. When obvious, we will use θ_{x_k} instead of $\theta^{(k)}_{x_k}$. Let $\mathcal{B}_\lambda$ be the set of rank-1 tensors such that each entry of the tensor has absolute values less than or equal to $\lambda \in \mathbb{R}_+$:

$$\mathcal{B}_\lambda = \{\psi : \psi_x = \lambda \cdot \textstyle\prod_{k=1}^{p} \theta_{x_k},\ \theta_{x_k} \in [-1, 1] \text{ for } x \in \mathcal{R}\}.$$

Then, the rank of a tensor is defined as:

$$\text{rank}(\psi) = \min\{q \mid \psi = \textstyle\sum_{k=1}^{q} \psi^k, \psi^k \in \mathcal{B}_\infty \text{ for } k \in [q]\},$$

and its CP decomposition is given by $\psi = \sum_{k=1}^{\text{rank}(\psi)} \psi^k$.

The tensor completion problem we consider begins with n observations of the tensor, which are denoted by the pairs $(x\langle i\rangle, y\langle i\rangle) \in \mathcal{R} \times \mathbb{R}$ for $i \in [n]$. Here, $y\langle i\rangle$ is the (possibly noisy) observation of the tensor entry $\psi_{x\langle i\rangle}$. We note that the $x\langle i\rangle$ are assumed to be independent and identically distributed in our model, which means that any given entry of the tensor may be observed multiple times within the n observations. Our approach is to solve the tensor completion problem using a least squares formulation:

$$\begin{aligned} \widehat{\psi} \in \arg\min_{\psi} \ & \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left(y\langle i\rangle - \psi_{x\langle i\rangle}\right)^2 \\ \text{s.t. } & \|\psi\|_{\pm} \leqslant \lambda \end{aligned} \tag{4.1}$$

where the constraint uses a gauge norm $\|\psi\|_{\pm}$ defined below.

4.3 Gauge Norm for Tensors

We construct a norm for general tensors using a gauge function [35, 87, 31]. Let $\mathcal{S}_\lambda$ be the set of rank-1 tensors such that each entry of the tensor has an absolute value of some $\lambda \in \mathbb{R}_+$:

$$\mathcal{S}_\lambda = \{\psi : \psi_x = \lambda \cdot \textstyle\prod_{k=1}^{p} \theta_{x_k},\ \theta_{x_k} \in \{-1, 1\} \text{ for } x \in \mathcal{R}\}.$$

Proposition 4.3.1. *The convex hulls of the sets $\mathcal{B}_\lambda$ and $\mathcal{C}_\lambda$ are the same, i.e. $\mathcal{C}_\lambda := \mathrm{conv}(\mathcal{B}_\lambda) = \mathrm{conv}(\mathcal{S}_\lambda)$.*

Remark 4.3.2. *Note $\mathcal{B}_\lambda = \lambda\mathcal{B}_1$, $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$, and $\mathcal{C}_\lambda = \lambda\mathcal{C}_1$.*

$\mathcal{C}_\lambda$ is useful because it is a convex polytope with its vertices as the points in $\mathcal{S}_\lambda$, and we use it to define a gauge norm.

Proposition 4.3.3. *The function defined as*

$$\|\psi\|_{\pm} := \inf\{\lambda \geqslant 0 \mid \psi \in \lambda\mathcal{C}_1\}$$

is a norm for all tensors $\psi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$.

To develop a practical algorithm using this norm, our next result provides an alternative characterization of the vertices of $\mathcal{C}_\lambda$ using linear inequality constraints of (binary) integer variables. This is unlike, say, the nuclear norm under the ℓ_2 norm that has a nonlinear (continuous) feasible region.

Proposition 4.3.4. *We have $\widehat{\mathcal{S}}_\lambda = \mathcal{S}_\lambda$ for the set*

$$
\begin{aligned}
\widehat{\mathcal{S}}_\lambda = \big\{\psi :\ & \psi_x = \lambda \cdot y_{x,1} && x \in \mathcal{R} \\
& y_{x,k} \geqslant (-\theta_{x_k} - y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \geqslant (\theta_{x_k} + y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \leqslant (\theta_{x_k} - y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \leqslant (-\theta_{x_k} + y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,p} = \theta_{x_p} && x \in \mathcal{R} \\
& \theta_{x_k} \in \{-1, 1\} && x \in \mathcal{R} \\
& \theta^k \in \mathbb{R}^{r_k},\ y_{x,k} \in \mathbb{R} && k \in [p], x \in \mathcal{R}\big\}.
\end{aligned}
$$

Remark 4.3.5. *The norm defined in (4.3.3) is equivalent to the nuclear-∞ norm by Proposition 4.3 of [66].*

Relation between Gauge Norm and Tensor Rank

The underlying issue is related to the fact that the best low-rank approximation problem is not well-posed for tensors [145], which is in sharp contrast to the case of nonnegative tensors for which the best low-rank approximation problem is well-posed [132]. We impose a regularity condition to eliminate such pathological behavior of tensors.

Assumption 4.3.1 (Regularity Condition)**.** *Consider a class of tensors defined by the set*

$$\Gamma = \big\{\psi : \exists \text{ CP } \textit{decomposition of } \psi \textit{ with terms } \psi^k \textit{s.t.} \|\psi^k\|_{\max} \leqslant \|\psi\|_{\max} \textit{ for } k \in [\mathrm{rank}(\psi)]\big\}.$$

This class is such that each tensor ψ has its largest entry at least as large as the largest entry of each CP *term ψ^k.*

Remark 4.3.6. *A* CP *decomposition always exists for a finite-valued tensor, but it may not be unique. The class defined above asks that the regularity condition holds for at least one* CP *decomposition, but does not make any statement about holding for all the possible* CP *decompositions.*

The following proposition suggests that the norm $\|\psi\|_\pm$, which is convex, can be a useful surrogate for tensor rank.

Proposition 4.3.7. *For any $\psi \in \Gamma$ that satisfies Assumption 4.3.1, we have $\|\psi\|_{\max} \leqslant \|\psi\|_\pm \leqslant \mathrm{rank}(\psi) \cdot \|\psi\|_{\max}$.*

Remark 4.3.8. *Tensor rank surrogacy is not always possible. It does not hold for a gauge norm defined using the l_1-ball.*

Complexity Analysis of Norm

We show that calculating the norm $\|\cdot\|_\pm$ is NP-complete.

Proposition 4.3.9 (Computational complexity). *It is in fact* NP*-complete to determine if* $\|\psi\|_\pm \leqslant K$ *for* $\psi \in \mathbb{R}^{r_1\times\cdots\times r_p}$.

Rademacher complexity, from computational learning theory, is used to characterize the richness of a class of functions [18, 150]. Roughly speaking, function classes with lower Rademacher complexity can be learned using less samples. From the following proposition, one can check that the norm $\|\cdot\|_\pm$ has an exponentially smaller Rademacher complexity than the max and Frobenius norms for tensors.

Proposition 4.3.10 (Stochastic Complexity). *We have the relation* $\mathsf{R}(\mathcal{C}_\lambda) \leqslant \mathsf{W}(\mathcal{C}_\lambda) \leqslant 2\lambda\sqrt{\rho/n}$, *where* $\mathsf{R}(\cdot)$ *and* $\mathsf{W}(\cdot)$ *are the Rademacher and worst case Rademacher complexities.*

4.4 Algorithm for Tensor Completion

Complexity Analysis of Tensor Completion

By interpreting the tensor completion problem (4.1) as a *convex aggregation* problem [124, 161, 105] for a finite set of functions, one can arrive at the generalization bound for the solution. Interestingly, these generalization bounds are the same in the special case of nonnegative tensors [31]. We believe this is because the proof for nonnegative tensors did not exploit the non-negativity, and hence could have led to non-tight generalization bounds. For completeness, we state these statistical guarantees in the following two results:

Proposition 4.4.1 ([105]). *Suppose* $|y| \leqslant b$ *almost surely. Given any* $\delta > 0$, *with probability at least* $1-4\delta$ *we have that* $\mathbb{E}\big((y-\psi_x)^2\big) \leqslant \min_{\varphi\in\mathcal{C}_\lambda}\mathbb{E}\big((y-\varphi_x)^2\big) + c_0\cdot\max\big[b^2,\lambda^2\big]\cdot\max\big[\zeta_n,\frac{\log(1/\delta)}{n}\big]$, *where* c_0 *is an absolute constant and*

$$\zeta_n = \begin{cases} \frac{2^\rho}{n}, & \text{if } 2^\rho \leqslant \sqrt{n} \\ \sqrt{\frac{1}{n}\log\left(\frac{e2^\rho}{\sqrt{n}}\right)}, & \text{if } 2^\rho > \sqrt{n} \end{cases}$$

Corollary 4.4.2 ([31]). *Suppose* $\psi \in \Gamma$ *is a tensor (satisfying Assumption 4.3.1) with* $\operatorname{rank}(\psi) = k$ *and* $\|\psi\|_{\max} \leqslant \mu$. *Under an additive noise model, if* $(x\langle i\rangle, y\langle i\rangle)$ *are independent and identically distributed with* $|y\langle i\rangle - \psi_{x\langle i\rangle}| \leqslant e$ *almost surely and* $\mathbb{E}y\langle i\rangle = \psi_{x\langle i\rangle}$. *Then given any* $\delta > 0$, *with probability at least* $1-4\delta$ *we have*

$$\mathbb{E}\big((y-\widehat{\psi}_x)^2\big) \leqslant e^2 + c_0\cdot\big(\mu k + e\big)^2\cdot\max\big[\zeta_n,\tfrac{\log(1/\delta)}{n}\big],$$

where ζ_n *is as in (4.4.1) and* c_0 *is an absolute constant.*

Remark 4.4.3. *The above result achieves the information-theoretic rate when the rank $k = O(1)$.*

Unsurprisingly, the tensor completion problem is NP-hard, because approximating the norm $\|\cdot\|_{\pm}$ is NP-hard and there is a polynomial-time reduction of the problem to the NP-hard weak membership problem [31].

Proposition 4.4.4. *Tensor completion is* NP*-hard to solve to arbitrary accuracy, and its decision version is* NP*-complete.*

Numerical Algorithm

Despite being NP-hard, the tensor completion optimization problem is convex. This enables application of various first-order convex optimization algorithms. In particular, we use a variant of the Frank-Wolfe algorithm called *Blended Conditional Gradients* (BCG) [29]. The main alternative approaches for iterative optimization seem to run into issues inherent to the implicit description of the feasible region; for instance, we do not have an effective barrier function available for interior point methods, nor do we have an efficient projection oracle that can be leveraged in projected gradient descent.

BCG makes iterate updates based on the gradient of the objective function at the current iterate. It needs a weak separation oracle to find a new vertex that reduces a gradient-based linear objective. The weak separation oracle is given in Algorithm 4. The output of the oracle should either give a vertex that accomplishes separation, or a certificate that separation is not possible. We design our weak separation oracle using two algorithms, an integer optimization problem in (4.2) and an alternating maximization heuristic.

Since integer optimization is likely to be more computationally expensive than the alternating heuristic, we first try the latter several times with different randomized initializations. The heuristic adapts Algorithm 5, where solutions are explored by toggling between ± 1 [31]. The following (nonlinear) separation objective is minimized over binary θ, given a BCG-generated parameter c: $z_M(\theta) := \sum_{x \in \mathcal{R}} \langle c_x, \psi_x - \lambda \cdot \prod_{k=1}^{p} \theta_{x_k} \rangle$. The heuristic considers one dimension at a time, setting the corresponding entries of θ to optimize within the local neighbourhood defined by the given dimension (which can be done by simply considering the signs of c). It runs in polynomial time (linear in the size of θ) and has been seen to speed up the computation in our simulations by reducing calls to the integer optimization solver. However, the heuristic is merely that and cannot in general give a certificate for the non-existence of separation, which requires global optimization. Indeed, even for a matrix ($p = 2$), z_M represents a generic form of NP-hard Quadratic Unconstrained Binary Optimization (QUBO) [109].

If the heuristic cannot yield a separating cut, we implement the following integer opti-

Algorithm 4 Weak Separation Oracle for $\mathcal{C}_\lambda$

Input: linear objective $c \in \mathbb{R}^{r_1 \times \cdots \times r_p}$, point $\psi \in \mathcal{C}_\lambda$, accuracy $K \geqslant 1$, gap estimate $\Phi > 0$, norm bound λ

Output: Either (1) vertex $\varphi \in \mathcal{S}_\lambda$ with $\langle c, \psi - \varphi \rangle \geqslant \Phi/K$, or (2) **false:** $\langle c, \psi - \varphi \rangle \leqslant \Phi$ for all $\varphi \in \mathcal{C}_\lambda$

mization. Note $\langle \cdot, \cdot \rangle$ is the dot product of tensors obtained by flattening them into vectors.

$$\begin{aligned}
\max_{\varphi,\theta,y} & \langle c, \psi - \varphi \rangle \\
\text{s.t. } & \varphi_x = \lambda y_{x,1} && x \in \mathcal{R} \\
& y_{x,k} \geqslant (-\theta_{x_k} - y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \geqslant (\theta_{x_k} + y_{x,k+1} - 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \leqslant (\theta_{x_k} - y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,k} \leqslant (-\theta_{x_k} + y_{x,k+1} + 1) && k \in [p-1], x \in \mathcal{R} \\
& y_{x,p} = \theta_{x_p} && x \in \mathcal{R} \\
& \theta_{x_k} \in \{-1, 1\} && k \in [p], x \in \mathcal{R}
\end{aligned} \tag{4.2}$$

Note that, in principle, for some binary variable $q \in \{-1, 1\}$, one can directly branch on the disjunction: $q = -1 \vee q = 1$. However, in implementation we choose instead to reformulate via 0-1 auxiliary variables, $\sigma := (q+1)/2$, as enforcing $\sigma \in \{0, 1\}$ obviates the need for imposing $q \in \{-1, 1\}$. Integer optimization solvers typically can only handle 0-1 binary variables by default instead of $\{-1, 1\}$ due to the more extensive set of techniques developed for 0-1 variables.

Since we are looking for weak separation, the integer optimization solver is made to terminate when a solution with an objective greater than Φ/K is found. In case of no such solution, the dual bound z from the solver serves as a no-separation certificate satisfying $\langle c, \psi - \varphi \rangle \leqslant z \leqslant \Phi$.

Proposition 4.4.5. *Using BCG, the tensor completion problem in (4.1) can be solved in a linear (in the required accuracy as represented in bits) number of calls to the Weak Separation Oracle that is presented in Algorithm 4.*

In fact, we find the BCG algorithm to be practically efficient for our problem since the weak linear separation oracle can accept (sufficiently good) suboptimal solutions that may be found at early termination of a global solver. When we design our weak separation oracle calls to an integer optimization problem, we explicitly define a tolerance for early termination. This works out well as integer optimization solvers usually discover near-optimal solutions fast and subsequently work hard to certify them.

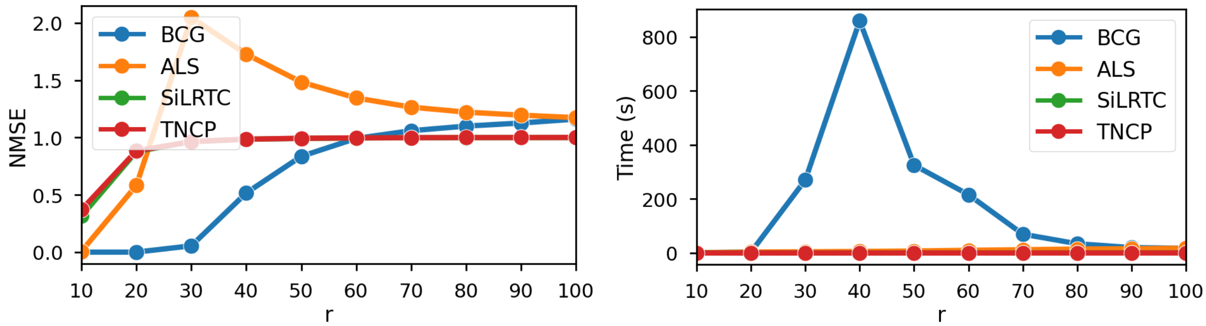


Figure 4.1: NMSE and computation time (in s) for order-3 tensors with size $r \times r \times r$ and $n = 1000$ samples.

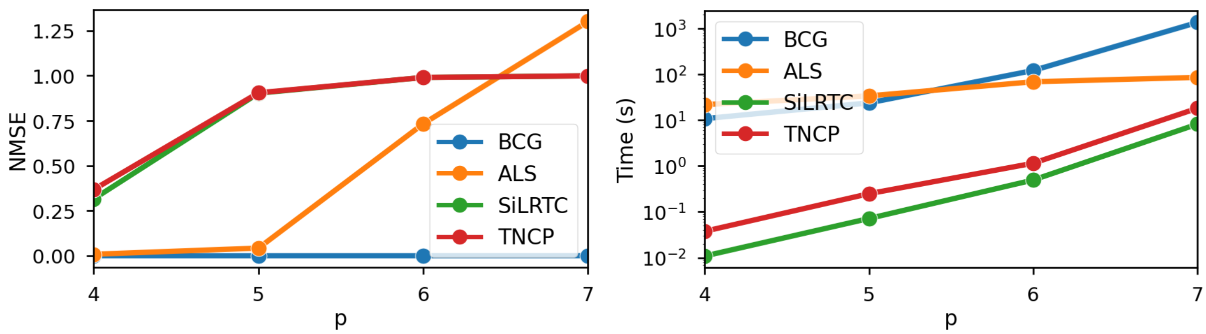


Figure 4.2: NMSE and computation time (in s) for increasing order tensors with size $10^{\times p}$ and $n = 10,000$ samples.

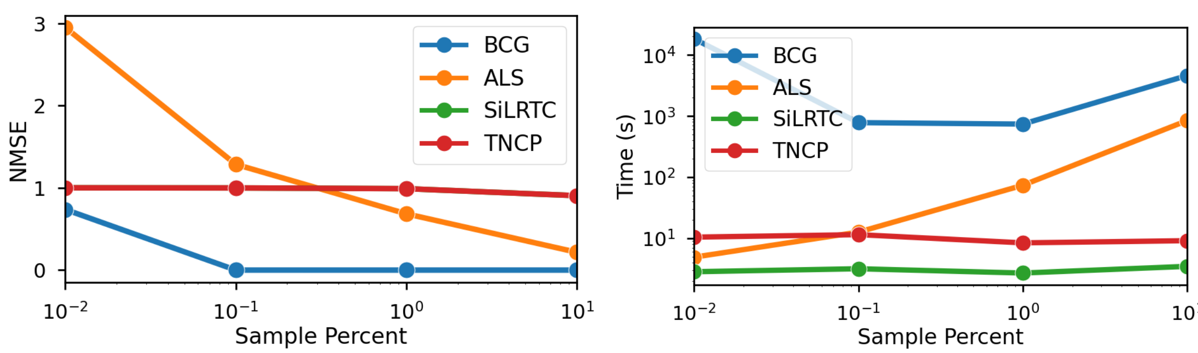


Figure 4.3: NMSE and computation time (in s) for tensors with size $10^{\times 7}$ and increasing n samples.

Algorithm 5 Alternating Maximization

Input: linear objective $c \in \mathbb{R}^{r_1 \times \cdots \times r_p}$, point $\psi \in \mathcal{C}_\lambda$, norm bound λ, incumbent solution $\hat{\theta} \in \mathcal{S}_\lambda$, objective function $z_M(\theta) := \sum_{x \in \mathcal{R}} \langle c_x, \psi_x - \lambda \cdot \prod_{k=1}^{p} \theta_{x_k} \rangle$.
Output: Best known solution θ

$\theta \leftarrow \hat{\theta}$
$z \leftarrow z_M(\theta)$
for $i = 1$ **to** p **do**
  for $k = 1$ **to** r_i **do**
    $\theta_k^{(i)} \leftarrow -1 \cdot \theta_k^{(i)}$
    if $z_M(\theta) > z$ **then**
      $z \leftarrow z_M(\theta)$
    else
      $\theta_k^{(i)} \leftarrow -1 \cdot \theta_k^{(i)}$
    end if
  end for
end for

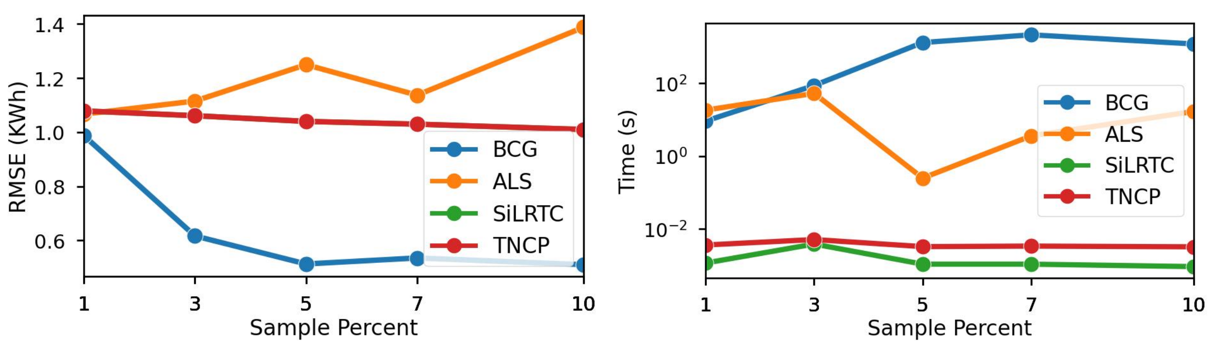


Figure 4.4: RMSE (kWh) and computation time(s) for the tensor model predicting energy recharged into storage batteries.

4.5 Numerical Experiments

We perform numerical experiments to demonstrate scalability and efficacy of our tensor completion algorithm[1]. We compare its performance with benchmark algorithms, including the often-called 'workhorse' for numerical tensor problems, alternating least squares (ALS) [99], and two state-of-the-art algorithms implemented in the PyTen[2], known as the simple

[1]Experiments were performed on a computer server running a Linux OS, with 16GB of RAM and an Intel Xeon Processor E5-2650L v3 (30M Cache, 1.80 GHz) that has 12 cores and became available in the year 2014. The algorithm was implemented in Python 3, and Gurobi v9.1 [80] was used as an integer optimization solver for the separation oracle (4.2).

[2]PyTen is available at `https://github.com/datamllab/pyten` package [149] under a GPL 2 license.

low-rank tensor completion (SiLRTC) algorithm [111] and the trace norm regularized CP decomposition (TNCP) algorithm [112].

Experiments on Simulated Tensor Data

We use simulated data to demonstrate the scalability of our algorithm with increasing tensor dimension and tensor order. The true tensor ψ is constructed by taking a random convex combination of a random set of ten points from $\mathcal{S}_1$ so that $\|\psi\|_{\pm} \leqslant 1$ and $\operatorname{rank}(\psi) \leqslant 10$. To minimize the influence of hyper-parameter selection in the results, we use these ground truth values for the value of λ in our algorithm and the value of k in ALS and TNCP. Since ALS tends to perform better under L2 regularization [123], we chose its corresponding hyperparameter to make it favorable to its accuracy. Each experiment was performed with 100 repetitions, and the normalized mean squared error (NMSE), i.e., $\|\widehat{\psi} - \psi\|_F^2/\|\psi\|_F^2$ was calculated. NMSE is a stricter measure than the error metric used in Corollary 4.4.2 because the statistical guarantee is not normalized.

In Figure 4.1, we use tensors of order $p = 3$ with dimensions increasing from $r = 10$ to $r = 100$. In each experiment, we observe $n = 1000$ samples, allowing for repetition in indices. Our approach yields greater accuracy in a lower size of r while all algorithms do not perform at a satisfactory level (NMSE below 1) as r increases close to 100.

In Figure 4.2, we consider tensors of increasing order p with dimension $r_i = 10$ for $i = 1, \ldots, p$. We observed $n = 10,000$ samples, again randomly sampled with replacement. we observe that our algorithm achieves consistently higher accuracy as compared to the other methods. In these plots, while our algorithm takes more computation time, it still converges within minutes even for tensors with 10^7 entries.

In Figure 4.3, we construct tensors of sizes $10^{\times 7}$. In each experiment, the sample size is increased by one order of magnitude as the percentage of total tensor entries, starting from 0.01% (using random sampling with replacement). The results show our algorithm achieves much higher accuracy while requiring greater computation time. Yet our algorithm converges within minutes for most cases, except for the case where the sample percent is 10^{-2}% (i.e., about three hours).

Experiments on Household Energy Storage Data

We run experiments on local weather and energy data from residential households in Ireland for the year 2020 [159]. Using tensor completion, we construct a nonparameteric model to predict how much energy (in kWh) is charged into storage batteries in an hour in one household. The inputs of this model are local temperature (in °C), percentage of battery charge, day of week (Monday-Friday) and hour of day (0-23). We map values of these inputs into 10, 5, 5, 24 uniform buckets, respectively, so that our nonparametric model is a $10\times5\times5\times24$ tensor. This model structure is beneficial for integration of model predictions for billion-device-scale systems because predictions have extremely low computational burden since model predictions simply become lookups into a table: the point at which a prediction is to be made is mapped to a set of indices based on the defined buckets, and the model prediction is simply the completed tensor value at that corresponding set of indices.

We randomly divide our dataset into non-overlapping train and test sets comprising 20% and 80% of total data, respectively. To run the experiments, we randomly sample (with replacement) $s = 1\%$ to $s = 10\%$ of total tensor entries in the nonparameteric model (i.e. $s \times 10 \times 5 \times 5 \times 24 = 6000 \times s$ entries). For each case of sample percentage s, we obtain the completed tensor $\widehat{\psi}$ and calculate the root mean square error (RMSE) on the test set X_{test}, i.e. $\sqrt{\sum_{x \in X_{\text{test}}} \|\widehat{\psi}_x - \psi_x^{\text{test}}\|^2 / \# X_{\text{test}}}$. To tune the hyperparameters of our and other algorithms, we do a grid search over the values in $[0.5 \cdot \|\psi^{\text{train}}\|_{\max}, 1.5 \cdot \|\psi^{\text{train}}\|_{\max}]$, where $\|\psi^{\text{train}}\|_{\max}$ is the maximum entry in the training set. The results plotted in Figure 4.4 show that our model outperforms other models at low sample percentages ranging from 1 to 10%, although with a slightly larger computation time for model training. However, we recall that model predictions have extremely low computational load, as described above.

4.6 Conclusion

We developed a (general) tensor completion algorithm by defining a tensor norm using the gauge of a specific polytope. The method balances computation and information-theoretic limits: it provably converges in a linear number of integer optimization oracle calls while achieving the information-theoretic sample complexity rate. Numerical experiments confirm its scalability and accuracy, and an application to household energy data shows its potential for building nonparametric models of battery charging. Unlike other nonparametric models (e.g., random forests, neural networks), our approach is extremely efficient at prediction, although computationally intensive to train, making it promising for billion-device–scale energy grid management. Future work could focus on accelerating training to further enhance performance.

