Price Optimization with Practical Constraints

Xiaojie Wang, Hsin-Chan Huang, Lanshan Han and Alvin Lim*

Nielsen Precima, LLC

April 21, 2021

Abstract

In this paper, we study a retailer price optimization problem which includes the practical constraints: maximum number of price changes and minimum amount of price change (if a change is recommended). We provide a closed-form formula for the Euclidean projection onto the feasible set defined by these two constraints, based on which a simple gradient projection algorithm is proposed to solve the price optimization problem. We study the convergence and solution quality of the proposed algorithm. We extend the base model to include upper/lower bounds on the individual product prices and solve it with some adjustments to the gradient projection algorithm. Numerical results are reported to demonstrate the performance of the proposed algorithm.

1 Introduction

Price optimization is a classical problem in the business operation of various product and service industries. It is well-known from economics that the demand of a product is affected by its price. Typically, for a consumer product, a lower price for the product leads to more demand for it. However, when multiple products are offered, the situation is often more complex due to the cross-product relationships. The products can be substitutes or complements of each other, and hence the demand on a product depends not only on its own price but also the prices of related products. We often use demand functions to quantify how demands of products are affected by their prices.

When price optimization is applied in various industries under different settings, different sets of business rules need to be imposed to guarantee that optimal prices are reasonable and practical to implement. In this paper, we are particularly interested in finding the optimal prices for a set of products in a brick-and-mortar grocer to maximize profits. In grocery retailing, certain specific business considerations need to be taken into account in setting prices of products. First, prices change frequently to reflect commodity cost changes, supplier incentives, response to consumer behavior, response to market conditions, etc., and price optimization is often implemented to obtain product shelf prices for various store locations. While a growing adoption of electronic shelf tags has been observed recently

---

*Corresponding author, alvin.lim@emory.edu

([Bhutani and Bhardwaj 2018](#)), the use of paper tags that are implemented using limited manual labor is still the norm. Due to the limited (and often expensive) resources, grocers only update in each week the prices of a limited number of products – those that have the most significant impact on revenues and profits. Second, while demand functions are typically continuous function defined in the nonnegative orthant and hence can describe the demand change expected from an arbitrary small change in price, consumers may not be sensitive enough to react to extremely small changes in price. So, when a price decrease is too small, the expected increase in demand for a product may not materialize. On the other hand, if the price increase is too small, while the demand will likely stay constant (leading to better profits), the cost of implementing the price increase may not justify the return. Therefore, a minimum price change is often imposed on products that are candidates for price changes. These two practical considerations significantly complicate the price optimization problem. In particular, they implicitly introduce certain combinatorial structures into the traditional price optimization problem, which is typically a continuous optimization problem (often also convex under reasonable assumptions) ([Gallego et al. 2019](#)). A mixed integer programming formulation is, therefore, needed to capture these two practical considerations. This type of optimization problem is notoriously challenging to solve to global optimality ([Wolsey 2007](#)).

This paper represents a first effort to include the aforementioned practical considerations into price optimization models. We explore various mathematical properties of the resulting formulation and investigate the use of a gradient projection algorithm to find a solution of the optimization problem. The algorithm is demonstrated to produce near-optimal solutions both theoretically and empirically. Moreover, we hope that the proposed approach may be applied to similar problems, such as marketing mix optimization ([Huang et al. 2021](#)) and portfolio optimization ([Bertsimas and Shioda 2009](#)), when analogous practical constraints are imposed.

The rest of the paper is organized as follows. In Section 2, we review existing research on price optimization. In Section 3, we provide mathematical models of the price optimization problem under investigation and present preliminaries regarding the mathematical properties of the models. In Section 4, we present a gradient projection algorithm to solve the price optimization problems as well as analyses on various properties of the algorithm such as convergence and performance bounds. Numerical results on randomly generated data sets as well as real-world retailer data sets are presented in Section 5. We conclude the paper with a summary of our contributions and a discussion on possible future research directions in Section 6.

2   Literature Review

In this section, we review the state of the art in retail price optimization under various settings. According to how the pricing decisions are made along the temporal dimension, the literature can be divided into "dynamic" pricing and "static" pricing categories. Dynamic pricing refers to the setting of time-varying prices over a selling period, and prices are typically set based on the changing expected availability of and expected demand on the product over time. Static pricing, on the other hand, refers to the setting of fixed prices over a selling period of products that are continuously replenished and virtually non-perishable ([Kunz and

Crone 2014). While dynamic pricing has seen applications in retail via e-commerce (Boyd and Bilegan 2003), we will focus on static pricing, which is the predominant practice among brick-and-mortar retailers.

Although more research efforts have been invested into dynamic pricing (Bitran and Caldentey 2003, Chiang et al. 2007, Goensch et al. 2009, Elmaghraby and Keskinocak 2003, Weatherford and Bodily 1992, McGill and Van Ryzin 1999, Gallego et al. 2004, Liu and Van Ryzin 2008, Bront et al. 2009), static pricing dominates retail pricing applications (Kunz and Crone 2014), as frequent price changes are not practical to implement at most brick-and-mortar retailers. Even at online retailers, where one may consider frequent price changes come at negligible costs, frequent changes in price may lead consumers' negative perception of an online retailer's brand image, and therefore not desirable for most retailers, particularly for fashion retailers (Ferreira et al. 2016, Levy et al. 2004). However, despite the prevalence of static pricing, limited studies on the static pricing problem exists in the operations research and econometric literature (Kunz and Crone 2014).

Kunz and Crone (2014) observed the research scarcity and importance of static retail pricing, including the case of grocery pricing. Levy et al. (2004) examined the practice of retail pricing and found seven critical elements that are essential for success in retail pricing: price sensitivity effects, substitution effects, dynamic effects of price promotions, segment-based pricing effects, cross-category effects, retailer costs and discounts, and retail competition. Several additional factors should also be considered in the implementation of retail pricing: market factors, grouping of products (category management), continuous learning, psychological price thresholds and reference effects, costs of changing prices, and the quality of data.

The classical method in static retail pricing employs rule-based approaches in category management, which is still the prevailing method used in the industry because of the lack of quality data and models to describe the complex interactions among products as well as the time and expertise to conduct in-depth analyses (Hall et al. 2003). Some studies on static retail pricing in the literature are concerned with "multi-product pricing", which focus on understanding interactions and substitutions among products, resulting in the formulation of the pricing problem as a mixed-integer program, which allows for a wide range of practical constraints but is NP-complete and its efficient solution remains a challenge (Dobson and Kalish 1988, Oren et al. 1987).

A demand model is considered the core of a static retail price optimization system that serves as a link between the input data and the optimization model (van Ryzin 2005). Types of demand models (Kunz and Crone 2014) include:

- absolute: linear, multiplicative, power, exponential, double-log, etc.;

- relative: multiple competitive interaction (MCI) and discrete choice model; and

- other forms: reference price, game theory, and hedonic pricing.

Many efforts were invested in discrete choice models to account for customer choice behaviors. A comprehensive review of choice-based pricing and revenue management can be found in Strauss et al. (2018). Discrete choice models are implemented using multinomial logit (MNL) (Song and Xue 2007, Dong et al. 2009), general attraction (Keller et al. 2014), nested logit

(Huh and Li 2015), exponomial choice (Alptekinoğlu and Semple 2016), generalized extreme value (GEV) (Zhang et al. 2018), and non-parameteric (Farias et al. 2013, Rusmevichientong et al. 2006) and other techniques. A more comprehensive review of demand functions in decision modeling can be found in Huang et al. (2013). We note that static retail pricing is mainly motivated by demand interdependencies and the consideration of cross price effects is therefore fundamental (Kunz and Crone 2014).

Estimation of the unknown parameters in the demand model is essential, because models without quality input data and parameter estimates have little value. In grocery retailing, the challenge of estimating a large number of parameters in a demand model likely contributed to the limited implementation of data-driven static pricing until the recent prominence in the adoption of data analytics. Bayesian estimation models are used to predict sales as a function of price and other factors (Kalyanam 1996, Seetharaman et al. 2005). An application at Rue La La, an online fashion retailer, uses machine learning techniques (bagged regression tree) to understand historical sales and predict future demands on changing prices (Ferreira et al. 2016). With the predicted demand, Ferreira et al. (2016) developed an integer programming model to maximize the total revenue, which is difficult to solve in large scale instances, and thus a linear programming bound algorithm was proposed to help achieve optimality.

The parameters of the demand function can also be estimated through online learning. A common framework in this area is to obtain information on demands through price experimentation based on some parametric demand choice models, and through experimentation, pricing polices are designed to minimize regret (Broder and Rusmevichientong 2012, Ferreira et al. 2018, Özer and Phillips 2012). More examples of demand function modeling in pricing and revenue management applications can be found in Mišić and Perakis (2020).

3 Problem Statement

Throughout this paper, we use lower and upper case letters to represent scalars, bold lower case letters to represent vectors, and bold upper case letters to represent matrices. We use superscript to index vectors or matrices and use subscripts to index scalars. All vectors are assumed to be column vectors. We denote the n-dimensional real vector space as $\mathbb{R}^n$, its nonnegative orthant as $\mathbb{R}^n_+$, and the interior of the nonnegative orthant as $\mathbb{R}^n_{++}$. For a vector $\boldsymbol{v} \in \mathbb{R}^n$, its i-th element is denoted by v_i and its transpose is written as $\boldsymbol{v}^T$. For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, its (i,j) element is denoted by a_{ij} and its transpose is written as $\boldsymbol{A}^T$. We denote the all zero vector and all one vector of length n by $\boldsymbol{0}_n$ and $\boldsymbol{1}_n$, respectively. We denote the $n \times n$ identity matrix by $\boldsymbol{I}_n$. The two-norm of a vector $\boldsymbol{v} \in \mathbb{R}^n$ is denoted by $\|\boldsymbol{v}\|_2$, i.e., $\|\boldsymbol{v}\|_2 = \sqrt{\boldsymbol{v}^T\boldsymbol{v}}$. The infinite-norm of a vector $\boldsymbol{v} \in \mathbb{R}^n$ is denoted by $\|\boldsymbol{v}\|_\infty$, i.e., $\|\boldsymbol{v}\|_\infty = \max_{i=1,\cdots,n}\{|v_i|\}$. We also define the zero-norm of a vector $\boldsymbol{v} \in \mathbb{R}^n$, denoted by $\|\boldsymbol{v}\|_0$, as the cardinality of the set of nonzero elements of $\boldsymbol{v}$. Given a function $g : \mathbb{R}^n \mapsto \mathbb{R}$, we let $\nabla g(\boldsymbol{x})$ be its gradient evaluated at $\boldsymbol{x}$, and let $\nabla_i g(\boldsymbol{x})$ be its i-th element. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and a subset $\alpha \subseteq \{1, \cdots, n\}$, we let $\boldsymbol{A}_{\alpha,\bullet}$ be the $|\alpha| \times n$ submatrix formed by the rows of $\boldsymbol{A}$ with indices in α, with $|\alpha|$ being the cardinality of α. Similarly we let $\boldsymbol{A}_{\bullet,\alpha}$ be the $n \times |\alpha|$ submatrix formed by the columns of $\boldsymbol{A}$ with indices in α, and $\boldsymbol{A}_{\alpha,\alpha}$ be the principal submatrix with column and row indices in α. In the special case where α is a singleton, i.e., $\alpha = \{i\}$, we simplify the notation by using $\boldsymbol{A}_{\bullet,i}$ and $\boldsymbol{A}_{i,\bullet}$ to represent the ith column and

ith row of $\boldsymbol{A}$, respectively. For a vector $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{v}_\alpha$ is the subvector of length $|\alpha|$ composed of elements of $\boldsymbol{v}$ with indices in α.

We consider a set of n products, indexed by $i = 1, \cdots, n$. Let $\boldsymbol{p} = (p_i)_{i=1}^n$ be the vector of the unit prices of the products, $\boldsymbol{c} = (c_i)_{i=1}^n$ be the vector of the unit costs of the products, and $v(\boldsymbol{p}) : \mathbb{R}^n \mapsto \mathbb{R}^n$ be the *(volume of) demand function* which estimates the expected volume of demand as a function of the prices of the products. The price optimization problem is to maximize the profit, possibly under various business rules. Mathematically, this problem is written as:

$$\max_{\boldsymbol{p} \in \mathcal{P}} \ Z(\boldsymbol{p}) \triangleq (\boldsymbol{p} - \boldsymbol{c})^T v(\boldsymbol{p}), \tag{1}$$

where $\mathcal{P}$ represents the set of all price vectors satisfying defined business rules. While there are many possible business rules such as the maximum percentage change from a baseline price, maximum price gap to a competitor's prices, or price relationship requirements for different package sizes of the same product, these rules can typically be written as linear inequality constraints. These rules create a convex set of feasible price vectors (a polyhedron), which is often considered benign in optimization approaches. On the other hand, there are other business rules that cause significant challenges by creating non-convex feasible set of price vector $\mathcal{P}$. In this paper, we consider two such constraints that are of significant practical relevance:

- **Maximum number of price changes**. For most retailers, the price optimization is conducted routinely with the most recent data collected. While it is possible for an online retailer to change the price of many or even all of their products in a very short time period, it is practically impossible for a traditional brick-and-mortar store to do so, as they only have limited resources to update the paper price tags on the shelves. Therefore, it is of crucial importance that we only change the prices of a priority subset of products that can generate the most incremental profits. Let the baseline price vector be $\boldsymbol{p}^0 = (p_i^0)_{i=1}^n$. Assuming that the resources available allow changing the prices of at most k products, this requirement can be translated to a constraint of the following format:

$$\|\boldsymbol{p} - \boldsymbol{p}^0\|_0 \ \leq \ k.$$

- **Minimum amount of price change**. As mentioned earlier, extremely small price changes may not deliver the expected return in profits due to the consumers insensitivity to small price changes in the case of a price decrease and the cost of implementing the change in the case of a price increase. In addition, many retailers also apply price ending rules as part of their price image strategy, such as prices ending with 9 cents. To address these two issues, in practice we often need to impose a minimum price change requirement. More specifically, product i's price can either stay the same or increase/decrease by a minimum amount $\delta_i \geq 0$. Mathematically, we have:

$$p_i \ \in \ \{p_i^0\} \cup [p_i^0 + \delta_i, \infty) \cup (-\infty, p_i^0 - \delta_i] \ \subseteq \ \mathbb{R}.$$

We write $\mathcal{P}_{\delta_i}^i \triangleq \{p_i^0\} \cup [p_i^0 + \delta_i, \infty) \cup (-\infty, p_i^0 - \delta_i]$ and $\mathcal{P}_{\boldsymbol{\delta}} \triangleq \bigotimes_{i=1}^n \mathcal{P}_{\delta_i}^i \subseteq \mathbb{R}^n$ with $\boldsymbol{\delta} \triangleq (\delta_i)_{i=1}^n$. Sometimes, we also have upper and lower bounds, denoted by u_i and l_i, on product i's price, in which case we write $\mathcal{P}_{\delta_i, l_i, u_i}^i \triangleq \{p_i^0\} \cup [p_i^0 + \delta_i, u_i] \cup [l_i, p_i^0 - \delta_i]$ and
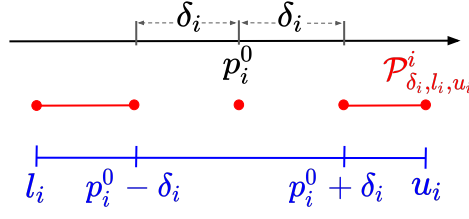
5

Figure 1: Set of allowed prices for a single product shown in red.

