January 2016

Cardinality Constrained Optimization Problems

Jinhak Kim
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Jinhak Kim

Entitled
Cardinality Constrained Optimization Problems

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Mohit Tawarmalani
Chair

Jean-Philippe P. Richard

Yanjun Li

Thanh Nguyen

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Mohit Tawarmalani

Approved by: Mark Bagnoli 7/18/2016

Head of the Departmental Graduate Program Date

CARDINALITY CONSTRAINED OPTIMIZATION PROBLEMS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Jinhak Kim

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2016

Purdue University

West Lafayette, Indiana

To my wife Jayoung and sons Timothy and Jeremy.

ACKNOWLEDGMENTS

I have been very fortunate to have two outstanding scholars, Prof. Mohit Tawarmalani and Prof. Jean-Philippe P. Richard as my advisors. Foremost, I express my best gratitude to Prof. Tawarmalani for his unstinted support throughout my Ph.D. study. His astonishing erudition and extraordinary passion have guided me to a stunning field, mathematical programming. No one can substitute for his heartfelt effort and ardent guidence for my Ph.D. study. I also would like to show my appreciation to my co-advisor, Prof. Richard for the thoughtful and fruitful advise on my research. I was particularly grateful to his sedulousness and earnestness on commenting my writings and presentations.

In addition to my advisors, I thank two other thesis committee members, Prof. Yanjun Li and Prof. Thanh Nguyen for their comments and questions on the thesis and the defense of the dissertation.

I also would like to acknowledge my family members for their patience, support, and sacrifice. I especially thank my lovely wife Jayoung Sohn for giving me two beloved sons, Timothy and Jeremy during our Ph.D. studies. I am blessed that I have the best women as the sole partner of my life. I next want to appreciate my mother, Yekeum Yoo and my parents in law, Inshik Sohn and Myungna Jeong for their diligence and effort on raising my children. This thesis will never exist without their dedication.

Finally, I gratefully acknowledge the partial support from National Science Foundation grants CMMI 0900065, 1234897, and 1235236 and a Purdue Research Foundation research assistantship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Kim, Jinhak Ph.D., Purdue University, August 2016. Cardinality constrained optimization problems. Major Professor: Mohit Tawarmalani.

In this thesis, we examine optimization problems with a constraint that allows for only a certain number of variables to be nonzero. This constraint, which is called a *cardinality constraint*, has received considerable attention in a number of areas such as machine learning, statistics, computational finance, and operations management. Despite their practical needs, most optimization problems with a cardinality constraints are hard to solve due to their nonconvexity. We focus on constructing tight convex relaxations to such problems.

We first study linear programs with a cardinality constraint (CCLPs). A procedure that yields cutting planes for any given vector that violates the cardinality constraint is developed. These cutting planes are derived from a disjunctive relaxation of the problem. The separation problem is recast as a network optimization problem where the network is constructed from a simplex tableau of the LP relaxation. We then present a procedure to generate a facet-defining inequality of the disjunctive relaxation using a variant of Prim's algorithm.

Second, we study an optimization formulation of sparse principal component analysis (sparse PCA). The formulation is a quadratically constrained quadratic problem with a cardinality constraint. The feasible set has a special structure which we call *permutation-invariance*. This structure allows us to construct the convex hull of the feasible set of the model. The convex hull is written through a majorization inequality that can be modeled using a polynomial number variables and linear inequalities. We then show that sparse PCA can be reformulated as a continuous convex maximization problem without a cardinality constraint. In addition, we derive SDP relaxations for

the reformulation. The relaxations are developed based on majorization arguments. The resulting relaxation is provably tighter than the prevalent SDP relaxation proposed in [24]. Our preliminary computational results show that our SDP relaxation has gaps 90% smaller than those of the classical SDP relaxation.

Third, we introduce other approaches for CCLPs. We first present a facial disjunctive reformulation for CCLPs and a finitely-convergent cutting plane algorithm. A generalized reformulation-linearization technique (RLT) is introduced to characterize the convex hull of the feasible set of CCLPs. As a special subclass of CCLP, we study the cardinality-constrainted knapsack problem (CCKP). We developed families of valid inequalities based on disjunctions for the cardinality constraint.

1. Introduction

Considerable attention has been paid to optimization problems with a constraint that allows only up to a certain number of variables to be nonzero. We call such a constraint a *cardinality constraint* and any optimization problem containing such a constraint a *cardinality constrained optimization problem* (CCOP).

In this thesis, we present relaxation strategies for certain classes of CCOPs using various techniques developed in the fields of mixed-integer linear programming, global optimization, convex and nonconvex optimization.

CCOPs arise in fields as diverse as computational finance, supply chain management, statistical data analysis, and machine learning. They are used in cardinality-constrained *optimal portfolio selection* problems in quantitative finance [14, 18, 22, 27, 43, 49, 52, 54]. These problems are variants of the Markowitz mean-variance model where the objective is to minimize a quadratic risk measure under linear constraints along with a restriction that the number of securities chosen for investment is sufficiently small. They also arise in *index tracking investment strategies* [10, 28, 41, 42, 60, 62]. These problems are modeled as time series optimization models where the objective is to minimize a quadratic tracking error under budget constraints and a restriction that the number of securities selected for investment is small. Facility location problems are classical supply chain management models where a company must decide where to locate facilities. The variant of the problem where at most p warehouses can be opened is known as the *p-median* problem, and has been extensively studied in the literature [1, 9, 21, 25, 38, 47, 55]. In statistical data analysis, principal component analysis (PCA) is a well-known technique for dimension reduction. It finds principal components as linear combinations of the original variables. When the coefficients of many variables in these linear combinations are nonzero, the principal components can be hard to interpret. In order to find principal

components that are easier to explain, a cardinality constraint (referred to as a *sparsity constraint*) is sometimes imposed on the original problem. The resulting problem is known as *sparse principal component analysis* (sparse PCA); see [24,35,46,75]. Ensemble pruning [73] and variable selection in multiple regression [12,13] are also often modeled as CCOPs.

Although CCOPs find uses in a variety of applications, they are hard to solve to global optimality. Perhaps the simplest of these problems, which involves optimizing a linear function over the intersection of a continuous knapsack polytope and a cardinality constraint, is already NP-hard [26]. Further, large instances of practical problems are computationally challenging to solve [14, 26, 54].

For a decision variable $x \in \mathbb{R}^n$, $\mathrm{card}(x)$ represents the number of nonzero components or the *cardinality* of x. A cardinality constraint is written as $\mathrm{card}(x) \leq K$ for some positive integer $K \in \{1, \ldots, n-1\}$. In this thesis, we assume that $K > 1$ because the problem is trivial when $K = 1$. Therefore, we also assume that $n \geq 3$. Various strategies have been proposed to model cardinality constraints, and to leverage classical MIP branch-and-cut methodologies in the solution of cardinality-constrained problems. When variables x are bounded, auxiliary binary variables can be introduced to model the cardinality constraint. That is, for bounds $l, u \in \mathbb{R}^n$ such that $l_i u_i \leq 0$ for $i = 1, \ldots, n$, constraints

$$\begin{cases} l \leq x \leq u, \\ \mathrm{card}(x) \leq K \end{cases}$$

can be replaced with

$$\begin{cases} l \circ z \leq x \leq u \circ z, \\ \mathbb{1}^\intercal z \leq K, \\ z \in \{0, 1\}^n \end{cases}$$

where $\circ$ is the Hardamard product and $\mathbb{1} \in \mathbb{R}^n$ is the vector whose components are all equal to one.

When the constraints of the initial problem are linear, such an approach allows the use of branch-and-cut algorithms developed for mixed integer programs (MIPs).

This reformulation also allows for the use of cutting planes derived for cardinality-constrained problems; see [71, 72].

In [8] a specialized branch-and-bound algorithm was proposed to solve problems with cardinality constraints where $K \in \{1, 2\}$. These techniques were adapted to logically constrained linear programs [48], mixed integer quadratic programs [14], and to cardinality-constrained knapsack problems in [26]. Moreover, [26] develops valid inequalities for cardinality constrained knapsack problems (CCKPs) that can be used for cardinality constraint linear programs (CCLPs).

This thesis is organized as follows. In Chapter 2, we develop a procedure that generates cutting planes for CCLPs. For a given LP relaxation of a CCLP and a basic feasible solution, we construct a disjunctive relaxation of the corresponding simplex tableau from which we derive the desired cuts. Specifically, if the given basic feasible solution violates the cardinality constraint, there exists at least $K + 1$ basic variables that correspond to nonzero components of the solution. Then, a disjunctive relaxation of the cardinality requirement can be obtained by imposing a disjunction that forces at least one of those basic variables to be nonpositive. We characterize the closed convex hull of this disjunctive relaxation by obtaining the extreme ray representation of each disjunct and by proving that facet-defining inequalities can be obtained by solving a dual network optimization problem. We further observe that the nontrivial facet-defining inequalities of the relaxation directly relate to a particular class of subgraphs, which we call *label-connected spanning trees*, of a bipartite network that can be constructed for the the simplex tableau. This procedure generalizes the E&R procedure [56] recently developed in the context of complementarity problems. We then describe a polynomial-time Prim-like algorithm that tightens any given valid inequality of the disjunctive relaxation into a facet-defining inequality. As a special case, we constructively show how the c-max cut, which is a well-known disjunctive cut, can be strengthened to a facet-defining inequality of the convex hull of our disjunctive relaxation when it is not.

Chapter 3 focuses on reformulation and relaxation techniques for an optimization formulation of sparse principal component analysis (sparse PCA). Sparse PCA seeks to find a sparse eigenvector of a given covariance matrix for a centered data matrix. We first provide an extended formulation for the convex hull of sparse PCA by studying the dual of the separation problem. The derivation of the convex hull is possible because of a special symmetry structure of the feasible set. More specifically, if a point is in the feasible set, so are all its permutations. We call this property *permutation-invariance*. Permutation-invariance enables us to represent the convex hull through a majorization inequality which can be modeled using a polynomial number of additional variables and linear constraints. The underlying idea is to construct the convex hull over the simplex $\{x : x_1 \geq \cdots \geq x_n \geq 0\}$ and replicate it onto the remaining region. By replacing the feasible set of sparse PCA with its convex hull, we relax sparse PCA to a convex maximization problem. We then show that the relaxation is a reformulation of sparse PCA by showing that any optimal solution to the reformulated problem that violates the cardinality constraint can always be transformed to a point that satisfies the cardinality constraint and achieves the same objective function value. In addition, we present semidefinite programming (SDP) relaxations for sparse PCA based on majorization arguments on matrix variables. We prove that our SDP relaxations are strictly tighter than a well-known SDP relaxation proposed in [24]. Preliminary computational results obtained for the `pitprops` dataset and for randomly generated covariance matrices show that our SDP relaxations have gaps 90% smaller than those of the classical relaxation.

In Chapter 4, we study other approaches to CCLPs. First, we formulate CCLPs as facial disjunctive programs by representing the cardinality constraint in conjunctive normal form. That is, $\text{card}(x) \leq K$ is equivalent to enforcing that every subset of $\{x_1, \ldots, x_n\}$ of size $K + 1$ includes at least one zero, or equivalently, $\bigwedge_{J \in \{1,\ldots,n\}, |J|=K+1} \bigvee_{j \in J} (x_j = 0)$. The facial structure of the disjunctive set enables us to apply the finitely-convergent cutting plane algorithm developed by Jeroslow [40]. We then present a generalized reformulation-linearization technique (RLT) to build

the convex hull of the feasible set of the CCLP. Based on the work of [67], we propose a product factor for generalized RLT that is a ratio of multilinear terms. In the remainder of the chapter, we focus on developing valid inequalities for CCKPs. The derivation is based on the following disjunction that is equivalent to the cardinality constraint $\text{card}(x) \leq K$: for a given $m \in \{0, \ldots, K\}$,

$$\text{card}(x_I) \leq m \quad \vee \quad \text{card}(x_{N \setminus I}) \leq K - m - 1$$

for any $I \subset N$ where x_I denotes the $|I|$-dimensional subvector of x corresponding to the index set I. This enables us to construct a new valid inequality from a given valid inequality. We show that the procedure generates a facet-defining inequality from a given facet-defining inequality under certain conditions. We also demonstrate that many valid inequalities proposed in [26] can be derived by our procedure. Finally, we show how to derive some valid inequalities using lifting arguments.

In the last chapter, we summarize the contributions of this dissertation and present directions for future research.

2. On cutting planes for cardinality constrained linear programming

In this chapter, we derive cutting planes for cardinality-constrained linear programs (CCLPs). These inequalities can be used to separate any basic feasible solution of an LP relaxation of the problem, assuming that this solution violates the cardinality requirement. To derive them, we first relax the given simplex tableau into a disjunctive set, expressed in the space of nonbasic variables. We establish that coefficients of valid inequalities for the closed convex hull of this set obey ratios that can be computed directly from the simplex tableau. We show that a transportation problem can be used to separate these inequalities. We then give a constructive procedure to generate violated facet-defining inequalities for the closed convex hull of the disjunctive set using a variant of Prim's algorithm.

2.1 Introduction

In this chapter, we focus on CCOPs, where the optimization problem is linear and refer to them as cardinality-constrained linear programs (CCLPs). A CCLP can be formulated as

$$\begin{aligned} \text{maximize} \quad & c^\intercal x + d^\intercal y \\ \text{subject to} \quad & Ax + By \leq b \\ & x, y \geq 0 \\ & \text{card}(x) \leq K, \end{aligned}$$

where c, $x \in \mathbb{R}^p$, d, $y \in \mathbb{R}^q$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times p}$, $B \in \mathbb{R}^{m \times q}$, and K is a fixed positive integer with $K < p$. Our treatment extends to problems with multiple cardinality

constraints. However, for the sake of simplicity in the exposition, we only consider a single cardinality constraint in this research.

We conduct a polyhedral study of CCLPs in the space of original problem variables. In particular, we use information contained in feasible simplex tableaux of LP relaxations of CCLPs to construct strong valid inequalities. Our underlying motivation is that, avoiding the introduction of unnecessary indicator variables will help maintain the original problem structure, and might lead to streamlined solution approaches for these problems. Although we are not aware of previous studies of tableau-based cuts for cardinality-constrained problems, such inequalities have been proposed in the literature in the context of MIPs, quadratic programming, concave programming, and linear complementarity problems [2, 7, 29, 32, 37, 58, 69].

The remainder of this chapter is organized as follows. In Section 2.2, we show that violated cuts for CCLPs can be generated from a disjunctive relaxation of any simplex tableau corresponding to a basic feasible solution violating the cardinality requirement. This disjunctive relaxation has $(K + 1)$ disjuncts, each with a single nontrivial constraint. We also show that the analysis of the closed convex hull of this set can be performed in the space of nonbasic variables. In Section 2.3, we give a characterization of the closed convex hull based on the extreme ray representation of each disjunct, without the use of disjunctive programming. This characterization relates coefficients of facet-defining inequalities to extreme points of a polyhedron, that we give in closed-form. In Section 2.4, we show that there exists a nonlinear transformation that establishes an isomorphism between the face-lattice of this polyhedron and that of the dual of a transportation problem. In Section 2.5, we prove that nontrivial facet-defining inequalities of the closed convex hull of the disjunctive set correspond to label-connected spanning trees of the bipartite network associated with the transportation problem. This result allows us to provide, in Section 2.6, a simple explicit constructive procedure for the derivation of nontrivial facet-defining inequalities. It also yields a polynomial time algorithm for the generation of such inequalities, and give a precise characterization of when a commonly used disjunctive

cut, which we refer to as c-max cut, is facet-defining for the disjunctive relaxation. We give concluding remarks in Section 2.7.

2.2 Disjunctive relaxation of a simplex tableau with a cardinality constraint

Given an LP relaxation of a CCLP, we next describe an approach to construct cardinality-based cutting planes. We assume that we know a basic feasible solution of this LP relaxation that violates the cardinality constraint, $\text{card}(x) \leq K$, together with an explicit description of the associated simplex tableau. Denoting the basic variables in this tableau by v (indexed by set $\mathcal{M}$), and the nonbasic variables by t (indexed by set $\mathcal{V}$), we then write the simplex tableau as

$$\begin{aligned} v_l &= v_l^* - \textstyle\sum_{i\in\mathcal{V}} f_{li} t_i, && \forall l \in \mathcal{M}, \\ v_l &\geq 0, && \forall l \in \mathcal{M}, \\ t_i &\geq 0, && \forall i \in \mathcal{V}, \end{aligned} \tag{2.1}$$

where $v_l^* \geq 0$ for $l \in \mathcal{M}$. Since we have assumed that the current basic solution $(v, t) = (v^*, 0)$ does not satisfy the cardinality constraint, there exists a subset $\mathcal{L} \subseteq \mathcal{M}$ of basic variables such that (i) $|\mathcal{L}| = K + 1$, (ii) variables v_l for $l \in \mathcal{L}$ appear in the cardinality constraint, and (iii) $v_l^* > 0$ for $l \in \mathcal{L}$. We construct the desired disjunctive relaxation by (i) relaxing the cardinality constraint $\text{card}(x) \leq K$ into the disjunction $\bigvee_{l\in\mathcal{L}}(v_l \leq 0)$, which forces one of the $K + 1$ variables in $\mathcal{L}$ to be nonpositive, (ii) removing the nonnegativity requirements on basic variables, and (iii) omitting the tableau constraints associated with basic variables $v_{\mathcal{M}\setminus\mathcal{L}}$. We therefore study

$$\bar{Q} := \left\{ (v,t) \in \mathbb{R}^{|\mathcal{L}|+|\mathcal{V}|} : \begin{array}{ll} v_l = v_l^* - \sum_{i\in\mathcal{V}} f_{li} t_i, & \forall l \in \mathcal{L}, \\ t_i \geq 0, & \forall i \in \mathcal{V} \\ \bigvee_{l\in\mathcal{L}}(v_l \leq 0) & \end{array} \right\}, \tag{2.2}$$

where each equality in the above set corresponds to a basic variable in $\mathcal{L}$, and represents it as an affine function of the nonbasic variables. If a nonbasic variable is a slack variable for a constraint in $Ax + By \leq b$, then an inequality valid for $\bar{Q}$ can be

written in the space of original problem variables using the defining inequality for the slack variable. We remark that the relaxations applied to the initial simplex tableau in order to obtain $\bar{Q}$ resemble those made to obtain the corner relaxation of an MIP; see [31].

Since $\bar{Q}$ is a finite union of polyhedra, $\operatorname{cl}\operatorname{conv}(\bar{Q})$ is a polyhedron; see Theorem 19.6 in [59] for instance.

Proposition 2.2.1 *The set* $\operatorname{cl}\operatorname{conv}(\bar{Q})$ *is a polyhedron.* ■

A linear inequality is valid for $\bar{Q}$ if and only if it is valid for $\operatorname{cl}\operatorname{conv}(\bar{Q})$. We therefore seek to characterize the valid inequalities of $\operatorname{cl}\operatorname{conv}(\bar{Q})$. We next show that this can be achieved by studying $\operatorname{cl}\operatorname{conv}(Q)$ where Q is the projection of $\bar{Q}$ onto the space of nonbasic variables t. Formally, for each $l \in \mathcal{L}$, define

$$Q_l := \left\{ t \in \mathbb{R}^{|\mathcal{V}|} : \sum_{i \in \mathcal{V}} f_{li} t_i \geq v_l^*, \;\; t_i \geq 0, \; \forall i \in \mathcal{V} \right\}$$

and set $Q := \bigcup_{l \in \mathcal{L}} Q_l$. We assume without loss of generality that $\mathcal{V} = \{1, \ldots, n\}$. Let $h^*(t) = \begin{pmatrix} v^* \\ 0 \end{pmatrix} + \begin{pmatrix} -F \\ I \end{pmatrix} t$, where the entry (l, i) of matrix F is f_{li}, as used in the definition of $\bar{Q}$ in (2.2). It is clear that $h^*(\cdot)$ is an affine map, and that $\bar{Q} = h^*(Q)$.

Lemma 1 *For $i = 1, \ldots, p$, let $P_i \in \mathbb{R}^n$ be nonempty polyhedra. Also let $h : \mathbb{R}^n \to \mathbb{R}^m$ be an affine map. Then*

$$h\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^p P_i\right)\right) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^p h(P_i)\right).$$

Proof It is easy to show that $h(\bigcup_{i=1}^p P_i) = \bigcup_{i=1}^p h(P_i)$. Then,

$$\begin{aligned} \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^p h(P_i)\right) &= \operatorname{cl}\operatorname{conv}\left(h\left(\bigcup_{i=1}^p P_i\right)\right) \\ &= \operatorname{cl}\left(h\left(\operatorname{conv}\left(\bigcup_{i=1}^p P_i\right)\right)\right) \\ &\supseteq h\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^p P_i\right)\right), \end{aligned}$$

where the second equality results from the fact that convex hull operators and affine maps commute, and the last inclusion follows from the continuity of h. On the other hand, it is straightforward that

$$\begin{aligned}
\operatorname{conv}\left(\bigcup_{i=1}^{p} h(P_i)\right) &= \operatorname{conv}\left(h\left(\bigcup_{i=1}^{p} P_i\right)\right) \\
&= h\left(\operatorname{conv}\left(\bigcup_{i=1}^{p} P_i\right)\right) \\
&\subseteq h\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^{p} P_i\right)\right).
\end{aligned}$$

By Theorem 19.6 in [59], $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^{p} P_i\right)$ is a polyhedron and hence its affine transformation is a polyhedron. This implies that $h\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^{p} P_i\right)\right)$ is a closed set and that $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^{p} h(P_i)\right) \subseteq h\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^{p} P_i\right)\right)$. ■

For the affine function $h^*(\cdot)$ described above, Lemma 1 implies

Proposition 2.2.2 *It holds that* $\operatorname{cl}\operatorname{conv}(\bar{Q}) = h^*\left(\operatorname{cl}\operatorname{conv}(Q)\right)$.

In the remainder of this chapter, we restrict our attention to the study of $\operatorname{cl}\operatorname{conv}(Q)$ since Proposition 2.2.2 shows that this is sufficient to characterize $\operatorname{cl}\operatorname{conv}(\bar{Q})$.

Since we assumed that $v_l^* > 0$ for $l \in \mathcal{L}$, we may scale each constraint so that $v_l^* = 1$. That is, for $l \in \mathcal{L}$,

$$Q_l = \left\{ t : \sum_{i \in \mathcal{V}} f_{li} t_i \geq 1, \quad t_i \geq 0, \forall i \in \mathcal{V} \right\}.$$

For each $l \in \mathcal{L}$, define

$$\begin{aligned}
\mathcal{I}_+^l &= \{i \in \mathcal{V} : f_{li} > 0\}, \\
\mathcal{I}_-^l &= \{i \in \mathcal{V} : f_{li} < 0\}, \\
\mathcal{I}_0^l &= \{i \in \mathcal{V} : f_{li} = 0\}.
\end{aligned}$$

Throughout the chapter, we assume without loss of generality that $Q_l \neq \emptyset$ for each $l \in \mathcal{L}$. In fact, if $Q_l = \emptyset$ for some $l \in \mathcal{L}$, then we can simply drop the corresponding set from the disjunction. It is simple to verify that $Q_l = \emptyset$ if and only if $f_{li} \leq 0$ for all $i \in \mathcal{V}$. For this reason, we make the following assumption in the rest of the chapter.

Assumption 1 *For each $l \in \mathcal{L}$, $\mathcal{I}_+^l \neq \emptyset$.*

Proposition 2.2.3 *Polyhedron Q_l is full-dimensional. Further,* $\operatorname{cl}\operatorname{conv}(Q)$ *is also full-dimensional.*

Proof By Assumption 1, $\mathcal{I}_+^l \neq \emptyset$. Choose $i \in \mathcal{I}_+^l$ and consider the point

$$\left(\frac{1}{f_{li}} + 1\right) e_i + \sum_{k \in \mathcal{V} \setminus \{i\}} \epsilon e_k,$$

where ϵ is positive but sufficiently small. This point is in the interior of Q_l and hence Q_l is full-dimensional. Further, since $Q_l \subseteq Q$, then $\operatorname{cl}\operatorname{conv}(Q)$ is also full-dimensional. ■

We next argue that there are valid inequalities of $\operatorname{cl}\operatorname{conv}(Q)$ that can be used to separate the basic feasible solution associated with the initial simplex tableau (2.1), if this solution violates the cardinality requirement. For instance, consider the inequality

$$\sum_{i \in \mathcal{V}} (\text{c}-\max)_i t_i \geq 1 \tag{2.3}$$

where $(\text{c}-\max)_i = \max\{f_{li} : l \in \mathcal{L}\}$ for $i \in \mathcal{V}$. This inequality, which we refer to hereafter as *c-max cut* was introduced in [37] for complementarity problems. Complementarity problems are special instances of cardinality problems requiring that at most one of the variables takes a nonzero value. The c-max cut can be easily seen to be valid for $\operatorname{cl}\operatorname{conv}(Q)$ because $\sum_{i \in \mathcal{V}} (\text{c}-\max)_i t_i \geq \sum_{i \in \mathcal{V}} f_{li} t_i \geq 1$ for all $t \in Q_l$ and $l \in \mathcal{L}$, *i.e.*, it is valid for each disjunct Q_l. Moreover, it separates the closed convex hull from $t = 0$ because this point violates (2.3). For the particular case when $|\mathcal{L}| = 2$, it is shown in [56] that the c-max cut is not always facet-defining for $\operatorname{cl}\operatorname{conv}(Q)$ and is not sufficient to provide a complete description of the nontrivial inequalities of $\operatorname{cl}\operatorname{conv}(Q)$. In this chapter, we provide a complete description of the nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q)$. We show that all nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q)$ cut off the current basic feasible solution of (2.1), and we precisely characterize when the c-max cut is strong.

2.3 A characterization of $\operatorname{cl}\operatorname{conv}(Q)$

In this section, we provide a characterization of the facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q)$. Recall that Minkowski-Weyl's theorem, see Theorem 7.13 in [36] for instance, establishes that a polyhedron can be represented in two forms, either using its vertices and extreme rays or as a finite intersection of half-spaces. Following [74], we refer to the former representation as a *$\mathcal{V}$-polyhedron*, and to the latter as a *$\mathcal{H}$-polyhedron*. Through the rest of the chapter, we alternate between these representations when studying $\operatorname{cl}\operatorname{conv}(Q)$. We also find it more convenient to study a certain homogenization of Q. We show in Proposition 2.3.4 that studying this homogenization is without loss of generality.

Let $\mathcal{V}_{\mathbf{0}} := \mathcal{V} \cup \{0\}$. Define Q_l^0 to be the homogenization of Q_l obtained as

$$Q_l^0 := \left\{ t := (t_1, \ldots, t_n, t_0) \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : \sum_{i \in \mathcal{V}} f_{li} t_i \geq t_0, \;\; t \geq 0 \right\}.$$

After defining $f_{l0} := -1$ and $f_l := (f_{l1}, \ldots, f_{ln}, f_{l0})^\intercal$, we can rewrite Q_l^0 as

$$Q_l^0 = \left\{ t \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : f_l^\intercal t \geq 0, \;\; t \geq 0 \right\}.$$

We refer to $f_l^\intercal t \geq 0$ as the nontrivial constraint of disjunct l. It is clear that Q_l^0 is a polyhedral cone. Referring to $\bigcup_{l \in \mathcal{L}} Q_l^0$ as Q^0, it is also clear that $\operatorname{cl}\operatorname{conv}(Q^0)$ is a cone. We next describe how these cones relate to the sets we originally introduced. For a nonempty convex set C, we define $K(C) := \{\lambda(d, 1) : d \in C, \lambda > 0\}$.

Proposition 2.3.1 *It holds that*

1. $Q_l^0 = \operatorname{cl}(K(Q_l))$.

2. $\operatorname{cl}\operatorname{conv}(Q^0) = \operatorname{cl}(K(\operatorname{cl}\operatorname{conv}(Q)))$.

Proof First, we prove 1. We refer to $K(Q_l)$ as $\mathcal{K}$. To show $\operatorname{cl}(\mathcal{K}) \subseteq Q_l^0$, consider $\lambda(d, 1) \in \mathcal{K}$ for some $\lambda > 0$ and $d \in Q_l$. Then, $\sum_{i \in \mathcal{V}} f_{li} d_i \geq 1$. Since $\lambda(d, 1) \geq 0$ and $f_l^\intercal \lambda(d, 1) = \lambda(\sum_{i \in \mathcal{V}} f_{li} d_i - 1) \geq 0$, then $\lambda(d, 1) \in Q_l^0$. Since Q_l^0 is a closed set,

$\mathrm{cl}(\mathcal{K}) \subseteq Q_l^0$. To show $Q_l^0 \subseteq \mathrm{cl}(\mathcal{K})$, consider $(d, d_0) \in Q_l^0$. If $d_0 > 0$, then $d/d_0 \in Q_l$ and hence $(d, d_0) = d_0(d/d_0, 1) \in \mathcal{K}$. Now, assume $d_0 = 0$. Then, $\sum_{i\in\mathcal{V}} f_{li} d_i \geq 0$. Since $Q_l \neq \emptyset$ by Assumption 1, we may select $d' \in Q_l$. Then, for any $\mu > 0$, $d' + \mu d \in Q_l$ because $\sum_{i\in\mathcal{V}} f_{li}(d' + \mu d)_i = \sum_{i\in\mathcal{V}} f_{li} d'_i + \mu \sum_{i\in\mathcal{V}} f_{li} d_i \geq 1$. Hence $(1/\mu)(d' + \mu d, 1) \in \mathcal{K}$. Observe that $\lim_{\mu\to\infty}(1/\mu)(d' + \mu d, 1) = (d, 0)$. Therefore, $(d, d_0) = (d, 0) \in \mathrm{cl}(\mathcal{K})$.

We next prove 2. Clearly, $\mathrm{cl\,conv}(Q^0) \supseteq K(\mathrm{cl\,conv}(Q))$ because $K(\mathrm{cl\,conv}(Q)) \subseteq \mathrm{cl\,conv}(K(Q)) \subseteq \mathrm{cl\,conv}(Q^0)$, where the first inclusion holds because $\mathrm{cl\,conv}(Q) \subseteq \mathrm{cl\,conv}(K(Q))$ and $\mathrm{cl\,conv}(K(Q))$ is a cone, and the second inclusion is because $K(Q) = \bigcup_{l\in\mathcal{L}} K(Q_l) \subseteq \bigcup_{l\in\mathcal{L}} Q_l^0 = Q^0$. For the reverse inclusion, observe that:

$$Q^0 = \bigcup_{l\in\mathcal{L}} Q_l^0 = \bigcup_{l\in\mathcal{L}} \mathrm{cl}(K(Q_l)) \subseteq \mathrm{cl}(K\,\mathrm{cl\,conv}(Q))),$$

where the first equality is by definition of Q^0, the second by Part 1, and the first inclusion is because $Q_l \subseteq \mathrm{cl\,conv}(Q)$. Since $\mathrm{cl}(K(\mathrm{cl\,conv}(Q)))$ is closed and convex, $\mathrm{cl\,conv}(Q^0) \subseteq \mathrm{cl}(K(\mathrm{cl\,conv}(Q)))$. ■

Propositions 2.2.3 and 2.3.1 directly yield

Corollary 1 *Polyhedron* $\mathrm{cl\,conv}(Q^0)$ *is full-dimensional.*

In Proposition 2.3.2, we present $\mathcal{V}$-polyhedron representations of Q_l^0 and Q_l. This result allows us to build $\mathcal{V}$-polyhedron representations of $\mathrm{cl\,conv}(Q^0)$ and $\mathrm{cl\,conv}(Q)$ in Corollary 2.

Proposition 2.3.2 *Define*

$$\begin{aligned}
R_l^0 &= \{f_{li}e_j - f_{lj}e_i \in \mathbb{R}^{|\mathcal{V}_0|} : i \in \mathcal{I}_+^l, j \in \mathcal{I}_-^l \cup \{0\}\} \cup \{e_k \in \mathbb{R}^{|\mathcal{V}_0|} : k \in \mathcal{I}_+^l \cup \mathcal{I}_0^l\}, \\
V_l &= \left\{\frac{1}{f_{li}} e_i \in \mathbb{R}^{|\mathcal{V}|} : i \in \mathcal{I}_+^l\right\}, \\
R_l &= \{f_{li}e_j - f_{lj}e_i \in \mathbb{R}^{|\mathcal{V}|} : i \in \mathcal{I}_+^l, j \in \mathcal{I}_-^l\} \cup \{e_k \in \mathbb{R}^{|\mathcal{V}|} : k \in \mathcal{I}_+^l \cup \mathcal{I}_0^l\}.
\end{aligned}$$

Then $Q_l^0 = \mathrm{cone}(R_l^0)$ *and* $Q_l = \mathrm{conv}\,(V_l) + \mathrm{cone}(R_l)$. *Further all the given points and rays are extremal.*

Proof We first show that $Q_l^0 = \text{cone}(R_l^0)$. Since Q_l^0 is a cone in the nonnegative orthant, it is pointed. This implies that all the points in the cone can be written as a conic combination of its extreme rays. Let r be a ray of Q_l^0. Then, r is extreme if and only if it belongs to the intersection of $n = |\mathcal{V}_\mathbf{0}| - 1$ independent hyperplanes among $\{t \in \mathbb{R}^{|\mathcal{V}_\mathbf{0}|} : f_l^\intercal t = 0\}$ and $\{t \in \mathbb{R}^{|\mathcal{V}_\mathbf{0}|} : t_k = 0\}$, for $k \in \mathcal{V}_\mathbf{0}$. First, for each $i \in \mathcal{V}_\mathbf{0}$, suppose that these n hyperplanes are $\{t : t_k = 0\}$ for $k \neq i$. Then $r_k = 0$ for all $k \neq i$ and hence $r = \rho e_i$ with $\rho > 0$. In order to be a ray, this vector must satisfy $f_l^\intercal r \geq 0$, *i.e.*, i must be chosen in $\mathcal{I}_+^l \cup \mathcal{I}_0^l$. Next, suppose that these n hyperplanes are $\{t \in \mathbb{R}^{|\mathcal{V}_\mathbf{0}|} : f_l^\intercal t = 0\}$ and $\{t \in \mathbb{R}^{|\mathcal{V}_\mathbf{0}|} : t_k = 0\}$ for $k \neq i, j$ for some $i, j \in \mathcal{V}_\mathbf{0}$. Then the face defined by the intersection of these hyperplanes is $F := \{t \in \mathbb{R}^{|\mathcal{V}_\mathbf{0}|} : f_{li}t_i + f_{lj}t_j = 0, t \geq 0, t_k = 0, \forall k \neq i, j\}$. In order for r to be a ray, $F \neq \{0\}$ and hence $f_{li}f_{lj} \leq 0$. By independence, $f_{li} \neq 0$ or $f_{lj} \neq 0$. If $f_{li} = 0$ or $f_{lj} = 0$ then we have that $r = e_k$ for some $k \in \mathcal{I}_0^l$. Now assume that $f_{li}f_{lj} < 0$. Without loss of generality, assume that $f_{li} > 0$ and $f_{lj} < 0$. Then, $r = f_{li}e_j - f_{lj}e_i$ where $i \in \mathcal{I}_+^l$ and $j \in \mathcal{I}_-^l \cup \{0\}$. We conclude that R_l^0 is precisely the collection of extreme rays of Q_l^0, and therefore $Q_l^0 = \text{cone}(R_l^0)$.

We next prove that $Q_l = \text{conv}\,(V_l) + \text{cone}(R_l)$. By Proposition 2.3.1 and by Lemma 5.41 of [36], we have that

$$Q_l = \text{conv}(V_l') + \text{cone}(R_l')$$

where V_l' is the set of vectors $\frac{t}{t_0}$ in $\mathbb{R}^{|\mathcal{V}|}$ obtained from rays (t, t_0) of Q_l^0 whose component t_0 is nonzero, and R_l' is the set of vectors t in $\mathbb{R}^{|\mathcal{V}|}$ obtained from rays (t, t_0) of Q_l^0 where $t_0 = 0$. Thus, $V_l' = V_l$ and $R_l' = R_l$. Extremality follows directly from the extremality of rays in R_l^0. ■

The result of Proposition 2.3.2 yields a $\mathcal{V}$-polyhedron representation for the closed convex hull of the union of the associated disjuncts.

Corollary 2 *The $\mathcal{V}$-polyhedron representations of* $\text{cl}\,\text{conv}(Q^0)$ *and* $\text{cl}\,\text{conv}(Q)$ *are*

$$\begin{aligned} \text{cl}\,\text{conv}(Q^0) &= \text{cone}(R^0), \\ \text{cl}\,\text{conv}(Q) &= \text{conv}(V) + \text{cone}(R), \end{aligned}$$

where $R^0 := \bigcup_{l\in\mathcal{L}} R^0_l, V := \bigcup_{l\in\mathcal{L}} V_l$, *and* $R := \bigcup_{l\in\mathcal{L}} R_l$.

We now seek to better understand the coefficient vectors $\beta \in \mathbb{R}^{|\mathcal{V}|}$ and $\beta' \in \mathbb{R}^{|\mathcal{V}_0|}$ that give rise to strong valid inequalities of $\operatorname{cl}\operatorname{conv}(Q)$ and $\operatorname{cl}\operatorname{conv}(Q^0)$, respectively. We show next that β and β' are closely related.

Proposition 2.3.3 *Inequality*

$$\sum_{i\in\mathcal{V}} \beta_i t_i \geq \gamma \tag{2.4}$$

is valid for $\operatorname{cl}\operatorname{conv}(Q)$ *if and only if inequality*

$$\sum_{i\in\mathcal{V}} \beta_i t_i \geq \gamma t_0 \tag{2.5}$$

is valid for $\operatorname{cl}\operatorname{conv}(Q^0)$.

Proof For the direct implication, suppose that (2.4) is valid for $\operatorname{cl}\operatorname{conv}(Q)$. Take any $(d, d_0) \in K(\operatorname{cl}\operatorname{conv}(Q))$. By definition, $(d, d_0) = d_0(\frac{d}{d_0}, 1)$ where $d_0 > 0$ and $\frac{d}{d_0} \in \operatorname{cl}\operatorname{conv}(Q)$. It follows that $\sum_{i\in\mathcal{V}} \beta_i d_i \geq \gamma d_0$. This shows that (2.5) is valid for $K(\operatorname{cl}\operatorname{conv}(Q))$, and therefore, by Proposition 2.3.1, for $\operatorname{cl}\operatorname{conv}(Q^0)$. For the reverse implication, suppose now that (2.5) is valid for $\operatorname{cl}\operatorname{conv}(Q^0)$. Take any $d \in Q$. Then $(d, 1) \in Q^0$. Since (2.5) is valid for $\operatorname{cl}\operatorname{conv}(Q^0)$, then $\sum_{i\in\mathcal{V}} \beta_i d_i \geq \gamma$. This shows that (2.4) is valid for $\operatorname{cl}\operatorname{conv}(Q)$. ■

The following result shows that characterizing the facets of $\operatorname{cl}\operatorname{conv}(Q^0)$ is equivalent to characterizing the facets of $\operatorname{cl}\operatorname{conv}(Q)$.