Part III

Strategic Data Governance and Sharing

Chapter 5

When would Online Platforms pay Data Dividends?

5.1 Introduction

The revenue models of many online platforms depend on collecting, analyzing, and selling users' data. A free-service and advertising-based revenue model can cause conflicts in the users' and platform's interests. Users may be concerned about their privacy and possible misuse of data, while platforms want to maximize their profits. Further, users' perception of a platform's ethics and their willingness to participate can be affected by the platform's revenue model, pricing decisions, and privacy practices[166, 151, 168].

Cybersecurity on Online Platforms

After the onset of the Internet of Things (IoT), there has been an increase in the variety, speed, and volume of users' data collected. Information from multiple sources including devices, sensors, and social networks is being used by platforms to assist users and collect data [130, 52].

Though users have become more aware of the rampant collection of their data, they often do not know about the proliferation of IoT devices in their everyday lives. For instance, in August 2022 the Australian federal court convicted a major search platform for collecting users' location data without their knowledge [10]. When users give consent and permissions to apps and platforms, they tend to underestimate the implications of it [76]. Reforms to protect users' privacy are very much needed across the world [122]. Users should be able to control, delete and transfer their data across different platforms and service providers. They should be asked for explicit consent every time a platform wants to use their data for a new purpose [130].

Privacy and Data Dividends

With growing user concerns, consumer privacy legislation has become an important topic for public discussion, and multiple new data privacy laws have been introduced. In Europe, the General Data Protection Regulation (GDPR) was introduced in 2016 to give consumers more control over their data [135]. In 2019, legislators in California announced their intent to introduce data dividends, which is a model in which platforms would pay users in exchange for use of users' data [165]. The same year, Oregon legislators introduced a bill, called the Health Information Property Act, to compensate consumers for monetizing their health data [148].

Implementing data dividends comes with its own challenges as companies holding the data are far more powerful than individual users. Further, there is a huge information asymmetry and only companies know the actual value of the users' data. The critics of data dividends argue that selling data would make it a commodity and be counter-productive in protecting users' privacy. They also feel that vulnerable groups – such as people of color and the poor – who are currently discriminated against should not be incentivized to pour more data into the system and further reduce their privacy [160, 19]. On the other hand, the proponents of data dividends argue that today's technology economy is hugely driven by monetizing users' data, and paying users a share of these benefits is only fair.

Recent studies have explored different ways of pricing data dividends for each user based on the value of their individual data [22, 23, 76]. Using the idea of Shapley and Owen values, they calculated the contribution made by each user to the platform's profits. For our work, we ask different questions: Should platforms pay data dividends at all, and why? In this chapter, we do our analysis with homogeneous users but our methods can be extended to analyze heterogeneous users too.

Contributions

Our work comes in the context of rising debates on data dividends. We try to understand when and how much online platforms would pay as data dividends. We consider users' privacy concerns as an important factor, for which a platform can invest in data protection. The platform can also pay users to incentivize taking risks and sharing their data. Our main contributions are to introduce a principal-agent model using a Stackelberg game to capture the platform-users dynamics and then use this to gain a better understanding of when the platform would pay users with data dividends.

5.2 Our Model

An online platform provides its users with free services and a data dividend in exchange for their data. The users are allowed to choose between two levels – high or low – of data sharing. For each level, the platform provides a different data dividend and set of services. As the platform collects users' data, it faces the risk of a possible data breach. If a data breach occurs, the platform loses its reputation and faces possible legal and financial complications, and the users are harmed by the misuse of their data. Hence, the platforms consider investing in protecting their users' data. This could include the costs of building better technological infrastructure and signing insurance contracts.

Platform's Utility

The platform has a fixed cost of S to maintain and provide its services. It invests I in users' data protection, reducing the probability of a data breach to $B(I)$. Here, B is assumed to be a twice differentiable function such that $B(I) > 0$ with $\lim_{I\to\infty} B(I) = 0$, $B'(I) < 0$ and $B''(I) > 0$ $\forall I$. Note that B is not a complementary cumulative distribution function. Instead, $B(I)$ gives the probability of data breach on investment I. In case of a data breach, the platform has a loss of F from legal cases and a lost reputation. When having access to k users' data, the platform makes a revenue of $U(k, b)$, where b of the total k users chose the low level of data sharing. The revenue may be due to selling the data to a third-party, selling advertisements to display to users or other sources. The platform pays a share of this revenue to users as data dividends, which are priced at two scales – p_0 and p_1 – for the low and high levels of data sharing respectively.

Considering these costs, revenue, and risks, the platform has a total expected utility of :

$$U(k, b) - B(I)F - I - S - p_0 b - p_1(k - b).$$

User's Utility

We consider k homogeneous users who have the same behavior and parameters. Each user values their personal data at V and faces an additional personal loss of L on a data breach. They also benefit from the platform's free services, which amount to a value of W. When a user chooses the low level of data sharing, these benefits and losses scale down by a factor $\alpha \in (0, 1)$. The platform pays users with data dividends at two scales of pay and thereby encourages a particular level of data sharing.

Let c_i be user i's decision variable so that $c_i = 0$ and 1 for the low and high levels of data sharing respectively. Then a user i's total expected utility is:

$$c_i(p_1 - \mathcal{V}) + (1 - c_i)(p_0 - \alpha\mathcal{V}),$$

where $\mathcal{V} = \bar{V} + B(I)L$ and $\bar{V} = V - W$.

Principal-Agent Formulation

We construct a Stackelberg game [146, 147, 20], in which players act sequentially with follower(s) acting after a leader. In our model, the platform first prices data dividends for different levels of data-sharing, and decides on investment for users' data protection. Given this information, the users decide how much data to share.

The optimal, equilibrium choices of the platform and the users are their Stackelberg strategies. We capture this using an optimization problem, which maximizes the platform's utility when the users are maximizing their utilities[108]. It is formulated as follows:

$$\begin{aligned} \max \;\; & U(k,b) - B(I)F - I - S - p_0 b - p_1(k-b) \\ s.t \;\; & c_i^* = \arg \max_{c_i \in \{0,1\}} c_i(p_1 - \mathcal{V}) + (1-c_i)(p_0 - \alpha\mathcal{V}) \\ & \qquad\qquad\qquad\qquad\qquad\qquad \forall i = \{1, \cdots, k\} \\ & c_i^*(p_1 - \mathcal{V}) + (1-c_i^*)(p_0 - \alpha\mathcal{V}) \geqslant 0 \\ & \qquad\qquad\qquad\qquad\qquad\qquad \forall i = \{1, \cdots, k\} \\ & I,\; p_0\;, p_1 \geqslant 0. \end{aligned}$$

The second constraint above ensures that the users do not have a net loss from using the platform, without which they would leave the platform.

5.3 Optimal Choices

Since the users are homogeneous, they make similar choices: either $c_i^* = 1$ or $c_i^* = 0$ for every user i. If users find both levels of data sharing to be utility-maximizing, then we assume that they all choose the level that most benefits the platform. We solve these two cases of c_i^* separately. The optimal solution is the best of the two cases and can vary with the numerical values of the parameters of the model.

Case 1: $c_i^* = 1 \;\; \forall i = \{1, \cdots, k\}$

To ensure users choose high level of data sharing, the platform must price the data dividends such that $p_1 - \mathcal{V} \geqslant p_0 - \alpha\mathcal{V}$. Note that $b = 0$ in this case. The optimization problem then reduces to:

$$\begin{aligned} \max \; & U(k,0) - B(I)F - I - p_1 k - S \\ s.t \;\; & p_1 - \mathcal{V} \geqslant p_0 - \alpha\mathcal{V} \\ & p_1 - \mathcal{V} \geqslant 0 \\ & I,\; p_0\;, p_1 \geqslant 0. \end{aligned}$$

Since $p_0 \geqslant 0$, $p_0^* = 0$ is optimal for the problem. Given this, one can conclude $p_1^* = \mathcal{V}$ when $\mathcal{V} \geqslant 0$ and $p_1^* = 0$ when $\mathcal{V} \leqslant 0$. Also, $U(k,0) - S$ can be treated as a constant. These observations further reduce this case to two sub-cases with optimization problems in a single variable I.

Sub-case 1: If $\mathcal{V} \geqslant 0$ then $p_1^* = \mathcal{V}$.

Set $p_1^* = \mathcal{V} = \bar{V} + B(I)L$ and add an additional constraint $\mathcal{V} \geqslant 0$.

$$\begin{aligned} &\min B(I)F_1 + I + \bar{V}k \\ &s.t\ \ \bar{V} + B(I)L \geqslant 0 \\ &\quad\ \ I \geqslant 0, \end{aligned}$$

where $F_1 = F + Lk$.
Let $I_1 = B'^{-1}(\frac{-1}{F_1})$ and $I_2 = B^{-1}(\frac{-\bar{V}}{L})$.

Proposition 5.3.1. *When $\bar{V} \geqslant 0$, $I^* = I_1$ if I_1 exists. Else, $I^* = 0$.*

Proposition 5.3.2. *When $\bar{V} \leqslant 0$,*

1. *If I_1 exists, $I^* = I_1$ when $B(I_1)L \geqslant -\bar{V}$. Else when $B(I_1)L < -\bar{V}$, $I^* = I_2$ if I_2 exists and no solution otherwise.*

2. *If I_1 doesn't exist, $I^* = 0$ when $B(0)L \geqslant -\bar{V}$ and no solution otherwise.*

Sub-case 2: If $\mathcal{V} \leqslant 0$ then $p_1^* = 0$.

Set $p_1 = 0$ and add an additional constraint $\mathcal{V} \leqslant 0$.

$$\begin{aligned} &\min B(I)F_0 + I \\ &s.t\ \ \bar{V} + B(I)L \leqslant 0 \\ &\quad\ \ I \geqslant 0 \end{aligned}$$

where $F_0 = F$.
Let $I_3 = B'^{-1}(\frac{-1}{F_0})$ and as before, $I_2 = B^{-1}(\frac{-\bar{V}}{L})$.

Proposition 5.3.3. *When $\bar{V} > 0$, there is no solution.*

Proposition 5.3.4. *When $\bar{V} \leqslant 0$,*

1. *If I_3 exists, $I^* = I_3$ when $-\bar{V} \geqslant B(I_3)L$. Else when $-\bar{V} < B(I_3)L$, $I^* = I_2$.*

2. *If I_3 doesn't exist, $I^* = 0$ when $-\bar{V} \geqslant B(0)L$ and $I^* = I_2$ otherwise.*

The results from the two sub-cases can be combined by finding the minimum objective value of the two sub-cases.

Final results of Case 1

Theorem 5.3.5. *For the case when all users choose the high level of data-sharing, when $\bar{V} \leqslant 0$ the optimal choices of the platform are as follows:*

1. *If I_1 and I_3 exist, ($I^* = I_1$, $p_1^* = \bar{V} + B(I_1)L$, $p_0^* = 0$) when $L \geqslant \frac{-\bar{V}}{B(I_1)}$, ($I^* = I_2$, $p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(I_1)} \geqslant L \geqslant \frac{-\bar{V}}{B(I_3)}$ and ($I^* = I_3$, $p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(I_3)} \geqslant L$.*
2. *If I_1 exists but not I_3, ($I^* = I_1$, $p_1^* = \bar{V} + B(I_1)L$, $p_0^* = 0$) when $L \geqslant \frac{-\bar{V}}{B(I_1)}$, ($I^* = I_2$, $p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(I_1)} \geqslant L \geqslant \frac{-\bar{V}}{B(0)}$ and ($I^* = 0$, $p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(0)} \geqslant L$.*
3. *If both I_1 and I_3 don't exist, ($I^* = 0$, $p_1^* = \bar{V} + B(0)L$, $p_0^* = 0$) when $L \geqslant \frac{-\bar{V}}{B(0)}$ and ($I^* = 0$, $p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(0)} \geqslant L$.*

Theorem 5.3.6. *For the case when all users choose the high level of data-sharing, when $\bar{V} \geqslant 0$ the optimal choices of the platform are as follows:*

1. *If I_1 exists then ($I^* = I_1$, $p_1^* = \bar{V} + B(I_1)L$, $p_0^* = 0$) is optimal.*
2. *If I_1 doesn't exist then ($I^* = 0$, $p_1^* = \bar{V} + B(0)L$, $p_0^* = 0$)*

The final results of case 1 are summarized in table I.

Case 2: $c_i^* = 0 \;\; \forall i = \{1, \cdots, k\}$

In this case, platform must price the data dividends such that $p_0 - \alpha\mathcal{V} \geqslant p_1 - \mathcal{V}$. Here, $b = k$. The optimization problem becomes:

$$\begin{aligned} &\max U(k,k) - B(I)F - I - p_0 k - S \\ &s.t \;\; p_0 - \alpha\mathcal{V} \geqslant p_1 - \mathcal{V} \\ &\quad\;\;\; p_0 - \alpha\mathcal{V} \geqslant 0 \\ &\quad\;\;\; I,\ p_0\ , p_1 \geqslant 0. \end{aligned}$$

Similar to case 1, it can be observed that $p_1^* = 0$ is optimal and $p_0^* = \alpha\mathcal{V}$ when $\mathcal{V} \geqslant 0$ and $p_0^* = (\alpha - 1)\mathcal{V}$ when $\mathcal{V} \leqslant 0$.

Let $I_4 = B'^{-1}\left(\frac{-1}{F+\alpha Lk}\right)$, $I_5 = B'^{-1}\left(\frac{-1}{F+(\alpha-1)Lk}\right)$ and as before, $I_2 = B^{-1}(\frac{-\bar{V}}{L})$.

Theorem 5.3.7. *For the case when all users choose the low level of data-sharing, when $\bar{V} \leqslant 0$ the optimal choices of the platform are as follows:*

1. *If I_4 and I_5 exist, ($I^* = I_4, p_1^* = 0, p_0^* = \alpha(\bar{V} + B(I_4)L)$) when $L \geqslant \frac{-\bar{V}}{B(I_4)}$, ($I^* = I_2, p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(I_4)} \geqslant L \geqslant \frac{-\bar{V}}{B(I_5)}$ and ($I^* = I_5$, $p_1^* = 0$, $p_0^* = (\alpha - 1)(\bar{V} + B(I_5)L)$ when $\frac{-\bar{V}}{B(I_5)} \geqslant L$.*

2. *If I_4 exists but not I_5, ($I^* = I_4$, $p_1^* = 0$, $p_0^* = \alpha(\bar{V} + B(I_4)L)$) when $L \geqslant \frac{-\bar{V}}{B(I_4)}$, ($I^* = I_2$, $p_1^* = 0$, $p_0^* = 0$) when $\frac{-\bar{V}}{B(I_4)} \geqslant L \geqslant \frac{-\bar{V}}{B(0)}$ and ($I^* = 0$, $p_1^* = 0$, $p_0^* = (\alpha - 1)(\bar{V} + B(0)L)$) when $\frac{-\bar{V}}{B(0)} \geqslant L$.*

3. *If both I_4 and I_5 don't exist, ($I^* = 0$, $p_1^* = 0$, $p_0^* = \alpha(\bar{V} + B(0)L)$) when $L \geqslant \frac{-\bar{V}}{B(0)}$ and ($I^* = 0$, $p_1^* = 0$, $p_0^* = (\alpha - 1)(\bar{V} + B(0)L)$) when $\frac{-\bar{V}}{B(0)} \geqslant L$.*

Theorem 5.3.8. *For the case when all users choose the low level of data-sharing, when $\bar{V} \geqslant 0$ the optimal choices of the platform are as follows:*

1. *If I_4 exists then ($I^* = I_4$, $p_1^* = 0$, $p_0^* = \alpha(\bar{V} + B(I_4)L)$) is optimal.*

2. *If I_4 doesn't exist then ($I^* = 0$, $p_1^* = 0$, $p_0^* = \alpha(\bar{V} + B(0)L)$)*

The final results of case 2 are summarized in table 5.1.

Insights from Optimal Choices

The platform uses data dividends to direct users to choose a particular level of data sharing. In general, the platform pays a higher data dividend for the high level of data sharing to compensate for users' larger risks. But it also earns a larger revenue when it collects more data. This trade-off between revenue and data dividend decides the optimal level of sharing.

It is optimal for the platform to pay for only one level of data sharing which maximizes its profit. However, sometimes it may prefer to not pay at all. The cases when this happens are outlined below.

Can $p_1^* = 0$ while encouraging high level of data sharing?

Firstly, $p_1^* \neq 0$ when $V > W$. That is, the platform always pays a data dividend whenever users value their data more than the platform's free services. When $V \leqslant W$, the users value the platform's free services more than their data. In this case, if the platform ensures that a data breach does not occur then it does not pay any data dividend. So if $B(I) \to 0\ \forall I$, meaning there is a low possibility of a data breach, or $L \to 0$ so that users face a small loss from a data breach, then no data dividend is paid. Interestingly, even if $F \to \infty$, which is if the platform has heavy losses from a data breach, then they do not pay users with data dividends and instead invest heavily in data protection ($I \to \infty$). In general, the platform does not pay a data dividend when $W - V - B(I^*)L \geqslant 0$, that is, the net users' utility before pay is positive. Currently, major technology companies do not pay data dividends. Our model suggests that these companies believe they are in the region described by the above discussion.

Table 5.1: Equilibrium results for Case 1 and Case 2

<table>
<tr><td colspan="3">Final results of Case 1 when $\bar{V} \leqslant 0$ (Note: $p_0^* = 0$ always)</td></tr>
<tr><td>I_1 and I_3 exist</td><td>I_1 exists, I_3 does not</td><td>I_1 and I_3 do not exist</td></tr>
<tr><td>If $L \geqslant \frac{-\bar{V}}{B(I_1)} : I^* = I_1$,
$p_1^* = \bar{V} + B(I_1)L$
If $\frac{-\bar{V}}{B(I_1)} \geqslant L \geqslant \frac{-\bar{V}}{B(I_3)} : I^* = I_2$,
$p_1^* = 0$
If $\frac{-\bar{V}}{B(I_3)} \geqslant L : I^* = I_3$,
$p_1^* = 0$</td><td>If $L \geqslant \frac{-\bar{V}}{B(I_1)} : I^* = I_1$,
$p_1^* = \bar{V} + B(I_1)L$
If $\frac{-\bar{V}}{B(I_1)} \geqslant L \geqslant \frac{-\bar{V}}{B(0)} : I^* = I_2$,
$p_1^* = 0$
If $\frac{-\bar{V}}{B(0)} \geqslant L : I^* = 0$,
$p_1^* = 0$</td><td>If $L \geqslant \frac{-\bar{V}}{B(0)} : I^* = 0$,
$p_1^* = \bar{V} + B(0)L$
If $\frac{-\bar{V}}{B(0)} \geqslant L : I^* = 0$,
$p_1^* = 0$</td></tr>
<tr><td colspan="3">Final results of Case 1 when $\bar{V} \geqslant 0$</td></tr>
<tr><td colspan="2">I_1 exists</td><td>I_1 doesn't exist</td></tr>
<tr><td colspan="2">$I^* = I_1, \quad p_1^* = \bar{V} + B(I_1)L$</td><td>$I^* = 0, \quad p_1^* = \bar{V} + B(0)L$</td></tr>
<tr><td colspan="3">Final results of Case 2 when $\bar{V} \leqslant 0$ (Note: $p_1^* = 0$ always)</td></tr>
<tr><td>I_4 and I_5 exist</td><td>I_4 exists, I_5 does not</td><td>I_4 and I_5 do not exist</td></tr>
<tr><td>If $L \geqslant \frac{-\bar{V}}{B(I_4)} : I^* = I_4$,
$p_0^* = \alpha(\bar{V} + B(I_4)L)$
If $\frac{-\bar{V}}{B(I_4)} \geqslant L \geqslant \frac{-\bar{V}}{B(I_5)} : I^* = I_2$,
$p_0^* = 0$
If $\frac{-\bar{V}}{B(I_5)} \geqslant L : I^* = I_5$,
$p_0^* = (\alpha - 1)(\bar{V} + B(I_5)L)$</td><td>If $L \geqslant \frac{-\bar{V}}{B(I_4)} : I^* = I_4$,
$p_0^* = \alpha(\bar{V} + B(I_4)L)$
If $\frac{-\bar{V}}{B(I_4)} \geqslant L \geqslant \frac{-\bar{V}}{B(0)} : I^* = I_2$,
$p_0^* = 0$
If $\frac{-\bar{V}}{B(0)} \geqslant L : I^* = 0$,
$p_0^* = (\alpha - 1)(\bar{V} + B(0)L)$</td><td>If $L \geqslant \frac{-\bar{V}}{B(0)} : I^* = 0$,
$p_0^* = \alpha(\bar{V} + B(0)L)$
If $\frac{-\bar{V}}{B(0)} \geqslant L : I^* = 0$,
$p_0^* = (\alpha - 1)(\bar{V} + B(0)L)$</td></tr>
<tr><td colspan="3">Final results of Case 2 when $\bar{V} \geqslant 0$</td></tr>
<tr><td colspan="2">I_4 exists</td><td>I_4 doesn't exist</td></tr>
<tr><td colspan="2">$I^* = I_4, \quad p_0^* = \alpha(\bar{V} + B(I_4)L)$</td><td>$I^* = 0, \quad p_0^* = \alpha(\bar{V} + B(0)L)$</td></tr>
</table>

Can $p_0^* = 0$ while encouraging low level of data sharing?

Like for p_1, if $V > W$ then $p_0^* \neq 0$. But when $V \geqslant W$, $p_0^* = 0$ only if $W - V - B(I^*)L = 0$, that is, the net users' utility is zero. This is strange because the platform pays a data dividend when $W - V - B(I^*)L > 0$, that is, the net users' utility before pay is positive. The reason is that in this case the users naturally prefer a high level of data sharing. So the platform has to pay a data dividend to incentivize the low level of data sharing. One could argue that the platform should rather remove the option of a high level of data sharing, and it probably would.

Finally, we also analyze the trends in the investment for users' data protection. If $B(I) \to 0\ \forall I$, meaning the possibility of a data breach is small, or if $F, L \to 0$ so that no party loses from a data breach, then the platform does not invest in data protection. If $B'(I) \to 0\ \forall I$, that is when the likelihood of a data breach is not very sensitive to investment, then the platform chooses to not invest at all.

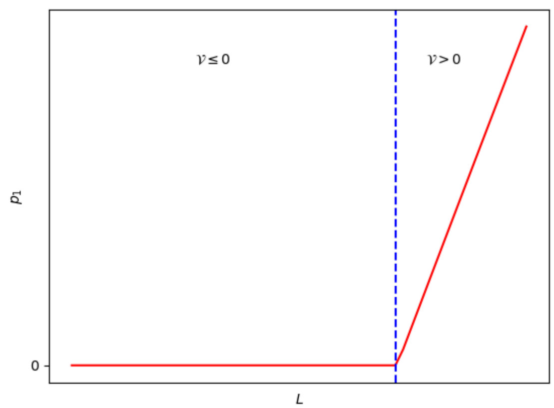


Figure 5.1: Here, users choose the high level of data sharing, I_1 exists, and $\bar{V} \leqslant 0$. Data dividend p_1 is plotted for increasing values of L, that is users' loss from a data breach. The platform pays a data dividend beyond a threshold when $\mathcal{V} > 0$, that is users face a total loss.

5.4 Conclusion

We studied the dynamics between an online platform and its users when the latter are paid with data dividends. We demonstrated how an online platform could use data dividends to incentivize its users to share more data and take more risks, while also giving them an option to share as much data as they want. If the platform wants to gather more data, it pays a data dividend only when the users perceive their risks and losses to be higher than their benefits. Further, when users feel that they do not lose much from a data breach, the platform prefers to invest in data protection rather than pay users for risking a breach. On the other hand, when users perceive a high loss from a breach, the platform has to both invest and pay. It would be interesting to analyze multiple (more than 2) levels of data sharing, each with a different data dividend, and where heterogeneous users are incentivized to share more or fewer data as is optimal for the platform. We leave this for future work.

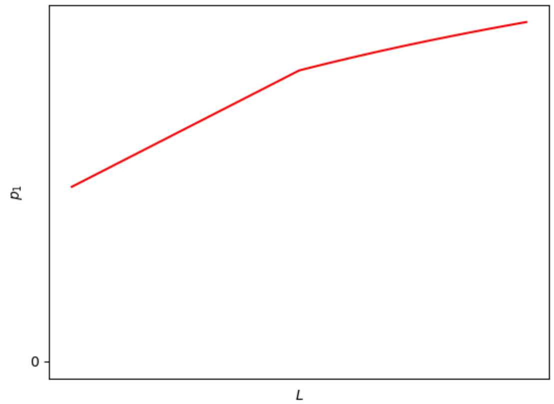


Figure 5.2: Here, users choose the high level of data sharing and $\bar{V} \geqslant 0$. Data dividend p_1 is plotted for increasing values of L, that is users' loss from a data breach. The platform always pays a data dividend of at least $\bar{V}$.

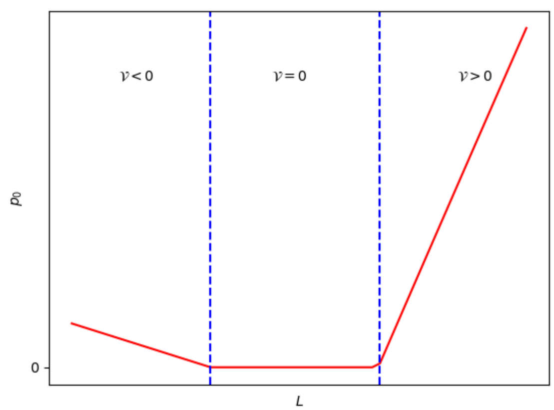


Figure 5.3: Here, users choose the low level of data sharing, I_4 exists, and $\bar{V} \leqslant 0$. Data dividend p_0 is plotted for increasing values of L, that is users' loss from a data breach. Interestingly, the platform doesn't pay a data dividend only when $\mathcal{V} = 0$, that is users neither have a net loss nor gain.

It is debatable whether data dividends could treat data as a commodity and exploit the poor and vulnerable. But if data dividends are introduced, our model shows that platforms cannot always use it to shirk responsibility for data protection. When users feel considerable harm from a data breach, platforms would both invest in data protection and pay higher data dividends. We believe that our insights could help a policymaker understand data dividends better.