$\mathcal{P}_{\boldsymbol{\delta},\boldsymbol{l},\boldsymbol{u}} \triangleq \bigotimes_{i=1}^n \mathcal{P}_{\delta_i,l_i,u_i}$, with the $\boldsymbol{u}$ and $\boldsymbol{l}$ being vectors of upper bounds and lower bounds, i.e., $\boldsymbol{u} \triangleq (u_i)_{i=1}^n$ and $\boldsymbol{l} \triangleq (l_i)_{i=1}^n$. An illustration of set $\mathcal{P}_{\delta_i,l_i,u_i}$ is given in Figure 1.

With the above two practical requirements, given baseline price vector $\boldsymbol{p}^0$ and price change threshold vector $\boldsymbol{\delta}$, we define

$$\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}} \triangleq \left\{ \boldsymbol{p} \mid \|\boldsymbol{p} - \boldsymbol{p}^0\|_0 \leq k,\, \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{\delta}} \right\},$$

which is the set of feasible price vectors. When the upper/lower bounds on individual products exist, this set of feasible price vectors is

$$\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta},\boldsymbol{l},\boldsymbol{u}} \triangleq \left\{ \boldsymbol{p} \mid \|\boldsymbol{p} - \boldsymbol{p}^0\|_0 \leq k,\, \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{\delta},\boldsymbol{l},\boldsymbol{u}} \right\}.$$

In this paper, we mainly focus on the case without bounds as follows:

$$\begin{aligned} \max_{\boldsymbol{p}} \quad & (\boldsymbol{p} - \boldsymbol{c})^T v(\boldsymbol{p}) \\ \text{s.t.} \quad & \boldsymbol{p} \in \Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}, \end{aligned} \tag{2}$$

but we will also consider the case with upper and lower bounds as a direct extension.

3.1 Linear Demand Functions

In this paper, we consider a linear demand function of the form $v(\boldsymbol{p}) = \boldsymbol{a} - \boldsymbol{D}\boldsymbol{p}$ with $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{D} = (d_{ij})_{i,j=1}^n \in \mathbb{R}^{n \times n}$, where for $i = 1, \cdots, n$, $-d_{ii}$ represents the effect of product i's own price on its demand v_i, and $-d_{ij}$ represents the effect of the cross price p_j of product j on the demand v_i of product i. Typically, $\boldsymbol{D}$ is not symmetric, and the symmetric part of $\boldsymbol{D}$ is denoted by $\boldsymbol{S} \triangleq (\boldsymbol{D} + \boldsymbol{D}^T)$. Often we make the following two assumptions on $\boldsymbol{D}$ or $\boldsymbol{S}$ (Gallego et al. 2019).

(A1). Own price effects are all negative and cross price effects are all positive. Mathematically, we have $\boldsymbol{S} = (s_{ij})_{i,j=1}^n$ is positive definite and $s_{ij} \leq 0$ for all $i \neq j$.

(A2). The demands when prices are all 0's are positive, the costs are all positive, and the demands when the prices are at the costs are nonnegative. Mathematically, we have $\boldsymbol{a} \in \mathbb{R}_{++}^n$, $\boldsymbol{c} \in \mathbb{R}_{++}^n$, and $\boldsymbol{a} - \boldsymbol{D}\boldsymbol{c} \in \mathbb{R}_+^n$.

6

Note that the cross price effects limitation of (A1) restricts us to consider only the price effects of substitute products. Under a linear demand function, it is easy to see that

$$Z(\boldsymbol{p}) = (\boldsymbol{p}-\boldsymbol{c})^T(\boldsymbol{a}-\boldsymbol{D}\boldsymbol{p}) = -\boldsymbol{p}^T\boldsymbol{D}\boldsymbol{p}+\boldsymbol{a}^T\boldsymbol{p}+\boldsymbol{c}^T\boldsymbol{D}\boldsymbol{p}-\boldsymbol{c}^T\boldsymbol{a} = -\frac{1}{2}\boldsymbol{p}^T\boldsymbol{S}\boldsymbol{p}+(\boldsymbol{a}+\boldsymbol{D}^T\boldsymbol{c})^T\boldsymbol{p}-\boldsymbol{c}^T\boldsymbol{a}.$$

Therefore, the base price optimization problem (2) can be written as the following equivalent minimization problem.

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & \mathcal{Q}(\boldsymbol{p}) \triangleq \frac{1}{2}\boldsymbol{p}^T\boldsymbol{S}\boldsymbol{p} - (\boldsymbol{a}+\boldsymbol{D}^T\boldsymbol{c})^T\boldsymbol{p} \\ \text{s.t.} \quad & \boldsymbol{p} \in \Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}. \end{aligned} \tag{3}$$

Define

$$\widehat{\boldsymbol{p}} \triangleq \arg\min_{\boldsymbol{p}} \ \mathcal{Q}(\boldsymbol{p}), \tag{4}$$

i.e., the unconstrained unique minimizer (due to the positive definiteness of $\boldsymbol{S}$) of the function $\mathcal{Q}(\boldsymbol{p})$. It is easy to see that

$$\widehat{\boldsymbol{p}} = \boldsymbol{S}^{-1}(\boldsymbol{a} + \boldsymbol{D}^T\boldsymbol{c}).$$

Let $\widehat{\mathcal{Q}} \triangleq \mathcal{Q}(\widehat{\boldsymbol{p}})$, i.e., the unconstrained optimal value. Let $\mathcal{Q}^*$ be the optimal value of (3). It is clear $\widehat{\mathcal{Q}} \leq \mathcal{Q}^*$. Note that the existence and uniqueness of $\widehat{\boldsymbol{p}}$ is guaranteed by the positive definiteness of $\boldsymbol{S}$. We have the following lemma, which is known in the literature. We include a short proof for completeness.

Lemma 1. *Assume (A1) and (A2) hold. Let $\widehat{\boldsymbol{p}}$ be defined by (4), then*

$$\widehat{\boldsymbol{p}} \geq \boldsymbol{c}.$$

Proof. It is easy to see that

$$\begin{aligned} \widehat{\boldsymbol{p}} - \boldsymbol{c} &= \boldsymbol{S}^{-1}(\boldsymbol{a} + \boldsymbol{D}^T\boldsymbol{c}) - \boldsymbol{c} = \boldsymbol{S}^{-1}(\boldsymbol{a} + \boldsymbol{D}^T\boldsymbol{c} - \boldsymbol{S}\boldsymbol{c}) \\ &= \boldsymbol{S}^{-1}(\boldsymbol{a} + \boldsymbol{D}^T\boldsymbol{c} - (\boldsymbol{D} + \boldsymbol{D}^T)\boldsymbol{c}) = \boldsymbol{S}^{-1}(\boldsymbol{a} - \boldsymbol{D}\boldsymbol{c}). \end{aligned}$$

By Assumption (A1), $\boldsymbol{S}$ is a *Stieltjes* matrix (named after Thomas Joannes Stieltjes), which is necessarily an M-matrix (Berman and Plemmons 1994). Therefore, $\boldsymbol{S}$ is always invertible and $\boldsymbol{S}^{-1}$ is a symmetric nonnegative matrix, i.e., all the elements of $\boldsymbol{S}^{-1}$ are non-negative. By (A2), $(\boldsymbol{a} - \boldsymbol{D}\boldsymbol{c}) \geq 0$. Therefore $\widehat{\boldsymbol{p}} - \boldsymbol{c} \geq 0$, and we have the desired result. □

Lemma 1 shows that when we don't have any constraints on the prices, assumptions (A1) and (A2) ensure that the optimal prices are profitable for each product. We call a price vector $\boldsymbol{p}$ *profitable* if $\boldsymbol{p} \geq \boldsymbol{c}$, where the $\geq$ relationship is element-wise. We next study the case of optimization problem (3). We make the following additional assumption:

(A3). $\boldsymbol{p}^0 - \boldsymbol{\delta}$ is profitable, i.e., $\boldsymbol{p}^0 - \boldsymbol{\delta} \geq \boldsymbol{c}$.

Before we present our main result about profitability, we first prove a technical lemma.

Lemma 2. *Let $\boldsymbol{p}^*$ be a global optimal solution to (3). Define an index set $\omega^* = \{i \,|\, \boldsymbol{p}^*_i = p^0_i\} \cup \{i \,|\, \boldsymbol{p}^*_i = p^0_i + \delta_i\} \cup \{i \,|\, \boldsymbol{p}^*_i = p^0_i - \delta_i\}$, and its complement set $\overline{\omega^*} = \{i \,|\, i \notin \omega^*\}$. Then*

$$\boldsymbol{p}^*_{\overline{\omega^*}} = \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\bullet,\overline{\omega^*}} \boldsymbol{c} - \boldsymbol{S}_{\overline{\omega^*},\omega^*} \boldsymbol{p}^*_{\omega^*} \right). \tag{5}$$

Proof. We claim that $\boldsymbol{p}^*_{\overline{\omega^*}}$ necessarily minimize the following quadratic function of $\boldsymbol{p}_{\overline{\omega^*}}$:

$$\min_{\boldsymbol{p}_{\overline{\omega^*}}} \frac{1}{2} \left[(\boldsymbol{p}^*_{\omega^*})^T \ (\boldsymbol{p}_{\overline{\omega^*}})^T \right] \begin{bmatrix} \boldsymbol{S}_{\omega^*,\omega^*} & \boldsymbol{S}_{\omega^*,\overline{\omega^*}} \\ \boldsymbol{S}_{\overline{\omega^*},\omega^*} & \boldsymbol{S}_{\overline{\omega^*},\overline{\omega^*}} \end{bmatrix} \begin{bmatrix} \boldsymbol{p}^*_{\omega^*} \\ \boldsymbol{p}_{\overline{\omega^*}} \end{bmatrix} + (\boldsymbol{a}^T_{\omega^*} + \boldsymbol{c}^T \boldsymbol{D}_{\bullet,\omega^*}) \boldsymbol{p}^*_{\omega^*} + (\boldsymbol{a}^T_{\overline{\omega^*}} + \boldsymbol{c}^T \boldsymbol{D}_{\bullet,\overline{\omega^*}}) \boldsymbol{p}_{\overline{\omega^*}} \tag{6}$$

Assume for the sake of contradiction that the above claim does not hold. Since $\boldsymbol{S}_{\overline{\omega^*},\overline{\omega^*}}$ is a principal submatrix of $\boldsymbol{S}$ and is hence positive definite, optimization problem (6) has a unique minimizer. Let $\boldsymbol{p}'_{\overline{\omega^*}}$ be the minimizer of (6). By the definition of ω^*, we know that there exists $\lambda > 0$ such that

$$\begin{bmatrix} \boldsymbol{p}_{\omega^*} \\ \boldsymbol{p}_{\overline{\omega^*}} + \lambda (\boldsymbol{p}'_{\overline{\omega^*}} - \boldsymbol{p}^*_{\overline{\omega^*}}) \end{bmatrix} \in \Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}},$$

and clearly this vector also results in an objective value less than that of p^*. This is a contradiction. So the claim holds. And simple algebra leads to Equation (5). $\square$

Theorem 3. *Suppose assumptions (A1)-(A3) hold. Let $\boldsymbol{p}^*$ be an optimal solution of (3), then $\boldsymbol{p}^*$ is profitable, i.e., $\boldsymbol{p}^* \geq \boldsymbol{c}$.*

Proof. Let ω^* be as defined in Lemma 2. Since $\boldsymbol{p}^0 - \boldsymbol{\delta} \geq \boldsymbol{c}$ and $\boldsymbol{\delta} \geq 0$, we must have $\boldsymbol{p}^0 + \boldsymbol{\delta} \geq \boldsymbol{p}^0 \geq \boldsymbol{c}$. Therefore, for any $i \in \omega^*$, $p^*_i \geq c_i$, or equivalently $\boldsymbol{p}^*_{\omega^*} - \boldsymbol{c}_{\omega^*} \geq 0$. For $i \in \overline{\omega^*}$, we know from Lemma 2 that

$$
\begin{aligned}
\boldsymbol{p}^*_{\overline{\omega^*}} - \boldsymbol{c}_{\overline{\omega^*}} &= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\bullet,\overline{\omega^*}} \boldsymbol{c} - \boldsymbol{S}_{\overline{\omega^*},\omega^*} \boldsymbol{p}^*_{\omega^*} \right) - \boldsymbol{c}_{\overline{\omega^*}} \\
&= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\bullet,\overline{\omega^*}} \boldsymbol{c} - \boldsymbol{S}_{\overline{\omega^*},\omega^*} \boldsymbol{p}^*_{\omega^*} - \boldsymbol{S}_{\overline{\omega^*},\overline{\omega^*}} \boldsymbol{c}_{\overline{\omega^*}} \right) \\
&= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}} \boldsymbol{c}_{\omega^*} + \boldsymbol{D}^T_{\overline{\omega^*},\overline{\omega^*}} \boldsymbol{c}_{\overline{\omega^*}} - (\boldsymbol{D} + \boldsymbol{D}^T)_{\overline{\omega^*},\omega^*} \boldsymbol{p}^*_{\omega^*} - (\boldsymbol{D} + \boldsymbol{D}^T)_{\overline{\omega^*},\overline{\omega^*}} \boldsymbol{c}_{\overline{\omega^*}} \right) \\
&= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}} \boldsymbol{c}_{\omega^*} - \boldsymbol{D}_{\overline{\omega^*},\omega^*} \boldsymbol{p}^*_{\omega^*} - \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}} \boldsymbol{p}^*_{\omega^*} - \boldsymbol{D}_{\overline{\omega^*},\overline{\omega^*}} \boldsymbol{c}_{\overline{\omega^*}} \right) \\
&= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}} (\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) - \boldsymbol{D}_{\overline{\omega^*},\omega^*} \boldsymbol{p}^*_{\omega^*} - \boldsymbol{D}_{\overline{\omega^*},\overline{\omega^*}} \boldsymbol{c}_{\overline{\omega^*}} \right) \\
&= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} + \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}} (\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) + \boldsymbol{D}_{\overline{\omega^*},\omega^*} (\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) - \boldsymbol{D}_{\overline{\omega^*},\omega^*} \boldsymbol{c}_{\omega^*} - \boldsymbol{D}_{\overline{\omega^*},\overline{\omega^*}} \boldsymbol{c}_{\overline{\omega^*}} \right) \\
&= \boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}} \left(\boldsymbol{a}_{\overline{\omega^*}} - \boldsymbol{D}_{\overline{\omega^*},\bullet} \boldsymbol{c} + \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}} (\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) + \boldsymbol{D}_{\overline{\omega^*},\omega^*} (\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) \right).
\end{aligned}
$$

