2013

Mixed Integer Second Order Cone Optimization, Disjunctive Conic Cuts: Theory and experiments

Julio Cesar Goez
Lehigh University

Mixed Integer Second Order Cone Optimization Disjunctive Conic Cuts: Theory and experiments

by

Julio C. Góez

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Industrial Engineering

Lehigh University

September 2013

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date

Dr. Tamás Terlaky
Dissertation Director

Accepted Date

Committee Members:

Dr. Tamás Terlaky, Committee Chair

Dr. Pietro Belotti

Dr. Frank Curtis

Dr. Imre Pólik

Dr. Ted Ralphs

Dr. Reha Tütüncü

Acknowledgements

The final result of this thesis would not be possible without the support of my family and the professors, the fellow graduate students, and the staff of the Industrial and Systems Engineering Department of Lehigh University. I thank all of them for making my life at Lehigh a memorable experience. I would like to give special thanks to my adviser Professor Tamás Terlaky for his patience and guidance through my Ph.D. Also to my beloved wife Diana C. Guerrero, for her love, patience, and support during the time of my Ph.D. I thank the members of my Ph.D. committee Dr. Pietro Belotti, Dr. Frank Curtis, Dr. Imre Pólik, Dr. Ted Ralphs, and Dr. Reha Tütücü, for their constructive feedback, guidance, and their contribution with ideas for the development of the result accomplished in this thesis. I happily acknowledge COR@l Lab for providing the hardware and software needed for the developments and testing of the ideas proposed here. I would like to thank my fellow graduate students Guillermo Bobadilla, Lynne Erickson, Gaurav Gulati, Ashutosh Mahajan, Camilo Mancilla, Melissa Marenus, Orest Pasichnyk, and Austin Szatkowski, for their support and kindness, which made my life at Lehigh a wonderful experience. Finally, I thank Dr. Germán Riaño and Dr. Andrés Medaglia for their support and encouragement to start my Ph.D. and Dr. Luis F. Zuluaga and Dr. Miguel F. Anjos for their encouragement during the last two years of my Ph.D.

Contents

List of Figures

Abstract

Mixed Integer Second Order Cone Optimization (MISOCO) problems allow practitioners to mathematically describe a wide variety of real world engineering problems including supply chain, finance, and networks design. A MISOCO problem minimizes a linear function over the set of solutions of a system of linear equations and the Cartesian product of second order cones of various dimensions, where a subset of the variables is constrained to be integer. This thesis presents a technique to derive inequalities that help to obtain a tighter mathematical description of the feasible set of a MISOCO problem. This improved description of the problem usually leads to accelerate the process of finding its optimal solution. In this work we extend the ideas of disjunctive programming, originally developed for mixed integer linear optimization, to the case of MISOCO problems. The extension presented here results in the derivation of a novel methodology that we call *disjunctive conic cuts* for MISOCO problems. The analysis developed in this thesis is separated in three parts. In the first part, we introduce the formal definition of disjunctive conic cuts. Additionally, we show that under some mild assumptions there is a necessary and sufficient condition that helps to identify a disjunctive conic cut for a given convex set. The main appeal of this condition is that it can be easily verified in the case of MISOCO problems. In the second part, we study the geometry of sets defined by a single quadratic inequality. We show that for some of these sets it is possible to derive a close form to build a disjunctive conic cut. In the third part, we show that the feasible set of a MISOCO problem

with a single cone can be characterized using sets that are defined by a single quadratic inequality. Then, we present the results that provide the criteria for the derivation of disjunctive conic cuts for MISOCO problems. Preliminary numerical experiments with our disjunctive conic cuts used in a branch-and-cut framework provide encouraging results where this novel methodology helped to solve MISOCO problems more efficiently. We close our discussion in this thesis providing some highlights about the questions that we consider worth pursuing for future research.

Notation

m, n, $\ell, \ldots$	Indices are natural numbers and they are denoted with lower case letters.
$\mathbb{R}$	The blackboard bold R denotes the set of real numbers.
$\mathbb{Z}$	The blackboard bold Z denotes the set of integer numbers.
α, $\beta, \ldots$	Scalars are denoted with Greek letters.
$\lvert\alpha\rvert$	The two vertical bars denote the absolute value of the scalar $\alpha \in \mathbb{R}$.
a, b, $x, \ldots$	Vectors are denoted with lower case letters and they are assumed to be column vectors.
x_i	The subindex in this notations denotes the i-th component of vector x.
$x_{2:n}$	The colon in this notation denotes the vector formed with components 2 to n of vector x.
$\lVert x\rVert$	The two double vertical bars denote the norm of vector x; all norms in this thesis are assumed to be Euclidean.
e_i	This denotes a column vector that has all its components equal to zero except for the i-th component that is equal to 1.
A, $B, \ldots$	Matrices are denoted with capital letters.

$A_{i:}$	The colon in this notation is used to denote the row i of matrix A.
$A_{:i}$	The colon in this notation is used to denote the column i of matrix A.
$P \succ 0$	The curly greater than denotes that the matrix P is positive definite.
$P \succeq 0$	The curly greater than or equal denotes that the matrix P is positive semi-definite.
ID1	This denotes that a matrix is indefinite with exactly one negative eigenvalue, and all other eigenvalues are positive.
$\mathcal{A}$, $\mathcal{B}, \ldots$	Sets are denoted with calligraphic letters.
$\overline{\mathcal{A}}$	The upper bar in this notation denotes the complement of the set $\mathcal{A}$.
$\mathbb{L}^n$	The blackboard bold L denotes a second order cone (Lorentz cone), the superscript gives the dimension of the cone.
(P, p, ρ)	This triplet represents the set $\{x \in \mathbb{R}^n \mid x^\top P x + 2p^\top x + \rho \leq 0\}$.

Chapter 1

Introduction

A Mixed Integer Second Order Cone Optimization (MISOCO) problem is that of minimizing a linear function over the set of solutions of a system of linear equations and the Cartesian product of second order cones of various dimensions, where a subset of the variables is constrained to be integer. Specifically, a MISOCO problem is given as

$$\begin{aligned} \text{minimize: } & c^\top x \\ \text{subject to: } & Ax = b \\ & x \in \mathcal{K} \\ & x \in \mathbb{Z}^d \times \mathbb{R}^{n-d}, \end{aligned} \tag{MISOCO}$$

where $A \in \mathbb{R}^{m\times n}$, with $\text{rank}(A) = m$; $c \in \mathbb{R}^n$; $b \in \mathbb{R}^m$; and $x = \left((x^1)^\top, (x^2)^\top, \ldots, (x^k)^\top\right)^\top$; $x^i \in \mathbb{R}^{n_i}$; $\mathcal{K} = \mathbb{L}_1^{n_1} \times \cdots \times \mathbb{L}_k^{n_k}$; $\mathbb{L}^{n_i} = \{x^i \in \mathbb{R}^{n_i} \mid x_1^i \geq \|x_{2:n_i}^i\|\}$, for $i = 1, \ldots, k$, with $\sum_{i=1}^k n_i = n$.

There are many areas of engineering in which applications of MISOCO problems arise. In computer vision models, Kumar et al. [2006] used a MISOCO problem as a relaxation of Markov random fields. Atamtürk et al. [2012] consider the design of a supply chain system

where a supplier ships products to different retailers, each with random demand.They reformulate these joint location-inventory models as MISOCO problems. Aktürk et al. [2009] strengthen the formulation of a machine-job assignment problem with separable convex costs using a polynomial number of conic quadratic constraints. The design of telecommunication networks with a minimum length connection network is a Euclidean Steiner tree problem, for which Fampa and Maculan [2004] present a MISOCO relaxation. Cheng et al. [2012] consider the problem of joint base station selection and multi-cell beam-forming and present a MISOCO formulation for this problem. In finance, cardinality-constrained portfolio optimization gives another example of applications of MISOCO problems [Bertsimas and Shioda, 2009]. Finally, turbine balancing is an engineering problem that can be formulated as a MISOCO problem as it is discussed by Drewes [2009], and White [1996]. These references represent a set of problems in the literature that motivate the research and underline the necessity of solving MISOCO problems.

During the last decade, a number of techniques have been developed to solve MISOCO problems. Most of these developments have aimed at extending results shown previously to be effective for Mixed Integer Linear Optimization (MILO) problems to the case of MISOCO problems. One of these approaches uses outer linear approximations of second order cones. Vielma et al. [2008], and Vielma [2009] used the polynomial-size polyhedral relaxation introduced by Ben-Tal and Nemirovski [2001b] in their "lifted linear programming" branch-and-bound algorithm for MISOCO problems. Krokhmal and Soberanis [2010] generalized this approach to integer p-order conic optimization. Drewes [2009] presented subgradient-based linear outer approximations for the second order cone constraints. This allows one to approximate the MISOCO problem by a MILO problem in a hybrid outer approximation branch-and-bound algorithm.

Stubbs and Mehrotra [1999] generalized the lift-and-project algorithm of Balas et al. [1993] for 0-1 MILO to 0-1 mixed integer convex optimization. Later, Çezik and Iyengar

[2005] investigated the generation of valid convex cuts for 0-1 Mixed Integer Conic Optimization (MICO) problems problems and discussed how to extend the Chvátal-Gomory [Gomory, 1958] procedure for generating linear cuts for MICO problems and the extension of *lift-and-project* techniques for MICO problems. In particular, they showed how to generate linear and convex quadratic valid inequalities using the relaxation obtained by a projection procedure. Recently, Drewes [2009] reviews the ideas proposed by Çezik and Iyengar [2005] and Stubbs and Mehrotra [1999] and applies them to MISOCO problems.

Atamtürk and Narayanan [2010, 2011] proposed two procedures for MISOCO problems to generate cuts. They first studied a generic lifting procedure for MICO, and then extended the *Mixed integer rounding* [Nemhauser and Wolsey, 1990, 1999] procedure to the MISOCO case. The main idea of the procedure is to reformulate a second order conic constraint using a set of two-dimensional second order cones. In this new reformulation the set of inequalities are called *polyhedral second-order conic constraint*. The authors used polyhedral analysis for studying these inequalities separately. This allowed the derivation of a mixed-integer rounding procedure, which yields a *nonlinear conic mixed-integer rounding*. A generalization of the use of polyhedral second-order conic constraints is presented by Masihabadi et al. Sanjeevi [2012].

Dadush et al. [2011] studied the *split closure* of a strictly convex body. In their work a *conic quadratic inequality* is presented as an example of a non-polyhedral split closure. In particular, the authors showed that it is necessary to consider conic quadratic inequalities in order to describe the split closure of an ellipsoid. This independently yielded a *conic quadratic inequality* that coincides with the cylindrical and conic cut for MISOCO problems presented in Chapter 4 for cases where the feasible set of the relaxation is an ellipsoid.

1.1 Background

In this section, we present some fundamental concepts of convex analysis and mixed integer optimization that are used in the developments of this thesis. This is not intended to be a comprehensive review of these areas. We provide some references for the reader interested in the proofs or additional results not contained in this section.

1.1.1 Convex analysis

We summarize the most relevant definitions and results of convex analysis that are used in this dissertation. For detailed presentation of convex analysis, the interested reader can review [Barvinok, 2002, Boyd and Vandenberghe, 2006, Rockafellar, 1970]. The results presented in this section are well known in the field of convex analysis. For that reason, they are stated here without proof.

We start with the definition of the concepts of affine combination and convex combination.

Definition 1.1 (Affine combination)**.** *Given a set of points* $u^1, \dots, u^t \in \mathbb{R}^n$*, a point*

$$v = \sum_{i=1}^{t} \lambda_i u^i, \text{ where } \lambda_i \in \mathbb{R}, \ i = 1, \dots, t, \text{ and } \sum_{i=1}^{t} \lambda_i = 1$$

is called an affine combination of $u^1, \dots, u^t$*.*

Geometrically speaking, the set of all affine combinations of two given vectors $u, v \in \mathbb{R}^n$ gives the line determined by these two points. Now, we can define an affine set based on Definition 1.1.

Definition 1.2 (Affine set)**.** *A set* $\mathcal{L} \in \mathbb{R}^n$ *is said to be affine if for any two given points*

$u, v \in \mathcal{L}$ and $\lambda \in \mathbb{R}$ we have that

$$(1-\lambda)u + \lambda v = u + \lambda(v-u) \in \mathcal{L}. \tag{1.1}$$

Definition 1.3 (Affine Hull)**.** *Given a set $\mathcal{S} \in \mathbb{R}^n$, the affine hull $\mathcal{L} = \text{aff}(\mathcal{S})$ of $\mathcal{S}$ is given by the set of all affine combinations of all points in $\mathcal{S}$.*

The following is a key theorem for the analysis of the feasible set of (MISOCO). The proof of Theorem 1.1 can be found in Rockafellar [1970].

Theorem 1.1. *Given $d \in \mathbb{R}^m$ and $D \in \mathbb{R}^{m \times n}$, the set*

$$\mathcal{L} = \{u \in \mathbb{R}^n \mid Du = d\} \tag{1.2}$$

is an affine set in $\mathbb{R}^n$. Furthermore, every affine set may be represented in this way.

Hence, the feasible set of (MISOCO) is the intersection of an affine set and the cone $\mathcal{K}$. Note also that a half-space is a special affine set given by a single linear equality, i.e., $A^{=} = \{u \in \mathbb{R}^n \mid a^\top u = \alpha\}$, for some $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Now we define an affine transformation.

Definition 1.4 (Affine transformation)**.** *A mapping $\mathsf{L} : \mathbb{R}^n \to \mathbb{R}^m$ is called an affine transformation if*

$$\mathsf{L}((1-\lambda)u + \lambda v) = (1-\lambda)\mathsf{L}(u) + \lambda \mathsf{L}(v),$$

for every $u, v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

An important property of affine transformations is that they preserve collinearity and ratios between distances. In other words, if two points lie in a line, they will still lie in a line after the transformation, and the mid point of a line segment will stay the midpoint after the transformation. Hence, after an affine transformation parallel lines will remain

parallel. The following theorem provides the explicit form of affine transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. The proof of this theorem is given in Rockafellar [1970].

Theorem 1.2. *The affine transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are the mappings L of the form $\mathsf{L}(u) = Mu + d$, where $u \in \mathbb{R}^n$, $M \in \mathbb{R}^{m\times n}$ and $d \in \mathbb{R}^m$.*

We now define a convex combination, which is a fundamental concept used in this dissertation.

Definition 1.5 (Convex combination)**.** *Given a set of vectors $u^1, \ldots, u^t \in \mathbb{R}^n$, the vector*

$$u = \sum_{i=1}^{t} \lambda_i u^i, \text{ where } \lambda_i \in \mathbb{R}, \ \lambda_i \geq 0, \ i = 1, \ldots, t, \text{ and } \sum_{i=1}^{t} \lambda_i = 1,$$

is called a convex combination of $u^1, \ldots, u^t$.

Geometrically speaking, the set of all convex combinations of two given vectors $u, v \in \mathbb{R}^n$ is the line segment connecting the two vectors. Now, we can define a convex set based on Definition 1.5.

Definition 1.6 (Convex set)**.** *A set $\mathcal{S} \in \mathbb{R}^n$ is said to be convex if for any two given vectors $u^1, u^2 \in \mathcal{S}$ and $0 \leq \lambda \leq 1$ we have that*

$$(1 - \lambda)u^1 + \lambda u^2 \in \mathcal{S}.$$

Definition 1.7 (Convex hull)**.** *The convex hull* $\text{conv}(\mathcal{S})$ *of a set $\mathcal{S} \in \mathbb{R}^n$ is given by the set of all convex combinations of all finite subsets of points of $\mathcal{S}$.*

Theorem 1.3. *The intersection of a given collection of convex sets is convex.*

Theorem 1.4. *Given a set $\mathcal{S} \in \mathbb{R}^n$, the set* $\text{conv}(\mathcal{S})$ *is the smallest convex set containing*

$\mathcal{S}$, *i.e.*,

$$\text{conv}(\mathcal{S}) = \bigcap_{\substack{\mathcal{U} \supseteq \mathcal{S} \\ \mathcal{U} \text{ convex}}} \mathcal{U}.$$

Theorem 1.4 describes a key property of convex hulls, which is used in the analysis of the conic cuts introduced in Chapter 2. Now, we introduced three standard results from convex analysis needed in Chapter 2. Before doing so, we recall that a hyperplane $\mathcal{A}^{=} = \{u \in \mathbb{R}^n \mid a^\top u = \alpha\}$ defines two half-spaces $\mathcal{A}^{+} = \{u \in \mathbb{R}^n \mid a^\top u \geq \alpha\}$ and $\mathcal{A}^{-} = \{u \in \mathbb{R}^n \mid a^\top u \leq \alpha\}$.

Definition 1.8 (Separation)**.** *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be non-empty sets in $\mathbb{R}^n$. A hyperplane $\mathcal{A}^{=}$ is said to* separate *$\mathcal{S}_1$ and $\mathcal{S}_2$ if $\mathcal{S}_1$ is contained in one of the closed half-spaces $\mathcal{A}^{+}$ or $\mathcal{A}^{-}$ and $\mathcal{S}_2$ lies in the other closed half-space. The hyperplane $\mathcal{A}^{=}$ is said to separate $\mathcal{S}_1$ and $\mathcal{S}_2$* properly *if not* both *$\mathcal{S}_1$ and $\mathcal{S}_2$ are contained in $\mathcal{A}^{=}$ itself.*

Definition 1.9 (Supporting half-space and supporting hyperplane)**.** *Let $\mathcal{S} \subset \mathbb{R}^n$ be a closed convex set. A supporting half-space $\mathcal{A}$ to $\mathcal{S}$ is a closed half-space which contains $\mathcal{S}$ and has a point of $\mathcal{S}$ in its boundary hyperplane $\mathcal{A}^{=}$. The hyperplane $\mathcal{A}^{=}$ is called a* supporting hyperplane *to $\mathcal{S}$.*

Definition 1.10 (Exposed Face)**.** *Let $\mathcal{S} \subset \mathbb{R}^n$ be a closed convex set. A set $\mathcal{S}_f \subset \mathcal{S}$ is called a* exposed face *of $\mathcal{S}$ if there exist a supporting hyperplane $\mathcal{A}^{=}$ such that $\mathcal{S}_f = \mathcal{S} \cap \mathcal{A}^{=}$. The set $\mathcal{S}_f$ may be empty. A non-empty face $\mathcal{S}_f \neq \mathcal{S}$ is called a proper face of $\mathcal{S}$.*

Before presenting the next separation result we need to introduce the definition of relative interior. The relative interior of a set $\mathcal{S} \in \mathbb{R}^n$, denoted by $\text{ri}(\mathcal{S})$, is defined as

$$\text{ri}(\mathcal{S}) = \{x \in \mathcal{S} \mid \exists \varepsilon > 0, \text{ such that } \mathcal{O}(x, \varepsilon) \cap \text{aff}(\mathcal{S}) \subset \mathcal{S}\},$$

where $\mathcal{O}(x, \varepsilon) = \{y \in \mathbb{R}^n \mid \|y - x\| \leq \varepsilon\}$ is the ball of radius ε with center $x \in \mathbb{R}^n$. In other words, the relative interior of $\mathcal{S}$ is its interior relative to $\text{aff}(\mathcal{S})$. We can now state the

following theorem about the existence of a separation hyperplane, the proof can be found in Rockafellar [1970].

Theorem 1.5. *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two non-empty convex sets in $\mathbb{R}^n$. Then, there exists a hyperplane separating $\mathcal{S}_1$ and $\mathcal{S}_2$ properly, if and only if* $\mathrm{ri}(\mathcal{S}_1) \cap \mathrm{ri}(\mathcal{S}_2) = \emptyset$.

We now introduce the definition of a cone, which is a fundamental concept used for the definition of the cuts discussed in this dissertation.

Definition 1.11 (Cone)**.** *We say that $\mathcal{K} \subseteq \mathbb{R}^n$ is a cone if $0 \in \mathcal{K}$ and if for all $x \in \mathcal{K}$ and $\lambda \geq 0$ we have $\lambda x \in \mathcal{K}$.*

If a cone $\mathcal{K}$ is a convex set, then it is called a convex cone. Alternative, we can use the following definition of a convex cone.

Definition 1.12 (Convex Cone)**.** *A set $\mathcal{K} \subseteq \mathbb{R}^n$ is a convex cone if for any two points $u, v \in \mathcal{K}$ and for any $\theta, \vartheta \geq 0$, we have $\theta u + \vartheta v \in \mathcal{K}$.*

Definition 1.13 (Ray)**.** *Given a cone $\mathcal{K} \in \mathbb{R}^n$ and $u \in \mathcal{K}$, then the set $\mathcal{R}_u = \{\lambda u \mid \lambda \geq 0\}$ is called a ray of $\mathcal{K}$.*

Definition 1.14 (Extreme Ray)**.** *Let $\mathcal{K} \in \mathbb{R}^n$ be a cone and $\mathcal{R}_u$ be a ray of $\mathcal{K}$. We say that $\mathcal{R}_u$ is an extreme ray of $\mathcal{K}$ if for any $u \in \mathcal{R}_u$ and any $v, w \in \mathcal{K}$, whenever $u = \lambda v + (1-\lambda)w$ for some $0 \leq \lambda \leq 1$, we must have $u, v \in \mathcal{R}_u$.*

Note that an extreme ray of a convex cone is always an exposed face of the convex cone. We use the concept of a pointed cone in the proofs presented in Chapter 2. Here we present its formal definition.

Definition 1.15 (Pointed Cone)**.** *A cone $\mathcal{K} \in \mathbb{R}^n$ is pointed if it contains no line, i.e., $u \in \mathcal{K}$ and $-u \in \mathcal{K}$ only if $u = 0$.*

We provide a separation theorem that is used in the proofs of Chapter 2.

Theorem 1.6. *Let $\mathcal{S}$ be a non-empty set in $\mathbb{R}^n$, and $\mathcal{K}$ be a cone in $\mathbb{R}^n$. If there exist a hyperplane that separates $\mathcal{S}$ and $\mathcal{K}$ properly, then there exists a hyperplane that separates $\mathcal{S}$ and $\mathcal{K}$ properly and the hyperplane passes through the origin.*

We close this summary with the definition of a convex cylinder as is used in Section 2.2, which a fundamental concept used for the definition of the cuts discussed in this dissertation.

Definition 1.16 (Convex Cylinder)**.** *Given a convex set $\mathcal{D} \subset \mathbb{R}^n$ and a vector $d^0 \in \mathbb{R}^n$, the set $\mathcal{C} = \{x \in \mathbb{R}^n \mid x = d + \lambda d^0, d \in \mathcal{D}, \lambda \in \mathbb{R}\}$ is called a convex cylinder in $\mathbb{R}^n$. The vector d^0 is called the direction of $\mathcal{C}$.*

1.1.2 Quadrics

In Chapter 3 we use the concept of quadric sets to analyze the geometry of the feasible set of MISOCO problems.

Definition 1.17 (Quadric)**.** *Let $P \in \mathbb{R}^{n\times n}$, $p, w \in \mathbb{R}^n$ and $\rho \in \mathbb{R}$, then the quadric $\mathcal{Q}$ is the set defined as*

$$\mathcal{Q} = \{w \in \mathbb{R}^n \mid w^\top P w + 2p^\top w + \rho \leq 0\}. \tag{1.3}$$

This definition is based on the definition of quadric surfaces. In this work, we limit our interest to quadrics where the matrix P is symmetric and it has at most one negative eigenvalue. More comprehensive study of quadric surfaces can be found in Snyder and Sisam Snyder and Sisam [1914], Cox et al. [Cox et al., 1997, Chp. 8], and in Harris [Harris, 1992, Chp. 22]. In what follows, we use the triplet (P, p, ρ) as a simplified representation of a quadric.

1.1.2.1 Shapes of quadrics where P has at most one non-positive eigenvalue

In the analysis of the feasible set of MISOCO problems we need the analysis of quadrics where the matrix P is symmetric and it has at most one non-positive eigenvalue. For that reason, in this section we describe the possible shapes for quadrics in this category. We start by introducing a key concept needed in the classification of such quadrics.

Definition 1.18 (Inertia Meyer [2000])**.** *The inertia* $\text{In}(P)$ *of a symmetric matrix* $P \in \mathbb{R}^{n\times n}$ *is defined by the triplet* (ϑ, ψ, ι)*, in which* ϑ *is the number of negative eigenvalues,* ψ *is the number of zero eigenvalues, and* ι *is the number of positive eigenvalues.*

The classification for the shapes of the quadric (P, p, ρ) is determined by two quantifiers: the inertia of matrix P and the quantity $p^\top P^{-1}p - \rho$. Note that if P has $k > 1$ zero eigenvalues, we can use the eigenvalue decomposition of P to express $\mathcal{Q}$ as the intersection of an affine set and a quadric in $\mathbb{R}^{n-k+1}$. For that reason, we assume w.l.o.g. that the zero eigenvalue has multiplicity 1. We consider three cases:

1. Let us assume first that P is non-singular. Then we can rewrite the defining inequality in (1.3) as

$$\left(w + P^{-1}p\right)^\top P \left(w + P^{-1}p\right) \leq p^\top P^{-1}p - \rho, \tag{1.4}$$

and either $P \succ 0$ or P is indefinite with exactly one negative eigenvalue (ID1). The possible shapes of the quadric (P, p, ρ) in this case are summarized in the following table:

	$p^\top P^{-1}p - \rho$		
	> 0	$= 0$	< 0
P is PD	ellipsoid	point	empty set
P is ID1	hyperboloid of one sheet	cone	hyperboloid of two sheets

Table 1.1: Shapes of the quadric $\mathcal{Q}$ when the matrix P is non-singular.