Chapter 6

Understanding the Impact of Coalitions between EV Charging Stations

6.1 Introduction

The proliferation of electric vehicles (EVs) has brought major changes to road transportation. EVs could play a significant role in the transition to a sustainable energy-based future and are projected to surpass traditional internal combustion engine-based vehicles in the coming decades [25]. The growth in EVs has led to the genesis of a new industry around building faster and more accessible charging infrastructure [30], comprising several electric vehicles charging companies (EVCCs) such as Tesla, EVgo, and Chargepoint. One major challenge associated with the design of charging infrastructure is that charging stations, especially fast-charging stations, draw a considerable amount of electricity when in operation [156, 5, 163, 59] and may adversely impact the grid infrastructure [4, 107, 58]. The ideal solution to this challenge is to coordinate the demands from all charging stations to distribute the load on the grid [4, 126]. However, this is hard to implement in practice due to the high communication and computational costs of such centralized approaches, and concerns about privacy of the information shared by EV charging stations [30]. Since coordination between all EV charging stations is impractical in most cases, we consider the scenario where only *some* of these stations coordinate and form a *coalition* (refer to Figure 6.1 for a schematic). While it might intuitively seem that some coordination is always better than no coordination, we give a counter-example to show this is not necessarily true. Particularly, we study the following question in this chapter:

> **Q:** *Could coordination between a few potentially heterogeneous charging stations lead to undesirable consequences?*

To answer this question, we model the interaction between charging station as a non-

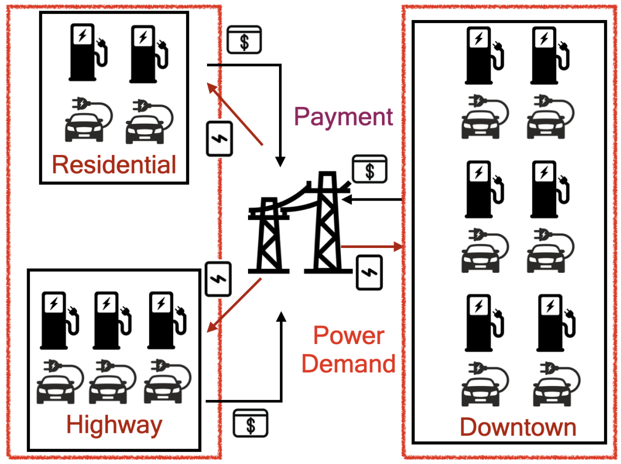


Figure 6.1: A schematic depiction of charging stations located in different areas (e.g. residential, downtown, highway) and owned by different EVCCs. Charging stations that are contained in the same red box are owned by same company (i.e. form a coalition). Each charging station demands charge from the grid, which determines the prices.

cooperative aggregative game. Charging stations are strategic agents that draw power from the grid over a finite time window, and have different location-specific charging demands and different sensitivities to deviations from a desired operating demand profile. Each charging station aims to minimize its cost comprising of (i) a payment for power demanded from the grid, and (ii) the deviations from their desired operating charging profile. In our model, we consider that the price per unit of power depends on the total power demanded by all charging stations (resulting in a game theoretic interaction between charging stations). Therefore, in our model charging stations act as "price-makers", rather than "price-takers" [116, 69, 127, 49]. Finally, some charging stations enter into a coalition (e.g. those owned by a single EVCC) to coordinate their total demand in order to minimize their total cost.

We compare the outcome of the game-theoretic interaction in two scenarios: *(a)* when some charging stations form a coalition, and *(b)* when each charging station operates independently without any coordination. We differentiate the equilibrium in these two scenarios as $\mathcal{C}-$Nash and Nash equilibrium respectively, and characterize them analytically in Theorem 6.3.1 and Corollary 6.3.5 respectively. In general, we observe that the equilibrium decomposes into two components: a charging profile uniformly distributed across time and a correction term to account for the coalition of charging stations (Theorem 6.3.1). We assess the scenarios *(a)*-*(b)* in terms of three metrics depending on the overall cost experienced by: *(i)* all charging stations (aka societal cost); *(ii)* all charging stations within a coalition, and *(iii)* all charging stations outside of coalition. We identify sufficient conditions on the game parameters under which the formation of a coalition will be worse than the independent operation of charging stations in terms of all three metrics *(i)* – *(iii)*, as presented in

Theorem 6.3.7 and 6.3.11. Furthermore, we present numerical examples that satisfy these conditions, demonstrating that coalitions can lead to worse outcomes for society, the coalition, and non-coalition charging stations alike. Notably, we find that these outcomes persist even when we relax the assumptions in our theoretical results, highlighting scenarios where the conventional intuition about coordinating charging stations does not hold.

Related Work

Aggregative games and EV charging. Several works have proposed EV charging as a non-cooperative game [116, 69, 127, 49]. These works mainly study existence, uniqueness, and computation of Nash equilibia of the EV charging game, and analyze these equilibria via measures such as the price of anarchy (PoA) [127]. Our key differentiation from this line of literature is to understand the impact of the formation of a coalition of a subset of EV stations.

Coalitions and equilibria. Strong Nash equilibrium is an outcome wherein no coalition of agents can collectively deviate from their strategy to improve their utilities, and was first introduced as a concept in [9]. Since this seminal paper, numerous studies have delved into necessary and sufficient conditions for its existence [125], computational properties [72], and its performance across various game classes, often measured by the *k-strong price of anarchy* [65, 57, 6, 64, 12, 39]. This metric quantifies the worst-case welfare loss at strong equilibrium compared to optimal welfare. Unlike strong Nash equilibrium, our notion of $\mathcal{C}-$Nash equilibrium (Definition 6.2.3) is a relaxation and only requires stability with respect to a specific coalition $\mathcal{C}$ and not all possible coalitions. Strong Nash equilibrium may not always exist in our setting but we provide an explicit characterization of $\mathcal{C}$-Nash equilibrium (Theorem 6.3.1).

Users as *price-takers*. Several studies have delved into the potential detrimental effects of uncoordinated Electric Vehicle (EV) charging on the power grid [133, 79]. In response, efforts such as those outlined in [4, 126, 68] have tackled the challenge of devising incentive mechanisms to coordinate small users, often categorized as "price-takers," to shift their charging windows and mitigate grid impacts. However, in contrast to these approaches, our work conceptualizes charging stations, which serve many EVs and thus can aggregate demand, as "price-makers" within electricity markets.

6.2 Our Model

Notations For any positive integer m, we define $[m] := \{1, 2, ..., m\}$. Consider two matrices $A = (a_{ij})_{i\in[m],j\in[n]} \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$. We define $A \otimes B \in \mathbb{R}^{pm\times qn}$ to be the Kronecker product of matrices A and B. For any finite set X, we define $\mathbf{1}_X \in \mathbb{R}^{|X|}$ to be a vector with all entries to be 1. For any vector $x \in \mathbb{R}^m$ and $p \subseteq [m]$, we use the notation $x_p \in R^{|p|}$ to denote the components of vector x corresponding to p. Additionally, we denote $x_{-p} \in R^{m-|p|}$ to denote the components of vector x corresponding to all entries that are not in p.

Charging stations as strategic entities Consider a game comprising of N charging stations, each operating as a strategic entity making decisions on their charging levels throughout a charging period spanning T units of time. Each station $i \in [N]$ is characterized by a *nominal charging profile* $(\bar{x}_i^t)_{t\in[T]}$, where the total charge demanded is d_i. That is, $\sum_{t\in[T]} \bar{x}_i^t = d_i$. The variation of charge demanded between stations reflect differences in the number of vehicles typically utilizing each charging facility.

The actual charging profile of any station may differ from the nominal charging profile due to externality imposed by other stations through electricity prices (to be defined shortly). We denote x_i^t to be the charge demanded by station i during hour t. Define $X_i = \{x_i \in \mathbb{R}^T : \sum_{t\in[T]} x_i^t = d_i\}$ to be the set of feasible charging profiles for station i [1]. With a slight abuse of notation, we define $x^t \in \mathbb{R}^N$ as a vector of charging profile of all stations at time step $t \in [T]$, and $x = (x_i^t)_{t\in T, i\in N}$ to be the joint charging profile of all stations.

In our model, we consider reactive prices where the price of electricity depends on the total charge demanded. More formally, the cost of incurred by any station i under a joint strategy x is represented as:

$$c_i(x) = \sum_{t\in[T]} p^t(x)x_i^t + \frac{\mu_i}{2}\|x_i - \bar{x}_i\|^2,$$

where (a) $\mu_i > 0$ is the sensitivity parameter of station i; (b) for every $t \in [T]$, $p^t(x)$ denotes the price per unit of electricity when the joint charging profile of all stations is x. It is through this price signal that the charging profiles of other charging stations impose externality on any station. In what follows we assume that the price function is linear, i.e. $p^t(x) = a^t + b^t \mathbf{1}_N^\top x^t$ for some $a^t, b^t > 0$ ([127, 49]). The parameters a^t represents the price fluctuations due to non-EV demand on the grid. In this chapter, we assume that for all time $t \in [T]$, $b^t = b$ for some positive scalar $b \in \mathbb{R}$.

Remark 6.2.1. *The heterogeneity in sensitivity parameters μ_i between different stations can be ascribed to the geographic location (cf. Figure 6.1). For instance, a station positioned in downtown can exhibit heightened sensitivity in meeting its demand requirements compared to one situated within a residential neighborhood.*

Nash equilibrium, defined below, is widely used to characterize the interaction in these kinds of charging games [116, 69, 127, 49].

Definition 6.2.2. *A joint charge profile x^* is a* Nash equilibrium *if $c_i(x_i^*, x_{-i}^*) \leqslant c_i(x_i, x_{-i}^*)$ for every $i \in [N], x_i \in X_i$.*

[1] We do not impose non-negativity constraints on the charging profile. This modeling decision is justified, as charging stations have the capability not only to draw power from the grid but also to inject power into the grid when necessary [96]. Correspondingly, we also do not impose upper bounds on the charging profiles; we assume that there is always enough charge available to meet the charging demand for a given problem instance. It is an interesting direction of future research to impose additional constraints on the set X which align more closely with real-world conditions.

Coalition between charging stations Coalitions between charging stations can be easily facilitated by the emerging electric vehicle charging companies (EVCCs) (such as Chargepoint, Tesla etc.) who operate multiple charging stations. We model that every coalition jointly decides their charging profiles. For ease of exposition, we consider only single coalition[2]. The goal of the coalition is to choose $x_{\mathcal{C}} \in \prod_{i\in\mathcal{C}} X_i$ that minimizes the coalition's cumulative cost function $c_{\mathcal{C}}(x) = \sum_{i\in\mathcal{C}} c_i(x)$. Next, we introduce a natural extension of Definition 6.2.2 to examine the outcomes of interactions in the presence of a coalition.

Definition 6.2.3. *A joint charge profile $x^\dagger \in X$ is a $\mathcal{C}$-*Nash equilibrium *if (i) $\sum_{i\in\mathcal{C}} c_i(x^\dagger) \leqslant \sum_{i\in\mathcal{C}} c_i(x'_{\mathcal{C}}, x^\dagger_{-\mathcal{C}})$ for every $x'_{\mathcal{C}} \in \prod_{j\in\mathcal{C}} X_j$, and (ii) $c_i(x^\dagger) \leqslant c_i(x'_i, x^\dagger_{-i})$ for every $x'_i \in X_i, i \notin \mathcal{C}$.*

When $|\mathcal{C}| = 1$, Definition 6.2.2 and 6.2.3 are equivalent. For the rest of the article, we shall denote $\mathcal{C}-$Nash equilibrium by $x^\dagger$ and Nash equilibrium when charging stations act autonomously by x^*.

6.3 Results

In Section 6.3, we analytically characterize the $\mathcal{C}-$Nash equilibrium in terms of game parameters. Next, in Section 6.3, we derive sufficient conditions when Nash equilibrium is preferred over $\mathcal{C}-$Nash equilibrium, in terms of overall cost experienced by all charging stations, by charging stations within the coalition, and the charging stations outside the coalition. Finally, in Section 6.3,we provide numerical instances where coalition is worse than Nash equilibrium.

Analytical Characterization of $\mathcal{C}-$Nash equilibrium

Theorem 6.3.1. *For any arbitrary coalition $\mathcal{C} \subset [N]$, when $b > 0$ and $\mu_i > 0$ $\forall i \in N$, the $\mathcal{C}-$Nash equilibrium exists, is unique, and takes the following form*

$$x^{\dagger t} = \frac{d}{T} + \sum_{t'\in[T]} \frac{T\delta^{tt'} - 1}{T} \left(b(\mathbf{1}_N \mathbf{1}_N^\top + \boldsymbol{C}) + \mu\right)^{-1} \cdot \left(\mu \bar{x}^{t'} - a^{t'} \mathbf{1}_N\right),$$

where $\delta^{tt'}$ is the Kronecker delta function ($\delta^{tt'} = 1$ when $t = t'$ and is 0 otherwise), $\mu =$ diag$([\mu_1, \cdots, \mu_N])$ and $\boldsymbol{C} = \begin{pmatrix} \mathbf{1}_{\mathcal{C}} \mathbf{1}_{\mathcal{C}}^\top & 0 \\ 0 & I_{N\setminus\mathcal{C}} \end{pmatrix}$.

Lemma 6.3.2. *Θ and Γ are positive definite and hence invertible matrices.*

We now make some remarks about Theorem 6.3.1.

[2] All results presented in this article can be extended to encompass scenarios involving multiple coalitions.

Remark 6.3.3. *The closed-form equilibrium solution in Theorem 6.3.1 is comprised of a fixed term and a correction term. The fixed term represents a charging profile that is uniform across time. The second term is the correction to account for the aggregate effects of the coalition, demand and charging preferences of the charging stations, and exogenous non-EV demand. As expected, when the time-dependent factors $a^{t'}$ and $\bar{x}^{t'}$ are constant across time, the correction term is zero and charging stations charge uniformly across time.*

Remark 6.3.4. *The result in Theorem 6.3.1 extends to the setting of multiple (non-overlapping) coalitions $\mathcal{C}_1, \mathcal{C}_2, ...\mathcal{C}_K$ by setting*

$$\boldsymbol{C} = \begin{pmatrix} \mathbf{1}_{\mathcal{C}_1}\mathbf{1}_{\mathcal{C}_1}^\top & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{\mathcal{C}_2}\mathbf{1}_{\mathcal{C}_2}^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \mathbf{1}_{\mathcal{C}_K}\mathbf{1}_{\mathcal{C}_K}^\top & \vdots \\ 0 & 0 & 0 & I_{N\setminus\cup_{i=1}^K \mathcal{C}_i} \end{pmatrix}.$$

In fact, all theoretical results in this article can be directly extended to the multiple coalition case by using the above $\boldsymbol{C}$ matrix. For the sake of clear presentation, we shall always work with $K = 1$ in the following text.

We conclude this section by stating the following specialization of Theorem 6.3.1 that characterizes Nash equilibrium.

Corollary 6.3.5. *By setting $\mathcal{C} = \{1\}$ in Theorem 6.3.1, the Nash equilibrium x^* takes the following form:*

$$x^{*t} = \frac{d}{T} + \sum_{t' \in [T]} \frac{T\delta^{tt'} - 1}{T} \left(b(\mathbf{1}_N\mathbf{1}_N^\top + I_N) + \mu\right)^{-1} \cdot \left(\mu\bar{x}^{t'} - a^{t'}\mathbf{1}_N\right),$$

where $I_N \in \mathbb{R}^{N\times N}$ is the identity matrix.

When is $\mathcal{C}-$Nash equilibrium beneficial?

In this subsection, we study conditions under which $\mathcal{C}-$Nash equilibrium will be preferred over Nash equilibrium and vice-versa. For a charging profile x, define the total cost incurred by a group of charging stations in $\mathcal{S} \subseteq [N]$ as

$$c_{\mathcal{S}}(x) := \sum_{i\in\mathcal{S}} c_i(x), \quad \forall\ x \in X.$$

In order to compare the outcome under Nash equilibrium and $\mathcal{C}-$Nash equilibrium, we use the following metrics:

$$\underbrace{\frac{c_{[N]}(x^*)}{c_{[N]}(x^\dagger)}}_{:=\mathsf{M}_{[N]}}, \quad \underbrace{\frac{c_{\mathcal{C}}(x^*)}{c_{\mathcal{C}}(x^\dagger)}}_{:=\mathsf{M}_{\mathcal{C}}}, \quad \underbrace{\frac{c_{[N]\setminus\mathcal{C}}(x^*)}{c_{[N]\setminus\mathcal{C}}(x^\dagger)}}_{:=\mathsf{M}_{[N]\setminus\mathcal{C}}}, \tag{Eval-Metric}$$

where, recall, x^* is the Nash equilibrium when there is no coalition, and $x^\dagger$ is the $\mathcal{C}$-Nash equilibrium. These metrics are the ratio of the total cost experienced by different groups of players at Nash equilibrium and at $\mathcal{C}-$Nash equilibrium. In particular, $\mathsf{M}_{[N]}$ is this ratio for all players, $\mathsf{M}_{\mathcal{C}}$ is this ratio for players within coalition $\mathcal{C}$, and $\mathsf{M}_{[N]\backslash\mathcal{C}}$ is this ratio of players outside of the coalition. For any $\mathcal{S} \in \{[N], \mathcal{C}, [N]\backslash\mathcal{C}\}$, if $\mathsf{M}_{\mathcal{S}} < 1$ then Nash equilibrium is preferred, otherwise $\mathcal{C}-$Nash equilibrium is preferred by the coalition $\mathcal{S}$. Next, we theoretically characterize these metrics in two cases:

Case A: Exogeneous price fluctuations a^t are uniform across time

Here, we study the case when $a^t = a \ \forall t \in [T]$ for some $a \in \mathbb{R}$. This can happen when the price is influenced by a constant non-EV demand throughout the day. Under this setting, the equilibria $x^\dagger$ and x^* are represented below

Proposition 6.3.6. *Suppose $a^t = a^{t'}$ for all $t, t' \in [T]$. Then for every $t \in [T]$,*

$$x^{*t} = \frac{d}{T} + \sum_{t' \in [T]} \frac{T\delta^{tt'} - 1}{T} \left(b(\mathbf{1}_N \mathbf{1}_N^\top + I_N) + \mu\right)^{-1} \mu \bar{x}^{t'},$$

$$x^{\dagger t} = \frac{d}{T} + \sum_{t' \in [T]} \frac{T\delta^{tt'} - 1}{T} \left(b(\mathbf{1}_N \mathbf{1}_N^\top + C) + \mu\right)^{-1} \mu \bar{x}^{t'}.$$

Further, consider the scenario when all the charging stations prepare for similar peak and low demand hours, and use charging rate recommendations from EV manufacturers to predict their demand requirements. That is, they desire similar demand profiles (up to a constant factor) across the day. This scenario is captured in Assumption 6.3.1.

Assumption 6.3.1. *The desired demand of each charging station $i \in [N]$ at time step $t \in [T]$ is $\bar{x}_i^t = d_i \alpha^t$, where $\sum_{t \in [T]} \alpha^t = 1$ and $\alpha^t \geqslant 0$ for all $t \in [T]$.*

Next, we delineate conditions under which the formation of a coalition is worse than the independent operation of charging stations in terms of verifiable conditions on game parameters.

Theorem 6.3.7. *Suppose Assumption 6.3.1 holds and $a^t = a^{t'}$ for all $t, t' \in [T]$. Any group $\mathcal{S} \subseteq [N]$ incurs a lower cost in Nash equilibrium compared to $\mathcal{C}-$Nash equilibrium iff*

$$\sum_{i \in \mathcal{S}} f_i^\dagger(\mu, b, d) - f_i^*(\mu, b, d) \geqslant 0,$$

where for every $i \in [N]$, $f_i^\dagger(\mu, b, d) := \Delta^\dagger \Delta_i^\dagger + \frac{\mu_i(\Delta_i^\dagger - d_i)^2}{2b}$, $f_i^*(\mu, b, d) := \Delta^* \Delta_i^* + \frac{\mu_i(\Delta_i^* - d_i)^2}{2b}$, $\Delta_i^\dagger := \left((b(1_N 1_N^\top + \boldsymbol{C}) + \mu)^{-1} \mu d\right)_i$, and $\Delta_i^* := \left((b(1_N 1_N^\top + I_N) + \mu)^{-1} \mu d\right)_i$. Additionally, $\Delta^\dagger := \sum_{i \in [N]} \Delta_i^\dagger$ and $\Delta^* := \sum_{i \in [N]} \Delta_i^*$.

Remark 6.3.8. *Interestingly, the condition in Theorem 6.3.7 is independent of* T, α^t, a *and depends only on* μ *and* d.

Corollary 6.3.9. *When* $c_i(x^\dagger), c_i(x^*) \geqslant 0 \ \forall i \in [N]$, *Theorem 6.3.7 can be used to analyze (Eval-Metric) as follows. For any* $\mathcal{S} \in \{[N], \mathcal{C}, [\mathcal{N}]\backslash\mathcal{C}\}$, $M_{\mathcal{S}} \leqslant 1$ *if and only if* $\sum_{i\in\mathcal{S}} f_i^\dagger(\mu, b, d) - f_i^*(\mu, b, d) \geqslant 0$.

Case B: The desired demand $\bar{x}^t$ is uniform across time

In this subsection, we analyze the setting when $\bar{x}^t$ is uniform in t. This denotes a uniform spread of desired demand by the stations. This resembles the nominal charging profile when the EV adoption will reach a critical mass when each charging station observes a constant flow of EV such that when averaged over any time window the nominal charge is uniformly spread.

Proposition 6.3.10. *If* $\bar{x}^t = d/T, \quad \forall\ t \in [T]$ *then*

$$x^{*t} = \frac{d}{T} - \sum_{t'\in[T]} \frac{T\delta^{tt'} - 1}{T} \left(b(\mathbf{1}_N\mathbf{1}_N^\top + I_N) + \mu\right)^{-1} \mathbf{1}_N a^{t'},$$

$$x^{\dagger t} = \frac{d}{T} - \sum_{t'\in[T]} \frac{T\delta^{tt'} - 1}{T} \left(b(\mathbf{1}_N\mathbf{1}_N^\top + C) + \mu\right)^{-1} \mathbf{1}_N a^{t'}.$$

Theorem 6.3.11. *Suppose* $\bar{x}^t = d/T$, *for every* $t \in [T]$. *Any group* $\mathcal{S} \subseteq [N]$ *incurs a lower cost in Nash equilibrium when compared to* $\mathcal{C}-$*Nash equilibrium if and only if* $g_{\mathcal{S}}^\dagger(\mu, b) - g_{\mathcal{S}}^*(\mu, b) \geqslant (\Gamma^\dagger - \Gamma^*)h(A)$, *where for every* $i \in [N]$,

$$g_{\mathcal{S}}^\dagger(\mu, b) := \frac{b}{T}\sum_{i\in\mathcal{S}} \Gamma^\dagger\Gamma_i^\dagger + \frac{\mu^i}{2T}(\Gamma_i^\dagger)^2$$

$$g_{\mathcal{S}}^*(\mu, b) := \frac{b}{T}\sum_{i\in\mathcal{S}} \Gamma^*\Gamma_i^* + \frac{\mu^i}{2T}(\Gamma_i^*)^2$$

$$h(A) := \frac{\sum_{t,t'\in[T]} a^t a^{t'} \left(T\delta^{tt'} - 1\right)}{\sum_{t\in[T]} \left(\sum_{t'\in[T]} a^{t'} \left(T\delta^{tt'} - 1\right)\right)^2},$$

$$\Gamma_i^* := \left((b(1_N 1_N^\top + I_N) + \mu)^{-1}\mathbf{1}_N\right)_i,$$

$$\Gamma_i^\dagger := \left(\left(b(\mathbf{1}_N\mathbf{1}_N^\top + C) + \mu\right)^{-1} \mathbf{1}_N\right)_i.$$

Additionally, $\Gamma^* := \sum_{i\in[N]} \Gamma_i^*$ *and* $\Gamma^\dagger := \sum_{i\in[N]} \Gamma_i^\dagger$.

Remark 6.3.12. *The condition in Theorem 6.3.11 does not depend on the heterogeneity of demand of charging stations* d.

Corollary 6.3.13. *When $c_i(x^\dagger), c_i(x^*) \geqslant 0 \ \forall i \in [N]$, Theorem 6.3.11 can be used to analyze (Eval-Metric) as follows.*

$$\mathsf{M}_{\mathcal{S}} \leqslant 1 \iff g^\dagger_{\mathcal{S}}(\mu, b) - g^*_{\mathcal{S}}(\mu, b) \geqslant (\Gamma^\dagger - \Gamma^*)h(A) \quad \forall \mathcal{S} \in \{[N], \mathcal{C}, [\mathcal{N}]\backslash\mathcal{C}\}.$$

Coalitions may not be always beneficial.

In this subsection, we construct instances where the formation of a coalition is not beneficial, which satisfy the setup discussed in Section 6.3. We set $N = 5, T = 10, b = 0.5$ and consider that each station can be of two types: Type H or Type L. For the purpose of this example, we consider that all stations in a coalition are of Type L and all stations outside the coalition are of Type H. A station $i \in [N]$ is said to be Type H if $d_i = 5$ and $\mu_i = 1$, and of Type L if $d_i = 1$ and $\mu_i = 0.1$. Additionally, we set $\alpha^t = \eta_1$ for $t \leqslant T/2$, and η_2 otherwise. Furthermore, we set $a_t = 0.5 + \delta\nu_t$ where $\nu_t \sim \mathsf{Unif}([0, 1])$. We study the impact of the size of the coalition on various metrics presented in (Eval-Metric). Particularly, in Figure 6.2, we study the case of $\eta_1 = 0.4/T, \eta_2 = 1.6/T$ and $\delta = 0$, which is aligned with the setup considered in Theorem 6.3.7. Meanwhile, in Figure 6.3, we study the case by setting $\eta_1 = 1/T, \eta_2 = 1/T$ and $\delta = 0.4$, which is aligned with the setup considered in Theorem 6.3.11. From these figures we conclude that there exist coalitions, where from the coalition's perspective (i.e. $\mathsf{M}_{\mathcal{C}}$) the outcome under Nash equilibrium is preferred to the outcome under coordination. Furthermore, in Figure 6.2, we find that there exist instances when the $\mathcal{C}-$Nash equilibrium is not preferred under every evaluation metric. Moreover, in Figure 6.3, we find that the formation of coalition not only adversely impacts the coalition but also provides advantage to stations outside coalitions.

6.4 Discussion

In this section we expand on the experimental results from Section 6.3[3] by relaxing the conditions imposed on game parameters.

Impact of simultaneous variations in a^t and α^t In Figure 6.4, we study the impact of the size of the coalition in terms of (Eval-Metric). In contrast to Section 6.3, we consider both a^t and α^t to be non-uniform and randomly assign stations to be of Type H with probability 0.2 if they are within the coalition. All stations outside of the coalition are assigned to be Type L. We find that if the size of coalition is small then Nash equilibrium turns out to be favorable along all metrics (Eval-Metric). However, as the size of coalition increases then it may be beneficial to form a coalition.