It is clear that $\boldsymbol{S}_{\overline{\omega^*},\overline{\omega^*}}$ is also an M-matrix and hence $\boldsymbol{S}^{-1}_{\overline{\omega^*},\overline{\omega^*}}$ is nonnegative (element-wise). By Assumption (A2), $\boldsymbol{a} - \boldsymbol{D}\boldsymbol{c} \geq \boldsymbol{0}_n$, and hence

$$(\boldsymbol{a} - \boldsymbol{D}\boldsymbol{c})_{\overline{\omega^*}} = \boldsymbol{a}_{\overline{\omega^*}} - \boldsymbol{D}_{\overline{\omega^*},\bullet} \boldsymbol{c} \geq 0.$$

As we have shown above, $\boldsymbol{p}^*_{\omega^*} - \boldsymbol{c}_{\omega^*} \geq 0$. Also, by Assumption (A1) $\boldsymbol{D}_{\overline{\omega^*},\omega^*} \leq 0$ (element-wise) since it is an off-diagonal submatrix. Therefore,

$$\boldsymbol{D}_{\overline{\omega^*},\omega^*} (\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) \geq 0.$$

8

Similarly,

$$\boldsymbol{D}^T_{\omega^*,\overline{\omega^*}}(\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) \geq 0.$$

Overall, we have

$$\boldsymbol{a}_{\overline{\omega^*}} - \boldsymbol{D}_{\overline{\omega^*},\bullet}\boldsymbol{c} + \boldsymbol{D}^T_{\omega^*,\overline{\omega^*}}(\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) + \boldsymbol{D}_{\overline{\omega^*},\omega^*}(\boldsymbol{c}_{\omega^*} - \boldsymbol{p}^*_{\omega^*}) \geq 0.$$

And it follows that $\boldsymbol{p}^*_{\overline{\omega^*}} - \boldsymbol{c}_{\overline{\omega^*}} \geq 0$. The theorem thus holds readily. $\square$

Since $\boldsymbol{S}$ is symmetric positive definite, it has n positive eigenvalues. Let its eignvalues be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We collect some useful results regarding function $\mathcal{Q}(\boldsymbol{p})$ in the following lemma.

Lemma 4. *Suppose that assumption (A1) holds. Then the following statements hold.*

(a). For any two price vectors $\boldsymbol{p}$ and $\boldsymbol{p}'$, and any $L \geq \lambda_1$

$$\mathcal{Q}(\boldsymbol{p}') \leq \mathcal{Q}(\boldsymbol{p}) + \nabla\mathcal{Q}(\boldsymbol{p})^T(\boldsymbol{p}' - \boldsymbol{p}) + \frac{1}{2}L\|\boldsymbol{p} - \boldsymbol{p}'\|^2_2.$$

(b). For any price vector $\boldsymbol{p}$,

$$\mathcal{Q}(\boldsymbol{p}) \leq \widehat{\mathcal{Q}} + \frac{1}{2\lambda_n}\|\boldsymbol{S}(\boldsymbol{p} - \widehat{\boldsymbol{p}})\|^2_2.$$

(c). The function $\mathcal{Q}(\boldsymbol{p})$ is coercive, i.e.

$$\mathcal{Q}(\boldsymbol{p}) \to +\infty \ as \ \|\boldsymbol{p}\|_2 \to \infty.$$

Proof. Statement (a) can be shown using second order Taylor series expansion of a quadratic function at $\boldsymbol{p}$. In fact, we have:

$$\mathcal{Q}(\boldsymbol{p}') = \mathcal{Q}(\boldsymbol{p}) + \nabla\mathcal{Q}(\boldsymbol{p})^T(\boldsymbol{p}' - \boldsymbol{p}) + \frac{1}{2}(\boldsymbol{p}' - \boldsymbol{p})^T H_{\mathcal{Q}}(\boldsymbol{p})(\boldsymbol{p}' - \boldsymbol{p}),$$

where $H_{\mathcal{Q}}(\boldsymbol{p})$ is the Hessian matrix at $\boldsymbol{p}$. It is clear that $H_{\mathcal{Q}}(\boldsymbol{p}) = \boldsymbol{S}$, and since $\boldsymbol{S}$ is symmetric positive definite with largest eigenvalue λ_1, we have

$$\mathcal{Q}(\boldsymbol{p}') \leq \mathcal{Q}(\boldsymbol{p}) + \nabla\mathcal{Q}(\boldsymbol{p})^T(\boldsymbol{p}' - \boldsymbol{p}) + \frac{1}{2}\lambda_1\|\boldsymbol{p}' - \boldsymbol{p}\|^2_2 \leq \mathcal{Q}(\boldsymbol{p}) + \nabla\mathcal{Q}(\boldsymbol{p})^T(\boldsymbol{p}' - \boldsymbol{p}) + \frac{1}{2}L\|\boldsymbol{p}' - \boldsymbol{p}\|^2_2.$$

To show (b), we look at the Taylor series expansion at $\widehat{\boldsymbol{p}}$. Note that since $\widehat{\boldsymbol{p}}$ is the unconstrained minimizer of $\mathcal{Q}(\boldsymbol{p})$, it follows that $\nabla\mathcal{Q}(\widehat{\boldsymbol{p}}) = \boldsymbol{0}_n$. We have

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{p}) &= \mathcal{Q}(\widehat{\boldsymbol{p}}) + \nabla\mathcal{Q}(\widehat{\boldsymbol{p}})^T(\boldsymbol{p} - \widehat{\boldsymbol{p}}) + \frac{1}{2}(\boldsymbol{p} - \widehat{\boldsymbol{p}})^T\boldsymbol{S}(\boldsymbol{p} - \widehat{\boldsymbol{p}}) \\
&= \mathcal{Q}(\widehat{\boldsymbol{p}}) + \frac{1}{2}\left[(\boldsymbol{p} - \widehat{\boldsymbol{p}})^T\boldsymbol{S}\right]\boldsymbol{S}^{-1}\left[\boldsymbol{S}(\boldsymbol{p} - \widehat{\boldsymbol{p}})\right] \\
&\leq \widehat{\mathcal{Q}} + \frac{1}{2\lambda_n}\|\boldsymbol{S}(\boldsymbol{p} - \widehat{\boldsymbol{p}})\|^2_2,
\end{aligned}$$

where the inequality is due to the fact that the eigenvalues of $\boldsymbol{S}^{-1}$ are $\frac{1}{\lambda_n} \geq \frac{1}{\lambda_{n-1}} \geq \cdots \geq \frac{1}{\lambda_1}$. Statement (c) is a direct consequence of the positive definiteness of $\boldsymbol{S}$. $\square$

Note that the coercivity of $\mathcal{Q}(\boldsymbol{p})$ implies that it has bounded level sets, i.e., for any given $c \in \mathbb{R}$, the level set $L_c \triangleq \{\boldsymbol{p} \in \mathbb{R}^n \,|\, \mathcal{Q}(\boldsymbol{p}) \leq c\}$ is bounded (and closed).

9

3.2 Combinatorial Nature

It is clear that $\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}$ is not convex. On the other hand, we can show that it is in fact the union of finitely many polyhedra. To see this, let a triplet of index sets (α, β, γ) be a partition of $\{1, 2, \cdots, n\}$, satisfying:

$$|\beta| + |\gamma| \leq k, \ \alpha \cup \beta \cup \gamma = \{1, 2, \cdots, n\}, \ \alpha \cap \beta = \emptyset, \ \beta \cap \gamma = \emptyset, \ \text{and} \ \alpha \cap \gamma = \emptyset. \quad (7)$$

Let the family of all partitions satisfying (7) be denoted by $\Im$. Let the polyhedron $\mathcal{F}_{\alpha,\beta,\gamma}$ be defined as:

$$\mathcal{F}_{\alpha,\beta,\gamma} \triangleq \left\{ \boldsymbol{p} \in \mathbb{R}^n \ \middle| \ \begin{array}{ll} p_i = p_i^0, & \forall i \in \alpha, \\ p_i \geq p_i^0 + \delta_i, & \forall i \in \beta, \\ p_i \leq p_i^0 - \delta_i, & \forall i \in \gamma. \end{array} \right\} \quad (8)$$

It is easy to verify that

$$\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}} = \bigcup_{(\alpha,\beta,\gamma) \in \Im} \mathcal{F}_{\alpha,\beta,\gamma}.$$

Throughout the rest of the paper, we make an additional assumption:

(A4). δ_i's are all positive, i.e., $\delta_i > 0$ for all $i = 1, \cdots, n$.

Assumption (A4) is not restrictive in practice. In fact, in reality any price change is at least by 1 cent, therefore (A4) is valid. Under Assumption (A4), $\mathcal{F}_{\alpha,\beta,\gamma}$'s are non-overlapping. We can define an optimization problem over each polyhedron $\mathcal{F}_{\alpha,\beta,\gamma}$ as follows.

$$\begin{aligned} \overline{\boldsymbol{p}}_{\alpha,\beta,\gamma} \triangleq \arg\min_{\boldsymbol{p}} \quad & \mathcal{Q}(\boldsymbol{p}) \\ \text{s.t.} \quad & p_i = p_i^0, & \forall i \in \alpha, \\ & p_i \geq p_i^0 + \delta_i, & \forall i \in \beta, \\ & p_i \leq p_i^0 - \delta_i, & \forall i \in \gamma. \end{aligned} \quad (9)$$

Due to Assumption (A1), optimization problem (9) has a unique solution, denoted by $\overline{\boldsymbol{p}}_{\alpha,\beta,\gamma}$. With Assumption (A4), it is easy to verify that $\overline{\boldsymbol{p}}_{\alpha,\beta,\gamma}$ is a local optimal solution of (3). Therefore, a global optimal solution of (3) can be found by enumerating all possible partitions satisfying (7). However, the number of partitions in $\Im$ is exponentially large with respect to the number of product n. In fact, the number of possible partitions in $\Im$ is

$$\sum_{i=1}^{k} \binom{n}{i} 2^i.$$

It is hence not practical to enumerate all the possible partitions. This observation demonstrates the combinational nature of the price optimization problem (3). Therefore, in the following section we study an algorithm that finds a good near-optimal solution.

4 A Gradient Projection Algorithm for Price Optimization

In this section, we study a gradient projection method for the price optimization problem (3). The gradient projection method, due to its simplicity, is a widely used optimization

algorithm. It can be regarded as a natural extension of the classical gradient descent method for unconstrained optimization problems to the constrained ones. Typically, for provable convergence to an optimal solution, the optimization problem needs to have both a convex objective function and a convex feasible set. In problem (3), the objective function is convex but the feasible set is not. On the other hand, as we will show, the feasible set, albeit nonconvex, allows for an easily computable (in polynomial time) algorithm to compute a projection onto it. In this research, we will show both theoretically and empirically that a gradient projection method can perform very well for problem (3) even though a provable optimal solution is not guaranteed to be found. Our method is motivated by Bertsimas et al. (2016), where the authors studied regression problems with zero norm constraints. To present our algorithm, we start with some results regarding projections onto the feasible set $\Omega_{k, \boldsymbol{p}^0, \boldsymbol{\delta}}$.

4.1 Projection Onto the Feasible Set

We first consider the one-dimensional projection problem onto $\mathcal{P}_{\delta_i}^i$. Given any $q_i \in \mathbb{R}$, we aim to project it onto $\mathcal{P}_{\delta_i}^i$, or equivalently find a $\widetilde{p}_i$ to

$$\min_{p_i \in \mathcal{P}_{\delta_i}^i} (p_i - q_i)^2. \tag{10}$$

Since $\mathcal{P}_{\delta_i}^i$ is not convex, the optimization problem (10) does not necessarily have a unique solution, i.e., the projection is not necessarily a singleton. Let the set of solutions to (10) be denoted as $\Pi_{\mathcal{P}_{\delta_i}^i}(q_i)$. It is clear that

$$\arg\min_{p_i \in \mathcal{P}_{\delta_i}^i} (p_i - q_i)^2 \triangleq \Pi_{\mathcal{P}_{\delta_i}^i}(q_i) = \begin{cases} \{q_i\} & \text{if } q_i \in \mathcal{P}_{\delta_i}^i \\ \{p_i^0 + \delta_i\} & \text{if } p_i^0 + \frac{1}{2}\delta_i < q_i < p_i^0 + \delta_i \\ \{p_i^0, p_i^0 + \delta_i\} & \text{if } q_i = p_i^0 + \frac{1}{2}\delta_i \\ \{p_i^0\} & \text{if } p_i^0 - \frac{1}{2}\delta_i < q_i < p_i^0 + \frac{1}{2}\delta_i \\ \{p_i^0, p_i^0 - \delta_i\} & \text{if } q_i = p_i^0 - \frac{1}{2}\delta_i \\ \{p_i^0 - \delta_i\} & \text{if } p_i^0 - \delta_i < q_i < p_i^0 - \frac{1}{2}\delta_i \end{cases} \tag{11}$$

We can see that for any $\boldsymbol{q} \in \mathbb{R}^n$, the projection of it onto $\mathcal{P}_{\boldsymbol{\delta}} \subseteq \mathbb{R}^n$ is

$$\Pi_{\mathcal{P}_{\boldsymbol{\delta}}}(\boldsymbol{q}) = \bigotimes_{i=1}^n \Pi_{\mathcal{P}_{\delta_i}^i}(q_i).$$

Since $\Pi_{\mathcal{P}_{\delta_i}^i}(q_i)$ is not necessarily a singleton, we define a set-valued map

$$\phi_i(q_i) : q_i \mapsto \Pi_{\mathcal{P}_{\delta_i}^i}(q_i).$$

The graph of ϕ_i is defined as

$$\text{gph } \phi_i \triangleq \{(q_i, g_i) \in \mathbb{R}^2 | g_i \in \phi_i(q_i)\}.$$

(a) Projection on to $\mathcal{P}_{\delta_i}^i$.
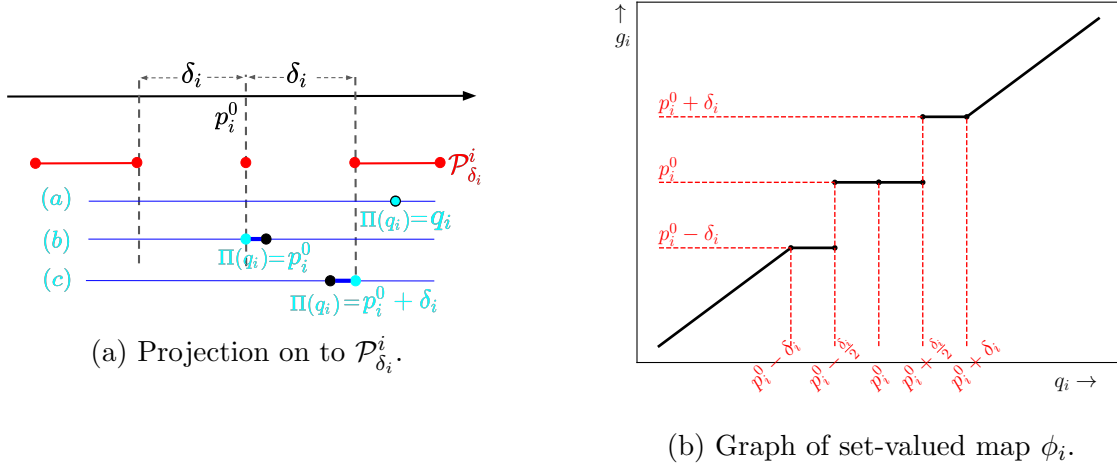

(b) Graph of set-valued map ϕ_i.