Proposition 2.3.4 *Inequality* (2.4) *is facet-defining for* $\operatorname{cl}\operatorname{conv}(Q)$ *if and only if inequality* (2.5) *is facet-defining for* $\operatorname{cl}\operatorname{conv}(Q^0)$ *and is not a scalar multiple of* $t_0 \geq 0$.

Proof The fact that validity is preserved is shown in Proposition 2.3.3. Suppose now that (2.4) is facet-defining for $\operatorname{cl}\operatorname{conv}(Q)$. Then there exist $n = |\mathcal{V}|$ affinely independent points $w^1, \ldots, w^n$ of $\operatorname{cl}\operatorname{conv}(Q)$ that satisfy (2.4) at equality. Points $(w^j, 1)$ belong to $\operatorname{cl}\operatorname{conv}(Q^0)$ for all $j \in \mathcal{V}$ and satisfy (2.5) at equality. Since $\{w^j :$

$j \in \mathcal{V}\}$ are affinely independent, $\{(w^j, 1) : j \in \mathcal{V}\}$ are linearly independent. This proves that (2.5) is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$. Clearly, (2.5) is not $t_0 \geq 0$ as otherwise (2.4) would be $0^\intercal t \geq -1$, which is not facet-defining for $\operatorname{cl}\operatorname{conv}(Q)$.

Conversely, suppose that (2.5) is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$. Since $\operatorname{cl}\operatorname{conv}(Q^0)$ is a full-dimensional polyhedral cone, there exist n linearly independent extreme rays (r^j, r_0^j) of $\operatorname{cl}\operatorname{conv}(Q^0)$ that satisfy (2.5) at equality. Suppose $r_0^j = 0$ for all $j \in \mathcal{V}$. Observe that $\{r^j : j \in \mathcal{V}\}$ are linearly independent and $\beta^\intercal r^j = 0$ for all $j \in \mathcal{V}$. This shows that $\beta = 0$. However, this is not possible as (2.5) would then correspond to the face of $\operatorname{cl}\operatorname{conv}(Q^0)$ induced by $t_0 \geq 0$. Therefore, there must exist $j \in \mathcal{V}$ such that $r_0^j \neq 0$. Define $I_1 = \{j' \in \mathcal{V} : r_0^{j'} \neq 0\}(\neq \emptyset)$ and $I_2 = \{j' \in \mathcal{V} : r_0^{j'} = 0\}$. Then, for $j \in I_1$, $\beta^\intercal \frac{r^j}{r_0^j} = \frac{1}{r_0^j}\beta^\intercal r^j = \gamma$. Further, for $k \in I_2$, $\beta^\intercal r^k = 0$. Fix $j_0 \in I_1$, and consider the sets of points

$$\left\{\frac{r^j}{r_0^j} : j \in I_1\right\} \cup \left\{\frac{r^{j_0}}{r_0^{j_0}} + r^k : k \in I_2\right\}.$$

It is clear that these points satisfy (2.4) at equality and that they belong to $\operatorname{cl}\operatorname{conv}(Q)$ by Proposition 2.3.1. It remains to prove that they are affinely independent, which can be established easily because the linear independence of vectors

$$\left\{\frac{r^j}{r_0^j} - \frac{r^{j_0}}{r_0^{j_0}} : j \in I_1 \setminus \{j_0\}\right\} \cup \left\{r^k : k \in I_2\right\}$$

follows from the assumed independence of $\{(r^j, r^0), j \in \mathcal{V}\}$. Therefore, (2.4) is facet-defining for $\operatorname{cl}\operatorname{conv}(Q)$. ■

In the remainder of this chapter, we prefer to study $\operatorname{cl}\operatorname{conv}(Q^0)$ because, being homogeneous, it allows for a unified treatment of the extreme points and extreme rays of Q, and thus permits a more streamlined presentation.

We are now ready to further investigate the structure of coefficient vectors associated with facet-defining inequalities (2.5) of $\operatorname{cl}\operatorname{conv}(Q^0)$. In particular, we will show in Proposition 2.3.6 that, except for some simple inequalities we describe next, most facet-defining inequalities are such that $\gamma > 0$.

For $i \in \mathcal{V}_{\mathbf{0}}$, we refer to the inequalities $t_i \geq 0$ of $\operatorname{cl}\operatorname{conv}(Q^0)$ as *trivial*. For notational convenience, we redefine $\mathcal{I}_-^l := \mathcal{I}_-^l \cup \{0\}$. Hence $\mathcal{I}_+^l$, $\mathcal{I}_-^l$, and $\mathcal{I}_0^l$ exclusively partition $\mathcal{V}_{\mathbf{0}}$. We also define

$$\begin{aligned}
\mathcal{I}_+ &= \{i \in \mathcal{V}_{\mathbf{0}} : f_{li} > 0 \text{ for some } l \in \mathcal{L}\} &&= \textstyle\bigcup_{l\in\mathcal{L}} \mathcal{I}_+^l, \\
\mathcal{I}_- &= \{i \in \mathcal{V}_{\mathbf{0}} : f_{li} < 0 \text{ for all } l \in \mathcal{L}\} &&= \textstyle\bigcap_{l\in\mathcal{L}} \mathcal{I}_-^l, \\
\mathcal{I}_0 &= \mathcal{V}_{\mathbf{0}} \setminus (\mathcal{I}_+ \cup \mathcal{I}_-).
\end{aligned}$$

It is clear that $0 \in \mathcal{I}_-$ and it follows from Assumption 1 that $\mathcal{I}_+ \neq \emptyset$. We establish next that trivial inequalities are typically facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$.

Proposition 2.3.5 *Trivial inequality $t_i \geq 0$ is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$ if and only if*

1. *$i \in \mathcal{I}_- \cup \mathcal{I}_0$, or*

2. *$i \in \mathcal{I}_+$ and $|\mathcal{I}_+| \geq 2$.*

Proof Inequality $t_i \geq 0$ is clearly valid for $\operatorname{cl}\operatorname{conv}(Q^0)$. Assume first that $i \in \mathcal{I}_- \cup \mathcal{I}_0$. Since $\mathcal{I}_+ \neq \emptyset$, there exists $j \in \mathcal{I}_+$ and $l \in \mathcal{L}$ such that $f_{lj} > 0$. Consider the point

$$\sum_{k\in\mathcal{V}_{\mathbf{0}}\setminus\{i,j\}} \epsilon e_k + \frac{1 - \sum_{k\in\mathcal{V}_{\mathbf{0}}\setminus\{i,j\}} \epsilon f_{lk}}{f_{lj}} e_j \tag{2.6}$$

for ϵ positive and sufficiently small. This point is in the relative interior of $Q_l^0 \cap \{t \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : t_i = 0\}$. Hence, it is in the relative interior of $\operatorname{cl}\operatorname{conv}(Q^0) \cap \{t \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : t_i = 0\}$. It follows that $t_i \geq 0$ is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$. Next, assume that $i \in \mathcal{I}_+$. If $|\mathcal{I}_+| \geq 2$, there exists $j \in \mathcal{I}_+ \setminus \{i\}$ and $l \in \mathcal{L}$ such that $f_{lj} > 0$. Then, (2.6) is an interior point of $Q_l^0 \cap \{t \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : t_i = 0\}$ and hence $t_i \geq 0$ is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$. Suppose $\mathcal{I}_+ = \{i\}$ and $j \in \mathcal{I}_-$. Then, for each $l \in \mathcal{L}$, every point in $Q_l^0 \cap \{t \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : t_i = 0\}$ satisfies $t_j = 0$. It follows that $t_i \geq 0$ defines a face of $\operatorname{cl}\operatorname{conv}(Q^0)$ of dimension at least two less than that of $\operatorname{cl}\operatorname{conv}(Q^0)$, showing that this inequality is not facet-defining. ■

Proposition 2.3.5 shows that trivial inequalities $t_i \geq 0$ are facet-defining unless $i \in \mathcal{I}_+$ and $|\mathcal{I}_+| = 1$. In the remainder of this chapter, we consider β to be a vector in $\mathbb{R}^{|\mathcal{V}_0|}$. We show next that the sign of the entries of coefficient vectors β for nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$ can be deduced directly from the sets $\mathcal{I}_+, \mathcal{I}_-$, and $\mathcal{I}_0$.

Proposition 2.3.6 *Let*

$$\sum_{i \in \mathcal{V}_0} \beta_i t_i \geq 0. \tag{2.7}$$

be a nontrivial facet-defining inequality for $\operatorname{cl}\operatorname{conv}(Q^0)$. *Then*

1. $\beta_i \geq -\max\{f_{li} : l \in \mathcal{L}\}\beta_0$ *for* $i \in \mathcal{I}_+$,
2. $\beta_j < 0$ *for* $j \in \mathcal{I}_-$,
3. $\beta_k = 0$ *if* $\max\{f_{lk} : l \in \mathcal{L}\} = 0$.

In particular, $\beta_i > 0$ *for* $i \in \mathcal{I}_+$.

Proof Consider a nontrivial facet-defining inequality (2.7). Observe that $\beta_i \geq 0$ for $i \in \mathcal{I}_+ \cup \mathcal{I}_0$ because e_i is a ray of Q^0.

We first prove 1. Choose $j' \in \mathcal{I}_-$ with $\beta_{j'} < 0$. Such a j' exists because otherwise, (2.7) is implied by trivial inequalities. Let $i \in \mathcal{I}_+^l$ for some $l \in \mathcal{L}$. Since $f_{li}e_{j'} - f_{lj'}e_i$ is a ray for $\operatorname{cl}\operatorname{conv}(Q^0)$, it follows that $\beta_i \geq \max\{f_{li} : l \in \mathcal{L}\}\frac{\beta_{j'}}{f_{lj'}} > 0$. Remember now that $0 \in \mathcal{I}_-$. If $\beta_0 < 0$, Part 1 follows easily since $f_{l0} = -1$. If $\beta_0 = 0$, Part 1 simply states that $\beta_i \geq 0$ while the inequality just proven for j' is stronger.

We now prove 2 and 3. Consider $j \in \mathcal{I}_- \cup \mathcal{I}_0$. There exists an extreme ray r of $\operatorname{cl}\operatorname{conv}(Q^0)$ such that $\beta^\intercal r = 0$ and $r_j > 0$ because otherwise, (2.7) is a trivial inequality. Proposition 2.3.2 shows that this ray can be of one of two forms. First assume that $r = f_{li}e_j - f_{lj}e_i$ for some $l \in \mathcal{L}$ and $i \in \mathcal{I}_+^l$. As shown above, $\beta_i > 0$. It follows from $\beta^\intercal r = 0$ that $\beta_j < 0$. This shows Part 2 when $j \in \mathcal{I}_-$ and shows that it is not the desired ray when $j \in \mathcal{I}_0$ as it contradicts the already established relation

$\beta_j \geq 0$. Now, consider $j \in \mathcal{I}_0$. We must have that $r = e_j$. This shows that $\beta_j = 0$ proving Part 3. ■

Example 1 *Consider the set Q^0 with disjuncts defined by the constraints*

$$\begin{array}{llllll} 5t_1 & -3t_2 & +0t_3 & +1t_4 & -5t_5 & -t_0 & \geq 0 \\ 3t_1 & -1t_2 & +2t_3 & -3t_4 & -3t_5 & -t_0 & \geq 0 \\ 4t_1 & -6t_2 & +4t_3 & -2t_4 & +0t_5 & -t_0 & \geq 0 \\ 2t_1 & -2t_2 & -2t_3 & +0t_4 & -2t_5 & -t_0 & \geq 0. \end{array} \tag{2.8}$$

Then

$$\mathcal{I}_+ = \{1, 3, 4\}, \quad \mathcal{I}_- = \{2, 0\}, \text{ and } \mathcal{I}_0 = \{5\}.$$

We use PORTA [19, 20] to obtain the extreme rays of each disjunct independently. We then run PORTA again based on this collection of extreme rays to obtain all facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$. *The resulting nontrivial facet-defining inequalities are*

$$\begin{array}{llllll} 5t_1 & -\frac{5}{3}t_2 & +4t_3 & +t_4 & +0t_5 & -t_0 & \geq 0 \\ 9t_1 & -3t_2 & +6t_3 & +t_4 & +0t_5 & -t_0 & \geq 0 \\ 6t_1 & -2t_2 & +4t_3 & +t_4 & +0t_5 & -t_0 & \geq 0. \end{array} \tag{2.9}$$

We observe that, as argued in Proposition 2.3.6, $\beta_i > 0$ for $i \in \mathcal{I}_+$, $\beta_i < 0$ for $i \in \mathcal{I}_-$ and $\beta_5 = 0$ in all nontrivial facet-defining inequalities (2.9).

Proposition 2.3.6 shows that nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$ are such that β_k is zero for each index k for which the tableau coefficients satisfy $f_{lk} \leq 0$ for all $l \in \mathcal{L}$ and $f_{l'k} = 0$ for some $l' \in \mathcal{L}$. Then, it is clear that $\operatorname{cl}\operatorname{conv}(Q) = \{t = (t_{-k}, t_k) : t_{-k} \in \operatorname{cl}\operatorname{conv}(Q_{-k}), t_k \in \mathbb{R}_+\}$ where t_{-k} is the vector obtained by dropping component t_k from t and $Q_{-k} := \operatorname{proj}_{t_{-k}}(Q)$. Thus, it is sufficient to study $\operatorname{cl}\operatorname{conv}(Q_{-k})$. For this reason, we make the following assumption in the remainder of the chapter.

Assumption 2 $\mathcal{I}_0 = \emptyset$.

With Assumption 2, it follows that $\mathcal{I}_+$ and $\mathcal{I}_-$ exclusively partition $\mathcal{V}_{\mathbf{0}}$.

We next derive an $\mathcal{H}$-polyhedron representation of $\operatorname{cl}\operatorname{conv}(Q^0)$. We obtain the linear inequalities of this representation by considering the dual cone of its $\mathcal{V}$-polyhedron representation, which was obtained in Corollary 2. For a given cone $C \subseteq \mathbb{R}^n$, we denote the dual cone of C by C^*. Recall that $C^* = \{y \in \mathbb{R}^n : y^\intercal x \geq 0,\ \forall x \in C\}$. As we established in Corollary 2 that $\operatorname{cl}\operatorname{conv}(Q^0) = \operatorname{cone}(R^0)$ where $R^0 := \bigcup_{l \in \mathcal{L}} R^0_l$, it is easy to see that $\beta^\intercal t \geq 0$ is a valid inequality for $\operatorname{cl}\operatorname{conv}(Q^0)$ if and only if $\beta^\intercal r \geq 0$ for all $r \in R^0$. Therefore, the coefficient vectors of valid inequalities for $\operatorname{cl}\operatorname{conv}(Q^0)$ belong to

$$B_1 = \left\{ \beta \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : \begin{array}{lll} f_{li}\beta_j - f_{lj}\beta_i \geq 0, & \forall (i,j) \in \mathcal{I}^l_+ \times \mathcal{I}^l_-, & l \in \mathcal{L} \\ \beta_k \geq 0, & \forall k \in \mathcal{I}^l_+ \cup \mathcal{I}^l_0, & l \in \mathcal{L} \end{array} \right\}, \tag{2.10}$$

where we use B_1 as a shorthand notation for $[\operatorname{cl}\operatorname{conv}(Q^0)]^*$.

Among the facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$, trivial inequalities are not useful in practice, since they do not cut off the basic solution associated with simplex tableau (2.1). We therefore concentrate on nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$, which have $\beta_0 < 0$ as shown in Proposition 2.3.6. Therefore, by scaling if necessary, we may assume that $\beta_0 = -1$. For this reason, we focus our ensuing study on $B_2 := B_1 \cap \{\beta \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : \beta_0 = -1\}$, and show that the description of this polyhedron requires fewer constraints than those given in (2.10).

Proposition 2.3.7 *For $(i,j) \in \mathcal{I}_+ \times \mathcal{I}_-$, define*

$$w_{ij} = \min \left\{ -\frac{f_{lj}}{f_{li}} : f_{li} > 0, l \in \mathcal{L} \right\}. \tag{2.11}$$

Then

$$B_2 = \left\{ \beta \in \mathbb{R}^{|\mathcal{V}_{\mathbf{0}}|} : \begin{array}{ll} \beta_j + w_{ij}\beta_i \geq 0, & \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_- \\ \beta_0 = -1 & \end{array} \right\}. \tag{2.12}$$

Proof Just as in the proof of Proposition 2.3.6, when $\beta_0 < 0$, the inequalities $\beta_k \geq 0$ for $k \in \mathcal{I}^l_+ \cup \mathcal{I}^l_0$ do not support B_1 and can therefore be dropped. Now, for any $i \in \mathcal{I}^l_+$ and $j \in \mathcal{I}^l_-$, $\beta_i \geq \frac{f_{li}}{f_{lj}}\beta_j$. This inequality is redundant if $j \in \mathcal{I}_+$ because, as argued above, $\beta_i > 0$. Therefore, $j \in \mathcal{I}_-$. Maximizing $\frac{f_{li}}{f_{lj}}\beta_j$ yields (2.12). ■

It is easy to see that the coefficients $\beta \in B_2$ are sign-constrained. Therefore, B_2 has no lines. Because B_2 does not have a line, it has at least one extreme point; see Corollary 18.5.3 in [59]. We mention that B_2 does also have rays, including vectors e_i for $i \in \mathcal{I}_+ \cup \mathcal{I}_- \backslash \{0\}$.

We next show that there is a one-to-one correspondence between the nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$ and the extreme points of B_2.

Theorem 2.3.1 *Any inequality $\beta^\intercal t \geq 0$ with $\beta_0 = -1$ is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$ if and only if β is an extreme point of B_2.*

Proof For a facet-defining inequality, $\beta^\intercal t \geq 0$ of $\operatorname{cl}\operatorname{conv}(Q^0)$, β is an extreme point of B_2 because of the n linearly independent tight constraints $\beta^\intercal r^j = 0$, one for each tight linearly independent extreme ray r^j of $\operatorname{cl}\operatorname{conv}(Q^0)$ and the equality constraint $\beta_0 = -1$. For the reverse inclusion, the tight constraints, besides $\beta_0 = -1$, each yield a linearly independent extreme ray tight for the inequality. ■

Extreme rays of B_2 also lead to valid inequalities for $\operatorname{cl}\operatorname{conv}(Q^0)$. In fact, consider a solution β and an extreme ray ρ of B_2. Clearly, $\rho_0 = 0$. For all $\tau \geq 0$, $\beta + \tau\rho \in B_2$, and therefore the inequality $(\beta + \tau\rho)^\intercal t \geq 0$ is valid for $\operatorname{cl}\operatorname{conv}(Q^0)$. Dividing throughout by τ and letting $\tau \to \infty$, we then conclude that $\rho^\intercal t \geq 0$, an inequality with $\rho_0 = 0$, is valid for $\operatorname{cl}\operatorname{conv}(Q^0)$. If this inequality is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$, then it must be one of the trivial ones. However, extreme rays, unlike extreme points, do not necessarily yield facet-defining inequalities for $\operatorname{cl}\operatorname{conv}(Q^0)$. We next illustrate these observations, together with the statement of Theorem 2.3.1.

Example 1 *(continued) For the set Q^0 with disjuncts defined by* (2.8) *and where variable t_5 has been removed, we compute that $w_{12} := \min\left\{\frac{3}{5}, \frac{1}{3}, \frac{6}{4}, \frac{2}{2}\right\} = \frac{1}{3}$, $w_{10} :=$*

$\min\left\{\frac{1}{5},\frac{1}{3},\frac{1}{4},\frac{1}{2}\right\}=\frac{1}{5}$, $w_{32}:=\min\left\{\frac{1}{2},\frac{6}{4}\right\}=\frac{1}{2}$, $w_{30}:=\min\left\{\frac{1}{2},\frac{1}{4}\right\}=\frac{1}{4}$, $w_{42}:=3$, *and* $w_{40}:=1$. *It then follows from Proposition 2.3.7 that*

$$B_2=\left\{(\beta_1,\beta_2,\beta_3,\beta_4,\beta_0)\in\mathbb{R}^5:\begin{array}{rcrcr}\beta_2 & + & \frac{1}{3}\beta_1 & \geq & 0\\ \beta_0 & + & \frac{1}{5}\beta_1 & \geq & 0\\ \beta_2 & + & \frac{1}{2}\beta_3 & \geq & 0\\ \beta_0 & + & \frac{1}{4}\beta_3 & \geq & 0\\ \beta_2 & + & 3\beta_4 & \geq & 0\\ \beta_0 & + & \beta_4 & \geq & 0\\ & & \beta_0 & = & -1\end{array}\right\}.$$

Coefficient vectors of all facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$ *that cut off the solution* $(0,0,0,0,1)$ *belong to* B_2. *For instance, the coefficient vector* $\beta=(5,-\frac{5}{3},4,1,-1)$ *belongs to* B_2. *Further, it satisfies the following system of linearly independent equations* $\beta_2+\frac{1}{3}\beta_1=0$, $\beta_0+\frac{1}{5}\beta_1=0$, $\beta_0+\frac{1}{4}\beta_3=0$, $\beta_0+\beta_4=0$, *and* $\beta_0=-1$. *Since the system has a unique solution,* β *is an extreme point of* B_2. *This extreme point is the coefficient vector of the first facet-defining inequality of* (2.9) *(where we have omitted the coefficient* β_5 *since* $\mathcal{I}_0=\{5\}$*). It can also be verified that* $(3,-1,2,\frac{1}{3},0)$ *is an extreme ray of* B_2. *It corresponds to the valid inequality* $3t_1-t_2+2t_3+\frac{1}{3}t_4\geq 0$, *which is not facet-defining for* $\operatorname{cl}\operatorname{conv}(Q^0)$ *since it can be obtained as a conic combination of the second facet-defining inequality of* (2.9) *and* $t_0\geq 0$ *with equal weights of* $\frac{1}{3}$.

2.4 Dual network formulation of B_2

In this section, we present a nonlinear transformation that maps (a subset of) the polyhedron B_2 to the feasible region of the dual of a transportation problem. We show that this transformation preserves the face-lattice of B_2 (see below for a definition.) We use these results in Section 2.5 to establish a correspondence between the extreme points of B_2, *i.e.*, the nontrivial facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$, and certain spanning trees of a suitably defined transportation network.

We have shown in Proposition 2.3.6 that if β is an extreme point of B_2, $\beta_i > 0$ for all $i \in \mathcal{I}_+$ and $\beta_j < 0$ for all $j \in \mathcal{I}_-$. Define $A = \{\beta \in \mathbb{R}^{|\mathcal{V}_0|} : \beta_i > 0, \beta_j < 0, \forall i \in \mathcal{I}_+, \forall j \in \mathcal{I}_-\}$. Observe that, for any $\beta \in B_2 \cap A$ and for $(i,j) \in \mathcal{I}_+ \times \mathcal{I}_-$,

$$\beta_j + w_{ij}\beta_i \geq 0 \quad \Longleftrightarrow \quad \frac{-\beta_j}{\beta_i} \leq w_{ij} \quad \Longleftrightarrow \quad \log(-\beta_j) - \log(\beta_i) \leq \log(w_{ij}).$$

All the logarithms computed above are well-defined under the conditions of A. Define $T : A \to \mathbb{R}^{|\mathcal{V}_0|}$ by

$$[T(\beta)]_k := \log|\beta_k| = \begin{cases} \log(\beta_k) & \text{if } k \in \mathcal{I}_+ \\ \log(-\beta_k) & \text{if } k \in \mathcal{I}_-. \end{cases}$$

Its inverse transformation T^{-1} is

$$[T^{-1}(\delta)]_k = \begin{cases} e^{\delta_k} & \text{if } k \in \mathcal{I}_+ \\ -e^{\delta_k} & \text{if } k \in \mathcal{I}_-. \end{cases}$$

After introducing the new variables $\delta_i = \log(\beta_i)$, for $i \in \mathcal{I}_+$ and $\delta_j = \log(-\beta_j)$, for $j \in \mathcal{I}_-$, and the constants $c_{ij} = \log(w_{ij})$, for $(i,j) \in \mathcal{I}_+ \times \mathcal{I}_-$, we define

$$\begin{aligned} D_1 &:= \left\{\delta \in \mathbb{R}^{|\mathcal{V}_0|} : \delta_j - \delta_i \leq c_{ij}, \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-\right\}, \\ D_2 &:= \left\{\delta \in \mathbb{R}^{|\mathcal{V}_0|} : \delta_j - \delta_i \leq c_{ij}, \delta_0 = 0, \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-\right\}. \end{aligned}$$

Proposition 2.4.1 *It holds that $T(B_2 \cap A) = D_2$.*

It is clear that for $\beta \in B_2 \cap A$ and $\delta = T(\beta) \in D_2$,

$$\beta_j + w_{ij}\beta_i = 0 \quad \Longleftrightarrow \quad \delta_j - \delta_i = c_{ij}, \tag{2.13}$$

$$\beta_j + w_{ij}\beta_i \leq 0 \quad \Longleftrightarrow \quad \delta_j - \delta_i \leq c_{ij}. \tag{2.14}$$

Let $H(E)$ be the subgraph of the complete bipartite graph $G := (\mathcal{I}_+, \mathcal{I}_-)$ with edge set $E \subseteq \mathcal{I}_+ \times \mathcal{I}_-$. Let $P, Q \in \mathbb{R}^{|\mathcal{I}_+| \times |\mathcal{I}_-|}$ be two matrices. We create the $|E| \times (n+1)$ matrix $M(H(E), P, Q)$ by fixing an ordering of the edges of E (say lexicographical) and by assigning the row of $M(H(E), P, Q)$ corresponding to edge $\{i,j\} \in E$ to be the vector $P_{ij}e_j^\intercal + Q_{ij}e_i^\intercal$.

Lemma 2 *Assume that $H(E)$ is a subforest of G. Assume also that $P_{ij} \neq 0$ and $Q_{ij} \neq 0$ for all $\{i,j\} \in E$. Then $M(H(E), P, Q)$ has full rank.*

Proof Suppose $H(E)$ is a subforest of G. Since $|E| < n+1$, we only need to prove independence of the rows of $M(H(E), P, Q)$. For a positive integer $k = 1, \ldots, |E|-1$, observe that the $(k+1)^{\text{th}}$ row of $M(H(E), P, Q)$ introduces a new nonzero entry, which was zero in the first k rows because $H(E)$ does not contain a cycle, $P_{ij} \neq 0$ and $Q_{ij} \neq 0$. This shows that the rows of $M(H(E), P, Q)$ are independent. ■

Define $\mathbb{J}$ to be the $|\mathcal{I}_+| \times |\mathcal{I}_-|$ matrix of ones and $\mathbb{W}$ to be the $|\mathcal{I}_+| \times |\mathcal{I}_-|$ matrix whose (i,j) entry is w_{ij}. For any $E \subseteq \mathcal{I}_+ \times \mathcal{I}_-$ such that $H(E)$ is a forest, Lemma 2 shows that both matrices $M(H(E), \mathbb{J}, -\mathbb{J})$ and $M(H(E), \mathbb{J}, \mathbb{W})$ have full rank.

Proposition 2.4.2 *Let $H(E)$ be a subgraph of G with $n = |\mathcal{V}_\mathbf{0}| - 1$ edges such that* $\text{rank}(M(H(E), \mathbb{J}, \mathbb{W})) = n$ *and $M(H(E), \mathbb{J}, \mathbb{W})\beta = 0$ for some $\beta \in B_2$. Then $H(E)$ is a tree of G.*

Proof Assume by contradiction that $H(E)$ has a cycle C, and let β^C be the components of β associated with nodes of C. Let M' be the $n \times n$ submatrix of $M(H(E), \mathbb{J}, \mathbb{W})$ associated with cycle C. Then it is easy to verify that M' is nonsingular and $M'\beta^C = 0$ which implies that $\beta^C = 0$. Since G is bipartite, C contains a node $k \in \mathcal{I}_+$, and so $\beta_k = 0$. Because $\beta \in B_2$, it satisfies $\beta_0 + w_{k0}\beta_k \geq 0$, which implies that $\beta_0 \geq 0$. This is a contradiction to the fact that $\beta_0 = -1$. Since $H(E)$ has n edges, $n+1$ nodes and no cycle, it is a tree.

■

A finite partially ordered set $(S, \leq)$, or *poset*, is the association of a finite set S with a relation "$\leq$" which is (i) reflexive: $x \leq x$ for all $x \in S$, (ii) transitive: $x \leq y$ and $y \leq z$ imply $x \leq z$, and (iii) antisymmetric: $x \leq y$ and $y \leq x$ imply $x = y$. The *face-lattice* of a polyhedron P is the poset of its faces, partially ordered by inclusion. We say that two posets $(S, \leq)$ and $(S', \preceq)$ are isomorphic if there is a bijection $T(\cdot)$ from S to S' such that $s_1 \leq s_2$ if and only if $T(s_1) \preceq T(s_2)$. Moreover, we say two *polyhedra are isomorphic* if their face-lattices are isomorphic.

Proposition 2.4.3 *Polyhedra D_2 and B_2 are isomorphic.*

Proof Given a polyhedron $P \subseteq \mathbb{R}^n$, we define $\mathcal{F}(P)$ to be the set of faces of P. Given $E \subseteq \mathcal{I}_+ \times \mathcal{I}_-$, we define

$$\begin{aligned} B_2|_E &= \{\beta \in B_2 \,|\, \beta_j + w_{ij}\beta_i = 0, \forall (i,j) \in E\} \\ D_2|_E &= \{\delta \in D_2 \,|\, \delta_j - \delta_i = c_{ij}, \forall (i,j) \in E\}. \end{aligned}$$

Clearly, $B_2|_E$ and $D_2|_E$ are (possibly empty) faces of B_2 and D_2, respectively. Given a nonempty face F of B_2, we denote by $\mathbb{E}(F)$ the largest subset $E \subseteq \mathcal{I}_+ \times \mathcal{I}_-$ such that $F = B_2|_E$. In particular, for every point β^* in the relative interior of F, $\beta^*_j + w_{ij}\beta^*_i < 0$ for $(i,j) \in (\mathcal{I}_+ \times \mathcal{I}_-)\backslash\mathbb{E}(F)$. Similarly, given a nonempty face F' of D_2, we denote by $\mathbb{E}'(F')$ the largest subset $E' \subseteq \mathcal{I}_+ \times \mathcal{I}_-$ such that $F' = D_2|_{E'}$. For every point δ^* in the relative interior of F', $\delta_j - \delta_i < c_{ij}$ for $(i,j) \in (\mathcal{I}_+ \times \mathcal{I}_-)\backslash\mathbb{E}'(F')$.

Next, we define $\varphi : \mathcal{F}(B_2) \mapsto \mathcal{F}(D_2)$ to be such that $\varphi(F) = D_2|_{\mathbb{E}(F)}$ for any nonempty face $F \in \mathcal{F}(B_2)$ and $\varphi(\emptyset) = \emptyset$. We show that φ is a bijection by constructing an inverse to φ. Define $\psi : \mathcal{F}(D_2) \mapsto \mathcal{F}(B_2)$ to be such that $\psi(F') = B_2|_{\mathbb{E}'(F')}$ for any nonempty face $F' \in \mathcal{F}(D_2)$ and $\psi(\emptyset) = \emptyset$. First, we argue that if $F \in \mathcal{F}(B_2)$ and $F' = \varphi(F)$, then $\mathbb{E}(F) = \mathbb{E}'(F')$. Consider a point $\bar{\beta}$ in the relative interior of F and an extreme point $\tilde{\beta}$ of that face. The line segment $[\bar{\beta}, \tilde{\beta})$ is in the relative interior of F; see Theorem 6.1 of [59]. Further, there exists a point β on this line segment, sufficiently close to $\tilde{\beta}$, that belongs to A. Define $\delta = T(\beta)$. It follows from (2.13) and (2.14) that δ belongs to the relative interior of F' and that $\mathbb{E}(F) = \mathbb{E}'(F')$. Similarly, if $F' \in \mathcal{F}(D_2)$ and $F'' = \psi(F')$, then $\mathbb{E}'(F') = \mathbb{E}(F'')$. Second, we argue that for each $F \in \mathcal{F}(B_2)$, $\psi(\varphi(F)) = F$. The result is clear when $F = \emptyset$. When $F \neq \emptyset$, define $F' = \varphi(F)$ and $F'' = \psi(F')$. It follows from the above discussion that $\mathbb{E}(F) = \mathbb{E}'(F') = \mathbb{E}(F'')$ Therefore $F'' = B_2|_{\mathbb{E}(F'')} = B_2|_{\mathbb{E}(F)} = F$.

To conclude the proof, consider two faces F_1 and F_2 of $\mathcal{F}(B_2)$ such that $F_1 \subseteq F_2$. Define $F'_1 = \varphi(F_1)$ and $F'_2 = \varphi(F_2)$. Since $F_1 \subseteq F_2$, then $\mathbb{E}(F_1) \supseteq \mathbb{E}(F_2)$. It follows from the above discussion that $\mathbb{E}'(F'_1) \supseteq \mathbb{E}'(F'_2)$, showing that $\varphi(F_1) = F'_1 \subseteq F'_2 = \varphi(F_2)$. ■

It is shown in Theorem 10.1 of [17] that two isomorphic polytopes have the same dimension, and that faces matched through the bijection $T(\cdot)$ have identical dimensions. The proof idea extends to our setting.

The proof of Proposition 2.4.3 shows that there is a one-to-one correspondence between the faces of dimension one of B_2 and D_2. We obtain

Corollary 3 *If u is an extreme point of D_2 then $T^{-1}(u)$ is an extreme point of B_2. Conversely, if v is an extreme point of B_2 then $T(v)$ is an extreme point of D_2.*

Extreme points of D_2 can be exposed as unique optimal solutions to certain linear programs (LPs) over D_2, or equivalently can be obtained from optimal solutions of certain LPs over D_1. In order for such LPs to have an optimal extreme point solution, the objective coefficient vector should be chosen in the polar cone of the recession cone of the feasible set. The recession cones of D_1 and D_2 are

$$\begin{aligned} \mathrm{rec}(D_1) &= \{\delta \in \mathbb{R}^{|\mathcal{V}_0|} : \delta_j - \delta_i \leq 0, \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-\}, \\ \mathrm{rec}(D_2) &= \{\delta \in \mathbb{R}^{|\mathcal{V}_0|} : \delta_j - \delta_i \leq 0, \delta_0 = 0, \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-\}. \end{aligned}$$

We next derive a $\mathcal{V}$-polyhedron description of $\mathrm{rec}(D_1)$ and $\mathrm{rec}(D_2)$. In this result, we let $\mathbb{1} = \sum_{k \in \mathcal{V}_0} e_k$. For a set of vectors V, we define $\mathrm{lin}(V)$ to be the linear subspace generated by V. For notational convenience, we write $\mathrm{lin}(v) = \mathrm{lin}(\{v\})$ for a vector v.

Proposition 2.4.4

1. *Let $R_2 = \{e_i : i \in \mathcal{I}_+\} \cup \{-e_j : j \in \mathcal{I}_- \setminus \{0\}\} \cup \{\mathbb{1} - e_0\}$. Then $\mathrm{rec}(D_2) = \mathrm{cone}(R_2)$.*

2. *Let $R_1 = \{e_i : i \in \mathcal{I}_+\} \cup \{-e_j : j \in \mathcal{I}_-\}$. Then $\mathrm{rec}(D_1) = \mathrm{cone}(R_1) + \mathrm{lin}(\mathbb{1})$.*

Proof Assume that $\delta \in \mathrm{rec}(D_2)$. Let $a = \min\{\delta_i : i \in \mathcal{I}_+\}$ and $b = \max\{\delta_j : j \in \mathcal{I}_-\}$. Then, $b \geq \delta_0 = 0$. Furthermore, $\delta_j \leq b \leq a \leq \delta_i$, for $i \in \mathcal{I}_+$ and $j \in \mathcal{I}_-$. We can then write

$$\delta = b(\mathbb{1} - e_0) + \sum_{j \in \mathcal{I}_- \setminus \{0\}} (b - \delta_j)(-e_j) + \sum_{i \in \mathcal{I}_+} (\delta_i - b) e_i,$$

which shows that $\text{rec}(D_2) \subseteq \text{cone}(R_2)$. Observe next that $R_2 \subseteq \text{rec}(D_2)$ since the elements of R_2 are rays of D_2. Therefore $\text{cone}(R_2) \subseteq \text{rec}(D_2)$, proving 1. We next show that $\text{rec}(D_1) = \text{rec}(D_2) + \text{lin}(\mathbb{1})$, which will prove 2 since $\mathbb{1} - e_0 \in -e_0 + \text{lin}(\mathbb{1})$. To prove the forward inclusion ($\subseteq$), consider δ' in $\text{rec}(D_1)$. Then $\delta' - \delta'_0\mathbb{1}$ belongs to $\text{rec}(D_2)$. To prove the reverse inclusion ($\supseteq$), consider δ' in $\text{rec}(D_2)$ and $t \in \mathbb{R}$. It is clear that $\delta' + t\mathbb{1} \in \text{rec}(D_1)$. ■

By definition of polar cone,

$$\begin{aligned}
(\text{rec}(D_1))^o &:= \{y : y^\intercal x \leq 0, x \in \text{rec}(D_1)\} \\
&= \{y : y^\intercal x \leq 0, x \in R_1 \cup \{-\mathbb{1}, \mathbb{1}\}\} \\
&= \left\{ y : \begin{array}{ll} y^\intercal e_i \leq 0, & \forall i \in \mathcal{I}_+ \\ y^\intercal(-e_j) \leq 0, & \forall j \in \mathcal{I}_- \\ y^\intercal \mathbb{1} = 0 & \end{array} \right\} \\
&= \left\{ y : \begin{array}{ll} y_i \leq 0, & \forall i \in \mathcal{I}_+ \\ y_j \geq 0, & \forall j \in \mathcal{I}_- \\ \sum_{k \in \mathcal{V}_0} y_k = 0 & \end{array} \right\}.
\end{aligned}$$

Similar to B_2, it is simple to verify that D_2 does not contain lines, and therefore has at least one extreme point. We next show that each extreme point of D_2 can be derived from an optimal solution of an LP over D_1 by setting an appropriate objective vector y in $\text{ri}\,(\text{rec}(D_1)^o)$. Define $\mathbf{s} := -y_{\mathcal{I}_+}$ and $\mathbf{d} := y_{\mathcal{I}_-}$. The desired LP is

$$\begin{aligned}
\max \quad & -\sum_{i \in \mathcal{I}_+} \mathbf{s}_i \delta_i + \sum_{j \in \mathcal{I}_-} \mathbf{d}_j \delta_j \\
\text{s.t.} \quad & \delta_j - \delta_i \leq c_{ij}, \qquad \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-.
\end{aligned} \tag{2.15}$$

Its dual is the transportation problem:

$$\begin{aligned}
\min \quad & \sum_{i \in \mathcal{I}_+} \sum_{j \in \mathcal{I}_-} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{j \in \mathcal{I}_-} x_{ij} = \mathbf{s}_i, \quad \forall i \in \mathcal{I}_+, \\
& \sum_{i \in \mathcal{I}_+} x_{ij} = \mathbf{d}_j, \quad \forall j \in \mathcal{I}_-, \\
& x_{ij} \geq 0 \qquad \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-.
\end{aligned} \tag{2.16}$$

We next argue that both primal (2.16) and dual (2.15) problems are feasible, thereby showing that optimal primal and dual solutions exist. The primal problem (2.16) is feasible because $\sum_{j\in\mathcal{I}_-}\mathbf{d}_j = \sum_{i\in\mathcal{I}_+}\mathbf{s}_i$. The c-max cut shows that the dual problem (2.16) is feasible. In fact, let c$-$max be the coefficient vector of the c-max cut. This vector is in B_2. Furthermore, $(\text{c}-\text{max})_i > 0$ for $i \in \mathcal{I}_+$ and $(\text{c}-\text{max})_j < 0$ for $j \in \mathcal{I}_-$. Therefore, $(\text{c}-\text{max}) \in B_2 \cap A$. It follows that $T(\text{c}-\text{max}) \in D_2 \subseteq D_1$. The fact that D_1 is nonempty also follows from Proposition 2.4.3.