In all of these cases, either the center of the ellipsoid or the intersection of the asymptotes of the hyperboloids is at $-P^{-1}p$.

2. Now, assume that P is positive semi-definite ($P \succeq 0$) but not positive definite, i.e., the smallest eigenvalue of P is 0 with multiplicity 1. Then, there are two cases:

   Case 1: If there is a vector w^c such that $Pw^c = -p$, then $\mathcal{Q}$ is:

   - **empty**, if $(w^c)^\top Pw^c - \rho < 0$;
   - **a line through** w^c in the direction of the eigenvector of the zero eigenvalue of P, if $(w^c)^\top Pw^c - \rho = 0$;
   - **a cylinder** with its center line through w^c in the direction of the eigenvector of the zero eigenvalue of P, if $(w^c)^\top Pw^c - \rho > 0$.

   Case 2: If there is no vector w^c such that $Pw^c = -p$, then $\mathcal{Q}$ is a **paraboloid**.

3. Finally, assume that P is indefinte and singular, where $\text{In}(P) = (1, 1, n-2)$. Then, there are two cases:

   Case 1: If there is a vector w^c such that $Pw^c = -p$, then $\mathcal{Q}$ is a cylinder in the direction of the eigenvector of the zero eigenvalue of P, and its cross section is:

   - **a hyperbolic cylinder of one sheet**, if $(w^c)^\top Pw^c - \rho > 0$;
   - **a conic cylinder**, if $(w^c)^\top Pw^c - \rho = 0$;
   - **a hyperbolic cylinder of two sheets**, if $(w^c)^\top Pw^c - \rho < 0$.

   Case 2: If there is no vector w^c such that $Pw^c = -p$, then $\mathcal{Q}$ is a **hyperbolic paraboloid**.

1.1.3 Disjunctive sets

Disjunctive sets are a fundamental concept needed for the development of the analysis in this dissertation. For that reason we give a brief introduction to the concept of disjunctive

sets here. Our discussion is based on the concepts described in Balas [1979], Mahajan [2009]. First we need to introduce the logical operator "or", denoted by $\vee$. In a disjunction $s_1 \vee s_2$ the operands s_1 and s_2 are called the disjuncts of the disjunction. The operands s_1 and s_2 are propositions such that if one of them is true, then the disjunction results in "true". In the context of linear systems, we can define a linear disjunction with respect a given $x \in \mathbb{R}^n$ as follows

$$\bigvee_{i \in \mathcal{I}} D^i x \geq d^i, \tag{1.5}$$

where $D^i \in \mathbb{R}^{m_i \times n}$, $d^i \in \mathbb{R}^{m_i}$, and $\mathcal{I}$ is an index set that may or may not be finite. Then, we say that the disjunction (1.5) is true for $\hat{x} \in \mathbb{R}^n$ if and only if there exist at least one $i \in \mathcal{I}$ such that $D^i \hat{x} \geq d^i$. A *disjunctive set* is given by the set of all points $x \in \mathbb{R}^n$ for which the disjunction (1.5) is true, i.e., it is given by

$$\bigcup_{i \in I} \left\{ x \in \mathbb{R}^n \mid D^i x \geq d^i \right\}.$$

A disjunction is said to be *valid* for (MISOCO) if

$$\left\{ x \in \mathbb{Z}^d \times \mathbb{R}^{n-d} \mid Ax = b, x \in \mathcal{K} \right\} \subseteq \bigcup_{i \in \mathcal{I}} \left\{ x \in \mathbb{R}^n \mid D^i x \geq d^i, Ax = b, x \in \mathcal{K} \right\}.$$

In this thesis we focus on valid linear disjunctions for (MISOCO) of the form

$$a^\top x \leq \alpha \bigvee h^\top x \geq \beta, x \in \mathbb{R}^n, \tag{1.6}$$

where $a, h \in \mathbb{R}^n$, $\alpha, \beta \in \mathbb{R}$, and the vectors $(a^\top, \alpha)$, $(h^\top, \beta)$ are not scalar multiples of each other. The disjunctive set over $\mathbb{R}^n$ associated with (1.6) is given by

$$\left\{ x \in \mathbb{R}^n \mid a^\top x \leq \alpha \right\} \bigcup \left\{ x \in \mathbb{R}^n \mid h^\top x \geq \beta \right\}. \tag{1.7}$$

Now, if we relax the integrality constrain in (MISOCO), then the intersection of the set (1.7) and the continuous relaxation of (MISOCO) is

$$\left\{x \in \mathbb{R}^n \mid Ax = b, a^\top x \leq \alpha, x \in \mathcal{K}\right\} \bigcup \left\{x \in \mathbb{R}^n \mid Ax = b, h^\top x \geq \beta, x \in \mathcal{K}\right\}. \tag{1.8}$$

We assume in this thesis that

$$\left\{x \in \mathbb{R}^n \mid Ax = b, a^\top x \leq \alpha, x \in \mathcal{K}\right\} \bigcap \left\{x \in \mathbb{R}^n \mid Ax = b, h^\top x \geq \beta, x \in \mathcal{K}\right\} = \emptyset. \tag{1.9}$$

Then, the set (1.8) is the union of two disjoint convex sets, which is a non-convex set. One of the goals in this dissertation is to characterize the convex hull of the set (1.8). We show in this work that this characterization yields the derivation of novel conic cuts for MISOCO problems.

1.1.4 Branch-and-Bound algorithm

The algorithm we use to solve a MISOCO problem in this thesis is based on the branch Branch-and-Bound (BB) algorithm. For that reason we present a brief description of the BB algorithm for MISOCO problems in this section.

Before describing the algorithm we need to introduce the continuous relaxation of the MISOCO problem. The main challenge faced when solving the problem (MISOCO) is associated with the integrality constraint in some of its variables. If we relax the integrality requirement in those variables we obtain the continuous relaxation of the MISOCO

problem. This relaxation is a second order cone optimization (SOCO) problem

$$\begin{aligned} \text{minimize: } & c^\top x \\ \text{subject to: } & Ax = b \\ & x \in \mathcal{K} \subset \mathbb{R}^n, \end{aligned} \tag{1.10}$$

where cone $\mathcal{K}$ is defined in page 5. For that reason we will refer to the continuous relaxation of the MISOCO problem as its SOCO relaxation. The SOCO is a well studied problem, and there are polynomial time algorithms for solving this problem, see, e.g., Alizadeh and Goldfarb [2003], Andersen et al. [2003], Ben-Tal and Nemirovski [2001a], Kuo and Mittelmann [2004], Lobo et al. [1998], Sturm [2002], Toh et al. [1999]. State of the art solvers, such as CPLEX [2011] and MOSEK [2011], can solve SOCO problems with thousands of variables fast and accurately.

The BB algorithm uses a "divide and conquer" approach, where the feasible region is divided into smaller sets, which define optimization problems over which we then recursively optimize. This algorithm has been broadly studied in the literature for the solution of mixed integer optimization problems, see, e.g., Belotti et al. [2009], Lawler and Wood [1966], Mahajan [2009], Nemhauser and Wolsey [1999], Stubbs and Mehrotra [1999], Vielma et al. [2008], Vielma [2009], Schrijver [1986]. In the case of a MISOCO problem, these sub-problems obtained after the division are MISOCO problems as well. For each of these sub-problems we solve its SOCO relaxation to iteratively improve the upper and lower bounds of the optimal value of the original MISOCO problem until these bounds are equal, at which point the algorithm stops. This allows us to exploit the capability to efficiently solve the SOCO relaxations. In particular, CPLEX [2011] and MOSEK [2011] use BB-based algorithms for the solution of MISOCO problems.

We start describing how the BB algorithm can be used for solving a MISOCO problem.

Let Π^0 denote a given MISOCO problem, $\mathcal{F}_r^0 = \{x \in \mathbb{R}^n \mid Ax = b, x \in \mathcal{K}\}$ be the feasible region of its SOCO relaxation, $\mathcal{F}^0 = \mathcal{F}_r^0 \cap (\mathbb{Z}^d \times \mathbb{R}^{n-d})$ be its feasible set, and ζ^* be the optimal value of Π^0. Throughout the algorithm, we maintain and update four elements:

- the best solution x^* found for Π^0, which is also known as the incumbent solution;
- the tightest upper bound known for ζ^*, denoted by $\overline{\zeta}$ and it is initialized to ∞;
- the tightest lower bound known for ζ^*, denoted by $\underline{\zeta}$ and it is initialized to $-\infty$;
- a set $\mathcal{M}$ of active MISOCO problems, which is initialized to $\mathcal{M} = \{\Pi^0\}$.

The BB algorithm starts by solving the SOCO relaxation of Π^0. Let x^r be the solution of the SOCO relaxation of Π^0 and ζ^r the optimal value of this relaxation. Then, ζ^r provides a lower bound for ζ^* and we can do the update $\underline{\zeta} = \zeta^r$. Now, any feasible solution to Π^0, i.e., a solution where the integrality constraints are satisfied, provides an upper bound for ζ^*. Hence, if $x^r \in \mathcal{F}^0$, then $\overline{\zeta} = \underline{\zeta} = \zeta^*$, and the algorithm terminates. Otherwise, let $\mathcal{P}^i$, $i = 1, \ldots, t$, denote t polyhedra such that $\cup_{i=1}^t \mathcal{P}^i$ is a valid linear disjunction for $\mathcal{F}^0$. Let Π^i, $i = 1, \ldots, t$, denote the problem of minimizing $c^\top x$ for $x \in \mathcal{F}^i = \mathcal{P}^i \cap \mathcal{F}^0$. Also, let ζ^i be the optimal value of the SOCO relaxation of Π^i, and x^i be the solution of this relaxation. We have that $\min_{i \in \{1,\ldots,t\}} \zeta^i$ is a lower bound for ζ^*, and also $\zeta^r \leq \min_{i \in \{1,\ldots,t\}} \zeta^i \leq \zeta^*$. Note that the first inequality may be strict if $x^r \notin \mathcal{F}^0 \cap (\cup_{i=1}^t \mathcal{P}^i)$. Hence, using the partition $\mathcal{F}^i$, $i = 1, \ldots, t$ we may be able to obtain a tighter lower bound for ζ^* with the update $\underline{\zeta} = \min_{i \in \mathcal{J}} \zeta^i$. If we apply this procedure recursively to each problem Π^i, $i = 1, \ldots, t$, we obtain a BB algorithm to solve MISOCO problems.

During the execution of the BB algorithm it is possible that some of the solutions x^i may be feasible to Π^0. However, none of these solutions may be certified as optimal until the gap between $\overline{\zeta}$ and $\underline{\zeta}$ is closed. Nevertheless, we use these intermediate solutions to improve the upper bound $\overline{\zeta}$ during the execution of the algorithm. Hence, if there is a

$x^i \in \mathcal{F}^i$, then x^i is a feasible solution for Π^0. Let $\mathcal{J}$ be the subset of indices $i = 1, \ldots, t$ such that $x^i \in \mathcal{F}^i$. Hence, we have that $\zeta^* \leq \min_{i \in \mathcal{J}} \zeta^i$. If $\min_{i \in \mathcal{J}} \zeta^i \leq \overline{\zeta}$, then the update $\overline{\zeta} = \min_{i \in \mathcal{J}} \zeta^i$ provides a tighter upper bound for ζ^*. Additionally, the incumbent solution is updated x^* with the solution x^j, where $j \in \mathcal{J}$ is the index of the problem Π^j corresponding to the $\min_{i \in \mathcal{J}} \zeta^i$. At termination, the algorithm either finds an optimal solution, which is returned in x^*, or declares the problem infeasible..

Most implementations of the BB algorithm use a partition in each recursion with $t = 2$. With this choice, a disjunction of the form $a^\top x \geq \alpha \bigvee h^\top x \leq \beta$ arise as the natural choice for creating the partition $\mathcal{F}^i = \mathcal{P}^i \cap \mathcal{F}^0$, which defines the branching step. In particular, a commons choice for branching is a disjunction of the form $x_i \leq \lfloor x_j \rfloor \bigvee x_i \geq \lceil x_j \rceil$, which is known as branching on variables. Algorithm 1 presents the steps of a BB algorithm when this is the branching choice, which is the base algorithm we use in this thesis to solve MISOCO problems. In this description Π^a denotes the problem over which we are currently executing the recursion, its feasible set is denoted by $\mathcal{F}^a$. The problems resulting from the partition defined by the valid disjunction are stored in the set of active problems denoted by $\mathcal{M}$. This set stores the problems for which are pending for applying the recursion that defines the BB algorithm.

In Algorithm 1 we have a number of algorithmic choices. On one hand we have the branching strategy. For this choice we need to decide what is the disjunctive set that would be used to do the partition process. The performance of the BB algorithm for different branching strategies in the solution of convex problems has been analyzed Achterberg et al. [2005], Bonami et al. [2011], Gupta and Ravindra [1985]. Another algorithmic choice in the BB algorithm is the search strategy, which defines how to choose the next problem to process with the recursive step from the set $\mathcal{M}$. Note that the Algorithm 1 does not specify the branching or the search strategy. The definition of these criteria and its effect in the algorithm performance for MISOCO problems is explored in some preliminary experiments

Algorithm 1 Branch-and-Bound algorithm with binary partitions

Data: $\mathcal{M} = \{\Pi^0\}$, $\overline{\zeta} = \infty$, $\underline{\zeta} = -\infty$, and the index set $\mathcal{I}$ of integer variables in Π^0
  while $\mathcal{M} \neq \emptyset$ **do**
    Select an active problem Π^a from $\mathcal{M}$, which has a feasible set $\mathcal{F}^a$
    $\mathcal{M} \leftarrow \mathcal{M} \setminus \Pi^a$
    Solve the SOCO relaxation of Π^a
    if If the SOCO relaxation of Π^a is feasible **then** ▷ (prune by infeasibility)
      $\zeta^r \leftarrow$ optimal value of SOCO relaxation of Π^a
      $x^r \leftarrow$ optimal solution of SOCO relaxation of Π^a
      if $\zeta^r \leq \overline{\zeta}$ **then** ▷ (prune by value dominance)
        if $x^r \in \mathcal{F}^a$ **then** ▷ (prune by integrality)
          $\overline{\zeta} \leftarrow \zeta^r$ ▷ (update upper bound)
          $x^* \leftarrow x^r$
          Delete all $\Pi^i \in \mathcal{M}$ such that $\zeta^i \geq \overline{\zeta}$
        else ▷ (branch)
          Select a branching variable $x_j \notin \mathbb{Z}$, $j \in \mathcal{I}$
          $\mathcal{M} \leftarrow \{\mathcal{M}, \min\{c^\top x \mid x \in \mathcal{F}^a, x_j \geq \lceil x_j \rceil\}, \min\{c^\top x \mid x \in \mathcal{F}^a, x_j \leq \lfloor x_j \rfloor\}\}$
          $\underline{\zeta} \leftarrow \min\{\zeta^i \mid \Pi^i \in \mathcal{M}\}$. ▷ (update lower bound)
        end if
      end if
    end if
    if $\overline{\zeta} - \underline{\zeta} = 0$ **then**
      $\mathcal{M} \leftarrow \emptyset$
    end if
  end while
  if $\overline{\zeta} < \infty$ **then**
    $\zeta^* \leftarrow \overline{\zeta}$
  else
    No feasible solution was found
  end if

Chapter 5. Finally, we consider the pruning strategy, which is based in detecting feasibility of the problem and the dominance of the upper bound $\overline{\zeta}$. This upper bound is as important as the branching strategy, and they are strongly related. In Algorithm 1 this bound is updated any time a new integer feasible solution is found.

1.1.5 Cutting-plane algorithms and disjunctive cuts

In this section we review the cutting-plane algorithm and the concept of disjunctive cuts for MILO problems. Recall that a MILO is a problem given as

$$\begin{aligned} \text{minimize: } & c^\top x \\ \text{subject to: } & Ax = b \\ & x \in \mathbb{Z}^d \times \mathbb{R}^{n-d}, \end{aligned} \tag{MILO}$$

where $A \in \mathbb{R}^{m\times n}$, with $\text{rank}(A) = m$; $c \in \mathbb{R}^n$; $b \in \mathbb{R}^m$; A, b have rational entries. Let $\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax = b\}$ and let $\mathcal{F} = \mathcal{P} \cap \mathbb{Z}^d \times \mathbb{R}^{n-d}$ be the feasible set of (MILO).

For describing the cutting-plane algorithm we need first to introduce the concept of valid inequalities and cuts, see e.g. Cornuéjols [2008]. We say that an inequality is valid for a set if it is satisfied by every point in the set. Now, a valid inequality is a cut with respect to a point $x \notin \text{conv}(\mathcal{F})$ if it is a valid inequality for $\text{conv}(\mathcal{F})$ that is violated by x.

The cutting-plane algorithm for a given MILO starts solving its continuous relaxation, which is also known as its Linear Optimization (LO) relaxation. If the LO relaxation is infeasible or unbounded, we stop and declare the MILO infeasible or unbounded. Otherwise, the solution of the LO relaxation provides a lower bound for the optimal value of the MILO problem. If the optimal solution of the LO relaxation, denoted by x^r, satisfies the integrality constraints in the MILO problem, then it is optimal for the MILO, and the algorithm stops. Otherwise, one can add a cut with respect to x^r to the LO relaxation

of the MILO problem, which is usually a cutting plane. This improved formulation is resolved, and the process is repeated until the solution of the improved LO problem is feasible to the MILO problem, or the LO problem becomes infeasible, in which case the MILO problem is declared infeasible.

Disjunctive programming Balas [1979] has been one of the most successful techniques used for generating cuts for (MILO). Let the disjunction $a^\top x_{1:d} \geq \alpha + 1 \bigvee a^\top x_{1:d} \leq \alpha$, $a \in \mathbb{Z}^d$ and $\alpha \in \mathbb{Z}$, be a valid disjunction for $\mathcal{F}$. Additionally, let $\mathcal{A} = \{x \in \mathbb{R}^n \mid a^\top x_{1:d} \geq \alpha + 1\}$ and $\mathcal{B} = \{x \in \mathbb{R}^n \mid a^\top x_{1:d} \leq \alpha\}$. Then, we have that $\mathcal{F} \subseteq \mathcal{P} \cap (\mathcal{A} \cup \mathcal{B})$. Thus, a inequality $h^\top x \leq \eta$ that is valid for $\mathcal{P} \cap (\mathcal{A} \cup \mathcal{B})$ is also valid for $\mathcal{F}$. An inequality $h^\top x \leq \eta$ is called a disjunctive inequality if there exist a $a \in \mathbb{Z}^d$ and a $\alpha \in \mathbb{Z}$ such that $h^\top x \leq \eta$ is valid for $\mathcal{P} \cap (\mathcal{A} \cup \mathcal{B})$, see, e.g., Cornuéjols [2008]. Additionally, if the inequality $h^\top x \leq \eta$ is a cut with respect to a point $x \notin \text{conv}(\mathcal{F})$, it is called a disjunctive cut.

In this thesis we define disjunctive cuts for the intersection of a closed convex set $\mathcal{S} \in \mathbb{R}^n$ and a disjunctive set as follows. Consider a disjunctive set $\mathcal{U} \cup \mathcal{V}$ where $\mathcal{U} = \{x \in \mathbb{R}^n \mid u^\top x \geq \varphi\}$ and $\mathcal{V} = \{x \in \mathbb{R}^n \mid v^\top x \leq \varpi\}$, $u^\top, v^\top \in \mathbb{R}^n$. We assume that $\mathcal{S} \cap (\mathcal{U} \cup \mathcal{V}) = \emptyset$. An inequality is a disjunctive inequality associated with $\mathcal{S}, \mathcal{U}$ and $\mathcal{V}$ if it is a valid inequality for $\mathcal{S} \cap (\mathcal{U} \cup \mathcal{V})$. Additionally, if the inequality is a cut with respect to a point $x \notin \text{conv}(\mathcal{S} \cap (\mathcal{U} \cup \mathcal{V}))$, we called a disjunctive cut. In Chapter 2 we analyze a special case of disjunctive cuts that is defined by a conic inequality.

One way to improve the bounding process in Algorithm 1 is to incorporate the use of cuts to strengthen the relaxation of Π^a solved in each iteration. In Algorithm 1, this can be done after solving the relaxation Π^a. The new algorithm resulting with the addition of this extra step is called a "*branch-and-cut algorithm*", which in MILO is a mixed between BB and cutting-planes. For MILO problems, the incorporation of linear cuts was essential in the development of efficient branch-and-cut algorithms Balas [1979], Cornuéjols [2008], Martin [2001], Nemhauser and Wolsey [1999], Schrijver [1986]. This technique has been

generalized to mixed integer nonlinear optimization problems as well Grossmann [2002]. In this thesis we research the development and usage of disjunctive cuts for MISOCO problems in a branch-and-cut algorithm. In Chapter 5 we present the details of the branch-and-cut algorithm used in this work and its performance is some preliminary experiments.

1.2 Dissertation overview

In this thesis, we study the derivation of Disjunctive Conic Cuts (DCCs) for MISOCO problems. Our main goal is to extend the ideas of disjunctive programming that have shown to be successful in the derivation of linear cuts for MILO. In this work, we describe how this ideas can be applied to MISOCO problems for generating conic cuts. Additionally, we present some preliminary numerical results that show that a certain class of cuts, which used in a branch-and-cut framework, can effectively help to accelerate the solution process.

In Chapter 2 we introduce the definition of disjunctive conic cuts and analyze some of their properties. In particular, we consider the intersection of a certain full dimensional closed convex set $\mathcal{E} \in \mathbb{R}^n$ and a disjunctive set in $\mathbb{R}^n$ of the form (1.6). We define a DCC as a cone such that its intersection with the set $\mathcal{E}$ is equal to the convex hull of the intersection of $\mathcal{E}$ with the disjunctive set. We then present the conditions under which a cone can be identified as a DCC, and we are able to prove that a cone satisfying these conditions is unique. We also show that the assumptions made in this chapter are required for uniqueness of that cone. In the second part of the chapter we define the Disjunctive Cylindrical Cones (DCyC) as cylinders that characterize the convex hull of the intersection of $\mathcal{E}$ and disjunctive set in $\mathbb{R}^n$ of the form (1.6). In this case we also provide the conditions to identify when a cylinder is a DCyC, and we are also able to prove that a cylinder satisfying these conditions is unique.

In Chapter 3 we present the analysis of the intersection of a quadric with two hyperplanes. The main result of this chapter is the characterization of the family of quadrics

having the same intersection with two given hyperplanes. This family is analyzed for two different cases: when the hyperplanes are parallel and when the hyperplanes are nonparallel. In the first case, we present a full characterization of the family when there is a quadric in it that is defined by a matrix with at most one non-positive eigenvalue. We prove that in this case, there is always a cylinder or a cone in the family. In the second case, we present a full characterization of the behavior of the family when there is a quadric in it that is defined by a positive definite matrix. We also prove that in this case there is always a cylinder or a cone in that family.

In Chapter 4, we present a procedure for the derivation of DCCs and DCyCs separating a given point from the feasible set of a MISOCO problem. This procedure that can be embedded in a branch-and-cut framework. First, we characterize the quadric associated with the feasible set of a MISOCO problem with a single cone. Second, we provide the theoretical basis for a procedure to derive either DCCs or DCyCs when the disjunctive set is defined by two parallel hyperplanes. Third, we provide the results for the derivation of DCCs and DCyCs when the disjunctive set is define by two nonparallel hyperplanes. In this case, the results are limited to cases for which the intersection of the hyperplanes with the feasible set of the relaxed MISOCO problem are bounded. We close the chapter with a comparison of the DCCs with nonlinear conic mixed-integer rounding inequalities.

In Chapter 5, we briefly describe our implementation of the procedure of Chapter 4. First, we provide a short description of the well known branch-and-cut algorithm. We also discuss briefly issues surrounding branching rules, node selection, and the selection of the seed to create the DCCs. Second, we describe how the procedure described in Chapter 4 can be adapted to cases when the MISOCO problem has more than one cone. We finish the chapter with a short description of our computational framework and some consideration about the implementation of DCCs. In Chapter 6, we describe the test sets used for the preliminary experimentation with the DCCs analyzed in this thesis. For each test set,

we give their main characteristics and present the results of the experiments performed. Additionally, we comment about the insights that these preliminary results provide about the effectiveness of the DCCs.

This thesis aims to provide a full analysis of the derivation of DCCs for MISOCO problems. For the reader interested in a full understanding of our results we suggest to read this document in the order it is provided. However, Chapters 2, and 3 are self-contained and they only depend on the well-known results presented in this introduction. For that reason, the reader may switch the order of reviewing these two chapters without affecting the understanding of their results. These two chapters are provided to support the main results of Chapter 4. For the reader interested specifically in the derivation of the DCCs, it is possible to skip Chapter 2, but it is recommended to first read Sections 3.2 and 3.3 of Chapter 3, which provide results essential to understanding the derivation of the DCCs and DCyCs described in Chapter 4. For the reader interested in knowing about the performance of the DCCs in our preliminary experiments we suggest reading Chapters 5 and 6 after becoming familiar with DCCs, since in these two chapters, details of the derivations are omitted. Finally, the results presented in Appendix A are provided to support the proofs of the results in Chapter 4. They are not essential to understand the derivation procedure. However, the interested reader is welcome to read them in order to acquire a full understanding of the proofs.