Figure 2: Projection onto the feasible set.

Recall that a set-valued map $\phi : \mathbb{R}^n \mapsto \mathbb{R}^n$ is closed on $\mathbb{R}^n$ if for any point $\boldsymbol{x} \in \mathbb{R}^n$, the following implication holds:

$$\left. \begin{array}{l} \lim_{j \to \infty} \boldsymbol{x}^j = \boldsymbol{x} \\ \lim_{j \to \infty} \boldsymbol{y}^j = \boldsymbol{y} \\ \boldsymbol{y}^j \in \phi(\boldsymbol{x}^j), \ \forall j \end{array} \right\} \Rightarrow \boldsymbol{y} \in \phi(\boldsymbol{x}). \tag{12}$$

It is well known that a set valued map ϕ is closed on $\mathbb{R}^n$ if and only if its graph gph ϕ is a closed set in $\mathbb{R}^n$. We have the following lemma regarding the closedness of ϕ_i.

Lemma 5. *For all* $i = 1, \cdots, n$, ϕ_i *is a closed map.*

Proof. The graph of ϕ_i on the q_i-g_i plane is the union of 2 rays and 3 line segments as follows:

$$\begin{aligned} \text{gph } \phi_i \ = \ & \{(q_i, g_i) | q_i = g_i, q_i \geq p_i^0 + \delta_i\} \cup \{(q_i, g_i) | q_i = g_i, q_i \leq p_i^0 - \delta_i\} \\ & \cup \{(q_i, g_i) | g_i = p_i^0 + \delta_i, p_i^0 + \frac{1}{2}\delta_i \leq q_i \leq p_i^0 + \delta_i\} \\ & \cup \{(q_i, g_i) | g_i = p_i^0, p_i^0 - \frac{1}{2}\delta_i \leq q_i \leq p_i^0 + \frac{1}{2}\delta_i\} \\ & \cup \{(q_i, g_i) | g_i = p_i^0 - \delta_i, p_i^0 - \delta_i \leq q_i \leq p_i^0 - \frac{1}{2}\delta_i\}. \end{aligned}$$

Since rays and line segments are closed, the union of these five sets is also closed. This concludes the proof. $\qquad \square$

Define $d(q_i, \mathcal{P}_{\delta_i}^i) \triangleq \min_{p_i \in \mathcal{P}_{\delta_i}^i} (p_i - q_i)^2$, i.e., the optimal value of (10) or the squared distance from q_i to set $\mathcal{P}_{\delta_i}^i$. It is straightforward to verify that

$$d(q_i, \mathcal{P}_{\delta_i}^i) \leq \frac{\delta_i^2}{4}. \tag{13}$$

12

In fact, we have

$$d(q_i, \mathcal{P}_{\delta_i}^i) = \begin{cases} 0 & \text{if } q_i \geq p_i^0 + \delta_i \\ (p_i^0 + \delta_i - q_i)^2 & \text{if } p_i^0 + \frac{1}{2}\delta_i \leq q_i \leq p_i^0 + \delta_i \\ (q_i - p_i^0)^2 & \text{if } p_i^0 \leq q_i \leq p_i^0 + \frac{1}{2}\delta_i \\ (p_i^0 - q_i)^2 & \text{if } p_i^0 - \frac{1}{2}\delta_i \leq q_i \leq p_i^0 \\ (q_i - p_i^0 + \delta_i)^2 & \text{if } p_i^0 - \delta_i \leq q_i \leq p_i^0 - \frac{1}{2}\delta_i \\ 0 & \text{if } q_i \leq p_i^0 - \delta_i \end{cases} \tag{14}$$

It is clear that $d(q_i, \mathcal{P}_{\delta_i}^i)$, when regarded as a function of q_i is continuous. For each $i \in \{1, \cdots, n\}$, define:

$$\Delta_i \triangleq \left(p_i^0 - q_i\right)^2 - d(q_i, \mathcal{P}_{\delta_i}^i). \tag{15}$$

By the definition of $d(q_i, \mathcal{P}_{\delta_i}^i)$, and the fact that $p_i^0 \in \mathcal{P}_{\delta_i}^i$, it is clear that $\Delta_i \geq 0$. We have the following lemma.

Lemma 6. *For each $i = 1, \cdots, n$, Let Δ_i be defined by (15), it holds that $\Delta_i = 0$ if and only if $p_i^0 - \frac{1}{2}\delta_i \leq q_i \leq p_i^0 + \frac{1}{2}\delta_i$.*

Proof. It is clear that $\Delta_i = 0$ if and only if $p_i^0 \in \Pi_{\mathcal{P}_{\delta_i}^i}(q_i)$. From Equation (11), it is clear that $p_i^0 \in \Pi_{\mathcal{P}_{\delta_i}^i}(q_i)$ if and only if $p_i^0 - \frac{1}{2}\delta_i \leq q_i \leq p_i^0 + \frac{1}{2}\delta_i$. $\square$

We next study the projection of a vector $\boldsymbol{q} \in \mathbb{R}^n$ onto the set $\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}$, i.e., the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & ||\boldsymbol{p} - \boldsymbol{q}||_2^2 \\ \text{s.t.} \quad & ||\boldsymbol{p} - \boldsymbol{p}^0||_0 \leq k \\ & \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{\delta}} \end{aligned} \tag{16}$$

Proposition 7. *Assume $\Delta_{(1)} \geq \Delta_{(2)} \geq \cdots \geq \Delta_{(n)}$, i.e., the ordered values of Δ_i's defined in (15), let $\widetilde{\boldsymbol{p}}$ satisfies*

$$\widetilde{p}_i \begin{cases} \in \Pi_{\mathcal{P}_{\delta_i}^i}(q_i), & \text{if } i \in \{(1), ..., (k)\}, \\ = p_i^0, & \text{otherwise.} \end{cases} \tag{17}$$

then $\widetilde{\boldsymbol{p}}$ is an optimal solution to (16).

Proof. It is straightforward to verify that $\widetilde{\boldsymbol{p}}$ given by (17) is a feasible solution to (16). Let $\mathcal{S}_k$ be the collection of all subsets of $\{1, \cdots, n\}$ with cardinality less than or equal to k. For each $\sigma \in \mathcal{S}_k$, we define the following optimization problem

$$\begin{aligned} V(\sigma) \triangleq \min_{\boldsymbol{p}} \quad & \sum_{i \notin \sigma}(q_i - p_i^0)^2 + \sum_{i \in \sigma}(q_i - p_i)^2 \\ \text{s.t.} \quad & p_i = p_i^0, \ \forall i \notin \sigma, \\ & \boldsymbol{p} \in \mathcal{P}_{\boldsymbol{\delta}}. \end{aligned} \tag{18}$$

Let $p(\sigma)$ be given by

$$p(\sigma)_i \begin{cases} \in \Pi_{\mathcal{P}_{\delta_i}}(q_i) & \text{if } i \in \sigma \\ = p_i^0 & \text{if } i \notin \sigma. \end{cases} \tag{19}$$

It is clear that $p(\sigma)$ is a feasible solution for (18). Moreover, given any vector $\boldsymbol{p}$ feasible for (18), we must have

$$\sum_{i \notin \sigma}(q_i - p_i^0)^2 + \sum_{i \in \sigma}(q_i - p_i)^2 \geq \sum_{i \notin \sigma}(q_i - p_i^0)^2 + \sum_{i \in \sigma} d(q_i, \mathcal{P}_{\delta_i}^i).$$

Therefore, $p(\sigma)$ is an optimal solution of (18), with the optimal value

$$
\begin{aligned}
V^*(\sigma) &= \sum_{i \notin \sigma}(q_i - p_i^0)^2 + \sum_{i \in \sigma} d(q_i, \mathcal{P}_{\delta_i}^i) \\
&= \sum_{i=1}^{n}(q_i - p_i^0)^2 + \sum_{i \in \sigma} \left[d(q_i, \mathcal{P}_{\delta_i}^i) - (q_i - p_i^0)^2 \right] \\
&= \|\boldsymbol{q} - \boldsymbol{p}^0\|_2^2 - \sum_{i \in \sigma} \Delta_i.
\end{aligned}
$$

Notice that (16) is equivalent to

$$\min_{\sigma \in \mathcal{S}_k} V^*(\sigma).$$

Since $\Delta_{(1)} \geq \Delta_{(2)} \geq \cdots \geq \Delta_{(n)}$, for any $\sigma \in \mathcal{S}_k$

$$V^*(\sigma) \geq \|\boldsymbol{q} - \boldsymbol{p}^0\|^2 - \sum_{i \in \{(1), \cdots, (k)\}} \Delta_i,$$

with the equality holding when σ is taken to be $\{(1), \cdots, (k)\}$. This implies that $\widetilde{\boldsymbol{p}}$ given in (17) is an optimal solution of (16). □

The above proposition provide an efficient algorithm to find a projection onto $\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}$ for any given $\boldsymbol{q} \in \mathbb{R}^n$. Due to the non-convexity of set $\Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}$, the solution set of (16) is not necessarily a singleton. For example, there can be ties among Δ_i's, leading to multiple index sets of k largest Δ_i's and hence multiple optimal solutions. In addition, the projection $\Pi_{\mathcal{P}_{\delta_i}^i}(q_i)$ itself is not necessarily a singleton. On the other hand, based on the proof of Proposition 7 it is straightforward to prove the next lemma, which provides a property of a solution to (16).

Lemma 8. *Let $\widetilde{\boldsymbol{p}} \in \mathbb{R}^n$ define $\sigma \triangleq \{i \mid \widetilde{p}_i \neq p_i^0\}$. The following statements hold.*

(i) *If $|\sigma| = k$, then $\widetilde{\boldsymbol{p}}$ is an optimal solution to (16) if and only if the following conditions hold:*
$$\begin{cases} \widetilde{p}_i \in \Pi_{\mathcal{P}_{\delta_i}^i}(q_i), & \forall\, i \in \sigma, \\ \Delta_i \geq \Delta_j, & \forall\, i \in \sigma, j \notin \sigma. \end{cases}$$

(ii) *If $|\sigma| < k$, then $\widetilde{\boldsymbol{p}}$ is an optimal solution to (16) if and only if the following conditions hold:*
$$\begin{cases} \widetilde{p}_i \in \Pi_{\mathcal{P}_{\delta_i}^i}(q_i), & \forall\, i \in \sigma, \\ \Delta_i = 0, & \forall\, i \notin \sigma. \end{cases}$$

In the rest of the paper, we denote the solution set to problem (16) by the notation $\mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\boldsymbol{q})$. We next study the set-valued map

$$\varphi(\boldsymbol{q}) : \boldsymbol{q} \mapsto \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\boldsymbol{q}).$$

We have the following result regarding $\varphi(\boldsymbol{q})$.

Lemma 9. $\varphi(\boldsymbol{q})$ *is a closed set-valued map on* $\mathbb{R}^n$.

Proof. We prove the lemma by verifying the definition. Take an arbitrary $\boldsymbol{q} \in \mathbb{R}^n$, let sequences $\{\boldsymbol{q}^j\}_{j=1,2,\cdots}$ and $\{\boldsymbol{g}^j\}_{j=1,2,\cdots}$ satisfy

$$\lim_{j \to \infty} \boldsymbol{q}^j = \boldsymbol{q}, \lim_{j \to \infty} \boldsymbol{g}^j = \boldsymbol{g}, \text{ and } \boldsymbol{g}^j \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\boldsymbol{q}^j) \ \forall j.$$

Define $\Delta_i^j \triangleq (p_i^0 - q_i^j)^2 - d(q_i^j, \mathcal{P}_{\delta_i}^i)$. For each $j = 1, 2, \cdots$, let $\sigma^j \triangleq \{i \,|\, g_i^j \neq p_i^0\}$, it is clear that $|\delta| \leq k$. By Lemma 8 it holds that

$$g_i^j \in \mathcal{P}_{\delta_i}^i(q_i^j), \forall i \in \sigma^j, \text{ and } \Delta_i^j \geq \Delta_{i'}^j, \forall i \in \sigma^j, i' \notin \sigma^j.$$

Since there are only finitely many possible subsets of $\{1, 2, \cdots, n\}$, at least one subset appears infinitely mainly times in the sequence $\{\sigma^j\}_{j=1,2,\cdots}$. Therefore, by working with a subsequence if necessary, we can assume without loss of generality that $\sigma^j = \sigma^*$ for all $j = 1, 2, \cdots$. It is clear that for any $i \in \sigma^*$ and $i' \notin \sigma^*$, by the continuity of $d(q_i, \mathcal{P}_{\delta_i}^i)$, we have

$$\begin{aligned}
\Delta_i &= (p_i^0 - q_i)^2 - d(q_i, \mathcal{P}_{\delta_i}^i) = \lim_{j \to \infty} \left[(p_i^0 - q_i^j)^2 - d(q_i^j, \mathcal{P}_{\delta_i}^i) \right] \\
&\geq \lim_{j \to \infty} \left[(p_{i'}^0 - q_{i'}^j)^2 - d(q_{i'}^j, \mathcal{P}_{\delta_{i'}}^{i'}) \right] = (p_{i'}^0 - q_{i'})^2 - d(q_{i'}, \mathcal{P}_{\delta_{i'}}^{i'}) = \Delta_{i'}
\end{aligned}$$

In addition, in case $|\sigma^*| < k$, we have:

$$\Delta_i = (p_i^0 - q_i)^2 - d(q_i, \mathcal{P}_{\delta_i}^i) = \lim_{j \to \infty} \left[(p_i^0 - q_i^j)^2 - d(q_i^j, \mathcal{P}_{\delta_i}^i) \right] = 0, \ \forall i \notin \sigma^*.$$

Moreover, by Lemma 5, i.e., the closedness of ϕ_i, we have

$$g_i \in \mathcal{P}_{\delta_i}^i(q_i), \forall i \in \sigma^*.$$

It is also clear that $g_{i'} = p_{i'}^0, \forall i' \notin \sigma^*$. Therefore, by Lemma 8, we have $\boldsymbol{g} \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\boldsymbol{q})$. This concludes the proof. $\square$

Remark 10. *Proposition 7 is a generalization of Proposition 3 of Bertsimas et al. (2016), that deals with the special case where* $\boldsymbol{\delta} = \boldsymbol{0}_n$ *and* $\boldsymbol{p}^0 = \boldsymbol{0}_n$.