Proposition 2.4.4 shows that D_1 has a lineality. It follows that the faces of D_1 of smallest dimension are edges. Because (2.15) has an optimal solution, it must therefore be that it has an edge of optimal solutions. Let δ' be a solution on this edge. There are n active constraints of D_1 at δ'. Now define $\delta^* = \delta' - \delta'_0\mathbb{1}$. Then,

$$\begin{aligned}
-\sum_{i\in\mathcal{I}_+}\mathbf{s}_i\delta_i^* + \sum_{j\in\mathcal{I}_-}\mathbf{d}_j\delta_j^* &= -\sum_{i\in\mathcal{I}_+}\mathbf{s}_i(\delta'_i - \delta_0) + \sum_{j\in\mathcal{I}_-}\mathbf{d}_j(\delta'_j - \delta_0) \\
&= -\sum_{i\in\mathcal{I}_+}\mathbf{s}_i\delta'_i + \sum_{j\in\mathcal{I}_-}\mathbf{d}_j\delta'_j + \delta_0(\sum_{i\in\mathcal{I}_+}\mathbf{s}_i - \sum_{j\in\mathcal{I}_-}\mathbf{d}_j) \\
&= -\sum_{i\in\mathcal{I}_+}\mathbf{s}_i\delta'_i + \sum_{j\in\mathcal{I}_-}\mathbf{d}_j\delta'_j.
\end{aligned}$$

Hence δ^* has the same objective function value as δ'. Moreover, δ^* satisfies all the constraints in (2.15) because $\delta_j^* - \delta_i^* = (\delta'_j - \delta'_0) - (\delta'_i - \delta'_0) = \delta'_j - \delta'_i \leq c_{ij}$ for all $(i,j) \in \mathcal{I}_+ \times \mathcal{I}_-$. Clearly, δ^* is an extreme point of D_2 since it satisfies $\delta_0^* = 0$ in addition to the n independent constraints active at δ'. Proposition 2.4.3 then implies that $\beta^* = T^{-1}(\delta^*)$ is an extreme point of B_2, *i.e.*, the coefficient vector of a facet-defining inequality for $\operatorname{cl}\operatorname{conv}(Q^0)$ that cuts off $(t_1,\ldots,t_n,t_0) = (0,\ldots,0,1)$.

Because basic feasible solutions of (2.16) correspond to certain spanning trees of G, it is natural to suspect that facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$ can be associated to those spanning trees. We explore this correspondence in the following section.

2.5 Label-connected trees and facet-defining inequalities of $\operatorname{cl}\operatorname{conv}(Q^0)$

In this section, we show that facet-defining inequalities for $\operatorname{cl}\operatorname{conv}(Q^0)$ correspond to certain subtrees of the complete undirected bipartite graph $G = (\mathcal{I}_+, \mathcal{I}_-)$. Recall that, in (2.11), we associated a *weight* w_{ij} to each arc $\{i, j\}$ where $i \in \mathcal{I}_+$ and $j \in \mathcal{I}_-$. To streamline notation, we define $w_{ji} := \frac{1}{w_{ij}}$ for $(i, j) \in \mathcal{I}_+ \times \mathcal{I}_-$.

Consider a spanning tree S of G. Then for any node $i \in \mathcal{I}_+ \cup \mathcal{I}_- \setminus \{0\}$ there exists a unique path from node 0 to node i in S. We denote this path P_{0i} by

$$(0 =)i_0 - i_1 - i_2 - \cdots - i_p(= i).$$

We say that an inequality $\beta^\intercal t \geq 0$ (or the associated coefficient vector β) is *induced* by the spanning tree S if

$$\beta_i := (-1)^p \beta_0 w_{i_0 i_1} w_{i_1 i_2} \ldots w_{i_{p-1} i_p}$$

for each $i \in \mathcal{V}$. It follows directly from the definition of induced inequality that, on path P_{0i}, if $0 \leq q < r \leq p$

$$\beta_{i_r} \quad = \quad (-1)^{r-q} \beta_{i_q} w_{i_q i_{q+1}} w_{i_{q+1} i_{q+2}} \ldots w_{i_{r-1} i_r}. \tag{2.17}$$

In particular, if two distinct spanning trees S and S' of G share the same path from node i_q to node i_r, then it follows from (2.17) that

$$\frac{\beta^S_{i_r}}{\beta^S_{i_q}} = \frac{\beta^{S'}_{i_r}}{\beta^{S'}_{i_q}} \tag{2.18}$$

where β^S and $\beta^{S'}$ represent the coefficient vectors induced by spanning trees S and S', respectively.

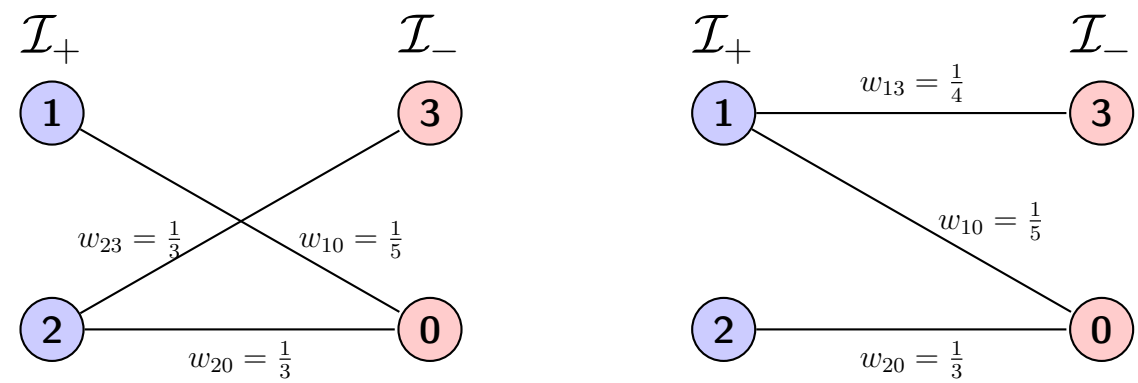
We will show in Proposition 2.5.1 that every facet-defining inequality is induced by a spanning tree of G. However, not all spanning trees of G induce valid inequalities, as we illustrate in the following example.

Example 2 *Consider the set Q^0 with disjuncts defined by the following inequalities*

$$\begin{array}{rrrrl} 4t_1 & +3t_2 & -t_3 & -t_0 & \geq 0 \\ 5t_1 & +t_2 & -2t_3 & -t_0 & \geq 0 \\ 5t_1 & +2t_2 & -2t_3 & -t_0 & \geq 0. \end{array}$$

We have that $\mathcal{I}_+ = \{1, 2\}$ *and* $\mathcal{I}_- = \{3, 0\}$. *Further, edge weights can be computed to be* $w_{13} = \frac{1}{4}$,$w_{10} = \frac{1}{5}$, $w_{23} = \frac{1}{3}$, $w_{20} = \frac{1}{3}$.

Two spanning trees of G *are shown in Figure 2.1. The inequality induced by the subtree of Figure 2.1(a) is* $5t_1 + 3t_2 - t_3 - t_0 \geq 0$. *This inequality is the c-max cut and, hence, is valid for* $\operatorname{cl}\operatorname{conv}(Q^0)$. *Furthermore, it can be verified to be facet-defining for this set. The inequality induced by the subtree of Figure 2.1(b) is* $5t_1 + 3t_2 - \frac{5}{4}t_3 - t_0 \geq 0$, *which is not valid because it cuts off the feasible point* $(0, 1, 3, 0) \in Q^0_1 \subseteq \operatorname{cl}\operatorname{conv}(Q^0)$.



(a) Tree inducing $5t_1 + 3t_2 - t_3 - t_0 \geq 0$

(b) Tree inducing $5t_1 + 3t_2 - \frac{5}{4}t_3 - t_0 \geq 0$

Figure 2.1.: Spanning trees and induced inequalities for Example 2

Example 2 shows that not all spanning trees of G induce a valid inequality. The reason is that the induced coefficients may violate an inequality corresponding to an edge that is not included in the spanning tree. We refer to a spanning tree that induces a valid inequality as a *feasible spanning tree*. We next show that any inequality induced by a feasible spanning tree is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$.

Proposition 2.5.1 *Inequality* $\beta^\intercal t \geq 0$ *with* $\beta_0 = -1$ *is facet-defining for* $\operatorname{cl}\operatorname{conv}(Q^0)$ *if and only if* β *is induced by a feasible spanning tree of* G.

Proof Let $\beta^\intercal t \geq 0$ with $\beta_0 = -1$ be a facet-defining inequality for $\operatorname{cl}\operatorname{conv}(Q^0)$. Then, by Theorem 2.3.1, β is an extreme point of B_2. Since β is an extreme point

of B_2, it belongs to $n = |\mathcal{V}_\mathbf{0}| - 1$ hyperplanes of the form $\{\beta \in \mathbb{R}^{|\mathcal{V}_\mathbf{0}|} : \beta_j + w_{ij}\beta_i = 0\}$ whose coefficient vectors are linearly independent, in addition to $\beta_0 = -1$. By Proposition 2.4.2, the subgraph with respect to β forms a spanning tree of G.

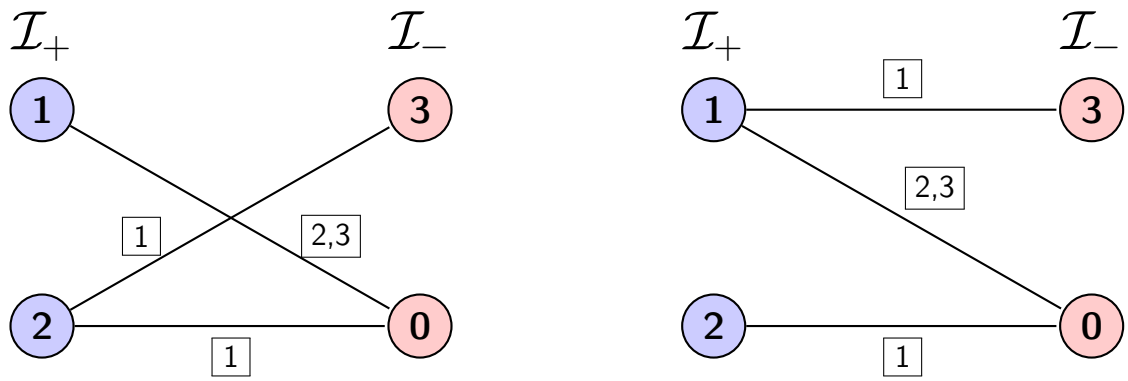
For the converse, suppose $\beta^\intercal t \geq 0$ with $\beta_0 = -1$ is induced by a feasible spanning tree. The validity of $\beta^\intercal t \geq 0$ follows directly from the definition of a feasible spanning tree. By construction, see (2.17), coefficients β satisfy n equations of the form $\beta_j + w_{ij}\beta_i = 0$, one for each edge of the tree. Lemma 2 shows that these n coefficient vectors are independent. Therefore, β is an extreme point of B_2. Hence, Theorem 2.3.1 implies that $\beta^\intercal t \geq 0$ is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$. ■

We next introduce the notion of *label-connectivity*. Let S be a spanning tree of G with edge set $E \subseteq \mathcal{I}_+ \times \mathcal{I}_-$. A function $L : E \to \mathcal{L}$ is called a *label-function* if

$$L(\{i,j\}) \in \left\{ l \in \mathcal{L} : f_{li} > 0, -\frac{f_{lj}}{f_{li}} = w_{ij} \right\}$$

for each $\{i,j\} \in E$. In words, $L(\{i,j\})$ returns the index l of an inequality in the description of Q^0 with $f_{li} > 0$ and the property that the ratio of the coefficient of t_j over that of t_i equals $-w_{ij}$. Because the ratio w_{ij} might be achieved in different rows, several label-functions might be associated with a single spanning tree. For this reason, we define the set of all the label-functions of spanning tree S by $\mathbb{L}(S)$. We write $S(E,L)$ to refer to a specific spanning tree with edge set E and label-function L. We say there is a *label-disconnection* for label l in $S(E,L)$ if the subgraph of $S(E,L)$ induced by the edges of label l is disconnected. It is easily seen that this definition is equivalent to stating that there exists a path in $S(E,L)$ where two edges with label l are connected within the tree using a path whose edges do not have label l. Finally, we say that a spanning tree S with edge set E is *label-connected* if there exists a label-function $L \in \mathbb{L}(S)$ such that $S(E,L)$ does not exhibit label-disconnection for any $l \in \mathcal{L}$. Otherwise it is *label-disconnected*.

Example 2 (continued) *In Figure 2.2, we add all possible valid edge labels to the edges of the spanning trees presented in Figure 2.1. In Figure 2.2(a), we observe that*



(a) Tree inducing $5t_1 + 3t_2 - t_3 - t_0 \geq 0$

(b) Tree inducing $5t_1 + 3t_2 - \frac{5}{4}t_3 - t_0 \geq 0$

Figure 2.2.: Possible edge labels for two spanning trees of Example 2

there are two possible labels for edge $\{1, 0\}$*, each of which determines that* $w_{10} = \frac{1}{5}$*. We see that, independent of the choice of label for edge* $\{1, 0\}$*, the spanning tree does not exhibit any label-disconnection. It is therefore label-connected. In Figure 2.2(b), we observe that independent of the choice of label for edge* $\{1, 0\}$*, the spanning tree will exhibit a label-disconnection for label 1 along the path* $3 - 1 - 0 - 2$*. We conclude that this spanning tree is label-disconnected.*

Label-connected spanning trees do not necessarily induce valid inequalities and not all feasible trees that induce a facet-defining inequality are label-connected. However, we show next via an example and later prove that, for facet-defining inequalities, there exists a feasible spanning tree that is label-connected.

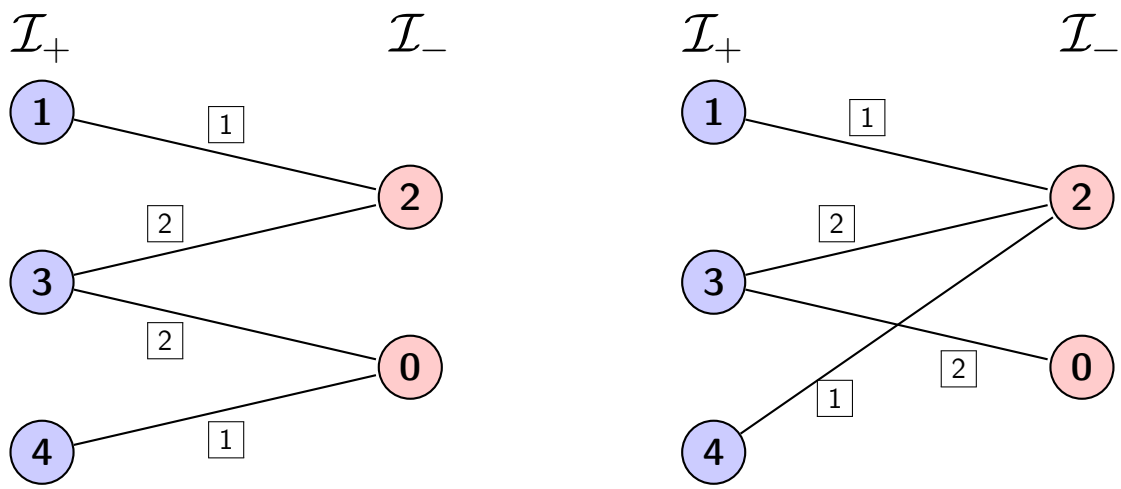
Example 3 *Consider the set* Q^0 *with disjuncts defined by the following inequalities*

$$\begin{array}{rrrrrl} \frac{25}{4}t_1 & -\frac{5}{2}t_2 & +\frac{5}{16}t_3 & +\frac{15}{4}t_4 & -t_0 & \geq 0 \\ 5t_1 & -\frac{5}{2}t_2 & +t_3 & +\frac{7}{2}t_4 & -t_0 & \geq 0. \end{array}$$

For this set, we have that $\mathcal{I}_+ = \{1, 3, 4\}$ *and* $\mathcal{I}_- = \{2, 0\}$*. We compute that* $w_{10} = \frac{4}{25}$*,* $w_{30} = 1$*,* $w_{40} = \frac{4}{15}$*,* $w_{12} = \frac{2}{5}$*,* $w_{32} = \frac{5}{2}$ *and* $w_{42} = \frac{2}{3}$*. Corresponding edge labels are* $l_{10} = 1$*,* $l_{30} = 2$*,* $l_{40} = 1$*,* $l_{12} = 1$*,* $l_{32} = 2$*, and* $l_{42} = 1$*. The spanning tree of*

Figure 2.3(a) is label-disconnected. The spanning tree of Figure 2.3(b), however, is label-connected. These spanning trees both induce the facet-defining inequality

$$\tfrac{1}{4}(25t_1 - 10t_2 + 4t_3 + 15t_4 - 4t_0) \geq 0. \tag{2.19}$$



(a) Label-disconnected spanning tree

(b) Label-connected spanning tree

Figure 2.3.: Two spanning trees inducing (2.19) in Example 3

Lemma 3 *Consider a facet-defining inequality induced by a spanning tree S for which there is a label-disconnection for label l. Let C_1 and C_2 be any two distinct components in the subgraph induced by edges with label l. Then, there exists a non-empty subtree of C_2 that can be detached from C_2 and attached to C_1, using an edge with label l, without changing the rest of the tree or the corresponding facet-defining inequality.*

Proof Since the given facet-defining inequality is induced by a spanning tree, there exists a unique path from a node in C_1 to a node in C_2 that contains no edge from C_1 or C_2. Let the starting node be $i_1 \in C_1$ and the ending node be $j_1 \in C_2$. Further, let i_2 be a neighbor of i_1 in C_1, and j_2 be a neighbor of j_1 in C_2. Let $i' \in \mathcal{I}_+ \cap \{i_1, i_2\}, j' \in \{i_1, i_2\} \setminus \{i'\}$ and let $i'' \in \mathcal{I}_+ \cap \{j_1, j_2\}, j'' \in \{j_1, j_2\} \setminus \{i''\}$. Since edges (i', j') and (i'', j'') have label l, it follows that

$$\beta_{j'} = \beta_{i'} \frac{f_{lj'}}{f_{li'}}, \text{ and } \beta_{j''} = \beta_{i''} \frac{f_{lj''}}{f_{li''}}. \tag{2.20}$$

Further, since the spanning tree yields a valid inequality,

$$\beta_{i''}\frac{f_{lj'}}{f_{li''}} \leq \beta_{j'}, \text{ and } \beta_{i'}\frac{f_{lj''}}{f_{li'}} \leq \beta_{j''}. \tag{2.21}$$

We write

$$\beta_{i''}\frac{f_{lj'}}{f_{li''}} \leq \beta_{j'} = \beta_{i'}\frac{f_{lj'}}{f_{li'}} = \beta_{i'}\frac{f_{lj''}}{f_{li'}}\frac{f_{lj'}}{f_{lj''}} \leq \beta_{j''}\frac{f_{lj'}}{f_{lj''}} = \beta_{i''}\frac{f_{lj'}}{f_{li''}}, \tag{2.22}$$

where the inequalities hold because of (2.21) and the equalities holds because of (2.20). Therefore, equality holds throughout and

$$\beta_{j''} = \beta_{i'}\frac{f_{lj''}}{f_{li'}}, \text{ and } \beta_{j'} = \beta_{i''}\frac{f_{lj'}}{f_{li''}}. \tag{2.23}$$

Now create a new spanning tree by deleting arc (j_1, j_2) from S and by connecting j_2 to the one node among i_1 and i_2 that belongs to the other partition of the bipartite graph. Call this node k and refer to the resulting spanning tree as S'. Clearly S' contains a label-connected component for label l that subsumes C_1 and has at least one more arc. Further, the label of both the edge added and the edge removed is l, while all other edges and their labels remain unchanged. For any node i, β_i is obtained by taking products of $-w_{i'j'}$ for edges $\langle i', j'\rangle$ along the path from 0 assuming $\beta_0 = -1$. We split this path into three parts from 0 to $\bar{i}$, $\bar{i}$ to $\bar{j}$, and $\bar{j}$ to i, where $\bar{i}$ (resp. $\bar{j}$) is the first (resp. last) of the nodes $\{i_1, i_2, j_1, j_2\}$ encountered along this path. Since the arcs from 0 to $\bar{i}$ remain untouched and so do the arcs from $\bar{j}$ to i, the ratios $\frac{\beta_{\bar{i}}}{\beta_0}$ and $\frac{\beta_i}{\beta_{\bar{j}}}$ are preserved. We have already shown that the tree preserves $\frac{\beta_{\bar{j}}}{\beta_{\bar{i}}}$. Taking a product, we see that β_i is preserved. ■

Example 3 (continued) *We have seen that the spanning tree of Figure 2.3(a) is feasible, but is label-disconnected. Label-1 disconnection occurs on the path* $1-2-3-0-4$, *as* $L(\{1,2\}) = L(\{4,0\}) = 1$ *and* $L(\{3,2\}) = L(\{3,0\}) = 2$. *Consider edge* $\{4,2\}$. *It is shown in Example 3 that* $L(\{4,2\}) = 1$. *Replacing edge* $\{4,0\}$ *with* $\{4,2\}$ *in the spanning tree does not change the induced inequality and yields the label-connected spanning tree shown in Figure 2.3(b).*

Theorem 2.5.1 *Let $\beta^\intercal t \geq 0$ be a non-trivial facet-defining inequality for* $\operatorname{cl}\operatorname{conv}(Q^0)$*. Then, there exists a label-connected feasible spanning tree that induces it.*

Proof If $\beta^\intercal t \geq 0$ is a nontrivial facet-defining inequality, Proposition 2.5.1 shows that it is induced by a feasible spanning tree. We prove the existence of a label-connected feasible spanning tree by contradiction. Let $\mathcal{T}$ be the set of all feasible spanning trees that induce this inequality. Note that $\mathcal{T} \neq \emptyset$, $\mathcal{T}$ is a finite set, and each tree in $\mathcal{T}$ is disconnected for some label. For any tree $T \in \mathcal{T}$, let $l(T)$ be the smallest label index for which it exhibits disconnection. Let $l' = \max\{l(T) : T \in \mathcal{T}\}$ and let $C(T, l)$ be the size of the largest connected component of label l in T. Choose

$$T' \in \operatorname{Arg\,max}\{C(T, l') : T \in \mathcal{T}, l(T) = l'\}.$$

Using Lemma 3, we can construct T'' from T' by choosing C_1 as a component of size $C(T', l')$. Since T'' is obtained without altering labels on any arc with labels other than l', labels that were previously connected remain connected. Further, T'' has a connected component for label l' of size larger than $C(T', l')$. The existence of T'' contradicts the definition of T', proving that there must exist a label-connected feasible spanning tree in $\mathcal{T}$. ■

Example 4 *Consider the set Q^0 defined in Example 1, where variable t_5 has been omitted. We record all spanning trees of $G(\mathcal{I}_+, \mathcal{I}_-)$ in Table 2.1. In particular, the columns of Table 2.1 contain the edges of each spanning tree, the coefficient β this spanning tree induces, and, in the case where the tree is infeasible, one edge that β violates. We conclude that* $\operatorname{cl}\operatorname{conv}(Q^0)$ *has only three nontrivial facet-defining inequalities, which were previously listed in* (2.9)*. It can be easily verified that the three feasible spanning trees are label-connected.*

Table 2.1: Feasible spanning trees for Example 4

	Edge 1	Edge 2	Edge 3	Edge 4	β	Violated Edge
Tree 1	$(1,2)$	$(1,0)$	$(3,2)$	$(4,2)$	$1/9(45,-15,30,5,-9)$	$(3,0)$
Tree 2	$(1,2)$	$(1,0)$	$(3,0)$	$(4,2)$	$1/9(45,-15,36,5,-9)$	$(4,0)$
Tree 3	$(1,2)$	$(3,2)$	$(3,0)$	$(4,2)$	$1/3(18,-6,12,2,-3)$	$(4,0)$
Tree 4	$(1,0)$	$(3,2)$	$(3,0)$	$(4,2)$	$1/3(15,-6,12,2,-3)$	$(4,0)$
Tree 5	$(1,2)$	$(1,0)$	$(3,2)$	$(4,0)$	$1/3(15,-5,10,3,-3)$	$(3,0)$
Tree 6	$(1,2)$	$(1,0)$	$(3,0)$	$(4,0)$	$1/3(15,-5,12,3,-3)$	–
Tree 7	$(1,2)$	$(3,2)$	$(3,0)$	$(4,0)$	$(6,-2,4,1,-1)$	–
Tree 8	$(1,0)$	$(3,2)$	$(3,0)$	$(4,0)$	$(5,-2,4,1,-1)$	$(1,2)$
Tree 9	$(1,2)$	$(3,2)$	$(4,2)$	$(4,0)$	$(9,-3,6,1,-1)$	–
Tree 10	$(1,0)$	$(3,2)$	$(4,2)$	$(4,0)$	$(5,-3,6,1,-1)$	$(1,2)$
Tree 11	$(1,2)$	$(3,0)$	$(4,2)$	$(4,0)$	$(9,-3,4,1,-1)$	$(3,2)$
Tree 12	$(1,0)$	$(3,0)$	$(4,2)$	$(4,0)$	$(5,-3,4,1,-1)$	$(3,2)$

2.6 Generalized Equate-and-Relax procedure for CCLPs

The Equate-and-Relax (E&R) procedure was recently proposed in [56] to construct the closed convex hull of

$$Q^c = \{x \in \mathbb{R}^n : a^\intercal x \geq 1, x \geq 0\} \cup \{x \in \mathbb{R}^n : b^\intercal x \geq 1, x \geq 0\},$$

where $a, b \in \mathbb{R}^n$, $a \not\leq 0$ and $b \not\leq 0$. Set Q^c occurs when relaxing, in a manner similar to that used here, a simplex tableau associated with the LP relaxation of a linear program with complementarity constraints.

The E&R procedure generates valid inequalities for $\operatorname{cl}\operatorname{conv}(Q^c)$. It has two steps. In the E-step, either the right-hand-side, or a variable x_i whose coefficients a_i and b_i are of the same sign is chosen. The nontrivial disjunct constraints $a^\intercal x \geq 1$ and $b^\intercal x \geq 1$ are then multiplied by suitable nonnegative scalars α and γ so that their right-hand-sides or the coefficients of variable x_i become equal, *i.e.*, $\alpha = \gamma$ or $\alpha a_i = \gamma b_i$. In the R-step, a valid inequality is created by setting the coefficient of each variable to be the maximum of its coefficients in the scaled disjunct inequalities. The right-hand-side of the inequality is set to the minimum of the right-hand-sides of the scaled inequalities. More precisely, the valid inequality produced is

$$\sum_{i=1}^{n} \max\{\alpha a_i, \gamma b_i\} x_i \geq \min\{\alpha, \gamma\}. \tag{2.24}$$

When $\alpha = \gamma > 0$, (2.24) is the c-max cut described in Section 2.2. It is shown in [56] that the family of E&R cuts characterizes $\operatorname{cl}\operatorname{conv}(Q^c)$.

The E&R result can be seen as a tightening of classical results in disjunctive programming. The fact that, for all $\alpha \geq 0$ and $\gamma \geq 0$, (2.24) is valid for $\operatorname{cl}\operatorname{conv}(Q^c)$ follows from [4]. The fact that every facet-defining inequality of $\operatorname{cl}\operatorname{conv}(Q^c)$ is of the form (2.24) for some α and γ also follows from LP duality. The requirement that $\alpha = \gamma$ or $\alpha a_i = \gamma b_i$ for some i, which is not explicit in traditional disjunctive programming constructs, allows the set of multipliers to be restricted to a finite collection. Although they collectively describe $\operatorname{cl}\operatorname{conv}(Q^c)$, not all E&R inequalities are facet-defining for $\operatorname{cl}\operatorname{conv}(Q^c)$. In [56], a partial characterization of when E&R

inequalities are facet-defining for complementarity problems is provided. Even for the c-max cut, no precise characterization of when it is facet-defining is available.

In this section, we provide a precise characterization of when an E&R inequality is facet-defining for $\operatorname{cl}\operatorname{conv}(Q^c)$. En route, we generalize E&R to the cardinality setting. We also describe a low-order polynomial time algorithm to strengthen a given inequality so that it becomes facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$.

Recall that we are interested in $Q^0 = \bigcup_{l\in\mathcal{L}} Q^0_l$ with $|\mathcal{L}| = K+1$ where $Q^0_l = \{t \in \mathbb{R}^{|\mathcal{V}_0|} : f_l^\intercal t \geq 0, t \geq 0\}$ and $f_{l0} = -1$ for all $l \in \mathcal{L}$. For multipliers $u_l \geq 0$, where $l \in \mathcal{L}$, we derive the following valid inequality

$$\sum_{i\in\mathcal{V}} \max_{l\in\mathcal{L}} \{u_l f_{li}\}\, t_i \geq 0. \tag{2.25}$$

It follows from [4] that the collection of inequalities of the form (2.25) characterizes $\operatorname{cl}\operatorname{conv}(Q^0)$. We show next that it is sufficient to consider weights associated with feasible label-connected spanning trees. We first illustrate the result on an example.

Example 5 *Consider the set Q^0 defined in Example 1, where variable t_5 has been omitted. Using multipliers $\left(1, \frac{5}{3}, 1, 1\right)$, we obtain*

$$5t_1 - \frac{5}{3}t_2 + 4t_3 + t_4 - t_0 \geq 0, \tag{2.26}$$

an inequality that is facet-defining inequality for $\operatorname{cl}\operatorname{conv}(Q^0)$*; see* (2.9)*a.*

Given a nontrivial facet-defining inequality, we next describe how to derive it using (2.25) by computing the appropriate multipliers u_l for $l \in \mathcal{L}$. If $\beta^\intercal t \geq 0$ is a nontrivial facet-defining inequality of $\operatorname{cl}\operatorname{conv}(Q^0)$, then by Theorem 2.5.1 there exists a label-connected feasible spanning tree T that induces it.

Consider 0 as the root node for the spanning tree T. For each label $l \in \mathcal{L}$ that appears in the tree, let $\jmath_l$ be the node with smallest distance (measured in number of arcs) from 0 among all nodes incident to an arc with label l. Let $\{\ell_n\}_{n=1,\ldots,r}$ be

the sequence of distinct labels encountered on the path from 0 to $\dot{\lambda}_l$ and let ℓ_{r+1} be l. Let

$$u_l = \begin{cases} \prod_{j=2}^{r+1} \frac{f_{\ell_{j-1}\dot{\lambda}_{\ell_j}}}{f_{\ell_j\dot{\lambda}_{\ell_j}}} & \text{if } r \geq 1 \\ 1 & \text{if } r = 0. \end{cases} \tag{2.27}$$

For labels l that do not appear in the spanning tree T, let

$$u_l \in \left[\max\left\{\frac{\beta_j}{f_{lj}} : j \in \mathcal{I}_-\right\}, \min\left\{\frac{\beta_i}{f_{li}} : i \in \mathcal{I}_+, f_{li} > 0\right\}\right]. \tag{2.28}$$

This procedure can be intuitively explained as follows. For each $l \in \mathcal{L}$, we obtain the subgraph induced by all arcs with label l. This subgraph is a (possibly empty) tree because T is label-connected. We refer to it as S_l and to its node set as $N(S_l)$. If S_l is empty, then the constraint of disjunct l does not play an active role in the derivation of the inequality. If S_l is not empty, the valid inequality produced is such that the coefficients of variables t_i for $i \in N(S_l)$ are a common multiple u_l of their coefficients in the nontrivial constraint of disjunct l. This multiple u_l is chosen to be 1 if $0 \in N(S_l)$. Otherwise u_l is computed so that the scaled coefficients in disjuncts l and ℓ_r of the variable $t_{\dot{\lambda}_l}$ are equal. In the complementarity case, where $|\mathcal{L}| = 2$, this procedure boils down to aggregating scaled constraints using (2.24) such that either the right-hand-sides match or one of the variables has the same coefficient in both constraints.

Proposition 2.6.1 *Any nontrivial facet-defining inequality of* $\operatorname{cl}\operatorname{conv}(Q^0)$ *induced by a feasible label-connected spanning tree* T *can be expressed as* (2.25) *by selecting weights* u_l *for* $l \in \mathcal{L}$ *as in* (2.27) *and* (2.28).

Proof First, we argue that weights u_l are well-defined for $l \in \mathcal{L}$. For labels l that appear in T, weights are uniquely defined by (2.27) since label-connectedness implies that $\dot{\lambda}_l$ is uniquely defined. For labels l that do not appear in T, the interval described in (2.28) is nonempty because T is feasible and therefore $f_{li}\beta_j - f_{lj}\beta_i \geq 0$ implies that $\frac{\beta_i}{f_{li}} \geq \frac{\beta_j}{f_{lj}}$ for all $i \in \mathcal{I}_+$ with $f_{li} > 0$ and $j \in \mathcal{I}_-$.

Second, we show that the given weights are nonnegative. When l does not appear in T, u_l is chosen according to (2.28). The lower bound of this interval is positive, proving the claim. Assume therefore that label l appears in the tree T. Let k be a node incident to an arc of label l. Assume that the path from 0 to k is $(k_0(=0), \ldots, k_s, k)$ with sequence of labels $\{\ell'_n\}_{n=1,\ldots,s+1}$ and sequence of distinct labels $\{\ell_n\}_{n=1,\ldots,r}$ where $\ell_{r+1} = l$. Note that the node associating ℓ_n and ℓ_{n+1} for $n = 1, \ldots, r$ is $\lambda_{\ell_{n+1}}$ and hence

$$\begin{aligned}
\beta_k &= (-1)^{s+1} w_{k_0,k_1} \ldots w_{k_s,k} \beta_0 \\
&= \frac{f_{\ell'_1 k_1}}{f_{\ell'_1 k_0}} \cdots \frac{f_{\ell'_s k_s}}{f_{\ell'_s k_{s-1}}} \frac{f_{\ell'_{s+1} k}}{f_{\ell'_{s+1} k_s}} \beta_0 \\
&= \frac{f_{\ell_1 \lambda_{\ell_2}}}{f_{\ell_1 \lambda_{\ell_1}}} \frac{f_{\ell_2 \lambda_{\ell_3}}}{f_{\ell_2 \lambda_{\ell_2}}} \cdots \frac{f_{\ell_r \lambda_{\ell_{r+1}}}}{f_{\ell_r \lambda_{\ell_r}}} \frac{f_{\ell_{r+1} k}}{f_{\ell_{r+1} \lambda_{\ell_{r+1}}}} \beta_0 \\
&= \frac{f_{\ell_1 \lambda_{\ell_2}}}{-1} \frac{f_{\ell_2 \lambda_{\ell_3}}}{f_{\ell_2 \lambda_{\ell_2}}} \cdots \frac{f_{\ell_r \lambda_{\ell_{r+1}}}}{f_{\ell_r \lambda_{\ell_r}}} \frac{f_{\ell_{r+1} k}}{f_{\ell_{r+1} \lambda_{\ell_{r+1}}}} (-1) \\
&= \begin{cases} \left(\prod_{j=2}^{r+1} \frac{f_{\ell_{j-1} \lambda_{\ell_j}}}{f_{\ell_j \lambda_{\ell_j}}} \right) f_{\ell_{r+1} k} & \text{if } r \geq 1 \\ f_{\ell_{r+1} k} & \text{if } r = 0 \end{cases} \\
&= u_l f_{lk}.
\end{aligned} \tag{2.29}$$

If $k \in \mathcal{I}_-$, $\beta_k < 0$ and $f_{lk} < 0$. It follows from (2.29) that $u_l = \frac{\beta_k}{f_k} > 0$. If $k \in \mathcal{I}_+$, then $\beta_k > 0$ and $f_{lk} > 0$. It follows from (2.29) that $u_l = \frac{\beta_k}{f_k} > 0$.

Finally, we show that with the given weights, (2.25) yields the desired inequality. Let $k \in \mathcal{V}_{\mathbf{0}}$. It follows directly from (2.29) that $\beta_k \leq \max_{l \in \mathcal{L}}\{u_l f_{lk}\}$. We next show that $\beta_k \geq \max\{u_l f_{lk} : l \in \mathcal{L}\}$. If l is not in the tree, the definition of the interval (2.28) directly implies that $\beta_k \geq u_l f_{lk}$. Consider therefore the situation where l is in the tree. Assume for a contradiction that $\beta_k < u_l f_{lk}$. Let C_1 be the set of nodes in the connected component for label l. Then, for any $i \in C_1$, $\beta_i = u_l f_{li}$. This shows that $k \notin C_1$. Choose a node k' in C_1 that belongs to $\mathcal{I}_-$ if $k \in \mathcal{I}_+$ and that belongs to $\mathcal{I}_+$ if $k \in \mathcal{I}_-$. Because $k' \in C_1$, $u_l = \frac{\beta_{k'}}{f_{lk'}}$. Our assumption then implies that $\beta_k < \frac{\beta_{k'}}{f_{lk'}} f_{lk}$, which is a contradiction to the fact that T is feasible. ■

Example 5 (continued) *Consider the spanning tree shown in Figure 2.4, which is label-connected and feasible. Because labels 1 and 3 are adjacent to node 0, we set*

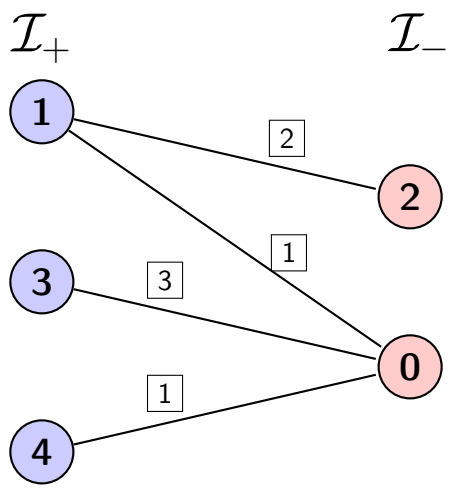


Figure 2.4.: Label-connected feasible spanning tree for Example 5

$u_1 = 1$ *and* $u_3 = 1$. *Then,* $u_2 = \frac{5}{3}$ *because* $f_{21} = 3$ *and* $f_{11} = 5$. *Finally,* u_4 *can be any value in* $[1, 5/2]$, *because label* 4 *does not appear in the tree. Using these weights yields* (2.26).