Chapter 2

Disjunctive conic cuts

In this chapter we present the definition of disjunctive conic cuts and some of their properties. Let $\mathcal{E} \subset \mathbb{R}^n$, $n > 1$ be a full dimensional closed convex set. Additionally, consider two half-spaces $\mathcal{A} = \{x \in \mathbb{R}^n | a^\top x \geq \alpha\}$ and $\mathcal{B} = \{x \in \mathbb{R}^n | b^\top x \leq \beta\}$, where $a, b \in \mathbb{R}^n$, and $(a^\top, \alpha)$, $(b^\top, \beta)$ are not scalar multiple of each other, i.e., $\nexists \eta \in \mathbb{R}$ such that $(a^\top, \alpha) = \eta(b^\top, \beta)$. Throughout this chapter we refer to the sets $\mathcal{A}^= = \{x \in \mathbb{R}^n | a^\top x = \alpha\}$ and $\mathcal{B}^= = \{x \in \mathbb{R}^n | b^\top x = \beta\}$, the boundary hyperplanes defining the half-spaces $\mathcal{A}$ and $\mathcal{B}$, respectively. Additionally, we assume the following about the set $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$:

Assumption 2.1. *The intersection $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E}$ is empty.*

We show that the set $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$ can be fully characterized using disjunctive conic cuts. The results presented here are based on the results from Belotti et al. [2012]. The characterization of the set $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$ is divided into two different cases. Section 2.1 presents the conditions to identify when $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$ is obtained by intersecting the set $\mathcal{E}$ with a cone, while Section 2.2 presents conditions in which the convex hull is obtained by intersecting $\mathcal{E}$ with a cylinder.

2.1 Disjunctive conic cuts

For discussing the case when the set $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$ can be obtained intersecting $\mathcal{E}$ with a cone $\mathcal{K}$ we need an additional assumption about the set $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$ to facilitate the proofs. We show later that this condition is required for uniqueness of the cone.

Assumption 2.2. *The intersections $\mathcal{E} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=}$ are nonempty and bounded.*

Recall Definition 1.12 of a convex cone. We introduce now the definition of a translated cone, which is a generalization of a standard cone.

Definition 2.1. *A set $\mathcal{K} \in \mathbb{R}^n$, is called a translated cone if there exists a vector $x^* \in \mathcal{K}$, called the* vertex *of $\mathcal{K}$, such that for every $\theta \geq 0$ and $x \in \mathcal{K}$, we have $x^* + \theta(x - x^*) \in \mathcal{K}$.*

Associated with any translated cone $\mathcal{K} \in \mathbb{R}^n$, is a set $\mathcal{K}^0 = \{y \in \mathbb{R}^n \mid \ y = x - x^t, x \in \mathcal{K}^t\}$, which is a cone in the sense of Definition 1.12. Although translated cones arise naturally in this setting, we assume w.l.o.g. that all cones have a vertex at the origin unless otherwise specified.

Definition 2.2. *A closed convex cone $\mathcal{K} \in \mathbb{R}^n$ with $\dim(\mathcal{K}) > 1$ is called a* Disjunctive Conic Cut *(DCC) for $\mathcal{E}$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$ if*

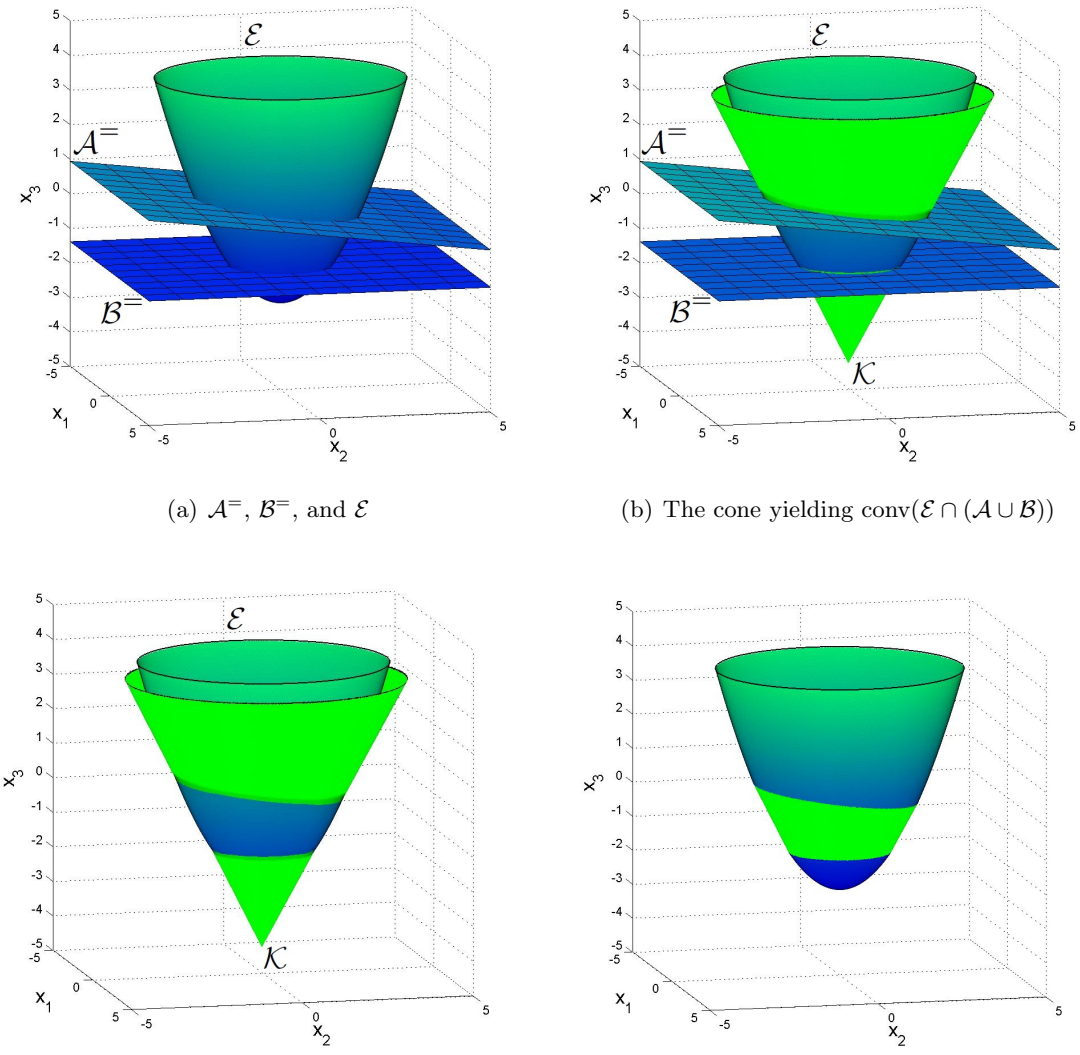
$$\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})) = \mathcal{E} \cap \mathcal{K}.$$

The following proposition gives a necessary and sufficient condition for a convex cone to be a DCC for the set $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$.

Proposition 2.1. *A closed convex cone $\mathcal{K} \in \mathbb{R}^n$ with $\dim(\mathcal{K}) > 1$ is a DCC for $\mathcal{E}$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$, if and only if,*

$$\mathcal{K} \cap \mathcal{A}^{=} = \mathcal{E} \cap \mathcal{A}^{=} \quad \textit{and} \quad \mathcal{K} \cap \mathcal{B}^{=} = \mathcal{E} \cap \mathcal{B}^{=}.$$

Figure 2.1 illustrates Proposition 2.1 where the set $\mathcal{E}$ is the epigraph of a paraboloid in $\mathbb{R}^3$. Before proving Proposition 2.1, see page 33 for the proof, we first provide a set of



(a) $\mathcal{A}^=$, $\mathcal{B}^=$, and $\mathcal{E}$

(b) The cone yielding $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$

(c) $\mathcal{E} \cap \mathcal{K}$

(d) $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$

Figure 2.1: Illustration of a disjunctive conic cut as specified in Proposition 2.1

lemmas that will make the proof more compact. To begin, let us recall the definition of a *base of a cone*, see Barvinok [Barvinok, 2002, page 66].

Definition 2.3 (Base of a cone Barvinok [2002]). *Let $\mathcal{K} \in \mathbb{R}^n$ be a convex cone. A set $\mathcal{V} \subset \mathcal{K}$ is called a* base *of $\mathcal{K}$ if $0 \notin \mathcal{V}$ and for every vector $u \in \mathcal{K}$, $u \neq 0$, there is a unique $v \in \mathcal{V}$ and $\lambda > 0$ such that $u = \lambda v$.*

Lemma 2.1 shows a key relationship between a DCC and the hyperplanes $\mathcal{A}^=$ and $\mathcal{B}^=$.

Lemma 2.1. *Consider a half space $\mathcal{G} = \{x \in \mathbb{R}^n \mid g^\top x \leq \varrho\}$. Assume that $\mathcal{E} \cap \mathcal{G}^=$ is nonempty, bounded, and does not contain the origin 0. If there exist a closed convex cone $\mathcal{K} \in \mathbb{R}^n$, with $\dim(\mathcal{K}) > 1$ and $\mathcal{K} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$, then $\mathcal{E} \cap \mathcal{G}^=$ is a base of $\mathcal{K}$.*

Proof. From the assumptions in the lemma, we have that $0 \notin \mathcal{K} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$. We may assume w.l.o.g. that $0 \in \mathcal{G}$. First, since $\mathcal{K} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$ is bounded we know that there exists no ray of $\mathcal{K}$ parallel to $\mathcal{G}^=$. Now, let us assume to the contrary that $\mathcal{E} \cap \mathcal{G}^=$ is not a base of $\mathcal{K}$. From Definition 2.3 we know that there must exists a vector $u \in \mathcal{K}$ such that the ray $\mathcal{R}_u = \{\lambda u \mid \lambda \geq 0\}$ does not intersect with $\mathcal{K} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$, i.e., there is a ray in $\mathcal{K}$ parallel to the hyperplane $\mathcal{G}^=$. This implies that the set $\mathcal{K} \cap \mathcal{G}^=$ is unbounded, which contradicts the boundedness Assumption 2.2. Therefore, $\mathcal{E} \cap \mathcal{G}^=$ is a base for $\mathcal{K}$. □

From the result of Lemma 2.1 we obtain that if there exists a DCC for $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$, then both $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are bases of that disjunctive cone. Now, we show that a convex cone that satisfy Lemma 2.1 is a pointed cone, which is an important result for our further development.

Lemma 2.2. *Any closed convex cone $\mathcal{K}$ satisfying Lemma 2.1 must be pointed.*

Proof. Recall Definition 1.15 of a pointed cone. Now, assume to the contrary that $\mathcal{K}$ is not pointed. This means that $\mathcal{K}$ contains a line. Hence, there exist two vectors $\hat{r}, \bar{r} \in \mathcal{K}$ such that $\hat{r} = -\bar{r}$. Additionally, from the convexity of $\mathcal{K}$ we have that $\mu\hat{r} + \nu\bar{r} \in \mathcal{K}$, for any $\mu, \nu > 0$. Now, since $\mathcal{E} \cap \mathcal{A}^=$ is a base of $\mathcal{K}$, there exists a vector $\hat{x} \in \mathcal{E} \cap \mathcal{A}^=$ in the ray $\mathcal{R}_{\hat{r}} = \{\mu\hat{r} \mid \mu \geq 0\}$ such that $\hat{x} = \mu\hat{r}$, for some $\mu > 0$. Similarly, there exists a vector

$\bar{x} \in \mathcal{E} \cap \mathcal{A}^{=}$ in the ray $\mathcal{R}_{\bar{r}} = \{\nu\bar{r} \mid \nu \geq 0\}$ such that $\bar{x} = \nu\bar{r}$, for some $\nu > 0$. Given that $\mathcal{A}^{=}$ is an affine set, we have

$$\gamma\hat{x} + (1-\gamma)\bar{x} \in \mathcal{A}^{=}, \quad \forall\gamma \in \mathbb{R}.$$

Expressing $\hat{x}$ and $\bar{x}$ in term of $\hat{r}$ and $\bar{r}$ gives

$$\begin{aligned}\gamma\hat{x} + (1-\gamma)\bar{x} &= \gamma(\mu\hat{r}) + (1-\gamma)(\nu\bar{r}) \\ &= -\gamma(\mu\bar{r}) + (1-\gamma)(\nu\bar{r}) \\ &= \nu\bar{r} - \gamma(\mu+\nu)\bar{r}.\end{aligned}$$

Hence, if $\gamma = 0$ then $\nu\bar{r} \in \mathcal{K}$. On the other hand, if $\gamma < 0$ we obtain that $\nu\bar{r} - \gamma(\mu+\nu)\bar{r} \in \mathcal{K}$, since it is a vector on the ray defined by $\bar{r}$. Finally, if $\gamma > 0$ then $\nu\bar{r} - \gamma(\mu+\nu)\bar{r} = \nu\bar{r} + \gamma(\mu+\nu)\hat{r} \in \mathcal{K}$, since it is a positive combination of two vectors in the cone $\mathcal{K}$. Hence, $\mathcal{K} \cap \mathcal{A}^{=}$ contains a whole line, which contradicts the assumption that $\mathcal{K} \cap \mathcal{A}^{=}$ is bounded. □

We can now prove that the vertex of a convex cone that satisfies Lemma 2.1 belongs exclusively to one of the half spaces $\mathcal{A}$ or $\mathcal{B}$. Observe that this does not imply that the set $\mathcal{A} \cap \mathcal{B}$ is empty. It only means that the vertex of the convex cone is not contained in $\mathcal{A} \cap \mathcal{B}$ if it is nonempty.

Lemma 2.3. *Let $\mathcal{K} \in \mathbb{R}^n$ be a closed convex cone, with $\dim(\mathcal{K}) > 1$, such that $\mathcal{E} \cap \mathcal{A}^{=} = \mathcal{K} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=} = \mathcal{K} \cap \mathcal{B}^{=}$. Then, the origin $x^0 = 0$ is either in $\mathcal{A}$, or in $\mathcal{B}$, but not in $\mathcal{A} \cap \mathcal{B}$.*

Proof. By Lemma 2.1 we have that $\mathcal{E} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=}$ are bases of the cone $\mathcal{K}$. Additionally, by Lemma 2.2 we know that the cone $\mathcal{K}$ is pointed. Consider the ray $\mathcal{R}_r$, where $r \in \mathcal{K}$ is such that $\|r\| = 1$. Then, there are two vectors $\hat{r} \in \mathcal{E} \cap \mathcal{A}^{=}$ and $\bar{r} \in \mathcal{E} \cap \mathcal{B}^{=}$ such that

$\hat{r} = \mu r$ and $\bar{r} = \nu r$ for some $\mu, \nu > 0$.

First we prove that $x^0 \in \mathcal{A} \cup \mathcal{B}$. Let us assume to the contrary that $x^0 \notin \mathcal{A} \cup \mathcal{B}$. Let $u \in \mathcal{R}_r$, then for any $\gamma < \min\{\nu, \mu\}$ we have that $u \in \overline{\mathcal{A}} \cap \overline{\mathcal{B}}$, and we may assume w.l.o.g. that $\nu < \mu$. Note that $\nu = \mu$ cannot happen as, by Assumption 2.1, $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E} = \emptyset$. Additionally, for any $\gamma \geq \nu$ we have that $u \in \mathcal{B}$, so the vector $\hat{r}$ is contained in the half space $\mathcal{B}$, and $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E} \neq \emptyset$, which contradicts Assumption 2.1.

Second we prove that $x^0 \notin \mathcal{A} \cap \mathcal{B}$. Let us assume now that $x^0 \in \mathcal{A} \cap \mathcal{B}$, and let $u = \gamma\bar{r} + (1-\gamma)\hat{r}$ for some $0 \leq \gamma \leq 1$. Then, we have that $u \in \mathcal{A}$ or $u \in \mathcal{B}$. When $\nu < \mu$ for $\gamma = 0$ we have $u \in \mathcal{A} \cap \mathcal{B}^{=} \cap \mathcal{E}$. Similarly, when $\mu < \nu$ for $\gamma = 1$ we have $u \in \mathcal{A}^{=} \cap \mathcal{B} \cap \mathcal{E}$. Hence, $x^0 \in \mathcal{A} \cap \mathcal{B}$ implies $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E} \neq \emptyset$, which contradicts Assumption 2.1. The proof is complete. □

We are able now to show that $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}) \subset \mathcal{K}$. This will facilitate the proof of the relation $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})) \subseteq \mathcal{E} \cap \mathcal{K}$.

Lemma 2.4. *Assume that there exist a closed convex cone $\mathcal{K}$ with $\dim(\mathcal{K}) > 1$, such that $\mathcal{E} \cap \mathcal{A}^{=} = \mathcal{K} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=} = \mathcal{K} \cap \mathcal{B}^{=}$. Then*

$$(\mathcal{E} \cap \mathcal{A}) \subset \mathcal{K} \quad \textit{and} \quad (\mathcal{E} \cap \mathcal{B}) \subset \mathcal{K}.$$

Proof. We prove that $(\mathcal{E} \cap \mathcal{A}) \subseteq \mathcal{K}$. Let us assume to the contrary that there exists a vector u such that $u \in (\mathcal{E} \cap \mathcal{A})$ but $u \notin \mathcal{K}$. First, by Theorem 1.5, there exists a hyperplane $\mathcal{H}$ separating u and $\mathcal{K}$ that contains a ray of $\mathcal{K}$ and does not contain u. Here, the assumption of $\dim(\mathcal{K}) > 1$ is needed, since the hyperplane $\mathcal{H}$ does not exist when $n = 1$.

Additionally, given the assumptions of the lemma, Assumption 2.2, and using Lemma 2.1, we have that the sets $\mathcal{E} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=}$ are bases for the cone $\mathcal{K}$. Hence, there exists a vector $v \in \mathcal{E} \cap \mathcal{B}^{=}$ such that $\mathcal{R}_v = \{\gamma v \mid \gamma \geq 0\} \subseteq \mathcal{H}$.

Given that the set $\mathcal{E}$ is convex, $\lambda u + (1-\lambda)v \in \mathcal{E}$ for all $0 \leq \lambda \leq 1$. On the other hand,

since v is a vector on a exposed face of $\mathcal{K}$, we have that $\lambda u + (1-\lambda)v \notin \mathcal{K}$ for $0 < \lambda \leq 1$. Furthermore, since $u \in (\mathcal{E} \cap \mathcal{A})$ and $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E} = \emptyset$, we have $a^\top u \leq \alpha$ and $a^\top v > \alpha$. Hence, from the equation $a^\top(\lambda u + (1-\lambda)v) = \lambda a^\top u + (1-\lambda)a^\top v$, we obtain that there exists a $\lambda \in (0, 1]$ such that $a^\top(\lambda u + (1-\lambda)v) = \alpha$. Therefore, there is a vector $w = \lambda u + (1-\lambda)v$ for some $\lambda \in (0, 1]$, such that $w \in \mathcal{E} \cap \mathcal{A}^=$, but $w \notin \mathcal{K}$, which contradicts the assumptions of the lemma. Hence, $(\mathcal{E} \cap \mathcal{A}) \subseteq \mathcal{K}$. One can prove $(\mathcal{E} \cap \mathcal{B}) \subseteq \mathcal{V}$ analogously.

Finally, recall that the sets $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are disjoint and nonempty. Then, given the assumptions of the lemma we have that $\mathcal{E} \cap \mathcal{A} \neq \mathcal{K}$ and $\mathcal{E} \cap \mathcal{B} \neq \mathcal{K}$, this completes the proof. □

Now we present the proof of Proposition 2.1.

Proof of Proposition 2.1. First, we prove that if $\mathcal{E} \cap \mathcal{A}^= = \mathcal{K} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^= = \mathcal{K} \cap \mathcal{B}^=$ then $\mathcal{K}$ is a disjunctive cone. Consider a vector $u \in (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. Then, from Lemma 2.4 we have that $u \in \mathcal{E} \cap \mathcal{K}$. Now, consider two given vectors $u, v \in (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. Then, since both $\mathcal{K}$ and $\mathcal{E}$ are convex, for any $0 \leq \lambda \leq 1$ we have $\lambda u + (1-\lambda)v \in \mathcal{E} \cap \mathcal{K}$. Hence, $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})) \subseteq \mathcal{E} \cap \mathcal{K}$.

We need to prove now that $\mathcal{E} \cap \mathcal{K} \subseteq \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$. Consider a vector $u \in \mathcal{E} \cap \mathcal{K}$. First, if $u \in \mathcal{E} \cap \mathcal{A}$ or $u \in \mathcal{E} \cap \mathcal{B}$, we have that $u \in \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$. Assume then that $u \notin (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$, which implies $u \in (\overline{\mathcal{A}} \cap \overline{\mathcal{B}} \cap \mathcal{K})$. Furthermore, by Lemma 2.1, there are two vectors $v \in \mathcal{E} \cap \mathcal{A}^=$ and $\bar{x} \in \mathcal{E} \cap \mathcal{B}^=$ such that, for some $\mu, \nu > 0$, $v = \mu u$ and $w = \nu x$. From Lemma 2.3, the vertex of the cone is either in $\mathcal{A}$ or $\mathcal{B}$ but not in both. Assume w.l.o.g. that the vertex of the cone is in $\mathcal{B}$. Then, $\nu < 1 < \mu$ and there exists a

$\gamma \in (0,1)$ such that $\gamma\nu + (1-\gamma)\mu = 1$. Hence, we can write

$$\begin{aligned} \gamma w + (1-\gamma)v &= \gamma\nu u + (1-\gamma)\mu u \\ &= (\gamma\nu + (1-\gamma)\mu)u \\ &= u. \end{aligned}$$

Therefore, u can be expressed as a linear convex combination of two vectors in $(\mathcal{E} \cap \mathcal{A}^{=}) \cup (\mathcal{E} \cap \mathcal{B}^{=})$. Hence, any vector $u \in \mathcal{E} \cap \mathcal{K}$ can be written as a linear convex combination of two vrctors in $(\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. Thus, $(\mathcal{E} \cap \mathcal{K}) \subseteq \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$. Finally, since the subset relation is valid in both directions, this proofs that $(\mathcal{E} \cap \mathcal{K}) = \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.

Second, we prove that if $\mathcal{K}$ is a DCC for $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$, then $\mathcal{E} \cap \mathcal{A}^{=} = \mathcal{K} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=} = \mathcal{K} \cap \mathcal{B}^{=}$. From Assumptions 2.1, 2.2 and Definition 2.2 we have that $\mathcal{E} \cap \mathcal{A}^{=} \subseteq \mathcal{E} \cap \mathcal{K}$, then for a given $x \in \mathcal{E} \cap \mathcal{A}^{=}$ we have that $x \in \mathcal{K} \cap \mathcal{A}^{=}$. Henceforth, $\mathcal{E} \cap \mathcal{A}^{=} \subseteq \mathcal{K} \cap \mathcal{A}^{=}$. We can show similarly that $\mathcal{E} \cap \mathcal{B}^{=} \subseteq \mathcal{K} \cap \mathcal{B}^{=}$.

Assume now that $\mathcal{E} \cap \mathcal{A}^{=}$ is a proper subset of $\mathcal{K} \cap \mathcal{A}^{=}$, i.e., $\mathcal{E} \cap \mathcal{A}^{=} \subset \mathcal{K} \cap \mathcal{A}^{=}$. Then, there is a vector $u \in \mathcal{E} \cap \mathcal{A}^{=}$ such that $u \notin \mathcal{K} \cap \mathcal{A}^{=}$, which implies that $u \notin \mathcal{K}$. Hence, $u \in \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$ but $u \notin \mathcal{E} \cap \mathcal{K}$, which violates Definition 2.2 of a DCC for $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$. Similarly, we can show that $\mathcal{E} \cap \mathcal{B}^{=}$ is not a proper subset of $\mathcal{K} \cap \mathcal{B}^{=}$. Therefore, we have that $\mathcal{E} \cap \mathcal{A}^{=} = \mathcal{K} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=} = \mathcal{K} \cap \mathcal{B}^{=}$. This completes the proof. □

We need to add an additional comment to complete this analysis. In Lemma 2.1, we assumed that none of the intersections $\mathcal{A}^{=} \cap \mathcal{E}$ and $\mathcal{B}^{=} \cap \mathcal{E}$ contain the vertex of the cone $\mathcal{K}$. However, if one of these intersections is a single point, then this single point must be the vertex of $\mathcal{K}$ and the other intersection must be a base of $\mathcal{K}$. In this case, we don't need Lemma 2.3, and the rest of the proof follows.

Lemma 2.5. *If a DCC $\mathcal{K} \in \mathbb{R}^n$ with $\dim(\mathcal{K}) > 1$ exists for the set $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$, then $\mathcal{K}$ is unique.*

Proof. Assume to the contrary that there are two different cones $\mathcal{K}_1$ and $\mathcal{K}_2$ that satisfy Definition 2.2. Thus, from Proposition 2.1 we have $\mathcal{K}_1 \cap \mathcal{A}^= = \mathcal{K}_2 \cap \mathcal{A}^=$ and $\mathcal{K}_1 \cap \mathcal{B}^= = \mathcal{K}_2 \cap \mathcal{B}^=$. Let $v^1 \in \mathcal{K}_1$ be the vertex of $\mathcal{K}_1$ and $v^2 \in \mathcal{K}_2$ be the vertex of $\mathcal{K}_2$. Now, assume w.l.o.g. that $v^1 = 0$ and that $v^2 \neq 0$, i.e., $\mathcal{K}_2$ is a translated cone.