Remark 11. *When upper and lower bounds on individual products exist, the projection of any* $q_i \in \mathbb{R}$ *on to set* $\mathcal{P}_{\delta_i, l_i, u_i}^i$ *is given by*

$$\Pi_{\mathcal{P}_{\delta_i, l_i, u_i}^i}(q_i) = \begin{cases} \{q_i\} & \text{if } q_i \in \mathcal{P}_{\delta_i, l_i, u_i} \\ \{l_i\} & \text{if } q_i < l_i \\ \{p_i^0 - \delta_i\} & \text{if } p_i^0 - \delta_i < q_i < p_i^0 - \frac{1}{2}\delta_i \\ \{p_i^0 - \delta_i, p_i^0\} & \text{if } q_i = p_i^0 - \frac{1}{2}\delta_i \\ \{p_i^0\} & \text{if } p_i^0 - \frac{1}{2}\delta_i < q_i < p_i^0 + \frac{1}{2}\delta_i \\ \{p_i^0, p_i^0 + \delta_i\} & \text{if } q_i = p_i^0 + \frac{1}{2}\delta_i \\ \{p_i^0 + \delta_i\} & \text{if } p_i^0 + \frac{1}{2}\delta_i \leq q_i \leq p_i^0 + \delta_i \\ \{u_i\} & \text{if } q_i > u_i \end{cases}$$

Proposition 7 as well as Lemma 9 can then be extended to include this case. The proofs are straightforward and hence omitted.

4.2   A Gradient Projection Algorithm

The gradient projection algorithm can be regarded as minimizing an upper approximation at current solution of $\mathcal{Q}(\boldsymbol{p})$ given by Lemma 4 item (a). In particular, at iteration t, let the current solution be $\boldsymbol{p}^t$, for an $L > \lambda_1$, we define

$$\widetilde{\mathcal{Q}}_L(\boldsymbol{p}; \boldsymbol{p}^t) \triangleq \mathcal{Q}(\boldsymbol{p}^t) + \nabla \mathcal{Q}(\boldsymbol{p}^t)^T(\boldsymbol{p} - \boldsymbol{p}^t) + \frac{1}{2}L\|\boldsymbol{p} - \boldsymbol{p}^t\|_2^2. \tag{20}$$

It is easy to see that

$$
\begin{aligned}
\arg\min_{\boldsymbol{p} \in \Omega_k, \boldsymbol{p}^0, \boldsymbol{\delta}} \widetilde{\mathcal{Q}}_L(\boldsymbol{p}; \boldsymbol{p}^t) &= \arg\min_{\boldsymbol{p} \in \Omega_k, \boldsymbol{p}^0, \boldsymbol{\delta}} \left(\frac{L}{2} \left\| \boldsymbol{p} - \left(\boldsymbol{\lambda}_1 - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^t) \right) \right\|_2^2 - \frac{1}{2L}\|\nabla\mathcal{Q}(\boldsymbol{p}^t)\|_2^2 + \mathcal{Q}(\boldsymbol{p}^t) \right) \\
&= \arg\min_{\boldsymbol{p} \in \Omega_k, \boldsymbol{p}^0, \boldsymbol{\delta}} \left\| \boldsymbol{p} - \left(\boldsymbol{p}^t - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^t) \right) \right\|_2^2 \\
&= \mathcal{H}_{k, \boldsymbol{\delta}, \boldsymbol{p}^0} \left(\boldsymbol{p}^t - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^t) \right),
\end{aligned} \tag{21}
$$

i.e., the projection of $\boldsymbol{p}^t - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^t)$ onto the set $\Omega_{k, \boldsymbol{p}^0, \boldsymbol{\delta}}$. The algorithm is given below.

Gradient Projection Algorithm (GPA)

Input: $\mathcal{Q}(\boldsymbol{p})$, parameter: $L > \lambda_1$ and convergence tolerance ε.
Output: A first-order stationary solution $\boldsymbol{p}^\infty$.

1. Initialize with $\boldsymbol{p}^1$ such that $\|\boldsymbol{p}^1 - \boldsymbol{p}^0\|_0 \le k$.

2. For $t \ge 1$, apply (17) to obtain $\boldsymbol{p}^{t+1}$, i.e.,

$$\boldsymbol{p}^{t+1} \in \mathcal{H}_{k, \boldsymbol{p}^0, \boldsymbol{\delta}} \left(\boldsymbol{p}^t - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^t) \right). \tag{22}$$

3. Repeat step 2, until $\mathcal{Q}(\boldsymbol{p}^t) - \mathcal{Q}(\boldsymbol{p}^{t+1}) \le \varepsilon$.

4.3   Analysis of the Gradient Projection Algorithm

In this subsection, we analyze the GPA. We aim to answer the following questions:

1. What kind of solutions is the GPA designed to find? How good are these solutions in terms of the objective values?

2. Does the GPA indeed converge to a solution it is designed to find?

It is clear that the GPA essentially aims to find a fixed point of the (set-valued) map

$$\boldsymbol{p} \mapsto \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}\left(\boldsymbol{p} - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p})\right).$$

We first study how such a fixed point is related to an optimal solution of problem (3). A technical lemma regarding a property of $\mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\cdot)$ is stated below.

Lemma 12. *Let* $\boldsymbol{p}' \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}\left(\boldsymbol{p} - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p})\right)$, *then*

$$\mathcal{Q}(\boldsymbol{p}) - \mathcal{Q}(\boldsymbol{p}') \geq \frac{L - \lambda_1}{2}\|\boldsymbol{p}' - \boldsymbol{p}\|_2^2. \tag{23}$$

Proof. As we can see from equation (21),

$$\boldsymbol{p}' \in \underset{\boldsymbol{\eta} \in \Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}}{\arg\min} \widetilde{\mathcal{Q}}_L(\boldsymbol{\eta};\boldsymbol{p}).$$

Therefore

$$
\begin{aligned}
\mathcal{Q}(\boldsymbol{p}) \;=\; \widetilde{\mathcal{Q}}_L(\boldsymbol{p};\boldsymbol{p}) \;\geq\;\; & \widetilde{\mathcal{Q}}_L(\boldsymbol{p}';\boldsymbol{p}) \\
=\;\; & \mathcal{Q}(\boldsymbol{p}) + (\nabla\mathcal{Q}(\boldsymbol{p}))^T(\boldsymbol{p}' - \boldsymbol{p}) + \frac{L}{2}\|\boldsymbol{p}' - \boldsymbol{p}\|_2^2 \\
=\;\; & \mathcal{Q}(\boldsymbol{p}) + (\nabla\mathcal{Q}(\boldsymbol{p}))^T(\boldsymbol{p}' - \boldsymbol{p}) + \frac{\lambda_1}{2}\|\boldsymbol{p}' - \boldsymbol{p}\|_2^2 + \frac{L - \lambda_1}{2}\|\boldsymbol{p}' - \boldsymbol{p}\|_2^2 \\
\geq\;\; & \mathcal{Q}(\boldsymbol{p}') + \frac{L - \lambda_1}{2}\|\boldsymbol{p}' - \boldsymbol{p}\|_2^2,
\end{aligned}
$$

where the second inequality is due to Lemma 4 item (a). Thus, we have

$$\mathcal{Q}(\boldsymbol{p}) - \mathcal{Q}(\boldsymbol{p}') \geq \frac{L - \lambda_1}{2}\|\boldsymbol{p}' - \boldsymbol{p}\|_2^2. \tag{24}$$

$\square$

We define a first-order stationary point as follows.

Definition 13. *[First-Order Stationary Point] Given an* $L > \lambda_1$, *a vector* $\boldsymbol{p}^\infty \in \mathbb{R}^n$ *is said to be a first-order stationary point of problem* (3) *if it satisfies the following fixed point equation:*

$$\boldsymbol{p}^\infty \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}\left(\boldsymbol{p}^\infty - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^\infty)\right). \tag{25}$$

Note that (25) implies that $\boldsymbol{p}^\infty \in \Omega_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}$ by the definition of $\mathcal{H}_{k,\boldsymbol{\delta},\boldsymbol{p}^0}(\cdot)$. Let

$$\alpha^\infty \triangleq \{i \mid p_i^\infty = p_i^0\}, \ \beta^\infty \triangleq \{i \mid p_i^\infty \geq p_i^0 + \delta_i\}, \text{ and } \gamma^\infty \triangleq \{i \mid p_i^\infty \leq p_i^0 - \delta_i\}. \tag{26}$$

Due to assumption (A4), the index sets α^∞, β^∞, and γ^∞ are uniquely defined given $\boldsymbol{p}^\infty$. With the index sets, define the following optimization problem

$$
\begin{aligned}
\min_{\boldsymbol{p}} \quad & \mathcal{Q}(\boldsymbol{p}) \\
\text{s.t.} \quad & p_i = p_i^0, && \forall\, i \in \alpha^\infty, \\
& p_i \geq p_i^0 + \delta_i, && \forall\, i \in \beta^\infty, \\
& p_i \leq p_i^0 - \delta_i, && \forall\, i \in \gamma^\infty.
\end{aligned}
\tag{27}
$$

Since the objective function in (27) is strongly convex (due to the positive definiteness of matrix S), and all the constraints are linear, there exists a unique optimal solution to (27). By introducing a multiplier μ_i for each constraint in (27), define the Lagrangian function:

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{\mu}) \triangleq \mathcal{Q}(\boldsymbol{p}) + \sum_{i \in \alpha^\infty} \mu_i(p_i - p_i^0) - \sum_{i \in \beta^\infty} \mu_i(p_i - p_i^0 - \delta_i) + \sum_{i \in \gamma^\infty} \mu_i(p_i - p_i^0 + \delta_i).$$

It is well known that a vector $\overline{\boldsymbol{p}}$ is the optimal solution of (27) if and only if there exists a vector $\boldsymbol{\mu}$ so that the pair $(\overline{\boldsymbol{p}}, \boldsymbol{\mu})$ satisfies:

$$
\begin{aligned}
\nabla_i \mathcal{Q}(\overline{\boldsymbol{p}}) + \mu_i &= 0, & \forall\, i \in \alpha^\infty \cup \gamma^\infty \\
\nabla_i \mathcal{Q}(\overline{\boldsymbol{p}}) - \mu_i &= 0, & \forall\, i \in \beta^\infty \\
\overline{p}_i - p_i^0 &= 0, & \forall\, i \in \alpha^\infty \\
0 \le \overline{p}_i - p_i^0 - \delta_i &\perp \mu_i \ge 0, & \forall\, i \in \beta^\infty \\
0 \le -\overline{p}_i + p_i^0 - \delta_i &\perp \mu_i \ge 0, & \forall\, i \in \gamma^\infty
\end{aligned}
\tag{28}
$$

where "$\perp$" means two vectors are perpendicular, i.e., $\boldsymbol{x} \perp \boldsymbol{y} \Leftrightarrow \boldsymbol{x}^T \boldsymbol{y} = 0$. The following lemma holds.

Lemma 14. *Let $\boldsymbol{p}^\infty \in \mathbb{R}^n$ be a first-order stationary point defined by (25). Then it is the optimal solution of (27) with the index sets α^∞, β^∞, and γ^∞ as defined in (26).*

Proof. It suffices to construct a vector $\boldsymbol{\mu}$ so that $(\boldsymbol{p}^\infty, \boldsymbol{\mu})$ satisfies (28). For each $i \in \alpha^\infty \cup \gamma^\infty$, we let $\mu_i = -\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty)$. For each $i \in \beta^\infty$, we let $\mu_i = \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty)$. It suffices to verify that the last two complementarity conditions in (28). In fact, for any $i \in \beta^\infty$, there are two cases as follows:

Case 1. $p_i^\infty - p_i^0 - \delta_i > 0$. In this case, it is clear that $p_i^\infty \ne p_i^0$. By Lemma 8, we have

$$p_i^\infty \in \Pi_{\mathcal{P}_{\delta_i}^i}\left(p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty)\right), \ \forall\, i \notin \alpha^\infty,$$

which implies that $p_i^\infty = p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \Rightarrow \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) = 0$. Hence, $\mu_i = 0$, and $0 \le p_i^\infty - p_i^0 - \delta_i \perp \mu_i \ge 0$ holds.

Case 2. $p_i^\infty - p_i^0 - \delta_i = 0$. In this case, we have $p_i^\infty \ne p_i^0$ due to assumption (A4). By Lemma 8, we still have

$$p_i^\infty \in \Pi_{\mathcal{P}_{\delta_i}^i}\left(p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty)\right), \ \forall\, i \notin \alpha^\infty.$$

By equation (11) we deduce that

$$p_i^0 + \frac{1}{2}\delta_i \le p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \le p_i^0 + \delta_i \Rightarrow p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \le p_i^\infty \Rightarrow \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \ge 0.$$

Thus $\mu_i = \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \ge 0$, and $0 \le p_i^\infty - p_i^0 - \delta_i \perp \mu_i \ge 0$ holds.