We next describe a procedure that starts from a valid inequality for $\operatorname{cl}\operatorname{conv}(Q^0)$ that is not facet-defining, and expresses it as a conic combination of "stronger" valid inequalities. In order to express this result, given a vector $\beta \in B_2$, we introduce the notation $d_{B_2}(\beta) = \dim(F)$, where F is the face of B_2 that contains β in its relative interior. Although this result can be proven in a more general setting, the specialized proof we give here has the advantage of yielding a low-order polynomial time algorithm for strengthening a valid inequality of $\operatorname{cl}\operatorname{conv}(Q^0)$ into a facet-defining inequality.

Proposition 2.6.2 *Let* $\beta^\intercal t \geq 0$ *be a valid inequality for* $\operatorname{cl}\operatorname{conv}(Q^0)$ *with* $\beta_0 < 0$ *that is not facet-defining,* i.e., $d_{B_2}(\beta) = k > 0$. *Then either*

1. *there exist two valid inequalities* $\bar{\beta}^\intercal t \geq 0$ *and* $\tilde{\beta}^\intercal t \geq 0$ *and* $\theta \in (0, 1)$ *such that* $\beta = \theta\bar{\beta} + (1-\theta)\tilde{\beta}$, $d_{B_2}(\bar{\beta}) < k$ *and* $d_{B_2}(\tilde{\beta}) < k$, *or*

2. there exists a valid inequality $\bar{\beta}^\intercal t \geq 0$ such that $\bar{\beta} \leq \beta$ and $d_{B_2}(\bar{\beta}) < k$.

Proof Let $\beta^\intercal t \geq 0$ be the given inequality. The coefficient vector β can be assumed to satisfy

$$\begin{aligned} \beta_j + w_{ij}\beta_i &\geq 0, & \forall (i,j) \in \mathcal{I}_+ \times \mathcal{I}_-, \\ \beta_i &\geq 0, & \forall i \in \mathcal{I}_+, \\ \beta_j &\leq 0, & \forall j \in \mathcal{I}_-, \\ \beta_0 &= -1, \end{aligned}$$

i.e., $\beta \in B_2$. For $i \in \mathcal{I}_+$, define

$$\delta_i = \begin{cases} \log \beta_i & \text{if } \beta_i > 0, \\ -\infty & \text{if } \beta_i = 0. \end{cases}$$

For $j \in \mathcal{I}_-$, define

$$\delta_j = \begin{cases} \log(-\beta_j) & \text{if } \beta_j < 0, \\ -\infty & \text{if } \beta_j = 0. \end{cases}$$

Given δ, we construct the subgraph $G_\delta(\mathcal{I}_+, \mathcal{I}_-)$ of $G(\mathcal{I}_+, \mathcal{I}_-)$ induced by the edges (i,j) for which inequality $\delta_j - \delta_i \leq c_{ij}$ is satisfied at equality by δ. Subgraph $G_\delta(\mathcal{I}_+, \mathcal{I}_-)$ is disconnected. In fact, if it was connected, any spanning tree would induce $\beta^\intercal t \geq 0$, and being feasible, would contradict the fact that this inequality is not facet-defining for $\operatorname{cl}\operatorname{conv}(Q^0)$; see Proposition 2.5.1.

Let C_1 and C_2 be the partition of $\mathcal{V}_\mathbf{0}$ where C_1 is the node set of the connected component of $G_\delta(\mathcal{I}_+, \mathcal{I}_-)$ that contains 0 and $C_2 = \mathcal{V}_\mathbf{0} \setminus C_1$. Compute

$$\begin{aligned} \Delta^+ &= \min \{\delta_i - \delta_j + c_{ij} : i \in C_1 \cap \mathcal{I}_+,\, j \in C_2 \cap \mathcal{I}_-\}, \\ \Delta^- &= \max \{-c_{ij} - \delta_i + \delta_j : i \in C_2 \cap \mathcal{I}_+,\, j \in C_1 \cap \mathcal{I}_-\}. \end{aligned}$$

There is at least one arc connecting C_1 with C_2 in $G(\mathcal{I}_+, \mathcal{I}_-)$. If not, $C_2 \cap \mathcal{I}_+ = \emptyset$, which means $C_1 \cap \mathcal{I}_+ \supseteq \mathcal{I}_+ \neq \emptyset$, yielding a contradiction to $C_2 \neq \emptyset$. Let $\chi(C)$ denote the indicator vector of C. Clearly, at least one of Δ^+ and Δ^- is well-defined. When Δ^+ is not well defined then $\chi(C_2)$ (resp. $-\chi(C_1)$) is a recession direction of

D_2 when $C_2 \cap \mathcal{I}_- = \emptyset$ (resp. $C_1 \cap \mathcal{I}_+ = \emptyset$) and we express $\delta = (\delta + \Delta^-\chi(C_2)) - \Delta^-\chi(C_2)$ (resp. $\delta = (\delta - \Delta^-\chi(C_1)) + \Delta^-\chi(C_1)$). Similarly, when Δ^- is not well defined, $C_2 \cap \mathcal{I}_+ = \emptyset$ then $-\chi(C_2)$ is a recession direction for D_2 and we express $\delta = (\delta + \Delta^+\chi(C_2)) - \Delta^+\chi(C_2)$. Finally, when Δ^+ and Δ^- are well defined, we express $\delta = \frac{\Delta^+}{\Delta^+-\Delta^-}(\delta + \Delta^-\chi(C_2)) - \frac{\Delta^-}{\Delta^+-\Delta^-}(\delta + \Delta^+\chi(C_2))$. The result still works after the transformation $\beta = e^\delta$ because for $\Delta', \Delta'' \geq 0$ and some set of nodes C, the perturbation $\delta + \Delta'\chi(C)$ (resp. $\delta - \Delta''\chi(C)$) yields an inequality $\beta' = e^{\Delta'\chi(C)} \circ \beta$ (resp. $\beta'' = e^{-\Delta''\chi(C)} \circ \beta$), where $\circ$ denotes Hadamard product. Since $e^{-\Delta''} \leq 1 \leq e^{\Delta'}$, β can be expressed as a convex combination of β' and β''. The case with recession cones also follows by letting $\Delta \to \infty$. Since the size of C_1 increases each time, the inequalities we use (if not the trivial recession directions) come from a smaller dimension face of D_2 and, hence of B_2. ■

We now illustrate the procedure used in the proof of Proposition 2.6.2.

Example 5 (continued) *Consider the inequality*

$$21t_1 - 7t_2 + 20t_3 + 4t_4 - 4t_0 \geq 0 \tag{2.30}$$

which is valid for $\operatorname{cl}\operatorname{conv}(Q^0)$ *since it can be obtained applying* (2.25) *using weight vector* $u = (4, 7, 5, 7)$. *We next express* (2.30) *as a weighted sum of facet-defining inequalities of* $\operatorname{cl}\operatorname{conv}(Q^0)$ *using the procedure underlying the proof of Proposition 2.6.2. Let* $\beta = (21, -7, 20, 4, -4)$. *For this vector* β, *only the two inequalities* $\beta_2 + w_{12}\beta_1 \geq 0$ *and* $\beta_0 + w_{40}\beta_4 \geq 0$ *are satisfied at equality. It follows that* $C_1 = \{4, 0\}$ *and* $C_2 = \{1, 2, 3\}$. *For* $f = {}^{12}/_{7}$ *and* $g = {}^{20}/_{21}$, *define* $\beta' = (21f, -7f, 20f, 4, -4) = (36, -12, {}^{240}/_{7}, 4, -4)$ *and* $\beta'' = (21g, -7g, 20g, 4, -4) = (20, -{}^{20}/_{3}, {}^{400}/_{21}, 4, -4)$ *and express*

$$\beta = \frac{1-g}{f-g}\beta' + \frac{f-1}{f-g}\beta'' = \frac{1}{16}\beta' + \frac{15}{16}\beta''. \tag{2.31}$$

We now apply the procedure again, for $\beta^1 = \beta'$ *and* $\beta^2 = \beta''$. *For* β^1, *only the three inequalities* $\beta_2 + w_{12}\beta_1 \geq 0$, $\beta_0 + w_{40}\beta_4 \geq 0$, *and* $\beta_2 + w_{42}\beta_4 \geq 0$ *are satisfied*

at equality. It follows that $C_1 = \{1, 2, 4, 0\}$ *and* $C_2 = \{3\}$*. For* $g = {}^{7}/_{10}$*, define* $\beta^{1,\prime\prime} = (36, -12, {}^{240}/_{7}\, g, 4, -4) = (36, -12, 24, 4, -4)$*. We write that*

$$\beta^1 = \beta^{1,\prime\prime} + (0, 0, {}^{240}/_{7}\,(1-g), 0, 0) = \beta^{1,\prime\prime} + (0, 0, {}^{72}/_{7}, 0, 0). \tag{2.32}$$

Then, $\beta^{1,\prime\prime}$ *is a non-trivial facet-defining inequality because the tight constraints form a spanning tree. For* β^2*, only the three inequalities* $\beta_2 + w_{12}\beta_1 \geq 0$*,* $\beta_0 + w_{40}\beta_4 \geq 0$*, and* $\beta_0 + w_{10}\beta_1 \geq 0$ *are satisfied at equality. It follows that* $C_1 = \{1, 2, 4, 0\}$ *and* $C_2 = \{3\}$*. For* $g = {}^{21}/_{25}$*, define* $\beta^{2,\prime\prime} = (20, -{}^{20}/_{3}, {}^{400}/_{21}\, g, 4, -4) = (20, -{}^{20}/_{3}, 16, 4, -4)$*. We write that*

$$\beta^2 = \beta^{2,\prime\prime} + (0, 0, {}^{400}/_{21}\,(1-g), 0, 0) = \beta^{2,\prime\prime} + (0, 0, {}^{64}/_{21}, 0, 0). \tag{2.33}$$

Again $\beta^{2,\prime\prime}$ *is facet-defining. Combining* (2.31)*,* (2.32) *and* (2.33)*, we obtain*

$$\beta = {}^{1}/_{16}\beta^{1,\prime\prime} + {}^{15}/_{16}\beta^{2,\prime\prime} + {}^{7}/_{2}(0, 0, 1, 0, 0).$$

In other words, (2.30) *can be expressed as a weighted combination of* (2.9)b*,* (2.9)a *and* $t_3 \geq 0$ *with weights* ${}^{4}/_{16}$*,* ${}^{60}/_{16}$*,* ${}^{7}/_{2}$*, respectively.*

It was observed in [56] that the c-max cut is not always facet-defining for the case where $|\mathcal{L}| = 2$. Using the algorithm underlying the proof of Proposition 2.6.2, we show next that, the coefficients of a c-max cut that is not facet-defining can always be strengthened to lead to a facet-defining inequality.

Proposition 2.6.3 *Any c-max cut can be expressed as a conic combination of a single nontrivial facet-defining inequality together with trivial inequalities. Moreover, the coefficients of the c-max cut and those of the single nontrivial facet-defining inequality are identical for each* $i \in \mathcal{I}_+$*.*

Proof In the proof of Proposition 2.6.2, if $C_1 \supseteq \mathcal{I}_+$, Δ^- is not defined, then the inequality is expressed as a conic combination of a tighter valid inequality and a trivial inequality, where the coefficients of variables in C_1 do not change. Therefore, we only need to show that throughout the procedure $C_1 \supseteq \mathcal{I}_+$. This is clearly true at

the beginning because the coefficient for each $i \in \mathcal{I}_+$ is derived from the inequality $l \in \operatorname{Arg\,max}_l f_{li}$. It is also true at each subsequent step because C_1 grows at each step of the procedure. ■

The question of when the c-max cut is facet-defining for $\operatorname{cl\,conv}(Q^0)$ with $|\mathcal{L}| = 2$, was raised but left open in [56]. The proofs of Propositions 2.6.2 and 2.6.3 answer this question in the general case where $|\mathcal{L}| \geq 2$. In fact, a c-max cut $\beta^\intercal t \geq 0$ is facet-defining for $\operatorname{cl\,conv}(Q^0)$ precisely when $C_2 = \emptyset$ at the first step in the proof of Proposition 2.6.2. Since $\mathcal{I}_+ \subseteq C_1$, this condition is equivalent to stating that each node $j \in \mathcal{I}_- \setminus \{0\}$ is such that $\beta_j + w_{ij}\beta_i = 0$ for some $i \in \mathcal{I}_+$. In terms of the initial problem formulation, this observation can be restated as follows.

Corollary 4 *A c-max cut $\beta^T t \geq 0$ is facet-defining for $\operatorname{cl\,conv}(Q^0)$ if and only if for each $j \in \mathcal{I}_- \setminus \{0\}$, there exists an $l \in \mathcal{L}$ and $i \in \mathcal{I}_+$ such that $\beta_i = f_{li}$ and $\beta_j = f_{lj}$.*

We may also use the constructive procedure used in the proof of Proposition 2.6.2 to design an algorithm to "tighten" a valid inequality for $\operatorname{cl\,conv}(Q^0)$ into a facet-defining of $\operatorname{cl\,conv}(Q^0)$. We choose to develop such an algorithm in the space of δ variables. A similar procedure could be developed in the space of β variables. The underlying idea is to expand the subgraph of tight equalities in D_1 for the given δ into a connected graph, while maintaining feasibility for the non-tight inequalities.

Algorithm 1 presents the pseudo-code for this constructive procedure. It is a slight variation on Prim's algorithm for minimum weight spanning trees; see [57]. It requires sets $\mathcal{I}_+$ and $\mathcal{I}_-$, a coefficient vector $\delta \in D_1$, and the matrix $C = [c_{ij}]$ where $c_{ij} = \log(w_{ij})$. Define $s_{ji} = s_{ij} = c_{ij} - \delta_j + \delta_i$ for $i \in \mathcal{I}_+$ and $j \in \mathcal{I}_-$. It is clear that $s_{ji} = s_{ij} \geq 0$ since $\delta \in D_1$. We will show later that $\delta + \text{key}^*$ corresponds to a facet-defining inequality where key^* is the key after the algorithm terminates.

For a given node set $X \subseteq \mathcal{V}_\mathbf{0}$, the operation Extract-Min(X) (resp. Extract-Max(X)) removes and returns the element of X with the smallest (resp. largest) key. We also define $\text{Min}(X) := \min_{i \in X} \text{key}[i]$ and $\text{Max}(X) := \max_{i \in X} \text{key}[i]$. We denote by $\mathcal{Q}_+$ (resp. $\mathcal{Q}_-$) the max-priority queue in $\mathcal{I}_+$ (resp. min-priority queue in $\mathcal{I}_-$)

```
k ← 0,  Q_+ ← I_+,  Q_− ← I_− \ {0}
key[i] ← −s_{i0}, ∀i ∈ Q_+,  key[i] ← ∞, ∀i ∈ Q_−,  key[0] ← 0
while Q_+ ≠ ∅ or Q_− ≠ ∅ do
    if Min(Q_−) − key[k] ≤ key[k] − Max(Q_+) then
        k ← Extract-Min(Q_−)
    else
        k ← Extract-Max(Q_+)
    end if
    for i ∈ Adj[k] do
        if i ∈ I_+ then
            key[i] ← max{key[i], −s_{ik} + key[k]}
        else
            key[i] ← min{key[i], s_{ki} + key[k]}
        end if
    end for
end while
```

Algorithm 1: Cut-Strengthening ($S = [s_{ij}], \mathcal{I}_+, \mathcal{I}_-$)

whose keys are not yet finalized. We let $\mathcal{Q} = \mathcal{Q}_+ \cup \mathcal{Q}_-$. For a node v, Adj$[v]$ is the set of nodes adjacent to v in $\mathcal{Q}$.

Let H represent the sequence of nodes extracted during the successive iterations of the **while** loop. For nodes i and j, we write $j \prec i$ if j occurs in H before i. We define key$[0] = 0$. Then, for $i \in \mathcal{Q}_+$ (resp. $\mathcal{Q}_- \backslash \{0\}$), we define key$[i]$ to be $\max_{j \in \mathcal{I}_- \backslash \mathcal{Q}_-, j \prec i} -s_{ij} + \text{key}[j]$ (resp. $\min_{j \in \mathcal{I}_+ \backslash \mathcal{Q}_+, j \prec i} s_{ij} + \text{key}[j]$). We use induction to show that keys follow this definition. The base case can be verified via the initial assignment of keys and the convention that $\min\{\emptyset\} = \infty$. If we assume that the keys satisfy the above definition before the iteration, then since $k \prec i$ for any $i \in \mathcal{Q} \backslash \{k\}$, the definition remains valid after the step 11 (resp. step 13). It is clear that once a node is extracted, the key of the node never changes in the remainder of the algorithm.

We first show that $\min(\mathcal{Q}_-) - \text{key}[k] \geq 0$ and $\text{key}[k] - \max(\mathcal{Q}_+) \geq 0$ at step 4 of the algorithm. This is clearly true for the base case. We assume that these inequalities are true and we choose to extract k' at either step 5 or step 7. We will only argue that the above inequalities hold for k' selected at step 5 because the other case is similar. We first argue that the result holds before step 9. If k' was selected at step 5, *i.e.*, $k' \in \mathcal{Q}_-$, then the first inequality holds because k' was chosen to be the node with minimum key in $\mathcal{Q}_-$. The second inequality holds because $\text{key}[k'] - \text{key}[k] \geq 0$ and $\text{key}[k] - \max(\mathcal{Q}_+) \geq 0$ by induction hypothesis. Now, we show that these inequalities continue to hold until the step 4 of the next iteration. In particular, observe that for $j \in \mathcal{Q}_-$ (resp. $j \in \mathcal{Q}_+$), since $\text{key}[j] \geq \text{key}[k']$ (resp. $\text{key}[j] \leq \text{key}[k']$) before the update in step 13 (resp. step 11) and $s_{jk'} \geq 0$, it remains so after the update as well.

Now, we show that at each iteration of the algorithm where node k' is extracted, $\delta + \sum_{j \preceq k'} \text{key}[j]\chi(\{j\}) + \text{key}[k']\chi(\mathcal{Q})$ defines a valid cut. This is trivially true for the base case. We now consider the case when k' is extracted. The incremental change to the vector is $(\text{key}[k'] - \text{key}[k])\chi(\mathcal{Q} \cup \{k'\})$, where k immediately precedes k' in H. Clearly, this change does not affect any inequality in D_1 expressed for nodes i and j which both precede k' or both succeed k'. Therefore, we only need to concern ourselves with an inequality with respect to i and j where $i \preceq k' \preceq j$. Assume

$j \in \mathcal{Q}_+$. If $k' \in \mathcal{I}_+$, then the result follows because $0 \leq s_{ij} + \text{key}[j] - \text{key}[i] \leq s_{ij} + \text{key}[k'] - \text{key}[i]$ because k' is the maximizer in $\mathcal{Q}_+$. On the other hand, if $k' \in \mathcal{I}_-$, then $s_{ij} + \text{key}[k'] - \text{key}[i] \geq 0$ if $k' = i$ and $s_{ij} + \text{key}[k'] - \text{key}[i] \geq s_{ij} + \text{key}[k] - \text{key}[i] \geq 0$, where the first inequality follows because $\text{key}[k'] - \text{key}[k] \geq 0$ by our earlier proof and the second inequality by the induction hypothesis and because $\text{key}[i]$ was not updated. The proof for the case $j \in \mathcal{Q}_-$ is similar.

It follows from the definition of keys that at least one of the inequalities with respect to k' and its predecessors becomes tight. Since the procedure only stops when all the nodes are visited, it follows that the graph of tight inequalities is connected at the end and $\delta + \sum_{j \in \mathcal{V}_0} \text{key}^*[j]\chi(\{j\})$ defines a facet-defining inequality.

We now show that all the tight inequalities remain tight during the procedure. In particular, assume $s_{ij} = 0$. Assume $j \preceq i$ where $j \in \mathcal{I}_-$ (the proof for $j \in \mathcal{I}_+$ is similar). Clearly, when $k = j$ at step 4, $\text{key}[j] \geq \text{key}[i]$. However, $\text{key}[i] \geq \text{key}[j]$ because of the previous update at step 11. Therefore, $\text{key}[j] - \text{key}[i] = 0$. Then, because of the condition in step 4, the keys added match $\text{key}[i]$. Therefore, there is no update to $\text{key}[i]$ because $\text{key}[i] \geq -s_{ik} + \text{key}[k]$ follows from $s_{ik} \geq 0$ and $\text{key}[i] = \text{key}[k]$.

The above algorithm can be implemented using heaps for both $\mathcal{Q}_+$ and $\mathcal{Q}_-$. If the graph $G(\mathcal{I}_+, \mathcal{I}_-)$ has n nodes and e edges, it requires $O(n)$ Insert, $O(n)$ Min, $O(n)$ Max, $O(n)$ Extract-Min and Extract-Max, and $O(e)$ Decrease-Key operations. With Fibonacci heaps, the running time is $O(e + n \log n)$ which exactly matches that of Prim's algorithm. We summarize the above discussion as follows.

Corollary 5 *From any valid inequality of* $\text{cl conv}(Q^0)$*, Algorithm 1 constructs a facet-defining inequality in time* $O(e + n \log n)$*. Further, the facet obtained contains the face defined by the initial inequality.*

2.7 Conclusion

Given an LP relaxation of the problem, and a basic solution that does not satisfy the cardinality requirement, we derive violated valid inequalities that can be generated

in polynomial time. These inequalities are facet-defining for a disjunctive relaxation of the problem. The separation is carried out by solving a network-flow problem in the original problem space instead of a higher-dimensional cut-generation LP typically used in disjunctive programming. We show that facet-defining inequalities can be associated with label-connected feasible spanning trees of a suitably defined bipartite graph and, consequently, derive various insights into their structure and validity. Using these insights, we modify the recently proposed E&R procedure, which generates cuts for complementarity problems, to the more general setting involving cardinality constraints. Our analysis reveals conditions under which the c-max cut, a cut widely used in the complementarity literature, is not facet-defining and can be improved using a simple procedure. More generally, we develop a fast separation procedure that tightens valid inequalities into facet-defining inequalities for our relaxation using a Prim-type combinatorial algorithm.

3. Semidefinite programming relaxations for sparse principal component analysis

In this chapter, we provide a characterization of the convex hull of the feasible set of an optimization formulation of sparse principal component analysis (sparse PCA). Sparse PCA seeks to find a linear combination of a small number of the variables of some data that explains most of its variance. We obtain a description of the convex hull in a lifted space by dualizing the separation problem and making use of majorization inequalities. This interpretation allows us to express each point in the convex hull as a convex combination of points that satisfy the cardinality constraint. Based on the convex hull result, we prove that sparse PCA can be reformulated as a continuous convex maximization problem. We next propose an SDP relaxation in a lifted space. Our preliminary computational experiments show that the gaps of our SDP relaxations are more than 90% smaller than those of the SDP relaxation proposed by d'Aspremont et al. [24] on the `pitprops` problem and on test problems with randomly generated covariance matrices.

3.1 Introduction

Principal component analysis (PCA) is a dimension reduction technique in exploratory multivariate statistical analysis with a wide variety of applications such as image compression, gene expression, portfolio hedging, and quality control. We refer the readers to [44] for additional applications of PCA. Given a dataset with intercorrelated variables, PCA generates a sequence of mutually uncorrelated variables called *principal components* (PCs) which are linear combinations of the variables in the data in a way that the first PC exhibits the largest variance of the dataset, and succeeding PCs exhibit the largest variance of the dataset under an orthogonality requirement

with the preceding PCs. PCs can be equivalently defined as unit eigenvectors of the centered covariance matrix of the dataset. More specifically, suppose the original dataset has n variables and assume that Σ is the centered covariance matrix for the dataset. Then, the ith PC is the eigenvector associated with the ith largest eigenvalue for $i = 1, \ldots, n$. Therefore, the first PC is an optimal solution to

$$\begin{array}{ll} \text{maximize} & x^T \Sigma x \\ \text{subject to} & \|x\| \leq 1. \end{array} \tag{3.1}$$

Principal components are linear combinations of most of the variables of the data. It is therefore often difficult to understand what each PC represents. To make the interpretation simpler, it is useful to find variables which are linear combinations of a small number of original variables. The problem of calculating such sparse loadings is called *sparse principal component analysis* (sparse PCA). By adding a sparsity (or cardinality) constraint to (3.1), sparse PCA can be formulated as

$$\begin{array}{ll} \text{maximize} & x^T \Sigma x \\ \text{subject to} & \|x\| \leq 1, \\ & \text{card}(x) \leq K \end{array} \tag{3.2}$$

where $1 < K < n$ and $\text{card}(x)$ represents the number of nonzero components of x. Tillmann and Pfetsch [68] showed that sparse PCA is NP-hard in the strong sense.

Sparse PCA has been extensively studied in the literature. Moghaddam et al. [53] used a greedy algorithm and branch-and-bound methods for sparse PCA and d'Aspremont et al. [23] proposed a modified greedy algorithm whose running time is $O(n^3)$. Jolliffe et al. [45] proposed SCoTLASS which adds a l_1-norm regularization similar to LASSO in regression. Zou et al. [75] formulated sparse PCA as a regression-type optimization problem and impose an elastic net constraint on the regression coefficients. To the best of the author's knowledge, the first SDP relaxation for sparse PCA was proposed by d'Aspremont et al. [24] which the authors solve using. Sriperumbudur et al. [66] considered generalized eigenvalue problems with sparse PCA as a special case and proposed a DC-based algorithm to find a local optimal

solution. Shen and Huang [63] proposed a method to extract sparse PCs by solving a low rank matrix approximation problem. Journée et al. [46] reformulated sparse PCA as maximizing a convex function over a compact set and generalized a power method to find a locally optimal solution. Lu and Zhang [50] developed an augmented Lagrangian method to find sparse and nearly uncorrelated components with orthogonal loading vectors which exhibit as much of the total variance as possible. More recently, an iterative thresholding approach was developed by Ma [51].

For every vector x (resp. matrix X), $|x|$ (resp. $|X|$) represents the vector (resp. matrix) of component-wise absolute values of x (resp. X). A *permutation* of $\{1,\ldots,n\}$ is defined as a bijection from $\{1,\ldots,n\}$ to itself. For a permutation σ of $\{1,\ldots,n\}$, we define the *permutation of a vector* $x \in \mathbb{R}^n$ *with respect to* σ as the vector whose ith component is $x_{\sigma(i)}$. Given a point $x \in \mathbb{R}^n$, the *permutahedron* with respect to x, which we denote by $\mathrm{Perm}(x)$, is the convex hull of permutations of x. The matrix corresponding to a permutation σ, which we denote by P_σ, is called the *permutation matrix* associated with σ. That is, for every $x \in \mathbb{R}^n$, $[P_\sigma x]_i = x_{\sigma(i)}$. The set of all the permutation matrices is denoted by $\mathcal{P}$. A set S is called *sign-invariant* if $x \in S$ and $|x| = |y|$ imply that $y \in S$. A set S is called *permutation-invariant* if and only if $x \in S$ implies $Px \in S$ for all $P \in \mathcal{P}$. The polar S° of a set $S \in \mathbb{R}^n$ is defined as $S^\circ := \{y \in \mathbb{R}^n : x^\intercal y \leq 1 \text{ for all } x \in S\}$. We denote the trace of a matrix X by $\mathrm{Tr}(X)$. Further, for a given matrix X, we denote by $\mathrm{diag}(X)$ the diagonal matrix whose diagonal elements match those of X. The diagonal matrix whose diagonal entries equal to the vector x is denoted by $\mathrm{diag}(x)$. For vector $x \in \mathbb{R}^n$, $x_{[i]}$ represents the ith largest component of x for all $i = 1,\ldots,n$. Given two vectors $x, y \in \mathbb{R}^n$, x *majorizes* y if and only if

$$\begin{array}{rcl} \sum_{i=1}^{j} x_{[i]} & \geq & \sum_{i=1}^{j} y_{[i]}, \quad j = 1,\ldots,n-1 \\ \sum_{i=1}^{n} x_i & = & \sum_{i=1}^{n} y_i \end{array} \tag{3.3}$$

and we denote the system (3.3) by $x \geq_m y$ and refer to it as a *majorization inequality*. We mention that majorization is typically written using "$\succeq$", but we reserve this symbol for positive semidefiniteness. Furthermore, we use "$\geq$" to state that a vector

is component-wise nonnegative. The set of n-by-n symmetric matrices is denoted by S^n. A matrix is called *doubly stochastic* if each rows and columns sums up to 1. We denote the p-by-q matrix of ones by $J(p, q)$.

Majorization has an elegant geometric interpretation. To describe it, we introduce Schur's result on the relationship between majorization and doubly stochastic matrices and Birkhoff's Theorem on the characterization of doubly stochastic matrices.

Theorem 3.1.1 (Schur [61], 1923) *$x \geq_m y$ if and only if $y = Dx$ for some doubly stochastic matrix D.* ■

Theorem 3.1.2 (Birkhoff Theorem [15], 1946) *The set of doubly stochastic matrices is the convex hull of the permutation matrices.* ■

Corollary 6 *$x \geq_m y$ if and only if y is a convex combination of permutations of x, or, equivalently $y \in \mathrm{Perm}(x)$.*

In this chapter, we first characterize the convex hull of the feasible set of (3.2) based on its sign- and permutation-invariance structure by dualizing a separation problem (Section 3.2). The convex hull is written through a *majorization inequality* which provides a geometric interpretation for points in the convex hull: each point in the convex hull can be written as a convex combination of points that satisfy the cardinality constraint. Based on the characterization of the convex hull, we reformulate sparse PCA as a continuous convex maximization problem by providing a procedure to convert a optimal to the convex maximization into an optimal solution that satisfies the cardinality constraint. We then introduce a new SDP relaxation based on the reformulation derived from majorization (Section 3.3). We show that our SDP relaxation generalizes that proposed by d'Aspremont et al. [24]. Lastly, we present preliminary computational results for `pitprops` problems and problems with randomly generated covariance matrices that show that the gaps of our SDP relaxations are more than 90% smaller than the gaps of the classical SDP relaxation (Section 3.4).

3.2 Characterization of the convex hull of sparse PCA

3.2.1 Separation problem

Let $F = \{x \in \mathbb{R}^n : \|x\| \leq 1, \mathrm{card}(x) \leq K\}$ be the feasible set of sparse PCA. Given $t \in \mathbb{R}^n$, consider the following separation problem:

$$z^*(t) = \max\{t^\intercal \beta : \beta^\intercal x \leq 1 \text{ is valid for } F\}. \tag{3.4}$$

Observe that $\pm e_i, 0 \in F$ for all $i = 1, \ldots, n$. This shows that 0 is in the interior of $\mathrm{conv}(F)$. Hence we assume without loss of generality that the right-hand-side of the valid inequality is 1 by scaling coefficients.

Lemma 4 (3.4) *has an optimal solution.*

Proof Observe first that $F^\circ = (\mathrm{conv}(F))^\circ$ since an inequality is valid for F if and only if it is valid for $\mathrm{conv}(F)$. Since F is a finite union of compact set, F is compact and hence so is $\mathrm{conv}(F)$ (see Proposition 1.3.2 of [11], for example). Since 0 is in the interior of $\mathrm{conv}(F)$, its dual F° is compact (see page 47 of [34], for example). The result follows. ■

It is straightforward that $t \in \mathrm{conv}(F)$ if and only if $z^*(t) \leq 1$. The next proposition gives a characterization of when an inequality is valid.

Proposition 3.2.1 *$\beta^\intercal x \leq 1$ is valid for F if and only if $|\beta|^2_{[1]} + \cdots + |\beta|^2_{[K]} \leq 1$.*

Proof Suppose $\beta^\intercal x \leq 1$ is valid for F. Let σ be a permutation such that $|\beta|_{\sigma(1)} \geq \cdots \geq |\beta|_{\sigma(n)}$. Define $\bar{\beta}$ by

$$\bar{\beta}_i = \begin{cases} \beta_i & \text{if } i \in \{\sigma(1), \ldots, \sigma(K)\}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\bar{x} := \bar{\beta}/\|\bar{\beta}\|$. Then, $\bar{x} \in F$ because $\|\bar{x}\| = 1$ and $\text{card}(\bar{x}) \leq K$. Further, $\beta^\intercal \bar{x} = \|\bar{\beta}\| \leq 1$. This shows that $|\beta|^2_{[1]} + \cdots + |\beta|^2_{[K]} = |\beta|^2_{\sigma(1)} + \cdots + |\beta|^2_{\sigma(K)} = \|\bar{\beta}\|^2 \leq 1$. Conversely, suppose $|\beta|^2_{[1]} + \cdots + |\beta|^2_{[K]} \leq 1$. Consider any $x \in F$. Then,

$$\begin{aligned} \beta^\intercal x &= \beta_1 x_1 + \cdots + \beta_n x_n \leq |\beta|_{[1]}|x|_{[1]} + \cdots + |\beta|_{[n]}|x|_{[n]} \\ &= |\beta|_{[1]}|x|_{[1]} + \cdots + |\beta|_{[K]}|x|_{[K]} \\ &\leq \|(|\beta|_{[1]}, \ldots, |\beta|_{[K]})\| \leq 1, \end{aligned}$$

where the first inequality holds by the rearrangement inequality, the second equality results from the the cardinality constraint, and the second inequality holds by Cauchy-Schwarz. We conclude that $\beta^\intercal x \leq 1$ is valid for F. ■

Corollary 7 *F° is sign- and permutation-invariant.* ■

Proposition 3.2.2

1. *Suppose β^* is an optimal solution to* (3.4)*. Then, $t_i\beta^*_i \geq 0$ for $i = 1, \ldots, n$,*

2. *Suppose that β^* is an optimal solution to* (3.4) *and that $|t|_{\sigma(1)} \geq \cdots \geq |t|_{\sigma(n)}$ for some permutation σ for $\mathbb{R}^n$. Then,*

$$|\beta^*|_{\sigma(1)} \geq \cdots \geq |\beta^*|_{\sigma(n)}.$$

3. *There exists an optimal solution to* (3.4) *β^{**} such that $|\beta^{**}|_{[K]} = \cdots = |\beta^{**}|_{[n]}$.*

Proof Suppose there exists i such that $t_i\beta^*_i < 0$. Then, replacing β^*_i with $-\beta^*_i$, the objective function value improves strictly, which produces the desired contradiction. Part 2 is directly from the rearrangement inequality and Corollary 7. For Part 3, consider an optimal solution β^* to (3.4) and assume that $|\beta^*|_{\sigma(1)} \geq \cdots \geq |\beta^*|_{\sigma(n)}$ for some permutation σ. Then, the vector β^{**} obtained from β^* by replacing $\beta^*_{\sigma(i)}$ for $i = K+1, \ldots, n$ with $\text{sign}(t_i)|\beta^*|_{\sigma(K)}$ is also optimal for (3.4) by Proposition 3.2.1. The result follows. ■

Suppose $|t|_{\sigma(1)} \geq \cdots \geq |t|_{\sigma(n)}$ for some permutation σ for $\mathbb{R}^n$. Proposition 3.2.2 shows that (3.4) can be rewritten as

$$\begin{aligned} z^*(t) \quad &= \text{maximize} && \textstyle\sum_{i=1}^{K-1} |t|_{\sigma(i)}|\beta|_{\sigma(i)} + \left(\sum_{i=K}^{n} |t|_{\sigma(i)}\right) |\beta|_{\sigma(K)} \\ &\quad \text{subject to} && |\beta|_{\sigma(1)} \geq \cdots \geq |\beta|_{\sigma(K)}, \\ & && |\beta|^2_{\sigma(1)} + \cdots + |\beta|^2_{\sigma(K)} \leq 1 \end{aligned} \tag{3.5}$$

or equivalently,

$$\begin{aligned} z^*(t) \quad &= \text{maximize} && \textstyle\sum_{i=1}^{K-1} |t|_{\sigma(i)}\gamma_i + \left(\sum_{i=K}^{n} |t|_{\sigma(i)}\right) \gamma_K \\ &\quad \text{subject to} && \gamma_1 \geq \cdots \geq \gamma_K \geq 0, \\ & && \gamma_1^2 + \cdots + \gamma_K^2 \leq 1. \end{aligned} \tag{3.6}$$

Recall that $t \in \text{conv}(F)$ if and only if $z^*(t) \leq 1$. When $z^*(t) > 1$, an inequality which cuts off t from $\text{conv}(F)$ can be obtained from an optimal solution to (3.6) by an appropriate permutation and sign-conversion based on Proposition 3.2.2. The separation problem not only gives us a separating scheme but it also provides us with a description of the convex hull. To see this, consider the dual of (3.6):

$$\begin{aligned} \text{minimize} \quad & \|u\|^2 \\ \text{subject to} \quad & u_1 = |t|_{\sigma(1)} + \lambda_1, \\ & u_i = |t|_{\sigma(i)} + \lambda_i - \lambda_{i-1}, \qquad i = 2, \ldots, K-1, \\ & u_K = \textstyle\sum_{i=K}^{n} |t|_{\sigma(i)} - \lambda_{K-1}, \\ & \lambda \geq 0. \end{aligned}$$

The objective function of the dual should be $\|u\|$. We replace it with $\|u\|^2$ since it simplifies the analysis and does not modify the set of optimal solutions to the problem. After simplifying further, we obtain

$$\begin{aligned} \text{minimize} \quad & \|u\|^2 \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} |t|_{\sigma(i)}, \quad j = 1, \ldots, K-1, \\ & \textstyle\sum_{i=1}^{K} u_i = \sum_{i=1}^{n} |t|_{\sigma(i)} \end{aligned}$$

or equivalently,

$$\begin{aligned} \text{minimize} \quad & \|u\|^2 \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} |t|_{[i]}, \quad j = 1, \ldots, K-1, \\ & \textstyle\sum_{i=1}^{K} u_i = \sum_{i=1}^{n} |t|_{i}. \end{aligned} \tag{3.7}$$

Observe that if u^* is an optimal solution to (3.7), then so is $(u^*_{[1]}, \ldots, u^*_{[K]})$. Therefore, we can add the constraints $u_1 \geq \cdots \geq u_K$ to the formulation.