First, we prove that if either $\mathcal{E} \cap \mathcal{A}^=$ or $\mathcal{E} \cap \mathcal{B}^=$ is a single point, then $\mathcal{K}_1 = \mathcal{K}_2$. Since $\dim(\mathcal{K}) > 1$ we have that $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ cannot be both single point sets. Let $u \in \mathcal{E}$, and assume that $\mathcal{E} \cap \mathcal{A}^= = \{u\}$, then $\mathcal{K}_1 \cap \mathcal{A}^= = \{u\}$ and $\mathcal{K}_2 \cap \mathcal{A}^= = \{u\}$. Now, if $u \neq v^1$, then we have that $\mathcal{K}_1 = \{\theta v^1 \mid \theta \geq 0\}$, which implies that the set $\mathcal{E} \cap \mathcal{B}^=$ is a single point. Thus, we have that $u = v^1$. On the other hand, if $u \neq v^2$, then we have that $\mathcal{K}_2 = \{v \in \mathbb{R}^n \mid y = v^2 + \theta(u - v^2), \theta \geq 0\}$, which also implies that the set $\mathcal{E} \cap \mathcal{B}^=$ is a single point. Hence, we have that $u = v^2$. Therefore, we have that $v^1 = v^2$. Finally, from Lemma 2.1 we know that $\mathcal{E} \cap \mathcal{B}^=$ is a base for $\mathcal{K}_2$ and $\mathcal{K}_1$. Therefore, we have that $\mathcal{K}_1 = \mathcal{K}_2$. The same argument would show that $\mathcal{K}_1 = \mathcal{K}_2$ if $\mathcal{E} \cap \mathcal{B}^= = \{z\}$.

Second, we show that if $\{v^1, v^2\} \cap (\mathcal{A}^= \cup \mathcal{B}^=) = \emptyset$, then $v^1 \in \mathcal{K}_2$ and $v^2 \in \mathcal{K}_1$. Assume to the contrary that $v^1 \notin \mathcal{K}_2$. Here use a similar argument to the one in the proof of Lemma 2.4. By the separation Theorem 1.5, there exists a hyperplane $\mathcal{H}$ separating v^1 and $\mathcal{K}_2$ properly and does not contain v^1. From Lemma 2.1, we know that the sets $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are bases for $\mathcal{K}_2$. Hence, there exists a vector $w \in \mathcal{E} \cap \mathcal{B}^=$ such that the extreme ray $\mathcal{R}_w = \{v^2 + \gamma(w - v^2) \mid \gamma \geq 0\}$ of $\mathcal{K}_2$ is in $\mathcal{H}$. Additionally, by Lemma 2.3 we have that v^1 is either in $\mathcal{A}$ or $\mathcal{B}$ but not in $\mathcal{A} \cap \mathcal{B}$. Assume w.l.o.g. that $v^1 \in \mathcal{A}$. Given that $\mathcal{K}_1$ is convex, $\lambda v^1 + (1 - \lambda)w \in \mathcal{K}_1$ for all $0 \leq \lambda \leq 1$. On the other hand, since w is a vector on an exposed face of $\mathcal{K}_2$, for $0 < \lambda \leq 1$ we have $\lambda v^1 + (1 - \lambda)w \notin \mathcal{K}_2$. Furthermore, since $v^1 \in \mathcal{A}$, by Assumption 2.1, we have $a^\top v^1 \leq \alpha$ and $a^\top w > \alpha$. Hence, from the equation $a^\top(\lambda v^1 + (1 - \lambda)w) = \lambda a^\top v^1 + (1 - \lambda)a^\top w$, we may obtain $0 < \lambda \leq 1$ such that $a^\top(\lambda x^* + (1 - \lambda)w) = \alpha$. Therefore, there exists a vector $u = \lambda v^1 + (1 - \lambda)w$ for some $0 < \lambda \leq 1$, such that $u \in \mathcal{K}_1 \cap \mathcal{A}^=$, but $u \notin \mathcal{K}_2$, which contradicts $\mathcal{K}_1 \cap \mathcal{A}^= = \mathcal{K}_2 \cap \mathcal{A}^=$.

Hence, we obtain that $v^1 \in \mathcal{K}_2$. Using a similar argument one can proof that $v^2 \in \mathcal{K}_1$.

Third, we show that if $v^1 \neq v^2$, then they cannot be both in $\mathcal{A}$ or in $\mathcal{B}$. Assume to the contrary that $v^1 \in \mathcal{A}$ and $v^2 \in \mathcal{A}$, then if $\alpha > 0$ we have $v_1 \notin \mathcal{A}$, thus $\alpha \leq 0$. On one hand, since $v^1 \in \mathcal{K}_2$ we have that $\mathcal{R}_{v^1} = \{(1-\theta)v^2 \mid \theta \geq 0\} \subseteq \mathcal{K}_2$. Hence, if $a^\top v^2 \leq 0$ then $R^{v^1} \in \mathcal{A}$ which implies that $\mathcal{A} \cap \mathcal{K}_1$ is unbounded. On the other hand, since $v^2 \in \mathcal{K}_1$ we have that $\mathcal{R}_{v^2} = \{\theta v^2 \mid \theta > 0\} \subseteq \mathcal{K}_1$. Hence, if $a^\top v^2 \geq 0$, then $\mathcal{R}_{v^2} \in \mathcal{A}$, which implies that $\mathcal{A} \cap \mathcal{K}_1$ is unbounded. Hence, if $v^1 \in \mathcal{A}$ and $v^2 \in \mathcal{A}$, then we obtain a contradicts to Assumption 2.2. Similarly, we can prove that if v^1 and v^2 cant not be simultaneously in $\mathcal{B}$.

Finally, we show that if v^1 and v^2 are in different halfspaces and $\{v^1, v^2\} \cap (\mathcal{A}^= \cup \mathcal{B}^=) = \emptyset$, then it contradicts the assumption that $\mathcal{K}_1 \cap \mathcal{A}^= = \mathcal{K}_2 \cap \mathcal{A}^=$ and $\mathcal{K}_1 \cap \mathcal{B}^= = \mathcal{K}_2 \cap \mathcal{B}^=$. Assume that $v^1 \in \mathcal{A}$ and $v^2 \in \mathcal{B}$. Recall that in this case $v^1 \in \mathcal{K}_2$ and $v^2 \in \mathcal{K}_1$, thus the set $\mathcal{R}_{v^1} = \{(1-\theta)v^2 \mid \theta \geq 0\} \subseteq \mathcal{K}_2$ and $\mathcal{R}_{v^2} = \{\theta v^2 \mid \theta > 0\} \subseteq \mathcal{K}_1$. Now, since $\dim(\mathcal{K}_1) > 1$ and $\mathcal{B}^= \cap \mathcal{K}_1$ is a base of $\mathcal{K}_1$, there is at least one extreme ray $\mathcal{R}_w = \{\gamma w \mid \gamma \geq 0\}$ of $\mathcal{K}_1$ such that $v^2 \notin \mathcal{R}_w$ and $w \in \mathcal{K}_1 \cap \mathcal{B}^=$ is a vector in the boundary of $\mathcal{K}_1$. Then, we have that $w \in \mathcal{K}_2 \cap \mathcal{B}^=$ and is a vector in the boundary of $\mathcal{K}_2$. This is true because if $w \in \text{ri}(\mathcal{K}_2)$, then since $\mathcal{K}_2 \cap \mathcal{B}^=$ is bounded and is a base of $\mathcal{K}_2$ we have that $w \in \text{ri}(\mathcal{K}_2 \cap \mathcal{B}^=)$. Thus, in that case there exist a vector $u \in \mathcal{K}_2 \cap \mathcal{B}^=$ such that $u \notin \mathcal{K}_1 \cap \mathcal{B}^=$, which contradicts $\mathcal{K}_1 \cap \mathcal{B}^= = \mathcal{K}_2 \cap \mathcal{B}^=$.

Now, since $w \in \mathcal{K}_2$ then $\{v^2 + \gamma(w - v^2) \mid \gamma \geq 0\} \in \mathcal{K}_2$. Even more, since $v^2 \in \mathcal{B}$ and $w \in \mathcal{B}^=$, there exist a $\hat{\gamma} > 1$ such that $a^\top(v^2 + \hat{\gamma}(w - v^2)) = \alpha$. However, since w is in the extreme ray $\mathcal{R}_w$ of $\mathcal{K}_1$ and $v^2 \notin \mathcal{R}_w$, then the vector $(v^2 + \hat{\gamma}(w - v^2)) \notin \mathcal{K}_1$. This contradicts the assumption $\mathcal{K}_1 \cap \mathcal{A}^= = \mathcal{K}_2 \cap \mathcal{A}^=$. The same contradiction is found if we assume that $v^1 \in \mathcal{B}$ and $v^2 \in \mathcal{A}$. Hence, since v^1 and v^2 cannot be in different halfspaces, then $v^1 = v^2$. In conclusion, we have that $\mathcal{K}_1 = \mathcal{K}_2$, since $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are bases for $\mathcal{K}_1$ and $\mathcal{K}_2$, which proof that the disjunctive conic cut is unique. □

Figure 2.2 illustrates how Lemma 2.5 fails when the intersections $\mathcal{E} \cap \mathcal{A}^{=}$ or $\mathcal{E} \cap \mathcal{B}^{=}$ are unbounded. In this case, one can see that only the cone $\mathcal{K}$ gives the convex hull of $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$. The other two cones $\mathcal{K}_1$ and $\mathcal{K}_2$ have the same intersections with $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ as the convex set $\mathcal{E}$. However, the intersections $\mathcal{K}_1 \cap \mathcal{E}$ and $\mathcal{K}_2 \cap \mathcal{E}$ fail to give $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.

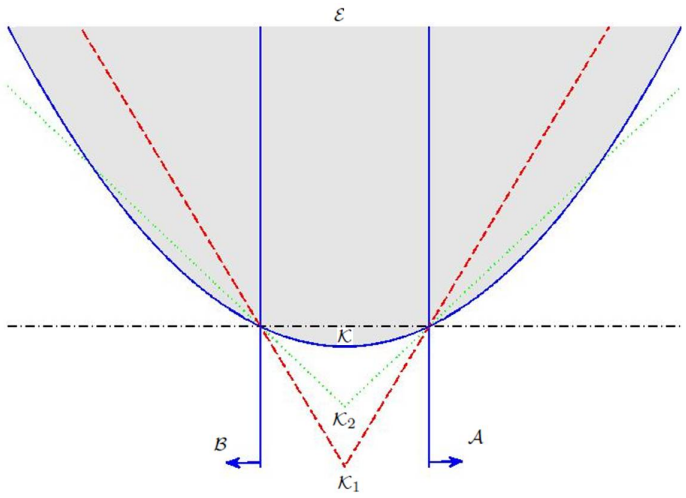
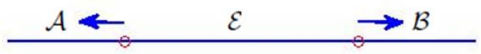


Figure 2.2: Example of unbounded intersections.

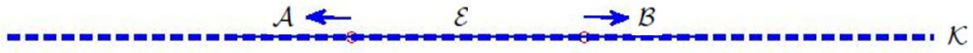
Another important case to consider here is when the set $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$ is of dimension $n = 1$. Figure 2.3(a) illustrates this case. Here, the set $\mathcal{E}$ is given by the solid line segment, and the sets $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ are given by the two circles. In particular, we can see that the uniqueness Lemma 2.5 fails in this case too. Observe the cone $\mathcal{K}_1$ in Figure 2.3(b) and the cone $\mathcal{K}_2$ in Figure 2.3(c), which are represented by two dashed half lines. These two cones have the same intersections with $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ as the set $\mathcal{E}$. However, the intersections $\mathcal{E} \cap \mathcal{K}_1$ and $\mathcal{E} \cap \mathcal{K}_2$ differ from $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$. In this case, only the cone $\mathcal{K}$ in Figure 2.3(d), given by a whole line, is such that $\mathcal{E} \cap \mathcal{K} = \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$. Finally, in these two examples the condition given in Proposition 2.1 is still necessary, but it is not a sufficient condition.



(a) $\mathcal{A}^{=}$, $\mathcal{B}^{=}$, and $\mathcal{E}$.

(b) Cone $\mathcal{K}_1$ does not give $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.

(c) Cone $\mathcal{K}_2$ does not give $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.



(d) $\mathcal{K} \cap \mathcal{E} = \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.

Figure 2.3: Example when the set $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$ has dimension $n = 1$.

2.2 Disjunctive cylindrical cuts

Let us first define the concept of a base of a cylinder, which is based on Definition 2.3 of a base of a cone and on Definition 1.16.

Definition 2.4 (Base of a Convex Cylinder)**.** *Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex cylinder with the direction $d^0 \in \mathbb{R}^n$. A set $\mathcal{D} \subset \mathcal{C}$ is called a* base *of $\mathcal{C}$ if for every vector $x \in \mathcal{C}$, there is a unique $d \in \mathcal{D}$ and $\sigma \in \mathbb{R}$ such that $x = d + \sigma d_0$.*

Definition 2.5 (Disjunctive Cylindrical Cut)**.** *Let $\mathcal{E}$ be a closed convex set. A closed convex cylinder $\mathcal{C}$ is a* Disjunctive Cylindrical Cut *(DCyC) for the set $\mathcal{E}$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$ if*

$$\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})) = \mathcal{C} \cap \mathcal{K}.$$

For the results in this section, Assumption 2.2 may be relaxed.

Assumption 2.3. *The intersections $\mathcal{E} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=}$ are nonempty.*

Assumption 2.1 and 2.3 are assumed to hold in the remainder of this section. The following proposition gives a necessary and sufficient condition for a convex cylinder $\mathcal{C}$ to be a DCyC.

Proposition 2.2. *A convex cylinder $\mathcal{C} \in \mathbb{R}^n$ with a unique direction $d^0 \in \mathbb{R}^n$, such that $a^\top d^0 \neq 0$ and $b^\top d^0 \neq 0$, is DCyC for $\mathcal{E}$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$ if and only if*

$$\mathcal{C} \cap \mathcal{A}^= = \mathcal{E} \cap \mathcal{A}^= \quad \text{and} \quad \mathcal{C} \cap \mathcal{B}^= = \mathcal{E} \cap \mathcal{B}^=. \tag{2.1}$$

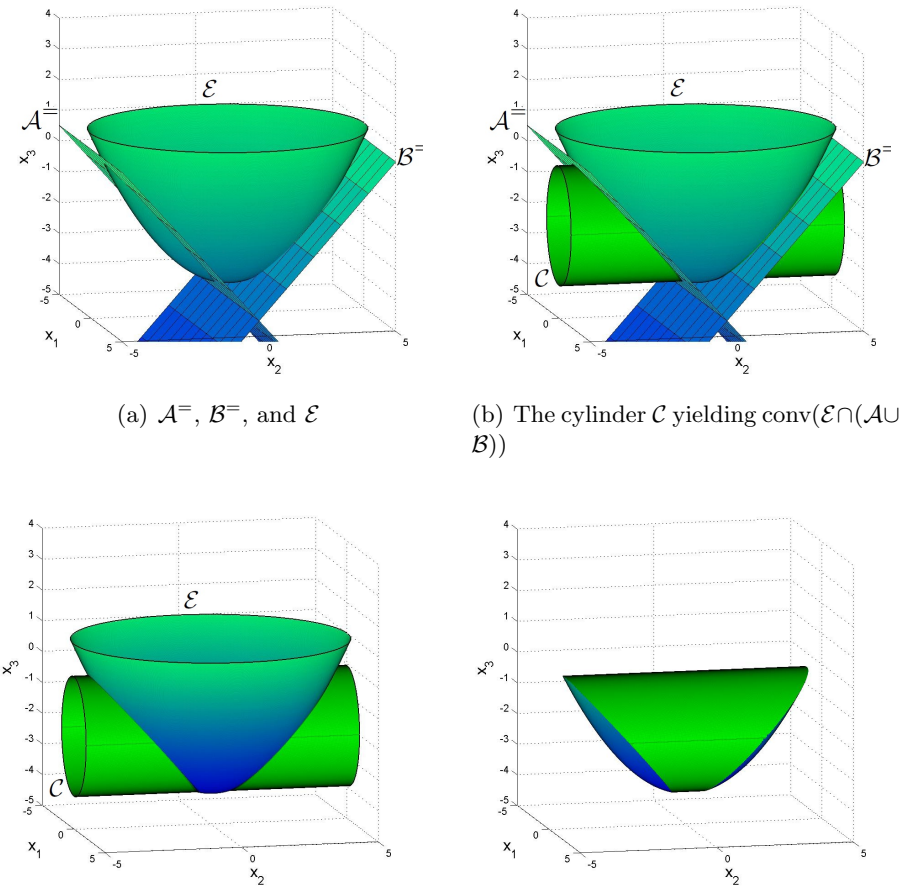
Figure 2.4 illustrates Proposition 2.2 where the set $\mathcal{E}$ is the epigraph of a paraboloid. Before proving Proposition 2.2, we first provide a set of lemmas that help to develop the proof.

Lemma 2.6. *Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex cylinder with a unique direction $d^0 \in \mathbb{R}^n$. Consider a half space $\mathcal{G} = \{x \in \mathbb{R}^n | g^\top x \leq \varrho\}$, such that $g^\top d^0 \neq 0$ and $\mathcal{E} \cap \mathcal{G}^=$ is nonempty. If $\mathcal{C} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$, then $\mathcal{E} \cap \mathcal{G}^=$ is a base for $\mathcal{C}$.*

Proof. Consider the convex cylinder $\mathcal{C} = \{x \in \mathbb{R}^n \mid x = d + \sigma d_0, d \in \mathcal{D}, \sigma \in \mathbb{R}\}$, where $\mathcal{D} \in \mathbb{R}^n$ is a convex set and $d^0 \in \mathbb{R}^n$. Assume that $\mathcal{C} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$. Observe that if $g^\top d_0 = 0$, then for any $u \in \mathcal{C}$ such that $u \notin \mathcal{G}^=$ we have that $\{v \in \mathbb{R}^n \mid v = u + \sigma d_0, \sigma \in \mathbb{R}\} \cap \mathcal{C} \cap \mathcal{G}^= = \emptyset$. Thus, if $g^\top d_0 = 0$ the set $\mathcal{C} \cap \mathcal{G}^=$ is not a base of $\mathcal{C}$. Assume now that if $g^\top d_0 \neq 0$, then for any vector $w \in \mathcal{C}$ there is a $\tilde{\sigma} \in \mathbb{R}$ such that $w + \tilde{\sigma} d^0 \in \mathcal{G}^= \cap \mathcal{C}$. That is, the set $\mathcal{G}^= \cap \mathcal{C}$ is a base of $\mathcal{C}$. Thus, since $\mathcal{C} \cap \mathcal{G}^= = \mathcal{E} \cap \mathcal{G}^=$ we have that $\mathcal{E} \cap \mathcal{G}^=$ is a base of $\mathcal{C}$. □

The next lemma states the relationship between cylinder $\mathcal{C}$ and the intersections of $\mathcal{E}$ with the half spaces $\mathcal{A}$ and $\mathcal{B}$.

Lemma 2.7. *Let $\mathcal{C}$ be a convex cylinder $\mathcal{C}$ with a unique direction $d^0 \in \mathbb{R}^n$, such that*



(a) $\mathcal{A}^=$, $\mathcal{B}^=$, and $\mathcal{E}$

(b) The cylinder $\mathcal{C}$ yielding $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$

(c) $\mathcal{E} \cap \mathcal{C}$

(d) $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$

Figure 2.4: Illustration of a disjunctive cylindrical cut as specified in Proposition 2.2

$a^\top d^0 \neq 0$ *and* $b^\top d^0 \neq 0$*, and for which condition* (2.1) *holds. Then*

$$(\mathcal{E} \cap \mathcal{A}) \subset \mathcal{C} \quad \textit{and} \quad (\mathcal{E} \cap \mathcal{B}) \subset \mathcal{C}.$$

Proof. Note that if $a^\top d^0 \neq 0$, then from condition (2.1) and Assumption 2.1 we have that $b^\top d^0 \neq 0$, otherwise $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E} \neq \emptyset$. Now, we prove first that $(\mathcal{E} \cap \mathcal{A}) \subseteq \mathcal{C}$. Let us assume to the contrary that $u \in (\mathcal{E} \cap \mathcal{A})$ but $u \notin \mathcal{C}$. First, by the separation theorem, there exists a hyperplane $\mathcal{H}$ properly separating u from $\mathcal{C}$. From the definition of $\mathcal{C}$

we have that $\mathcal{H}$ is parallel to d_0. Now, let $\mathcal{H}$ be a supporting hyperplane of $\mathcal{C}$, which implies that $\mathcal{H} \cap \mathcal{C}$ is an exposed face of $\mathcal{C}$. Note that for any $v \in \mathcal{H} \cap \mathcal{C}$ the inclusion $\{w \in \mathbb{R}^n \mid w = v + \sigma d_0, \sigma \in \mathbb{R}\} \subseteq \mathcal{H} \cap \mathcal{C}$ holds. Additionally, since $a^\top d^0 \neq 0$, then by condition (2.1), and Lemma 2.6, the sets $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are bases for $\mathcal{C}$. Hence, there exists a vector $\hat{w} \in \mathcal{E} \cap \mathcal{B}^=$ such that $\hat{w} \in \mathcal{H}$, and $\hat{w}$ is in an exposed face of $\mathcal{C}$.

Convexity of $\mathcal{E}$ implies $\lambda u + (1-\lambda)\hat{w} \in \mathcal{E}$ for any $\lambda \in [0,1]$. On the other hand, the vector $\hat{w}$ is in an exposed face of $\mathcal{C}$, then by convexity of $\mathcal{C}$ we obtain that $\lambda u + (1-\lambda)\hat{w} \notin \mathcal{C}$ for $0 < \lambda \leq 1$. Since $u \in (\mathcal{E} \cap \mathcal{A})$ and $\mathcal{A} \cap \mathcal{B} \cap \mathcal{E} = \emptyset$, we have that $a^\top u \leq \alpha$ and $a^\top \hat{w} > \alpha$. Hence, from the equation $a^\top(\lambda u + (1-\lambda)\hat{w}) = \lambda a^\top u + (1-\lambda) a^\top \hat{w}$, we obtain that there exists a value $0 < \lambda \leq 1$ such that $a^\top(\lambda u + (1-\lambda)\hat{w}) = \alpha$. Therefore, there is a $0 < \lambda \leq 1$ such that $\tilde{u} = \lambda u + (1-\lambda)\hat{w} \in \mathcal{E} \cap \mathcal{A}^=$, but $\tilde{u} \notin \mathcal{C}$, which contradicts condition (2.1). Hence, $(\mathcal{E} \cap \mathcal{A}) \subseteq \mathcal{C}$. One can prove $(\mathcal{E} \cap \mathcal{B}) \subseteq \mathcal{C}$ analogously.

Recall that the sets $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are disjoint and nonempty. Then, Definition 2.5 implies that $\mathcal{E} \cap \mathcal{A} \neq \mathcal{C}$ and $\mathcal{E} \cap \mathcal{B} \neq \mathcal{C}$, this proofs the lemma. □

We can now present the proof of Proposition 2.2.

Proof of Proposition 2.2. First, consider a vector $u \in (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. Then, Lemma 2.7 implies that $u \in \mathcal{E} \cap \mathcal{C}$. Consider any two vectors $u, v \in (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. Then, since both $\mathcal{C}$ and $\mathcal{E}$ are convex, for all $0 \leq \lambda \leq 1$ the convex combination $\lambda x + (1-\lambda) y \in \mathcal{E} \cap \mathcal{C}$. Hence, $\text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})) \subseteq (\mathcal{E} \cap \mathcal{C})$.

Consider now a vector $u \in (\mathcal{E} \cap \mathcal{C})$. First, if $u \in (\mathcal{E} \cap \mathcal{A})$ or $u \in (\mathcal{E} \cap \mathcal{B})$, we have that $u \in \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$. Suppose now that $u \notin (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. Then, $u \in (\overline{\mathcal{A}} \cap \overline{\mathcal{B}} \cap \mathcal{C})$. Furthermore, by Lemma 2.6 there are two vectors $v \in \mathcal{E} \cap \mathcal{A}^=$ and $w \in \mathcal{E} \cap \mathcal{B}^=$ such that $u = v + \mu d_0$ and $u = w + \nu d_0$, for some $\mu, \nu \in \mathbb{R}$. Thus, given that $u \notin (\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$ we can assume w.l.o.g. that $\nu > 0$ and $\mu < 0$. Then, we have that $u = \lambda v + (1-\lambda) w$, where $\lambda = \nu/(\nu - \mu)$ and $0 < \lambda < 1$. In other words, u is a convex combination of v and

w. Since u is an arbitrary vector we have that any vector $u \in (\mathcal{E} \cap \mathcal{C})$ can be written as a convex combination of two vectors in $(\mathcal{E} \cap \mathcal{A}) \cup (\mathcal{E} \cap \mathcal{B})$. As a conclusion, we have that $(\mathcal{E} \cap \mathcal{C}) \subset \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.

Now we prove that if $\mathcal{C}$ is a disjunctive cylindrical cut for $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$, then $\mathcal{E} \cap \mathcal{A}^{=} = \mathcal{C} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=} = \mathcal{C} \cap \mathcal{B}^{=}$. From Assumption 2.1 and Definition 2.5 we have that $\mathcal{E} \cap \mathcal{A}^{=} \subseteq \mathcal{C} \cap \mathcal{A}^{=}$, then for a given $u \in \mathcal{E} \cap \mathcal{A}^{=}$ we have that $u \in \mathcal{C} \cap \mathcal{A}^{=}$. Henceforth, $\mathcal{E} \cap \mathcal{A}^{=} \subseteq \mathcal{C} \cap \mathcal{A}^{=}$. We can show similarly that $\mathcal{E} \cap \mathcal{B}^{=} \subseteq \mathcal{C} \cap \mathcal{B}^{=}$.