We can similarly show that the second complementarity condition in (28) holds. This concludes the proof. $\square$

Lemma 14 indicates that a first-order stationary point satisfying (25) must be the optimal solution of (9) with the partition being $(\alpha^\infty, \beta^\infty, \gamma^\infty)$ and hence a local optimal solution of (3). On the other hand, not every local optimal solution of (3) is a first-order stationary point. As we have discussed in Section 3, each partition $(\alpha, \beta, \gamma) \in \Im$ corresponds to a local optimal solution. We have the following example of a local optimal solution that is not a first-order stationary point.

Example 15. *Consider the case where* $n = 2$ *and* $k = 1$, $\boldsymbol{p}^0 = [0, 0]^T$, $\boldsymbol{\delta} = [0.5, 0.5]^T$, $\boldsymbol{S} = \begin{bmatrix} 2 & -a \\ -a & 2 \end{bmatrix}$, $\boldsymbol{r} = [-6, -1]^T$, *and the problem is*

$$\min_{\boldsymbol{p}} \quad \frac{1}{2}\boldsymbol{p}^T \boldsymbol{S} \boldsymbol{p} + \boldsymbol{r}^T \boldsymbol{p}. \tag{29}$$

We can show that the local optimal solution with $\alpha = \{1\}$, $\beta = \{2\}$ *and* $\gamma = \emptyset$ *is not a stationary point. In fact, the local optimal solution with partition* $(\{1\}, \{2\}, \emptyset)$ *is* $[0, 0.5]^T$. *Then, we can show that* $[0, 0.5]^T$ *is not a stationary point when* $0 \leq a < 1$. *In fact, when* $\boldsymbol{p} = [0, 0.5]^T$, *and* $L = 3$ *we can calculate* $\boldsymbol{q} = \boldsymbol{p} - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p})$ *as follows:*

$$\begin{aligned} q_1 &= p_1 - \tfrac{1}{3}(s_{11}p_1 + s_{12}p_2 - 6) &= \tfrac{a}{3}p_2 + 2 > 2 \\ q_2 &= p_2 - \tfrac{1}{3}(s_{21}p_1 + s_{22}p_2 - 1) &= \tfrac{1}{3}p_2 + \tfrac{1}{3} = 0.5 \end{aligned} \tag{30}$$

To calculate $\mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\boldsymbol{q})$, *we have:*

$$\begin{aligned} \Delta_1 &= (p_1^0 - q_1)^2 - d(q_1, \mathcal{P}_{\delta_1}^1) &= q_1^2 \\ \Delta_2 &= (p_2^0 - q_2)^2 - d(q_2, \mathcal{P}_{\delta_2}^2) &= 0.25 \end{aligned} \tag{31}$$

Clearly, $\Delta_1 > \Delta_2$ *and hence* $[0, 0.5]^T \notin \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}(\boldsymbol{q}) = \{[q_1, 0]^T\}$. *Therefore, the local optimal solution with partition* $(\{1\}, \{2\}, \emptyset)$ *is not a stationary point.*

While not every local optimal solution of (3) is a first-order stationary point, the following proposition guarantees that a global optimal solution must be a first-order stationary point.

Proposition 16. *Let* $\boldsymbol{p}^*$ *be a global optimal solution of* (3), *then* $\boldsymbol{p}^*$ *is a first order stationary point.*

Proof. By (23) in Lemma 12, any $\boldsymbol{\eta} \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}\left(\boldsymbol{p}^* - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^*)\right)$ satisfies

$$\mathcal{Q}(\boldsymbol{p}^*) - \mathcal{Q}(\boldsymbol{\eta}) \geq \frac{L - \lambda_1}{2}\|\boldsymbol{p}^* - \eta\|_2^2.$$

Since $\boldsymbol{p}^*$ is a global optimal solution of (3), we must have $\mathcal{Q}(\boldsymbol{p}^*) - \mathcal{Q}(\boldsymbol{\eta}) \leq 0$, therefore $\|\boldsymbol{p}^* - \boldsymbol{\eta}\|_2^2 = 0$. Hence $\boldsymbol{p}^* = \boldsymbol{\eta} \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}\left(\boldsymbol{p}^* - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^*)\right)$. This concludes the proof. $\square$

In summary, we have shown the following relationship regarding (3):

{global optimal solutions} $\subseteq$ {first-order stationary points} $\subseteq$ {local optimal solutions}.

We next study the quality of a first-order stationary point in terms of the objective value. Let $\Delta_i^\infty \triangleq (p_i^0 - q_i^\infty)^2 - d(q_i^\infty, \mathcal{P}_{\delta_i}^i)$, with $\boldsymbol{q}^\infty \triangleq \boldsymbol{p}^\infty - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^\infty)$. Assume $\Delta_{(1)}^\infty \geq \Delta_{(2)}^\infty \geq \cdots \geq \Delta_{(n)}^\infty$. The next result provides a performance bound of a first-order stationary point.

Proposition 17. *Let $\boldsymbol{p}^\infty$ satisfy the first-order stationary condition* (25), *let* $\sigma^\infty \triangleq \{i \mid p_i^\infty \neq p_i^0\}$ *and* $\kappa = |\sigma^\infty|$, *then the following two statements hold.*

(i).

$$\|\nabla \mathcal{Q}(\boldsymbol{p}^\infty)\|_2^2 \leq L^2 \sum_{i=\kappa+1}^n \Delta_{(i)}^\infty + \frac{1}{4}L^2 \sum_{i=1}^n \delta_i^2.$$

(ii).

$$\mathcal{Q}(\boldsymbol{p}^\infty) - \mathcal{Q}^* \leq \frac{L^2}{2\lambda_n} \sum_{i=\kappa+1}^n \Delta_{(i)}^\infty + \frac{L^2}{8\lambda_n} \sum_{i=1}^n \delta_i^2.$$

Proof. By definition, for each $i \notin \sigma^\infty$, we have $p_i^\infty = p_i^0$. Hence, for each $i \notin \sigma^\infty$

$$\Delta_i^\infty = \left(p_i^0 - p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \right)^2 - d(q_i, \mathcal{P}_{\delta_i}^i) = \frac{1}{L^2} \left(\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \right)^2 - d(q_i, \mathcal{P}_{\delta_i}^i)$$

$$\Rightarrow \left(\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \right)^2 = L^2 \left(\Delta_i^\infty + d(q_i, \mathcal{P}_{\delta_i}^i) \right) \leq L^2 \left(\Delta_i^\infty + \frac{\delta_i^2}{4} \right),$$

where the inequality is due to inequality (13). For any $i \in \sigma^\infty$, by Lemma 8 it is clear that

$$p_i^\infty \in \Pi_{\mathcal{P}_{\delta_i}^i} \left(p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \right).$$

Notice that for $p_i^\infty \in \mathcal{P}_{\delta_i}^i$, there are five cases to consider:

Case 1. $p_i^\infty > p_i^0 + \delta_i$. In this case, we have

$$p_i^\infty = p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \Rightarrow \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) = 0.$$

Case 2. $p_i^\infty = p_i^0 + \delta_i$. In this case, we have

$$p_i^0 + \frac{1}{2}\delta_i \leq p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \leq p_i^0 + \delta_i \Rightarrow p_i^0 + \frac{1}{2}\delta_i \leq p_i^0 + \delta_i - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \leq p_i^0 + \delta_i.$$

Therefore, we have:

$$0 \leq \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \leq \frac{1}{2}L\delta_i.$$

and hence $(\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty))^2 \leq \frac{1}{4}L^2 \delta_i^2$.

Case 3. $p_i^\infty = p_i^0$. In this case, $p_i^0 \in \Pi_{\mathcal{P}_{\delta_i}^i}(q_i)$, which, by equation (11), implies that

$$p_i^0 - \frac{1}{2}\delta_i \leq p_i^\infty - \frac{1}{L}\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \leq p_i^0 + \frac{1}{2}\delta_i \Rightarrow (\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty))^2 \leq \frac{1}{4}L^2 \delta_i^2.$$

Case 4. $p_i^\infty < p_i^0 - \delta_i$. Similar to Case 1, we have $\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) = 0$.

20

Case 5. $p_i^\infty = p_i^0 - \delta_i$. Similar to Case 2, we have

$$-\frac{1}{2}L\delta_i \leq \nabla_i \mathcal{Q}(\boldsymbol{p}^\infty) \leq 0 \Rightarrow (\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty))^2 \leq \frac{1}{4}L^2\delta_i^2.$$

Overall, for all $i \in \sigma^\infty$,

$$(\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty))^2 \leq \frac{1}{4}L^2\delta_i^2.$$

Hence, we have

$$\|\nabla \mathcal{Q}(\boldsymbol{p}^\infty)\|_2^2 = \sum_{i=1}^n (\nabla_i \mathcal{Q}(\boldsymbol{p}^\infty))^2 \leq \sum_{i \in \{(1),\cdots,(\kappa)\}} \frac{1}{4}L^2\delta_i^2 + \sum_{i \in \{(\kappa+1),\cdots,(n)\}} L^2\left(\Delta_i^\infty + \frac{1}{4}\delta_i^2\right)$$

$$= L^2 \sum_{i=\kappa+1}^n \Delta_{(i)}^\infty + \frac{1}{4}L^2 \sum_{i=1}^n \delta_i^2,$$

and (i) holds readily. To show (ii), we notice that $\mathcal{Q}(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T S \boldsymbol{p} - \boldsymbol{f}^T \boldsymbol{p}$, with $\boldsymbol{f} \triangleq \boldsymbol{a} + \boldsymbol{D}^T \boldsymbol{c}$. It is easy to see that $\nabla \mathcal{Q}(\boldsymbol{p}) = S\boldsymbol{p} - \boldsymbol{f}$. Recall that $\widehat{\boldsymbol{p}}$ is the unconstrained minimizer of $\mathcal{Q}(\boldsymbol{p})$, and hence $\nabla \mathcal{Q}(\boldsymbol{p}) = \boldsymbol{0}_n$. Therefore, we have

$$\|S(\boldsymbol{p}^\infty - \widehat{\boldsymbol{p}})\|_2 = \|\nabla \mathcal{Q}(\boldsymbol{p}^\infty) - \nabla \mathcal{Q}(\widehat{\boldsymbol{p}})\|_2 = \|\nabla \mathcal{Q}(\boldsymbol{p}^\infty)\|_2.$$

By Lemma 4 item (b), we know that

$$\mathcal{Q}(\boldsymbol{p}^\infty) - \widehat{\mathcal{Q}} \leq \frac{1}{2\lambda_n}\|S(\boldsymbol{p}^\infty - \widehat{\boldsymbol{p}})\|_2^2 = \frac{1}{2\lambda_n}\|\nabla \mathcal{Q}(\boldsymbol{p}^\infty)\|_2^2 \leq \frac{L^2}{2\lambda_n}\sum_{i=\kappa+1}^n \Delta_{(i)}^\infty + \frac{L^2}{8\lambda_n}\sum_{i=1}^n \delta_i^2.$$

$\square$

The next result is a corollary to Proposition 17 when $\boldsymbol{p}^\infty$ satisfies the zero-norm constraint strictly.

Corollary 18. *Let $\boldsymbol{p}^\infty$ satisfy the first-order stationary condition (25) and σ^∞ and κ be defined in Proposition 17. If $|\sigma^\infty| < k$, then the following two statements hold.*

(i).

$$\|\nabla \mathcal{Q}(\boldsymbol{p}^\infty)\|_2^2 \leq \frac{1}{4}L^2 \sum_{i=1}^n \delta_i^2.$$

(ii).

$$\mathcal{Q}(\boldsymbol{p}^\infty) - \mathcal{Q}^* \leq \frac{L^2}{8\lambda_n}\sum_{i=1}^n \delta_i^2.$$

Proof. For each $i \in \alpha^\infty$, by Lemma 8 and the assumption that $|\sigma^\infty| < k$, we deduce that $\Delta_{(i)}^\infty = 0$ for all $i = \kappa + 1, \cdots, n$. The corollary therefore follows readily. $\square$

Remark 19. *In case when $\delta_i = 0$ for all $i = 1, \cdots, n$, Corollary 18 indicates that if $\|\boldsymbol{p}^\infty - \boldsymbol{p}^0\|_0 < k$ then $\boldsymbol{p}^\infty$ is a global optimal solution of (3). This is an extension of Proposition 5 of Bertsimas et al. (2016).*

We next study the convergence of the GPA. We first state a technical lemma regarding the decrease in the objective value from iteration to iteration.

Lemma 20. *Let $\{\boldsymbol{p}^t\}_{t=1,2,\cdots}$ be a sequence generated by the GPA, for any $L > \lambda_1$. Then the sequence $\{\mathcal{Q}(\boldsymbol{p}^t)\}_{t=1,2,\cdots}$ satisfies*

$$\mathcal{Q}(\boldsymbol{p}^t) - \mathcal{Q}(\boldsymbol{p}^{t+1})) \geq \frac{L - \lambda_1}{2}\|\boldsymbol{p}^{t+1} - \boldsymbol{p}^t\|_2^2, \ \forall\, t = 1, \cdots. \tag{32}$$

Moveover,

$$\lim_{t\to\infty} \|\boldsymbol{p}^{t+1} - \boldsymbol{p}^t\|_2 \to 0.$$

Proof. By equation (23) in Lemma 12 it is clear that

$$\mathcal{Q}(\boldsymbol{p}^t) - \mathcal{Q}(\boldsymbol{p}^{t+1}) \geq \frac{L - \lambda_1}{2}\|\boldsymbol{p}^{t+1} - \boldsymbol{p}^t\|_2^2, \ \forall\, t = 1, \cdots. \tag{33}$$

Therefore the sequence $\{\mathcal{Q}(\boldsymbol{p}^t)\}_{t=1,2,\cdots}$ is monotonically decreasing and is bounded from below by $\widehat{\mathcal{Q}}$, and hence is convergent, i.e.,

$$\lim_{t\to\infty} \mathcal{Q}(\boldsymbol{p}^t) = \mathcal{Q}^\infty.$$

Hence, $\lim_{t\to\infty} \|\boldsymbol{p}^{t+1} - \boldsymbol{p}^t\| = 0$. $\qquad\square$

Next we prove a result regarding the stability of certain key index sets when the number of iterations is large enough. Let $\{\boldsymbol{p}^t\}_{t=1,2,\cdots}$ be the sequence generated by the GPA. For each t, we define

$$\alpha^t \triangleq \{i \,|\, p_i^t = p_i^0\}, \ \beta^t \triangleq \{i \,|\, p_i^t \geq p_i^0 + \delta_i\}, \ \text{and} \ \gamma^t \triangleq \{i \,|\, p_i^t \leq p_i^0 - \delta_i\}. \tag{34}$$

As we can see, since assumption (A4) holds, for each $t = 1, 2, \cdots$, the index sets α^t, β^t, and γ^t are mutually exclusive and satisfy $\alpha^t \cup \gamma^t \cup \gamma^t = \{1, 2, \cdots, n\}$.