Proposition 3.2.3 *Suppose u^* is an optimal solution to* (3.7)*. Then, $u^* \geq 0$.*

Proof First, it is clear that $t = 0$ if and only if $u^* = 0$. Hence we assume that $t \neq 0$ and hence $u^*_1 > 0$. Assume by contradiction that there exists an integer $M \in \{1, \ldots, K-1\}$ such that $u^*_1 \geq \cdots \geq u^*_M \geq 0 > u^*_{M+1} \geq \cdots \geq u^*_K$. Then, as $\sum_{j=1}^K u^*_j > 0$, there exists a unique $i \in \{1, \ldots, M\}$ such that $\sum_{j=i}^K u^*_j \geq 0$ and $\sum_{j=i+1}^K u^*_j < 0$. Define $\bar{u}$ as

$$\bar{u}_j = \begin{cases} u^*_j, & j \in \{1, \ldots, i-1\}, \\ \sum_{j=i}^K u^*_j & j = i, \\ 0, & j \in \{i+1, \ldots, K\}. \end{cases}$$

Then, $\bar{u} \geq 0$ and $\bar{u}$ is feasible for (3.7). Moreover, it is easy to show that $u^* \geq_m \bar{u}$. By strict convexity of the objective function of (3.7), $\|u^*\| > \|\bar{u}\|$. This is the desired contradiction. ■

Consequently, we can reformulate (3.7) as

$$\begin{array}{ll} \text{minimize} & \|u\|^2 \\ \text{subject to} & u_1 \geq \cdots \geq u_K \geq 0, \\ & \sum_{i=1}^j u_i \geq \sum_{i=1}^j |t|_{[i]}, \quad j = 1, \ldots, K-1, \\ & \sum_{i=1}^K u_i = \sum_{i=1}^n |t|_i, \end{array} \tag{3.8}$$

We now assume that t is a variable and recall the fact that $t \in \text{conv}(F)$ if and only if $z^*(t) \leq 1$.

3.2.2 Characterization of the convex hull

We next augment $u \in \mathbb{R}^K$ into $(u, 0) \in \mathbb{R}^n$ and redefine $u := (u, 0)$. Define $\Delta = \{u : u_1 \geq \cdots \geq u_n \geq 0\}$. The following result gives a characterization of $\text{conv}(F)$.

Theorem 3.2.1 $\operatorname{conv}(F) = \{t : u \in F \cap \Delta, u \geq_m |t|\}$.

Proof Define $G = \{t : u \in F \cap \Delta, u \geq_m |t|\}$. First, observe that

$$F \cap \Delta = \left\{ u : \begin{array}{l} \|u\| \leq 1, \\ \operatorname{card}(u) \leq K, \\ u_1 \geq \cdots \geq u_n \geq 0 \end{array} \right\} = \left\{ u : \begin{array}{l} \|u\| \leq 1, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0 \end{array} \right\}$$

and hence it is convex. Since it is clear that Slater's condition holds, by strong duality, $t \in \operatorname{conv}(F)$ if and only if there exists $u \in \mathbb{R}^K$ such that $\|u\| \leq 1$ and (t, u) is feasible for (3.8). Therefore, we have that

$$\operatorname{conv}(F) = \left\{ t : \begin{array}{l} u \in F \cap \Delta, \\ \sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} |t|_{[i]}, \;\; j = 1, \ldots, K-1, \\ \sum_{i=1}^{K} u_i = \sum_{i=1}^{n} |t|_i \end{array} \right\}. \tag{3.9}$$

First, suppose $t \in \operatorname{conv}(F)$ so that there exists u such that (t, u) satisfies all the inequalities in (3.9). Then, for $j = K, \ldots, n-1$,

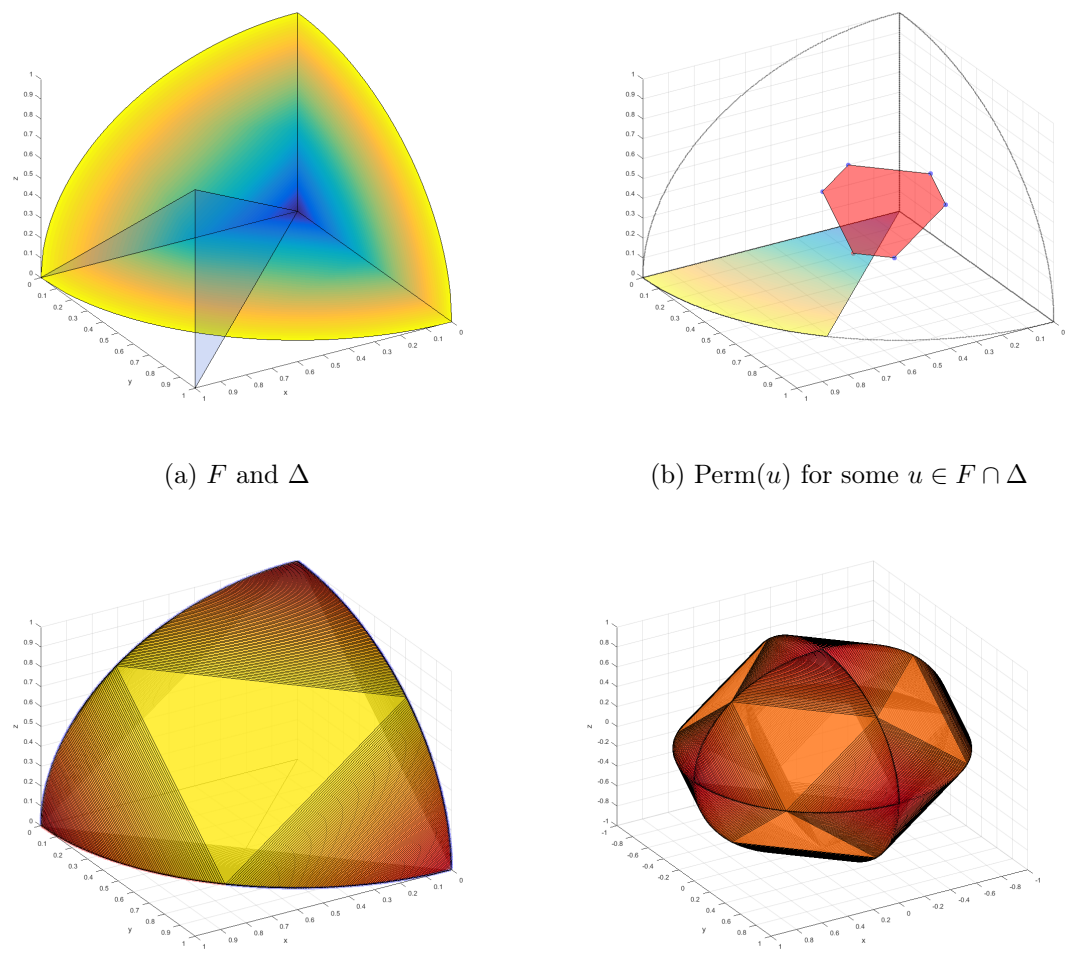
$$\sum_{i=1}^{j} u_i = \sum_{i=1}^{K} u_i = \sum_{i=1}^{n} |t|_i \geq \sum_{i=1}^{j} |t|_{[i]}.$$

Similarly, $\sum_{i=1}^{n} u_i = \sum_{i=1}^{n} |t|_i$ and hence $t \in G$. Conversely, suppose $t \in G$. Then, there exists $u \in F \cap \Delta$ such that $u \geq_m |t|$. It follows that $\sum_{i=1}^{K} u_i = \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} |t|_i$ and hence $t \in \operatorname{conv}(F)$. Therefore, $\operatorname{conv}(F) = \{t : u \in F \cap \Delta, u \geq_m |t|\}$. ■

The fact that (3.9) is convex can be verified directly since the functions $t \mapsto \sum_{i=1}^{j} |t|_{[i]}$ for $j = 1, \ldots, K-1$ are convex.

The majorization inequality yields the following geometric interpretation. It is easy to show that $x \in \operatorname{conv}(F)$ implies that $|x| \in \operatorname{conv}(F \cap \mathbb{R}^n_+)$. Therefore, the convex hull of the entire region is obtained by replicating the convex hull over $\mathbb{R}^n_+$. Observe that $t \in \operatorname{conv}(F \cap \mathbb{R}^n_+)$ if and only if t can be written as a convex combination of some point $u \in F \cap \Delta$ and its permutations. In other words, $t \in \operatorname{Perm}(u)$ for some $u \in F \cap \Delta$.

Example 6 *Consider the case where $n = 3$ and $K = 2$. Figure 3.1a shows the F and the simplicial cone Δ in the first quadrant. For a fixed $u \in F \cap \Delta$, the permutahedron generated by u is shown Figure 3.1b. The convex hull of sparse PCA in the first quadrant is then obtained by taking the union of all the possible permutahedra (see Figure 3.1c), that is,* $\mathrm{conv}(F \cap \mathbb{R}^3_+) = \bigcup_{u \in F \cap \Delta} \mathrm{Perm}(u)$. *By replicating the result in the first quadrant, we obtain the convex hull in $\mathbb{R}^3$ as shown in Figure 3.1d.*



(a) F and Δ

(b) $\mathrm{Perm}(u)$ for some $u \in F \cap \Delta$

(c) $\mathrm{conv}(F \cap \mathbb{R}^3_+)$

(d) $\mathrm{conv}(F)$

We next introduce variables v and w to model $|t|$ in the above convex hull formulation. That is, we replace $|t|$ with y and add constraints $y = v + w$, $t = v - w$, and $v, w \geq 0$ to the system. We obtain

$$\operatorname{conv}(F) = \left\{ t \in \mathbb{R}^n : \begin{array}{l} \|u\| \leq 1, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ \sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} y_{[i]}, \;\; j = 1, \ldots, K-1, \\ \sum_{i=1}^{K} u_i = \sum_{i=1}^{n} y_i, \\ y = v + w, \;\; t = v - w, \\ v, w \geq 0 \end{array} \right\}. \tag{3.10}$$

The following result shows an alternative characterization for the convex hull obtained by modeling the sum of j-largest components of y for all $j = 1, \ldots, K-1$.

Theorem 3.2.2

$$\operatorname{conv}(F) = \left\{ t \in \mathbb{R}^n : \begin{array}{l} \|u\| \leq 1, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ \sum_{i=1}^{j} u_i \geq j r_j + \sum_{i=1}^{n} s_i^j, \;\; j = 1, \ldots, K-1, \\ \sum_{i=1}^{K} u_i = \sum_{i=1}^{n} y_i, \\ y = v + w, t = v - w, \\ v, w \geq 0, \\ y_i \leq r_j + s_i^j, j \in \{1, \ldots, K-1\}, \;\; i \in 1, \ldots, n, \\ s \geq 0 \end{array} \right\}. \tag{3.11}$$

Proof Denote the set on the right-hand-side of (3.11) by G. It is clear that G is convex. Suppose $t \in \operatorname{conv}(F)$. Then, there exist $u, y, v, w \in \mathbb{R}^n$ that satisfy the

constraints in (3.10). For a given $j \in \{1, \ldots, K-1\}$ and a set of real numbers $\{y_1, \ldots, y_n\}$, consider the following optimization problem

$$\begin{array}{ll} \text{maximize} & y_1 z_1 + \cdots + y_n z_n \\ \text{subject to} & z_1 + \cdots + z_n = j, \ . \\ & 0 \leq z \leq 1. \end{array} \tag{3.12}$$

Observe that (3.12) returns $\sum_{i=1}^{j} y_{[i]}$ and, by strong duality, so does its dual

$$\begin{array}{ll} \text{minimize} & jr + \sum_{i=1}^{n} s_i \\ \text{subject to} & y_i \leq s_i + r, \quad i = 1, \ldots, n, \ . \\ & s \geq 0. \end{array} \tag{3.13}$$

It follows that there exist $s \in \mathbb{R}^n$ and $r \in \mathbb{R}$ such that

$$\begin{array}{l} jr_j + \sum_{i=1}^{n} s_i^j = \sum_{i=1}^{j} y_{[i]}, \\ y_i \leq s_i^j + r_j, \\ s^j \geq 0. \end{array}$$

Since j can be chosen arbitrarily in $\{1, \ldots, K-1\}$, $\text{conv}(F) \subseteq \text{proj}_t(G)$.

Conversely, suppose $t \in \text{proj}_t(G)$. Then, there exist u, s, r, y, v, and w that satisfy the constraints in (3.11). Then, for each $j \in \{1, \ldots, K-1\}$, weak duality implies that $jr_j + \sum_{i=1}^{n} s_i^j \geq \sum_{i=1}^{j} y_{[i]}$. Therefore, $\sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} y_{[i]}$ for all $j \in \{1, \ldots, K-1\}$ which implies that $\text{conv}(F) \supseteq \text{proj}_t(G)$. ■

Alternatively, we can make use of the formulation of the permutahedron proposed by Goemans [30] to model the majorization inequality. Then, the majorization inequality $x \geq_m y$ can be modeled as $y \in \text{Perm}(x)$ which corresponds to a set of linear inequalities.

3.2.3 Recovery of an optimal solution

The convex hull result enables us to build the following relaxation of sparse PCA as follows:

$$\max\{x^\intercal \Sigma x : x \in G\} \tag{3.14}$$

where G is defined as the right-hand-side of (3.11).

Proposition 3.2.4 *Suppose that $\Sigma \succ 0$. Then, an optimal solution x^* to* (3.14) *is an optimal solution to* (3.2).

Proof Assume by contradiction that $\text{card}(x^*) > K$. Then, since F is a disjunctive set and $x^* \notin F$, there exist $v_i \in F$ and $\lambda_i > 0$ for $i = 1, \ldots, t$ where $t \geq 2$ such that $x^* = \sum_{i=1}^t \lambda_i u^i$ and $\sum_{i=1}^t \lambda_i = 1$. Let $f(x) = x^\intercal \Sigma x$. Using the fact that f is strictly convex and that x^* is a global maximizer, we write

$$f(x^*) < \sum_{i=1}^t \lambda_i f(v_i) \leq \sum_{i=1}^t \lambda_i f(x^*) = f(x^*),$$

which is the desired contradiction. Therefore, $\text{card}(x^*) \leq K$ and hence $x^* \in F$. ■

Now we study the case where $\Sigma \succeq 0$. Consider a permutation $\sigma \in \mathcal{P}$ such that $|x^*|_{\sigma(1)} \geq \cdots \geq |x^*|_{\sigma(n)}$ and define $\bar{x}$ to be the vector with components $\bar{x}_i = |x^*|_{\sigma(i)}$ for $i = 1, \ldots, n$. Furthermore, define $\delta \in \mathbb{R}^K$ by $\delta_i = \frac{1}{K-i+1} \sum_{k=i}^n \bar{x}_k$ for $i \in \{1, \ldots, K\}$ and let $m \in \operatorname{Arg\,min}\{\delta_1, \ldots, \delta_K\}$. We next define $\bar{u}$ as

$$\bar{u}_i = \begin{cases} \bar{x}_i, & i = 1, \ldots, m-1, \\ \delta_m, & i = m, \ldots, K, \\ 0, & i = K+1, \ldots, n. \end{cases}$$

When $m = 1$, define $\bar{u}_1 = \delta_1$. For any vector $v \in \mathbb{R}^n$ and permutation $\sigma \in \mathcal{P}$, we define $F_\sigma(v) \in \mathbb{R}^n$ as

$$[F_\sigma(v)]_i = \text{sign}(x_i^*) v_{\sigma^{-1}(i)}$$

for $i = 1, \ldots, n$. Observe that $F_\sigma(\bar{x}) = x^*$ and define $u^* := F_\sigma(\bar{u})$.

Proposition 3.2.5 *Suppose x^* is an optimal solution to* (3.14) *and $\sigma \in \mathcal{P}, \bar{x} \in \mathbb{R}^n, \delta \in \mathbb{R}^K, m \in \{1, \ldots, K\}, \bar{u} \in \mathbb{R}^n$, and $u^* \in \mathbb{R}^n$ are constructed as above. Then,*

1. $\delta_{i+1} - \delta_i = \frac{1}{K-i+1}(\delta_{i+1} - \bar{x}_i) = \frac{1}{K-i}(\delta_i - \bar{x}_i)$ *for* $i = 1, \ldots, K-1$.

2. $\delta_1 \geq \cdots \geq \delta_m$ *and* $\delta_m \leq \cdots \leq \delta_K$.

3. $\bar{u} \geq_m \bar{x}$.

4. $\bar{u} \in F \cap \Delta$.

Proof For $i = 1, \ldots, K-1$,

$$\begin{aligned}
\delta_{i+1} - \delta_i &= \frac{1}{K-i}\sum_{k=i+1}^{n} \bar{x}_k - \frac{1}{K-i+1}\sum_{k=i}^{n} \bar{x}_k \\
&= \frac{1}{(K-i)(K-i+1)}\left\{(K-i+1)\sum_{k=i+1}^{n} \bar{x}_k - (K-i)\sum_{k=i}^{n} \bar{x}_k\right\} \\
&= \frac{1}{(K-i)(K-i+1)}\left\{\sum_{k=i+1}^{n} \bar{x}_k - (K-i)\bar{x}_i\right\} \\
&= \frac{1}{K-i+1}(\delta_{i+1} - \bar{x}_i) \\
&= \frac{1}{(K-i)(K-i+1)}\left\{\sum_{k=i}^{n} \bar{x}_k - (K-i+1)\bar{x}_i\right\} \\
&= \frac{1}{K-i}(\delta_i - \bar{x}_i).
\end{aligned}$$

The first part follows. For the second part, observe that $\delta_{i+1} \geq \delta_i$ implies that $\delta_{i+1} \geq \bar{x}_i \geq \bar{x}_{i+1}$ by the first part and hence $\delta_{i+2} \geq \delta_{i+1}$ is obtained by the first part. Similarly, $\delta_{i+1} \leq \delta_i$ implies that $\delta_i \leq \bar{x}_{i-1}$ and hence $\delta_i \leq \delta_{i-1}$. This proves the second part.

We next show that $\bar{u}$ majorizes $\bar{x}$. Observe that $\bar{x}_{m-1} \geq \delta_m$ because of the fact $\delta_m \leq \delta_{m-1}$ and part 1. This shows that $\bar{u}_i, i = 1, \ldots, n$ are nonincreasing. By construction, $\sum_{k=1}^{n} \bar{u}_k = \sum_{k=1}^{n} \bar{x}_k$ and $\sum_{k=1}^{j} \bar{u}_k = \sum_{k=1}^{j} \bar{x}_k$ for each $j = 1, \ldots, m-1$. By part 1, $\delta_{m+1} \geq \delta_m$ implies that $\delta_m \geq \bar{x}_m \geq \cdots \geq \bar{x}_K$. This shows that $\bar{u} \geq_m \bar{x}$. Since $\bar{u} \in \Delta$ and $\mathrm{card}(\bar{u}) \leq K$ by its construction, it remains to show that $\|\bar{u}\| \leq 1$. Define $\bar{\gamma} \in \mathbb{R}^n$ as follows:

$$\bar{\gamma}_i = \begin{cases} \bar{u}_i/\|\bar{u}\|, & i \in \{1, \ldots, K\}, \\ \bar{u}_K/\|\bar{u}\|, & i \in \{K+1, \ldots, n\}. \end{cases}$$

Now, consider an optimal solution γ to (3.6) with respect to $t = \bar{x}$ and define $\bar{\gamma}$ as $\bar{\gamma}_i = \gamma_i$ for $i = 1, \ldots, K-1$ and $\bar{\gamma}_i = \gamma_K$ for $i = K, \ldots, n$. Observe that

$$\begin{aligned}
\sum_{i=1}^{n} \bar{x}_i \bar{\gamma}_i &= \sum_{i=1}^{K-1} \bar{x}_i \bar{\gamma}_i + \left(\sum_{i=K}^{n} \bar{x}_i \right) \bar{\gamma}_K \\
&= \frac{1}{\|\bar{u}\|} \left\{ \sum_{i=1}^{K} \bar{x}_i \bar{u}_i + \left(\sum_{i=K+1}^{n} \bar{x}_i \right) \bar{u}_K \right\} \\
&= \frac{1}{\|\bar{u}\|} \left\{ \sum_{i=1}^{m-1} \bar{x}_i \bar{u}_i + \sum_{i=m}^{K} \bar{x}_i \bar{u}_i + \left(\sum_{i=K+1}^{n} \bar{x}_i \right) \bar{u}_K \right\} \\
&= \frac{1}{\|\bar{u}\|} \left\{ \sum_{i=1}^{m-1} \bar{u}_i^2 + \left(\sum_{i=m}^{K} \bar{x}_i \right) \delta_m + \left(\sum_{i=K+1}^{n} \bar{x}_i \right) \delta_m \right\} \\
&= \frac{1}{\|\bar{u}\|} \left\{ \sum_{i=1}^{m-1} \bar{u}_i^2 + \left(\sum_{i=m}^{n} \bar{x}_i \right) \delta_m \right\} \\
&= \frac{1}{\|\bar{u}\|} \left\{ \sum_{i=1}^{m-1} \bar{u}_i^2 + (K-m+1)\delta_m^2 \right\} \\
&= \frac{1}{\|\bar{u}\|} \sum_{i=1}^{K} \bar{u}_i^2 = \|\bar{u}\|,
\end{aligned}$$

where the fourth equality holds by the definition of $\bar{u}$ and the sixth equality holds because of the definition of δ_m as $\frac{1}{K-m+1} \sum_{i=m}^{n} \bar{x}_i$. Since $x^* \in \mathrm{conv}(F)$ and $\bar{\gamma}$ is feasible for (3.6),

$$\sum_{i=1}^{n} \bar{x}_i \bar{\gamma}_i = \sum_{i=1}^{K-1} |x^*|_{\sigma(i)} \bar{\gamma}_i + \left(\sum_{i=K}^{n} |x^*|_{\sigma(i)} \right) \bar{\gamma}_K \leq 1.$$

This shows that $\|\bar{u}\| \leq 1$. It follows $\bar{u} \in F \cap \Delta$. ■

Theorem 3.2.3 *Suppose x^*, σ, and u^* are defined as previously. Then, u^* is an optimal solution to* (3.2).

Proof Consider the separation problem

$$\max\{(x^*)^\intercal \beta : \beta^\intercal x \leq 1 \text{ is valid for } F\}. \tag{3.15}$$

It is clear that the coefficient vector of any valid inequality which passes through x^* is an optimal solution to the separation problem and vice-versa. Let β^* be an optimal

solution to (3.15). Then, By Proposition 3.2.2, $|\beta^*|_{\sigma(1)} \geq \cdots \geq |\beta^*|_{\sigma(n)}$ and $x_i^*\beta_i^* \geq 0$ for $i = 1, \ldots, n$. Since $u^* \geq_m x^*$, we can write $x^* = \sum_{P\in\mathcal{P}} \lambda_P(Pu^*)$ where $\lambda_P \geq 0$ and $\sum_{P\in\mathcal{P}} \lambda_P = 1$. Suppose $(\beta^*)^\intercal x^* > (\beta^*)^\intercal(Pu^*)$ for all permutation matrices $P \in \mathcal{P}$. Then,

$$(\beta^*)^\intercal x^* = (\beta^*)^\intercal \sum_{P\in\mathcal{P}} \lambda_P(Pu^*) = \sum_{P\in\mathcal{P}} \lambda_P(\beta^*)^\intercal(Pu^*) < \sum_{P\in\mathcal{P}} \lambda_P(\beta^*)^\intercal x^* = (\beta^*)^\intercal x^*,$$

yielding a contradiction. Therefore, there exists $P \in \mathcal{P}$ such that $(\beta^*)^\intercal x^* = (\beta^*)^\intercal(Pu^*)$. But, by the rearrangement inequality, $(\beta^*)^\intercal u^* \geq (\beta^*)^\intercal(Pu^*)$ and hence $(\beta^*)^\intercal x^* = (\beta^*)^\intercal u^* = 1$. This implies that if x^* satisfies a valid inequality at equality then u^* also satisfies the inequality at equality. We next argue that x^* can be written as a convex combination of points including u^* with a positive coefficient. By the previous argument, we consider a minimal dimensional face F of $\mathrm{Perm}(u^*)$ that contains x^*. Since F is of minimum dimension, x^* is in its relative interior. It follows from Minkowski-Carathéodory Theorem (see Theorem 8.11 of [65], for example) that there exist extreme points $v^i, i = 1, \ldots, t$ of F such that

$$\begin{cases} x^* = \sum_{i=1}^t \lambda_i v^i, \\ \lambda_i > 0, \quad i = 1, \ldots, t, \\ \sum_{i=1}^t \lambda_i = 1, \\ u^* \in \{v^i : i = 1, \ldots, t\}. \end{cases}$$

Convexity of the objective function $f(x) = x^\intercal \Sigma x$ then implies that

$$f(x^*) = f\left(\sum_{i=1}^t \lambda_i v^t\right) \leq \sum_{i=1}^t \lambda_i f(v^i) \leq f(x^*).$$

Since all above inequalities must hold at equality and $\lambda_i > 0, i = 1, \ldots, t$, we conclude that $f(x^*) = f(v^i)$ for all $i = 1, \ldots, t$. Therefore, $(x^*)^\intercal \Sigma(x^*) = (u^*)^\intercal \Sigma(u^*)$. ■

3.3 SDP relaxation for sparse PCA

In the preceding sections, we have discussed how to reformulate sparse PCA as a non-convex QCQP (convex maximization problem) by characterizing the convex hull

of its feasible set, and by showing that an optimal solution that satisfies the cardinality requirement can be created from one that does not. In this section, we present an SDP relaxation for the reformulation and show that the relaxation generalizes the SDP relaxation proposed in [24], which is the tightest SDP relaxation in the literature.

For notational clarity, we regard $u \in \mathbb{R}^K$ and write $(u, 0)$ to represent the lifted vector in $\mathbb{R}^n$.

We first lift vector variables x and u to matrix variables $X \in S^n$ and $U \in S^K$, representing $xx^\intercal$ and $uu^\intercal$, respectively. We consider the relaxations $X \succeq 0$ and $U \succeq 0$. The constraint $\|u\| \leq 1$ can be imposed as $\mathrm{Tr}(U) \leq 1$ under the condition that $U = uu^\intercal$.

Recall from Theorem 3.2.1 that $\mathrm{conv}(F) = \{x : u \in F \cap \Delta, u \geq_m |x|\}$. Suppose x^* is an optimal solution to sparse PCA. Define u^* by $u^*_i = |x^*|_{[i]}$ for $i \in \{1, \ldots, K\}$. Then, it is clear that $(u^*, 0) \in F \cap \Delta$ and $(u^*, 0) \geq_m |x^*|$. Therefore, we can narrow down our focus on vectors u and x such that $|x|$ is a permutation of $(u, 0) \in F \cap \Delta$. Hence, any convex constraint implied under this premise can be imposed. In matrix variable space, we can assume that $|X|$ equals to $\begin{bmatrix} U & 0 \\ 0 & 0 \end{bmatrix}$ after permuting rows and columns appropriately. It is clear that such U and X satisfy that $\mathrm{Tr}(U) = \mathrm{Tr}(|X|)$ and $\mathbb{1}^\intercal U \mathbb{1} = \mathbb{1}^\intercal |X| \mathbb{1}$. When $X \succeq 0$, we have that $\mathrm{Tr}(|X|) = \mathrm{Tr}(X)$. Therefore,

$$\mathrm{Tr}(U) = \mathrm{Tr}(X) \tag{3.16}$$

can be imposed. The equality $\mathbb{1}^\intercal U \mathbb{1} = \mathbb{1}^\intercal |X| \mathbb{1}$, however, is nonconvex and hence we impose the convex relaxation

$$\mathbb{1}^\intercal U \mathbb{1} \geq \mathbb{1}^\intercal |X| \mathbb{1}. \tag{3.17}$$

The above discussion yields the following SDP relaxation:

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}(\Sigma X) \\ \text{subject to} & \mathrm{Tr}(U) \leq 1, \\ & \mathrm{Tr}(U) = \mathrm{Tr}(X), \\ & \mathbb{1}^\intercal U \mathbb{1} \geq \mathbb{1}^\intercal |X| \mathbb{1}, \\ & X \succeq 0, U \succeq 0, \\ & X \in S^n, U \in S^K. \end{array} \tag{3.18}$$

Suppose (X^*, U^*) is a feasible solution to (3.18) and assume that $\mathrm{Tr}(U^*) < 1$. Then, we can scale U^* and X^* by a positive scalar $\lambda > 1$ so that $\mathrm{Tr}(\lambda U^*) = 1$, while still satisfying all constraints of (3.18). Further, $\mathrm{Tr}(\Sigma(\lambda X^*)) = \lambda \mathrm{Tr}(\Sigma X^*) > \mathrm{Tr}(\Sigma X^*)$. We conclude that optimal solutions (X^*, U^*) to (3.18) satisfy $\mathrm{Tr}(U^*) = 1$. Consequently, we obtain the following SDP relaxation:

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}(\Sigma X) \\ \text{subject to} & \mathrm{Tr}(U) = 1, \\ & \mathrm{Tr}(U) = \mathrm{Tr}(X), \\ & \mathbb{1}^\intercal U \mathbb{1} \geq \mathbb{1}^\intercal |X| \mathbb{1}, \\ & X \succeq 0, U \succeq 0, \\ & X \in S^n, U \in S^K. \end{array} \tag{3.19}$$

We denote the feasible set of (3.19) by F_{basic}.

The most commonly used SDP relaxation for sparse PCA was introduced in [24] and is given by:

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}(\Sigma X) \\ \text{subject to} & \mathrm{Tr}(X) = 1, \\ & \mathbb{1}^\intercal |X| \mathbb{1} \leq K, \\ & X \succeq 0. \end{array} \tag{3.20}$$

We denote the feasible set of (3.20) by F_d.

Lemma 5 *Let A be an $n \times n$ matrix with $A_{ii} = a$ for all i and $A_{ij} = b$ for all $i \neq j$. Then, the eigenvalues of A are $a - b$ (with multiplicity $n - 1$) and $a + (n-1)b$.*

Proof It is known that eigenvalues of $J(n,n)$ are $n, 0, \ldots, 0$. Since $A = bJ(n,n) + (a-b)I$, the eigenvalues of A are $bn + (a-b), a-b, \ldots, a-b$. The result follows. ■

Proposition 3.3.1 $\mathrm{proj}_X F_{\mathrm{basic}} = F_d$.

Proof Let $(X, U) \in F_{\mathrm{basic}}$. Then, $\mathrm{Tr}(X) = \mathrm{Tr}(U) = 1$ and $X \succeq 0$. We next show that $\mathbb{1}^\intercal |X| \mathbb{1} \leq K$. Consider the function $p(x) = x^\intercal U x$. Then, by the convexity of p, we have that

$$\begin{aligned} \mathbb{1}^\intercal |X| \mathbb{1} &\leq \mathbb{1}^\intercal U \mathbb{1} = p(\mathbb{1}) = K^2 p\left(\frac{1}{K}\mathbb{1}\right) = K^2 p\left(\sum_{i=1}^{K} \frac{1}{K} e_i\right) \leq K^2 \frac{1}{K} \sum_{i=1}^{K} p(e_i) \\ &= K\mathrm{Tr}(U) \leq K. \end{aligned}$$

Next, suppose $X \in F_d$. Define U by $U_{ii} = \frac{1}{K}$ for $i = 1, \ldots, K$ and $U_{ij} = \frac{\mathbb{1}^\intercal |X| \mathbb{1} - 1}{K(K-1)}$ for $i \neq j$. Observe that $U \geq 0$. Furthermore, $\frac{1}{K} \geq \frac{\mathbb{1}^\intercal |X| \mathbb{1} - 1}{K(K-1)}$ because $\mathbb{1}^\intercal |X| \mathbb{1} \leq K$. By construction, $\mathrm{Tr}(U) = 1 = \mathrm{Tr}(X)$ and $\mathbb{1}^\intercal U \mathbb{1} = \mathbb{1}^\intercal |X| \mathbb{1}$. It remains to show that $U \succeq 0$. Observe that U is of the form of A in Lemma 5 and hence the eigenvalues of U are $a + (K-1)b, a-b, \ldots, a-b$ where $a = \frac{1}{K}$ and $b = \frac{\mathbb{1}^\intercal |X| \mathbb{1} - 1}{K(K-1)}$. Since all its eigenvalues are nonnegative, we have that $U \succeq 0$. ■

We next construct additional inequalities to tighten the feasible set of (3.19). Since u is nonnegative and in nonincreasing order, $uu^\intercal$ is nonnegative and each of its row or column is in nonincreasing order. Therefore, we can impose the constraints:

$$\begin{aligned} &U \geq 0, \\ &U_{i,j} \geq U_{i,j+1}, \qquad i \in \{1, \ldots, K\}, j \in \{1, \ldots, K-1\}, \\ &U_{i,j} \geq U_{i+1,j}, \qquad i \in \{1, \ldots, K-1\}, j \in \{1, \ldots, K\}. \end{aligned} \tag{3.21}$$

Given matrix X, denote the jth largest component of the ith row of X by $X_{i,[j]}$. We model $|X|$ by replacing it to Y and adding constraints $Y = V + W$ and $X = V - W$ for nonnegative matrix variables V and W. Under the premise that U equals to $|X| (= Y)$ after permuting rows and columns appropriately, we can impose the following constraints:

Diagonal majorization: It is clear that we can impose

$$\operatorname{diag}(U) \geq_m \operatorname{diag}(Y), \tag{3.22}$$

which we call *diagonal majorization.*

Upper-sum majorization: For each $p \in \{1, \ldots, n\}$, define $\overline{U}^p \in \mathbb{R}^n$ and $\overline{Y}^p \in \mathbb{R}^n$ as follows:

$$\begin{aligned} \overline{U}_i^p &:= \textstyle\sum_{j=1}^p U_{i,[j]} = \sum_{j=1}^p U_{i,j}, \\ \overline{Y}_i^p &:= \textstyle\sum_{j=1}^p Y_{i,[j]} \end{aligned}$$

under the constraints (3.21). Then, we impose inequalities

$$\overline{U}^p \geq_m \overline{Y}^p, \quad p = 1, \ldots, n \tag{3.23}$$

which we call *upper-sum majorization inequality.*

Observe that each $\overline{Y}_i^p$ is the sum of p-largest components of ith row of Y. It can therefore be modeled using the technique used in Theorem 3.2.2. In particular, each inequality in (3.23) can be decomposed into $n-1$ inequalities and one equality. We only need to model the right-hand-sides of the inequalities since components of U are already in nonincreasing order. Since each of the right-hand-sides of the inequalities is a sum of q-largest components of the vector $(\overline{Y}_1^p, \ldots, \overline{Y}_n^p)$ for some $q = 1, \ldots, K-1$, it can be modeled in a similar fashion. As a special case of upper-sum majorization, consider the case where $p = n$,

$$\overline{U}^n \geq_m \overline{Y}^n. \tag{3.24}$$

While upper-sum majorization inequalites for $p = 1, \ldots, K-1$ require applying the modeling technique twice, only one step of modeling is needed when $p = n$. We call this constraint *row-sum majorization inequality.* We refer to the relaxation (3.19) with additional constraints (3.21) and (3.22) as *diagonal relaxation* and denote its feasible set as F_{diag}. The feasible set of the SDP relaxation obtained by replacing the diagonal majorization inequality with the row-sum majorization inequality are denoted by F_{rowsum}. After imposing all the constraints, we obtain an SDP relaxation which we call (mSDP). We refer to its feasible set as F_{mSDP}.

We next argue that F_{diag} and F_{rowsum} are proper subsets of F_d after projection.

Lemma 6 *For any $U \in S^K$, suppose $\text{Tr}(U) \leq 1, \mathbb{1}^\intercal U \mathbb{1} \geq K$, and $U \succeq 0$. Then, $U = \frac{1}{K} J(K, K)$.*

Proof For any square matrix X, we denote the vector of eigenvalues in nonincreasing order by $\lambda(X)$ Then,

$$K \leq \mathbb{1}^\intercal U \mathbb{1} = \text{Tr}(UJ(K,K)) \leq \lambda(U)^\intercal \lambda(J(K,K)) \leq K\lambda(U)_1 \leq K$$

where the second inequality is from Fan's inequality (see Theorem 1.2.1 of [16], for example), the third inequality is from the fact that the eigenvalues of $J(1,1)$ are $K, 0, \ldots, 0$, and the last inequality is from the conditions $\text{Tr}(U) \leq 1$ and $U \succeq 0$. It follows that the eigenvalues of U are $1, 0, \ldots, 0$. Further, the second inequality holds at equality if and only if there exists an orthogonal matrix V such that $U = V^\intercal \text{diag}(\lambda(U))V$ and $J(K,K) = V^\intercal \text{diag}(\lambda(J(K,K))V$. Since $\text{diag}(\lambda(U)) = \frac{1}{K}\text{diag}(\lambda(J(K,K))$, then $U = \frac{1}{K} J(K,K)$. ■

Proposition 3.3.2 $\text{proj}_X F_D \subsetneq \bar{F}_d$ *and* $\text{proj}_X F_R \subsetneq \bar{F}_d$.

Proof We consider the following system

$$\begin{cases} x_1 + \cdots + x_n = \sqrt{K}, \\ x_1^2 + \cdots + x_n^2 = 1, \\ x_i \geq 0, \quad i = 1, \ldots, n. \end{cases}$$

It is clear that all the permutations of u defined as $u_i = \frac{1}{\sqrt{K}}$ for $i = 1, \ldots, K$ and $u_i = 0$ for $i = K+1, \ldots, n$ are solutions to the system. Define $w(c) \in \mathbb{R}^n$ as $w(c)_1 = c$ and $w(c)_i = \frac{1-c}{n-1}$ for $i = 2, \ldots, n$. Consider the strictly concave function $g(x) = \sum_{i=1}^n \sqrt{x_i}$. Observe that $g(w(1)) = 1$ and $g(w(\frac{1}{n})) = \sqrt{n}$. By the Intermediate Value Theorem, there exists $c' \in (\frac{1}{n}, 1)$ such that $g(w(c')) = \sqrt{K}$. By taking the component-wise square root of $w(c')$, we obtain a solution x to the system that is not a permutation of u. Define $X = xx^\intercal$. Suppose there exists U such that

$(X, U) \in F_{\text{basic}}$ and $(\text{diag}(U), 0) \geq_m \text{diag}(|X|)$. By Lemma 6, $U = \frac{1}{K}J(K, K)$. Then, $\text{diag}(|X|) \in \text{Perm}(\text{diag}(U))$. Moreover, $\text{diag}(|X|)$ is not a permutation of $\text{diag}(U)$. Therefore, by strict concavity of g and the fact that $g(P\text{diag}(U)) = \sqrt{K}$ for all permutation matrices P, $\sqrt{K} < g(\text{diag}(|X|)) = x_1 + \cdots + x_n = \sqrt{K}$, yielding the desired contradiction.

For Part 2, we consider the same matrix U and X described in the proof of the first part and write them as $uu^\intercal$ and $xx^\intercal$ respectively. Suppose that they satisfy the row-sum majorization. By row-sum majorization and the fact that U and X are of rank 1, we have that $(\sum_{i=1}^n u_i)u \geq_m (\sum_{i=1}^n x_i)x$. This implies that $u \geq_m x$, concluding that x is a convex combination of u and its permutations. consider the strictly convex function $h(x) = \|x\|^2$. Since x is not a permutation of u and $h(Pu) = 1$ for all permutations P, $1 = h(u) > h(x) = 1$, providing a contradiction. ■

We next present an illustrative example in $\mathbb{R}^3$ for which our SDP relaxation returns the global optimal solution to the sparse PCA while (3.20) does not.