Assume now that $\mathcal{E} \cap \mathcal{A}^{=}$ is a proper subset of $\mathcal{C} \cap \mathcal{A}^{=}$, i.e., $\mathcal{E} \cap \mathcal{A}^{=} \subset \mathcal{C} \cap \mathcal{A}^{=}$. Then, there is a vector $u \in \mathcal{E} \cap \mathcal{A}^{=}$ such that $u \notin \mathcal{C} \cap \mathcal{A}^{=}$, which implies that $u \notin \mathcal{C}$. Hence, $u \in \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$ but $u \notin \mathcal{E} \cap \mathcal{C}$, which violates Definition 2.5 of a disjunctive cylinder for $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$. Similarly, we can show that $\mathcal{E} \cap \mathcal{B}^{=}$ is not a proper subset of $\mathcal{C} \cap \mathcal{B}^{=}$. Therefore, we have that $\mathcal{E} \cap \mathcal{A}^{=} = \mathcal{C} \cap \mathcal{A}^{=}$ and $\mathcal{E} \cap \mathcal{B}^{=} = \mathcal{C} \cap \mathcal{B}^{=}$. □

Lemma 2.8. *If a DCyC $\mathcal{C} \in \mathbb{R}^n$ with a direction $d^0 \in \mathbb{R}^n$ exists for $\mathcal{E}$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$, such that $a^\top d^0 \neq 0$ and $b^\top d^0 \neq 0$, then $\mathcal{C}$ is unique.*

Proof. Assume that there exist two different DCyC $\mathcal{C}_1 = \{x \in \mathbb{R}^n \mid x = d + \sigma d^0, d \in \mathcal{D}_1, \sigma \in \mathbb{R}\}$ and $\mathcal{C}_2 = \{x \in \mathbb{R}^n \mid x = h + \gamma h^0, h \in \mathcal{D}_2, \gamma \in \mathbb{R}\}$ such that $a^\top d^0 \neq 0$ and $a^\top h^0 \neq 0$. Then, we have that $\mathcal{C}_1 \cap \mathcal{A}^{=} = \mathcal{C}_2 \cap \mathcal{A}^{=}$ and $\mathcal{C}_1 \cap \mathcal{B}^{=} = \mathcal{C}_2 \cap \mathcal{B}^{=}$.

Given that $\mathcal{C}_1 \neq \mathcal{C}_2$ there must exist a vector u that belongs only to one cylinder, and w.l.o.g. we assume that $u \in \mathcal{C}_1$ and $u \notin \mathcal{C}_2$. Observe that, by Assumption 2.1, $u \notin \mathcal{A} \cap \mathcal{B}$.

Let us begin assuming that $u \in \bar{\mathcal{A}} \cap \bar{\mathcal{B}}$. Then, given that $\mathcal{E} \cap \mathcal{A}^{=}$ is a base for both cylinders there exists a $\sigma_1 \in \mathbb{R}$ such that $u = d^1 + \sigma_1 d^0$ for some $d^1 \in \mathcal{E} \cap \mathcal{A}^{=} = \mathcal{C}_1 \cap \mathcal{A}^{=} = \mathcal{C}_2 \cap \mathcal{A}^{=}$. On the other hand, since $\mathcal{E} \cap \mathcal{B}^{=}$ is a base for $\mathcal{C}_1$, there exist $\sigma_2 \in \mathbb{R}$ such that $u = d^2 + d^0\sigma_2$ for some $d^2 \in \mathcal{E} \cap \mathcal{B}^{=} = \mathcal{C}_1 \cap \mathcal{B}^{=}$. Hence, $u = \lambda d^1 + (1 - \lambda)d^2$ where $\lambda = \sigma_1/(\sigma_1 - \sigma_2) \leq 1$, since σ_1 and σ_2 must have opposite signs. Additionally, given that the two cylinders are convex we obtain that $d^2 \notin \mathcal{C}_2$. Then, $\mathcal{C}_1 \cap \mathcal{B}^{=} \neq \mathcal{C}_2 \cap \mathcal{B}^{=}$, a contradiction.

Let us assume now that $u \in \mathcal{A}$ and $u \notin \mathcal{B}$. By the separation Theorem 1.5, there exists a hyperplane $\mathcal{H}$ properly separating u from $\mathcal{C}_2$. From Definition 1.16 of a cylinder, we have $\mathcal{H}$ is parallel h^0. Now, let $\mathcal{H}$ be a supporting hyperplane of $\mathcal{C}_2$, which implies that $\mathcal{H} \cap \mathcal{C}_2$ is an exposed face of $\mathcal{C}_2$. Note that for any $v \in \mathcal{H} \cap \mathcal{C}_2$ we have that $\{w \in \mathbb{R}^n \mid w = v + \sigma h^0, \sigma \in \mathbb{R}\} \subseteq \mathcal{H} \cap \mathcal{C}_2$. Additionally, we know that the sets $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are bases of $\mathcal{C}_2$. Hence, there exists a vector $\bar{w} \in \mathcal{E} \cap \mathcal{B}^=$ such that $w \in \mathcal{H}$, and w is contained in an exposed face of $\mathcal{C}_2$.

Convexity of $\mathcal{C}_1$ implies that for any $\lambda \in [0, 1]$, $\lambda u + (1 - \lambda)\bar{w} \in \mathcal{C}_1$. On the other hand, since $\bar{w} \in \mathcal{H}$ is on exposed face of $\mathcal{C}_2$, $\lambda u + (1 - \lambda)\bar{w} \notin \mathcal{C}_2$ for $0 < \lambda \leq 1$. Since $u \in \mathcal{A} \cap \mathcal{C}_1$ and $\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}_1 = \emptyset$, we have that $a^\top u \leq \alpha$ and $a^\top \bar{w} > \alpha$. Hence, from the equation $a^\top(\lambda u + (1 - \lambda)\bar{w}) = \lambda a^\top u + (1 - \lambda)a^\top \bar{w}$, there exists a value $0 < \lambda \leq 1$ such that $a^\top(\lambda u + (1 - \lambda)\bar{w}) = \alpha$. Therefore, there exists a vector $\bar{u} = \lambda u + (1 - \lambda)w$ for some $0 < \lambda \leq 1$, such that $\bar{u} \in \mathcal{C}_1 \cap \mathcal{A}^=$, but $\bar{u} \notin \mathcal{C}_2$, which is a contradiction. An analogous argument can be used when $u \in \mathcal{B}$ and $u \notin \mathcal{A}$. This proof the lemma. □

In this section we kept the assumption that $\dim(\mathcal{E}) \geq 2$. This assumption is needed in the proofs of Proposition 2.2 and Lemma 2.7 for the sake of the separation argument. However, this assumption excludes the case when the DCyC is be a line. In this case, the separation argument is not needed and the result becomes trivial. In particular if $\mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{E} \cap \mathcal{B}^=$ are to points, then the disjunctive cylindrical cut $\mathcal{C}$ must be a line. In this case, the sets $\mathcal{E} \cap \mathcal{A}$ and $\mathcal{E} \cap \mathcal{B}$ must be either two half lines or two points, since $\mathcal{E}$ is a convex set. As a direct consequence we have that $\mathcal{E} \subset \mathcal{C}$, which implies that $\mathcal{E} \cap \mathcal{C} = \mathcal{E} = \text{conv}(\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B}))$.

Chapter 3

Analysis of quadrics

In this chapter, we consider a given quadric $\mathcal{Q}$ represented by (P, p, ρ), where $P \in \mathbb{R}^{\ell \times \ell}$ is a symmetric matrix with at most one non-positive eigenvalue, $p \in \mathbb{R}^{\ell}$ and $\rho \in \mathbb{R}$. In particular, we focus on the intersection of the quadric represented by (P, p, ρ) an two given hyperplanes. The results described here are based on those reported in Belotti et al. [2013b] and Belotti et al. [2013a]. We begin the discussion with a description of some affine transformations in Section 3.1 that will simplify the algebra in the analysis of Sections 3.2 and 3.3. In Section 3.2, we consider the case when the two hyperplanes are parallel. Finally, in Section 3.3, we analyze the case when the two hyperplanes are in general position. In this chapter, we use a set of well known results of linear algebra. The interested reader can find an extensive review of these concepts and results in textbooks, such as like Golub and Van Loan [1996], Lancaster and Tismenetsky [1985], Searle [1982].

3.1 Affine transformations of quadrics

In this section, we define some convenient affine transformations that will simplify the algebra in Sections 3.2.3 and 3.3.2. The affine transformation defined here can be applied to the quadric represented by (P, p, ρ) to transform it into a simpler object. This will

allow us to focus our analysis on the geometry of this simpler object independent of its representation.

Since P is a real symmetric matrix, it is well known that P can be factorized as $P = VDV^\top$, where $V \in \mathbb{R}^{\ell\times\ell}$ is an orthonormal matrix and $D \in \mathbb{R}^{\ell\times\ell}$ is a diagonal matrix Searle [1982]. This factorization gives the eigenvalue and eigenvector decomposition of P, where the diagonal elements of D are the eigenvalues of P and the columns of V are the normalized eigenvectors of P. Now, recall the concept of inertia of a matrix given in Definition 1.18. Then, it is clear that $\text{In}(P) = \text{In}(D)$.

The rest of the discussion about affine transformations defined in this section is separated into two cases. We discuss the case when P is a non-singular matrix in Section 3.1.1 and the case when P is a singular matrix in Section 3.1.2.

3.1.1 The matrix P is non-singular

Consider the case when the matrix P is non-singular and recall expression (1.4). Then the quadric represented by (P, p, ρ) can be given by

$$\left\{ w \in \mathbb{R}^\ell \mid (w + P^{-1}q)^\top P(w + P^{-1}p) \leq pP^{-1}p - \rho \right\}. \tag{3.1}$$

We can use the eigenvalue and eigenvector decomposition $P = VDV^\top$ to rewrite (3.1) in terms of V and D. First, let $\tilde{J}$ be the $\ell \times \ell$ diagonal matrix defined as

$$\tilde{J}_{i,i} = \frac{D_{i,i}}{|D_{i,i}|}, i = 1, \ldots, \ell. \tag{3.2}$$

Observe that $\tilde{J}$ is the identity if $P \succ 0$. On the other hand, if $D_{k,k} < 0$ for some $k \in \{1, \ldots, \ell\}$, then we have that $\tilde{J}_{k,k} = -1$, thus we have that $\text{In}(\tilde{J}) = \text{In}(P)$. Now, let $\tilde{D} \in \mathbb{R}^{\ell\times\ell}$ be a diagonal matrix defined as $\tilde{D}_{i,i} = |D_{i,i}|$, $i = 1, \ldots, \ell$. Therefore, the set

(3.1) has the following equivalent description

$$\{w \in \mathbb{R}^\ell \mid \left(w + P^{-1}p\right)^\top (V\tilde{D}^{\frac{1}{2}})\tilde{J}(\tilde{D}^{\frac{1}{2}}V^\top)\left(w + P^{-1}p\right) \leq p^\top P^{-1}p - \rho\}. \tag{3.3}$$

Consider an affine transformation $\mathsf{L} : \mathbb{R}^\ell \mapsto \mathbb{R}^\ell$ defined by

$$\mathsf{L}(w) = \tilde{D}^{\frac{1}{2}}V^\top \left(w + P^{-1}p\right). \tag{3.4}$$

Recall that V is a an orthonormal matrix, and we know that $\tilde{D}$ is non-singular by definition. Hence, the matrix $\tilde{D}^{\frac{1}{2}}V^\top$ is non-singular. Using (3.4) we show that there is a one-to-one mapping between every element of (3.3) to elements in the quadric

$$\left\{u \in \mathbb{R}^n \mid u^\top \tilde{J}u \leq \delta\right\}, \tag{3.5}$$

where the value of the scalar δ depends on the quantity $p^\top P^{-1}p - \rho$.

In first place, if $p^\top P^{-1}p - \rho \neq 0$, then this quantity can be either positive or negative. With this in mind, let us define

$$u = \frac{1}{\sqrt{|p^\top P^{-1}p - \rho|}}\mathsf{L}(w) \quad \text{and} \quad \delta = \frac{p^\top P^{-1}p - \rho}{|p^\top P^{-1}p - \rho|}. \tag{3.6}$$

Then, since $\tilde{D}^{\frac{1}{2}}V^\top$ is non-singular, using (3.6) we obtain a one-to-one mapping between every element of the set (3.3) and the set (3.5). On the other hand, if $p^\top P^{-1}p - \rho = 0$, define

$$u = \mathsf{L}(w) \quad \text{and} \quad \delta = 0. \tag{3.7}$$

In this case, using (3.7) we obtain a one-to-one mapping between the sets (3.3) and (3.5) as well.

A consequence of using transformation (3.4) is that the classification of the quadrics

(3.1) and (3.5) in Table 1.1 is the same. The shape of the quadrics is determined by the inertia of P, which is the same of $\tilde{J}$, and the sign of the quantities $p^\top P^{-1}p - \rho$, which is the same as the sign of δ. Hence, these two conditions show that (3.1) and (3.5) have the same classification in Table 1.1.

3.1.2 The matrix P is singular

Now, consider the case when P is a singular matrix. Recall that P has at most one non-positive eigenvalue. Also, recall the representation of a quadric given in Definition 1.17, which is

$$\left\{ w \in \mathbb{R}^\ell \mid w^\top P w + 2p^\top w + \rho \leq 0 \right\}. \tag{3.8}$$

If P is singular, its non-positive eigenvalue is zero. Hence, there exist a $j \in \{1, \ldots, \ell\}$ such that $D_{j,j} = 0$ for the matrix D from the diagonalization of P. In this case we can define a diagonal matrix $\bar{D} \in \mathbb{R}^{\ell \times \ell}$ as $\bar{D}_{i,i} = D_{i,i}$ for $i \in \{1, \ldots, \ell\} \setminus j$ and $\bar{D}_{j,j} = 1$. Additionally, let $\bar{J} \in \mathbb{R}^{\ell \times \ell}$ be a diagonal matrix defined as

$$\bar{J}_{i,i} = 1, i \in \{1, \ldots, \ell\} \setminus j, \quad \text{and} \quad \bar{J}_{j,j} = 0. \tag{3.9}$$

Thus, for the set (3.8) we have the following equivalent description

$$\left\{ w \in \mathbb{R}^\ell \mid w^\top V \bar{D}^{\frac{1}{2}} \bar{J} \bar{D}^{\frac{1}{2}} V^\top w + 2(p^\top V \bar{D}^{-\frac{1}{2}})(\bar{D}^{\frac{1}{2}} V^\top w) + \rho \leq 0 \right\}, \tag{3.10}$$

Consider an affine transformation $\mathsf{L} : \mathbb{R}^\ell \mapsto \mathbb{R}^\ell$ defined by

$$\mathsf{L}(w) = \bar{D}^{\frac{1}{2}} V^\top w. \tag{3.11}$$

Recall that V is a an orthonormal matrix, and we know that $\bar{D}$ is non-singular by definition. Hence, the matrix $\bar{D}^{\frac{1}{2}} V^\top$ is non-singular. We show now that there is a one-to-one mapping

between every element of (3.10) and the quadric

$$\left\{u \in \mathbb{R}^{\ell} \mid u^{\top}\bar{J}u + 2\bar{p}^{\top}u + \omega \leq 0\right\}, \tag{3.12}$$

where the definition of $\bar{p}$ and the scalar ω depend on ρ.

In the first place, if $\rho \neq 0$, then it can be either positive or negative. Consequently, let us define

$$u = \frac{1}{\sqrt{|\rho|}}\mathsf{L}(w), \quad \bar{p} = \frac{1}{\sqrt{|\rho|}}\bar{D}^{-\frac{1}{2}}V^{\top}p, \quad \omega = \frac{\rho}{|\rho|}. \tag{3.13}$$

Then, since $\bar{D}^{\frac{1}{2}}V^{\top}$ is non-singular, using (3.13) we obtain a one-to-one mapping between every element of (3.10) and the elements of the set (3.12). Now, if $\rho = 0$, then we can define

$$u = \mathsf{L}(w), \quad \bar{p} = \bar{D}^{-\frac{1}{2}}V^{\top}p, \quad \omega = 0. \tag{3.14}$$

In this case, using (3.14) we obtain a one-to-one mapping between the sets (3.10) and (3.12) as well.

We need to verify now that the sets (3.8) and (3.12) have the same shape. From Section 1.1.2.1 we know that the shape of these quadrics depends on one hand on the inertia of P, which is the same of $\bar{J}$. The next criteria given in Section 1.1.2.1 is to verify is if there exist both a vector $w^{\mathrm{c}} \in \mathbb{R}^{\ell}$ such that $Pw^{\mathrm{c}} = -p$, and a vector $u^{\mathrm{c}} \in \mathbb{R}^{\ell}$ such that $\bar{J}u^{\mathrm{c}} = -\bar{p}$. Let us assume first that there is no vector x^{c} such that $Px^{\mathrm{c}} = -p$. Then, from the system $VDV^{\top}x^{\mathrm{c}} = -p$ we have that $DV^{\top}x^{\mathrm{c}} = -V^{\top}p$, and we may conclude that p is not orthogonal to the eigenvector associated with the zero eigenvalue of P. As a consequence, we have that $\bar{p}_j \neq 0$, and there is no vector u^{c} such that $\bar{J}u^{\mathrm{c}} = -\bar{p}$. Thus, from **Case 2** in Section 1.1.2.1 we have that (3.8) and (3.12) are two paraboloids.

Let us assume now that a vector w^{c} such that $Pw^{\mathrm{c}} = -p$ exists. Then p must be orthogonal to the eigenvector associated with the zero eigenvalue of P. As a consequence,

we have that $\bar{p}_j = 0$, which ensures the existence of a vector $u^c \in \mathbb{R}^\ell$ such that $\bar{J}z^c = -\bar{p}$. We know from **Case 1** in Section 1.1.2.1 that in this case the set (3.8) is empty if $(w^c)^\top P w^c - \rho < 0$, is a line if $(w^c)^\top P w^c - \rho = 0$, and a cylinder if $(w^c)^\top P w^c - \rho > 0$. Similarly, the shape of the set (3.12) is determined by the quantity $(u^c)^\top \bar{J} u^c - \omega$. Thus, if $\rho = \omega = 0$, using (3.14) we can define $u^c = \bar{D}^{\frac{1}{2}} V^\top w^c$, and we obtain

$$(u^c)^\top \bar{J} u^c = (w^c)^\top V \bar{D}^{\frac{1}{2}} \bar{J} \bar{D}^{\frac{1}{2}} V^\top w^c = (w^c)^\top P w^c.$$

Hence, $(u^c)^\top \bar{J} u^c = (w^c)^\top P w^c \geq 0$, since $P \succeq 0$, and from (3.14) we have that $\bar{J} u^c = \bar{p}$. In this situation we have from **Case 1** in Section 1.1.2.1 that (3.8) and (3.12) are two cylinders if $(w^c)^\top P w^c > 0$, and two lines if $(w^c)^\top P w^c = 0$. On the other hand, if $\rho \neq 0$, then from (3.14) we obtain that $\omega \neq 0$, and the two quantities share the same sign. Additionally, we can define $u^c = \frac{1}{\sqrt{|\rho|}} \bar{D}^{\frac{1}{2}} V^\top w^c$, and we obtain

$$(u^c)^\top \bar{J} u^c - \omega = \frac{1}{|\rho|} \left((w^c)^\top P w^c - \rho \right).$$

Thus, from **Case 1** in Section 1.1.2.1 we know that (3.8) and (3.12) are empty if $(w^c)^\top P w^c - \rho < 0$; are lines if $(w^c)^\top P w^c - \rho = 0$; and cylinders if $(w^c)^\top P w^c - \rho > 0$. In brief, this shows that using the transformation 3.11 the classification of the quadrics (3.8) and (3.12) according to **Case 1** or **Case 2** of Section 1.1.2.1 is always the same.

Tow final remarks about the transformations described in Sections 3.1.2 and 3.1.1 are needed. An advantage of using transformations (3.4) and (3.11) is that affine transformations preserve straight lines and ratios between distances. Thus, after an affine transformation parallel hyperplanes will remain parallel. Recall that the goal of this chapter is to analyze the geometric properties of the intersection of quadrics of the form (3.1) and (3.8) and two given hyperplanes. Then, if P is non-singular we can analyze these properties using the intersection of a quadric of the form (3.5) and the two given hyper-

planes transformed using (3.6) or (3.7). On the other hand, if P is singular we can analyze these properties using the intersection of a quadric of the form (3.12) and the two given hyperplanes transformed using (3.13) or (3.14).

Another important advantage of the transformations (3.4) and (3.11) is that the matrices $\tilde{D}^{\frac{1}{2}}V^\top$ and $\bar{D}^{\frac{1}{2}}V^\top$ are non-singular. Thus, the inverse mapping $\mathsf{L}(x)^{-1}$ is well defined in both cases. This allows us to translate the results of the analysis in the transformed sets to the original sets. This last property is important for the practical application of this results in Chapter 4.

3.2 Intersections with parallel hyperplanes

In this section, we investigate the intersection of the quadric $\mathcal{Q}$ with two parallel hyperplanes. For the sake of simplifying the algebra, w.l.o.g we may assume throughout this section that the quadric $\mathcal{Q}$ is one of the sets (3.5) or (3.12). Recall that the results obtained for this two sets can be generalized using the inverse transformations (3.4) or (3.11). Now, let $\mathcal{A}^= = \{w \in \mathbb{R}^\ell \mid a^\top w = \alpha\}$ and $\mathcal{B}^= = \{w \in \mathbb{R}^\ell \mid a^\top w = \beta\}$ be two given parallel hyperplanes for some $a \in \mathbb{R}^\ell$ and $\alpha, \beta \in \mathbb{R}$, where $\alpha \neq \beta$, and w.l.o.g. we may assume that $\|a\| = 1$. Additionally, assume that the intersections $\mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q} \cap \mathcal{B}^=$ are nonempty. We first present in Section 3.2.1 a theorem that characterizes a family of quadrics having the same intersection with the hyperplanes $\mathcal{A}^=$ and $\mathcal{B}^=$ as the quadric $\mathcal{Q}$. Then, in Section 3.2.2, we recall some results from linear algebra about the eigenvalues of a diagonal matrix modified by a rank one update. Finally, we analyze the family of quadrics to show that there is always a quadric in the family that is either a cone or a cylinder. The analysis of the family is divided in three parts. First, in Section 3.2.3, we consider the case when $P \succ 0$. Then, in Section 3.2.4, we analyze the case when P has one zero eigenvalue. Finally, in Section 3.2.5, we consider the case when P has one negative eigenvalue.

3.2.1 The family of quadrics with fixed parallel planar sections

First we recall the definition of a pencil of quadrics as it is given in Snyder and Sisam [1914].

Definition 3.1. *Consider two given quadrics represented by* (P_1, p_1, ρ_1) *and* (P_2, p_2, ρ_2)*, for* $P_1, P_2 \in \mathbb{R}^{\ell\times\ell}$*,* $p_1, p_2 \in \mathbb{R}^{\ell}$ *and* $\rho_1, \rho_2 \in \mathbb{R}$*. The family of quadrics* $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ *is called a* pencil of quadrics*, where* $\mathcal{Q}(\tau)$ *is represented by* $\hat{P}(\tau) = P_1 + \tau P_2$*,* $\hat{p}(\tau) = p_1 + \tau\tilde{p}_2$*, and* $\hat{\rho}(\tau) = \rho_1 + \tau\tilde{\rho}_2$*.*

Now we characterize a family of quadrics having the same intersection with two hyperplanes $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ as the quadric $\mathcal{Q}$.