Lemma 21. *Let $\{\boldsymbol{p}^t\}_{t=1,2,\cdots}$ be a sequence generated by the GPA. Let the index sets α^t, β^t, and γ^t be defined in (34). There exists an integer T, and mutually exclusive sets α^T, β^T, and γ^T such that*

$$\alpha^t = \alpha^T, \ \beta^t = \beta^T, \ \text{and} \ \gamma^t = \gamma^T$$

for all $t \geq T$.

Proof. Let $\delta_{\min} = \min\{\delta_1, \cdots, \delta_n\}$, the smallest among δ_i's. By Assumption (A4), $\delta_{\min} > 0$. By Lemma 20, there exists an integer T such that

$$\|\boldsymbol{p}^t - \boldsymbol{p}^{t+1}\|_2^2 \leq \frac{\delta_{\min}^2}{2}.$$

For any $t \geq T$, if $\alpha^t \neq \alpha^{t+1}$, then there are two possibilities. (i) There exists an index i such that $i \in \alpha^t$ and $i \notin \alpha^{t+1}$. (ii) There exists an index i such that $i \in \alpha^{t+1}$ and $i \notin \alpha^t$. In either case, $|p_i^t - p_i^{t+1}| \geq \delta_i \geq \delta_{\min}$, and hence $\|\boldsymbol{p}^t - \boldsymbol{p}^{t+1}\|_2^2 \geq \delta_{\min}^2$, which is a contradiction. Therefore, for all $t \geq T$, $\alpha^t = \alpha^{t+1} = \cdots$. Similarly, we have $\beta^t = \beta^{t+1} = \cdots$ and $\gamma^t = \gamma^{t+1} = \cdots$ for all $t \geq T$. $\qquad\square$

Theorem 22. *Let $\{\boldsymbol{p}^t\}_{t=1,2,\cdots}$ be a sequence generated by the GPA. It follows that the sequence $\{\boldsymbol{p}^t\}_{t=1,2,\cdots}$ converges to a first-order stationary point, i.e.,*

$$\lim_{t\to\infty} \boldsymbol{p}^t = \boldsymbol{p}^\infty,$$

and $\boldsymbol{p}^\infty$ is a first-order stationary point satisfying (25).

Proof. By Lemma 21, there exists a $T \geq 1$ such that

$$\alpha^t = \alpha^T, \ \beta^t = \beta^T, \ \text{and} \ \gamma^t = \gamma^T$$

for all $t \geq T$. Therefore, the GPA is equivalent to applying a gradient projection algorithm on the following convex quadratic program

$$
\begin{aligned}
\overline{\boldsymbol{p}}_{\alpha,\beta,\gamma} \triangleq \arg\min_{\boldsymbol{p}} \quad & \mathcal{Q}(\boldsymbol{p}) \\
\text{s.t.} \quad & p_i = p_i^0, && \forall\, i \in \alpha^T, \\
& p_i \geq p_i^0 + \delta_i, && \forall\, i \in \beta^T, \\
& p_i \leq p_i^0 - \delta_i, && \forall\, i \in \gamma^T.
\end{aligned}
\tag{35}
$$

By the coercivity of $\mathcal{Q}(\boldsymbol{p})$ (part (c) of Lemma 4), and Lemma 20, it is clear that $\{\boldsymbol{p}^i\}_{t=1,2,\cdots}$ is bounded. Hence, the convergence follows (Nesterov 2013). Next, we show that the GPA converges to a first-order stationary point satisfying (25). Because of the continuity of the $\nabla\mathcal{Q}(\boldsymbol{p})$ of a quadratic function,

$$\lim_{t\to\infty}\left[\boldsymbol{p}^t - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^t)\right] = \boldsymbol{p}^\infty - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^\infty).$$

Then, by taking the limit on both sides of (22), because of the closedness of $\varphi(\boldsymbol{q})$ by Lemma 9, we know that $\boldsymbol{p}^\infty \in \mathcal{H}_{k,\boldsymbol{p}^0,\boldsymbol{\delta}}\left(\boldsymbol{p}^\infty - \frac{1}{L}\nabla\mathcal{Q}(\boldsymbol{p}^\infty)\right)$. This completes the proof. $\qquad\square$

Remark 23. *From Lemma 21 and Theorem 22 we can see that once the GPA progresses to the point where $\|\boldsymbol{p}^{t+1} - \boldsymbol{p}^t\|_2^2 \leq \frac{\delta_{\min}^2}{2}$, it is equivalent to applying a gradient projection algorithm to the convex quadratic program (35). Since the gradient projection algorithm is a first-order algorithm which converges at most linearly, we can speed up the convergence from that point by switching to a more efficient convex quadratic programming algorithm, such as an interior point algorithm.*

5   Numerical Experiments

We conduct numerical experiments on both randomly generated data as well as real-world data. On randomly generated data, we conduct two different groups of numerical experiments. In the first group, we consider the base formulation (3). We include upper/lower bounds on the individual prices in the second group. For both groups and the real-world data, we demonstrate the performance of the GPA by comparing its results to those from Gurobi.

5.1   Randomly Generated Data Sets

Given that the number of products in a typical grocery store in North America is about 50,000 and about 20,000 in Europe, we generate five sets of data corresponding to 10,000, 25,000, 50,000, 75,000 and 100,000 products as follows:

1. Generate matrix $\boldsymbol{D}$. For each row, we first generate a uniformly distributed random number between $[1, 10]$ as the diagonal element, and then generate up to five negative off-diagonal elements of magnitudes less than 0.2 times that of the diagonal element in that row.

2. Generate vectors $\boldsymbol{a}$, $\boldsymbol{c}$, $\boldsymbol{\delta}$ and $\boldsymbol{p}^0$. For each product i, p_i^0 is set to be a random number between $[1, 10]$. The vector $-\boldsymbol{a} - \boldsymbol{D}^T \boldsymbol{c}$ is generated as a random vector with all elements in $[1, 10]$. $\boldsymbol{\delta}$ is set as $\mathbf{1}_n$ for \$1.00 and $\boldsymbol{\delta} = 0.5 \cdot \mathbf{1}_n$ for \$0.50.

3. Set k. We allow up to 10% of the products to change prices.

Counting the number of data sets and the two possibilities of $\boldsymbol{\delta}$ for each case, we have a total of 10 instances of the price optimization problem. For each instance, we use five distinct initial solutions to start the GPA: (1) $\boldsymbol{p}^0$, (2),(3),(4) are three random vectors, and (5) the solution obtained by using $\boldsymbol{p}^0$ with a step length larger than $\frac{1}{L}$ for some L greater than the largest eigenvalue of $\boldsymbol{D}$. The motivation of using initial solution (5) is that by allowing the algorithm to take a longer step, we hope that it can move away from the local optimal solution. In practice, we observe that the best solution often comes from (5). Together with the five initial solutions, we have a total of 50 optimization problems that are solved sequentially.

5.2   Benchmark Formulation

We use the results from Gurobi as a benchmark to demonstrate the performance of the GPA. To solve the price optimization (3) using Gurobi, for each product i we introduce a triplet of binary variables, z_i^P, z_i^R, z_i^L indicating if the price of product i stays at p_i^0, is at least $p_i^0 + \delta_i$, or is up to $p_i^0 - \delta_i$, respectively. We use the following mixed integer programming

formulation:

$$
\begin{aligned}
\max_{\boldsymbol{p}, \boldsymbol{z}^P, \boldsymbol{z}^R, \boldsymbol{z}^L} \quad & -\frac{1}{2}\boldsymbol{p}^T S \boldsymbol{p} + (\boldsymbol{a} + \boldsymbol{D}^T \boldsymbol{c})^T \boldsymbol{p} \\
\text{s.t.} \quad & p_i \geq p_i^0 z_i^P - M z_i^L + (p_i^0 + \delta_i) z_i^R, \quad \forall\, i = 1, \cdots, n, \quad -- \text{(lower bound)} \\
& p_i \leq p_i^0 z_i^P + (p_i^0 - \delta_i) z_i^L + M z_i^R, \quad \forall\, i = 1, \cdots, n, \quad -- \text{(upper bound)} \\
& z_i^P + z_i^L + z_i^R = 1, \qquad\qquad\qquad \forall\, i = 1, \cdots, n, \\
& \sum_{i=1}^{n} (z_i^L + z_i^R) \leq k, \\
& z_i^P, z_i^L, z_i^R \in \{0, 1\}, \qquad\qquad\quad \forall\, i = 1, \cdots, n,
\end{aligned}
\tag{36}
$$

where M is a large positive integer chosen to be relatively small to help with the performance of Gurobi. When we have lower/upper bounds $l_i \leq p_i^0 - \delta_i$ and $u_i \geq p_i^0 + \delta_i$ on the individual prices p_i, we use the following formulation, which is similar to the above formulation with $-M$ and M being replaced by the lower bounds and upper bounds, respectively.

$$
\begin{aligned}
\max_{\boldsymbol{p}, \boldsymbol{z}^P, \boldsymbol{z}^R, \boldsymbol{z}^L} \quad & -\frac{1}{2}\boldsymbol{p}^T S \boldsymbol{p} + (\boldsymbol{a} + \boldsymbol{D}^T \boldsymbol{c})^T \boldsymbol{p} \\
\text{s.t.} \quad & p_i \geq p_i^0 z_i^P + l_i z_i^L + (p_i^0 + \delta_i) z_i^R, \quad \forall\, i = 1, \cdots, n, \quad -- \text{(lower bound)} \\
& p_i \leq p_i^0 z_i^P + (p_i^0 - \delta_i) z_i^L + u_i z_i^R, \quad \forall\, i = 1, \cdots, n, \quad -- \text{(upper bound)} \\
& z_i^P + z_i^L + z_i^R = 1, \qquad\qquad\qquad \forall\, i = 1, \cdots, n, \\
& \sum_{i=1}^{n} (z_i^L + z_i^R) \leq k, \\
& z_i^P, z_i^L, z_i^R \in \{0, 1\}, \qquad\qquad\quad \forall\, i = 1, \cdots, n.
\end{aligned}
\tag{37}
$$

To compare the solution quality, we used the adjusted objective gap as the performance measure:

$$
\text{Adjusted Objective Gap} \triangleq \frac{100(Z(\boldsymbol{p}^B) - Z(\boldsymbol{p}^G))}{|Z(\boldsymbol{p}^0)|},
$$

where $\boldsymbol{p}^0$ is the baseline price, $\boldsymbol{p}^G$ is the Gurobi solution and $\boldsymbol{p}^B$ is the GPA solution. When the adjusted objective gap is positive, the GPA achieved a better solution than Gurobi, whereas when it is negative, Gurobi obtained a better solution than the GPA. We use Gurobi 7.0.2 configured as follows: the maximum run time is set to 1 hour for experiments labeled Gurobi-1 and 4 hours for Gurobi-4, the acceptable optimality gap is set to 0, and all other parameters are at their default values. The number of computing threads is set to 1 to allow for a fair comparison with the GPA, which uses only 1 thread. Both Gurobi and GPA implementations were executed on Amazon Elastic Computing Cloud running Amazon Linux AMI 2018.03 with dual Intel® Xeon® 16-core E5-2686 v4 CPU @ 2.3GHz and 244GB memory.

5.3 Results from the Randomly Generated Data Sets

For a reasonable total run time of 205 minutes solving the 50 optimization problems for the unbounded case, Table 1 reports on the combined results for cases when the minimum change

Table 1: Adjusted objective gaps on random data sets.

Bounds	n	Gurobi-1			Gurobi-4		
		Mean	StDev	Range	Mean	StDev	Range
no bounds	10,000	-3.91	3.97	[-6.24, 4.62]	-5.86	0.34	[-6.25, -5.39]
	25,000	29.37	0.56	[28.53, 30.10]	29.37	0.56	[28.53, 30.10]
	50,000	29.58	0.70	[28.26, 30.49]	29.58	0.70	[28.26, 30.49]
	75,000	29.19	0.52	[28.46, 30.08]	29.19	0.52	[28.46, 30.08]
	100,000	29.25	0.44	[28.68, 30.14]	29.25	0.44	[28.68, 30.14]
$l_i \in [1, 5]$ $u_i \in [5, 10]$	10,000	-2.19	0.18	[-2.48, -1.99]	-2.19	0.18	[-2.48, -2.00]
	25,000	-2.26	0.11	[-2.43, -2.10]	-2.26	0.11	[-2.43, -2.11]
	50,000	22.89	13.26	[-2.33, 29.51]	-2.25	0.06	[-2.35, -2.17]
	75,000	28.86	0.24	[28.39, 29.16]	4.08	13.12	[-2.36, 29.16]
	100,000	28.86	0.25	[28.48, 29.29]	28.86	0.25	[28.48, 29.29]
$l_i \in [1, 5]$ $u_i \in [10, 15]$	10,000	-4.00	0.11	[-4.15, -3.89]	-4.02	0.11	[-4.16, -3.91]
	25,000	-3.76	0.18	[-3.96, -3.35]	-4.13	0.04	[-4.18, -4.05]
	50,000	26.62	0.43	[26.01, 27.32]	26.62	0.43	[26.01, 27.32]
	75,000	26.48	0.26	[26.00, 26.85]	26.48	0.26	[26.00, 26.85]
	100,000	26.51	0.28	[25.91, 26.94]	26.51	0.28	[25.91, 26.94]
$l_i \in [1, 5]$ $u_i \in [15, 20]$	10,000	-4.33	0.19	[-4.56, -3.94]	-4.39	0.14	[-4.58, -4.20]
	25,000	20.26	12.09	[-2.85, 26.35]	1.86	12.64	[-4.43, 25.93]
	50,000	26.13	0.35	[25.43, 26.57]	26.13	0.35	[25.43, 26.57]
	75,000	26.08	0.29	[25.67, 26.56]	26.08	0.29	[25.67, 26.56]
	100,000	26.05	0.22	[25.69, 26.38]	26.05	0.22	[25.69, 26.38]

is $0.50 and $1.00. The results are combined as they are comparable. The GPA significantly outperforms both Gurobi-1 and Gurobi-4 when $n > 10,000$, leading to incremental profits of around 29% from the baseline. In fact, Gurobi barely improved the objective value from 1 hour to 4 hours, achieving about 35% MIP gap for $n = 10,000$ but no meaningful MIP gaps for larger cases at the end of 4 hours. For $n = 10,000$, the average adjusted objective gap compared with Gurobi-1 is -3.91% with a maximum of -6.24%, and the average adjusted objecive gap compared with Gurobi-4 is -5.86% with a maximum of -6.25%. These results indicate that the GPA is better suited for real-world applications, especially when the number of products is large.