[3]The code to generate all figures in this section is available at https://github.com/kkulk/coalition-ev

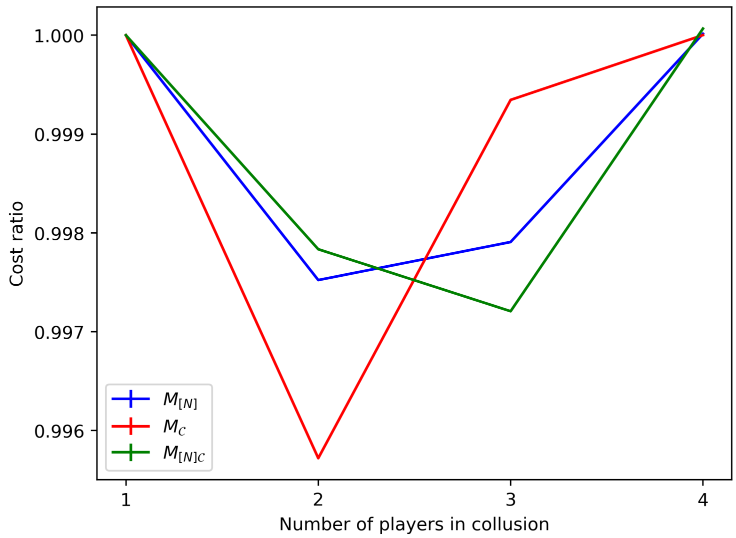


Figure 6.2: Setting $\eta_1 = 0.4/T, \eta_2 = 1.6/T$ and $\delta = 0$

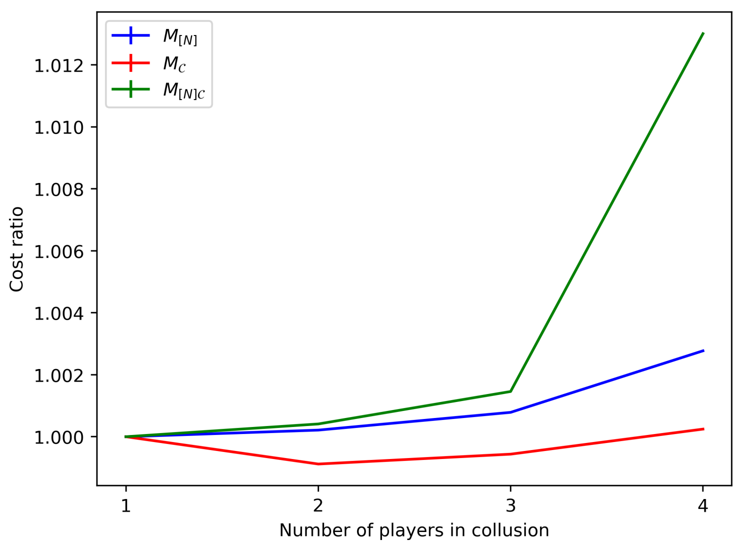


Figure 6.3: Setting $\eta_1 = 1/T, \eta_2 = 1/T$ and $\delta = 0.4$

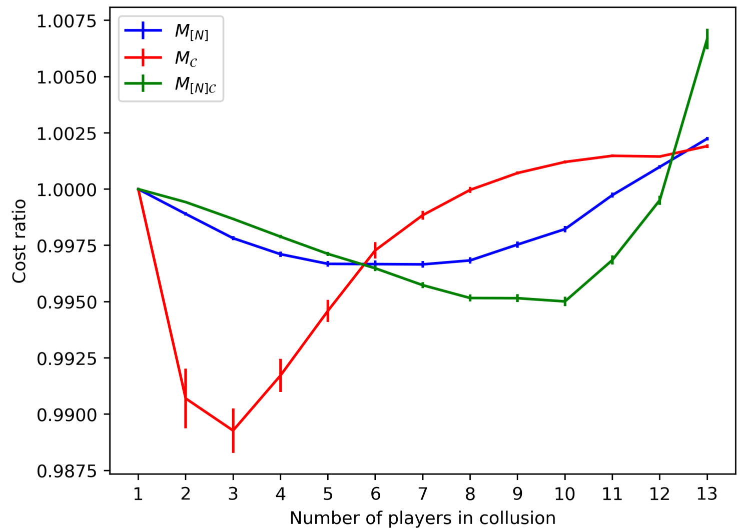


Figure 6.4: Comparison between Nash equilibrium and $\mathcal{C}-$Nash equilibrium with respect to (Eval-Metric) under different size of coalition.

Impact of the composition of coalitions Here we show that not only the size of the coalition, but also the composition of the coalition, plays an important role in deciding whether the coalition is beneficial. To illustrate this point, we examine an example with $N = 3, T = 10, b = 1$ and $a^t = 10$ for all $t \in [T]$. We posit a scenario where the first two stations form a coalition. Each station is one of two types, namely Type H or Type L. We call a station $i \in [N]$ to be of Type H if $d_i = 5$ and $\mu_i = 5$, and of Type L if $d_i = 1$ and $\mu_i = 1$. In Figure 6.5, we present a comparative analysis, evaluating how different coalitions perform in terms of metric presented in (Eval-Metric). We find under some circumstances $\mathsf{M}_{\mathcal{S}} > 1$ for all $\mathcal{S} \in \{[N], \mathcal{C}, [N]\backslash[\mathcal{C}]\}$. For instance, this is the case when the coalition is comprised of atleast one Type H station and the station outside coalition is of Type H. On the contrary, perhaps surprisingly, there are also instances when $\mathsf{M}_{\mathcal{S}} < 1$ for all $\mathcal{S} \in \{[N], \mathcal{C}, [N]\backslash[\mathcal{C}]\}$. For instance, this is the case when the stations within a coalition are of Type H and the station outside the coalition is of Type L.

6.5 Conclusion

In this chapter, we initiate a study to understand the impact of coalitions between charging stations as charging infrastructure continues to grow in coming years. As charging stations draw a substantial amount of electricity from the grid, they will be "price-makers" on the electricity grid. More often than not, multiple charging stations are operated by same electric vehicle charging company (EVCC), which could facilitate coordination between charging stations. We analytically characterize the equilibrium outcome in the presence of coalitions.

Types in Coalition

Type Outside Coalition	HH	HL	LL
H	Coalition better	Coalition better	Nash better
L	Nash better	Coalition better	Coalition better

(a) The phrase "Coalition better" (resp. "Nash better") implies $\mathsf{M}_{[N]} > 1$ (resp. $\mathsf{M}_{[N]} < 1$).

Types within Coalition

Type Outside Coalition	HH	HL	LL
H	Coalition better	Coalition better	Coalition better
L	Nash better	Coalition better	Nash better

(b) The phrase "Coalition better" (resp. "Nash better") implies $\mathsf{M}_{\mathcal{C}} > 1$ (resp. $\mathsf{M}_{\mathcal{C}} < 1$).

Types within Coalition

Type Outside Coalition	HH	HL	LL
H	Coalition better	Coalition better	Nash better
L	Nash better	Nash better	Coalition better

(c) The phrase "Coalition better" (resp. "Nash better") implies $\mathsf{M}_{[N]\setminus\mathcal{C}} > 1$ (resp. $\mathsf{M}_{[N]\setminus\mathcal{C}} < 1$).

Figure 6.5: Comparison of societal (overall) cost, coalitional cost, and the cost of the station outside the coalition in the three station game.

Our analysis hints at potential losses encountered by EVCCs if they coordinate all charging stations owned by them in the presence of heterogeneity in charging demand and user preferences.

There are several interesting questions for future research in understanding the impact of coalitions between charging stations. In the current model we assume the overall demand of a charging station is fixed, but this demand could be affected by electricity prices [107, 58, 170]. We also assume the price functions are linear; future work may extend this to generic nonlinear functions. Furthermore, there are additional operational constraints such as limited capacity of the energy infrastructure and bounds on charging rates of EVs that could be accounted for while computing the equilibrium.

Chapter 7

Conclusion

This dissertation has examined how game theory, mechanism design, statistics, and optimization can be combined to address emerging challenges in digital markets, data governance, and energy systems populated by autonomous, strategic agents. Across the different chapters, we developed new mechanisms, algorithms, and models that (i) explicitly account for strategic behavior and information structure, and (ii) provide rigorous performance guarantees, and relate them to governance outcomes. Together, these contributions advance the design of systems in which algorithmic decision-makers interact with each other and with human stakeholders in complex, data-rich environments.

Each chapter of this dissertation also opens directions for future research. In auction design, a natural extension of V-PoP is to settings with heterogeneous items or online bidder arrivals, where the Item Split Oracle must be redesigned, and collusion detection must operate in an online fashion. For the trilevel optimization algorithm, it would be natural to extend the framework beyond quadratic objectives and convex feasible regions at each level, for example, by tackling nonlinear or nonconvex structures via carefully designed convex approximations. For data economics, characterizing how platform competition affects dividend policy and how dividends should be shared among various heterogeneous user populations are interesting questions. Finally, for EV charging coordination, extending the coalition–Nash framework to dynamic settings with multiple coalitions over time, and to markets with nonlinear pricing, would bring the model closer to real-world grid management.

More broadly, the intersection of mechanism design and machine learning offers a rich frontier for future research. As AI agents increasingly participate in markets through pricing goods, bidding in auctions, and making allocation decisions, the robustness of classical mechanisms to adaptive, learning-based strategies must be re-examined. The frameworks developed in this dissertation provide a foundation for this line of investigation, and we hope they will be useful both as standalone contributions and as building blocks for future work on the design of efficient, equitable, and trustworthy algorithmic systems.

Bibliography

[1] Evrim Acar et al. "Scalable Tensor Factorizations for Incomplete Data". In: *Chemometrics and Intelligent Laboratory Systems* 106.1 (2011), pp. 41–56.

[2] Priyank Agrawal et al. "Learning-Augmented Mechanism Design: Leveraging Predictions for Facility Location". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22. Association for Computing Machinery, 2022, pp. 497–528.

[3] Alessandro Alessio and Alberto Bemporad. "A Survey on Explicit Model Predictive Control". In: *Nonlinear Model Predictive Control: Towards New Challenging Applications*. Ed. by Lalo Magni, Davide Martino Raimondo, and Frank Allgöwer. Berlin, Heidelberg: Springer, 2009, pp. 345–369.

[4] Polina Alexeenko and Eilyan Bitar. "Achieving Reliable Coordination of Residential Plug-In Electric Vehicle Charging: A Pilot Study". In: *Transportation Research Part D: Transport and Environment* 118 (2023), p. 103658.

[5] Electrify America. *How Electric Vehicle (EV) Charging Works*. 2021.

[6] Nir Andelman, Michal Feldman, and Yishay Mansour. "Strong Price of Anarchy". In: *Games and Economic Behavior* 65.2 (2009), pp. 289–317.

[7] Stephanie Assad et al. "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market". In: *Journal of Political Economy* 132.3 (Mar. 2024), pp. 723–771.

[8] Anil Aswani. "Low-rank approximation and completion of positive tensors". In: *SIMAX* 37.3 (2016), pp. 1337–1364.

[9] Robert J. Aumann. "Acceptable Points in General Cooperative N-Person Games". In: *Contributions to the Theory of Games* 4.40 (1959), pp. 287–324.

[10] Australian Associated Press. "Google to Pay $60m Fine for Misleading Australians About Collecting Location Data". In: *The Guardian* (Aug. 2022).

[11] Yoram Bachrach, Morteza Zadimoghaddam, and Peter Key. "A cooperative approach to collusion in auctions". In: *ACM SIGecom Exchanges* 10.1 (2011), pp. 17–22.

[12] Yoram Bachrach et al. "Strong Price of Anarchy, Utility Games and Coalitional Dynamics". In: *Algorithmic Game Theory: 7th International Symposium, SAGT 2014, Haifa, Israel, September 30–October 2, 2014. Proceedings 7*. Springer. 2014, pp. 218–230.

[13] M. J. Bagajewicz and Vassilis Manousiouthakis. "On the Generalized Benders Decomposition". In: *Computers & Chemical Engineering* 15.10 (Oct. 1991), pp. 691–700.

[14] Maria-Florina Balcan, Siddharth Prasad, and Tuomas Sandholm. "Bicriteria multidimensional mechanism design with side information". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2023.

[15] Eric Balkanski, Vasilis Gkatzelis, and Xizhi Tan. "Strategyproof Scheduling with Predictions". In: *14th Innovations in Theoretical Computer Science Conference (ITCS)* 251 (2023).

[16] Boaz Barak and Ankur Moitra. "Noisy tensor completion via the sum-of-squares hierarchy". In: *COLT*. PMLR. 2016, pp. 417–445.

[17] Jonathan F. Bard. "An Investigation of the Linear Three Level Programming Problem". In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-14.5 (Sept. 1984), pp. 711–717.

[18] Peter Bartlett and Shahar Mendelson. "Rademacher and Gaussian Complexities: Risk Bounds and Structural Results". In: *J. Mach. Learn. Res.* (2002).

[19] Robert Bartlett et al. "Consumer-Lending Discrimination in the FinTech Era". In: *Journal of Financial Economics* 143.1 (Jan. 2022), pp. 30–56.

[20] Tamer Basar. "On the Relative Leadership Property of Stackelberg Strategies". In: *Journal of Optimization Theory and Applications* 11.6 (June 1973), pp. 655–661.

[21] Ranojoy Basu and Conan Mukherjee. "Strategy-Proof Multidimensional Mechanism Design". In: *Mathematics of Operations Research* 49.4 (2024), pp. 2768–2785.

[22] Eric Bax. "Computing a Data Dividend". In: *arXiv* (2019).

[23] Eric Bax et al. "Data Consortia". In: *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 2*. Cham: Springer International Publishing, 2021, pp. 489–498.

[24] Juan Andrés Bazerque, Gonzalo Mateos, and Georgios B. Giannakis. "Rank Regularization and Bayesian Inference for Tensor Completion and Extrapolation". In: *IEEE Transactions on Signal Processing* 61.22 (2013), pp. 5689–5703.

[25] NEF Bloomberg. "Hitting the EV Inflection Point". In: *Electric Vehicle Price Parity Phasing Out Combustion Vehicle Sales in Europe; Bloomberg: New York, NY, USA* (2021).

[26] Jérôme Bolte et al. *Bilevel Gradient Methods and the Morse Parametric Qualification Condition*. Mar. 2026.

[27] Suresh Bolusani and Ted K. Ralphs. "A Framework for Generalized Benders' Decomposition and Its Application to Multilevel Optimization". In: *Mathematical Programming* 196.1 (Nov. 2022), pp. 389–426.

[28] Trevor Bonjour, Vaneet Aggarwal, and Bharat Bhargava. "Information theoretic approach to detect collusion in multi-agent games". In: *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*. Vol. 180. PMLR, 2022, pp. 223–232.

[29] Gábor Braun et al. "Blended conditonal gradients". In: *ICML*. PMLR. 2019, pp. 735–743.

[30] Abby Brown et al. *Electric Vehicle Charging Infrastructure Trends from the Alternative Fueling Station Locator: Third Quarter 2023*. Tech. rep. Golden, CO, United States: National Renewable Energy Laboratory (NREL), 2024.

[31] Caleb Bugg, Chen Chen, and Anil Aswani. "Nonnegative Tensor Completion via Integer Optimization". In: *NeurIPS* 35 (2022), pp. 10008–10020.

[32] Jeremy Bulow and Paul Klemperer. "Auctions versus Negotiations". In: *American Economic Review* 86.1 (1996), pp. 180–94.

[33] Changxiao Cai et al. "Nonconvex Low-Rank Tensor Completion from Noisy Data". In: *NeurIPS* 32 (2019).

[34] Emilio Calvano et al. "Artificial Intelligence, Algorithmic Pricing and Collusion". In: *SSRN Electronic Journal* (2018).

[35] Venkat Chandrasekaran et al. "The convex geometry of linear inverse problems". In: *Foundations of Computational Mathematics* 12.6 (2012), pp. 805–849.

[36] Sylvain Chassang et al. "Robust Screens for Noncompetitive Bidding in Procurement Auctions". In: *Econometrica* 90.1 (2022), pp. 315–346.

[37] Xin Chen et al. "Tensor Completion via Integer Optimization". In: *2025 IEEE 64th Conference on Decision and Control (CDC)*. 2025.

[38] Yukun Cheng, Wei Yu, and Guochuan Zhang. "Strategy-proof approximation mechanisms for an obnoxious facility game on networks". In: *Theoretical Computer Science*. Combinatorial Algorithms and Applications 497 (2013), pp. 154–163.

[39] Steve Chien and Alistair Sinclair. "Strong and Pareto Price of Anarchy in Congestion Games". In: *International Colloquium on Automata, Languages, and Programming*. Springer. 2009, pp. 279–291.

[40] Daranith Choeum and Dae-Hyun Choi. "Trilevel Smart Meter Hardening Strategy for Mitigating Cyber Attacks Against Volt/VAR Optimization in Smart Power Distribution Systems". In: *Applied Energy* 304 (2021), p. 117710.

[41] Edward Clarke. “Multipart pricing of public goods”. In: *Public Choice* 11.1 (1971), pp. 17–33.

[42] Richard Cole and Tim Roughgarden. “The sample complexity of revenue maximization”. In: *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. ACM, May 2014, pp. 243–252.

[43] Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. *Integer Programming*. Vol. 271. Springer International Publishing, 2014.

[44] Timothy G. Conley and Francesco Decarolis. “Detecting Bidders Groups in Collusive Auctions”. In: *American Economic Journal: Microeconomics* 8.2 (May 2016), pp. 1–38.

[45] Herbert David and Haikady Nagaraja. *Order Statistics*. 1st ed. Wiley Series in Probability and Statistics. Wiley, 2003. ISBN: 9780471389262 9780471722168.

[46] Robert Day and Paul Milgrom. “Core-selecting Package Auctions”. In: *International Journal of Game Theory* 36.3 (2008), pp. 393–407.

[47] Stephan Dempe and Alain B. Zemkoho. “On the Karush–Kuhn–Tucker Reformulation of the Bilevel Optimization Problem”. In: *Nonlinear Analysis: Theory, Methods & Applications* 75.3 (Feb. 2012), pp. 1202–1218.

[48] Stephan Dempe and Alain B. Zemkoho. “The Bilevel Programming Problem: Reformulations, Constraint Qualifications and Optimality Conditions”. In: *Mathematical Programming* 138.1 (2013), pp. 447–473.

[49] Luca Deori, Kostas Margellos, and Maria Prandini. “On the Connection Between Nash Equilibria and Social Optima in Electric Vehicle Charging Control Games”. In: *IFAC-PapersOnLine* 50.1 (2017), pp. 14320–14325.

[50] Nikhil Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. “The sample complexity of auctions with side information”. In: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. STOC ’16. Association for Computing Machinery, 2016, pp. 426–439.

[51] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. “Revenue maximization with a single sample”. In: *Games and Economic Behavior* 91 (May 2015), pp. 318–333.

[52] Roy Dong et al. “Quantifying the Utility–Privacy Tradeoff in the Internet of Things”. In: *ACM Transactions on Cyber-Physical Systems* 2.2 (Apr. 2018), pp. 1–28.

[53] RF Drenick. “Multilinear programming: Duality theories”. In: *JOTA* 72.3 (1992), pp. 459–486.

[54] Marco F. Duarte and Richard G. Baraniuk. “Kronecker Compressive Sensing”. In: *IEEE Transactions on Image Processing* 21.2 (2012), pp. 494–504.

[55] Shaddin Dughmi, Tim Roughgarden, and Mukund Sundararajan. “Revenue Submodularity”. In: *Theory of Computing* 8 (2012), pp. 95–119.

[56] Adam N. Elmachtoub and Paul Grigas. “Smart Predict, Then Optimize”. In: *Management Science* 68.1 (2022).

[57] Amir Epstein, Michal Feldman, and Yishay Mansour. “Strong Equilibrium in Cost Sharing Connection Games”. In: *Proceedings of the 8th ACM Conference on Electronic Commerce*. 2007, pp. 84–92.

[58] J. Joaquin Escudero-Garzas and Gonzalo Seco-Granados. “Charging Station Selection Optimization for Plug-In Electric Vehicles: An Oligopolistic Game-Theoretic Framework”. In: *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. IEEE. 2012.

[59] EVgo. *What is EVgo — Electric Vehicle (EV) Fast Charging Stations*. 2021.

[60] Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

[61] Nuno P. Faísca et al. “A Multi-Parametric Programming Approach for Multilevel Hierarchical and Decentralised Optimisation Problems”. In: *Computational Management Science* 6.4 (Oct. 2009), pp. 377–397.

[62] Ramy Fakhry et al. “Tri-Level Mixed-Binary Linear Programming: Solution Approaches and Application in Defending Critical Infrastructure”. In: *European Journal of Operational Research* 298.3 (May 2022), pp. 1114–1131.

[63] Amir Mohammad Fathollahi Fard and Mostafa Hajaghaei-Keshteli. “A Tri-Level Location-Allocation Model for Forward/Reverse Supply Chain”. In: *Applied Soft Computing* 62 (2018), pp. 328–346.

[64] Michal Feldman and Ophir Friedler. “A Unified Framework for Strong Price of Anarchy in Clustering Games”. In: *International Colloquium on Automata, Languages, and Programming*. Springer. 2015, pp. 601–613.

[65] Bryce L. Ferguson et al. “Collaborative Coalitions in Multi-Agent Systems: Quantifying the Strong Price of Anarchy for Resource Allocation Games”. In: *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE. 2023, pp. 3238–3243.

[66] Shmuel Friedland and Lek-Heng Lim. “Nuclear norm of higher-order tensors”. In: *Mathematics of Computation* 87.311 (2017).

[67] Shmuel Friedland and Lek-Heng Lim. “The computational complexity of duality”. In: *SIOPT* 26.4 (2016), pp. 2378–2393.

[68] Zhengtang Fu, Peiwu Dong, and Yanbing Ju. “An Intelligent Electric Vehicle Charging System for New Energy Companies Based on Consortium Blockchain”. In: *Journal of Cleaner Production* 261 (2020), p. 121219.

[69] Lingwen Gan, Ufuk Topcu, and Steven H. Low. “Optimal Decentralized Protocol for Electric Vehicle Charging”. In: *IEEE Transactions on Power Systems* 28.2 (2012), pp. 940–951.

[70] Silvia Gandy, Benjamin Recht, and Isao Yamada. “Tensor completion and low-n-rank tensor recovery via convex optimization”. In: *Inverse problems* 27.2 (2011), p. 025010.

[71] Manuel J. García Rodríguez et al. "Collusion detection in public procurement auctions with machine learning algorithms". In: *Automation in Construction* 133 (2022).

[72] Nicola Gatti, Marco Rocco, and Tuomas Sandholm. "On the Verification and Computation of Strong Nash Equilibrium". In: *arXiv preprint arXiv:1711.06318* (2017).

[73] Hancheng Ge, James Caverlee, and Haokai Lu. "TAPER: A Contextual Tensor-Based Approach for Personalized Expert Recommendation". In: *RecSys '16*. 2016, pp. 261–268.

[74] A. M. Geoffrion. "Generalized Benders Decomposition". In: *Journal of Optimization Theory and Applications* 10.4 (Oct. 1972), pp. 237–260.

[75] Nafiseh Ghorbani-Renani et al. "Protection-Interdiction-Restoration: Tri-Level Optimization for Enhancing Interdependent Network Resilience". In: *Reliability Engineering & System Safety* 199 (2020), p. 106907.

[76] Jennifer Golbeck and Matthew Mauriello. "User Perception of Facebook App Data Access: A Comparison of Methods and Privacy Concerns". In: *Future Internet* 8.2 (Mar. 2016), p. 9.

[77] Andrew Goldberg and Jason Hartline. "Collusion-Resistant Mechanisms for Single-Parameter Agents". In: *Proc. 16th Symp. on Discrete Alg.* ACM/SIAM, May 2004, p. 15.

[78] Theodore Groves. "Incentives in Teams". In: *Econometrica* 41.4 (1973), p. 617.

[79] Giambattista Gruosso. "Analysis of Impact of Electrical Vehicle Charging on Low Voltage Power Grid". In: *2016 International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC)*. IEEE. 2016, pp. 1–6.

[80] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2021.

[81] Joseph E Harrington. "Developing Competition Law for Collusion by Autonomous Artificial Agents". In: *Journal of Competition Law & Economics* 14.3 (2018), pp. 331–363.

[82] Jason D. Hartline, Sheng Long, and Chenhao Zhang. "Regulation of Algorithmic Collusion". In: *Proceedings of the Symposium on Computer Science and Law*. ACM, Mar. 2024, pp. 98–108.

[83] Jason D. Hartline and Tim Roughgarden. "Simple versus optimal mechanisms". In: *Proceedings of the 10th ACM Conference on Electronic Commerce*. EC '09. 2009, pp. 225–234.

[84] Julien M. Hendrickx and Alex Olshevsky. "Matrix p -Norms Are NP-Hard to Approximate If $p \neq 1, 2, \infty$". In: *SIMAX* 31.5 (2010), pp. 2802–2812. DOI: `10.1137/09076773X`.

[85] Christopher J. Hillar and Lek-Heng Lim. "Most Tensor Problems Are NP-Hard". In: *Journal of the ACM* 60.6 (Nov. 2013), 45:1–45:39.

[86] Zhanxuan Hu et al. "Low Rank Regularization: A review". In: *Neural Networks* 136 (2021), pp. 218–232.

[87] Martin Jaggi. "Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization". In: *ICML*. 2013, pp. 427–435.

[88] Swayambhoo Jain, Alexander Gutierrez, and Jarvis Haupt. "Noisy Tensor Completion for Tensors with a Sparse Canonical Polyadic Factor". In: *IEEE International Symposium on Information Theory*. 2017, pp. 2153–2157.

[89] Sajid Javed, Thierry Bouwmans, and Soon Ki Jung. "Stochastic Decomposition into Low Rank and Sparse Tensor for Robust Background Subtraction". In: *ICDP-15*. 2015, pp. 1–6.

[90] Robert G. Jeroslow. "The Polynomial Hierarchy and a Simple Model for Competitive Analysis". In: *Mathematical Programming* 32.2 (June 1985), pp. 146–164.

[91] Yang Jiao, Kai Yang, and Chengtao Jian. "DTZO: Distributed Trilevel Zeroth Order Learning with Provable Non-Asymptotic Convergence". In: *Proceedings of the 42nd International Conference on Machine Learning*. PMLR, Oct. 2025, pp. 27873–27916.