Example 7 *For* $\Sigma = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}$ *and consider the following sparse PCA*

$$\begin{array}{ll} \text{maximize} & x^T \Sigma x \\ \text{subject to} & \|x\| \leq 1, \\ & \text{card}(x) \leq 2 \end{array}$$

A global optimal solution can be obtained by finding the 2×2 *principal submatrix which maximizes the leading eigenvalue. Then, the eigenvector of the optimal principal submatrix gives the optimal solution to the sparse PCA and the optimum is the maximal leading eigenvalue. It can be verified that the optimum is attained when* $x_3 = 0$ *(or* $x_2 = 0$*) with optimal solution* $\begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \end{bmatrix}^\intercal$ *and the optimal value is* $z_E^* = 5$*. The optimal solution corresponding to the SDP relaxation* (3.20) *is*

$$X_D^* = \begin{bmatrix} 8/9 & 2/9 & 2/9 \\ 2/9 & 1/18 & 1/18 \\ 2/9 & 1/18 & 1/18 \end{bmatrix} = (x_D^*)(x_D^*)^\intercal$$

where $x_D^* = \begin{bmatrix} \frac{2\sqrt{2}}{3} & \frac{\sqrt{2}}{6} & \frac{\sqrt{2}}{6} \end{bmatrix}^\intercal$. *While the optimal solution satisfies the rank-1 constraint, the corresponding optimal value is* $z_D^* = \mathrm{Tr}(\Sigma X_D^*) = 50/9 \approx 5.556$. *Observe that the optimal value of the separation problem* (3.6) *is* 1.0541 *which exceeds 1 and it indicates that* $x_D^* \notin \mathrm{conv}(F)$. *Furthermore, the optimal solution* $\gamma^* = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}^\intercal$ *to* (3.6) *corresponds to the cutting plane* $\beta^\intercal x \leq 1$ *in the original variable space where*

$$\beta = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}^\intercal.$$

The corresponding cut in the matrix space is

$$\mathrm{Tr}(BX) \leq 1 \tag{3.25}$$

where $B = \beta\beta^\intercal$. *In this example, the relaxation* (3.20) *with additional constraint* (3.25) *gives the optimal value* 5. *On the other hand, our row-sum relaxation returns the global optimal value* $z_{rowsum}^* = 5$. *The corresponding optimal solution is*

$$X_{rowsum}^* = \begin{bmatrix} 4/5 & 1/5 & 1/5 \\ 1/5 & 1/10 & 0 \\ 1/5 & 0 & 1/10 \end{bmatrix}.$$

3.4 Preliminary computational results

In this section, we report the results of preliminary computational tests with `CVX` 2.1 [33] for problems of small dimension. We compare the tightness of our SDP relaxations and of (3.20). To obtain global optimal solutions to our test problems, we implemented an exhaustive search algorithm that compares eigenvalues of all $\binom{n}{K}$ $K \times K$ submatrices.

3.4.1 `pitprops` problem

`pitprops` [39] is one of the most commonly used problems for sparse PCA algorithms. The instance has 13 variables and 180 observations.

For the sake of exposition, we report test results for the row-sum relaxation and upper-sum relaxation in Table 3.1 and 3.2. z_E^* represents the global optimal value

Table 3.1: Optimal values and gaps closed for the test problem `pitprops`

K	z_E^*	z_D^*	z_{rowsum}^*	Gap Closed (%)	z_{mSDP}^*	Gap closed (%)
3	2.475	2.522	2.495	57.86	[1]2.475	100.00
4	2.937	3.017	[1]2.967	62.83	[1]2.948	87.15
5	3.406	3.458	3.407	97.97	[1]3.406	100.00
6	3.771	3.814	3.771	100.00	[1]3.771	100.00
7	3.996	4.032	3.996	100.00	[1]3.996	100.00
8	4.069	4.145	4.073	94.22	[1]4.072	95.48
9	4.139	4.206	[1]4.139	100.00	[1]4.139	100.00
10	4.173	4.219	4.177	91.32	[1]4.177	91.41
			Average	88.025	Average	96.76

for the sparse PCA and z_D^* represents the optimal value for SDP relaxation (3.20). We denote the optimal value for the SDP relaxation with constraints (3.21), (3.22), and (3.24) by z_{rowsum}^*. The optimal value for the SDP relaxation after imposing all the constraints is denoted by z_{mSDP}^*. Table 3.1 shows the test results for cardinality $K = 3, \ldots, 10$. To measure the relative tightness of a relaxation when compared to (3.20), we calculate "gap closed" as

$$\left(\frac{z_D^* - z_{SDP}^*}{z_D^* - z_E^*}\right) \times 100.$$

where z_{SDP}^* is one of z_D^*, z_{rowsum}^*, and z_{mSDP}^*.

The output status `Inaccurate/Solved` indicates that `CVX` could not determine the solution within the default numerical tolerance, but returned a solution using a relaxed tolerance.

For this particular test problems, the relaxation with feasible set F_{rowsum} reduces the gaps of (3.20) by more than 88% and with F_{mSDP} reduces by more than 96% on average, returning global optimal solutions for some problems.

[1]"`Inaccurate/Solved`" `CVX` output

3.4.2 Test results with randomly generated matrices

We next report test results for randomly generated covariance matrices. Random matrices are generated as follows:

1. Choose a random integer $m \in \{1, \ldots, n\}$ for the number of nonzero eigenvalues of the matrix by setting $m = \lceil nU \rceil$ where $U \sim \mathcal{U}(0, 1)$.

2. Generate m random vectors $v_i \in \mathbb{R}^n \sim \mathcal{N}(0, I_n), i = 1, \ldots, m$ for rank-1 matrices.

3. Generate m positive random eigenvalues $\lambda_i \sim \mathcal{U}(0, 1), i = 1, \ldots, m$.

4. Then, construct the desired random covariance matrix as $\Sigma = \sum_{i=1}^{m} \lambda_i v_i v_i^\intercal$.

The tests are performed for problems with size $n \in \{4, \ldots, 10\}$ and cardinalities $K \in \{2, \ldots, \lfloor n/3 \rfloor\}$. Note that the reported results are based on the test problems with `CVX` outputs status "`Solved`" or "`Inaccurate/Solved`". See Table 3.2. We observe that our SDP relaxations improve the gaps of the SDP relaxation (3.20) by more than 90% (on average).

3.5 Conclusion

Sparse principal component analysis was introduced as a way to resolve interpretability issues in principal component analysis and has received considerable attention by researchers in machine learning, statistics, and optimization. This problem is known to be NP-hard and the main difficulty resides in the cardinality constraint which allows for only a certain number of loadings to be nonzero. In this chapter, we derive the convex hull of an optimization formulation for finding the first sparse principal component by considering the separation problem and its dual. The convex hull is written through a majorization inequality which can be modeled using linear inequalities in a higher dimensional space. The majorization inequality allows us to

Table 3.2: Test results for randomly generated covariance matrices

			Average gap closed (%)	
n	K	# Test Problems	z^*_{rowsum}	z^*_{mSDP}
4	2	100	94.993	95.459
5	2	100	94.184	96.689
6	2	100	91.454	95.163
7	2	50	88.892	93.179
7	3	50	90.285	93.086
8	2	50	88.689	92.481
8	3	20	93.434	95.053
9	2	20	87.928	94.963
9	3	20	78.115	87.835
10	2	20	75.478	85.015
10	3	20	85.036	88.827
10	4	20	77.327	81.311
		Overall Average	90.180	93.559

interpret each point of the convex hull as a convex combination of points that satisfy the cardinality constraint. Furthermore, we show that the relaxation obtained by replacing the feasible set by its convex hull is a reformulation of the problem by showing that, for any optimal solution of the relaxation that does not satisfy the cardinality constraint, we can recover an equivalent optimal solution that satisfies the cardinality constraint. We next study a SDP relaxation. Under the fact that optimal solutions X can be written as $xx^\intercal$ where $\text{card}(x) \leq K$, we derive cuts which represent the natural majorization relationship between X and the sorted version of X. In particular, we derive diagonal majorization and upper-sum majorization. Our preliminary computational results show that the gaps of our SDP relaxation are more than 90% (on average) smaller than those of (3.20).

4. Facial disjunctive programming formulation and generalized RLT for cardinality constrained linear programming

In this chapter, we study convexification techniques for linear programs with a cardinality constraint. A facial disjunctive program formulation is developed to construct a finitely convergent cutting plane algorithm. We also use a ratio of multilinear terms as product factors to generalize the reformulation-linearization technique (RLT) to problems with a cardinality constraint. Using this approach we develop relaxation schemes that converge to the convex hull of solutions when the feasible region is compact. We then develop valid inequalities for the feasible set of cardinality-constrained knapsack problems based on disjunctive equivalents of the cardinality constraint.

4.1 Introduction

We study the optimization problem:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ & 0 \leq x \leq 1 \\ & \text{card}(x) \leq K \end{aligned} \tag{4.1}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^m \times \mathbb{R}^n$, and $b \in \mathbb{R}^m$. Observe that (4.1) is trivial when $K = 1$ or $K = n$. Therefore, we assume that $1 < K < n$. de Farias and Nemhauser [26] derived lifted inequalities for the case where $m = 1$, where (4.1) is called the *cardinality-constrained knapsack problem* (CCKP). They showed that CCKP is NP-hard.

Typically, (4.1) is modeled as a 0-1 MILP by introducing auxiliary binary variables:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \\ & 0 \leq x \leq z \\ & \mathbb{1}^\intercal z \leq K \\ & z \in \{0,1\}^n. \end{aligned} \tag{4.2}$$

Although generic MILP solvers can be used to solve (4.2), it is often desirable to construct formulations without using integer variables. Some potential benefits of this approach are discussed in [26].

In this chapter, we model (4.1) as a *facial disjunctive program*. This particular structure enables us to construct a finitely convergent cutting-plane algorithm using the seminal work by Jeroslow [40].

4.2 Facial Disjunctive Program Formulation

4.2.1 Formulation and sequential convexification

A *disjunctive set* is a set of points satisfying inequalities connected by $\wedge$ (conjunctions) and $\vee$ (disjunctions). A *disjunctive programming* is an optimization problem with linear objective whose feasible set is a disjunctive set. Any disjunctive set has an equivalent *conjunctive normal form* $\left\{x \in \mathbb{R}^n : \bigwedge_{h=1}^{q} \left(\bigvee_{j \in J_h} (d^j x \geq d_0^j)\right)\right\}$. For a given polyhedron F_0, a disjunction $\bigvee_{j \in J} (d^j x \geq d_0^j)$ is called *facial* with respect to F_0 if $F_0 \cap \{x \in \mathbb{R}^n : d^j x \geq d_0^j\}$ is a face of F_0, for all $j \in J$. A disjunctive program is called *facial* if all the disjunctions of the conjunctive normal form of the feasible set are facial with respect to F_0. Balas [3, 4] showed that the convex hull of the set $F = \left\{x \in F_0 : \bigwedge_{h=1}^{q} \left(\bigvee_{j \in J_h} (d^j x \geq d_0^j)\right)\right\}$ can be obtained by sequentially imposing

disjunctions in the conjunctive normal form on F_0 if F is facial. That is, if we define $S_0 = F_0$ and

$$S_h := \operatorname{conv}\left(S_{h-1} \cap \left\{x : \bigvee_{j \in J_h} (d^j x \geq d_0^j)\right\}\right).$$

Then $S_q = \operatorname{conv}(F)$. We refer to [4] for more details about disjunctive programs.

We next reformulate (4.1) as the following optimization problem

$$\begin{array}{ll} \text{maximize} & c^\intercal x \\ \text{subject to} & Ax \leq b \\ & 0 \leq x \leq 1 \\ & \prod_{j \in J} x_j = 0, \ \forall J \in \mathcal{J}_{K+1}, \end{array} \tag{4.3}$$

where $\mathcal{J}_i := \{A \subseteq \{1, \ldots, n\} : |A| = i\}$ for $i = 1, \ldots, n$.

Problem (4.3) is a facial disjunctive program because the constraint $\prod_{j \in J} x_j = 0$ can be written as $\bigvee_{j \in J}(x_j = 0)$ and it is a facial constraint because the constraints $x \geq 0$ are valid for $F_0 = \{x \in \mathbb{R}^n : Ax \leq b, 0 \leq x \leq 1\}$. In the next proposition, we prove the equivalence of the two formulations.

Proposition 4.2.1 (4.1) *and* (4.3) *are equivalent.*

Proof Suppose that $\operatorname{card}(x) \leq K$. Then, for any choice of a set of $K+1$ components of x, there exists at least one zero component. This proves that $\prod_{j \in J} x_j = 0$ for all $J \in \mathcal{J}_{K+1}$. Conversely, suppose that x satisfies $\prod_{j \in J} x_j = 0$ for all $J \in \mathcal{J}_{K+1}$ and that $\operatorname{card}(x) > K$. By choosing an index set $J \in \mathcal{J}_{K+1}$ in the support of x, we have $\prod_{j \in J} x_j \neq 0$ which yields the desired contradiction. ■

4.2.2 Finitely convergent cutting plane algorithm

In this section, we propose a finitely convergent cutting plane algorithm to solve (4.1). Recall that

$$\begin{aligned} F_0 &= \{x \in \mathbb{R}^n : Ax \leq b, 0 \leq x \leq 1\}, \\ F &= F_0 \cap \left(\bigcap_{J \in \mathcal{J}_{K+1}} \{x : \textstyle\prod_{j \in J} x_j = 0\}\right). \end{aligned}$$

For each $J \in \mathcal{J}_{K+1}$, define $F_J := \{x \in F_0 : \prod_{j \in J} x_j = 0\}$. Since F_j is a union of polyhedra, its convex hull can be obtained in a higher dimensional space using disjunctive programming [3, 5]. Let $\bar{A} = [A^\intercal \ \ -I \ \ I]^\intercal$ and $\bar{b} = [b \ \ \mathbb{0} \ \ \mathbb{1}]$. For a given polyhedron $B = \{x : \bar{A}x \leq \bar{b}\}$, define

$$\begin{aligned} P_J^*(B) &= \{(\alpha, \beta) : \alpha = u_j^\intercal \bar{A} + u_j^0 e_j, \beta = u_j^0 \bar{b}, u_j \geq 0, \forall j \in J\} \\ P_J(B) &= \{x : \alpha^\intercal x \leq \beta, \ \forall (\alpha, \beta) \in P_J^*(B)\}. \end{aligned}$$

Suppose $t \in \operatorname{conv}(B \setminus \bigcup_{j \in J} \{x : x_j = 0\})$. Consider the following linear program:

$$\begin{aligned} \text{maximize} \quad & t^\intercal \alpha - \beta \\ \text{subject to} \quad & (\alpha, \beta) \in P_J^*(B) \cap S \end{aligned} \tag{4.4}$$

where S is a normalization set. An optimal solution to (4.4) defines a face of $\operatorname{conv}(B \cap \bigcup_{j \in J} \{x : x_j = 0\})$ and the corresponding inequality cuts off t. We denote the vertex set of a polyhedron P by $\operatorname{vert}(P)$.

A general cutting plane procedure to solve (4.1) is given below:

1: $G = \{x \in \mathbb{R}^n : Ax \leq b, 0 \leq x \leq 1\}$ and $t \in \operatorname{Arg\,max}\{c^\intercal x : x \in G\} \cap \operatorname{vert}(G)$
2: **while** $\operatorname{card}(t) > K$ **do**
3: Let $J \in \mathcal{J}_{K+1}$ be such that $\prod_{j \in J} t_j \neq 0$.
4: Let (α, β) be an extreme point optimal solution to (4.4) with respect to t and $P_J(G)$.
5: $G \leftarrow G \cap \{x : \alpha^\intercal x \leq \beta\}$.
6: $t \in \operatorname{Arg\,max}\{c^\intercal x : x \in G\}$.
7: **end while**

Algorithm 2: General cutting plane algorithm

The procedure of Algorithm 2 does not converge in finite time in general. Jeroslow [40] developed a cutting plane algorithm for facial disjunctive programs that terminates in finite time. One of the important insights of the procedure is that, one does

not use G in the cut-generating LP, but a superset, and the facial structure of the disjunction enables generate valid inequalities that cut off t and result in an algorithm that only performs a finitely number of iterations. Following the same idea as Jeroslow, we propose the following cutting plane algorithm. For use in this new algorithm, we use notations analogous to those of Balas [6]. We first label those $q := |\mathcal{J}_{K+1}|$ index sets in $\mathcal{J}_{K+1}$ in a certain order $\leq^L$:

$$J_1 \leq^L J_2 \leq^L \ldots \leq^L J_q.$$

One may choose the lexicographical order, for instance. In each iteration of procedure, the current polyhedron G is defined by F_0 intersected with a set of half spaces corresponding to the cuts introduced so far. For $j = 1, \ldots, q$, a cut that appears in the definition of G is called *k-cut* if it was generated as a cut using the kth disjunction, $\bigvee_{j \in J_k}(x_j = 0)$. Let G_k be F_0 intersected by half spaces corresponding to i-cuts for $i = 1, \ldots, k$. We define $G_0 = F_0$. Then, the following algorithm returns an optimal solution to (4.1) in a finite number of iterations.

$G = \{x \in \mathbb{R}^n : Ax \leq b, 0 \leq x \leq 1\}$ and $t \in \operatorname{Arg\,max}\{c^\intercal x : x \in G\} \cap \operatorname{vert}(G)$
while $\operatorname{card}(t) > K$ **do**
    Let $J_k \in \operatorname{Arg\,max}\{J \in \mathcal{J}_{K+1} : \prod_{j \in J} t_j \neq 0\}$ under the order $\leq^L$.
    Let (α, β) be an extreme point optimal solution to (4.4) with respect to t and $P_{J_k}(G_{k-1})$.
    $G \leftarrow G \cap \{x : \alpha^\intercal x \leq \beta\}$.
    $t \in \operatorname{Arg\,max}\{c^\intercal x : x \in G\}$.
end while

Algorithm 3: Finitely convergent cutting plane algorithm

Theorem 4.2.1 *Algorithm 3 finds an optimal solution to* (4.1) *in a finite number of iterations.*

Proof We first prove that t is a vertex of G_k in the beginning of each iteration. If $k = q$ then $G_k = G$, and hence t is a vertex of G_k. If we assume $k < q$, then t does not violate the qth constraint and hence $\prod_{j \in J_q} t_j = 0$. Therefore, we have

$$t \in G_{q-1} \cap \left\{ x : \prod_{j \in J_q} x_j = 0 \right\} \subseteq \operatorname{conv}\left(G_{q-1} \cap \left\{ x : \prod_{j \in J_q} x_j = 0 \right\} \right) \subseteq G_q.$$

The last inclusion holds because any q-cut is valid for $\operatorname{conv}(G_{q-1} \cap \{x : \prod_{j \in J_q} x_j = 0\})$ by its construction. This shows that t is a vertex of $G_{q-1} \cap \{x : \prod_{j \in J_q} x_j = 0\}$. Because of the facial structure, it is also a vertex of G_{q-1}. Consequently, by induction, we obtain, t is a vertex of G_k. Next, we show that $t \notin P_{J_k}(G_{k-1})$ to conclude that the cutting plane cuts of t. Suppose $t \in P_{J_k}(G_{k-1})$. Since $P_{J_k}(G_{k-1}) \subseteq G_k$ and $t \in \operatorname{vert}(G_k)$, $t \in \operatorname{vert}(P_{J_k}(G_{k-1}))$. This implies that t satisfies kth disjunction and it produces the desired contradiction. It remains to show that only a finite number of iterations is needed to obtain an optimal solution. To show this, we only need to prove that there are only finitely many k-cuts for $k = 1, \ldots, q$. To use induction, consider first $k = 1$. Since a 1-cut is generated by solving a linear program constructed from the disjunction $\bigvee_{j \in J_1}(x_j = 0)$ together with $G_0 = F_0$, all the possible 1-cuts correspond to the vertices of $P^*_{J_1}(B) \cap S$. Since this feasible set is independent of iteration steps, there exist only finitely many 1-cuts. Now assume that the number of i-cuts is finite for $i = 1, \ldots, k-1$. A k-cut is obtained by the disjunction $\bigvee_{j \in J_k}(x_j = 0)$ together with G_{k-1}. Since the number of i-cuts are finite for $i = 1, \ldots, k-1$ after a sufficient number of iterations, G_{k-1} will no longer be updated and a k-cuts will be obtained as a vertex of $P_{J_k}(G_{k-1})$. Thus, there are only finitely many k-cuts. This shows that the algorithm is finitely convergent. ■

4.3 Generalized Reformulation-Linearization Technique

4.3.1 Barycentric coordinates

Definition 4.3.1 *Let P be a polytope. Real valued functions $b_v : P \to \mathbb{R},\ v \in v(P)$ are called* barycentric coordinates *if*

1. *(non-negativity)* $b_v(x) \geq 0, \quad v \in v(P),\ x \in P,$

2. *(partition of unity)* $\sum_{v \in v(P)} b_v(x) = 1, x \in P,$

3. *(linear precision)* $\sum_{v \in v(P)} b_v(x) \cdot v = x.$

In short, barycentric coordinates are the coefficients of a convex combination of vertices of P that can be used to obtain $x \in P$.

Example 8 (Barycentric coordinates for the hyper cube in $\mathbb{R}^n$) *It is easy to show that n-dimensional unit cube has 2^n vertices, each of which corresponds to a subset of $J = \{1, \ldots, n\}$. Let $v_A = \sum_{i \in A} e_i$. Then $\{v_A \mid A \subseteq J\}$ is the vertex set of the hypercube. It is easy to show that barycentric coordinates are*

$$b_{v_A}(x) = \prod_{i \in A} x_i \prod_{i \in J \setminus A} (1 - x_i).$$

Barycentric coordinates are identical to the multipliers used in reformulation-linearization technique [64] of level-n.

Warren [70] developed explicit barycentric coordinates for general convex sets.

4.3.2 Inclusion certificates

Tawarmalani [67] defined the concept of inclusion certificates as a probability measure. We use a restricted definition in the context of disjunctive programming. Let $\Pi_1, \ldots, \Pi_p$ be convex sets. Suppose W is such that $\bigcup_{i=1}^p \Pi_i \subseteq W \subseteq \text{conv}(\bigcup_{i=1}^p \Pi_i)$. For any $x \in W$, a p-tuple of functions $(b_1(x), \ldots, b_p(x))$ is called an *inclusion certificate* if there exist $v_i(x) \in \Pi_i,\ i = 1, \ldots, p$ such that $b_i(x) \geq 0$, and $b_i(x),\ i = 1, \ldots, p$ are barycentric coordinates of $\text{conv}\{v_1(x), \ldots, v_p(x)\}$. Barycentric coordinates form a special case of inclusion certificates where each Π_i is either the empty set or a singleton. In general, inclusion certificates are not unique.

Example 9 *Let $\Pi_1 = \{x \in \mathbb{R}^2_+ : x_2 = 0\}$ and $\Pi_2 = \{x \in \mathbb{R}^2_+ : x_1 = 0\}$. It is clear that conv$(\bigcup_{i=1}^2 \Pi_i) = \mathbb{R}^2_+$. For any $x \in \mathbb{R}^2_+$, we define $b_i(x) = \frac{x_i}{x_1 + x_2}, i = 1, 2$ and*

$v_1(x) = (x_1 + x_2, 0)$ and $v_2(x) = (0, x_1 + x_2)$. It is clear that $\sum_{i=1}^n b_i(x) = 1$ and $b_i(x) \geq 0$. Moreover,

$$b_1(x)v_1(x) + b_2(x)v_2(x) = \frac{x_1}{x_1 + x_2}(x_1 + x_2, 0) + \frac{x_2}{x_1 + x_2}(0, x_1 + x_2) = (x_1, x_2) = x.$$

Thus, $(b_1(x), b_2(x))$ is an inclusion certificate.

For $m = 1, \ldots, n$ and $J \in \mathcal{J}_m$, define $\Pi_J = \{x \in \mathbb{R}^n_+ : x_j = 0, j \notin J\}$. Observe that $\bigcup_{J \in \mathcal{J}_m} \Pi_J = \{x \in \mathbb{R}^n : \mathrm{card}(x) \leq m\}$.

Proposition 4.3.1 *Define $b_J(x) = \frac{\prod_{j \in J} x_j}{\sum_{J \in \mathcal{J}_m} \prod_{j \in J} x_j}$. Then $(\{b_J(x) : J \in \mathcal{J}_m\})$ is an inclusion certificate for $\bigcup_{J \in \mathcal{J}_m} \Pi_J$.*

Proof Define $v_J(x)$ by

$$(v_J(x))_i = \begin{cases} 0 & \text{if } i \notin J \\ x_i + \frac{1}{m} \sum_{k \notin J} x_k & \text{if } i \in J. \end{cases}$$

Let $\mathcal{J}_m^i = \{J \in \mathcal{J}_m : i \in J\}$. Observe that each term of

$$\sum_{J \in \mathcal{J}_m^i} \prod_{j \in J \setminus \{i\}} x_j \sum_{k \notin J} x_k \tag{4.5}$$

is of the form $\prod_{j \in I} x_j$ where $I \in \mathcal{J}_m \setminus \mathcal{J}_m^i$. For any $I \in \mathcal{J}_m \setminus \mathcal{J}_m^i$, I can be written as $(I \setminus \{h\}) \cup \{h\}$ for $h \in I$. Let $J_h = (I \setminus \{h\}) \cup \{i\}$. Then $J_h \in \mathcal{J}_m^i$ and $\prod_{j \in I} x_j = (\prod_{j \in J_h \setminus \{i\}} x_j) x_h$. This shows that $\prod_{j \in I} x_j$ is one of the terms of (4.5). Since any $h \in I$ be chosen, there are $m = |I|$ identical terms that can be found in (4.5). This shows that

$$\sum_{J \in \mathcal{J}_m^i} \prod_{j \in J \setminus \{i\}} x_j \sum_{k \notin J} x_k = m \sum_{J \in \mathcal{J}_m \setminus \mathcal{J}_m^i} \prod_{j \in J} x_j.$$

Therefore,

$$\begin{aligned}
\sum_{J\in\mathcal{J}_m^i} b_J(x)(v_J(x))_i &= \sum_{J\in\mathcal{J}_m^i} \frac{\prod_{j\in J} x_j}{\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j}\left(x_i + \frac{1}{m}\sum_{k\notin J} x_k\right) \\
&= \left(\frac{1}{\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j}\right)\sum_{J\in\mathcal{J}_m^i}\left\{\prod_{j\in J} x_j\left(x_i + \frac{1}{m}\sum_{k\notin J} x_k\right)\right\} \\
&= \left(\frac{1}{\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j}\right)\left(\sum_{J\in\mathcal{J}_m^i} x_i\prod_{j\in J} x_j + \frac{1}{m}\sum_{J\in\mathcal{J}_m^i}\prod_{j\in J} x_j\sum_{k\notin J} x_k\right) \\
&= x_i\left(\frac{1}{\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j}\right)\left(\sum_{J\in\mathcal{J}_m^i}\prod_{j\in J} x_j + \frac{1}{m}\sum_{J\in\mathcal{J}_m^i}\prod_{j\in J\setminus\{i\}} x_j\sum_{k\notin J} x_k\right) \\
&= x_i\left(\frac{1}{\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j}\right)\left(\sum_{J\in\mathcal{J}_m^i}\prod_{j\in J} x_j + \sum_{J\in\mathcal{J}_m\setminus\mathcal{J}_m^i}\prod_{j\in J} x_j\right) \\
&= x_i\left(\frac{1}{\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j}\right)\sum_{J\in\mathcal{J}_m}\prod_{j\in J} x_j = x_i
\end{aligned}$$

This shows that $\sum_{J\in\mathcal{J}} b_J(x)v_J(x) = x$. ■

Inclusion certificates play an important role in disjunctive programming because, by specifying inclusion certificates and the corresponding vertex points of each disjunction, each point in the convex set can be specified. We first introduce a representation of the convex hull of a disjunctive set. Let $\Pi_i, i = 1, \ldots, p$ be convex sets. It follows from the definition of convex hull that

$$C := \operatorname{conv}\left(\bigcup_{i=1}^{p}\Pi_i\right) = \operatorname{conv}\left\{x : \begin{array}{l} x = \sum_{i=1}^{p}\lambda_i x_i \\ x_i \in \Pi_i,\ \ \forall i = 1, \ldots, p \\ \sum_{i=1}^{p}\lambda_i = 1 \\ \lambda_i \geq 0 \end{array}\right\}.$$

Balas [5] proved that, in the case where each Π_i is a polyhedron, the closed convex hull can be represented as a projection of a higher dimensional polyhedron. For general sets Π_i, we have the same result.

Proposition 4.3.2 *Define*

$$C' := \left\{ x : \begin{array}{ll} x = \sum_{i=1}^{p} \mathbf{z}_i, & \\ \mathbf{z}_i \in \lambda_i \Pi_i, & \forall i = 1, \ldots, p \\ \sum_{i=1}^{p} \lambda_i = 1, & \\ \lambda_i \geq 0, & \forall i = 1, \ldots, p \end{array} \right\}.$$

Then C' is convex and $C' = C$.

Proof Consider a convex combination $x = \sum_{i=1}^{p} \alpha_i x_i \in C$. That is, we have $x_i \in \Pi_i$, $\alpha_i \geq 0$, and $\sum_{i=1}^{p} \alpha_i = 1$. Define $\lambda_i = \alpha_i$ and $\mathbf{z}_i = \alpha_i x_i$. Then $\sum_{i=1}^{p} \mathbf{z}_i = \sum_{i=1}^{p} \alpha_i x_i = x$, $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \alpha_i = 1$, and $\lambda_i = \alpha_i \geq 0$. Moreover, $\mathbf{z}_i = \alpha_i x_i \in \alpha_i \Pi_i$ because $x_i \in \Pi_i$. This shows that $C \subseteq C'$. Conversely, let $x \in C'$ so that there exist $\mathbf{z}_i$ and $\lambda_i \geq 0$ such that $\sum_{i=1}^{p} \lambda_i = 1$, $\mathbf{z}_i \in \lambda_i \Pi_i$, and $x = \sum_{i=1}^{p} \mathbf{z}_i$. Let $I = \{i = 1, \ldots, p : \lambda_i \neq 0\}$. Define $x_i := \frac{\mathbf{z}_i}{\lambda_i}$ for $i \in I$. Then $x_i \in \Pi_i$. Observe that for $i \notin I$, $\mathbf{z}_i = 0$ because $\mathbf{z}_i \in 0 \cdot \Pi_i$. Therefore, $x = \sum_{i=1}^{p} \mathbf{z}_i = \sum_{i \in I} \mathbf{z}_i = \sum_{i \in I} \lambda_i x_i$. Thus $x \in \operatorname{conv}(\bigcup_{i \in I} \Pi_i) \subseteq \operatorname{conv}(\bigcup_{i=1}^{p} \Pi_i) = C$. ■

Theorem 4.3.1 *Let $(b_1(x), \ldots, b_p(x))$ be an inclusion certificate with vertex points $v_i(x)$, $i = 1, \ldots, p$ for $\bigcup_{i=1}^{p} \Pi_i$. Then*

$$C = \bar{C} := \{x : b_i(x) v_i(x) \in b_i(x) \Pi_i, i = 1, \ldots, p\}.$$

Proof Since $(b_1(x), \ldots, b_p(x))$ is an inclusion certificate, $v_i(x) \geq 0$ and $\sum_{i=1}^{p} v_i(x) = 1$. By setting $\lambda_i = b_i(x)$ and $\mathbf{z}_i = b_i(x) v_i(x)$, we can show that $\bar{C} \subseteq C'$. We next show that $C \subseteq \bar{C}$. By definition, for any $x \in C$, $v_i(x) \in \Pi_i$. Thus, $b_i(x) v_i(x) \in b_i(x) \Pi_i$ for any $i = 1, \ldots, p$ and hence $x \in \bar{C}$. Since $C = C'$, we have $C \subseteq \bar{C} \subseteq C' = C$ and thus $C = \bar{C}$. ■

The above theorem shows that a specific choice of inclusion certificate is enough to describe all points in the convex hull.

4.3.3 Convexification using generalized RLT

Tawarmalani [67] proved that inclusion certificates of disjunctive set can be used as product factors for convexifying a compact disjunctive set under some technical conditions. Based on the inclusion certificates in Proposition 4.3.1, we develop a generalized reformulation-linearization technique. Recall that

$$b_J(x) = \frac{\prod_{j\in J} x_j}{\sum_{I\in\mathcal{J}_K}\prod_{j\in I} x_j}, \quad J \in \mathcal{J}_K$$

are inclusion certificates for $x \in \mathbb{R}^n_+$ with support $\bigcup_{I\in\mathcal{J}_K} \Pi_I$.

The generalized reformulation-linearization technique is as follows:

Step 1. Reformulation Step: Multiply all the constraints for LP by $b_J(x), J \in \mathcal{J}_K$. In doing so, a list of rational inequalities is obtained.

Step 2. Linearization Step: For each $i = 1, \ldots, n$, linearize the rational inequalities by substituting new variables y_J for $b_J(x)$ and new variables $y(i, J)$ for $b_J(x)x_i$. Set $y(i, J) = 0$ if $i \notin J$. Impose additional constraints

$$\begin{cases} \sum_{J\in\mathcal{J}_K} y_J = 1 \\ \sum_{j\in J\in\mathcal{J}_K} y(j, J) = x_j, \ \ j \in N \end{cases}$$

Call the resulting polyhedron Y.

Step 3. Projection Step: Project Y onto the space of x-variables. Call the resulting polyhedron X.

Theorem 4.3.2 $X = \text{conv}(F)$.

Proof We first show that $X \supseteq \text{conv}(F)$. Since X is a polyhedron (hence convex), it suffices to show that $X \supseteq F$. Let $x^0 \in F$ and hence $\text{card}(x^0) \leq K$. Define

$$y^0_J = b_J(x^0), \quad y^0(j, J) = b_J(x^0)x^0_j, \quad j \in J \in \mathcal{J}_K.$$

Define $y(J) = [y(1,J)\dots y(n,J)]$ and $y^0(J) = [y^0(1,J)\dots y^0(n,J)]$. Suppose $LP = \{x : Dx \le d\}$. Then $Y = \{(x, y(J), yJ : J \in \mathcal{J}_K) : Dy(J) \le dy_J\}$. It suffices to show that $(x^0, y^0(J), y^0_J : J \in \mathcal{J}_K) \in Y$. Observe that

$$Dy^0(J) = Db_J(x^0)x^0 \le b_J(x^0)d = dy^0_J.$$

Thus, $X \supseteq \text{conv}(F)$.

We next show that $X \subseteq \text{conv}(F)$ Let $\Pi_J = LP \cap \{x : x_j = 0, j \notin J\}$ and define

$$C_1 = \left\{ x : \begin{array}{l} x = \sum_{J\in\mathcal{J}_K} z_J, \\ z_J \in \lambda_J \Pi_J, \quad J \in \mathcal{J}_K \\ \sum_{J\in\mathcal{J}_K} \lambda_J = 1, \\ \lambda_J \ge 0, \quad J \in \mathcal{J}_K \end{array} \right\},$$

Notice that C_1 deals with all generic forms of z_J and λ_J where $y(J)$ and y_J are a special case. Therefore, $X \subseteq C_1$. By Balas' lifting theorem [5], $C_1 = \text{conv}(\bigcup_{J\in\mathcal{J}_K} \Pi_J)$. Therefore, $X \subseteq \text{conv}(F)$. ■

4.4 Valid inequalities for cardinality constrained knapsack problem

In this section, we derive valid inequalities for the feasible set of the CCKP:

$$\begin{array}{ll} \text{maximize} & c^\intercal x \\ \text{subject to} & a^\intercal x \le b \\ & 0 \le x \le 1, \\ & \text{card}(x) \le K \end{array} \tag{4.6}$$

where $c, a, x \in \mathbb{R}^n$ and $b \in \mathbb{R}$. We let F be the feasible set of (4.6), LPS be its linear relaxation and $PS = \text{conv}(F)$ to be notationally consistent with [26]. Given any disjoint index sets I, J and any set $P \subseteq \mathbb{R}^n$, we define $P(I,J) = P \cap \{x : x_I = 0, x_J = \mathbb{1}\}$. de Farias and Nemhauser [26] proved that CCKP is NP-hard and derived valid and facet-defining inequalities for PS in the original space of variables. They used the notion of cover and cover inequality to derive these inequalities. Even though

they presented various explicit forms for facet-defining inequalities, many others are yet to be identified.

We propose a procedure to create a valid inequality for F directly from another valid inequality for F. Under certain conditions, the procedure can be shown to generate a facet-defining inequality from a different facet-defining inequality. This enables us to construct hierarchies of facet-defining inequalities and to build families of facet-defining inequalities for PS.

Given any vector x, we denote its ith largest component by $x_{[i]}$. If $\sum_{i=1}^{K} a_{[i]} \leq b$. Then, the cardinality constraint is redundant and (4.6) is a standard continuous knapsack problem. For this reason, we make the following assumption.

Assumption 3 $\sum_{i=1}^{K} a_{[i]} > b$.

4.4.1 Preprocessing

It is easy to show that an optimal solution to CCKP has at most one fractional component. It is also clear that if $c_i \geq c_j, a_i \leq a_j$, and $x_j^* \neq 0$, then $x_i^* = 1$. When $c_i \geq c_j$ and $a_i \leq a_j$ we say that i is *preferred* over j and we denote it by $i \gg j$. We refer to $i_1 \gg \cdots \gg i_k$ as a *preference chain* and define k to be the *length* of the chain. We list some properties of preference chains

1. If $i \gg j$ and $x_j^* > 0$, then $x_i^* = 1$.
2. If $i \gg j$ and $x_i^* < 1$, then $x_j^* = 0$.
3. If there exist at least K indices that are preferred to j, then $x_j^* = 0$.