Theorem 3.1. *Let a quadric* $\mathcal{Q} \in \mathbb{R}^{\ell}$ *represented by* (P, p, ρ)*, where* $P \in R^{\ell\times\ell}$*,* $p \in \mathbb{R}^{\ell}$*,* $\rho \in \mathbb{R}$*, and two parallel hyperplanes* $\mathcal{A}^{=} = \{x \in \mathbb{R}^{\ell} \mid a^{\top}x = \alpha\}$ *and* $\mathcal{B}^{=} = \{x \in \mathbb{R}^{\ell} \mid a^{\top}x = \beta\}$ *be given. The uni-parametric family of quadrics having the same intersection with* $\mathcal{A}^{=}$ *and* $\mathcal{B}^{=}$ *as the quadric* $\mathcal{Q}$ *is defined by the pencil of quadrics* $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$*, where* $\mathcal{Q}(\tau)$ *is represented by* $(P(\tau), p(\tau), \rho(\tau))$*, and*

$$\begin{aligned} P(\tau) &= P + \tau a a^{\top}, \\ p(\tau) &= p - \tau\frac{(\alpha+\beta)}{2}a, \\ \rho(\tau) &= \rho + \tau\alpha\beta. \end{aligned}$$

Proof. Consider the set $\mathcal{A}^{=} \cup \mathcal{B}^{=}$, which can be described as

$$\{x \in \mathbb{R}^{\ell} \mid (a^{\top}x - \alpha)(a^{\top}x - \beta) = 0\},$$

and observe that

$$(a^{\top}x - \alpha)(a^{\top}x - \beta) = x^{\top}aa^{\top}x - (\alpha+\beta)a^{\top}x + \alpha\beta = 0. \tag{3.15}$$

Now, let

$$\tilde{P} = aa^\top, \qquad \tilde{p} = -\frac{(\alpha+\beta)}{2}a, \qquad \tilde{\rho} = \alpha\beta.$$

Then, the set of solutions of equation (3.15) can be written as a quadric surface $\tilde{\mathcal{Q}}$ represented by $(\tilde{P}, \tilde{p}, \tilde{\rho})$. Now, consider a pencil $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$, where $\mathcal{Q}(\tau)$ is represented by $\hat{P}(\tau) = P + \tau\tilde{P}$, $\hat{p}(\tau) = p + \tau\tilde{p}$, and $\hat{\rho}(\tau) = \rho + \tau\tilde{\rho}$. Let $\bar{x} \in \mathbb{R}^\ell$ be a given vector satisfying $\bar{x}^\top\tilde{P}\bar{x} + 2\tilde{p}^\top\bar{x} + \tilde{\rho} = 0$. Then, for $\tau \in \mathbb{R}$ we have $\bar{x} \in \mathcal{Q}(\tau)$ if and only if

$$\bar{x}^\top(P + \tau\tilde{P})\bar{x} + 2(p + \tau\tilde{p})^\top\bar{x} + (\rho + \tau\tilde{\rho}) = \bar{x}^\top P\bar{x} + 2p^\top\bar{x} + \rho \leq 0.$$

Hence, we have $\bar{x} \in \mathcal{Q}(\tau) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$ if and only if $\bar{x} \in \mathcal{Q} \cap (\mathcal{A}^= \cup \mathcal{B}^=)$ for $\tau \in \mathbb{R}$. □

We call the attention of the reader to the fact that Theorem 3.1 is rather general in that $\mathcal{Q}$ does not need to be constrained to one of the quadrics represented by $(\tilde{J}, 0, \delta)$ or $(\bar{J}, \bar{q}, \omega)$. This assumption is made for the sake of simplifying the algebra in the analysis of the subsequent sections.

3.2.2 Eigenvalues of a diagonal matrix modified by a rank one update

A key component in the analysis of Section 3.2.3 is the inertia of the matrix $P(\tau)$ of Theorem 3.1. Recall that in this chapter we assume that the matrix $P \in \mathbb{R}^{\ell\times\ell}$ is symmetric and has at most one non-positive eigenvalue. In this section we provide exact formulas for the computation of the eigenvalues of the matrix

$$P + \tau aa^\top,$$

where $\tau \in \mathbb{R}$, and $a \in \mathbb{R}^\ell$ is a vector with $\|a\| = 1$.

The eigenvalues of $P + \tau aa^\top$ can be computed finding the roots of the equation

$$\det(P + \tau aa^\top - \lambda I) = 0,$$

which is shown in Golub [1973] to be equivalent to the characteristic equation

$$\prod_{i=1}^{n}(P_{i,i} - \lambda) + \tau \sum_{i=1}^{n} a_i^2 \prod_{\substack{j=1 \\ j \neq i}}^{n}(P_{i,i} - \lambda) = 0. \tag{3.16}$$

We use this equation in Sections 3.2.3, 3.2.4, and 3.2.5 to find a specific formula for the eigenvalues of $P + \tau aa^\top$ suitable for each case.

3.2.3 Classification of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ when $P \succ 0$

Here we focus on the case when $P \succ 0$, i.e., we assume that the quadric $\mathcal{Q}$ is an ellipsoid. A consequence of this assumption is that the sets $\mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q} \cap \mathcal{B}^=$ are bounded. We may assume w.l.o.g. that the quadric $\mathcal{Q}$ is not a single point, since otherwise $\alpha = \beta$. Recall the affine transformation defined in Section 3.1.1. Hence, to simplify the algebra, we may assume w.l.o.g. that $\mathcal{Q}$ is a unit hypersphere centered at the origin, and recall that $\|a\| = 1$. Thus, we have that $P = I$, $p = 0$, and $\rho = -1$, and the representation of $\mathcal{Q}(\tau)$ in this section is defined by

$$P(\tau) = I + \tau aa^\top, \quad p(\tau) = -\tau \frac{\alpha + \beta}{2} a, \quad \rho(\tau) = -1 + \tau\alpha\beta. \tag{3.17}$$

Our goal is to characterize the behavior of the family $\{\mathcal{Q}(\tau) | \tau \in \mathbb{R}\}$ defined by (3.17) as a function of the parameter τ. First, we need a result about the inertia of $P(\tau)$.

Lemma 3.1. *The matrix $P(\tau)$ can be classified as a function of parameter τ as follows:*

- $P(\tau) \succ 0$ *if* $\tau > -1$,

- $P(\tau) \succeq 0$ *with one zero eigenvalue if* $\tau = -1$,
- $P(\tau)$ *is ID1 if* $\tau < -1$.

Proof. If $P = I$, then the characteristic polynomial (3.16) simplifies to

$$(1-\lambda)^{n-1}(1-\lambda+\tau\|a\|^2) = (1-\lambda)^{n-1}(1-\lambda+\tau) = 0.$$

Thus, in this case the eigenvalues are 1 with multiplicity $n-1$ and $1+\tau$. Hence, the eigenvalues of $P(\tau)$ are all positive if $\tau > -1$. The matrix $P(\tau)$ has $n-1$ positive eigenvalues and one zero eigenvalue if $\tau = 1$. Finally, the matrix $P(\tau)$ has $n-1$ positive eigenvalues and one negative eigenvalue if $\tau < -1$. This proves the lemma. □

Given the result in Lemma 3.1 we have two cases to analyze: $P(\tau)$ is non-singular and $P(\tau)$ is singular. In the following sections, we analyze these two cases separately.

3.2.3.1 $P(\tau)$ is non-singular

If $\tau \neq -1$, we obtain from Lemma 3.1 that $P(\tau)$ is non-singular, which relates to the cases in Table 1.1 in the background section. Hence, we show here the existence of a $\tau \in \mathbb{R}$ for which $p(\tau)^\top P(\tau)^{-1} p(\tau) - \bar{\rho}(\tau) = 0$, i.e., for which $\mathcal{Q}(\tau)$ is a cone.

We use the Sherman-Morrison-Woodbury formula Golub and Van Loan [1996] to compute the inverse of $P(\tau)$:

$$P(\tau)^{-1} = \left(I + \tau a a^\top\right)^{-1} = I - \frac{\tau}{1+\tau} a a^\top. \tag{3.18}$$

As expected from Lemma 3.1, the inverse does not exist if $\tau = -1$. This case is discussed in Section 3.2.3.2.

Now, using (3.18) and the expressions in (3.17) we can express the quantity $p(\tau)^\top P(\tau)^{-1} p(\tau) -$

$\rho(\tau)$ in terms of α an β. Then, we have:

$$\begin{aligned}
p(\tau)^\top P(\tau)^{-1} p(\tau) - \rho(\tau) &= \left(-\frac{\tau(\alpha+\beta)}{2}\, a\right)^\top \left(I + \tau a a^\top\right)^{-1} \left(-\frac{\tau(\alpha+\beta)}{2}\, a\right) - (-1+\tau\alpha\beta) \\
&= \frac{\tau^2\,(\alpha+\beta)^2}{4} a^\top \left(I - \frac{\tau}{1+\tau} a a^\top\right) a - (\tau\alpha\beta - 1) \\
&= \frac{\tau^2(\alpha+\beta)^2}{4}\left(1 - \frac{\tau}{1+\tau}\right) - (\tau\alpha\beta - 1) \\
&= \frac{\tau^2(\alpha+\beta)^2}{4}\left(\frac{1}{1+\tau}\right) - (\tau\alpha\beta - 1) \\
&= \frac{\tau^2(\alpha+\beta)^2 - 4\tau\alpha\beta + 4 - 4\tau^2\alpha\beta + 4\tau}{4(1+\tau)} \\
&= \frac{4\tau^2\left(\frac{(\alpha+\beta)^2}{4} - \alpha\beta\right) + 4\tau(1-\alpha\beta) + 4}{4(1+\tau)} \\
&= \frac{\tau^2\frac{(\alpha-\beta)^2}{4} + \tau(1-\alpha\beta) + 1}{(1+\tau)}.
\end{aligned} \tag{3.19}$$

Since $\tau \neq -1$, then the denominator in (3.19) is non-zero. Hence, we need to focus only on the roots of the numerator in (3.19). Let $f : \mathbb{R} \mapsto \mathbb{R}$ the function defined as value is

$$f(\tau) = \tau^2 \frac{(\alpha-\beta)^2}{4} + \tau(1-\alpha\beta) + 1. \tag{3.20}$$

Clearly $f(\tau)$ is a quadratic function of τ, and let $\bar{\tau}_1$ and $\bar{\tau}_2$ represent the roots of f.

The discriminant of f is:

$$(1-\alpha\beta)^2 - 4\left(\frac{(\alpha-\beta)^2}{4}\right) = (1-\alpha^2)(1-\beta^2). \tag{3.21}$$

Therefore, if $(1-\alpha^2) \geq 0$ and $(1-\beta^2) \geq 0$, then f has real roots. Thus, since $\mathcal{Q}$ is a unit sphere we have that f has real roots when $\mathcal{Q} \cap \mathcal{A}^= \neq \emptyset$ and $\mathcal{Q} \cap \mathcal{B}^= \neq \emptyset$, which were our assumptions. Note that if either $(1-\alpha^2) = 0$ or $(1-\beta^2) = 0$ but not both, then one of the hyperplanes is tangent to $\mathcal{Q}$. Now, observe that when $(1-\alpha^2) = 0$ and $(1-\beta^2) = 0$,

then $|\alpha| = |\beta| = 1$ and there are two particular cases to consider. First, if α and β have the same sign, then $f(\tau) = 1$ for $\tau \in \mathbb{R}$. In other words, the set $\mathcal{Q} \cap \mathcal{A}^{=} \cap \mathcal{B}^{=}$ is a single point, however this case is not possible with the assumption $\alpha \neq \beta$. Second, if α and β have opposite signs, then $f(\tau) = (\tau + 1)^2$ and the two roots of f are equal to -1. This last case is cover with the discussion in Section 3.2.3.2.

Now, given two different hyperplanes, i.e., $\alpha \neq \beta$, the coefficient of τ^2 in $f(\tau)$ is positive. For the coefficient of τ in $f(\tau)$ we have $1 - \alpha\beta \geq 0$, where the inequality is implied by the assumption that $\mathcal{Q} \cap \mathcal{A}^{=} \neq \emptyset$ and $\mathcal{Q} \cap \mathcal{B}^{=} \neq \emptyset$. This shows that all three coefficients in $f(\tau)$ are non-negative. Hence, we have $\bar{\tau}_1 < 0$ and $\bar{\tau}_2 < 0$.

Let us see how the two roots of f compare to -1, at which value $P(\tau)$ becomes singular. We have

$$f(-1) = \frac{(\alpha - \beta)^2}{4} - (1 - \alpha\beta) + 1 = \frac{(\alpha + \beta)^2}{4} \geq 0, \tag{3.22}$$

thus -1 is not between the two roots of f. Next, we check the derivative of $f(-1)$ to decide on which branch of f the value -1 lies. We have

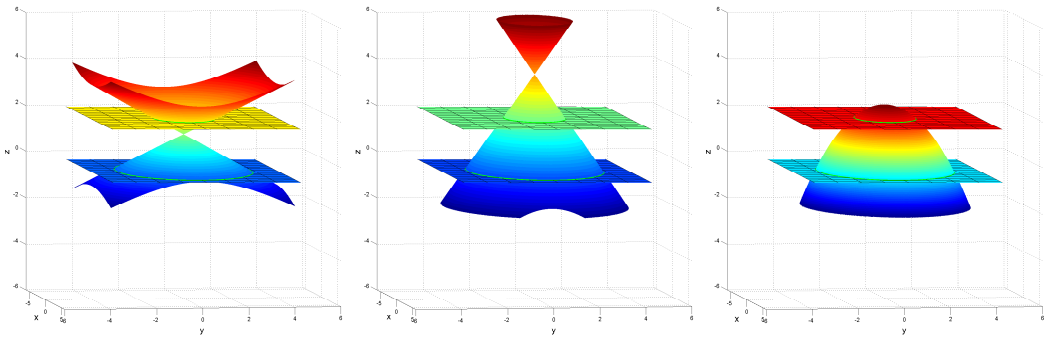
$$f'(-1) = -\frac{(\alpha - \beta)^2}{2} + 1 - \alpha\beta = 1 - \frac{\alpha^2 + \beta^2}{2} \geq 0,$$

where the inequality follows from the assumption that $\mathcal{Q} \cap \mathcal{A}^{=} \neq \emptyset$ and $\mathcal{Q} \cap \mathcal{B}^{=} \neq \emptyset$. This shows that both $\bar{\tau}_1 < -1$ and $\bar{\tau}_2 < -1$. As a result, $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_1)$ are both cones.

Summary of shapes According to the values of the discriminant (3.21) we can classify the shapes of $\mathcal{Q}(\tau)$ at the roots of f. Recall that $\tau \neq -1$, $\bar{\tau}_1 \neq -1$, and $\bar{\tau}_2 \neq -1$. We may further assume w.l.o.g. that $\bar{\tau}_1 \leq \bar{\tau}_2$. We have the following cases:

- If the discriminant (3.21) is not equal to zero, then $-1 > \bar{\tau}_2 > \bar{\tau}_1$, and there are two different cones at $\tau = \bar{\tau}_1$ and $\tau = \bar{\tau}_2$ in the family $\mathcal{Q}(\tau)$. For illustrations see Figure

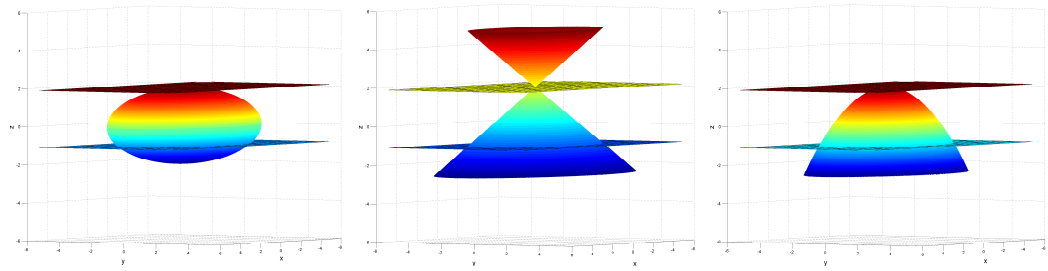
3.1.



(a) $\tau = \bar{\tau}_1$ (b) $\tau = \bar{\tau}_2$ (c) $\tau = -1$

Figure 3.1: $f(\tau)$ has two distinct roots which do not coincide with -1.

- If the discriminant(3.21) is equal to zero, then $-1 > \bar{\tau}_2 = \bar{\tau}_1$, and there is a unique cone in the family $\mathcal{Q}(\tau)$ at $\tau = \bar{\tau}_1 = \bar{\tau}_2$. Observe that in this case, if either $\alpha^2 = 1$ or $\beta^2 = 1$, then one of the hyperplanes is tangent to the ellipsoid. See Figure 3.2.



(a) One hyperplane is tangent to the ellipsoid. (b) $\tau = \bar{\tau}_1 = \bar{\tau}_2$ (c) $\tau = -1$

Figure 3.2: The two roots of $f(\tau)$ coincide, but are different from $\hat{\tau}$.

3.2.3.2 $P(\tau)$ is singular

It follows from Lemma 3.1 that $P(\tau)$ is singular when $\tau = -1$. In this case we have that $P(-1) \succeq 0$ with one zero eigenvalue. Thus, from Section 1.1.2.1 we have that $\mathcal{Q}(-1)$ is

either a line, a cylinder, or a paraboloid. The shape of $\mathcal{Q}(-1)$ can be decided by verifying if $p(-1)$ is in the range of $P(-1)$. Particularly, we have that $P(-1)a = (I - aa^\top)a = a - a = 0$, then a is an eigenvector of $P(-1)$ associated to the zero eigenvalue of $P(-1)$. Then, $p(-1)$ is in the range of $P(-1)$ if $p(-1)^\top a = 0$. We have

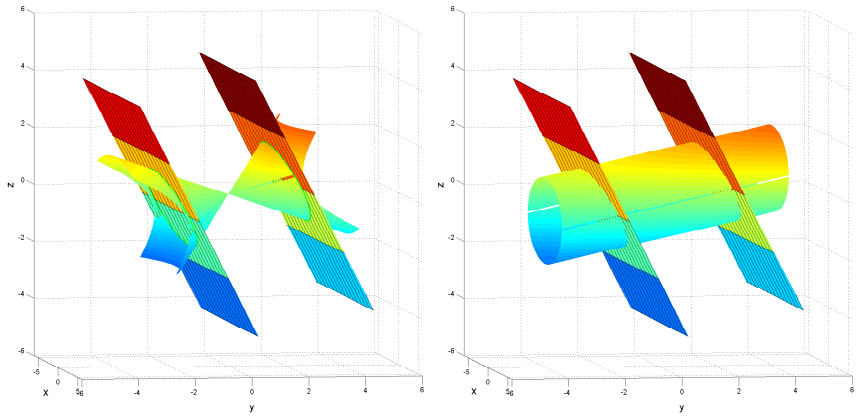
$$p(-1)^\top a = \left(\frac{\alpha+\beta}{2}a\right)^\top a = \frac{\alpha+\beta}{2}. \tag{3.23}$$

Hence, $p(-1)^\top a$ is zero if and only if $\alpha = -\beta$, i.e., the two hyperplanes $\mathcal{A}^=$ and $\mathcal{B}^=$ are symmetric about the center of the hypersphere $\mathcal{Q}$. Therefore, if $\alpha = -\beta$ any vector $x^c = \eta a$, for all $\eta \in \mathbb{R}$, satisfies the condition $P(-1)x^c = p(-1)$ of **Case 1** in Section 1.1.2.1. On the other hand, if $\alpha \neq -\beta$, then $p(-1)$ is not orthogonal to a, and there is no x^c such that $P(-1)x^c = -p(-1)$. Recall that this is true because a is an eigenvector corresponding to the zero eigenvalue of $P(-1)$. Then, from **Case 2** in Section 1.1.2.1 we conclude that $\mathcal{Q}(-1)$ is a paraboloid. For illustrations, see Figures 3.1(c) and 3.2(c).

Summary of shapes According to equation (3.23) and the values of the discriminant (3.21) we can classify the shapes of $\mathcal{Q}(\tau)$ at -1, $\bar{\tau}_1$, $\bar{\tau}_2$ when $p(-1)^\top a = 0$. We may assume w.l.o.g. that $\bar{\tau}_1 \leq \bar{\tau}_2$. We may have the following cases:

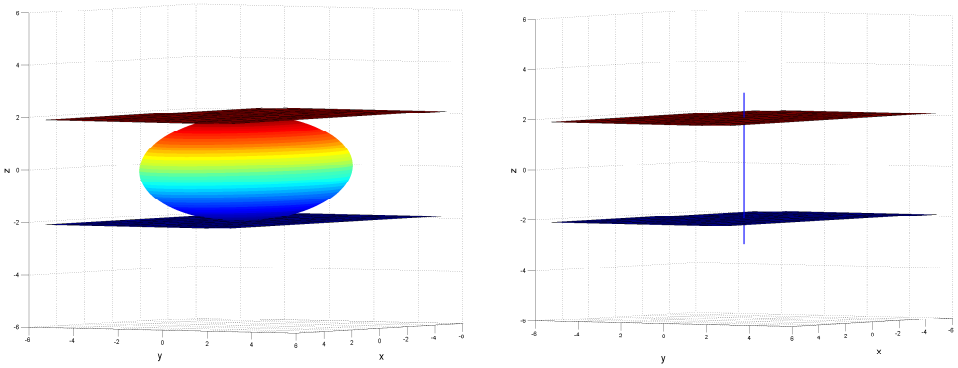
- If the discriminant (3.21) is not equal to zero and $-1 = \bar{\tau}_2 > \bar{\tau}_1$, then for the vector $x^c = a$ we obtain from (3.23) that $(x^c)^\top P(-1)x^c - \rho(-1) = (1 - \alpha^2) > 0$, and from **Case 1** in Section 1.1.2.1 we have that $\mathcal{Q}(-1)$ is a cylinder. Additionally, $\mathcal{Q}(\bar{\tau}_1)$ is a cone. For illustrations see Figure 3.3.

- If the discriminant (3.21) is zero and $-1 = \bar{\tau}_2 = \bar{\tau}_1$, then for the vector $x^c = a$ from (3.23) we obtain that $(x^c)^\top P(-1)x^c - \rho(-1) = (1 - \alpha^2) = 0$, and from **Case 1** in §1.1.2.1 we have that $\mathcal{Q}(-1)$ is a line. For illustrations see Figure 3.4.



(a) $\tau = \bar{\tau}_1$ (b) $\tau = -1$

Figure 3.3: $f(\tau)$ has two distinct roots, but the larger root coincides with -1.



(a) The two hyperplanes are tangent to the ellipsoid. (b) $\tau = -1$

Figure 3.4: The two roots of $f(\tau)$ coincide with -1.

3.2.3.3 Summarizing the shapes of $\mathcal{Q}(\tau)$

We can summarize the shapes of the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ using $\bar{\tau}_1$, and $\bar{\tau}_2$ in the following theorem. We assume w.l.o.g. that $\bar{\tau}_1 \leq \bar{\tau}_2$.

Theorem 3.2. *The following cases may occur for the shape of $\mathcal{Q}(\tau)$:*

- $\bar{\tau}_1 < \bar{\tau}_2 < -1$*: $\mathcal{Q}(-1)$ is a paraboloid, and $\mathcal{Q}(\bar{\tau}_1)$, $\mathcal{Q}(\bar{\tau}_2)$ are two cones.*
- $\bar{\tau}_1 = \bar{\tau}_2 < -1$*: $\mathcal{Q}(-1)$ is a paraboloid and $\mathcal{Q}(\bar{\tau}_1)$ is a cone.*
- $\bar{\tau}_1 < \bar{\tau}_1 = -1$*: $\mathcal{Q}(-1)$ is a cylinder and $\mathcal{Q}(\bar{\tau}_1)$ is cone.*
- $\bar{\tau}_1 = \bar{\tau}_2 = -1$*: $\mathcal{Q}(-1)$ is a line.*

This completes the description of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of quadrics when $P \succ 0$ and $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ are parallel.

3.2.4 Classification of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ when P is singular

Recall our assumption that the matrix $P \in \mathbb{R}^{\ell \times \ell}$ is symmetric and has at most one non-positive eigenvalue. Here we focus on the case when $P \succeq 0$ and has one zero eigenvalue, i.e., we assume that the quadric $\mathcal{Q}$ is either a paraboloid or a cylinder. Recall the affine transformation defined in Section 3.1.2. To simplify the algebra we may assume w.l.o.g. that $\mathcal{Q}$ is given as a quadric in the form (3.12), where $P_{1,1} = 0$. Note that in the discussion of Section 3.1.2 we considered the general case where $P_{j,j} = 0$ for some $j \in \{1, \ldots, n\}$. For the sake of simplifying the discussion, here we assume that $j = 1$. Thus, we have that $P = \bar{J}$, and the representation of $\mathcal{Q}(\tau)$ in this section is defined by

$$P(\tau) = \bar{J} + \tau a a^{\top}, \quad p(\tau) = p - \tau \frac{\alpha + \beta}{2} a, \quad \rho(\tau) = \rho + \tau \alpha \beta. \tag{3.24}$$

Recall that $\|a\| = 1$, $\rho \in \{-1, 0, 1\}$, and $\bar{J} \in \mathbb{R}^{n \times n}$ is defined as

$$\bar{J} = \begin{bmatrix} 0 & 0^\top \\ 0 & I \end{bmatrix}.$$

We now characterize the behavior of the family $\{\mathcal{Q}(\tau) | \tau \in \mathbb{R}\}$ defined by (3.24) as a function of parameter τ. For this characterization we need first to discuss the inertia of $P(\tau)$ in Section 3.2.4.1. Then we divide the analysis of the classification in three cases. In Section 3.2.4.2 we discuss the case when a is such that $a_1 \neq 0$. In Section 3.2.4.3 we discuss the case when there exist a vector $x^c \in \mathbb{R}^\ell$ solving the system $Px^c = -p$. Finally, in Section 3.2.4.4 we discuss the case when there is no vector x^c solving the system $Px^c = -p$.

3.2.4.1 The eigenvalues of $P(\tau)$

In this case the characteristic polynomial (3.16) simplifies to

$$(1-\lambda)^{n-2}(\lambda^2 - \lambda(1 + \tau\|a\|^2) + \tau a_1^2) = (1-\lambda)^{n-2}(\lambda^2 - \lambda(1+\tau) + \tau a_1^2) = 0.$$

Thus, 1 is an eigenvalue of P with multiplicity $n-2$. The other two eigenvalues are given by the roots of $\lambda^2 - \lambda(1+\tau) + \tau a_1^2 = 0$, which are

$$\frac{(1+\tau) \pm \sqrt{(1+\tau)^2 - 4\tau a_1^2}}{2}. \tag{3.25}$$

The inertia of $P + \tau a a^\top$ in this case is defined by the signs of the two roots in (3.25). Since $0 \leq |a_1| \leq 1$, we have that $(1+\tau)^2 - 4\tau a_1^2 \geq 0$, and both eigenvalues are reals.

3.2.4.2 Classification when $a_1 \neq 0$

The behavior of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case can be characterized analogous to the analysis developed in Section 3.2.3. First, consider the following results about the

subfamily $\{\mathcal{Q}(\tau) \mid \tau > 0\}$.