[92] Ruben Juarez. "Group strategyproof cost sharing: The role of indifferences". In: *Games and Economic Behavior* 82 (2013), pp. 218–239.

[93] Alexandros Karatzoglou et al. "Multiverse Recommendation: N-Dimensional Tensor Factorization for Context-Aware Collaborative Filtering". In: *RecSys '10*. 2010, pp. 79–86.

[94] Abay Molla Kassa and Semu Mitiku Kassa. "Deterministic Solution Approach for Some Classes of Nonlinear Multilevel Programs with Multiple Followers". In: *Journal of Global Optimization* 68.4 (Aug. 2017), pp. 729–747.

[95] Çağatay Kayı and Eve Ramaekers. "Characterizations of Pareto-efficient, fair, and strategy-proof allocation rules in queueing problems". In: *Games and Economic Behavior* 68.1 (2010), pp. 220–232.

[96] Wajahat Khan, Furkan Ahmad, and Mohammad Saad Alam. "Fast EV Charging Station Integration with Grid Ensuring Optimal and Quality Power Exchange". In: *Engineering Science and Technology, an International Journal* 22.1 (2019), pp. 143–152.

[97] Jip Kim et al. "Strategic Policymaking for Implementing Renewable Portfolio Standards: A Tri-Level Optimization Approach". In: *IEEE Transactions on Power Systems* 36.6 (2021), pp. 4915–4927.

[98] Thomas Kleinert et al. "A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization". In: *EURO Journal on Computational Optimization* 9 (2021), p. 100007.

[99] Tamara Kolda and Brett Bader. "Tensor Decompositions and Applications". In: *SIAM Review* 51.3 (2009), pp. 455–500.

[100] Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. "Low-Rank Tensor Completion by Riemannian Optimization". In: *BIT Numerical Mathematics* 54.2 (2014), pp. 447–468.

[101] Vijay Krishna and Motty Perry. "Efficient Mechanism Design". In: *SSRN Electronic Journal* (1997).

[102] Sukanya Kudva and Anil Aswani. "When Would Online Platforms Pay Data Dividends?" In: *IEEE Control Systems Letters* (2023).

[103] Sukanya Kudva, Edward Dowling, and Anil Aswani. "Collusion-proof Auction Design using Side Information". Preprint.

[104] Sukanya Kudva et al. "Understanding the Impact of Coalitions between EV Charging Stations". In: *2024 IEEE 63rd Conference on Decision and Control (CDC)*. 2024.

[105] Guillaume Lecué. "Empirical risk minimization is optimal for the convex aggregation problem". In: *Bernoulli* 19.5B (2013), pp. 2153–2166.

[106] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

[107] Woongsup Lee et al. "Electric Vehicle Charging Stations with Renewable Power Generators: A Game Theoretical Analysis". In: *IEEE Transactions on Smart Grid* 6.2 (2014), pp. 608–617.

[108] Yoon Lee and Anil Aswani. "Optimally Designing Cybersecurity Insurance Contracts to Encourage the Sharing of Medical Data". In: *2022 IEEE 61st Conference on Decision and Control (CDC)*. 2022, pp. 6782–6787.

[109] Mark Lewis and Fred Glover. "Quadratic unconstrained binary optimization problem preprocessing: Theory and empirical analysis". In: *Networks* 70.2 (2017), pp. 79–97.

[110] Jingqi Li et al. "The Computation of Approximate Feedback Stackelberg Equilibria in Multiplayer Nonlinear Constrained Dynamic Games". In: *SIAM Journal on Optimization* 34.4 (Dec. 2024), pp. 3723–3749.

[111] Ji Liu et al. "Tensor completion for estimating missing values in visual data". In: *IEEE transactions on pattern analysis and machine intelligence* 35.1 (2012), pp. 208–220.

[112] Yuanyuan Liu et al. "Trace norm regularized CANDECOMP/PARAFAC decomposition with missing data". In: *IEEE Trans. Cybern.* 45.11 (2014), pp. 2437–2448.

[113] Jie Lu et al. "Multifollower Trilevel Decision Making Models and System". In: *IEEE Transactions on Industrial Informatics* 8.4 (2012), pp. 974–985.

[114] Jie Lu et al. "Multilevel Decision-Making: A Survey". In: *Information Sciences* 346–347 (June 2016), pp. 463–487.

[115] Marcos S. Lyra et al. "Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods". In: *Applied Network Science* 7.1 (Dec. 2022), p. 83.

[116] Zhongjing Ma, Duncan S. Callaway, and Ian A. Hiskens. "Decentralized Charging Control of Large Populations of Plug-In Electric Vehicles". In: *IEEE Transactions on Control Systems Technology* 21.1 (2011), pp. 67–78.

[117] George J Mailath and Peter Zemsky. "Collusion in second price auctions with heterogeneous bidders". In: *Games and Economic Behavior* 3.4 (1991), pp. 467–486.

[118] Pascal Massart. "Some Applications of Concentration Inequalities to Statistics". In: *Annales de la faculté des sciences de Toulouse Sér. 6* 9.2 (2000), pp. 245–303.

[119] Manipushpak Mitra and Suresh Mutuswami. "Group strategyproofness in queueing models". In: *Games and Economic Behavior* 72.1 (May 2011), pp. 242–254.

[120] Alexander Mitsos. "Global Solution of Nonlinear Mixed-Integer Bilevel Programs". In: *Journal of Global Optimization*. 2010.

[121] Andrea Montanari and Nike Sun. "Spectral Algorithms for Tensor Completion". In: *Communications on Pure and Applied Mathematics* 71.11 (2018), pp. 2381–2425.

[122] Liangyuan Na et al. "Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets from Which Protected Health Information Has Been Removed with Use of Machine Learning". In: *JAMA Network Open* 1.8 (Dec. 2018), e186040.

[123] Carmeliza Navasca, Lieven De Lathauwer, and Stefan Kindermann. "Swamp reducing technique for tensor decomposition". In: *2008 16th European Signal Processing Conference*. IEEE. 2008, pp. 1–5.

[124] Arkadi Nemirovski. "Topics in non-parametric statistics". In: *Ecole d'Eté de Probabilités de Saint-Flour* 28 (2000), p. 85.

[125] Rabia Nessah and Guoqiang Tian. "On the Existence of Strong Nash Equilibria". In: *Journal of Mathematical Analysis and Applications* 414.2 (2014), pp. 871–885.

[126] Hassan Obeid et al. "Learning and Optimizing Charging Behavior at PEV Charging Stations: Randomized Pricing Experiments, and Joint Power and Price Optimization". In: *Applied Energy* 351 (2023), p. 121862.

[127] Dario Paccagnan, Francesca Parise, and John Lygeros. "On the Efficiency of Nash Equilibria in Aggregative Charging Games". In: *IEEE Control Systems Letters* 2.4 (2018), pp. 629–634.

[128] M. Pal and E. Tardos. "Group strategy proof mechanisms via primal-dual algorithms". In: *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* IEEE Computer. Soc, 2003, pp. 584–593.

[129] Wenhao Pan, Anil Aswani, and Chen Chen. "Accelerated non-negative tensor completion via integer programming". In: *Frontiers in Applied Mathematics and Statistics* 9 (2023).

[130] Charith Perera et al. "Big Data Privacy in the Internet of Things Era". In: *IT Professional* 17.3 (May 2015), pp. 32–39.

[131] Efstratios N. Pistikopoulos et al. “Theoretical and Algorithmic Advances in Multi-Parametric Programming and Control”. In: *Computational Management Science* 9.2 (May 2012), pp. 183–203.

[132] Yang Qi, Pierre Comon, and Lek-Heng Lim. “Uniqueness of Nonnegative Tensor Approximations”. In: *IEEE Transactions on Information Theory* 62.4 (2016), pp. 2170–2183.

[133] Jairo Quiros-Tortos, Luis Ochoa, and Timothy Butler. “How Electric Vehicles and the Grid Work Together: Lessons Learned from One of the Largest Electric Vehicle Trials in the World”. In: *IEEE Power and Energy Magazine* 16.6 (2018), pp. 64–76.

[134] Nikhil Rao, Parikshit Shah, and Stephen Wright. “Forward–Backward Greedy Algorithms for Atomic Norm Regularization”. In: *IEEE Transactions on Signal Processing* 63.21 (2015), pp. 5798–5811.

[135] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016*. June 2018.

[136] Marc S. Robinson. “Collusion and the Choice of Auction”. In: *The RAND Journal of Economics* 16.1 (1985), p. 141.

[137] R. Tyrrell Rockafellar and Roger Wets. *Variational Analysis*. Springer, 2009.

[138] Ralph Tyrrell Rockafellar. “Convex Analysis: (PMS-28)”. In: *Convex Analysis*. Princeton University Press, Apr. 2015.

[139] Jiří Rohn. “Computing the norm $\|A\|_{\infty,1}$ is NP-hard”. In: *Linear and Multilinear Algebra* 47.3 (2000), pp. 195–204. DOI: 10.1080/03081080008818644.

[140] Ryo Sato, Mirai Tanaka, and Akiko Takeda. “A Gradient Method for Multilevel Optimization”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 7522–7533.

[141] Kathrin Schacke. “On the Kronecker Product”. In: *Master's thesis, University of Waterloo* (2004).

[142] Allahkaram Shafiei, Vyacheslav Kungurtsev, and Jakub Marecek. “Trilevel and Multilevel Optimization Using Monotone Operator Theory”. In: *Mathematical Methods of Operations Research* 99.1 (Apr. 2024), pp. 77–114.

[143] Itai Sher. “Optimal shill bidding in the VCG mechanism”. In: *Economic Theory* 50.2 (June 2012), pp. 341–387.

[144] Marco Signoretto et al. “Tensor Versus Matrix Completion: A Comparison with Application to Spectral Data”. In: *IEEE Signal Processing Letters* 18.7 (2011), pp. 403–406.

[145] Vin de Silva and Lek-Heng Lim. “Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem”. In: *SIAM Journal on Matrix Analysis and Applications* 30.3 (2008), pp. 1084–1127.

[146] M. Simaan. "Equilibrium Properties of the Nash and Stackelberg Strategies". In: *Automatica* 13.6 (Nov. 1977), pp. 635–636.

[147] M. Simaan and J. B. Cruz. "On the Stackelberg Strategy in Nonzero-Sum Games". In: *Journal of Optimization Theory and Applications* 11.5 (May 1973), pp. 533–555.

[148] Alexander Slagg. *What Oregon's Move to Redefine Data Privacy Means for PHI*. Feb. 2019.

[149] Qingquan Song et al. "Tensor completion algorithms in big data analytics". In: *ACM TKDD* 13.1 (2019), pp. 1–48.

[150] Nathan Srebro, Karthik Sridharan, and Ambuj Tewari. "Smoothness, Low Noise and Fast Rates". In: *NeurIPS*. 2010, pp. 2199–2207.

[151] Yi Su and Liyin Jin. "The Impact of Online Platforms' Revenue Model on Consumers' Ethical Inferences". In: *Journal of Business Ethics* 178.2 (June 2022), pp. 555–569.

[152] Huachun Tan et al. "Low-n-Rank Tensor Recovery Based on Multi-Linear Augmented Lagrange Multiplier Method". In: *Neurocomputing* 119 (2013), pp. 144–152.

[153] Huachun Tan et al. "Short-Term Traffic Prediction Based on Dynamic Tensor Completion". In: *IEEE Transactions on Intelligent Transportation Systems* 17.8 (Aug. 2016), pp. 2123–2133.

[154] Pingzhong Tang, Dingli Yu, and Shengyu Zhao. "Characterization of Group strategyproof Mechanisms for Facility Location in Strictly Convex Space". In: *Proceedings of the 21st ACM Conference on Economics and Computation*. ACM, 2020, pp. 133–157.

[155] Ahmad Tavakoli et al. "Impacts of grid integration of solar PV and electric vehicle on grid stability, power quality and energy economics: a review". In: *IET Energy Systems Integration* 2.3 (Sept. 2020), pp. 243–260.

[156] Tesla. *Introducing V3 Supercharging*. 2019.

[157] Giorgio Tomasi and Rasmus Bro. "PARAFAC and Missing Values". In: *Chemometrics and Intelligent Laboratory Systems* 75.2 (2005), pp. 163–180.

[158] Petter Tøndel, Tor Arne Johansen, and Alberto Bemporad. "An Algorithm for Multi-Parametric Quadratic Programming and Explicit MPC Solutions". In: *Automatica* 39.3 (Mar. 2003), pp. 489–497.

[159] Rohit Trivedi et al. "Comprehensive Dataset on Electrical Load Profiles for Energy Community in Ireland". In: *Scientific Data* 11.1 (June 2024), p. 621.

[160] Hayley Tsukayama. *Why Getting Paid for Your Data Is a Bad Deal*. Oct. 2020.

[161] Alexandre B Tsybakov. "Optimal rates of aggregation". In: *Learning theory and kernel machines*. Springer, 2003, pp. 303–313.

[162] William Vickrey. "Counterspeculation, Auctions, and Competitive Sealed Tenders". In: *The Journal of Finance* 16.1 (1961), pp. 8–37.

[163] John Voelcker. *Porsche Claims It Can Double Tesla's Fast-Charging Rate*. 2021.

[164] Xuan Wu and Antonio J. Conejo. "An Efficient Tri-Level Optimization Model for Electric Grid Defense Planning". In: *IEEE Transactions on Power Systems* 32.4 (July 2017), pp. 2984–2994.

[165] X. *Newsom Wants Companies Collecting Personal Data to Share the Wealth with Californians*. May 2019.

[166] Lan Xia, Kent B. Monroe, and Jennifer L. Cox. "The Price is Unfair! A Conceptual Framework of Price Fairness Perceptions". In: *Journal of Marketing* 68.4 (Oct. 2004), pp. 1–15.

[167] Chenyang Xu and Pinyan Lu. "Mechanism Design with Predictions". In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. 2022, pp. 571–577.

[168] Lu Yu et al. "A Meta-Analysis to Explore Privacy Cognition and Information Disclosure of Internet Users". In: *International Journal of Information Management* 51 (Apr. 2020), p. 102015.

[169] Ming Yuan and Cun-Hui Zhang. "On Tensor Completion via Nuclear Norm Minimization". In: *Foundations of Computational Mathematics* 16.4 (2016), pp. 1031–1068.

[170] Wei Yuan, Jianwei Huang, and Ying Jun Angela Zhang. "Competitive Charging Station Pricing for Plug-In Electric Vehicles". In: *IEEE Transactions on Smart Grid* 8.2 (2015), pp. 627–639.

[171] Rehman Zafar et al. "Prosumer based energy management and sharing in smart grid". In: *Renew. Sustain. Energy Rev.* 82 (2018), pp. 1675–1684.

[172] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. *Social Media Mining: An Introduction*. 1st ed. Cambridge University Press, Apr. 2014.

[173] Zihan Zhang et al. "Tensor-Based Temporal Control for Partially Observed High-Dimensional Streaming Data". In: *Technometrics* 66.2 (Apr. 2024), pp. 227–239.

Appendix A

Appendix for Chapter 2

The omitted proofs from Chapter 2 are provided here.

Proof of Lemma 2.3.1

First, we characterize the colluding bidders' optimal strategy in the following lemma.

Lemma A.0.1 (Colluders' Strategy in VCG)**.** *Since the VCG mechanism is DSIC for non-colluding bidders, their bids $\mathbf{b}^{\mathrm{N}}$ can be assumed to be their true valuations $\mathbf{v}^{\mathrm{N}}$. The colluding bidders' optimal strategy to maximize their joint utility is as follows:*

1. *Given that a fixed number of items r_{C} are allotted to colluding bidders (and the remaining $r_{\mathrm{N}} = r - r_{\mathrm{C}}$ items are allotted to non-colluding bidders), their optimal bids are $b_i^{\mathrm{C}} = \max\{v_i^{\mathrm{C}}, b_{r_{\mathrm{N}}+1}^{\mathrm{N}} + \epsilon\}$ $\forall i \in \{1, \cdots, r_{\mathrm{C}}\}$, and $b_i^{\mathrm{C}} = 0$ $\forall i \in \{r_{\mathrm{C}} + 1, \cdots, C\}$, for some small $\epsilon > 0$. Here, $b_{r_{\mathrm{N}}+1}^{\mathrm{N}}$ is the largest losing bid among non-colluding bidders.*

   Define $u_{\mathrm{C}}(\mathbf{v}^{\mathrm{N}}; r_{\mathrm{C}})$ to be the optimal utility achieved through this strategy. The maximum utility is:

   $$u_{\mathrm{C}}(\mathbf{v}^{\mathrm{N}}; r_{\mathrm{C}}) := \max_{\mathbf{b}^{\mathrm{C}}} \{u_{\mathrm{C}}(\mathbf{b}^{\mathrm{C}}, \mathbf{v}^{\mathrm{N}}) \mid \textstyle\sum_{i\in \mathrm{C}} x_i(\mathbf{b}^{\mathrm{C}}, \mathbf{v}^{\mathrm{N}}) = r_{\mathrm{C}}\} = \textstyle\sum_{i=1}^{r_{\mathrm{C}}} v_i^{\mathrm{C}} - r_{\mathrm{C}} \cdot b_{r_{\mathrm{N}}+1}^{\mathrm{N}}.$$

2. *Let r_{C}^* and r_{N}^* be the VCG allocation to colluding and non-colluding bidders, respectively, in the absence of collusion. Even when manipulating their bids, the colluding bidders do not desire more than r_{C}^* items. That is, $u_{\mathrm{C}}(\mathbf{v}^{\mathrm{N}}; r_{\mathrm{C}}^*) \geqslant u_{\mathrm{C}}(\mathbf{v}^{\mathrm{N}}; r_{\mathrm{C}}^* + \Delta r)$, $\forall \Delta r > 0$.*

Proof. The colluders maximize $u_{\mathrm{C}}(b^{\mathrm{C}}, v^{\mathrm{N}}) = \sum_{i\in \mathrm{C}} v_i x_i - p_i$. For $\sum_{i\in \mathrm{C}} x_i = r_{\mathrm{C}}$, the r_{C} largest colluder bids win. The VCG payment p_i is $\max\{b_{r_{\mathrm{C}}+1}^{\mathrm{C}}, b_{r_{\mathrm{N}}+1}^{\mathrm{N}}\}$.
Optimal Bids. The choice $b_i^{\mathrm{C}} = 0$ for $i \in \{r_{\mathrm{C}} + 1, \ldots, C\}$ minimizes the payment, since $b_{r_{\mathrm{C}}+1}^{\mathrm{C}} = 0$. The winning bids $b_i^{\mathrm{C}} = \max\{v_i^{\mathrm{C}}, b_{r_{\mathrm{N}}+1}^{\mathrm{N}} + \epsilon\}$ for $i \in \{1, \ldots, r_{\mathrm{C}}\}$ ensure allocation and $b_i^{\mathrm{C}} > b_{r_{\mathrm{N}}+1}^{\mathrm{N}}$ for small $\epsilon > 0$. Since payment $p_i = b_{r_{\mathrm{N}}+1}^{\mathrm{N}}$, the maximum utility for fixed r_{C} is $u_{\mathrm{C}}(b^{\mathrm{C}}, v^{\mathrm{N}}) = \sum_{i=1}^{r_{\mathrm{C}}} v_i^{\mathrm{C}} - r_{\mathrm{C}} \cdot b_{r_{\mathrm{N}}+1}^{\mathrm{N}}$.

Optimal Item Count. The welfare-maximizing allocation is $(r^*_{\mathrm{C}}, r^*_{\mathrm{N}})$. For any $\Delta r > 0$:

$$\begin{aligned} u_{\mathrm{C}}(v^{\mathrm{N}}; r^*_{\mathrm{C}} + \Delta r) - u_{\mathrm{C}}(v^{\mathrm{N}}; r^*_{\mathrm{C}}) &= \textstyle\sum_{i=r^*_{\mathrm{C}}+1}^{r^*_{\mathrm{C}}+\Delta r} v^{\mathrm{C}}_i - (r^*_{\mathrm{C}} + \Delta r)\, b^{\mathrm{N}}_{r^*_{\mathrm{N}}-\Delta r+1} + r^*_{\mathrm{C}} b^{\mathrm{N}}_{r^*_{\mathrm{N}}+1} \\ &= \textstyle\sum_{i=1}^{\Delta r} \left(v^{\mathrm{C}}_{r^*_{\mathrm{C}}+i} - b^{\mathrm{N}}_{r^*_{\mathrm{N}}-\Delta r+1}\right) + r^*_{\mathrm{C}} \left(b^{\mathrm{N}}_{r^*_{\mathrm{N}}+1} - b^{\mathrm{N}}_{r^*_{\mathrm{N}}-\Delta r+1}\right) \\ &\leqslant 0. \end{aligned}$$

This holds because: (1) $v^{\mathrm{C}}_{r^*_{\mathrm{C}}+i} \leqslant b^{\mathrm{N}}_{r^*_{\mathrm{N}}}$ for $i \geqslant 1$, as these are items the colluders didn't win in the truthful VCG allocation, and $b^{\mathrm{N}}_{r^*_{\mathrm{N}}-\Delta r+1} \geqslant b^{\mathrm{N}}_{r^*_{\mathrm{N}}}$; thus $v^{\mathrm{C}}_{r^*_{\mathrm{C}}+i} \leqslant b^{\mathrm{N}}_{r^*_{\mathrm{N}}-\Delta r+1}$. (2) b^{N} is sorted in descending order, and $b^{\mathrm{N}}_{r^*_{\mathrm{N}}+1} \leqslant b^{\mathrm{N}}_{r^*_{\mathrm{N}}-\Delta r+1}$ since $\Delta r > 0$. Both parts of the sum are non-positive. □

Next, we prove the main results about equilibirum allocation, welfare and revenue for VCG with collusion.

Allocation and Bids. By Lemma A.0.1, $r^{Col}_{\mathrm{C}} \leqslant r^*_{\mathrm{C}}$. Hence, $r^{Col}_{\mathrm{N}} \geqslant r^*_{\mathrm{N}}$ and $v^{\mathrm{C}}_i > b^{\mathrm{N}}_{r^{Col}_{\mathrm{N}}+1}$ (the payment price for winning items). The optimal winning bids are $b^{\mathrm{C},Col}_i = \max\{v^{\mathrm{C}}_i, b^{\mathrm{N}}_{r^{Col}_{\mathrm{N}}+1} + \epsilon\} = v^{\mathrm{C}}_i$ for $i \in \{1, \ldots, r^{Col}_{\mathrm{C}}\}$ and losing bids $b^{\mathrm{C},Col}_i = 0$ for $i \in \{r^{Col}_{\mathrm{C}} + 1, \ldots, C\}$. That is, colluding bidders are incentivized to bid as low as possible to win, and always bid either v^{C}_i or 0 ensuring $b^{\mathrm{C},Col}_i \leqslant v^{\mathrm{C}}_i$ (bid shading).

Utility. Since $r^{Col}_{\mathrm{N}} \geqslant r^*_{\mathrm{N}}$,the non-colluders' allocation is non-decreasing: $x_i(b^{Col}) \geqslant x_i(b^*)$ $\forall i \in \mathrm{N}$. The collusive payment $p(b^{Col}) = b^{\mathrm{N}}_{r^{Col}_{\mathrm{N}}+1}$ and truthful payment $p(b^*) = \max\{v^{\mathrm{C}}_{r^*_{\mathrm{C}}+1}, b^{\mathrm{N}}_{r^*_{\mathrm{N}}+1}\}$. Since $r^{Col}_{\mathrm{N}} \geqslant r^*_{\mathrm{N}}$, $b^{\mathrm{N}}_{r^{Col}_{\mathrm{N}}+1} \leqslant b^{\mathrm{N}}_{r^*_{\mathrm{N}}+1}$, so $p(b^{Col}) \leqslant p(b^*)$.

Non-Colluder Utility: $u_i(b^{Col}) = x_i(b^{Col})v_i - p(b^{Col}) \geqslant x_i(b^*)v_i - p(b^*) = u_i(b^*)$.

Colluder Utility: $u_{\mathrm{C}}(b^{Col}) = u_{\mathrm{C}}(v^{\mathrm{N}}; r^{Col}_{\mathrm{C}}) \geqslant u_{\mathrm{C}}(v^{\mathrm{N}}; r^*_{\mathrm{C}}) = u_{\mathrm{C}}(b^*)$, by definition of r^{Col}_{C}.

Auctioneer Revenue: $u_a(b^{Col}) = r \cdot p(b^{Col}) \leqslant r \cdot p(b^*) = u_a(b^*)$.

Welfare and Revenue. VCG without collusion maximizes social welfare, Welf(b^*). Any deviation, b^{Col}, that results in a different allocation must yield Welf$(b^{Col}) \leqslant$ Welf(b^*). This, combined with Rev$(b^{Col}) \leqslant$ Rev(b^*) (from the auctioneer's utility), confirms that both welfare and revenue are non-increasing.

Proof of Theorem 2.3.3

First, we prove the claim:

$$\mathrm{Welf}_{VCG}(\mathsf{N} \cup \mathsf{C}) \geqslant \mathrm{Welf}_{VCG}(\mathsf{N})$$

From Lemma 2.3.1, when all $\mathsf{N} \cup \mathsf{C}$ bidders participate in a VCG mechanism, the non-colluding bidders bid v^{N} (i.e. truthfully) and the colluding bidders bid $b^{\mathrm{C},Col} \leqslant v^{\mathrm{C}}$ (i.e. either bid-shade to 0 or report their true valuation). The welfare achieved is Welf$(v^{\mathrm{N}}, b^{\mathrm{C},Col})$. Since the VCG mechanism allocates items to the top bidders, when a colluding bidder i bid shades to 0, an item could be allocated to a bidder j who values it less than i leading to a welfare loss. In the worst case, all the colluders bid-shade to 0 achieving a welfare $\text{Welf}(v^{\text{N}}, 0)$. Hence, $\text{Welf}(v^{\text{N}}, b^{\text{C},Col}) \geqslant \text{Welf}(v^{\text{N}}, 0)$.

Now, consider a VCG auction with only the non-colluding bidders N achieving a welfare $\text{Welf}(v^{\text{N}})$. Since the bidders have non-negative valuations, adding $|\mathsf{C}|$ more bids of 0 cannot hurt the welfare, i.e. $\text{Welf}(v^{\text{N}}) \leqslant \text{Welf}(v^{\text{N}}, 0)$. The result follows from $\text{Welf}(v^{\text{N}}, b^{\text{C},Col}) \geqslant \text{Welf}(v^{\text{N}}, 0) \geqslant \text{Welf}(v^{\text{N}})$.