Example 10 *Consider the following CCKP*

$$\begin{aligned} \text{maximize} \quad & 2x_1 + 3x_2 + x_3 + 4x_4 \\ \text{subject to} \quad & 10x_1 + 5x_2 + 4x_3 + x_4 \leq 7 \\ & \text{card}(x) \leq 2 \\ & 0 \leq x \leq 1. \end{aligned}$$

We observe that $c_4 > c_2 > c_1 > c_3$ *and* $a_1 > a_2 > a_3 > a_4$. *Thus,* $4 \gg 2 \gg 1$ *and* $4 \gg 3$. *This proves that* $x_1^* = 0$. *Furthermore, since* $x = 0$ *is not an optimal solution,* $x_4^* > 0$. *If* x_4^* *is fractional* (< 1) *then all other components of optimal solution must be zero. By projecting the original problem onto* x_4*-space, we obtain a trivial optimal solution* $x_4^* = 1$ *since* $a_4 \leq b$, *yielding a contradiction to the fact that* x_4^* *is fractional. Thus,* $x_4^* = 1$. *After projecting the problem over the space with* $x_4 = 1$ *and* $x_1 = 0$, *the CCKP reduces to*

$$\begin{array}{ll} \text{maximize} & 3x_2 + x_3 \\ \text{subject to} & 5x_2 + 4x_3 \leq 6 \\ & \text{card}(x) \leq 1 \\ & 0 \leq x \leq 1. \end{array}$$

4.4.2 Two-term disjunction method: δ-inequalities and its variants

Proposition 4.4.1 *For* $m = 0, 1, \ldots, K$, $\text{card}(x) \leq K$ *if and only if*

$$\text{card}(x_I) \leq m \quad \vee \quad \text{card}(x_{N \setminus I}) \leq K - m - 1$$

for all $I \subseteq N$.

Proof Suppose $\text{card}(x) \leq K$ and assume that $\text{card}(x_I) > m$ and $\text{card}(x_{N \setminus I}) > K - m - 1$ for some $I \subseteq N$. Since cardinality is integer, $\text{card}(x_I) \geq m + 1$ and $\text{card}(x_{N \setminus I}) \geq K - m$. It follows that $\text{card}(x) = \text{card}(x_I) + \text{card}(x_{N \setminus I}) \geq K + 1$. This contradicts the fact that $\text{card}(x) \leq K$.

Conversely, suppose $\text{card}(x_I) \leq m$ or $\text{card}(x_{N \setminus I}) \leq K - m - 1$ for all $I \subseteq N$ and assume by contradiction that $\text{card}(x) \geq K + 1$. Then there exists an index set I_0 such that $I_0 \subseteq \text{supp}(x)$ and $|I_0| = m + 1$ where $\text{supp}(x)$ represents the set of indices i with $x_i \neq 0$. Then $\text{card}(x_{N \setminus I_0}) \geq K - m$ yields the desired contradiction. ■

Thus, for a fixed $I \subseteq \{1, \ldots, n\}$, we can consider the disjunction to represent F as shown in Table 4.1. Let A_I and B_I be the sets corresponding to the disjunct 1 and

Table 4.1: Two-term disjunction for F

	Disjunct 1	Disjunct 2
cardinality constraint	$\text{card}(x_I) \leq m$	$\text{card}(x_{N\setminus I}) \leq K - m - 1$
knapsack inequality	$\sum_{j\in N} a_j x_j \leq b$	$\sum_{j\in N} a_j x_j \leq b$
box constraints	$x_i \leq 1, \quad \forall i = 1, \ldots, n$	$x_i \leq 1, \quad \forall i = 1, \ldots, n$
	$-x_i \leq 0, \quad \forall i = 1, \ldots, n$	$-x_i \leq 0, \quad \forall i = 1, \ldots, n$

2 respectively. For an index set I and a positive integer k, define a_I^k as

$$a_I^k := \begin{cases} (a_I)_{[1]} + \cdots + (a_I)_{[k]} & \text{if } k \leq |I| \\ \sum_{j\in I} a_j & \text{if } k > |I|. \end{cases}$$

For an index set $I \subseteq N$, the cardinality constraint $\text{card}(x_I) \leq m$ can be relaxed into the linear inequality $\sum_{j\in I} x_j \leq m$. Furthermore, for any valid inequality $\sum_{j\in I} f_j x_j \leq f_0$, it is easy to show that $\sum_{j\in I} f_j x_j \leq \min\{f_0, f_I^m\}$ is implied by $\sum_{j\in I} x_j \leq m$ and $0 \leq x_j \leq 1$ for $j \in I$. Consequently, disjunct 1 can be relaxed without the use of a cardinality constraint as

$$\begin{cases} \sum_{j\in I} x_j \leq m \\ \sum_{j\in N} a_j x_j \leq b \\ x_i \leq 1, \quad \forall i = 1, \ldots, n \\ -x_i \leq 0, \quad \forall i = 1, \ldots, n. \end{cases}$$

Let RA_I be the set corresponding to the relaxation. Therefore, any valid inequality for RA_I can be expressed as

$$\sum_{j\in I} (\alpha^A + \beta^A a_j + \delta_i^A - \epsilon_i^A) x_j + \sum_{j\in N\setminus I} (\beta^A a_j + \delta_i^A - \epsilon_i^A) x_j \leq \alpha^A m + \beta^A b + \sum_{i\in N} \delta_i^A$$

where all the multipliers are nonnegative.

Similarly, disjunct 2 can be relaxed without the use of a cardinality constraint as

$$\begin{cases} \sum_{j\in N\setminus I} x_j \leq K-m-1 \\ \sum_{j\in N} a_j x_j \leq b \\ x_i \leq 1, \quad \forall i = 1,\ldots,n \\ -x_i \leq 0, \quad \forall i = 1,\ldots,n. \end{cases}$$

Any valid inequality for RB_I can be expressed as

$$\sum_{j\in I}(\beta^B a_j+\delta_i^B-\epsilon_i^B)x_j+\sum_{j\in N\setminus I}(\alpha^B+\beta^B a_j+\delta_i^B-\epsilon_i^B)x_j \leq \alpha^B(K-m-1)+\beta^B b+\sum_{i\in N}\delta_i^B.$$

Therefore, based on the given disjunction, we have a generic form of valid inequalities for $PS \subseteq \text{conv}(RA_I \cup RB_I)$ whose coefficients are solutions to the following system of equations:

$$\begin{cases} \alpha^A+\beta^A a_j+\delta_i^A-\epsilon_i^A &= \beta^B a_j+\delta_i^B-\epsilon_i^B, \qquad j\in I \\ \beta^A a_j+\delta_i^A-\epsilon_i^A &= \alpha^B+\beta^B a_j+\delta_i^B-\epsilon_i^B, \qquad j\in N\setminus I \\ \alpha^A m+\beta^A b+\sum_{i\in N}\delta_i^A &= \alpha^B(K-m-1)+\beta^B b+\sum_{i\in N}\delta_i^B. \end{cases}$$

We pay particular attention to the case where $m = 0$ since $\text{card}(x_I) = 0$ implies that $\text{card}(x_{N\setminus I}) \leq K$ and this enables us to add extra inequality $\sum_{j\in N\setminus I} x_j \leq K$ which strengthen the relaxation. More generally, consider the disjunction $\bigvee_{p=0}^{m}(\text{card}(x_I) = p)$ for $\text{card}(x_I) \leq m$. We can add $\sum_{j\in N\setminus I} x_j \leq K-p$ in the set of inequalities of the disjunct $\text{card}(x_I) = p$ to strengthen the relaxation. The resulting inequality is clearly valid for PS. We will show that it produces a strong valid inequality under some technical conditions.

The following application of the disjunctive argument produces a valid inequality for PS from another valid inequality for PS without considering the knapsack inequality defining CCKP. When we start from a facet-defining inequality, the resulting valid inequality may be another facet-defining inequality under some technical conditions.

Theorem 4.4.1 *Suppose that*

$$\sum_{j\in N} f_j x_j \leq f_0 \tag{4.7}$$

is a valid inequality for PS. Let J be a subset of N such that $|J| = K$ and $\sum_{j\in J} f_j < f_0$. Define $\delta = f_0 - \sum_{j\in J} f_j$. Let

$$\begin{aligned} T(J) &= \{j \in N : f_j \le f_J\}, \\ H(J) &= N \setminus (J \cup T(J)) \end{aligned}$$

where $f_J = \min\{f_j : j \in J\}$. Then

$$\sum_{j\in H(J)} f_j x_j + \sum_{j\in J\cup T(J)} (f_j + \delta)x_j \le f_0 + (K-1)\delta \tag{4.8}$$

is a valid inequality for PS.

Proof We derive (4.8) using disjunctive arguments. We denote $x = (x_{H(J)}, x_J, x_{T(J)})$ where x_A consists of components of x whose indices belong to A. From the cardinality constraint, we consider the disjunction

$$\mathrm{card}(x_{H(J)}) \le 0 \quad \vee \quad \mathrm{card}(x_{J\cup T(J)}) \le K-1. \tag{4.9}$$

We use (4.9) instead of the cardinality constraint. We will show that (4.8) is valid for both the disjuncts $PS \cap \{x : \mathrm{card}(x_{H(J)}) \le 0\}$ and $PS \cap \{x : \mathrm{card}(x_{J\cup T(J)}) \le K-1\}$.

First, consider $PS \cap \{x : \mathrm{card}(x_{H(J)}) \le 0\}$. Then, together with the condition $\sum_{j\in J\cup T(J)} f_j x_j \le \sum_{j\in J} f_j$, we have

$$\sum_{j\in J\cup T(J)} (f_j + \delta)x_j = \sum_{j\in J\cup T(J)} f_j x_j + \delta \sum_{j\in J\cup T(J)} x_j \le \sum_{j\in J} f_j + K\delta = f_0 + (K-1)\delta.$$

Since $\mathrm{card}(x_{H(J)}) = 0$, $\sum_{j\in H(J)} f_j x_j = 0$. Thus we conclude that (4.8) is valid for $PS \cap \{x : \mathrm{card}(x_{H(J)}) \le 0\}$.

On the other hand, consider $PS \cap \{x : \mathrm{card}(x_{J\cup H(J)}) \le K-1\}$. The inequality defining the disjunct shows that $\sum_{j\in J\cup T(J)} x_j \le K-1$ and hence $\sum_{j\in J\cup T(J)} \delta x_j \le (K-1)\delta$. Adding this inequality to (4.7) yields (4.8). ■

We call (4.8) a *δ-inequality* with respect to (4.7). Starting from the knapsack inequality,k which is trivially valid for PS, one can sequentially apply Theorem 4.4.1 to obtain many valid inequalities for PS.

We next present a result that describes when a δ-inequality is facet-defining. For $A \subset J \cup T(J)$ and $i \in H(J)$ such that $|A| = K-1$ and $f_i + \sum_{j\in A} f_j \geq f_0$, we define

$$v^{(i,A)} := e_A + \frac{b - \sum_{j\in A} f_j}{f_i} e_i.$$

Suppose $m \in \operatorname{Arg\,min}\{f_j : j \in J\}$ and define $M := \{j \in N : f_j = f_m\}$. It is clear that $M \subseteq J \cup T(J)$. Consider the following sets:

$$\begin{aligned} V_J^1 &= \{e_{(J\setminus M)\cup B} : B \subseteq M, |B| = |J \cap M|\}, \\ V_J^2 &= \left\{v^{(i,A)} : A \subset J \cup T(J),\ |A| = K-1,\ f_i + \textstyle\sum_{j\in A} f_j \geq f_0\right\}, \\ V_J &= V_J^1 \,\dot\cup\, V_J^2. \end{aligned}$$

We next argue that V_J is the set of vertices of PS on the face induced by the δ-inequality.

Theorem 4.4.2 *Define $F_J = PS \cap \{x : x$ satisfies (4.8) at equality$\}$. Then $\operatorname{ext}(F_J) = V_J$.*

Proof We first show that $\operatorname{ext}(F_J) \subseteq V_J$. Since the δ-inequality is valid for PS, it defines a (possibly empty) face of PS. Therefore, $\operatorname{ext}(F_J) \subseteq \operatorname{vert}(PS)$. Suppose $v \in \operatorname{ext}(F_J)$. Then

$$\sum_{j\in H(J)} f_j v_j + \sum_{j\in J\cup T(J)} (f_j + \delta) v_j = f_0 + (K-1)\delta. \tag{4.10}$$

We consider two cases. Assume first that $\sum_{j\in N} f_j v_j = f_0$. Subtracting this relation from (4.10) shows that $\sum_{j\in J\cup T(J)} v_j = K-1$. Since v contains at most one fractional component, $v_{J\cup T(J)}$ is binary. Then, $v_{H(J)} \neq 0$ because

$$\sum_{j\in H(J)} f_j v_j = \sum_{j\in N} f_j v_j - \sum_{j\in J\cup T(J)} f_j v_j \geq f_0 - \left(\sum_{j\in J} f_j - f_m\right) = \delta + f_m > 0.$$

Hence $\operatorname{card}(v_{H(J)}) = 1$. Pick $i \in H(J)$ and $A \subseteq J \cup T(J)$ such that $|A| = K-1, v_i \neq 0$, and $f_i v_i + \sum_{j\in A} f_j = f_0$. Hence $v_i = \frac{f_0 - \sum_{j\in A} f_j}{f_i}$. Therefore, $v = e_A + \frac{f_0 - \sum_{j\in A} f_j}{f_i} e_i = v^{(i,A)} \in V_J$. Assume second that $\sum_{j\in N} f_j v_j < f_0$. Then, since v satisfies (4.8) at equality,

$$0 > \sum_{j\in N} f_j v_j - f_0 = \delta \left(K - 1 - \sum_{j\in J\cup T(J)} v_j\right).$$

This shows that $\sum_{j\in J\cup T(J)} v_j > K-1$ and hence $v_{J\cup T(J)}$ carries all the cardinality of v. Then,

$$f_0 + (K-1)\delta = \sum_{j\in H(J)} f_j v_j + \sum_{j\in J\cup T(J)} (f_j+\delta)v_j = \sum_{j\in J\cup T(J)} (f_j+\delta)v_j \leq f_0 + (K-1)\delta.$$

The last inequality holds at equality only if $\sum_{j\in J\cup T(J)}(f_j+\delta)v_j = \sum_{j\in J}(f_j+\delta)$. This is equivalent to stating that

$$\sum_{j\in J\cup T(J)} f_j v_j = (\text{sum of } K \text{ largest } f_j\text{s in } \{f_j : j \in J\cup T(J)\}).$$

Therefore, $v_j = 1$ for all $j \in (J\setminus M)\cup B$ where $B \subseteq M$ and $|B| = |J\cap M|$. Therefore, $v = e_{(J\setminus M)\cup B} \in V_J$.

We next prove that $\text{ext}(F_J) \supseteq V_J$. It suffices to show that any point in V_j is a vertex of PS that satisfies (4.4.1) at equality. For A and i where $v^{(i,A)}$ is well-defined, $v^{(i,A)}$ is a vertex of PS because $v^{(i,A)}$ has cardinality K, contains at most one fractional component and satisfies

$$\sum_{j\in N} f_j (v^{(i,A)})_j = \sum_{j\in A} f_j + f_i \left(\frac{b - \sum_{j\in A} f_j}{f_i}\right) = f_0.$$

This shows that $v^{(i,A)}$ is a vertex of PS. Moreover, $v^{(i,A)}$ satisfies the δ-inequality at equality because

$$\sum_{j\in H(J)} f_j (v^{(i,A)})_j + \sum_{j\in J\cup T(J)} (f_j+\delta)(v^{(i,A)})_j = \sum_{j\in A}(f_j+\delta) + f_i \left(\frac{b - \sum_{j\in A} f_j}{f_i}\right) = f_0 + (K-1)\delta.$$

Consider now any point $e_{(J\setminus M)\cup B}$,

$$\begin{aligned}\sum_{j\in N} f_j (e_{(J\setminus M)\cup B})_j &= \sum_{j\in (J\setminus M)\cup B} f_j = \sum_{j\in J} f_j - \sum_{j\in J\cap M} f_m + \sum_{j\in B} f_m \\ &= \textstyle\sum_{j\in J} f_j - |J\cap M| f_m + |B| f_m = \sum_{j\in J} f_j = f_0 - \delta < f_0.\end{aligned}$$

Since it has cardinality K, it is a vertex of PS. For the δ-inequality,

$$\begin{aligned}
&\sum_{j\in H(J)} f_j(e_{(J\setminus M)\cup B})_j + \sum_{j\in J\cup T(J)} (f_j+\delta)(e_{(J\setminus M)\cup B})_j = \sum_{j\in J\cup T(J)} (f_j+\delta)(e_{(J\setminus M)\cup B})_j \\
= &\sum_{j\in (J\setminus M)\cup B} (f_j+\delta) = \sum_{j\in J}(f_j+\delta) - \sum_{j\in J\cap M}(f_j+\delta) + \sum_{j\in B}(f_j+\delta) \\
= &\sum_{j\in J}(f_j+\delta) - \sum_{j\in J\cap M}(f_m+\delta) + \sum_{j\in B}(f_m+\delta) \\
= &\sum_{j\in J} f_j + K\delta - |J\cap M|(f_m+\delta) + |B|(f_m+\delta) = b - \delta + K\delta = f_0 + (K-1)\delta
\end{aligned}$$

Thus, $\mathrm{ext}(F_J) = V_J$. ■

Corollary 8 *(4.8) is facet-defining if and only if V_J contains n affinely independent vectors.*

Proposition 4.4.2 *If $\min\{f_j : j \in H(J)\} - f_m < \delta$ then (4.8) is not facet-defining.*

Proof Let $m_H = \operatorname{Arg\,min}\{f_j : j \in H(J)\}$ and hence $f_{m_H} - f_m < \delta$. Let $v \in V_J$. If $v \in V_J^1$ then it is clear that $v_{m_H} = 0$. We next consider $v \in V_J^2$. Assume by contradiction that $v_{m_H} > 0$. Since $v \in V_J^2$, $f_{m_H} + \sum_{j\in A} f_j \geq f_0$ for some $A \subseteq J \cup T(J)$ with $|A| = K-1$. Then

$$f_{m_H} - f_m \geq f_0 - \sum_{j\in A} f_j - f_m \geq f_0 - \sum_{j\in J\setminus\{m\}} f_j - f_m = \delta,$$

which yields a contradiction. Therefore, $v_{m_H} = 0$. If there exists n affinely independent points in V_J, then the facet-defining inequality corresponding to V_J should be $x_{m_H} \geq 0$, which inequality (4.8) is not. Hence there does not exist n affinely independent points in V_J. This shows that (4.8) is not facet-defining. ■

Proposition 4.4.3 *Let $m_2 \in \operatorname{Arg\,min}\{f_j : j \in J \setminus \{m\}\}$, $M_H \in \operatorname{Arg\,max}\{f_j : j \in H(J)\}$, and $m_T \in \operatorname{Arg\,min}\{f_j : j \in T(J)\}$. Suppose $f_{m_T} < f_m$. If $f_{M_H} - f_m - f_{m_2} - f_{m_T} < \delta$ then (4.8) is not facet-defining.*

Proof Let $v \in V_J$. Since $f_{m_T} < f_m$, $v_{m_T} = 0$ if $v \in V_J^1$. Suppose $v \in V_J^2$ and assume that $v_{m_T} = 1$. Then there exists $i \in H(J)$ and $A \subseteq J \cup T(J)$ with $m_T \in A$ such that $f_i + \sum_{j \in A} f_j \geq f_0$. Hence

$$\begin{aligned} f_0 &\leq f_i + \sum_{j \in A \setminus \{m_T\}} f_j + f_{m_T} \\ &\leq f_{M_H} + \sum_{j \in J \setminus \{m, m_2\}} f_j + f_{m_T} \\ &= f_{M_H} + \sum_{j \in J} f_j - f_m - f_{m_2} + f_{m_T} \\ &= f_{M_H} + f_0 - \delta - f_m - f_{m_2} + f_{m_T}. \end{aligned}$$

Therefore, $f_{M_H} - f_m - f_{m_2} + f_{m_T} \geq \delta$, which contradicts the assumption. Hence $v_{m_T} = 0$. This shows that (4.8) is not facet-defining. ■

The following result is obtained by considering $m = 1$ in the disjunctive argument.

Proposition 4.4.4 *Suppose $K \geq 2$. Let (4.7) be a valid inequality for PS. Let $J \subseteq N$ be such that $|J| = K$ and $\sum_{j \in J} f_j < b$. Let $\delta = f_0 - \sum_{j \in J} f_j$. Let $j^* \in \operatorname{Arg\,max}\{f_j : j \in N\}$ and define β by*

$$\beta = \begin{cases} \max\{f_j : j \in T(J)\} & \text{if } T(J) \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

Suppose $j^* \in J$ and $f_{j^*} - \beta \geq \delta$. Then

$$\sum_{j \in H(J) \cup \{j^*\}} f_j x_j + \sum_{j \in J \cup T(J) \setminus \{j^*\}} (f_j + \delta) x_j \leq f_0 + (K-2)\delta \tag{4.11}$$

is valid for PS.

Proof Consider the following disjunction for the cardinality constraint:

$$\operatorname{card}(x_{H(J) \cup \{j^*\}}) = 0 \ \vee \ \operatorname{card}(x_{H(J) \cup \{j^*\}}) = 1 \ \vee \ \operatorname{card}(x_{J \cup T(J) \setminus \{j^*\}}) \leq K - 2. \tag{4.12}$$

The first two disjuncts can be written as $\operatorname{card}(x_{H(J) \cup \{j^*\}}) \leq 1$. The negation of (4.12) is therefore

$$\operatorname{card}(x_{H(J) \cup \{j^*\}}) \geq 2 \quad \wedge \quad \operatorname{card}(x_{J \cup T(J) \setminus \{j^*\}}) \geq K - 1,$$

which is equivalent to $\text{card}(x) \geq K+1$. Hence (4.12) is equivalent to the cardinality constraint. For convenience, we denote the sets that represent the three disjuncts by A, B, and C respectively.

We first consider $PS \cap A$. Since $\text{card}(x_{H(J)\cup\{j^*\}}) = 0$, $\text{card}(x_{J\cup T(J)\setminus\{j^*\}}) \leq K$ and hence $\sum_{j\in J\cup T(J)\setminus\{j^*\}} x_j \leq K$, we have that

$$\sum_{j\in J\cup T(J)\setminus\{j^*\}} f_j x_j \leq \sum_{j\in J} f_j - f_{j^*} + \beta = f_0 - \delta - f_{j^*} + \beta \leq f_0 - 2\delta.$$

Therefore,

$$\begin{aligned}\sum_{j\in H(J)\cup\{j^*\}} f_j x_j + \sum_{j\in J\cup T(J)\setminus\{j^*\}} (f_j+\delta)x_j &= \sum_{j\in J\cup T(J)\setminus\{j^*\}} f_j x_j + \delta\left(\sum_{j\in J\cup T(J)\setminus\{j^*\}} x_j\right)\\ &\leq f_0 - 2\delta + K\delta = f_0 + (K-2)\delta.\end{aligned}$$

Next, from disjunct $PS \cap B$, we have that $\text{card}(x_{J\cup T(J)\setminus\{j^*\}}) \leq K-1$. It follows that $\sum_{j\in J\cup T(J)\setminus\{j^*\}} \delta x_j \leq (K-1)\delta$ and

$$\sum_{j\in J\cup T(J)\setminus\{j^*\}} f_j x_j \leq \sum_{j\in J} f_j - f_{j^*} = f_0 - \delta - f_{j^*}.$$

We also have that $\sum_{j\in H(J)\cup\{j^*\}} f_j x_j \leq f_{j^*}$. Therefore,

$$\sum_{j\in H(J)\cup\{j^*\}} f_j x_j + \sum_{j\in J\cup T(J)\setminus\{j^*\}} (f_j+\delta)x_j \leq f_{j^*} + f_0 - \delta - f_{j^*} + (K-1)\delta = f_0 + (K-2)\delta.$$

Finally, we consider $PS \cap C$. From the cardinality constraint of the disjunct, we have that

$$\sum_{j\in J\cup T(J)\setminus\{j^*\}} x_j \leq K-2.$$

Adding a multiple of this inequality to (4.7), we obtain that (4.11) is valid for $PS \cap C$.

■

Proposition 4.4.4 can be generalized based on the number of largest coefficients J contains.

Proposition 4.4.5 *Let $m \in \{1, \ldots, K\}$. Let (4.7) be a valid inequality for PS. Without loss of generality, assume that $f_1 \geq f_2 \geq \cdots \geq f_n$. Let $J \subseteq N$ be such that $|J| = K$ and $\sum_{j \in J} f_j < b$. Let $\delta = f_0 - \sum_{j \in J} f_j$. Define*

$$\beta_1 = \begin{cases} \max\{f_k : k \in T(J)\} & \text{if } |T(J)| \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

For $j \geq 2$, let

$$\beta_j = \begin{cases} \max\{f_k : k \in T(J) \setminus \{1, \ldots, j-1\}\} & \text{if } |T(J)| \geq j \\ 0 & \text{otherwise} \end{cases}$$

Suppose $\{1, \ldots, m\} \subseteq J$ and $\sum_{j=i+1}^{m} f_j - \sum_{j=1}^{m-i} \beta_j \geq (m-i)\delta$ for $i = 0, \ldots, m-1$. Then

$$\sum_{H(J) \cup \{1,\ldots,m\}} f_j x_j + \sum_{j \in J \cup T(J) \setminus \{1,\ldots,m\}} (f_j + \delta) x_j \leq f_0 + (K - m - 1)\delta \tag{4.13}$$

is valid for PS.

Proof We omit the proof because it is similar to that of Proposition 4.4.4.

■

Example 11 *Consider CCKP with $K = 3$ and knapsack inequality*

$$12x_1 + 6x_2 + 4x_3 + 2x_4 + 2x_5 \leq 19.$$

We first consider the knapsack inequality itself as a valid inequality of PS. Choosing $J = \{1, 3, 4\}$, we have $H(J) = \emptyset, T(J) = \{5\}$, and $\delta = 1$. Set J contains 1 and we compute that $f_1 - \beta_1 = f_1 - f_5 = 10 > \delta$. Hence we can apply Proposition 4.4.5 to obtain that

$$12x_1 + 6x_2 + 5x_3 + 3x_4 + 3x_5 \leq 20 \tag{4.14}$$

is valid for PS. This inequality is facet-defining for PS but cannot be obtained using Proposition 4.4.1. Moreover, (4.14) is stronger than the δ-inequality $13x_1 + 6x_2 + 5x_3 + 3x_4 + 3x_5 \leq 21$ because $x_1 \leq 1$ is valid for PS. Consider (4.14) as a valid

inequality for PS. We can apply Proposition 4.4.5 with $J = \{1, 4, 5\}$. In this case, $\delta = 2$ and $\beta_1 = 0$. Further, $f_1 - \beta_1 = f_1 = 12 \geq \delta = 2$. Hence the resulting valid inequality is

$$12x_1 + 6x_2 + 5x_3 + 5x_4 + 5x_5 \leq 22. \tag{4.15}$$

This inequality is facet-defining for PS and stronger than the δ-inequality, $14x_1 + 6x_2 + 5x_3 + 5x_4 + 5x_5 \leq 24$. Further (4.15) cannot be obtained using δ-method.

We next derive valid inequalities directly from the knapsack constraint $\sum_{j=1}^{n} a_j x_j \leq b$ that defines CCKP.

Proposition 4.4.6 *Suppose $K \geq 2$. Define $m \in \operatorname{Arg\,min}\{\sum_{j=n-t+1}^{n} a_j : b - a_1 < \sum_{j=n-t+1}^{n} a_j < b,\ t = 1, \ldots, K-1\}$. Assume that there exists $j \leq n - m$ such that $a_j + \sum_{k=n-m+1}^{n} a_k < b$ and let $j^* \in \operatorname{Arg\,max}\{a_j : a_j + \sum_{k=n-m+1}^{n} a_k < b\}$. Let $J = \{j^*, n-m+1, \ldots, n\}$ and $\delta = b - \sum_{j \in J} a_j$. Also assume that $b + \delta \leq a_{j^*}$. Then*

$$\left(\frac{a_{j^*}}{\delta + a_{j^*}}\right) a_1 x_1 + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \leq b + (K-2)\delta \tag{4.16}$$

is valid for PS.

Proof For any $x \in \operatorname{vert}(PS)$, if $x_1 = 0$, we only need to consider the inequality

$$\sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \leq b + (K-2)\delta \tag{4.17}$$

in $PS(\{1\}, \emptyset)$. Since $1 \notin J$, by Proposition 4.4.4, (4.17) is valid for $PS(\{1\}, \emptyset)$.

Suppose $x_1 = 1$ and assume $K \geq 3$. Then $\sum_{j \in N \setminus \{1\}} a_j x_j \leq b - a_1$. By Proposition 4.4.4,

$$\sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \leq b - a_1 + (K-3)\delta$$

is valid for $PS(\emptyset, \{1\})$. Thus

$$\begin{aligned}
& \left(\frac{a_{j^*}}{\delta + a_{j^*}}\right) a_1 x_1 + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \\
= \ & \frac{a_{j^*} a_1}{\delta + a_{j^*}} + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \\
\leq \ & \frac{a_{j^*} a_1}{\delta + a_{j^*}} + b - a_1 + (K-3)\delta \leq b + (K-2)\delta.
\end{aligned}$$

If $x_1 = 1$ and $K = 2$, by construction $m = 1$ and $J = \{j^*, n\}$. Hence $\delta = b - a_j^* - a_n > 0$ and $a_1 + a_n > b$. This shows that x_n is fractional and hence x satisfies the knapsack constraint at equality. That is, $x_n = \frac{b-a_1}{a_n}$. Therefore,

$$\begin{aligned}
& \left(\frac{a_{j^*}}{\delta + a_{j^*}}\right) a_1 x_1 + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \\
= \quad & \left(\frac{a_{j^*}}{\delta + a_{j^*}}\right) a_1 + (a_n + \delta) \frac{b - a_1}{a_n} = \frac{a_1 a_j^*}{b - a_n} + \frac{(b - a_1)(b - a_j^*)}{a_n} \\
= \quad & \frac{a_1 a_j^* a_n + (b - a_1)(b - a_j^*)(b - a_n)}{a_n (b - a_n)} = b + \frac{b(a_1 + a_n - b)(a_j^* + a_n - b)}{a_n (b - a_n)} \leq b \\
= \quad & b + (K - 2)\delta.
\end{aligned}$$

Now suppose x_1 is fractional and $K \geq 2$. We consider the following disjunction for the cardinality constraint:

$$\text{card}\left(x_{(N \setminus \{1\}) \setminus (J \setminus \{j^*\})}\right) \leq 0 \quad \vee \quad \text{card}\left(x_{J \setminus \{j^*\}}\right) \leq K - 2.$$

We denote the above disjuncts by A, and B respectively.

We first consider the disjunct $PS \cap A$. Observe that $\text{card}(x_{(N \setminus \{1\}) \setminus (J \setminus \{j^*\})}) = 0$ implies that $\text{card}(x_{J \setminus \{j^*\}}) \leq K - 1$. Thus, we have that $\sum_{j \in J \setminus \{j^*\}} a_j x_j \leq \sum_{j \in J} a_j - a_{j^*} = b - \delta - a_{j^*}$ and $\sum_{j \in J \setminus \{j^*\}} x_j \leq K - 1$. Since x_1 is fractional, x satisfies the knapsack inequality at equality. That is,

$$a_1 x_1 + \sum_{j \in J \setminus \{j^*\}} a_j x_j = b.$$

Hence

$$a_1 x_1 = b - \sum_{j \in J \setminus \{j^*\}} a_j x_j = \delta + a_{j^*}.$$

Therefore,

$$\begin{aligned}
& \left(\frac{a_{j^*}}{\delta + a_{j^*}}\right) a_1 x_1 + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \\
= \quad & a_{j^*} + \sum_{j \in J \setminus \{j^*\}} a_j x_j + \delta \sum_{j \in J \setminus \{j^*\}} x_j \\
\leq \quad & a_{j^*} + b - \delta - a_{j^*} + (K - 1)\delta = b + (K - 2)\delta.
\end{aligned}$$

Next, we consider $PS \cap B$. Since the knapsack inequality is valid, we have that $\sum_{j \in N} a_j x_j \leq b$. The disjunct constraint imposes that $\sum_{j \in J \setminus \{j^*\}} x_j \leq K - 2$. Hence

$$\begin{aligned} & \left(\frac{a_{j^*}}{\delta + a_{j^*}}\right) a_1 x_1 + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \\ \leq \quad & a_1 x_1 + \sum_{j \in (N \setminus \{1\}) \setminus (J \setminus \{j^*\})} a_j x_j + \sum_{j \in J \setminus \{j^*\}} (a_j + \delta) x_j \leq b + (K-2)\delta. \end{aligned} \tag{4.18}$$

■

Example 12 *Consider CCKP with $n = 5, K = 2$, and knapsack inequality*

$$30x_1 + 27x_2 + 10x_3 + 7x_4 + 4x_5 \leq 32.$$

Let $J = \{2, 5\}$. Then, $\delta = 1$. Inequality (4.16) takes its form

$$\frac{405}{14}x_1 + 27x_2 + 10x_3 + 7x_4 + 5x_5 \leq 32 \quad \Longleftrightarrow \quad 405x_1 + 378x_2 + 140x_3 + 98x_4 + 70x_5 \leq 448.$$

The above inequality is facet-defining for PS.

4.4.3 Derivation of known inequalities for the literature

In this section, we derive some inequalities introduced in [26] using our disjunctive arguments.

Proposition 4.4.7 (Theorem 2, [26]) *The inequality*

$$\sum_{j \in N} \max\left\{a_j, b - \sum_{i=1}^{K-1} a_i\right\} x_j \leq b \tag{4.19}$$

is valid for PS.

Proof We recall the assumption that $a_1 \geq \cdots \geq a_n$ and $\sum_{i=1}^{K} a_i > b$. If $a_j + \sum_{i=1}^{K-1} a_i \geq b$ for all $j \in N$, then (4.19) is nothing but the knapsack inequality. Hence here we assume that there exists $j \in N$ such that $a_j + \sum_{i=1}^{K-1} a_i < b$. Define $\alpha = b - \sum_{i=1}^{K-1} a_i$. Since $\sum_{i=1}^{K} a_i > b$, then $a_K > \alpha$. We first define an index j^* as follow:

$$j^* = \min\{j : a_j < \alpha,\ j \in N\}.$$

Let $A = \{1, \ldots, j^* - 1\}$. Consider the following disjunction on the cardinality constraint:

$$\begin{aligned} & (\mathrm{card}(x_A) \leq K-1) \quad \vee \quad (\mathrm{card}(x_{N\setminus A}) \leq 0) \\ \iff & \bigvee_{k=0}^{K-1} (\mathrm{card}(x_A) = k) \quad \vee \quad (\mathrm{card}(x_{N\setminus A}) = 0). \end{aligned}$$

For every $k = 0, \ldots, K-1$, consider the disjunct $\mathrm{card}(x_A) = k$ and the corresponding disjunct for PS. For any vertex x of PS, we have that $\mathrm{card}(x_{N\setminus A}) \leq K-k$. Then

$$\begin{aligned} \sum_{i=1}^{j^*-1} a_j x_j + \sum_{i=j^*}^{n} \alpha x_j &\leq \sum_{i=1}^{k} a_j + \alpha(K-k) \\ &= \sum_{i=1}^{K-1} a_j - \sum_{i=k+1}^{K-1} a_j + \alpha(K-k) \\ &= b - \alpha - \sum_{i=k+1}^{K-1} a_j + \alpha(K-k) \\ &\leq b - (K-k-1)(a_{K-1} - \alpha) \leq b. \end{aligned}$$

For the disjunct $\mathrm{card}(x_{N\setminus A}) = 0$,

$$\sum_{i=1}^{j^*-1} a_j x_j + \sum_{i=j^*}^{n} \alpha x_j = \sum_{i=1}^{j^*-1} a_j x_j \leq \sum_{i=1}^{n} a_j x_j \leq b.$$

Therefore, (4.4.7) is valid for PS. ■

Proposition 4.4.8 (Theorem 3, [26]) *Let* $\alpha = b - \sum_{j=n-K+2}^{n} a_j$ *and suppose* $a_{n-K} + a_{n-K+1} - a_n \leq \alpha$. *Then*

$$\sum_{j \in N} \max\{a_j, \alpha\} x_j \leq \alpha K \tag{4.20}$$

is valid for PS.

Proof Define $j^* := \mathrm{Arg\,min}\{j : a_j \leq \alpha,\ j \in N\}$ and denote $A = \{1, \ldots, j^* - 1\}$ and $B = \{j^*, \ldots, n\}$. We consider the following disjunction for the cardinality constraint:

$$(\mathrm{card}(x_A) \leq 0) \quad \vee \quad \bigvee_{k=0}^{K-1} \big(\mathrm{card}(x_{N\setminus A}) = k\big)$$

Consider further the case where a vertex x of PS satisfies the first disjunct $\text{card}(x_A) = 0$. Then $\text{card}(x_B) \leq K$ and hence $\sum_{j\in A} a_j x_j + \alpha \sum_{j\in B} x_j = \alpha \sum_{j=j^*}^{n} x_j \leq \alpha K$.

Consider the second case where x satisfies $\text{card}(x_B) = k$ for some $k = 0, \ldots, K-1$. Then $\sum_{j\in A} a_j x_j \leq b - \sum_{j=n-k+1}^{n} a_j$ and $\sum_{j\in B} x_j \leq k$. Moreover, for any $j \geq n-K+2$,

$$\alpha - a_j \geq a_{n-K} + a_{n-K+1} - a_n - a_j \geq 0.$$

Therefore,

$$\begin{aligned}
\sum_{j\in A} a_j x_j + \alpha \sum_{j\in B} x_j \leq \quad & b - \sum_{j=n-k+1}^{n} a_j + \alpha k \\
= \quad & b - \sum_{j=n-K+2}^{n} a_j + \sum_{j=n-K+2}^{n-k} a_j + \alpha k \\
\leq \quad & \alpha + \alpha(K-k-1) + \alpha k = \alpha K.
\end{aligned}$$

■

Observe that Proposition 4.4.8 does not require the condition $\alpha < a_1$ which is imposed in Theorem 3 of [26]. The condition $\alpha < a_1$ implies that $\max\{a_1, \alpha\} = a_1$ and hence it is consistent with [26]. When $\alpha \geq a_j$ for $j \geq n-K+2$, (4.20) is still valid for PS.

4.4.4 Derivation of the δ-inequality via lifting

In this subsection, we derive (4.8) using lifting arguments. We first relabel the indices of J as $n_1, \ldots, n_K$ so that $f_{n_1} \geq \cdots \geq f_{n_K}$. Consider the following valid inequality

$$\begin{aligned}
& \sum_{j\in J} (f_j + \delta) x_j \leq f_0 + (K-1)\delta \\
\Longleftrightarrow & \sum_{j=1}^{K} (f_{n_j} + \delta) x_{n_j} \leq f_0 + (K-1)\delta
\end{aligned}$$

for $PS(N \setminus J, \emptyset)$ where $\sum_{j\in J} f_j < f_0$ and $\delta = f_0 - \sum_{j\in J} f_j$.

For some $k \in N \setminus J$, we want to find g_k such that

$$\sum_{j=1}^{K} (f_{n_j} + \delta) x_{n_j} + g_k x_k \leq f_0 + (K-1)\delta \tag{4.21}$$

is valid for $PS(N \setminus (J \cup \{k\}), \emptyset)$.