Lemma 3.2. *If $a_1 \neq 0$, then for any $\tau > 0$ the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ represented by (3.24) are ellipsoids.*

Proof. First, we have that $1 + \tau > 0$ and $4\tau a_1^2 > 0$ for $\tau > 0$. Hence, the two eigenvalues of the matrix $P(\tau)$ in (3.24) given by (3.25) are positive. As a consequence, any quadric $\mathcal{Q}(\tau)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ represented by (3.24) is an ellipsoid for $\tau > 0$. □

Thus, if $a_1 \neq 0$, then Lemma 3.2 proves that the quadrics in the subfamily $\{\mathcal{Q}(\tau) \mid \tau > 0\}$ are ellipsoids. Now, since the domain of τ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is the whole real line, we can analyze its behavior using any of the quadrics in the family. Particularly, for the characterization of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case, we can use the quadric $\mathcal{Q}(1)$. Since this quadric is an ellipsoid, this family is characterized already by the analysis developed in Section 3.2.3. Then, Theorem 3.2 summarizes the possible shapes for the family in this case.

3.2.4.3 Classification when $Px = -p$ is solvable and $a_1 = 0$

Here we characterize the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ defined by (3.24) when $a_1 = 0$ and there exists a vector $w^c \in \mathbb{R}^\ell$ solving the system $Pw^c = -p$. The behavior of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case is strongly related to the analysis developed in Section 3.2.3. First, we need to know the inertia of the matrix $P(\tau)$.

Lemma 3.3. *If $a_1 = 0$, then one of the eigenvalues of the matrix $P(\tau)$ is always zero for all the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$.*

Proof. From (3.25), if $a_1 = 0$, then we have that for $\tau \in \mathbb{R}$ the two eigenvalues given by this expression are 0 and $1 + \tau$. □

Lemma 3.3 tells us that the characterization of the behavior of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is determined by the eigenvalue $1 + \tau$ of $P(\tau)$. This eigenvalue will be positive for

$\tau > -1$, zero for $\tau = -1$, and negative for $\tau < -1$. Now, we have that a solution w^c to the system $Px = -p$ exists only if $p_1 = 0$. Additionally, observe that for $\tau \in \mathbb{R}$ the first row and column of $P(\tau)$ are always zero vectors. Hence, we have that any quadric $\mathcal{Q}(\tau)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is a cylinder in the direction of $(1, 0^\top)$, where $0 \in \mathbb{R}^{\ell-1}$ is the all zeros vector. In other words, a quadric $\mathcal{Q}(\tau)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is given by the set

$$\bigcup_{\delta \in \mathbb{R}} \Big\{ \{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\} \cap \{x \in \mathbb{R}^\ell \mid (1, 0^\top)x = \delta\} \Big\}$$

.

Thus, to classify the shape of these cylinders for $\tau \in \mathbb{R}$, it is enough to analyze the shapes of a base of these cylinders on the hyperplane $\{w \in \mathbb{R}^\ell \mid (1, 0^\top)w = 0\}$. A base of any cylinder in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this hyperplane is a quadric in $\mathbb{R}^{n-1}$, represented by

$$\tilde{P}(\tau) = I + \tau a_{2:n} a_{2:n}^\top, \quad \tilde{p}(\tau) = p_{2:n} - \tau \frac{\alpha + \beta}{2} a_{2:n}, \quad \rho(\tau) = \rho + \tau\alpha\beta.$$

Note that these sets are fully analyzed in Section 3.2.3. Therefore, we know that a base of a cylinder in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ would be an ellipsoid if $\tau > -1$, a hyperboloid or a cone if $\tau \leq -1$, and a paraboloid or a cylinder if $\tau = -1$.

3.2.4.4 Classification when $Px = -p$ is not solvable and $a_1 = 0$

Here we characterize the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ define by (3.24) when $a_1 = 0$ and there is no vector $x^c \in \mathbb{R}^\ell$ such that $Px^c = -p$. We show that in this case there is a parabolic cylinder in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$.

First, we can easily characterize the behavior of the family for $\tau \neq -1$. Note that the system $Px = -p$ is not solvable only if $p_1 \neq 0$. Then, since $a_1 = 0$ we have that $p(\tau)_1 \neq 0$. On the other hand, from Lemma 3.3 we know that one eigenvalue of $P(\tau)$ is

always zero. Then, since 1 is an eigenvalue of $P(\tau)$ with multiplicity $\ell - 2$, the behavior of the family is determined by the eigenvalue $1 + \tau$. Thus, the quadrics in the subfamily $\{\mathcal{Q}(\tau) \mid \tau > -1\}$ are paraboloids, and the quadrics in the subfamily $\{\mathcal{Q}(\tau) \mid \tau < -1\}$ are hyperbolic paraboloids.

To complete the characterization we need to analyze the quadric $\mathcal{Q}(-1)$, which is classified in Lemma 3.4. Figure 3.5 illustrates this result.

Lemma 3.4. *If $\mathcal{Q}$ is a paraboloid and $a_1 = 0$, then the quadric $\mathcal{Q}(-1)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ defined by (3.24) is a parabolic cylinder.*

Proof. First of all, we have that zero is an eigenvalue of the matrix $P(\tau)$, with multiplicity 2. We perform the proof in three steps. First, we find a basis for the null space of $P(-1)$. Then, we find a direction in that space that is orthogonal to $p(-1)$. Finally, we show that $\mathcal{F}(\tau)$ is a cylinder in that direction.

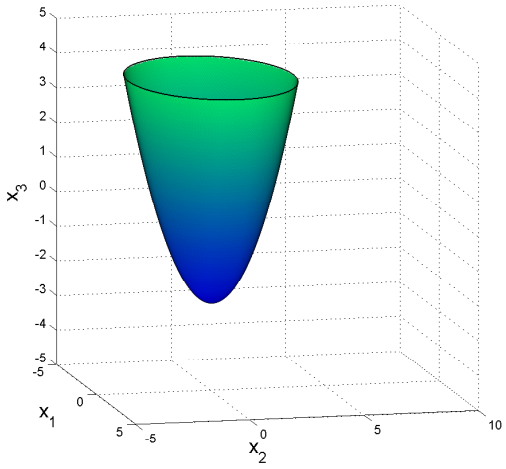
Recall that for $\tau \in \mathbb{R}$ the first row and column of $P(-1)$ are zero vectors. Since $\|a\| = 1$, and $a_1 = 0$, we have that

$$P(-1)a = \left(\bar{J} - \tau a a^\top\right) a = a - a = 0.$$

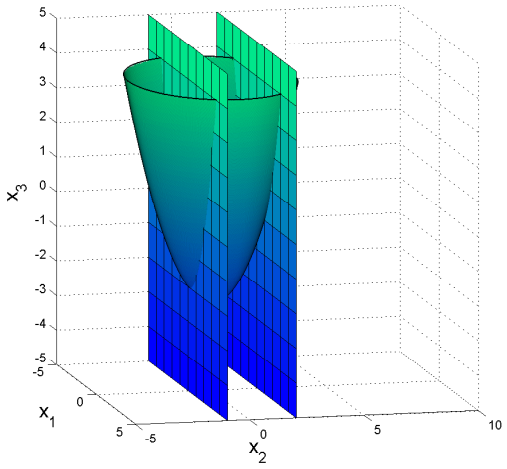
Thus, a and $(1, 0^\top)$ are eigenvectors of $P(-1)$ associated with the 0 eigenvalue, and form a basis for the null space of $P(-1)$. Hence, any vector of the form $(\gamma, a_{2:n}^\top)$, $\gamma \in \mathbb{R}$, belongs to the null space of $P(-1)$.

Define $\tilde{\gamma}$ as

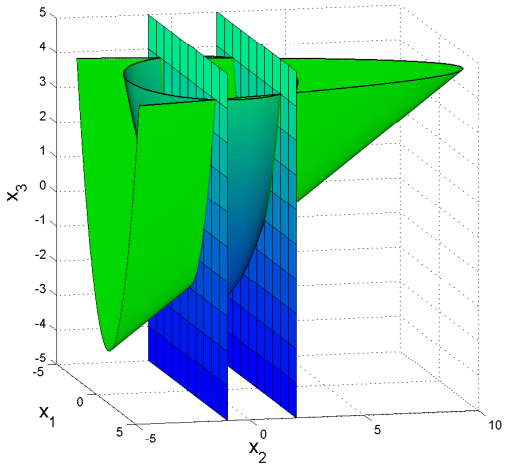
$$\tilde{\gamma} = \frac{-p_{2:n}^\top a_{2:n} - \frac{\alpha+\beta}{2}}{p_1}.$$



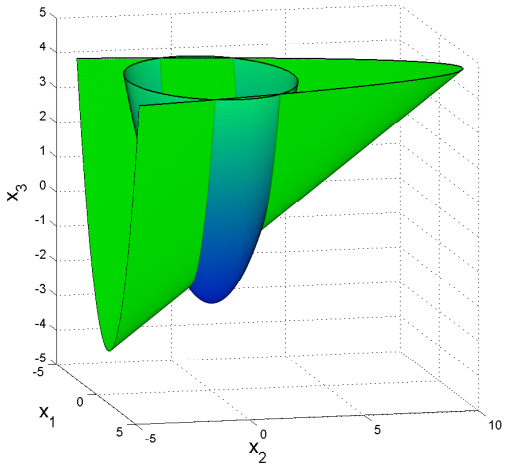
(a) Set $\mathcal{Q}$



(b) Set $\mathcal{Q}$ intersected with two parallel hyperplanes



(c) Quadric $\mathcal{Q}(-1)$, sharing the same intersection with the hyperplanes that $\mathcal{Q}$ has



(d) Set $\mathcal{Q}$ intersected with the quadric $\mathcal{Q}(-1)$

Figure 3.5: Illustration of Lemma 3.4.

The vector $(\tilde{\gamma}, a_{2:n}^\top)$ is orthogonal to $p(-1)$, since

$$p(-1)^\top \begin{bmatrix} \tilde{\gamma} \\ a_{2:n} \end{bmatrix} = \left(p^\top + \frac{\alpha+\beta}{2}a\right) \begin{bmatrix} \frac{-p_{2:n}^\top a_{2:n} - \frac{\alpha+\beta}{2}}{p_1} \\ a_{2:n} \end{bmatrix}$$
$$= -p_{2:n}^\top a_{2:n} - \frac{\alpha+\beta}{2} + \frac{\alpha+\beta}{2} + p_{2:n}^\top a_{2:n} = 0.$$

Let $\tilde{w} \in \mathbb{R}^\ell$ be a vector such that $\tilde{w} \in \mathcal{Q}(-1) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$, then we have that

$$\tilde{w}^\top P(-1)\tilde{w} + 2p(-1)^\top \tilde{w} + \rho(-1) \leq 0.$$

Now, let $\tilde{u}^\top = \tilde{w}^\top + \theta(\tilde{\gamma}, a_{2:n}^\top)$ for some $\theta \in \mathbb{R}$, then we have that

$$\begin{aligned}
&\tilde{u}^\top P(-1)\tilde{u} + 2p(-1)^\top \tilde{u} + \rho(-1) \\
&\quad = \tilde{w}^\top P(-1)\tilde{w} + \theta(\tilde{\gamma}, a_{2:n}^\top)P(-1)\tilde{w} + \theta^2(\tilde{\gamma}, a_{2:n}^\top)P(-1)\begin{bmatrix} \tilde{\gamma} \\ a_{2:n} \end{bmatrix} + 2p(-1)^\top \tilde{u} + \rho(-1) \\
&\quad = \tilde{w}^\top P(-1)\tilde{w} + 2p(-1)^\top \tilde{w}^\top + 2\theta p(-1)^\top \begin{bmatrix} \tilde{\gamma} \\ a_{2:n} \end{bmatrix} + \rho(-1) \\
&\quad = \tilde{w}^\top P(-1)\tilde{w} + 2p(-1)^\top \tilde{w}^\top + \rho(-1) \leq 0,
\end{aligned}$$

where the last inequality follows from the assumption $\tilde{x} \in \mathcal{Q}(-1) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$. Hence, the distance of a vector $\tilde{z}^\top$ to the boundary of $\mathcal{Q}(-1)$ is constant for any $\theta \in \mathbb{R}$. Therefore, $\mathcal{Q}(-1)$ is a parabolic cylinder in the direction $(\tilde{\gamma}, a_{2:n}^\top)^\top$. □

3.2.5 Classification of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ when P is indefinite

Here we focus on the case when P is an indefinite matrix with one negative eigenvalue and $\ell - 1$ positive eigenvalues. Recall the affine transformation defined in Section 3.1.1. Hence,

to simplify the algebra we may assume w.l.o.g. that $\mathcal{Q}$ is a quadric of the form (3.5), where $P_{1,1} = -1$. Note that in the discussion of Section 3.1.1 we considered the general case where $P_{j,j} = -1$ for some $j \in \{1, \ldots, n\}$. For the sake of simplifying the discussion, here we assume that $j = 1$. Thus, we have that $P = \tilde{J}$, $p = 0$, and the representation of the quadrics $\mathcal{Q}(\tau)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this section is defined by

$$P(\tau) = \tilde{J} + \tau a a^\top, \quad p(\tau) = -\tau \frac{\alpha + \beta}{2} a, \quad \rho(\tau) = \rho + \tau \alpha \beta. \tag{3.26}$$

Recall that $\|a\| = 1$, $\rho \in \{-1, 0, 1\}$, and $\tilde{J} \in \mathbb{R}^{n \times n}$ is defined as

$$\tilde{J} = \begin{bmatrix} -1 & \mathbf{0}^\top \\ \mathbf{0} & I \end{bmatrix}.$$

We characterize the behavior of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ defined by (3.26) as a function of the parameter τ. For this characterization we discuss in Section 3.2.5.1 the inertia of $P(\tau)$. Based on the eigenvalues of $P(\tau)$, we divide the analysis in two cases: 1) $a_1 > 1/2$; and 2) $a_1 \leq 1/2$. In Section 3.2.5.2 we discuss the case when $a_1 > 1/2$. To analyze the case when $a_1 \leq 1/2$ we discuss in Section 3.2.5.3 the function $g(\tau) = p(\tau)^\top P(\tau)^{-1} p(\tau) - \rho(\tau)$. Based on the analysis of $f(\tau)$ we identify four cases that need to be considered in order to completely analyze the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ when $P(\tau)$ is indefinite. First, in Section 3.2.5.4 we consider the case when $a_1 = 1/2$. Second, in Section 3.2.5.5 we analyze the case when $a_1 < 1/2$ and $\rho = 0$, i.e., when $\mathcal{Q}$ is a cone. Third, in Section 3.2.5.6 we characterize the case when $a_1 < 1/2$ and $\rho = 1$. Finally, in Section 3.2.5.7 we complete the characterization of the family with the case when $a_1 < 1/2$ and $\rho = -1$.

3.2.5.1 The eigenvalues of $P(\tau)$

Here we provide a closed formula to compute the eigenvalues and the inertia of $P(\tau)$ when P is indefinite. Recall that $P_{1,1} = -1$, then in this case the characteristic polynomial (3.16) simplifies to

$$(1-\lambda)^{\ell-2}\left(\lambda^2 - \lambda\tau\left\|a\right\|^2 + (\tau a_1^2 - \tau\sum_{i=2}^{\ell} a_1^2 - 1)\right) = (1-\lambda)^{\ell-2}(\lambda^2 - \lambda\tau + (2\tau a_1^2 - \tau - 1)) = 0.$$

Thus, 1 is an eigenvalue of P with multiplicity $\ell - 2$. The other two eigenvalues are given by the roots of $\lambda^2 - \lambda\tau + (2\tau a_1^2 - \tau - 1) = 0$, which are

$$\frac{\tau \pm \sqrt{\tau^2 + 4 + 4\tau(1 - 2a_1^2)}}{2}. \tag{3.27}$$

The inertia of $P + \tau aa^\top$ in this case is defined by the signs of the two roots given by (3.27). Since $0 \leq |a_1| \leq 1$, we have that $(2+\tau)^2 - 8\tau a_1^2 \geq 0$, thus both of those eigenvalues are real. We need to consider three cases:

1. If $a_1^2 > \frac{1}{2}$, $(1 - 2a_1^2) < 0$. Thus, we have:
   - if $\tau > -\frac{1}{(1-2a_1^2)}$, then both eigenvalues are positive,
   - if $\tau = -\frac{1}{(1-2a_1^2)}$, then there is a zero and a negative eigenvalue,
   - if $\tau < -\frac{1}{(1-2a_1^2)}$, then there is a positive and a negative eigenvalue.
2. If $a_1^2 = \frac{1}{2}$, then the eigenvalues are

$$\frac{\tau \pm \sqrt{\tau^2 + 4}}{2}. \tag{3.28}$$

   In this case we have a positive eigenvalue and a negative eigenvalue.

3. If $a_1^2 < \frac{1}{2}$, then $(1 - 2a_1^2) > 0$. Thus, we have:

- if $\tau > -\frac{1}{(1-2a_1^2)}$, then there is a positive and a negative eigenvalue,
- if $\tau = -\frac{1}{(1-2a_1^2)}$, then there is a zero and a negative eigenvalue,
- if $\tau < -\frac{1}{(1-2a_1^2)}$, then both eigenvalues are negative.

3.2.5.2 Classification when $a_1^2 > \frac{1}{2}$

The behavior of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case can be characterized by utilizing the analysis developed in Section 3.2.3. Recall from Section 3.2.5.1 that the eigenvalues of $P(\tau)$ are 1 with multiplicity $\ell - 2$ and the two values in (3.27). We have the following result.

Lemma 3.5. *If $a_1 > \frac{1}{2}$, then the quadrics $\{\mathcal{Q}(\tau) \mid \tau > -\frac{1}{(1-2a_1^2)}\}$ are ellipsoids, where $\mathcal{Q}(\tau)$ is defined by (3.26).*

Proof. From Section 3.2.5.1, we know that if $a_1^2 > \frac{1}{2}$, then the values in (3.27) are positive for $\tau > -\frac{1}{(1-2a_1^2)}$. Thus, if $a_1^2 > \frac{1}{2}$, then Lemma 3.5 shows that the quadrics in the subfamily $\{\mathcal{Q}(\tau) \mid \tau > -\frac{1}{(1-2a_1^2)}\}$ are ellipsoids. □

Observe that, since the domain of τ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is the whole real line, we can analyze its behavior using any of the quadrics in the family. Then, for the characterization of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case, we can use the quadric $\mathcal{Q}(1 - \frac{1}{(1-2a_1^2)})$. Since this quadric is an ellipsoid, the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case is characterized already by the analysis developed in Section 3.2.3. In particular, Theorem 3.2 summarizes the possible shapes for the family in this case.

3.2.5.3 Analysis of the function $g(\tau) = p(\tau)^\top P(\tau)^{-1} p(\tau) - \rho(\tau)$

Before discussing the classification of the quadrics when $a_1^2 \leq \frac{1}{2}$, we need to analyze the function

$$g(\tau) = p(\tau)^\top P(\tau)^{-1} p(\tau) - \rho(\tau) = \left(-\tau \frac{\alpha+\beta}{2} a\right)^\top \left(J + \tau a a^\top\right)^{-1} \left(-\tau \frac{\alpha+\beta}{2} a\right) - (\rho + \tau\alpha\beta).$$

For the classification of the cases when $a_1^2 \leq \frac{1}{2}$ we use similar steps to those in the analysis in Section 3.2.3. First, we show in this section that the roots of the function $g(\tau)$ coincide with the roots of a quadratic polynomial of τ.

Using the Sherman-Morrison-Woodbury formula Golub and Van Loan [1996] to compute the inverse of $P(\tau)$ we obtain:

$$P(\tau)^{-1} = \left(\tilde{J} + \tau a a^\top\right)^{-1} = \tilde{J} - \frac{\tau \begin{bmatrix} -a_1 \\ a_{2:n} \end{bmatrix} \begin{bmatrix} -a_1 & a_{2:n} \end{bmatrix}}{1 + \tau(\|a_{2:n}\|^2 - a_1^2)}. \tag{3.29}$$

Note that for $\tau = -\frac{1}{(1-2a_1^2)} = -\frac{1}{(\|a_{2:n}\|^2 - a_1^2)}$ the inverse does not exist, as was expected from case 3 in Section 3.2.5.1. On the other hand, when $a_1^2 = \frac{1}{2}$ the inverse of $P(\tau)$ always exists. Now, for computing the polynomial in τ we assume that $P(\tau)$ is non-singular, i.e., that $\tau \neq -\frac{1}{(1-2a_1^2)}$. The cases when $\tau = -\frac{1}{(1-2a_1^2)}$ will be consider explicitly Sections 3.2.5.4, 3.2.5.5, 3.2.5.6, and 3.2.5.7.

When we substitute $P(\tau)^{-1}$, $p(\tau)$, and $\rho(\tau)$ in $g(\tau)$ using the expressions (3.26) and

(3.29), we obtain

$$\begin{aligned}
g(\tau) &= \left(-\tau\frac{\alpha+\beta}{2}a\right)^\top \left(J+\tau aa^\top\right)^{-1}\left(-\tau\frac{\alpha+\beta}{2}a\right) - (\rho+\tau\alpha\beta) \\
&= \frac{\tau^2(\alpha+\beta)^2}{4}a^\top\left(\tilde{J} - \frac{\tau\begin{bmatrix}-a_1\\ a_{2:n}\end{bmatrix}\begin{bmatrix}-a_1 & a_{2:n}\end{bmatrix}}{1+\tau(1-2a_1^2)}\right)a - (\rho+\tau\alpha\beta) \\
&= \frac{\tau^2(\alpha+\beta)^2}{4}\left((1-2a_1^2) - \frac{\tau(1-2a_1^2)^2}{1+\tau(1-2a_1^2)}\right) - (\rho+\tau\alpha\beta) \\
&= \frac{\tau^2(\alpha+\beta)^2}{4}\left(\frac{1-2a_1^2}{1+\tau(1-2a_1^2)}\right) - (\rho+\tau\alpha\beta) \\
&= \frac{\tau^2(1-2a_1^2)((\alpha+\beta)^2-4\alpha\beta) - 4\tau(\rho(1-2a_1^2)+\alpha\beta) - 4\rho}{4(1+\tau(1-2a_1^2))} \\
&= \frac{\tau^2(1-2a_1^2)\frac{(\alpha-\beta)^2}{4} - \tau(\rho(1-2a_1^2)+\alpha\beta) - \rho}{1+\tau(1-2a_1^2)}.
\end{aligned} \tag{3.30}$$

Obviously $g(\tau) = 0$ when the numerator of (3.30) is zero. Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined by

$$f(\tau) = \tau^2(1-2a_1^2)\frac{(\alpha-\beta)^2}{4} - \tau(\rho(1-2a_1^2)+\alpha\beta) - \rho \tag{3.31}$$

which is a quadratic function of τ. Recall that $\alpha \neq \beta$. If $a_1^2 = 1/2$, then f is a linear function of τ. Let us assume now that $a_1^2 < 1/2$ and let $\bar{\tau}_1$ and $\bar{\tau}_2$ be the two roots of f.

We can now check the value of $f(\hat{\tau})$, and the value of the derivative $f'(\hat{\tau})$, which are used to compare the two roots of f with the critical value $\hat{\tau} = -1/(1-2a_1^2)$. First, we have

$$f(\hat{\tau}) = \frac{(\alpha-\beta)^2}{4(1-2a_1^2)} - \rho + \frac{\alpha\beta}{(1-2a_1^2)} + \rho = \frac{(\alpha-\beta)^2+4\alpha\beta}{4(1-2a_1^2)} = \frac{(\alpha+\beta)^2}{4(1-2a_1^2)} \geq 0, \tag{3.32}$$

and since the coefficient of τ^2 is positive, this implies that $\hat{\tau}$ is not between the two roots

of f. For the derivative we have

$$f'(\hat{\tau}) = -\frac{(\alpha-\beta)^2}{2} - (\rho(1-2a_1^2)+\alpha\beta), \tag{3.33}$$

which depends on the scalar ρ.

Finally, the discriminant of f is:

$$\begin{aligned}
&(\rho(1-2a_1^2)+\alpha\beta)^2 + \rho(1-2a_1^2)(\alpha-\beta)^2 \\
&= \rho^2(1-2a_1^2)^2 + \alpha^2\beta^2 + \rho(1-2a_1^2)\alpha^2 + \rho(1-2a_1^2)\beta^2 \\
&= \rho(1-2a_1^2)(\rho(1-2a_1^2)+\beta^2) + \alpha^2(\rho(1-2a_1^2)+\beta^2) \\
&= (\alpha^2+\rho(1-2a_1^2))(\beta^2+\rho(1-2a_1^2)),
\end{aligned} \tag{3.34}$$

which also depends on the scalar ρ.

Based on the values of a_1^2 and ρ, and the discriminant (3.34) of f, we have to separate the analysis in the following cases:

- If $a_1^2 = 1/2$, then f is a linear function of τ, this case is analyzed in Section 3.2.5.4.
- If $a_1^2 < 1/2$ and $\rho = 0$, then the roots of f are real since its discriminant is positive, this case is analyzed in Section 3.2.5.5.
- If $a_1^2 < 1/2$ and $\rho = 1$, then the roots of f are real since its discriminant is positive, this case is analyzed in Section 3.2.5.6.
- If $a_1^2 < 1/2$ and $\rho = -1$, then f has real roots only if:
  - $\beta^2 \leq 1-2a_1^2$ and $\alpha^2 \leq 1-2a_1^2$, or
  - $\beta^2 \geq 1-2a_1^2$ and $\alpha^2 \geq 1-2a_1^2$.