Next, we introduce a set of new bidders $\widetilde{\mathsf{N}}$ (all of whom are assumed to be non-colluding and truthful), such that $\left|\widetilde{\mathsf{N}}\right| = \left|\mathsf{C}\right|$. From the claim we proved earlier, $\text{Wel}_{VCG}(\mathsf{N} \cup \widetilde{\mathsf{N}} \cup \mathsf{C}) \geqslant \text{Wel}_{VCG}(\mathsf{N} \cup \widetilde{\mathsf{N}})$. The key observation is that VCG is a welfare-maximizing auction when bidders are truthful, and it has the largest revenue among the welfare-maximizing auctions. Let us use OPT to denote a welfare-maximizing auction. Hence, $\mathbb{E}[\text{Welf}_{VCG}(\mathsf{N} \cup \widetilde{\mathsf{N}})] = \mathbb{E}[\text{Welf}_{OPT}(\mathsf{N} \cup \widetilde{\mathsf{N}})] \geqslant \mathbb{E}[\text{Welf}_{OPT}(\mathsf{N} \cup \mathsf{C})]$, where the last inequality is because $\left|\widetilde{\mathsf{N}}\right| = \left|\mathsf{C}\right|$ and the true valuations of all bidders come from an i.i.d distribution. Putting them all together, we get $\mathbb{E}\big[\text{Welf}_{VCG}(\mathsf{N} \cup \widetilde{\mathsf{N}} \cup \mathsf{C})\big] \geqslant \mathbb{E}\big[\text{Welf}_{VCG}(\mathsf{N} \cup \widetilde{\mathsf{N}})\big] \geqslant \mathbb{E}\big[\text{Welf}_{OPT}(\mathsf{N} \cup \mathsf{C})\big]$.

Proof of Theorem 2.4.2

Consider bidder $i \in \mathsf{N}$. Phase 1 applies the VCG mechanism to allocate k items among non-colluding bidders, where k is pre-determined and independent of all bids. In VCG, it is a dominant strategy to bid truthfully whenever $p_i(b^{\text{N}}) = b^{\text{N}}_{k+1} x_i(b^{\text{N}})$, and individual rationality also holds [78, 41].

Consider bidder $i \in \mathsf{C}$. Phase 2 allocates the remaining $r - k$ items to colluding bidders via the posted price b^{N}_{k+1}. In our model for colluding bidders, we consider them to be maximizing their total utility $\sum_{i \in \text{C}} u_i(b)$. Since they exchange monetary payments and items amongst themselves, it suffices to consider the top bidding colluders while calculating their net utility. Hence, $\sum_{i \in \text{C}} u_i(b) = \sum_{i=1}^{r-k} \left(v^{\text{C}}_i - b^{\text{N}}_{k+1}\right) 1\{b^{\text{C}}_i > b^{\text{N}}_{k+1}\}$. For fixed b^{N}_{k+1}, this net utility is maximized when $v^{\text{C}}_i > b^{\text{N}}_{k+1} \Rightarrow 1\{b^{\text{C}}_i > b^{\text{N}}_{k+1}\} = 1$. Hence, bidding truthfully, i.e., $v^{\text{C}}_i = b^{\text{C}}_i$ $\forall i \in C$ is a utility-maximizing strategy.

Hence, for every bidder, truthful reporting is a weakly dominant strategy, and the mechanism is (ex-post) DSIC.

Proof of Lemma 2.4.3

First, we calculate the conditional expected welfare, conditioned on all the non-colluding $\bar{V}^{\text{N}}_k$ and note that this is equivalent to conditioning on highest non-colluding losing bid V^{N}_{k+1}. Also, since the colluding bidders can exchange both items and monetary payments among themselves, the specific allocation of items to individual bidders is irrelevant. For welfare computation, it suffices to consider the highest colluding bids among the winners. Using Assumption 2.4.1 and probability integral transform of continuous random variables, $V^{\text{N}}_k = \mathsf{Q}(U^{\text{N}}_k)$ (see Equation (2.3.7) of [45]), where U^{N}_k is $k-$th largest order statistic of N samples of standard uniform random variables. Using $Q(\cdot) = F^{-1}(\cdot)$, we get $F(V^{\text{N}}_k) = U^{\text{N}}_k$. We use these facts to simplify the expression $\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};k)\mid \bar{V}_k^{\mathrm{N}}\right]=\mathbb{E}\left[\sum_{i\in\mathsf{N}\cup\mathsf{C}} V_i x_i(V)\mid V_{k+1}^{\mathrm{N}}\right]$.

$$\begin{aligned}
&\mathbb{E}\big[\textstyle\sum_{i\in\mathsf{N}\cup\mathsf{C}} V_i x_i(V)\mid V_{k+1}^{\mathrm{N}}\big]\\
&=\mathbb{E}\big[\textstyle\sum_{i=1}^{k} V_i^{\mathrm{N}}\mid V_{k+1}^{\mathrm{N}}\big]+\sum_{q=1}^{r-k-1}\mathbb{E}\big[\sum_{i=1}^{q} V_i^{\mathrm{C}}\mid V_{k+1}^{\mathrm{N}}\big]\cdot\mathbb{P}\big(\big|\hat{\mathsf{W}}_{\mathrm{C}}\big|=q\big)+\mathbb{E}\big(\sum_{i=1}^{r-k} V_i^{\mathrm{C}}\big)\\
&\quad\cdot\mathbb{P}\big(\|\hat{\mathsf{W}}_{\mathrm{C}}|\geqslant r-k\big)\\
&=\mathbb{E}\big[\textstyle\sum_{i=1}^{k} V_i^{\mathrm{N}}\mid V_{k+1}^{\mathrm{N}}\big]+\sum_{q=1}^{r-k-1}\mathbb{E}\big[\sum_{i=1}^{q} V_i^{\mathrm{C}}\mid V_{k+1}^{\mathrm{N}}\big]\cdot\binom{\mathrm{C}}{q}\big(1-F(V_{k+1}^{\mathrm{N}})\big)^q\big(F(V_{k+1}^{\mathrm{N}})\big)^{\mathrm{C}-q}+\\
&\qquad\mathbb{E}\big[\textstyle\sum_{i=1}^{r-k} V_i^{\mathrm{C}}\mid V_{k+1}^{\mathrm{N}}\big]\sum_{q=r-k}^{\mathrm{C}}\binom{\mathrm{C}}{q}\big(1-F(V_{k+1}^{\mathrm{N}})\big)^q\big(F(V_{k+1}^{\mathrm{N}})\big)^{\mathrm{C}-q}\\
&=\mathbb{E}\big[\textstyle\sum_{i=1}^{k} Q(U_i^{\mathrm{N}})\mid U_{k+1}^{\mathrm{N}}\big]+\mathbb{E}\big[\sum_{i=1}^{r-k} Q(U_i^{\mathrm{C}})\mid U_{k+1}^{\mathrm{N}}\big]\cdot\sum_{q=r-k}^{\mathrm{C}}\binom{\mathrm{C}}{q}\big(1-F(V_{k+1}^{\mathrm{N}})\big)^q\\
&\quad\big(F(V_{k+1}^{\mathrm{N}})\big)^{\mathrm{C}-q}+\textstyle\sum_{q=1}^{r-k-1}\mathbb{E}\big[\sum_{i=1}^{q} Q(U_i^{\mathrm{C}})\mid U_{k+1}^{\mathrm{N}}\big]\binom{\mathrm{C}}{q}\big(1-F(V_{k+1}^{\mathrm{N}})\big)^q\big(F(V_{k+1}^{\mathrm{N}})\big)^{\mathrm{C}-q}\\
&=k\cdot\mathbb{E}\big[Q(U)\mid U\geqslant F(V_{k+1}^{\mathrm{N}})\big]+\textstyle\sum_{q=1}^{r-k-1} q\cdot\mathbb{E}\big[Q(U)\mid U\geqslant F(V_{k+1}^{\mathrm{N}})\big]\cdot\binom{\mathrm{C}}{q}\big(1-F(V_{k+1}^{\mathrm{N}})\big)^q\\
&\quad\big(F(V_{k+1}^{\mathrm{N}})\big)^{\mathrm{C}-q}+\textstyle\sum_{q=r-k}^{\mathrm{C}}\mathbb{E}\big[\sum_{i=1}^{r-k} Q(U_i^{q})\mid U_q^{q}\geqslant F(V_{k+1}^{\mathrm{N}})\big]\cdot\binom{\mathrm{C}}{q}\big(1-F(V_{k+1}^{\mathrm{N}})\big)^q\big(F(V_{k+1}^{\mathrm{N}})\big)^{\mathrm{C}-q}\\
&=\mathbb{E}\big[Q(U)\mid U\geqslant F(V_{k+1}^{\mathrm{N}})\big]\cdot\Big(k+\mathbb{E}\Big[G\left(|\hat{\mathsf{W}}_{\mathrm{C}}|,V_{k+1}^{\mathrm{N}};k\right)\mid V_{k+1}^{\mathrm{N}}\Big]\Big)
\end{aligned}$$

Above, $U\sim\mathrm{Uniform}[0,1]$ and $|\hat{\mathsf{W}}_{\mathrm{C}}|\mid V_{k+1}^{\mathrm{N}}\sim\mathrm{Binomial}\left(C,1-F\left(V_{k+1}^{\mathrm{N}}\right)\right)$ are the total number of winning colluding bidders and $G\left(|\hat{\mathsf{W}}_{\mathrm{C}}|,V_{k+1}^{\mathrm{N}};k\right)=|\hat{\mathsf{W}}_{\mathrm{C}}|1\left\{|\hat{\mathsf{W}}_{\mathrm{C}}|\leqslant r-k\right\}+$ $\frac{\mathbb{E}\left[\sum_{i=1}^{r-k} Q(U_i^{|\hat{\mathsf{W}}_{\mathrm{C}}|})\mid U_{|\hat{\mathsf{W}}_{\mathrm{C}}|}^{|\hat{\mathsf{W}}_{\mathrm{C}}|}\geqslant F(V_{k+1}^{\mathrm{N}})\right]}{\mathbb{E}\left[Q(U)\mid U\geqslant F(V_{k+1}^{\mathrm{N}})\right]}1\left\{|\hat{\mathsf{W}}_{\mathrm{C}}|>r-k\right\}$. Taking an expectation over V_{k+1}^{N} is the above expression gives the required result.

Proof of Theorem 2.4.4

By Lemma 2.4.2, all bids are known to be truthful. When $k=r$, all the items are allocated to non-colluding bidders, and V-PoP simply runs the usual VCG mechanism on the non-colluding bidders. Hence, $\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};r)=\mathrm{Welf}_{VCG}(\mathsf{N})$. Since V-PoP uses $k^*=\arg\max_k M(k)$, setting $M(k)=\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};k)\right]$ maximizes the expected welfare, and $\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};k^*)\right]\geqslant\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};r)\right]$. Combining it all, we get $\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};k^*)\right]\geqslant\mathbb{E}\left[\mathrm{Welf}_{V-PoP}(\mathsf{N}\cup\mathsf{C};r)\right]=\mathbb{E}\left[\mathrm{Welf}_{VCG}(\mathsf{N})\right]$.

Next, we prove the revenue guarantees. If $k^*=r$ then $\mathrm{Rev}_{V-PoP}(\mathsf{N}\cup\mathsf{C};k^*)=\mathrm{Rev}_{VCG}(\mathsf{N})$. If, instead, $k^*<r$ then the price set by the mechanism $b_{k^*+1}^{\mathrm{N}}$ is higher than the VCG price with just the non-colluding bidders b_{r+1}^{N}, i.e., $b_{k^*+1}^{\mathrm{N}}>b_{r+1}^{\mathrm{N}}$. Hence, when all items are sold by V-PoP mechanism then the revenue achieved is higher than that of VCG with the non-colluding bidders as the price cut-off is higher. All items are sold with the probability

$\mathbb{P}\Big(|\hat{\mathsf{W}}_{\mathrm{C}}| > r - k\Big)$ calculated as follows:

$$\begin{aligned}&\mathbb{E}\left[\mathbb{P}\Big(|\hat{\mathsf{W}}_{\mathrm{C}}| > r - k \mid V^{\mathrm{N}}_{k^*+1}\Big)\right] = \mathbb{E}\left[\textstyle\sum_{q=r-k^*}^{\mathrm{C}} \binom{\mathrm{C}}{q}\big(1 - F(V^{\mathrm{N}}_{k^*+1})\big)^{q}\big(F(V^{\mathrm{N}}_{k^*+1})\big)^{\mathrm{C}-q}\right]\\ &= \textstyle\sum_{q=r-k^*}^{\mathrm{C}} \binom{\mathrm{C}}{q}\mathbb{E}\left[\big(1 - U^{\mathrm{N}}_{k^*+1}\big)^{q}\big(U^{\mathrm{N}}_{k^*+1}\big)^{\mathrm{C}-q}\right]\\ &= \textstyle\sum_{q=r-k^*}^{\mathrm{C}} \binom{\mathrm{C}}{q}\frac{B(C+N-k^*-q,q+k^*+1)}{B(N-k^*,k^*+1)},\end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function. The last equality above uses the property that for $X \sim$Beta(α,β), $\mathbb{E}\left[X^a(1-X)^b\right] = \frac{B(\alpha+a,\beta+b)}{B(\alpha,\beta)}$.

Proof of Lemma 2.4.6

The result is a consequence of the probability integral transform of continuous random variables. By equation (2.3.7) of [45], $V^{\mathrm{S}}_k = \mathsf{Q}(U^{\mathrm{S}}_k)$. $LU^{\mathrm{S}}_k \leqslant V^{\mathrm{S}}_k$ follows from Assumption 2.4.2. Using $Q(\cdot) = F^{-1}(\cdot)$ from Assumption 2.4.1, we get $F(V^{\mathrm{N}}_k) = U^{\mathrm{N}}_k$.

Proof of Lemma 2.4.7

By the proof in Lemma 2.4.3, $\mathbb{E}\big[\mathrm{Welf}_{V\text{-PoP}}(\mathsf{N} \cup \mathsf{C}; k) \mid (V^{\mathrm{N}}_{k+1}, \ldots, V^{\mathrm{N}}_{\mathrm{N}})\big] = \mathbb{E}\big[Q(U) \mid U \geqslant F(V^{\mathrm{N}}_{k+1})\big] \cdot \big(k + \mathbb{E}\big[G(|\hat{\mathsf{W}}_{\mathrm{C}}|, V^{\mathrm{N}}_{k+1}; k) \mid V^{\mathrm{N}}_{k+1}\big]\big)$, where $U \sim \mathrm{Uniform}[0,1]$, $|\hat{\mathsf{W}}_{\mathrm{C}}| \mid V^{\mathrm{N}}_{k+1} \sim$ Binomial$(C, 1 - F(V^{\mathrm{N}}_{k+1}))$ and $G\left(|\hat{\mathsf{W}}_{\mathrm{C}}|, V^{\mathrm{N}}_{k+1}; k\right) = |\hat{\mathsf{W}}_{\mathrm{C}}| 1\left\{|\hat{\mathsf{W}}_{\mathrm{C}}| \leqslant r - k\right\} +$ $\frac{\mathbb{E}\left[\sum_{i=1}^{r-k} Q(U_i^{|\hat{\mathsf{W}}_{\mathrm{C}}|}) | U_{|\hat{\mathsf{W}}_{\mathrm{C}}|}^{|\hat{\mathsf{W}}_{\mathrm{C}}|} \geqslant F(V^{\mathrm{N}}_{k+1})\right]}{\mathbb{E}\left[Q(U) | U \geqslant F(V^{\mathrm{N}}_{k+1})\right]} 1\left\{|\hat{\mathsf{W}}_{\mathrm{C}}| > r - k\right\}$. Using Assumption 2.4.2 and Lemma 2.4.6, we bound the above terms above separately as follows.

$$\begin{aligned}
&\mathbb{E}\big[Q(U) \mid U \geqslant F\big(V^{\mathrm{N}}_{k+1}\big)\big] \cdot \left(k + \mathbb{E}\Big[G\left(|\hat{\mathsf{W}}_{\mathrm{C}}|, V^{\mathrm{N}}_{k+1}; k\right)\Big|\, V^{\mathrm{N}}_{k+1}\Big]\right)\\
&= \mathbb{E}\big[V \mid V \geqslant V^{\mathrm{N}}_{k+1}\big] \cdot \left(k + \textstyle\sum_{q=1}^{r-k} q\binom{\mathrm{C}}{q}\big(1 - F(V^{\mathrm{N}}_{k+1})\big)^q\big(F(V^{\mathrm{N}}_{k+1})\big)^{\mathrm{C}-q}\right)\\
&\qquad + \textstyle\sum_{q=r-k+1}^{\mathrm{C}} \mathbb{E}\left[\sum_{i=1}^{r-k} Q(U_i^q) \mid U_q^q \geqslant F(V^{\mathrm{N}}_{k+1})\right]\binom{\mathrm{C}}{q}\big(1 - F(V^{\mathrm{N}}_{k+1})\big)^q\big(F(V^{\mathrm{N}}_{k+1})\big)^{\mathrm{C}-q}\\
&\geqslant \mathbb{E}\big[LU \mid U \geqslant U^{\mathrm{N}}_{k+1}\big]\left(k + \textstyle\sum_{q=1}^{r-k} q\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\right)\\
&\qquad + \textstyle\sum_{q=r-k+1}^{\mathrm{C}} \mathbb{E}\left[\sum_{i=1}^{r-k} LU_i^q \mid U_q^q \geqslant U^{\mathrm{N}}_{k+1}\right]\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\\
&\geqslant \tfrac{L\left(1+U^{\mathrm{N}}_{k+1}\right)}{2}\left(k + \textstyle\sum_{q=1}^{r-k} q\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\right)\\
&\qquad + L\textstyle\sum_{q=r-k+1}^{\mathrm{C}} \mathbb{E}\left[\sum_{i=1}^{r-k}(1 - U^{\mathrm{N}}_{k+1})U_i^q + U^{\mathrm{N}}_{k+1}\right]\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\\
&\geqslant \tfrac{L\left(1+U^{\mathrm{N}}_{k+1}\right)}{2}\left(k + \textstyle\sum_{q=1}^{r-k} q\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\right)\\
&\qquad + \textstyle\sum_{q=r-k+1}^{\mathrm{C}} L(r-k)\left(1 - \frac{\left(1-U^{\mathrm{N}}_{k+1}\right)(r-k+1)}{2(q+1)}\right)\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\\
&= \tfrac{L\left(1+U^{\mathrm{N}}_{k+1}\right)}{2}\cdot r(k, U^{\mathrm{N}}_{k+1})\\
&\geqslant \tfrac{L(1+\frac{1}{H}V^n_{k+1})}{2}\left(k + \textstyle\sum_{q=1}^{r-k} q\binom{\mathrm{C}}{q}\left(\max\{0, 1 - \frac{1}{L}V^n_{k+1}\}\right)^q\left(\min\{1, \frac{1}{H}V^n_{k+1}\}\right)^{\mathrm{C}-q}\right) + \textstyle\sum_{q=r-k+1}^{\mathrm{C}} L\cdot\\
&(r-k)\left(1 - \tfrac{(\max\{0,1-\frac{1}{H}V^n_{k+1}\})(r-k+1)}{2(q+1)}\right)\binom{\mathrm{C}}{q}\left(\max\{0, 1 - \tfrac{1}{L}V^n_{k+1}\}\right)^q\left(\min\{1, \tfrac{1}{H}V^n_{k+1}\}\right)^{\mathrm{C}-q}.
\end{aligned}$$

The second inequality is because $Q(U) = V \geqslant LU$ by Lemma 2.4.6 and $F(V^{\mathrm{N}}_{k+1}) = U^{\mathrm{N}}_{k+1}$. The third equality is using $\mathbb{E}\left[U \mid U \geqslant U^{\mathrm{N}}_{k+1}\right] = (1 + U^{\mathrm{N}}_{k+1})/2$ for $U \sim \text{Uniform}[0, 1]$, and $\mathbb{E}\left[U_i^q \mid U_q^q \geqslant U^{\mathrm{N}}_{k+1}\right] = \left(1 - U^{\mathrm{N}}_{k+1}\right)\mathbb{E}\left[U_i^q\right] + U^{\mathrm{N}}_{k+1}$. The last inequality is because the summation of expected values of uniform order statistics $\sum_{i=1}^{r-k}\mathbb{E}\left[U_i^q\right] = (r-k)\frac{(2q-r+k+1)}{2(q+1)}$ when $r - k \leqslant q$. Finally, note that $U^{\mathrm{N}}_{k+1} \sim \text{Beta}\left(N - k, k + 1\right)$ and is $(k + 1)-$th highest order statistic from N i.i.d. standard uniform random variables.

Proof of Lemma 2.4.8

The proof is very similar to Lemma 2.4.7.

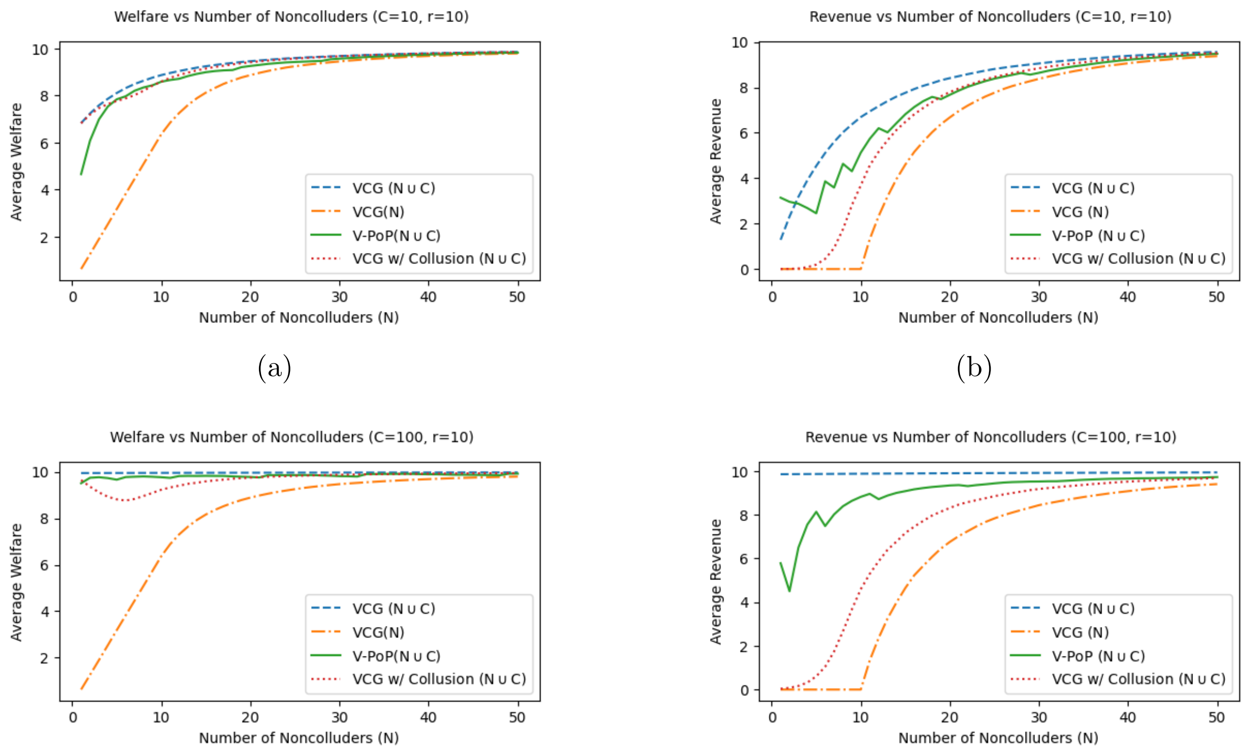
$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Rev}_{V-PoP}(\mathsf{N} \cup \mathsf{C}; k) \mid (V^{\mathrm{N}}_{k+1}, \cdots, V^{\mathrm{N}}_{\mathrm{N}})\right] = V^{\mathrm{N}}_{k+1}\left(k + \mathbb{E}\Big[\min\{|\hat{\mathsf{W}}_{\mathrm{C}}|, r-k\}\Big]\right)\\
&\geqslant V^{\mathrm{N}}_{k+1}\left(k + \textstyle\sum_{q=0}^{\mathrm{C}} \min\{r-k, q\}\binom{\mathrm{C}}{q}\big(1 - U^{\mathrm{N}}_{k+1}\big)^q\big(U^{\mathrm{N}}_{k+1}\big)^{\mathrm{C}-q}\right)\\
&\geqslant LU^{\mathrm{N}}_{k+1}\tilde{r}(k, U^{\mathrm{N}}_{k+1}).
\end{aligned}$$

Proof of Theorem 2.4.9

The proof is very similar to the derivation of revenue guarantees in Theorem 2.4.4.

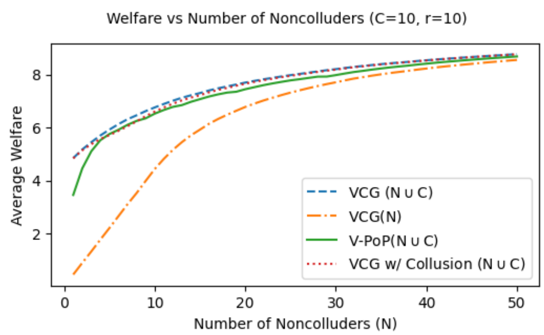
Other plots from numerical experiments

The plots from numerical simulations for the distributions- sinusoidal, trapezoidal, and quadratic- are below in Figures A.1, A.2 and A.3 respectively.

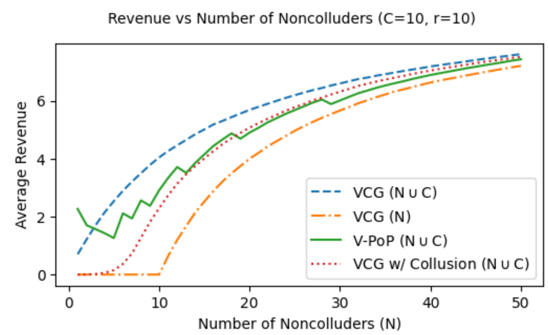


(a) (b) (c) (d)

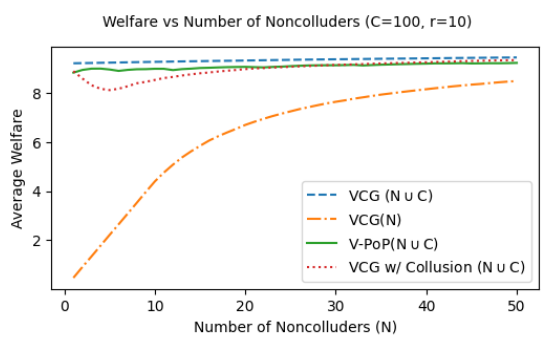
Figure A.1: Numerical results for Sinusoidal distribution.