The lifting function is

$$\begin{aligned}
\Psi(z) &= \min \left\{ f_0 + (K-1)\delta - \textstyle\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} : \begin{array}{l} \sum_{j\in N} f_j x_j \le f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \le K-1 \end{array} \right\} \\
&= f_0 + (K-1)\delta - \max \left\{ \textstyle\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} : \begin{array}{l} \sum_{j=1}^{K} f_{n_j} x_{n_j} \le f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \le K-1 \end{array} \right\}.
\end{aligned}$$

Proposition 4.4.9 $z \le \Psi(z)$.

Proof Consider the feasible set of the maximization problem in the definition of $\Psi(z)$. Any x in the set satisfies $\sum_{j=1}^{K} x_{n_j} \le K-1$ because of the cardinality constraint. It follows that

$$\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} = \sum_{j=1}^{K} f_{n_j}x_{n_j} + \delta \sum_{j=1}^{K} x_{n_j} \le f_0 - z + \delta(K-1)$$

Therefore, we have that $\Psi(z) \ge f_0 + (K-1)\delta - (f_0 - z + \delta(K-1)) = z$. ■

It is easy to verify that if g_k satisfies $g_k x_k \le \Psi(f_k x_k)$ for all $x_k \in [0,1]$ then (4.21) is valid for $PS(N \setminus (J \cup \{k\}), \emptyset)$. Therefore, the lifted inequality

$$\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} + f_k x_k \le f_0 + (K-1)\delta \tag{4.22}$$

is valid for $PS(N \setminus (J \cup \{k\}), \emptyset)$.

Next, we pick a variable x_i other than $x_{n_1}, \ldots, x_{n_K}$, and x_k. We want to find g_k such that $\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} + f_k x_k + g_i x_i \le f_0 + (K-1)\delta$ is valid for $PS(N \setminus (J \cup \{k,i\}), \emptyset)$.

The lifting function is

$$\begin{aligned}
\Psi(z) &= \min\left\{ f_0 + (K-1)\delta - \textstyle\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} - f_k x_k : \begin{array}{l} \sum_{j\in N} f_j x_j \le f_0 - z, \\ x \in [0,1]^n, \\ \mathrm{card}(x) \le K-1 \end{array} \right\} \\
&= f_0 + (K-1)\delta \\
&\quad - \max\left\{ \textstyle\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} + f_k x_k : \begin{array}{l} \sum_{j=1}^{K} f_{n_j} x_{n_j} + f_k x_k \le f_0 - z, \\ x \in [0,1]^n, \\ \mathrm{card}(x) \le K-1 \end{array} \right\}.
\end{aligned}$$

Note that $\Psi(0) = 0$. It is obvious that $\Psi(z) \ge z$ and hence $\sum_{j=1}^{K}(f_{n_j}+\delta)x_{n_j} + f_k x_k + f_i x_i \le f_0 + (K-1)\delta$ is valid for $PS(N \setminus (J \cup \{k,i\}), \emptyset)$. Iterating this process, we obtain that

$$\sum_{j\in J}(f_j+\delta)x_j + \sum_{j\in N\setminus J} f_j x_j \le f_0 + (K-1)\delta$$

is valid inequality for PS.

4.5 New valid inequalities via lifting

Consider the valid inequality

$$\sum_{j\in J} f_j x_j \le f_0 - \delta$$

for $PS(N \setminus J, \emptyset)$. By the cardinality constraint,

$$\sum_{j\in J\cup T(J)} f_j x_j \le f_0 - \delta. \tag{4.23}$$

is valid for $PS(N \setminus (J \cup T(J)), \emptyset)$.

For $k \notin J \cup T(J)$, we need to find g_k such that

$$\sum_{j\in J\cup T(J)} f_j x_j + g_k x_k \le f_0 - \delta.$$

is valid for $PS(N \setminus (J \cup T(J) \cup \{k\}), \emptyset)$. The lifting function here is

$$\begin{aligned}
\Psi(z) &= \min \left\{ f_0 - \delta - \textstyle\sum_{j \in J \cup T(J)} f_j x_j : \begin{array}{l} \sum_{j \in N} f_j x_j \leq f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \leq K - 1 \end{array} \right\} \\
&= f_0 - \delta - \max \left\{ \textstyle\sum_{j \in J \cup T(J)} f_j x_j : \begin{array}{l} \sum_{j \in J \cup T(J)} f_j x_j \leq f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \leq K - 1 \end{array} \right\}.
\end{aligned}$$

Let $m \in \operatorname{Arg\,min}\{f_j : j \in J\}$ and assume $0 < z \leq \delta + f_m$. Notice that $\sum_{j \in J} f_j - f_m$ is an upper bound of $\sum_{j \in J \cup T(J)} f_j x_j$ and $\mathbf{e}^{J-m} := \sum_{j \in J} e_j - e_m$ achieves its optimal value. Moreover, this solution is feasible because $\mathrm{card}(\mathbf{e}^{J-m}) = K - 1$ and

$$\sum_{j \in J \cup T(J)} f_j \mathbf{e}_j^{J-m} = \sum_{j \in J} f_j - f_m = f_0 - \delta - f_m \leq f_0 - z.$$

This shows that $\Psi(z) = f_0 - \delta - (f_0 - \delta - f_m) = f_m$.

Suppose $\delta + f_m \leq z \leq f_0$. We have

$$\sum_{j \in J} f_j - f_m = f_0 - \delta - f_m \geq f_0 - z.$$

This shows that there exists x^0 with cardinality $K - 1$ such that $\sum_{j \in J \cup T(J)} f_j x_j^0 = f_0 - z$ and hence

$$\max \left\{ \sum_{j \in J \cup T(J)} f_j x_j : \begin{array}{l} \sum_{j \in J \cup T(J)} f_j x_j \leq f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \leq K - 1 \end{array} \right\} = f_0 - z.$$

Therefore we obtain

$$\Psi(z) = f_0 - \delta - (f_0 - z) = z - \delta$$

for $\delta + f_m \leq z \leq f_0$. If $z > f_0$ the problem is infeasible and therefore $\Psi(z) = \infty$. Thus, we have

$$\Psi(z) = \begin{cases} 0 & z \in (-\infty, 0] \\ f_m & z \in (0, \delta + f_m] \\ z - \delta & z \in [\delta + f_m, f_0] \\ \infty & z \in (f_0, \infty). \end{cases}$$

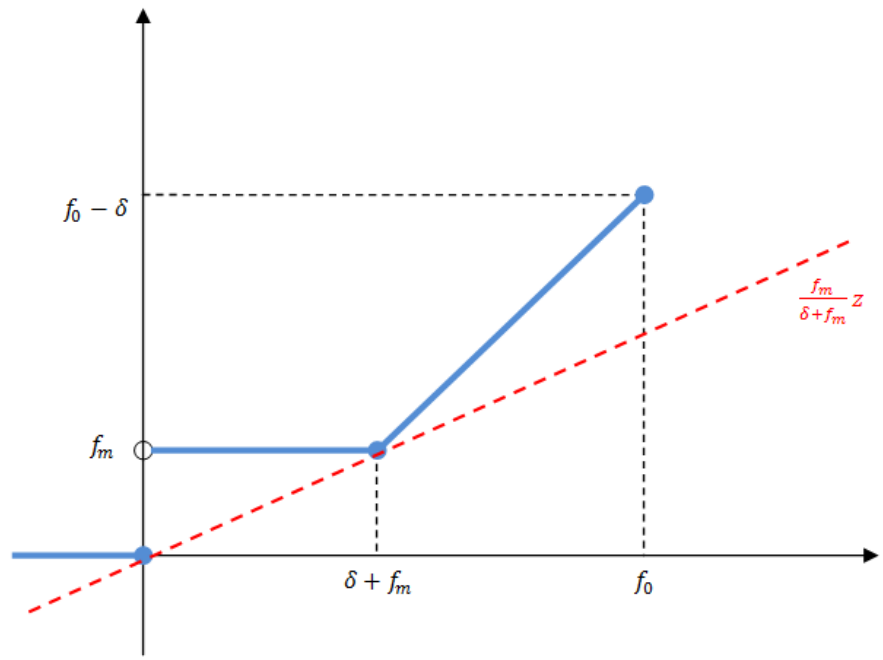


Figure 4.1.: $\Psi(z)$ and its linear under-estimator for (4.23)

It is clear that $\frac{f_m}{\delta+f_m} z \leq \Psi(z)$. It follows that $\frac{f_m}{\delta+f_m} f_k x_k \leq \Psi(f_k x_k)$. We conclude that $g_k = \frac{f_m}{\delta+f_m} f_k$ is a valid lifting coefficient. We obtain

$$\sum_{j \in J \cup T(J)} f_j x_j + \frac{f_m f_k}{\delta + f_m} x_k \leq f_0 - \delta$$

or equivalently,

$$(\delta + f_m) \sum_{j \in J \cup T(J)} f_j x_j + f_m f_k x_k \leq (\delta + f_m)(f_0 - \delta)$$

is a valid inequality for $PS(N \setminus (J \cup T(J) \cup \{k\}), \emptyset)$.

Define $f^{(1)}$ by $f_j^{(1)} = f_j$ for $j \in J \cup T(J)$ and $f_k^{(1)} = \frac{f_m f_k}{\delta+f_m}$. Next, for $i \notin (J \cup T(J) \cup \{k\})$, we need to find g_i such that $\sum_{j \in J \cup T(J)} f_j^{(1)} x_j + f_k^{(1)} x_k + g_i x_i \leq f_0 - \delta$ is a valid inequality for $PS(N \setminus (J \cup T(J) \cup \{k, i\}), \emptyset)$. Consider the lifting function $\Psi(z)$:

$$\begin{aligned}
\Psi(z) &= \min\left\{ f_0 - \delta - \textstyle\sum_{j\in J\cup T(J)} f_j^{(1)} x_j - f_k^{(1)} x_k : \begin{array}{l} \sum_{j\in N} f_j^{(1)} x_j \le f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \le K-1 \end{array} \right\} \\
&= f_0 - \delta \\
&\quad - \max\left\{ \textstyle\sum_{j\in J\cup T(J)} f_j^{(1)} x_j + f_k^{(1)} x_k : \begin{array}{l} \sum_{j\in J\cup T(J)} f_j^{(1)} x_j + f_k^{(1)} x_k \le f_0 - z \\ x \in [0,1]^n \\ \mathrm{card}(x) \le K-1 \end{array} \right\}.
\end{aligned}$$

Let L be the set of indices in $J \cup T(J) \cup \{k\}$ for the $(K-1)$-largest $f_j^{(1)}$s. Suppose $0 < z \le f_0 - \sum_{j\in L} f_j^{(1)}$. Then $e^L := \sum_{j\in L} e_j$ is feasible for the maximization problem because $\mathrm{card}(e^L) = K-1$ and

$$\sum_{j\in J\cup T(J)} f_j^{(1)} (e^L)_j + f_k^{(1)} (e^L)_k = \sum_{j\in L} f_j^{(1)} \le f_0 - z.$$

By the maximality of L, the maximum is $\sum_{j\in L} f_j^{(1)}$ and hence

$$\Psi(z) = f_0 - \delta - \sum_{j\in L} f_j^{(1)}.$$

For $f_0 - \sum_{j\in L} f_j^{(1)} \le z \le f_0$, we observe that $\sum_{j\in L} f^{(1)} \ge f_0 - z$. This shows that there exists x^0 such that

$$\sum_{j\in J\cup T(J)} f_j^{(1)} x_j^0 + f_k^{(1)} x_k^0 = f_0 - z.$$

Hence we have $\Psi(z) = f_0 - \delta - (f_0 - z) = z - \delta$. Therefore, we have

$$\Psi(z) = \begin{cases} 0 & z \in (-\infty, 0] \\ f_0 - \delta - \sum_{j\in L} f_j^{(1)} & z \in (0, f_0 - \sum_{j\in L} f_j^{(1)}] \\ z - \delta & z \in [f_0 - \sum_{j\in L} f_j^{(1)}, f_0] \\ \infty & z \in (f_0, \infty). \end{cases}$$

Figure 4.2 shows the graph of the lifting function and a linear under-estimator.

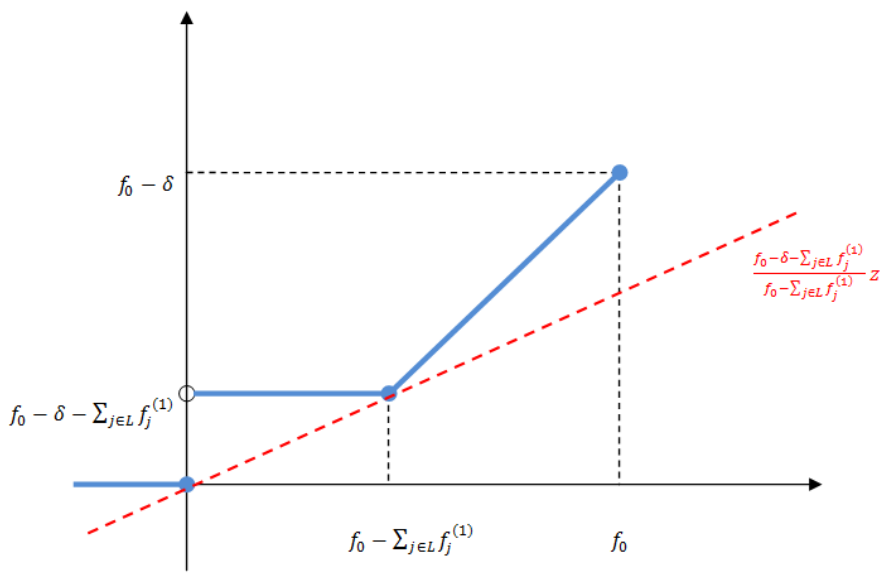


Figure 4.2.: $\Psi(z)$ and its linear under-estimator for Proposition 4.5.1

Thus, we have that $\Psi(z) \geq \left(\frac{f_0-\delta-\sum_{j\in L} f_j^{(1)}}{f_0-\sum_{j\in L} f_j^{(1)}}\right) z$. Therefore, we have

$$g_i = \frac{f_0 - \delta - \sum_{j\in L} f_j^{(1)}}{f_0 - \sum_{j\in L} f_j^{(1)}} f_i$$

is a valid lifting coefficient. We obtain the following valid inequality for $PS(N - \backslash (J \cup T(J) \cup \{k, i\}, \emptyset)$:

$$\sum_{j\in J\cup T(J)} f_j^{(1)} x_j + f_k^{(1)} x_k + \left(\frac{f_0 - \delta - \sum_{j\in L} f_j^{(1)}}{f_0 - \sum_{j\in L} f_j^{(1)}} f_i\right) x_i \leq f_0 - \delta.$$

Proposition 4.5.1 *By iterating the above procedure we will obtain a valid inequality for PS.*

Example 13 *Consider CCKP with $n = 7$, cardinality $K = 4$ and knapsack inequality*

$$15x_1 + 11x_2 + 6x_3 + 5x_4 + 3x_5 + 3x_6 + x_7 \leq 25.$$

Set $J = \{1, 4, 5, 7\}$ and hence $T(J) = \emptyset$ and $\delta = 25 - (15 + 5 + 3 + 1) = 1$. The seed inequality

$$15x_1 + 5x_4 + 3x_5 + x_7 \leq 24$$

is valid for $PS(\{2,3,6\},\emptyset)$*. For variable* x_6*, we want to find* g_6 *such that* $15x_1+5x_4+3x_5+g_6x_6+x_7 \leq 24$ *is valid for* $PS(\{2,3\},\emptyset)$*. We have* $g_6 = \frac{25-1-23}{25-23}3 = 3/2$*. This shows that*

$$15x_1 + 5x_4 + 3x_5 + \frac{3}{2}x_6 + x_7 \leq 24$$

is valid for $PS(\{2,3\},\emptyset)$*. For variable* x_3*, we want to find* g_3 *such that* $15x_1 + g_3x_3 + 5x_4 + 3x_5 + \frac{3}{2}x_6 + x_7 \leq 24$ *is valid for* $PS(\{2\},\emptyset)$*. We compute that* $g_3 = \frac{1}{2}6 = 3$*. This shows that*

$$15x_1 + 3x_3 + 5x_4 + 3x_5 + \frac{3}{2}x_6 + x_7 \leq 24$$

is valid for $PS(\{2\},\emptyset)$*. For variable* x_2*, we want to find* g_2 *such that* $15x_1 + g_2x_2 + 3x_3 + 5x_4 + 3x_5 + \frac{3}{2}x_6 + x_7 \leq 24$ *is valid for* PS*. We compute that* $g_2 = \frac{1}{2}11$*. This shows that*

$$\begin{aligned} & 15x_1 + \tfrac{11}{2}x_2 + 3x_3 + 5x_4 + 3x_5 + \tfrac{3}{2}x_6 + x_7 \leq 24 \\ \iff\ & 30x_1 + 11x_2 + 6x_3 + 10x_4 + 6x_5 + 3x_6 + 2x_7 \leq 48 \end{aligned}$$

is valid for PS*. Even though the resulting inequality is not facet-defining, it is satisfied at equality by the following extreme points of its feasible region*

$$(1,0,0,1,1,\frac{2}{3},0),(1,0,\frac{1}{3},1,1,0,0),(1,\frac{2}{11},0,1,1,0,0),(1,0,0,1,1,0,1).$$

5. Concluding remarks

5.1 Conclusion

In this thesis, we have studied certain classes of cardinality constrained optimization problems. We first designed a cut-generating procedure for CCLPs based on the simplex tableau associated with a basic feasible solution of the linear relaxation. To this end, we characterized the closed convex hull of a disjunctive relaxation of the tableau. This disjunctive relaxation is obtained by taking $K+1$ non-zero components of the basic feasible solution and imposing that the corresponding basic variables are nonpositive. The result can be used to improve the c-max cut, a popular disjunctive cut in the literature, and to generalize the E&R procedure recently developed for complementarity problems [56]. Facet-defining inequalities for the closed convex hull were shown to correspond to spanning trees with a special structure we call *label-connectivity*. This construction enables us to design a polynomial time cut-strengthening algorithm.

We next studied sparse PCA, which is the problem of finding a sparse eigenvector that explains most of the variance of some data. The original optimization problem was reformulated as a convex maximization problem and semidefinite relaxations are introduced. The construction is based on the fact that the feasible set is permutation-invariant. The convex hull was written through a majorization inequality that can be modeled with a polynomial number of additional variables and linear constraints. Our SDP relaxations were shown to be tighter than that proposed in [24] and preliminary computational experiments showed that considerable portion of gaps remaining in the SDP relaxation in [24] are eliminated by our SDP relaxations.

Lastly, we studied CCLPs. A facial disjunctive program reformulation was presented to take advantage of Jeroslow's finite-convergent cutting plane algorithm [40].

We also generalized RLT to the cardinality setting and proposed to use product factors as ratios of multilinear terms. As a special type of CCLPs, we investigated valid inequalities for CCKPs based on disjunctions for the cardinality constraint.

5.2 Future research directions

Implementing and improving the tableau-based cut-generating procedure

While the procedure we described in Chapter 2 generates a valid inequality that cuts off any given basic feasible solution to an LP relaxation that does not satisfy the cardinality constraint, the convergence of the resulting cutting plane algorithm to an optimal solution to the CCLP has not yet been studied. Further, it would be valuable to conduct computational experiments to evaluate the strength of the generated cutting planes. Finally, the choice of disjunctive relaxation for the tableau depends on which $K+1$ basic variables are selected. It would be interesting to study which choices of $K+1$ basic variables can be shown to be superior to others. Another avenue of future work is to generalize disjunction we studied to one that forces m out of $K+m$ variables to be nonpositive.

Sparse vector recovery from an optimal matrix solution

We showed that sparse PCA can be reformulated as a convex maximization problem by verifying that a sparse vector can be recovered from a non-sparse vector solution. In SDP relaxations, however, a matrix optimal solution possesses more useful information about the global solution and returns a better objective value. Therefore, it would be desirable to find a sparse vector that explains as much variance as the matrix solution does. This question is closely related to sparse rank-1 matrix recovery.

Relaxation for optimization for multiple sparse eigenvectors

Sparse PCA aims to find a single sparse vector that explains most of the variance of some data. In practice, it is often important to compute multiple sparse eigenvectors that explains a majority of the total variance. Current methods to find multiple sparse eigenvectors are mostly based on greedy or heuristic algorithms. According to the definition of total variance explained by a set of variables in [63], we can formulate this problem as

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr}(\Sigma P) \\ \text{subject to} \quad & P = V(V^\intercal V)^{-1}V^\intercal, \;\; V = \begin{bmatrix} v_1 & \dots & v_m \end{bmatrix}, \\ & \operatorname{card}(v_i) \leq K, i = 1, \dots, m, \\ & \|v_i\| = 1, i = 1, \dots, m. \end{aligned}$$

When $m = 2$, it can be written as

$$\begin{aligned} \text{maximize} \quad & \frac{1}{1-s^2}[x^\intercal \Sigma x + y^\intercal \Sigma y - s(x^\intercal \Sigma y + y^\intercal \Sigma x)] \\ \text{subject to} \quad & s = x^\intercal y, \\ & \operatorname{card}(x) \leq K, \;\; \operatorname{card}(y) \leq K, \\ & \|x\| = \|y\| = 1 \\ & x, y \in \mathbb{R}^n, s \in \mathbb{R}. \end{aligned}$$

One of our next goals is to design tractable convex relaxations for this formulation.

Convexification of CCKPs

In spite of its structural simplicity, little is known about the convex hull of a CCKP. PORTA outputs for small-sized problems show that many of its facet-defining inequalities can be derived using the two-term disjunction approaches we used in the thesis. We plan on investigating further the polyhedral structure of CCKPs to develop valid inequalities for its convex hull.

LIST OF REFERENCES

LIST OF REFERENCES

[1] Pasquale Avella and Antonio Sassano. On the p-median polytope. *Mathematical Programming*, 89(3):395–411, 2001.

[2] Egon Balas. Intersection cuts–a new type of cutting planes for integer programming. *Operations Research*, 19(1):19–39, 1971.

[3] Egon Balas. Disjunctive programming: Properties of the convex hull of feasible points. Technical report, DTIC Document, 1974.

[4] Egon Balas. Disjunctive programming. *Annals of Discrete Mathematics*, 5:3–51, 1979.

[5] Egon Balas. Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. *SIAM Journal on Algebraic Discrete Methods*, 6(3):466–486, 1985.

[6] Egon Balas, Sebastián Ceria, and Gérard Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0–1 programs. *Mathematical Programming*, 58(1-3):295–324, 1993.

[7] Mokhtar S Bazaraa and Hanif D Sherali. On the use of exact and heuristic cutting plane methods for the quadratic assignment problem. *Journal of the Operational Research Society*, 33(11):991–1003, 1982.

[8] Evelyn Martin Lansdowne Beale and John A Tomlin. Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. In *J. Lawrence, ed., Prodeedings of the Fifth International Conference on Operational Research*, pages 447–454. Tavistock Publications, London, 1970.

[9] John E Beasley. A note on solving large p-median problems. *European Journal of Operational Research*, 21(2):270–273, 1985.

[10] John E Beasley, Nigel Meade, and T-J Chang. An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research*, 148(3):621–643, 2003.

[11] Dimitri P Bertsekas, Angelia Nedi, Asuman E Ozdaglar, et al. Convex analysis and optimization. 2003.

[12] Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *Annals of statistics*, 44(2):813–852, 2016.

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

[14] Daniel Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74(2):121–140, 1996.

[15] Garrett Birkhoff. Tres observaciones sobre el algebra lineal. *Universidad Nacional de Tucumán Revista Series A*, 5:147–151, 1946.

[16] Jonathan M Borwein and Adrian S Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

[17] Arne Brøndsted. *An introduction to convex polytopes*, volume 90. Springer, 1983.

[18] T.-J. Chang, Nigel Meade, John E Beasley, and Yazid M Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302, 2000.

[19] Thomas Christof. *Ein verfahren zur transformation zwischen polyederdarstellungen*. PhD thesis, Diplomarbeit, Universität Augsburg, 1991.

[20] Thomas Christof and Andreas Löbel. Porta–polyhedron representation transformation algorithm, 2008. *URL: http://www. zib. de/Optimization/Software/Porta*, 2012.

[21] Nicos Christofides and John E Beasley. A tree search algorithm for the p-median problem. *European Journal of Operational Research*, 10(2):196–204, 1982.

[22] Yves Crama and Michaël Schyns. Simulated annealing for complex portfolio selection problems. *European Journal of Operational Research*, 150(3):546–571, 2003.

[23] Alexandre d'Aspremont, Francis Bach, and Laurent El Ghaoui. Optimal solutions for sparse principal component analysis. *The Journal of Machine Learning Research*, 9:1269–1294, 2008.

[24] Alexandre d'Aspremont, Laurent El Ghaoui, Michael I Jordan, and Gert RG Lanckriet. A direct formulation for sparse PCA using semidefinite programming. *SIAM review*, 49(3):434–448, 2007.

[25] Ismael R de Farias Jr. A family of facets for the uncapacitated p-median polytope. *Operations Research Letters*, 28(4):161–167, 2001.

[26] Ismael R de Farias Jr and George L Nemhauser. A polyhedral study of the cardinality constrained knapsack problem. *Mathematical Programming*, 96(3):439–467, 2003.

[27] Matthias Ehrgott, Kathrin Klamroth, and Christian Schwehm. An MCDM approach to portfolio optimization. *European Journal of Operational Research*, 155(3):752–770, 2004.

[28] Manfred Gilli and Evis Këllezi. The threshold accepting heuristic for index tracking. In *Financial Engineering, E-Commerce and Supply Chain*, pages 1–18. Springer, 2002.

[29] Fred Glover. Convexity cuts and cut search. *Operations Research*, 21(1):123–134, 1973.

[30] Michel X Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming*, 153:5–11, 2015.

[31] R. E. Gomory. Some polyhedra related to combinatorial problems. *Linear Algebra and its Application*, 2:341–375, 1969.

[32] Ralph Gomory. An algorithm for the mixed integer problem. Technical report, DTIC Document, 1960.

[33] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[34] Branko Grunbaum, Victor Klee, Micha A Perles, and Geoffrey Colin Shephard. *Convex polytopes*. Springer, 1967.

[35] Yue Guan and Jennifer G Dy. Sparse probabilistic principal component analysis. In *International Conference on Artificial Intelligence and Statistics*, pages 185–192, 2009.

[36] Osman Güler. *Foundations of optimization*, volume 258. Springer Science & Business Media, 2010.

[37] Toshihide Ibaraki. The use of cuts in complementary programming: Technical note. *Operations Research*, 21(1):353–359, 1973.

[38] Pertti Järvinen, Jaakko Rajala, and Heikki Sinervo. A branch-and-bound algorithm for seeking the p-median: Technical note. *Operations Research*, 20(1):173–178, 1972.

[39] J. N. R. Jeffers. Two case studies in the application of principal component analysis. *Applied Statistics*, pages 225–236, 1967.

[40] Robert G Jeroslow. A cutting-plane game for facial disjunctive programs. *SIAM Journal on Control and Optimization*, 18(3):264–281, 1980.

[41] Roland Jeurissen and Jan Van den Berg. Index tracking using a hybrid genetic algorithm. In *2005 ICSC Congress on Computational Intelligence Methods and Applications*, pages 6 pp.–. IEEE, 2005.

[42] Roland Jeurissen and Jan van den Berg. Optimized index tracking using a hybrid genetic algorithm. In *IEEE Congress on Evolutionary Computation (CEC 2008)*, pages 2327–2334. IEEE, 2008.

[43] Norbert J Jobst, Michael D Horniman, Cormac A Lucas, and Gautam Mitra. Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. *Quantitative Finance*, 1(5):489–501, 2001.

[44] Ian T Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[45] Ian T Jolliffe, Nickolay T Trendafilov, and Mudassir Uddin. A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531–547, 2003.

[46] Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre. Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research*, 11:517–553, 2010.

[47] Oded Kariv and S Louis Hakimi. An algorithmic approach to network location problems. ii: The p-medians. *SIAM Journal on Applied Mathematics*, 37(3):539–560, 1979.

[48] RS Laundy. *Some logically constrained mathematical programming problems*. PhD thesis, University of Southampton, 1983.

[49] Andrea Loraschi, Marco Tomassini, A Tettamanzi, and Paolo Verda. Distributed genetic algorithms with an application to portfolio selection problems. In *Artificial neural nets and genetic algorithms*, pages 384–387. Springer, 1995.

[50] Zhaosong Lu and Yong Zhang. An augmented lagrangian approach for sparse principal component analysis. *Mathematical Programming*, pages 1–45, 2012.

[51] Zongming Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772–801, 2013.

[52] Renata Mansini and Maria Grazia Speranza. Heuristic algorithms for the portfolio selection problem with minimum transaction lots. *European Journal of Operational Research*, 114(2):219–233, 1999.

[53] Baback Moghaddam, Yair Weiss, and Shai Avidan. Spectral bounds for sparse pca: Exact and greedy algorithms. In *Advances in neural information processing systems*, pages 915–922, 2005.

[54] Rafael Moral-Escudero, Rubén Ruiz-Torrubiano, and Alberto Suarez. Selection of optimal investment portfolios with cardinality constraints. In *IEEE Congress on Evolutionary Computation (CEC 2006)*, pages 2382–2388. IEEE, 2006.

[55] Subhash C Narula, Ugonnaya I Ogbu, and Haakon M Samuelsson. Technical note - an algorithm for the p-median problem. *Operations Research*, 25(4):709–713, 1977.

[56] Trang T Nguyen. *Convexification Techniques for Complementarity and Multilinear Constraints*. PhD thesis, University of Florida, 2013.

[57] Robert C. Prim. Shortest connection networks and some generalizations. *The Bell Systems Technical Journal*, 36(6):1389–1401, 1957.

[58] Madabhushi Raghavachari. On connections between zero-one integer programming and concave programming under linear constraints. *Operations Research*, 17(4):680–684, 1969.

[59] R Tyrrell Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1997.

[60] Rubén Ruiz-Torrubiano and Alberto Suárez. A hybrid optimization approach to index tracking. *Annals of Operations Research*, 166(1):57–71, 2009.

[61] Issai Schur. Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.

[62] John Shapcott. Index tracking: genetic algorithms for investment portfolio selection. *Edinburgh Parallel Computing Centre*, 1992.

[63] Haipeng Shen and Jianhua Z Huang. Sparse principal component analysis via regularized low rank matrix approximation. *Journal of multivariate analysis*, 99(6):1015–1034, 2008.

[64] Hanif D Sherali and Warren P Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

[65] Barry Simon. *Convexity: An analytic viewpoint*, volume 187. Cambridge University Press, 2011.

[66] Bharath K Sriperumbudur, David A Torres, and Gert RG Lanckriet. Sparse eigen methods by DC programming. In *Proceedings of the 24th international conference on Machine learning*, pages 831–838. ACM, 2007.

[67] Mohit Tawarmalani. Inclusion certificates and simultaneous convexification of functions. 2012. Working paper.

[68] Andreas M Tillmann and Marc E Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, 2014.

[69] Hoang Tuy. Concave programming under linear constraints. *Soviet Mathematics*, 5:1437–1440, 1964.

[70] Joe Warren, Scott Schaefer, Anil N Hirani, and Mathieu Desbrun. Barycentric coordinates for convex sets. *Advances in computational mathematics*, 27(3):319–338, 2007.

[71] B. Zeng and J.-P. P. Richard. A polyhedral study on 0-1 knapsack problems with disjoint cardinality constraints: facet-defining inequalities by sequential lifting. *Discrete Optimization*, 8:277–301, 2011.

[72] B. Zeng and J.-P. P. Richard. A polyhedral study on 0-1 knapsack problems with disjoint cardinality constraints: strong valid inequalities by sequence-independent lifting. *Discrete Optimization*, 8:259–276, 2011.

[73] Yi Zhang, Samuel Burer, and W Nick Street. Ensemble pruning via semi-definite programming. *The Journal of Machine Learning Research*, 7:1315–1338, 2006.

[74] Günter M Ziegler. *Lectures on polytopes*, volume 152. Springer, 1995.

[75] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.

VITA

VITA

Education

Krannert School of Management, Purdue University West Lafayette, IN

Ph.D. in Management (Expected) August 2016

- Major: Quantitative Methods (Minor in Operations Management)
- Advisors: Dr. Mohit Tawarmalani and Dr. Jean-Philippe P. Richard

Department of Statistics, Purdue University West Lafayette, IN

M.S. in Statistics (Mathematical Statistics) May 2011

Department of Mathematics, Purdue University West Lafayette, IN

M.S. in Mathematics (Applied Mathematics) May 2009

Mathematics and Statistics, Chonnam National University South Korea

B.S. in Mathematics, *Summa Cum Laude* August 2006

Research Interests

Convex and nonconvex optimization, operations research with applications in business, economics, and engineering. Research emphasis is on optimization problems with cardinality (sparsity) constraints

Work Experiences

- **Mitchell College of Business, University of South Alabama** Mobile, AL
  - **Assistant Professor of Management** (tenure-track) (Expected) 2016-

- **Krannert School of Management, Purdue University** West Lafayette, IN
  - **Research Assistant** 2011-2016
    - Under NSF CMMI 0900065 (Principal Investigator: Dr. Mohit Tawarmalani)
      Collaborative Research: Generating Stronger Cuts for Nonlinear Programs Via Orthogonal Disjunctions and Lifting Techniques, 2011-2013
    - Under NSF CMMI 1234897 (Principal Investigator: Dr. Mohit Tawarmalani)
      Collaborative Research: Novel Tighter Relaxations for Complementarity Constraints with Applications to Nonlinear and Bilevel Programming, 2013-2014
    - Under research grant from Purdue Research Foundation
      Convexification Techniques for Cardinality Constrainted Optimization Problems, 2014-2015
  - **Graduate Teaching Instructor:** Business statistics 2013 Summer, 2014 Summer
- **Department of Mathematics, Purdue University** West Lafayette, IN
  - **Graduate Teaching Instructor** 2007-2011
    - Lecturer and reciatation instructor
    - Subjects: engineering calculus, business calculus, algebra
    - Peer mentor for teaching assistants (TA screening committee)
- **Department of Mathematics, Chonnam National University** South Korea
  - **Research Assistant** 2003-2006
    - Parallel Computing Lab under Brain Korea 21

Teaching Experiences

Term	Course number and subject		Role	Evaluation
2008 Spring	MA 261	Multivariate Calculus	Recitation Instructor	4.6/5.0
2008 Fall	MA 162	Calculus II	Recitation Instructor	4.7/5.0
2009 Spring	MA 153	Algebra and Trigonometry I	Lecturer	4.3/5.0
2009 Fall	MA 153	Algebra and Trigonometry I	Lecturer	4.7/5.0
2010 Spring	MA 261	Multivariate Calculus	Recitation Instructor	4.7/5.0
2010 Fall	MA 162	Calculus II	Recitation Instructor	4.7/5.0
2011 Spring	MA 261	Multivariate Calculus	Recitation Instructor	4.9/5.0
2013 Summer	MGMT 305	Business Statistics	Lecturer	4.5/5.0
2014 Summer	MGMT 305	Business Statistics	Lecturer	4.6/5.0

- Role as a lecturer: Design syllabus, prepare teaching material and exams, perform office hours, and give final grades
- Role as a recitation lecturer: Prepare quizzes, answer homework questions, perform office hours

Publication

- Kim J., Tawarmalani M., Richard J.-P. P. , *On Cutting Planes for Cardinality Constrained Optimization Problems*, submitted to *Mathematical Programming*
- Kim J., Tawarmalani M., Richard J.-P. P. , *Convexification of permutation-invariant sets and application*, work in progress.
- Kim J., Tawarmalani M., Richard J.-P. P. , *Semidefinite programming relaxation for sparse principal component analysis*, work in progress.

Presentation

- Kim J., Tawarmalani M., and Richard J.-P. P. , *Cardinality constrained linear program: facial disjunctive formulation and valid inequalities*, INFORMS Annual Meeting, Minneapolis, MN, 2013
- Kim J., Tawarmalani M., and Richard J.-P. P., *On cutting planes for cardinality constrained optimization problems*, INFORMS Annual Meeting, San Francisco, 2014
- Kim J., Tawarmalani M., and Richard J.-P. P., *A cut generating procedure for cardinality constrained optimization problems (CCOP)*, ISMP, Pittsburgh, 2015
- Tawarmalani M., Kim J., and Richard J.-P. P., *Convex hull of permutation invariant sets and applications*, ISIM, West Lafayette, 2015
- Kim J., Tawarmalani M., and Richard J.-P. P., *Sparse principal component analysis (sparse PCA) via convexification*, INFORMS Annual Meeting, Philadelphia, 2015
- Tawarmalani M., Kim J., and Richard J.-P. P., *Exploiting permutation invariance to construct tight relaxations*, 2016 MILP Workshop, Coral Gables, 2016

Honors and Awards

- Krannert Certificates for Distinguished and Outstanding Teaching Awards, MGMT 305 Business Statistics, Summer 2013 and Summer 2014

Grants

- Purdue Research Foundation research grant, .50 FTE Assistantship, 2014-2015
  Title: *Convexification Techniques for Cardinality Constrainted Optimization Problems*
  Authored the proposal under supervision of Dr. Mohit Tawarmalani

Professional Memberships

- INFORMS Student Member, Fall 2013 to Present
- INFORMS Optimization Society, Fall 2013 to Present
- INFORMS Computing Society, Fall 2013 to Present
- INFORMS Chapter at Purdue University

Graduate Coursework

- **Operations Research:** Integer Programming, Convex Optimization, Combinatorial Optimization, Convex and Nonconvex Optimization, Nonlinear Optimization Algorithms and Models, Theoretical Foundations of Optimization, Computational Methods in Optimization, Optimization Modeling with Spreadsheets, Optimization Game Theory and Uncertainty, Network and Game Theory, Mathematical Analysis for Economists, Stochastic methods in Operations Management, Operations Management, Stochastic Models in Operations Research
- **Mathematics:** Linear Algebra, Numerical Analysis, Functional Analysis, Real Analysis and Measure Theory, Methods of Applied Mathematics, Abstract Algebra, Mathematics of Finance, Probability Theory, Partial Differential Equations, Complex Analysis
- **Statistics:** Applied Regression Analysis, Mathematical Statistics, Computational Statistics, Stochastic Processes, Theory of Linear Models, Design of Experiment, Statistical Methodology, Time Series and Applications, Applied Multivariate Analysis

Computational Experience

- Programming: C, R, Matlab
- Optimization Software: IBM CPLEX Callable Library, Gurobi, cvx, Sedumi, PORTA
- Statistics: SAS, R, SPSS, Minitab

References

- **Dr. Mohit Tawarmalani**

  Professor and Allison and Nancy Schleicher Chair of Management

  Krannert School of Management

  Purdue University

  `http://web.ics.purdue.edu/~mtawarma/`

- **Dr. Jean-Philippe P. Richard**

  Professor

  Industrial and Systems Engineering

  University of Florida

  `http://www.ise.ufl.edu/richard/`

- **Dr. Yanjun Li**

  Professor

  Krannert School of Management

  Purdue University

  `http://www.krannert.purdue.edu/directory/bio.php?username=li14`