Next, we report on experiments with upper/lower bounds on the prices of individual products. For each product, we generate a random lower bound between $l_i \in [1, 5]$ and three increasing random upper bounds within $u_i \in [5, 10]$, $u_i \in [10, 15]$ and $u_i \in [15, 20]$. p_i^0 is set to be a random number between $[l_i, u_i]$. The purpose of considering three different upper bounds is to investigate the impact of the width of $[l_i, u_i]$ on the performance of the GPA.

As expected, the bounds significantly reduced the feasible regions of the problem instances, and so both the GPA and Gurobi solved the problems faster. The GPA took less time to converge. For example, with $u_i \in [5, 10]$, the GPA took up to 19 minutes on the 50 problems, whereas Gurobi-1 and Gurobi-4 ended up with smaller MIP gaps (6%) for the smaller cases ($n <= 25,000$ for Gurobi-1 and $n <= 50,000$ for Gurobi-4). As the upper bounds increase, the running time of the GPA increases, i.e., up to 23 minutes total run time for $u_i \in [10, 15]$ and up to 30 minutes for $u_i \in [15, 20]$. The maximum run times are all for $n = 100,000$.

Similar to the unbounded case, the GPA performed better than Gurobi on larger instances on all three different upper bounds. For example when $u_i \in [5, 10]$, Table 1 shows that the GPA outperforms Gurobi-1 when $n \geq 50,000$, and outperforms Gurobi-4 when $n \geq 75,000$. The critical number of products at which GPA outperforms Gurobi decreases as the upper bounds increase, i.e., the GPA performs better than Gurobi with a larger feasible region.

It is also worth noting that for the instances where Gurobi outperforms the GPA, its advantage as measured by the adjusted objective gap generally decreases with the feasible region. This means that for smaller problem instances with reasonably tight bounds, the GPA can reliably provide good quality solutions in significantly shorter time.

5.4   Numerical Experiments on Real-World Cases

We conduct experiments on real-world data sets from a grocery chain with 300 stores in four regions in Europe. Based on a rule where the prices of each product is set uniformly across stores in each region but may differ across regions, we have 59,000 prices to optimize. We also consider both substitute and complement products, i.e., positive and negative off-diagonal elements in matrix $\boldsymbol{D}$. This relaxes assumption (A1), but the matrix $\boldsymbol{S}$ remains positive definite.

First, we consider three cases of fixed minimum price change thresholds across all products: 1 cent, 10 cents and 50 cents. We also consider a cases of varying minimum price changes set to 10% of the baseline price of each products, i.e., $\delta_i = 10\% \times p^0 i$. With the four cases of minimum price change, maximum number of prices that can be changed is limited to 3,000, 5,000 and 8,000. In these experiments, the GPA solutions for each case are obtained

27

Table 2: Results on real-world cases.

| Min. Amount of Price Change | Max. No. of Price Changes | GPA | | Gurobi-4 |
		Improve from Base (million €)	Run Time (minutes)	Improve from Base (million €)
1 cent	3,000	615 (15.4%)	10.3	0
	5,000	696 (17.5%)	7.1	0
	8,000	777 (19.5%)	10.0	546 (13.7%)
10 cents	3,000	611 (15.3%)	10.1	0
	5,000	693 (17.4%)	7.3	0
	8,000	773 (19.4%)	10.3	567 (14.2%)
50 cents	3,000	569 (14.3%)	10.0	0
	5,000	635 (15.9%)	9.0	0
	8,000	708 (17.7%)	9.3	554 (13.9%)
10% of p_i^0	3,000	611 (15.3%)	10.0	0
	5,000	691 (17.3%)	6.9	0
	8,000	772 (19.3%)	9.7	560 (14.0%)

in about 10 minutes on sequential runs with 5 initial solutions. The results are summarized in Table 2. In comparison, Gurobi, initialized with the baseline prices, did not provide better solutions than p^0 for cases with 3,000 and 5,000 maximum number of price changes and did not provide better solutions than the GPA in any of the cases within 4 hours.

We also observed that increasing the maximum number of price changes provide diminishing returns on maximum profit (Figure 3). Using as reference the profit improvement obtained by the GPA with minimum price change threshold of 1 cent and $k = 59,000$ (i.e., 100% of potential profit improvement), we then observe that $k = 3,000$ (5.1%) can achieve more than 65% of potential profit improvement, $k = 8,000$ (13.5%) can achieve more than 80% potential profit improvement, and $k = 24,000$ (40.7%) can achieve more than 95% potential profit improvement. Notice that since the GPA only finds local optima, it may not be the global maximum improvement. Moreover, a minimum change of 1 cent, 10 cents or 10% of the baseline price p^0 lead to similar results, while a 50 cent minimum change threshold would reduce the profit by 5-11%.

6   Conclusion

In this paper, we study a brick-and-mortar retailer price optimization problem with maximum number of price changes and minimum amount of price change constrains. These two constraints are crucial to provide actionable price recommendations to retailers. Although the feasible region of this optimization problem is not convex, we showed that the Euclidean projection onto it can be obtained very efficiently. Therefore, a gradient projection algorithm is proposed to solve this problem. We demonstrated desired theoretical properties of the algorithm and verified its efficiency using computational experiments.

There are several future research directions stemming from this work. One immediate
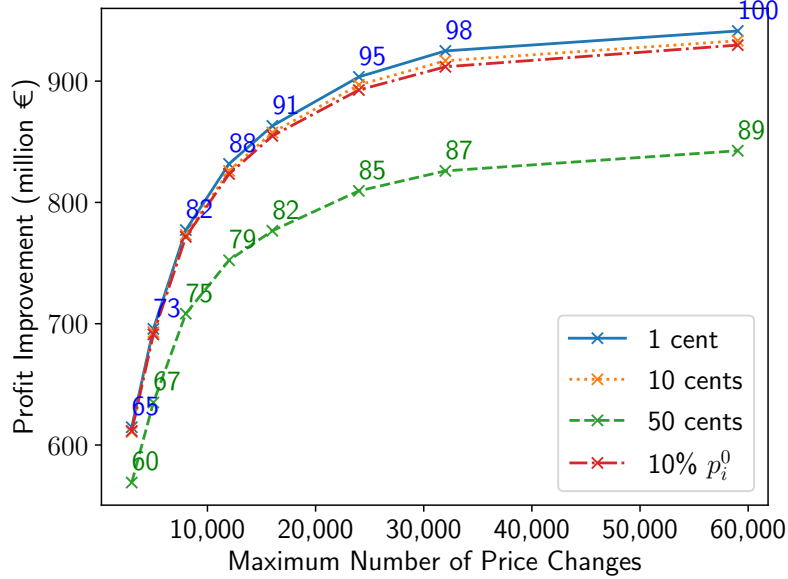
Figure 3: Real-world application of the GPA: Diminishing returns over maximum number of price changes.

extension is to include other more standard business rules, such as those that can be written as linear inequalities. With those additional constraints, it is not easy to compute the projection of an arbitrary point to the feasible region, and therefore the GPA in this paper is not directly applicable. However, one can consider decomposition approaches, where the GPA can be applied to solve sub problems. Theoretical properties of such decomposition approaches as well as their empirical performance are worth further investigation. It is also worth investigating if existing algorithms for solving mathematical programs with cardinality constraints, such as those reported in Burdakov et al. (2016), can be extended to solve the price optimization problem with the two practical constraints. Another potential extension is to consider other demand functions, such as those resulting from discrete choice models. These demand functions are not necessarily convex, which adds another layer of challenge to this problem.

As we have mentioned before, beyond price optimization problems, the two practical constraints studied are considerably common in various applications. Therefore, we believe that optimization problems with these two practical constraints deserve special attention. Mathematical properties of the feasible set defined by these two constraints as well as sophisticated algorithms taking advantage of those properties can be the focus of future research.

References

Aydın Alptekinoğlu and John H Semple. The exponomial choice model: A new alternative for assortment and price optimization. *Operations Research*, 64(1):79–93, 2016.

Abraham Berman and Robert J Plemmons. *Nonnegative matrices in the mathematical sciences.* SIAM, 1994.

Dimitris Bertsimas and Romy Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 43(1):1–22, 2009.

Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, pages 813–852, 2016.

A. Bhutani and P. Bhardwaj. Electronic Shelf Label (ESL) Market Size By Product (LCD ESL, Segmented E-Paper ESL, Full Graphic E-Paper ESL), By Technology (RF, IR, NFC), By Application (Industrial, Commercial [Hypermarket, Supermarket, Non-Food Retail]), Industry Analysis Report, Regional Outlook (U.S., Canada, UK, Germany, France, Italy, Russia, Spain, China, Japan, South Korea, India, Singapore, Malaysia, Australia, Brazil, Mexico, Saudi Arabia, UAE, South Africa), Growth Potential, Competitive Market Share & Forecast, 2018 – 2024, 2018. URL https://www.gminsights.com/industry-analysis/electronic-shelf-label-esl-market.

Gabriel Bitran and René Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

E Andrew Boyd and Ioana C Bilegan. Revenue management and e-commerce. *Management Science*, 49(10):1363–1386, 2003.

Josef Broder and Paat Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

Juan José Miranda Bront, Isabel Méndez-Díaz, and Gustavo Vulcano. A column generation algorithm for choice-based network revenue management. *Operations Research*, 57(3):769–784, 2009.

Oleg P Burdakov, Christian Kanzow, and Alexandra Schwartz. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26(1):397–425, 2016.

Wen-Chyuan Chiang, Jason CH Chen, and Xiaojing Xu. An overview of research on revenue management: current issues and future research. *International Journal of Revenue Management*, 1(1):97–128, 2007.

Gregory Dobson and Shlomo Kalish. Positioning and pricing a product line. *Marketing Science*, 7 (2):107–125, 1988.

Lingxiu Dong, Panos Kouvelis, and Zhongjun Tian. Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management*, 11(2):317–339, 2009.

Wedad Elmaghraby and Pınar Keskinocak. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, 49 (10):1287–1309, 2003.

Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322, 2013.

Kris Johnson Ferreira, Bin Hong Alex Lee, and David Simchi-Levi. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88, 2016.

Kris Johnson Ferreira, David Simchi-Levi, and He Wang. Online network revenue management using Thompson sampling. *Operations Research*, 66(6):1586–1602, 2018.

G Gallego, G Iyengar, R Phillips, and A Dubey. Managing flexible products on a network. depart-

ment of industrial engineering and operations research, columbia university. Technical report, CORC Technical Report TR-2004-01, 2004.

Guillermo Gallego, Huseyin Topaloglu, et al. *Revenue management and pricing analytics*, volume 209. Springer, 2019.

Jochen Goensch, Robert Klein, and Claudius Steinhardt. Dynamic pricing–state-of-the-art. *Zeitschrift für Betriebswirtschaft, Ergänzungsheft*, 3:1–40, 2009.

Joseph M Hall, Praveen K Kopalle, and Aradhna Krishna. A category management model of retailer dynamic pricing and ordering decisions: Normative and empirical analysis. Working paper, Tuck School of Business at Dartmouth, Hanover, New Hampshire, 2003.

Hsin-Chan Huang, Jiefeng Xu, and Alvin Lim. Marketing mix optimization with practical constraints. *arXiv preprint arXiv:2101.03663*, 2021.

Jian Huang, Mingming Leng, and Mahmut Parlar. Demand functions in decision modeling: A comprehensive survey and research directions. *Decision Sciences*, 44(3):557–609, 2013.

Woonghee Tim Huh and Hongmin Li. Pricing under the nested attraction model with a multistage choice structure. *Operations Research*, 63(4):840–850, 2015.

Kirthi Kalyanam. Pricing decisions under demand uncertainty: A bayesian mixture model approach. *Marketing Science*, 15(3):207–221, 1996.

Philipp W Keller, Retsef Levi, and Georgia Perakis. Efficient formulations for pricing under attraction demand models. *Mathematical Programming*, 145(1-2):223–261, 2014.

Timo P Kunz and Sven F Crone. Demand models for the static retail price optimization problem-a revenue management perspective. In *4th Student Conference on Operational Research*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

Michael Levy, Dhruv Grewal, Praveen K Kopalle, and James D Hess. Emerging trends in retail pricing practice: implications for research, 2004.

Qian Liu and Garrett Van Ryzin. On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310, 2008.

Jeffrey I McGill and Garrett J Van Ryzin. Revenue management: Research overview and prospects. *Transportation Science*, 33(2):233–256, 1999.

Velibor V Mišić and Georgia Perakis. Data analytics in operations management: A review. *Manufacturing & Service Operations Management*, 22(1):158–169, 2020.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

Shmuel S Oren, Stephen A Smith, and Robert B Wilson. Multi-product pricing for electric power. *Energy Economics*, 9(2):104–114, 1987.

Özalp Özer and Robert Phillips. *The Oxford handbook of pricing management*. Oxford University Press, 2012.

Paat Rusmevichientong, Benjamin Van Roy, and Peter W Glynn. A nonparametric approach to multiproduct pricing. *Operations Research*, 54(1):82–98, 2006.

PB Seetharaman, Siddhartha Chib, Andrew Ainslie, Peter Boatwright, Tat Chan, Sachin Gupta, Nitin Mehta, Vithala Rao, and Andrei Strijnev. Models of multi-category choice behavior. *Marketing Letters*, 16(3-4):239–254, 2005.

Jing-Sheng Song and Zhengliang Xue. Demand management and inventory control for substitutable products. Working paper, Duke University, Durham, North Carolina, 2007.

Arne K Strauss, Robert Klein, and Claudius Steinhardt. A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research*, 271(2):375–387, 2018.

Garrett J. van Ryzin. Models of demand. *Journal of Revenue & Pricing Management*, 4(2):204–210, 2005.

Lawrence R Weatherford and Samuel E Bodily. A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking, and pricing. *Operations Research*, 40(5):831–844, 1992.

Laurence A Wolsey. Mixed integer programming. *Wiley Encyclopedia of Computer Science and Engineering*, pages 1–10, 2007.

Heng Zhang, Paat Rusmevichientong, and Huseyin Topaloglu. Multiproduct pricing under the generalized extreme value models with homogeneous price sensitivity parameters. *Operations Research*, 66(6):1559–1570, 2018.