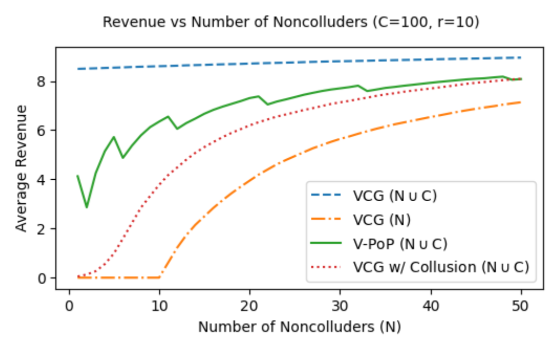
(a)



(b)

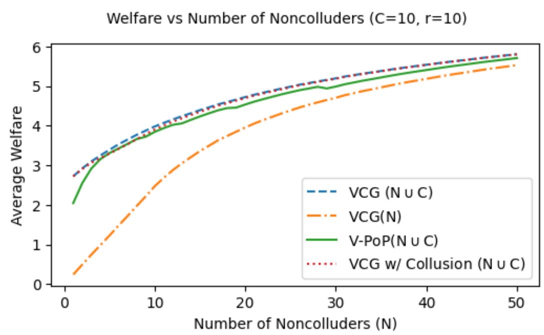


(c)

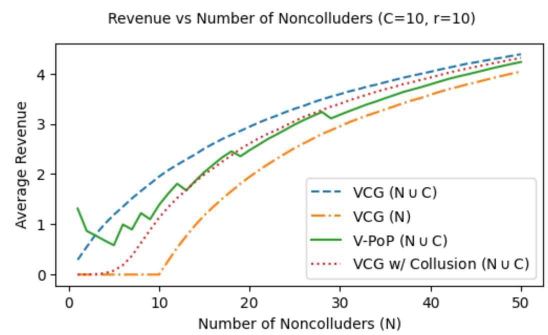


(d)

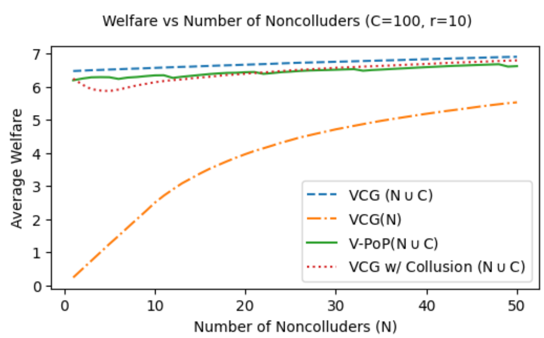
Figure A.2: Numerical results for Trapezoidal distribution.



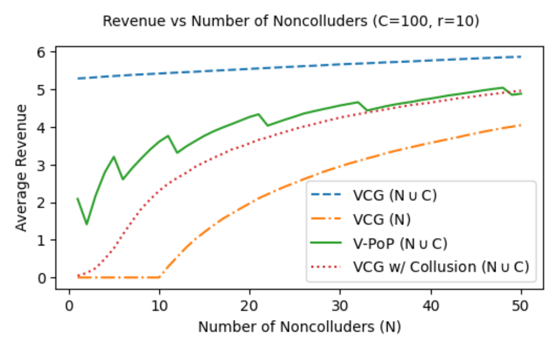
(a)



(b)



(c)



(d)

Figure A.3: Numerical results for Quadratic distribution.

Appendix B

Appendix for Chapter 3

The omitted proofs from Chapter 3 are provided here.

Proof of Lemma 3.3.4

Fix a feasible trilevel point $x = (x^{(1)}, x^{(2)}, x^{(3)}) \in \Omega_{\text{true}}$. Then, by definition of feasibility of the trilevel problem, the component $x^{(3)}$ is an optimal solution of the level 3 subproblem given the upper level decisions $(x^{(1)}, x^{(2)})$. Since the level 3 objective is strongly convex and the constraints are linear, the level 3 subproblem is a convex program with a unique primal optimal point. Also, the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality.

Let A_3 denote the active constraint set at the optimizer. The KKT system corresponding to A_3 can be written as a linear system in the primal-dual variables $(x^{(3)}, \lambda_3)$. Solving this system yields the affine representation

$$\hat{x}^{(3)} = K_3^{(\mathsf{A}_3)\dagger(x)}\big(N_3^{(\mathsf{A}_3)} + M_3^{(\mathsf{A}_3)} x^{(12)}\big),$$

together with the dual and primal validity constraints in (3.5). These constraints define a region of validity on which the above affine mapping coincides with the exact optimal response of the level 3 problem.

Because $x \in \Omega_{\text{true}}$, the level 3 optimizer at $(x^{(1)}, x^{(2)})$ satisfies the KKT system and therefore satisfies the generated optimality constraint. Hence the level-3 optimality constraint is valid for Ω_{true}.

Proof of Lemma 3.3.5

We first prove (1). Take any $(x^{(2)}, x^{(3)}) \in \bar{\Omega}_2(x^{(1)})$. By definition, this point satisfies the level 2 constraints and also the level 3 optimality constraints (3.4) and (3.5). The constraint (3.5) ensures that the active set A_3' used to derive the cut is valid at $(x^{(1)}, x^{(2)})$. Since the level 3 subproblem is a strongly convex, quadratic program with linear constraints, its KKT conditions are necessary and sufficient for optimality. Therefore, (3.4) represents the exact optimal response of level 3 in its region of validity defined by constraint (3.5). Hence $x^{(3)}$ is the true lower-level best response to $(x^{(1)}, x^{(2)})$, and $(x^{(2)}, x^{(3)})$ is feasible for the original level 2 subproblem (3.6). This proves that $\bar{\Omega}_2(x^{(1)}) \subseteq \Omega_2(x^{(1)})$.

We now prove (2). By definition, $\hat{x}^{(23)}$ is feasible for (3.6) at the given $x^{(1)}$ and optimal for the level 2 subproblem. In particular, the component $\hat{x}^{(3)}$ solves the level 3 subproblem at $(x^{(1)}, \hat{x}^{(2)})$. Let $\hat{\lambda}_3$ be any dual multiplier associated with this level 3 optimum, and let A'_3 be the active constraint set at $(\hat{x}^{(3)}, \hat{\lambda}_3)$. Using KKT conditions and active set A'_3, when evaluated at $(x^{(1)}, \hat{x}^{(2)}, \hat{x}^{(3)})$ and $\hat{\lambda}_3$, the constraints (3.4) and (3.5) are satisfied. Since $\hat{x}^{(23)}$ also satisfies all level 2 constraints from (3.6), it satisfies every constraint defining $P_2^{\mathrm{UBD}}(x^{(1)})$. Therefore, $\hat{x}^{(23)} \in \bar{\Omega}_2(x^{(1)})$, which proves (2).

Proof of Lemma 3.3.8

Let $v_2(x^{(1)})$ be the optimal value of the level 2 subproblem (3.6), and $\bar{v}_2(x^{(1)})$ be the optimal value of $P_2^{\mathrm{UBD}}(x^{(1)})$. Define $\Omega_2(x^{(1)})$ and $\bar{\Omega}_2(x^{(1)})$ as the feasible sets of these problems, respectively. By Lemma 3.3.5, for the specific active sets A_3 and A_2 used to construct $P_2^{\mathrm{UBD}}(x^{(1)})$, we have $\bar{\Omega}_2(x^{(1)}) \subseteq \Omega_2(x^{(1)})$ and $v_2(x^{(1)}) = \bar{v}_2(x^{(1)})$. In particular, the inequality $v_2(x^{(1)}) \leqslant \bar{v}_2(x^{(1)})$ is tight at the bilevel optimal point and for the chosen active sets.

Now consider any globally feasible trilevel point $x = (x^{(1)}, x^{(2)}, x^{(3)}) \in \Omega_{\mathrm{true}}$. By definition, $(x^{(2)}, x^{(3)}) \in \Omega_2(x^{(1)})$ and it is also optimal for the level 2 subproblem. Hence, its optimal objective value $v_2(x^{(1)})$ satisfies $v_2(x^{(1)}) = \bar{v}_2(x^{(1)})$. Similar to the proof in Lemma (3.3.4), the upper bound problem $P_2^{\mathrm{UBD}}(x^{(1)})$ is quadratic program with strongly convex objective and KKT conditions are necessary and sufficent. Given the active constraints set A_2, its unique primal optimal point $x_{UB}^{(23)}$ is given by (3.7) within the validity polytope (3.8). The optimal point $x_{UB}^{(23)}$ can be used to calculate $\bar{v}_2(x^{(1)})$ within the validity polytope (3.8). This along with $v_2(x^{(1)}) \leqslant \bar{v}_2(x^{(1)})$ implies that the optimality constraint holds for the $x \in \Omega_{\mathrm{true}}$. Therefore, every globally feasible trilevel point satisfies the generated level 2 optimality constraints.

Proof of Lemma 3.3.9

Say, the level 2 subproblem is infeasible for a given $x^{(1)}$. For any $x^{(2)}$ for which the third level subproblem is feasible, an upper bound problem similar to P_2^{UBD} can be constructed using the corresponding third level active set A_3. Let this upper bound problem have a consolidated constraint system (3.9) with feasible region $\Omega(x^{(1)})$. Then, similar to Lemma (3.3.5), $\Omega(x^{(1)}) \subseteq \Omega_2(x^{(1)})$. Since $\Omega_2(x^{(1)}) = \varnothing$ due to the infeasibility of level 2 subproblem, $\Omega(x^{(1)}) = \varnothing$ too. By Farkas' lemma, there exists a multiplier $\omega \geqslant 0$ satisfying $\omega^\top G_2 = 0$ and $\omega^\top (W_2 + S_2 x^{(1)}) < 0$. Therefore, for $\Omega(x^{(1)})$ to be non-empty, $(\omega^\top S_2) x^{(1)} \geqslant -\omega^\top W_2$ must be satisfied. Thus the feasibility cut is also valid for Ω_{true}.

Proof of Lemma 3.3.10

We first verify feasibility of x_u for the trilevel problem (3.1). In the upper bound problem (3.11) at $x_k^{(1)}$, the level 1 and level 2 feasibility constraints $A_i x_u \leqslant B_i$ and $l_i \leqslant x_u^{(i)} \leqslant u_i$ for $i = 1, 2$ are satisfied by construction. The last constraint block in (3.11) lists the level 3 subproblem ensuring level 3 feasibility and optimality, i.e. at $(x_k^{(1)}, x_u^{(2)})$, $x_u^{(3)}$ is both level 3 feasible and optimal. Finally, $f_2^\star(x_k^{(1)})$ is the optimal value of the level 2 subproblem (3.6) at $x_k^{(1)}$ and for any feasible $(x^{(2)}, x^{(3)})$ of the level 2 bilevel problem, we must have $x^{(23)\top} H_2^{(23,23)} x^{(23)} + (2x_k^{(1)\top} H_2^{(1,23)} + C_2^{(23)\top})x^{(23)} \geqslant f_2^\star(x_k^{(1)})$. In the upper-bound problem (3.11), we impose the additional inequality $x^{(23)\top} H_2^{(23,23)} x^{(23)} + (2x_k^{(1)\top} H_2^{(1,23)} + C_2^{(23)\top})x^{(23)} \leqslant f_2^*(x_k^{(1)})$. Combining this with the previous inequality shows that any feasible $(x^{(2)}, x^{(3)})$ of (3.11) must satisfy $x^{(23)\top} H_2^{(23,23)} x^{(23)} + (2x_k^{(1)\top} H_2^{(1,23)} + C_2^{(23)\top})x^{(23)} = f_2^\star(x_k^{(1)})$. Thus, $(x_u^{(2)}, x_u^{(3)})$ is an optimal solution of the level 2 bilevel problem for the fixed first level decision $x_k^{(1)}$. Hence, x_u is feasible for level 1,2,3 constraints and optimal at levels 2,3, and so feasible for the original trilevel problem (3.1).

Finally, the objective of the upper bound problem (3.11) is exactly the level 1 objective $C_1^\top x + x^\top H_1 x$ evaluated at x_u. Since x_u is a feasible solution of the trilevel problem, its objective value is an upper bound on the optimal trilevel value. This completes the proof.

Proof of Lemma 3.3.11

Since the constraints (3.4)–(3.5) were added to the master problem at iteration k, and $x_{k'}$ lies in their validity polytope with the same level 3 active set A_3, the pair $(x_{k'}^{(3)}, \lambda_3)$ satisfies the KKT system of the level 3 subproblem corresponding to A_3. Because the level 3 subproblem is a strictly convex quadratic program with linear constraints, these KKT conditions are necessary and sufficient, implying that $x_{k'}^{(3)}$ is the unique optimal response of level 3 to $(x_{k'}^{(1)}, x_{k'}^{(2)})$.

Similarly, the constraints (3.7)–(3.8) were added using the level 2 active sets $(\mathsf{A}_2, \mathsf{A}_3')$ at iteration k. By the lemma (3.3.5), the corresponding upper bound problem $P_2^{\mathrm{UBD}}(x_k^{(1)})$ is an exact restriction of the level 2 bilevel problem on its validity polytope, and its optimal value equals $v_2(x_k^{(1)})$. Since $x_{k'}$ lies in the same validity polytope with the same active sets $(\mathsf{A}_2, \mathsf{A}_3')$, the inequality (3.7) enforces $f_2(x_{k'}^{(1)}, x_{k'}^{(23)}) \leqslant f_2\big(x_{k'}^{(1)}, \hat{x}_{\mathrm{UB}}^{(23)}(x_{k'}^{(1)})\big) = v_2(x_{k'}^{(1)})$. By definition of $v_2(x_{k'}^{(1)})$ as a minimum, we also have $f_2(x_{k'}^{(1)}, x_{k'}^{(23)}) \geqslant v_2(x_{k'}^{(1)})$. Thus $f_2(x_{k'}^{(1)}, x_{k'}^{(23)}) = v_2(x_{k'}^{(1)})$, and $(x_{k'}^{(2)}, x_{k'}^{(3)})$ is optimal for the level 2 subproblem at $x_{k'}^{(1)}$.

Combining these facts, $x_{k'}$ satisfies all level 1 and level 2 physical constraints, the level 3 and level 2 optimality conditions of the trilevel problem (3.1). Hence $x_{k'}$ is feasible for the original trilevel problem.

Finally, the lower bound from the master problem at iteration k' equals $f_1(x_{k'})$. The upper bound problem, constructed at fixed $x_{k'}^{(1)}$, also has $x_{k'}$ as a feasible point, yielding the same optimal objective value $f_1(x_{k'})$. Thus, the lower and upper bounds coincide at iteration k', and the optimality gap closes.

Appendix C

Appendix for Chapter 4

The omitted proofs from Chapter 4 are provided here.

Proof of Proposition 4.3.1

The proof for this proposition is similar to that of Proposition 2.1 of [31], but where a multilinear optimization problem is formulated by restricting the entries to $[-1, 1]$ instead of $[0, 1]$ as in (C). We proceed by showing two set inclusions $\text{conv}(\mathcal{S}_\lambda) \subseteq \text{conv}(\mathcal{B}_\lambda)$ and $\text{conv}(\mathcal{B}_\lambda) \subseteq \text{conv}(\mathcal{S}_\lambda)$. The first inclusion is immediate by definition since we have $\mathcal{S}_\lambda \subset \mathcal{B}_\lambda$. The second inclusion can be proved by contradiction. Suppose instead that $\text{conv}(B_\lambda) \not\subset \text{conv}(\mathcal{S}_\lambda)$. Then there exists a tensor $\psi' \in B_\lambda$ with $\psi' \notin \text{conv}(\mathcal{S}_\lambda)$. By the hyperplane separation theorem, there exists $\varphi \in \mathbb{R}^{r_1 \times \cdots \times r_p}$ and $\delta > 0$ such that $\langle \varphi, \psi' \rangle \geqslant \langle \varphi, \psi \rangle + \delta$ for all $\psi \in \text{conv}(\mathcal{S}_\lambda)$, where $\langle \cdot, \cdot \rangle$ is the usual inner product that is defined as the summation of elementwise multiplication. Now consider the multilinear optimization problem

$$\begin{aligned} &\max \langle \varphi, \psi \rangle \\ &\text{s.t. } \psi_x = \lambda \cdot \textstyle\prod_{k=1}^p \theta_{x_k}, \quad &&\text{for } x \in \mathcal{R} \\ &\qquad \theta_{x_k} \in [-1, 1], &&\text{for } x \in \mathcal{R} \end{aligned}$$

Proposition 2.1 of [53] shows there exists a global optimum ψ'' of (C) with $\psi'' \in \mathcal{S}_\lambda$. By construction, $\langle \varphi, \psi'' \rangle \geqslant \langle \varphi, \psi' \rangle$, which implies $\langle \varphi, \psi'' \rangle \geqslant \langle \varphi, \psi \rangle + \delta$ for $\psi \in \text{conv}(\mathcal{S}_\lambda)$. This statement is a contradiction since $\psi'' \in \mathcal{S}_\lambda \subseteq \text{conv}(\mathcal{S}_\lambda)$.

Proof of Proposition 4.3.3

We use the result from Example 3.50 of [137] to conclude that $\| \cdot \|_\pm$ is a norm. To apply the result, we check that the required conditions on $\mathcal{C}_1$ hold.

By definition $\mathcal{C}_1$ is convex, closed, and bounded. $\mathcal{C}_1$ is also symmetric since for every $a \in \mathcal{C}_1$, $-a \in \mathcal{C}_1$ is also true. To see this, let $a = \sum_i \lambda_i \psi_i$ where $\psi_i \in \mathcal{S}_1$, $\lambda_i \in [0, 1]$ and $\sum_i \lambda_i = 1$ and notice $-\psi_i \in \mathcal{S}_1$. Symmetry and convexity also ensure $0 \in \mathcal{C}_1$ since for any $a \in \mathcal{C}_1$, $\frac{1}{2}(a + (-a)) = 0$.

The final, non-trivial condition required is that $\mathcal{C}_1$ has a non-empty interior. To prove this, we use Theorem 2.4 of [138] that the dimension of $\mathcal{C}_1$ is the maximum of dimensions of simplices included in it. We construct a simplex in $\mathcal{C}_1$ that has dimension $\pi = \Pi_i r_i$, and hence $\mathcal{C}_1$ has a dimension of at least π. But the flattened vector space of tensors also has dimension π; so, the dimension of $\mathcal{C}_1$ is at most π. Hence, $\mathcal{C}_1$ has a dimension π, which is the full dimension of the space, implying that the set $\mathcal{C}_1$ must have a non-empty interior.

The rest of the proof constructs such a simplex of dimension π. Consider a polytope $D = \operatorname{conv}(0 \cup \{d^x\}_{x\in\mathcal{R}})$. Here, for any $x = (x_1, \cdots, x_p) \in \mathcal{R}$, consider $d^x = \bigotimes_{k=1}^p \beta^{x_k}$ with each vector $\beta^{x_k} \in \mathbb{R}^{r_k}$ defined as $\beta^{x_k} = \mathbb{1}$ if $x_k = 1$ and $\beta^{x_k} = f_{x_k}$ if $x_k \neq 1$, where $\mathbb{1}$ is a vector of one's and f_j is a vector with -1 in position j and one's elsewhere. One can verify that $\{\beta^{x_k}\}_{x_k\in[r_k]}$ are linearly independent vectors and make a complete basis for $\mathbb{R}^{r_k}$. Since the tensor product of linearly independent vectors gives linearly independent tensors, the tensors $\{d^x\}_{x\in\mathcal{R}}$ are all linearly independent. Consequently, $\{d^x - 0\}_{x\in\mathcal{R}}$ are linearly independent and $\{d^x\}_{x\in\mathcal{R}} \cup 0$ are affinely independent. By definition, the polytope D, which is a convex hull of $|\mathcal{R}| + 1 = \pi + 1$ affinely independent points, is a simplex of dimension π. Note that all the points are in $\mathcal{C}_1$, and hence so is the simplex.

With this, $\mathcal{C}_1$ is shown to satisfy all the required conditions, which means that the proposition holds.

Proof of Proposition 4.3.4

We show the constraints defining sets $\mathcal{S}_\lambda$ and $\widehat{\mathcal{S}}_\lambda$ are equivalent. Consider some $x \in \mathcal{R}$. From the definition of $\mathcal{S}_\lambda$ in (4.3), we have $\psi_x = \lambda \prod_{i=1}^p \theta_{x_k}$. Define $y_{x,k} = \prod_{i=k}^p \theta_{x_k}$, for $k \in [p]$ so that $\psi_x = \lambda y_{x,1}$. Or equivalently, in a recursive relationship, $y_{x,k} = \theta_{x_k} y_{x,k+1}$ for $k \in [p-1]$ and $y_{x,p} = \theta_{x_p}$. The recursive constraints can be thought of as a negated-XOR relation, and linearized by transformations for conjunctive and disjunctive statements (see Section 2.5 of [43]). These linearized constraints correspond to constraints 2-5 in the definition of $\widehat{\mathcal{S}}_\lambda$ as given above. So for each $x \in \mathcal{R}$, the tensor ψ_x is defined the same in both $\mathcal{S}_\lambda$ and $\widehat{\mathcal{S}}_\lambda$.

Proof of Proposition 4.3.7

By the CP decomposition and the triangle inequality: $\|\psi\|_\pm \leqslant \sum_{k=1}^{\operatorname{rank}(\psi)} \|\psi^k\|_\pm = \sum_{k=1}^{\operatorname{rank}(\psi)} \|\psi^k\|_{\max}$, where $\psi^k \in \mathcal{B}_\infty$. The last equality follows from $\|\psi^k\|_\pm = \|\psi^k\|_{\max}$ when $\psi^k \in \mathcal{B}_\infty$. Using $\|\psi^k\|_{\max} \leqslant \|\psi\|_{\max}$ from Assumption 4.3.1 gives the right-side inequality.

The proof of the left-side inequality is similar to Proposition 2.4 of [31]. For any $\lambda \geqslant 0$, if $\psi \in \mathcal{S}_\lambda$, then by definition $\|\psi\|_{\max} = \lambda$. By the convexity of norms we have that: if $\psi \in \mathcal{C}_\lambda$, then $\|\psi\|_{\max} \leqslant \lambda$. This means that $\forall\, \lambda \geqslant 0$ we have $\mathcal{C}_\lambda \subseteq \mathcal{U}_\lambda := \{\psi : \|\psi\|_{\max} \leqslant \lambda\}$, and thus $\inf\{\lambda \mid \psi \in \lambda\mathcal{U}_1\} \leqslant \inf\{\lambda \mid \psi \in \lambda\mathcal{C}_1\}$. But this is equivalent to $\|\psi\|_{\max} \leqslant \|\psi\|_\pm$, which proves the left-side inequality.

Proof of Proposition 4.3.9

Note that $\|\varphi\|_\circ = \sup\{|\langle\varphi,\psi\rangle| \mid \|\psi\|_\pm \leqslant 1\} = \sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{C}_1\}$ is the dual norm for $\|\cdot\|_\pm$. The approximation of $\|\cdot\|_\circ$ can be reduced to approximation of the norm $\|\cdot\|_\pm$ in polynomial time from theorems 3 and 10 of [67]. Our main idea is to give a polynomial-time reduction of an NP-Hard problem to an approximation of $\|\cdot\|_\circ$. In particular, we prove that calculating the $\infty, 1$ subordinate matrix norm is polynomial-time reducible to $\sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{S}_1\}$. Without loss of generality, assume $p = 2$ and $d := r_1 = r_2$. The decision version of $\sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{S}_1\}$ is: *Question:* Does there exist $\theta_{x_k} \in \{-1, 1\}$ for all $x_k \in [d]$ with $k = 1, 2$ such that for a given L we have $\sum_{x_1=1}^d \sum_{x_2=1}^d \varphi_{x_1x_2}\theta_{x_1}\theta_{x_2} \geqslant L$? From Proposition 1 of [139], we have $\|W\|_{\infty,1} = \{\max \sum_{i,j} W_{ij}x_iy_j \,|\, x_i, y_j \in \{-1, 1\}\}$ for a matrix W. The decision version of the $\infty, 1$ subordinate matrix norm for the special case of M-matrices can be written as: *Question:* Does there exist $x_i, y_i \in \{-1, 1\}$ for all $i \in [d]$ such that for a given $L' \geqslant 0$ and a *symmetric, positive definite* matrix $W \in \mathbb{R}^{d\times d}$ satisfying $w_{ij} \leqslant 0$ for all $i \neq j$ and $\sum_{i=1}^d \sum_{j=1}^d w_{ij}x_iy_j \geqslant L'$? Clearly, setting $L = L'$ and $\varphi = W$ is a valid polynomial time reduction. Since $\mathcal{C}_1 = \mathrm{conv}(\mathcal{S}_1)$ and $\langle\varphi,\psi\rangle$ is linear, $\|\varphi\|_\circ = \sup\{\langle\varphi,\psi\rangle \mid \psi \in \mathcal{S}_1\}$. The result now follows since approximately solving any ∞, p subordinate matrix norm, where $p \in [1, \infty)$, to arbitrary accuracy is NP-hard [84]. In particular, Theorem 5 of [139] shows that it is an NP-hard problem for M-matrices. When combined with Corollary 3.2 in [31], this establishes NP-completeness.

Proof of Proposition 4.3.10

The proof is similar to Proposition 3.3 of [31]. By definition, $\mathsf{R}(\mathcal{C}_\lambda) = \mathbb{E}_\sigma(\sup_{\Psi\in\mathcal{C}_\lambda} \frac{1}{n}|\sum_{i=1}^n \sigma_i \cdot \psi_{x\langle i\rangle}|)$ and $\mathsf{W}(\mathcal{C}_\lambda) = \sup_X \mathsf{R}(\mathcal{C}_\lambda)$, where σ_i are independent and identically distributed (i.i.d.) Rademacher random variables (i.e., $\sigma_i = \pm 1$ with probability $\frac{1}{2}$) [18, 150]. Hence, $\mathsf{R}(\mathcal{C}_\lambda) \leqslant \mathsf{W}(\mathcal{C}_\lambda)$. Recall that $\mathcal{C}_\lambda = \mathrm{conv}(\mathcal{S}_\lambda)$ by Proposition 4.3.1. This means $\mathsf{W}(\mathcal{C}_\lambda) = \mathsf{W}(\mathcal{S}_\lambda)$ [106, 18]. Next, observe that

$$\begin{aligned}
\mathsf{W}(\mathcal{C}_\lambda) &= \mathsf{W}(\mathcal{S}_\lambda) \\
&= \sup_X \mathbb{E}_\sigma\big(\sup_{\psi\in\mathcal{S}_\lambda} \tfrac{1}{n}\big|\textstyle\sum_{i=1}^n \sigma_i \cdot \psi_{x\langle i\rangle}\big|\big) \\
&= \sup_X \mathbb{E}_\sigma\big(\max_{\psi\in\mathcal{S}_\lambda} \tfrac{1}{n} \cdot \textstyle\sum_{i=1}^n \sigma_i \cdot \psi_{x\langle i\rangle}\big) \\
&\leqslant \sup_X r\sqrt{2\log \#\mathcal{S}_\lambda}/n
\end{aligned}$$

where in the last line since the set $\mathcal{S}_\lambda$ is finite, we used the Finite Class Lemma [118] with $r = \max_{\psi\in\mathcal{S}_\lambda} \sqrt{\sum_{i=1}^n (\psi_{x\langle i\rangle})^2} \leqslant \lambda\sqrt{n}$, and replaced the supremum with a maximum. This inequality on r is because $\mathcal{S}_\lambda$ consists of tensors whose entries are from $\{-\lambda, \lambda\}$. Thus $\mathsf{W}(\mathcal{C}_\lambda) \leqslant \lambda\sqrt{(2\log 2) \cdot \rho/n} \leqslant \lambda \cdot 2\sqrt{\rho/n}$.

Proof of Proposition 4.4.4

The proof for this proposition is similar to the proof of Proposition 4.4 of [31]. Define the ball of radius $\delta > 0$ centered at a tensor ψ to be $B(\psi, \delta) = \{\varphi : \|\varphi - \psi\|_F \leqslant \delta\}$. Next define $W(\mathcal{C}_1, \delta) = \bigcup_{\psi \in \mathcal{C}_1} B(\psi, \delta)$ and $W(\mathcal{C}_1, -\delta) = \{\psi \in \mathcal{C}_1 : B(\psi, \delta) \subseteq \mathcal{C}_1\}$. The weak membership problem for $\mathcal{C}_1$ is that given a tensor ψ and a $\delta > 0$ decide whether $\psi \in W(\mathcal{C}_1, \delta)$ or $\psi \notin W(\mathcal{C}_1, -\delta)$. Theorem 10 of [67] shows that approximation of the norm $\|\cdot\|_\pm$ is polynomial-time reducible to the weak membership problem for $\mathcal{C}_1$. Since Proposition 4.3.9 shows that approximation of the norm $\|\cdot\|_\pm$ is NP-hard, the result follows if we can reduce the weak membership problem to the tensor completion problem.

Suppose we are given inputs ψ and δ for the weak membership problem. Choose $x\langle i\rangle$ for $i = 1, \ldots, \pi$ such that each element in $\mathcal{R}$ is enumerated exactly once. Next choose $y\langle i\rangle = \psi_{x\langle i\rangle}$ and $\lambda = 1$. Finally, note if we solve (4.1) and the minimum objective value is less than or equal to δ, then we have $\psi \in W(\mathcal{C}_1, \delta)$; otherwise, we have $\psi \notin W(\mathcal{C}_1, -\delta)$. The result follows since this was the desired reduction.

Proof of Proposition 4.4.5

By Theorem 3.1 of [29], BCG converges linearly when the feasible set is a polytope and the objective function is smooth and strongly convex. Note the feasible set $\|\psi\|_\pm \leqslant \lambda$ in Problem (4.1) is a polytope. The objective is smooth and it can be made strongly convex by projecting the feasible space onto the set of unique, observed tensor entries U [31]. Specifically, we use the equivalent reformulation in which we change the feasible set from $\{\psi : \|\psi\|_\pm \leqslant \lambda\} = C_\lambda$ to $\mathrm{Proj}_U(\mathcal{C}_\lambda)$ where the projection is done over the unique indices specified by the set U. Strong convexity of the objective results from strict convexity and quadraticity of the objective on the compact projected space. Since the conditions for linear convergence are satisfied by our problem, our algorithm terminates in a linear number of calls to Algorithm 4.

Appendix D

Appendix for Chapter 5

The omitted proofs from Chapter 5 are provided here.

Proof for Proposition 5.3.1

$B(I)L \geqslant -\bar{V}$ holds $\forall I$ because $\bar{V} \geqslant 0$. When I_1 exists, it's the optimal solution to the unconstrained problem. Since it also satisfies $I_1 \geqslant 0$, it's the optimal solution for the constrained problem too. When I_1 doesn't exist, the objective in the proposition is an increasing function for $I \geqslant 0$. So the smallest feasible value of I, i.e 0, is optimal.

Proof for Proposition 5.3.2

When I_1 exists and $B(I_1)L \geqslant -\bar{V}$ holds, I_1, which is the unconstrained optimal solution, satisfies both constraints and is optimal. When $B(I_1)L < -\bar{V}$, $I \geqslant I_1$ doesn't satisfy the first constraint and the objective is a decreasing function for $I < I_1$. If I_2 exists, it is the largest $I < I_1$ satisfying $\bar{V} + B(I_2)L \geqslant 0$ and is optimal. If I_2, which satisfies $\bar{V} + B(I_2)L = 0$, doesn't exist then there is no solution as $B(I)$ is a strictly decreasing function and the constraint $B(I)L \geqslant -\bar{V}$ can never be satisfied.

When I_1 doesn't exist, the objective is an increasing function for $I \geqslant 0$. If $B(0)L \geqslant -\bar{V}$ then $I = 0$ satisfies both constraints of the problem and is optimal. If $B(0)L < -\bar{V}$ then there is no solution as $B(I)$ is strictly decreasing and the constraint $B(I)L \geqslant -\bar{V}$ can never hold.

Proof for Proposition 5.3.3

$-\bar{V} \geqslant B(I)L$ can never hold for any I as $\bar{V} > 0$.

Proof for Proposition 5.3.4

When I_3 exists and $-\bar{V} \geqslant B(I_3)L$ holds, I_3, which is the unconstrained optimal solution, satisfies both constraints and is optimal. When $-\bar{V} < B(I_3)L$, $I \leqslant I_3$ doesn't satisfy -$-\bar{V} \geqslant B(I)L$ as $B(I)$ is strictly decreasing. Since the objective is increasing for $I > I_3$, I_2 which is the smallest I satisfying $-\bar{V} \geqslant B(I)L$ is optimal. Here, I_2 exists because $I_2 > I_3$ and $B(I)$ is strictly decreasing.

When I_3 doesn't exist, the objective is increasing for $I \geqslant 0$ and hence the smallest feasible I is optimal. If $-\bar{V} \geqslant B(0)L$ then $I = 0$ satisfies both the constraints and is optimal. If $-\bar{V} < B(0)\,L$ then $I = I_2$ is the smallest feasible I with $\bar{V} + B(I)L = 0$ and is optimal.

Proof for Theorem 5.3.5

The results from Propositions 5.3.1-5.3.4 are combined by finding the optimal of the two sub-cases.

1. When I_1 and I_3 exist :

If $L \geqslant \frac{-\bar{V}}{B(I_1)}$, the optimal solution is one of ($I = I_1$, $p_1 = \bar{V} + B(I_1)L$, $p_0 = 0$) and ($I = I_2$, $p_1 = 0$, $p_0 = 0$) with objective function values $B(I_1)F_1 + I_1 + \bar{V}k$ and $B(I_2)F_0 + I_2$ respectively. The former solution is optimal if:

$$\begin{aligned} B(I_2)F_0 + I_2 &\geqslant B(I_1)F_1 + I_1 + \bar{V}k \\ \iff I_2 - I_1 &\geqslant F_1(B(I_1) - B(I_2)) \\ \iff \frac{-1}{F_1} &\leqslant \frac{B(I_1) - B(I_2)}{I_1 - I_2} \end{aligned}$$

Since $B(I_1) \geqslant \frac{-\bar{V}}{L} = B(I_2)$, $I_1 \leqslant I_2$ as B is strictly decreasing. Using the mean value theorem, $\exists I'$ s.t $I_1 \leqslant I' \leqslant I_2$ and $B'(I') = \frac{B(I_1)-B(I_2)}{I_1-I_2}$. As $B'' > 0$, $B'(I_1) \geqslant B'(I') \implies \frac{-1}{F_1} \leqslant \frac{B(I_1)-B(I_2)}{I_1-I_2}$. Hence, the above equation is true.

If $\frac{-\bar{V}}{B(I_1)} \geqslant L \geqslant \frac{-\bar{V}}{B(I_3)}$ then both sub-cases give the same optimal solution ($I^* = I_2$, $p_1^* = 0$, $p_0^* = 0$).

If $\frac{-\bar{V}}{B(I_3)} \geqslant L \geqslant \frac{-\bar{V}}{B(0)}$ then the optimal solution is one of ($I = I_3$, $p_1 = 0$, $p_0 = 0$) and ($I = I_2$, $p_1 = 0$, $p_0 = 0$) with objective function values $B(I_3)F_0 + I_3$ and $B(I_2)F_0 + I_2$ respectively. The former solution is optimal if:

$$\begin{aligned} B(I_2)F_0 + I_2 &\geqslant B(I_3)F_0 + I_3 \\ \iff \frac{-1}{F_0} &\geqslant \frac{B(I_3) - B(I_2)}{I_3 - I_2} \end{aligned}$$

Since $B(I_3) \leqslant \frac{-\bar{V}}{L} = B(I_2)$, $I_3 \geqslant I_2$ as B is strictly decreasing. Using the mean value theorem, $\exists I'$ s.t $I_2 \leqslant I' \leqslant I_3$ and $B'(I') = \frac{B(I_3)-B(I_2)}{I_3-I_2}$. As $B'' > 0$, $B'(I_3) \geqslant B'(I') \implies \frac{-1}{F_0} \geqslant \frac{B(I_3)-B(I_2)}{I_3-I_2}$. Hence, the above equation is true.

If $\frac{-\bar{V}}{B(0)} \geqslant L$ then $(I^* = I_3,\ p_1^* = 0,\ p_0^* = 0)$ is only solution from the two sub-cases as one of them was infeasible.

2. When I_1 exists but not I_3:

If $L \geqslant \frac{-\bar{V}}{B(I_1)}$ then $(I = I_1, p_1 = \bar{V} + B(I_1)L, p_0 = 0)$ and $(I = I_2, p_1 = 0, p_0 = 0)$ are compared and $B(I_2)F_0 + I_2 \geqslant B(I_1)F_1 + I_1 + \bar{V}k$ holds. If $\frac{-\bar{V}}{B(I_1)} \geqslant L \geqslant \frac{-\bar{V}}{B(0)}$ then both sub-cases give the same optimal solution $(I^* = I_2,\ p_1^* = 0,\ p_0^* = 0)$. If $\frac{-\bar{V}}{B(0)} \geqslant L$ then $(I^* = 0,\ p_1^* = 0,\ p_0^* = 0)$ is only solution from the two sub-cases as one of them was infeasible.

3. When both I_1 and I_3 don't exist:

If $L \geqslant \frac{-\bar{V}}{B(0)}$ then $(I = 0,\ p_1 = \bar{V} + B(0)L,\ p_0 = 0)$ and $(I = I_2,\ p_1 = 0,\ p_0 = 0)$ with objective function values $B(0)F_1 + \bar{V}k$ and $B(I_2)F_0 + I_2$ respectively are compared. The former solution is optimal if:

$$
\begin{aligned}
& B(I_2)F_0 + I_2 \geqslant B(0)F_1 + \bar{V}k \\
\iff\ & (B(I_2) - B(0))F_1 \geqslant -I_2 \\
\iff\ & \frac{B(I_2) - B(0)}{I_2} \geqslant \frac{-1}{F_1}
\end{aligned}
$$

Using the mean value theorem, $\exists I'$ s.t $I_2 \leqslant I' \leqslant 0$ and $B'(I') = \frac{B(I_2)-B(0)}{I_2}$. And $B'(I') \geqslant \frac{-1}{F_1}$ as I_1 doesn't exist. Hence, the above equation is true.

If $\frac{-\bar{V}}{B(0)} \geqslant L$ then $(I^* = 0,\ p_1^* = 0,\ p_0^* = 0)$ is only solution from the two sub-cases as one of them was infeasible.

Proof for Theorem 5.3.6

The results follow from Propositions 5.3.1-5.3.4, and the above solutions are the only feasible solutions from the two sub-cases.

Proof for Theorem 5.3.7

Propositions similar to 5.3.1-5.3.4 can be derived for case 2. The proof is then similar to Theorem 5.3.5.

Proof for Theorem 5.3.8

The proof is similar to Theorem 5.3.6.

Appendix E

Appendix for Chapter 6

The omitted proofs from Chapter 6 are provided here.

Proof for Theorem 6.3.1

Before presenting the proof, we define some useful notation. Define $A = [a^1, a^2, ..., a^T]^\top$ and the concatenated vector of charging profiles $x^\dagger := [x_1^{1^\dagger}, \cdots, x_N^{1^\dagger}, \cdots\cdots, x_1^{t^\dagger}, \cdots, x_N^{t^\dagger}]^\top$.

First, we show that for any coalition $\mathcal{C} \subset [N]$, the resulting game is a strongly monotone game. To ensure this it is sufficient to verify that (i) the strategy set is convex, and (ii) the game Jacobian is positive definite [60]. The convexity of strategy sets hold because the strategy sets X_i are simplices. Next, to compute the game Jacobian, we define $\mathcal{G}_i(x) = \frac{\partial c_{\mathcal{C}}(x)}{\partial x_i}$ if $i \in \mathcal{C}$, and $\mathcal{G}_i(x) = \frac{\partial c_i(x)}{\partial x_i}$ if $i \notin \mathcal{C}$.

It can be verified that the game Jacobian, $J(x)$, is $J(x) = \nabla \mathcal{G}(x) = \Theta$, where $\Theta := I_T \otimes \big(b(\mathbf{1}_N \mathbf{1}_N^\top + \mathbf{C}) + \mu\big)$, which is guaranteed to be a positive definite matrix by Lemma 6.3.2. Thus, for any $\mathcal{C}$, the game is strongly monotone, and the equilibrium is unique.

We now introduce two optimization problems:

$$
\begin{aligned}
P_{\mathcal{C}}(x_{-\mathcal{C}}^\dagger) : \min_{x_{\mathcal{C}} \in X_{\mathcal{C}}} & \sum_{j \in \mathcal{C}} c_j(x_{\mathcal{C}}, x_{-\mathcal{C}}^\dagger) \\
\text{s.t.} & \sum_{t' \in [T]} x_j^{t'} = d_j \quad \forall j \in \mathcal{C}, \\
\forall\ i \in [N] \backslash \mathcal{C},\ P_i(x_{-i}^\dagger) : \min_{x_i \in X_i} & \ c_i(x_i, x_{-i}^\dagger) \\
\text{s.t.} & \sum_{t' \in [T]} x_i^{t'} = d_i.
\end{aligned}
$$

By definition of $\mathcal{C}-$Nash equilibrium, charging stations within $\mathcal{C}$ jointly solve the optimization problem $P_{\mathcal{C}}(x_{-\mathcal{C}}^\dagger)$ when the charging profiles of other stations $x_{-\mathcal{C}}^\dagger$ is known. Similarly, each player $i \in [N]\backslash\mathcal{C}$ solves the optimization problem $P_i(x_{-i}^\dagger)$ when the charging profile of other players $x_{-i}^\dagger$ is known. Note that each of the optimization problems has linear constraints, and hence Karush–Kuhn–Tucker (KKT) conditions are necessary for optimality. We solve the KKT conditions of all these problems simultaneously to get a unique solution. Since we established that the equilibrium is unique, this unique solution to the KKT conditions is the $\mathcal{C}-$Nash equilibrium.

The KKT conditions can be written in a compact form as in (E.1), where $\lambda = (\lambda_i)_{i\in[N]}$ are the Lagrange multipliers associated with the linear constraint of every charging station.

$$\begin{aligned}
&\Theta x^\dagger = (I_T \otimes \mu)\bar{x} - (I_T \otimes \mathbf{1}_N)A - (\mathbf{1}_T \otimes I_N)\lambda \\
&(\mathbf{1}_T^\top \otimes I_N)x^\dagger = d,
\end{aligned}$$

where $\Gamma := (\mathbf{1}_T^\top \otimes I_N)\Theta^{-1}(\mathbf{1}_T \otimes I_N)$. We prove in Lemma 6.3.2 that Γ and Θ are invertible. Using this fact and solving for λ in the above equations, we obtain a closed-form expression for the $\mathcal{C}-$Nash equilibrium below.

$$x^\dagger = \Psi_1(\Gamma,\Theta)\left((I_T \otimes \mu)\bar{x} - (I_T \otimes \mathbf{1}_N)A\right) + \Psi_2(\Gamma,\Theta)d,$$

with $\Psi_1(\Gamma,\Theta) = \left(I_{NT} - \Theta^{-1}(\mathbf{1}_T \otimes I_N)\Gamma^{-1}(\mathbf{1}_T^\top \otimes I_N)\right)\Theta^{-1}$ and $\Psi_2(\Gamma,\Theta) = \Theta^{-1}(\mathbf{1}_T \otimes I_N)\Gamma^{-1}$. Expanding the terms in terms of the time index, it can be verified that the above closed-form solution is equivalent to the one stated in the theorem, thereby completing the proof.

Proof of Lemma 6.3.2

Note that $\Theta = I_T \otimes \left(b(\mathbf{1}_N\mathbf{1}_N^\top + \mathbf{C}) + \mu\right)$ is positive definite if and only if each of the terms in the Kronecker product is positive definite [141]. Hence, it is sufficient to show that $Q := b(\mathbf{1}_N\mathbf{1}_N^\top + \mathbf{C}) + \mu$ is a positive definite matrix. First, note that Q is a symmetric matrix. Additionally, for any $z \in \mathbb{R}^N$, it holds that

$$z^\top Q z = b(\mathbf{1}_N^\top z)^2 + b(\sum_{i\in\mathcal{C}} z_i)^2 + \sum_{i\in\mathcal{C}} \mu_i z_i^2 + \sum_{i\in N\setminus\mathcal{C}} (b+\mu_i) z_i^2,$$

which is strictly positive for all $z \neq 0$.

Next, we show that Γ is invertible by calculating its inverse. Indeed,

$$\begin{aligned}
\Gamma^{-1} &= \left((\mathbf{1}_T^\top \otimes I_N)\Theta^{-1}(\mathbf{1}_T \otimes I_N)\right)^{-1} \\
&= \left(\mathbf{1}_T^\top\mathbf{1}_T \otimes \left(b(\mathbf{1}_N\mathbf{1}_N^\top + C) + \mu\right)^{-1}\right)^{-1} \\
&= \frac{1}{T}\left(b(\mathbf{1}_N\mathbf{1}_N^\top + C) + \mu\right) = \frac{1}{T}Q.
\end{aligned}$$

This completes the proof.

Proof of Theorem 6.3.7

For a subset $\mathcal{S}$ of charging stations, the total cost under Nash equilibrium is lower than $\mathcal{C}-$Nash equilibrium if and only if

$$\sum_{i\in\mathcal{S}} c_i(x^*) \leqslant \sum_{i\in\mathcal{S}} c_i(x^\dagger).$$

In the rest of the proof, we calculate these costs in terms of the game parameters. For every $t \in [T]$, define $F^t := T\alpha^t - \sum_{t'\in[T]} \alpha^{t'}$. Since $\sum_{t\in[T]} \alpha^t = 1$, it holds that $F^t = T\alpha^t - 1$ and $\sum_{t\in[T]} F^t = 0$. Using Assumption 6.3.1 along with Proposition 6.3.6, we get the following for $\mathcal{C}-$Nash equilibrium:

$$x_i^{\dagger t} = \frac{d_i}{T} + \frac{F^t}{T}\Delta_i^\dagger, \qquad \mathbf{1}_N^\top x^{\dagger t} = \frac{D}{T} + \frac{F^t}{T}\Delta^\dagger$$
$$x_i^{\dagger t} - \bar{x}_i^t = \frac{F^t}{T}(\Delta_i^\dagger - d_i),$$

where $D = \sum_{i\in[N]} d_i$. Using the above equations and $\sum_{t\in[T]} F^t = 0$, the cost of station i at $\mathcal{C}-$Nash equilibrium is:

$$\begin{aligned} c_i(x^\dagger) &= ad_i + \left(b \sum_{t\in[T]} (1_N^\top x^{t\dagger}) x_i^{t\dagger} \right) + \frac{\mu_i}{2}\|x_i^\dagger - \bar{x}_i\|^2 \\ &= ad_i + \frac{b}{T^2}\left(TDd_i + \sum_{t\in[T]} (F^t)^2 \Delta^\dagger \Delta_i^\dagger \right) + \frac{\mu_i}{2T^2} \sum_{t\in[T]} (F^t)^2 (\Delta_i^* - d_i)^2. \end{aligned}$$

Consequently, the total cost of charging stations in $\mathcal{S}$ is:

$$\sum_{i\in\mathcal{S}} c_i(x^\dagger) = \left(aD_{\mathcal{S}} + \frac{bDD_{\mathcal{S}}}{T} \right) + \frac{\mathcal{F}}{T^2}\left(\sum_{i\in\mathcal{S}} b\Delta^\dagger\Delta_i^\dagger + \frac{\mu^i}{2}(\Delta_i^\dagger - d_i)^2 \right),$$

where $D_{\mathcal{S}} := \sum_{i\in\mathcal{S}} d_i$, $\mathcal{F} := \sum_{t\in[T]}(F^t)^2$. Analogously, we can also compute the total cost at Nash equilibrium. Using these results, we conclude

$$\begin{aligned} & aD_{\mathcal{S}} + \frac{bDD_{\mathcal{S}}}{T} + \frac{\mathcal{F}}{T^2}\left(\sum_{i\in\mathcal{S}} b\Delta^\dagger\Delta_i^\dagger + \frac{\mu^i}{2}(\Delta_i^\dagger - d_i)^2 \right) \\ & \geqslant aD_{\mathcal{S}} + \frac{bDD_{\mathcal{S}}}{T} + \frac{\mathcal{F}}{T^2}\left(\sum_{i\in\mathcal{S}} b\Delta^*\Delta_i^* + \frac{\mu^i}{2}(\Delta_i^* - d_i)^2 \right) \\ & \iff \sum_{i\in\mathcal{S}} f_i^\dagger(\mu, b, d) - f_i^*(\mu, b, d) \geqslant 0. \end{aligned}$$

Proof of Theorem 6.3.11

For every $t \in [T]$, define $G^t := \sum_{t' \in [T]} a^{t'} \left(T\delta^{tt'} - 1\right)$. Using this, the expression for $x^{\dagger t}$ from Proposition 6.3.10 can be re-written as follows

$$x^{\dagger t} = \frac{d}{T} - \frac{1}{T}\left(b(\mathbf{1}_N \mathbf{1}_N^\top + C) + \mu\right)^{-1} \mathbf{1}_N G^t. \tag{E.2}$$

Consequently, it holds that

$$\begin{aligned} x_i^{\dagger t} &= \frac{d_i}{T} - \frac{G^t}{T}\Gamma_i^\dagger \\ \mathbf{1}_N^\top x^{\dagger t} &= \frac{D}{T} - \frac{G^t}{T}\Gamma^\dagger \\ x_i^{\dagger t} - \bar{x}_i^t &= -\frac{1}{T}\Gamma_i^\dagger G^t. \end{aligned} \tag{E.3}$$

Consequently, combining previous equations with (**??**), for every $i \in [N]$,

$$\begin{aligned} c_i(x^\dagger) &= \sum_{t \in [T]} a^t x_i^{\dagger t} + \frac{b}{T^2} \sum_{t \in [T]} \left(D - G^t \Gamma^\dagger\right)\left(d_i - G^t \Gamma_i^\dagger\right) \\ &\quad + \frac{\mu_i (\Gamma_i^\dagger)^2}{2T^2} \sum_{t \in [T]} (G^t)^2 \\ &= \sum_{t \in [T]} a^t \left(\frac{d_i}{T} - \frac{G^t}{T}\Gamma_i^\dagger\right) + \frac{b}{T^2}\left(TDd_i + \Gamma^\dagger \Gamma_i^\dagger \sum_{t \in [T]} (G^t)^2\right) \\ &\quad + \frac{\mu_i (\Gamma_i^\dagger)^2}{2T^2} \sum_{t \in [T]} (G^t)^2, \end{aligned} \tag{E.4}$$

where the last equality is because $\sum_{t \in [T]} G^t = \sum_{t,t' \in [T]} a^{t'} \left(T\delta^{tt'} - 1\right) = 0$. Additionally, summing (E.4) for all $i \in \mathcal{S}$, we obtain

$$\begin{aligned} \sum_{i \in \mathcal{S}} c_i(x^\dagger) &= \sum_{t \in [T]} a^t \left(\frac{D_{\mathcal{S}}}{T} - \frac{G^t}{T} \sum_{i \in \mathcal{S}} \Gamma_i^\dagger\right) + \frac{bDD_{\mathcal{S}}}{T} \\ &\quad + \frac{b\Gamma^\dagger \sum_{i \in \mathcal{S}} \Gamma_i^\dagger \sum_{t \in [T]} (G^t)^2}{T^2} + \frac{b}{2T^2} \sum_{t \in [T]} (G^t)^2 \sum_{i \in \mathcal{S}} \frac{\mu^i}{b} (\Gamma_i^\dagger)^2 \\ &= \left(\frac{AD_{\mathcal{S}} - \sum_{i \in \mathcal{S}} \Gamma_i^\dagger \sum_{t \in [T]} a^t G^t}{T} + \frac{bDD_{\mathcal{S}}}{T}\right) \\ &\qquad + \frac{b \sum_{t \in [T]} (G^t)^2}{T^2} \left(\Gamma^\dagger \sum_{i \in \mathcal{S}} \Gamma_i^\dagger + \sum_{i \in \mathcal{S}} \frac{\mu^i}{2b} (\Gamma_i^\dagger)^2\right), \end{aligned}$$

where $A := \sum_{t\in[T]} a^t$. Analogously, we obtain the cost at Nash equilibrium as follows

$$\sum_{i\in\mathcal{S}} c_i(x^*) = \left(\frac{AD_{\mathcal{S}} - \sum_{i\in\mathcal{S}} \Gamma_i^* \sum_{t\in[T]} a^t G^t}{T} + \frac{bDD_{\mathcal{S}}}{T}\right)$$
$$+ \frac{b\sum_{t\in[T]}(G^t)^2}{T^2}\left(\Gamma^* \sum_{i\in\mathcal{S}} \Gamma_i^* + \sum_{i\in\mathcal{S}} \frac{\mu^i}{2b}(\Gamma_i^*)^2\right).$$

Using the above equations, we obtain that $\sum_{i\in\mathcal{S}} c_i(x^\dagger) > \sum_{i\in\mathcal{S}} c_i(x^*)$ if and only if $g_{\mathcal{S}}^\dagger(\mu, b) - g_{\mathcal{S}}^*(\mu, b) \geqslant (\Gamma^\dagger - \Gamma^*)h(A)$. This completes the proof.