  This case is analyzed in Section 3.2.5.7.

3.2.5.4 Classification when $a_1^2 = \frac{1}{2}$

In this case we can show that there is a single cone in the entire family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$. We show that this cone occurs when $\tau = -\frac{\rho}{\alpha\beta}$.

Recall that in this case the function f is linear an is given by

$$f(\tau) = -\alpha\beta\tau - \rho,$$

Moreover, if $a_1^2 = \frac{1}{2}$, then from Section 3.2.5.1 we know that $P(\tau)$ has $\ell - 1$ positive eigenvalues and a negative eigenvalue for $\tau \in R$. Now, if either $\alpha = 0$ or $\beta = 0$, then $f(\tau) = \rho$ for $\tau \in R$. Thus, if $\rho \neq 0$, and either $\alpha = 0$ or $\beta = 0$, then all the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ are hyperboloids. On the other hand, if $\rho = 0$, and either $\alpha = 0$ or $\beta = 0$, then all the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ are scaled second order cones. The following lemma characterizes the shape of the quadric found at the root of f when $a_1^2 = \frac{1}{2}$, $\alpha \neq 0$, and $\beta \neq 0$.

Lemma 3.6. *If $a_1^2 = \frac{1}{2}$, $\alpha \neq 0$, and $\beta \neq 0$, then for $\bar{\tau} = -\frac{\rho}{\alpha\beta}$ the quadric $\mathcal{Q}(\bar{\tau})$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is a cone.*

Proof. In this case we have from Section 3.2.5.1 that $P(\tau)$ is always an invertible matrix with one negative eigenvalue. On the other hand, we have from (3.31) that $p(\bar{\tau})P(\bar{\tau})^{-1}p(\bar{\tau}) - \rho(\bar{\tau}) = 0$ for $\bar{\tau} = -\rho/\alpha\beta$. Hence, from the classification in Table (3.31) we have that the quadric $\mathcal{Q}(\bar{\tau})$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is a cone. □

3.2.5.5 Classification when $a_1^2 < 1/2$ and $\rho = 0$

In this case we show first that one of the roots of $f(\tau)$ is always zero. Then, we prove that the matrix $P(\tau)$ has at most one negative eigenvalue when τ equals to the non-zero root of $f(\tau)$. Finally, we characterize the shape of the quadric at the non-zero root of f.

As discussed, the roots of $f(\tau)$ coincide with the roots of $g(\tau)$. In this case (3.31) simplifies to

$$f(\tau) = \tau^2(1-2a_1^2)\frac{(\alpha-\beta)^2}{4} - \tau\alpha\beta.$$

Recall that if $\alpha \neq \beta$, then the roots of $f(\tau)$ are

$$2\left(\frac{\alpha\beta \pm |\alpha\beta|}{(1-2a_1^2)(\alpha-\beta)^2}\right). \tag{3.35}$$

Hence, one root is zero and the other root, denoted by $\bar{\tau}$, can be positive or negative depending on the sign of the product $\alpha\beta$.

Lemma 3.7. *If $a_1^2 < \frac{1}{2}$ and $\rho = 0$, then the matrix $P(\tau)$ has at most one negative eigenvalue at the non-zero root of f.*

Proof. We first show that

$$\bar{\tau} \geq -\frac{1}{(1-2a_1^2)}.$$

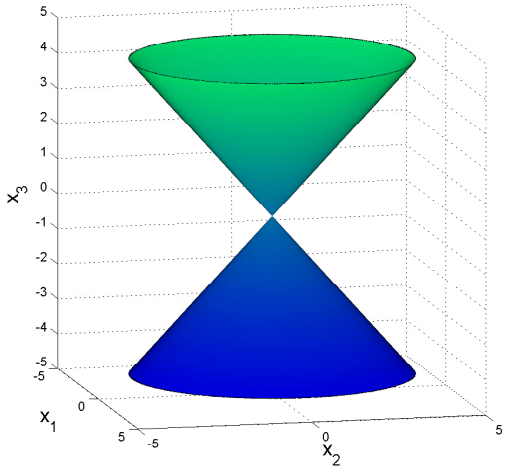
The most negative value $\bar{\tau}$ can take is achieved when $\alpha\beta < 0$. We have

$$\frac{-4\,|\alpha\beta|}{(1-2a_1^2)(\alpha-\beta)^2} = \left(\frac{-1}{(1-2a_1^2)}\right)\left(\frac{4\,|\alpha\beta|}{(\alpha-\beta)^2}\right) \geq \frac{-1}{(1-2a_1^2)}. \tag{3.36}$$

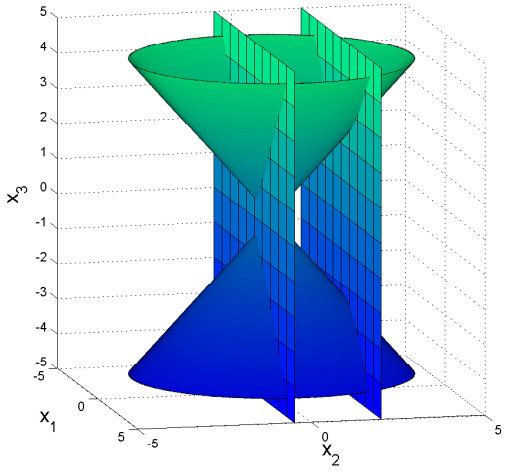
The last inequality follows because if $\alpha\beta < 0$, then $\alpha^2 - 2\alpha\beta + \beta^2 \geq 4\,|\alpha\beta|$ since $\alpha^2 + \beta^2 \geq 2\,|\alpha\beta|$.

Now, from Section 3.2.5.1 we know that $P(\bar{\tau})$ has one negative eigenvalue and $n-1$ positive eigenvalues if the inequality (3.36) is strict. If (3.36) is satisfied with equality, then $P(\bar{\tau})$ has one negative eigenvalue, one zero eigenvalue, and $n-2$ positive eigenvalues. □

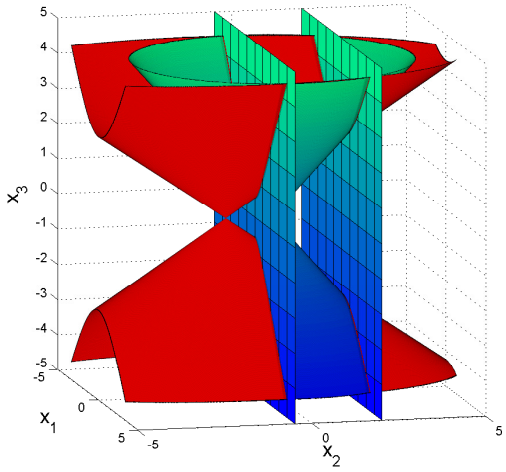
Now we can characterize the shapes of $\mathcal{Q}(\bar{\tau})$. Figure 3.6 illustrates the result in Theorem 3.3.



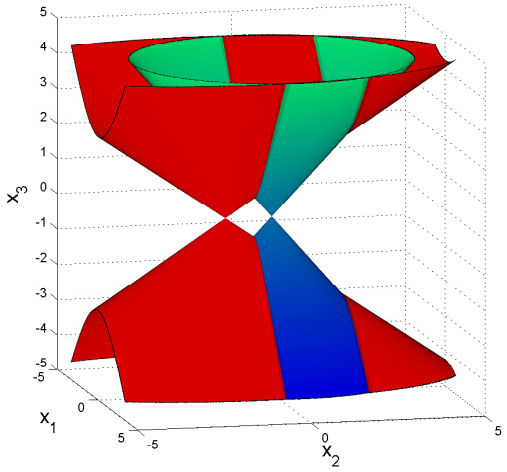
(a) The set $\mathcal{Q}$.



(b) Set $\mathcal{Q}$ intersected with two parallel hyperplanes.



(c) Quadric $\mathcal{Q}(\bar{\tau})$, sharing the same intersection with the hyperplanes that $\mathcal{Q}$ has.



(d) Set $\mathcal{Q}$ intersected with the quadric $\mathcal{Q}(\bar{\tau})$.

Figure 3.6: Illustration of Theorem 3.3.

Theorem 3.3. *If $a_1^2 < \frac{1}{2}$ and $\rho = 0$ in (3.26), then the shape of the quadric $\mathcal{Q}(\bar{\tau})$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ is:*

- *a cone if the $\bar{\tau} > -\frac{1}{(1-2a_1^2)}$,*
- *a hyperbolic cylinder of two sheets if $\bar{\tau} = -\frac{1}{(1-2a_1^2)}$.*

Proof. If $\bar{\tau} > -\frac{1}{(1-2a_1^2)}$, then from Lemma 3.7 and Section 3.2.5.1 we obtain that $\mathcal{Q}(\bar{\tau})$ is a cone. Now, we analyze the case when $\bar{\tau} = -\frac{1}{(1-2a_1^2)}$, which by (3.36) can happen only when $\beta = -\alpha$. In this case $P(\bar{\tau})$ is singular, $p(\bar{\tau}) = 0$, and $\rho(\bar{\tau}) > 0$. Recall that since $P(\bar{\tau})$ is symmetric, then there exist $D(\bar{\tau}) \in \mathbb{R}^{\ell\times\ell}$ and $V(\bar{\tau}) \in \mathbb{R}^{\ell\times\ell}$ such that $P(\bar{\tau}) = V(\bar{\tau})^\top D(\bar{\tau}) V(\bar{\tau})$.

Let us now characterize the shape of the quadric $\mathcal{Q}(\bar{\tau})$. First, recall that when $\bar{\tau} = -\frac{1}{(1-2a_1^2)}$ then $P(\bar{\tau})$ has one negative eigenvalue, one zero eigenvalue, and $\ell - 2$ positive eigenvalues. We may assume w.l.o.g. that $D_{1,1}(\bar{\tau}) < 0$, $D_{2,2}(\bar{\tau}) = 0$, and $D_{i,i}(\bar{\tau}) > 0$, $i \in \{3, \dots n\}$. Then $P(\bar{\tau}) = V(\bar{\tau})\hat{D}(\bar{\tau})^{\frac{1}{2}}\hat{J}\hat{D}(\bar{\tau})^{\frac{1}{2}}V(\bar{\tau})^\top$, where $\hat{D}(\bar{\tau})$ is a diagonal matrix with $\hat{D}_{i,i}(\bar{\tau}) = |D_{i,i}(\bar{\tau})|$, $i \in \{1, \dots n\} \setminus \{2\}$, and $\hat{D}_{2,2}(\bar{\tau}) = 1$. Additionally, $\hat{J}$ is a diagonal matrix defined as $\hat{J}_{1,1} = -1$, $\hat{J}_{2,2} = 0$, and $\hat{J}_{i,i} = 1$, $i \in \{3, \dots n\}$. Thus, using the transformation

$$u = \frac{\hat{D}(\tau)^{\frac{1}{2}}V(\tau)^\top w}{\sqrt{\rho(\bar{\tau})}}, \ \forall w \in \mathcal{Q}(\bar{\tau}),$$

we obtain that $\mathcal{Q}(\bar{\tau})$ is an affine transformation of the set

$$\{u \in \mathbb{R}^\ell \mid u^\top \hat{J} z u \leq -1\}, \tag{3.37}$$

which is a hyperbolic cylinder of two sheets. The right hand side of the quadratic equation in (3.37) is -1 because

$$\rho(\bar{\tau}) = -\frac{\beta\alpha}{(1-2a_1^2)} = \frac{\alpha^2}{(1-2a_1^2)} > 0.$$

Finally, given that $\hat{J}$ and $P(\bar{\tau})$ have the same inertia, we have shown that $\mathcal{Q}(\bar{\tau})$ is a hyperbolic cylinder of two sheets. □

3.2.5.6 Classification when $a_1^2 < 1/2$ and $\rho = 1$

In this case we begin again computing the two roots of the function f. Then, we compare these roots with the critical value $\hat{\tau} = -\frac{1}{(1-2a_1^2)}$. Based on this comparison, we close this section by classifying the shapes of the quadrics $\mathcal{Q}(\tau)$ at the two roots of f. In this case (3.31) simplifies to

$$f(\tau) = \tau^2(1-2a_1^2)\frac{(\alpha-\beta)^2}{4} - \tau(1-2a_1^2+\alpha\beta) - 1, \quad \forall \tau \in \mathbb{R}. \tag{3.38}$$

Recall that $\alpha \neq \beta$, and then the roots $\bar{\tau}_1$ and $\bar{\tau}_2$ of $f(\tau)$ are

$$\frac{2\left(1-2a_1^2+\alpha\beta \pm \sqrt{(1-2a_1^2+\alpha\beta)^2 + (1-2a_1^2)(\alpha-\beta)^2}\right)}{(1-2a_1^2)(\alpha-\beta)^2}. \tag{3.39}$$

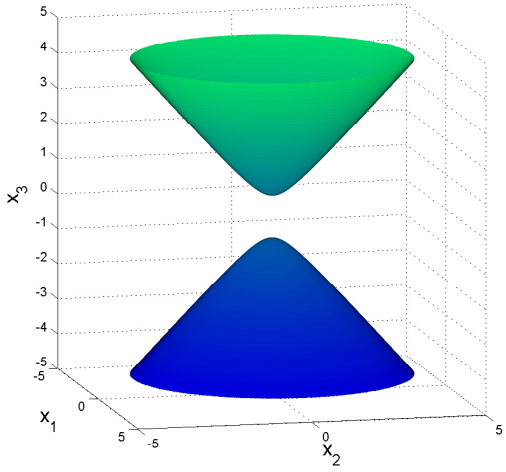
Hence, since $(1-2a_1^2)(\alpha-\beta)^2 > 0$, we have that one root is positive and the other is negative. We may assume w.l.o.g. that $\bar{\tau}_1 \leq \bar{\tau}_2$. Also, observe that in this case the roots are always different since the discriminant (3.34) of f is never zero for $a_1^2 < 1/2$.

Let us compare these two roots with the critical value $\hat{\tau} = -\frac{1}{(1-2a_1^2)}$. First of all, we know from (3.32) that $\hat{\tau}$ is not between the two roots. Additionally, if $\alpha \neq -\beta$, then the inequality in (3.32) is strict, i.e., $f(\hat{\tau}) > 0$. To complete the comparison we need to check the value of the derivative $f'(\hat{\tau})$ to verify in which branch of f the value $\hat{\tau}$ lies. We have that

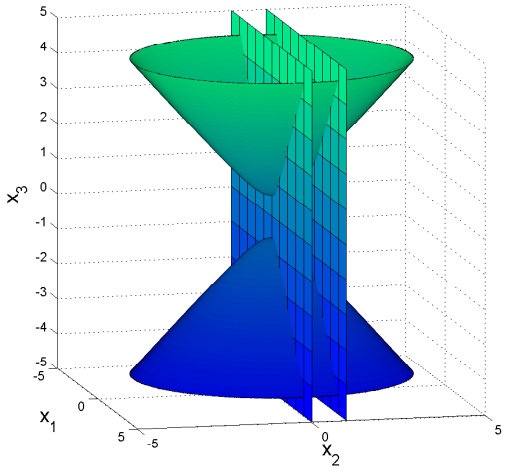
$$f'(\hat{\tau}) = -\frac{(\alpha-\beta)^2}{2} - (1-2a_1^2+\alpha\beta) = -\frac{(\alpha^2+\beta^2)}{2} - (1-2a_1^2) \leq 0.$$

Hence, the inequality $\hat{\tau} \leq \bar{\tau}_1$ is always satisfied, and it is strict if $\alpha \neq -\beta$.

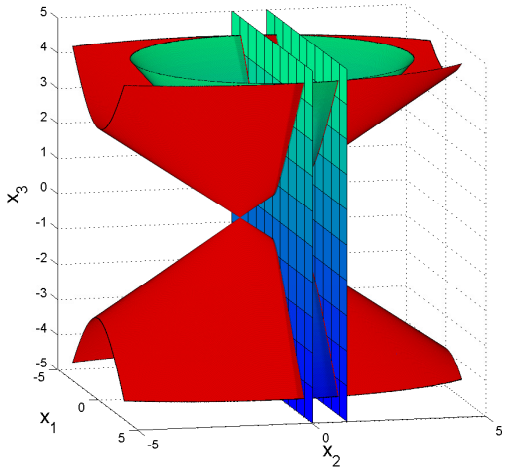
From Section 3.2.5.1, we know that if $\hat{\tau} < \bar{\tau}_1$, then $\operatorname{In} P(\bar{\tau}_1) = \{1, 0, \ell-1\}$ and $\operatorname{In} P(\bar{\tau}_2) =$



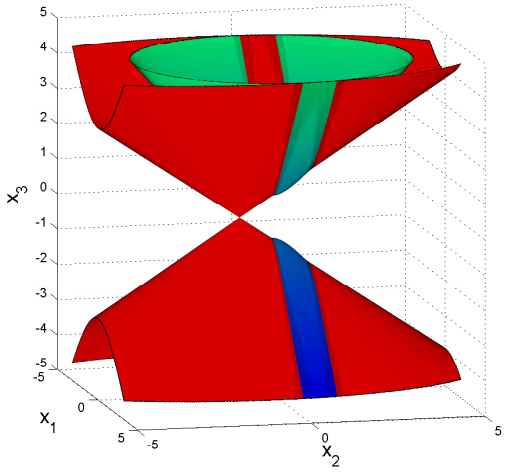
(a) The set $\mathcal{Q}$.



(b) Set $\mathcal{Q}$ intersected with two parallel hyperplanes.

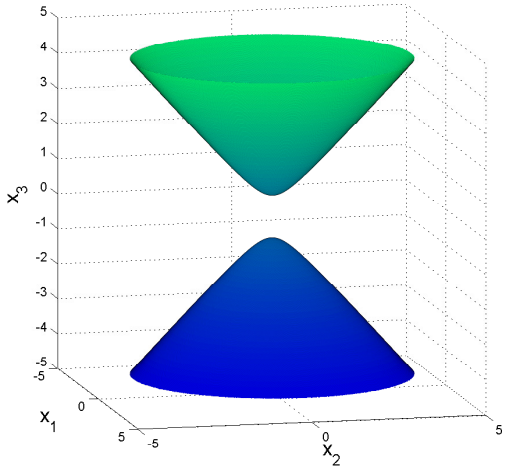


(c) Quadric $\mathcal{Q}(\bar{\tau}_1)$, sharing the same intersection with the hyperplanes that $\mathcal{Q}$ has.

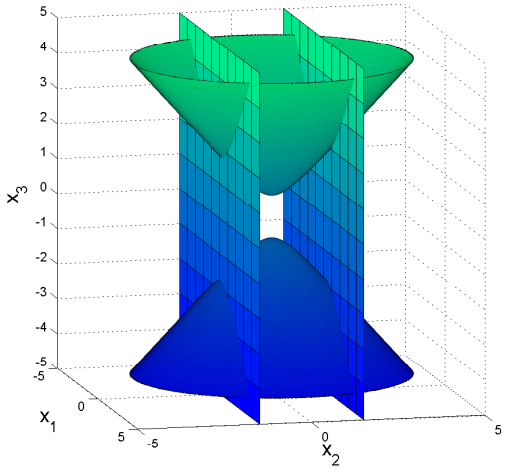


(d) Set $\mathcal{Q}$ intersected with the quadric $\mathcal{Q}(\bar{\tau}_1)$.

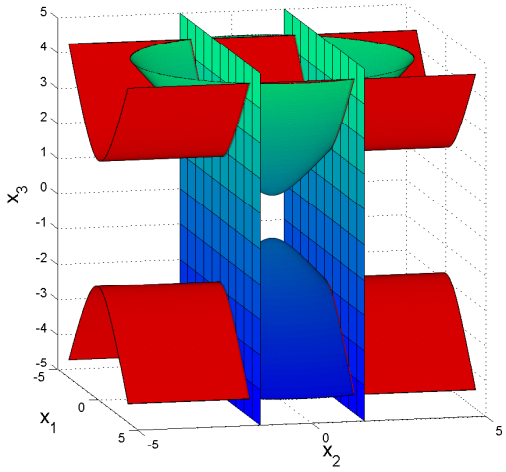
Figure 3.7: Illustration of Theorem 3.4.



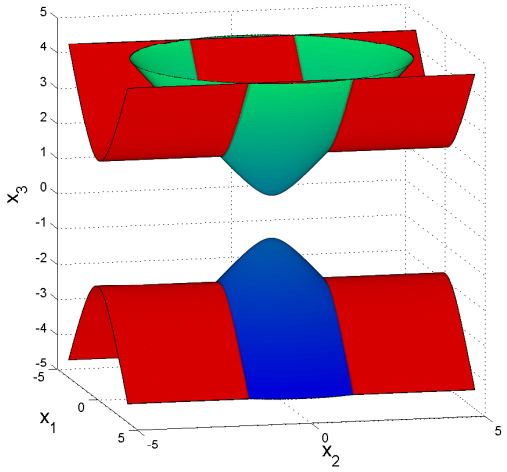
(a) The set $\mathcal{Q}$.



(b) Set $\mathcal{Q}$ intersected with two parallel hyperplanes.



(c) Quadric $\mathcal{Q}(\bar{\tau}_1)$, sharing the same intersection with the hyperplanes that $\mathcal{Q}$ has.



(d) Set $\mathcal{Q}$ intersected with the quadric $\mathcal{Q}(\bar{\tau}_1)$.

Figure 3.8: Illustration of Theorem 3.4.

$\{1, 0, \ell - 1\}$, i.e, they have $\ell - 1$ positive eigenvalues and one negative eigenvalue. As a result, $\mathcal{Q}(\tau_1)$ and $\mathcal{Q}(\tau_2)$ are two different scaled second order cones. On the other hand, if $\alpha = -\beta$, then the roots of f are given by

$$\begin{aligned}&\frac{2\left(1 - 2a_1^2 - \alpha^2 \pm \sqrt{(1 - 2a_1^2 - \alpha^2)^2 + 4(1 - 2a_1^2)\alpha^2}\right)}{4(1 - 2a_1^2)\alpha^2}\\ &= \frac{2\left(1 - 2a_1^2 - \alpha^2 \pm \sqrt{(1 - 2a_1^2 + \alpha^2)^2}\right)}{4(1 - 2a_1^2)\alpha^2}.\end{aligned}$$

Thus, $\hat{\tau} = \bar{\tau}_1$ when the hyperplanes are symmetric with respect to the origin. From Section 3.2.5.1 we know that $\operatorname{In} P(\bar{\tau}_1) = \{1, 1, \ell - 2\}$. Additionally, note that

$$\rho(\bar{\tau}_1) = 1 + \frac{\alpha^2}{(1 - 2a_1^2)} > 0.$$

Thus, similarly to the proof of Theorem 3.3, one can use the eigenvalue decomposition of $P(\bar{\tau}_1)$ to show that $\mathcal{Q}(\bar{\tau}_1)$ is an affine transformation of the set (3.37). Thus, $\mathcal{Q}(\bar{\tau}_1)$ is a cylindrical hyperboloid of two sheets. Finally, since $\bar{\tau}_1 < \bar{\tau}_2$, then $\operatorname{In} P(\bar{\tau}_2) = \{1, 0, \ell - 1\}$, and we obtain from Table 1.1 that $\mathcal{Q}(\bar{\tau}_2)$ is a cone. The results are summarize in Theorem 3.4, and they are illustrated in Figures 3.7 and 3.8.

Theorem 3.4. *If $a_1^2 < 1/2$ and $\rho = 1$ in (3.26), then for the shape of the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ we have the following cases:*

- *if $\beta \neq \alpha$, then both quadrics are cones,*
- *if $\beta = \alpha$, then $\mathcal{Q}(\bar{\tau}_1)$ is a hyperbolic cylinder of two sheets and $\mathcal{Q}(\bar{\tau}_2)$ is a cone.*

3.2.5.7 Classification when $a_1 < 1/2$ and $\rho = -1$

The structure of this section follows the same order given in Section 3.2.5.6. We begin computing the two roots of the function f. Then, we compare these roots to the critical

value $\hat{\tau} = -\frac{1}{(1-2a_1^2)}$. Based on this comparison, we close this section by classifying the shapes of the quadrics $\mathcal{Q}(\tau)$ at the two roots of f. In this case (3.31) simplifies to

$$f(\tau) = \tau^2(1-2a_1^2)\frac{(\alpha-\beta)^2}{4} + \tau((1-2a_1^2)-\alpha\beta) + 1. \tag{3.40}$$

Recall that if $\alpha \neq \beta$, then the two roots $\bar{\tau}_1$ and $\bar{\tau}_2$ are

$$\frac{2\left(\alpha\beta - (1-2a_1^2) \pm \sqrt{(\alpha\beta-(1-2a_1^2))^2 - (1-2a_1^2)(\alpha-\beta)^2}\right)}{(1-2a_1^2)(\alpha-\beta)^2}.$$

We may assume w.l.o.g. that $\bar{\tau}_1 < \bar{\tau}_2$. The discriminant of $f(\tau)$ can be factorized as follows

$$(\alpha\beta-(1-2a_1^2))^2 - (1-2a_1^2)(\alpha-\beta)^2 = (\alpha^2-(1-2a_1^2))(\beta^2-(1-2a_1^2)).$$

Hence, the roots $\bar{\tau}_1$ and $\bar{\tau}_2$ are reals only if:

- $\beta^2 \leq 1-2a_1^2$ and $\alpha^2 \leq 1-2a_1^2$, or
- $\beta^2 \geq 1-2a_1^2$ and $\alpha^2 \geq 1-2a_1^2$.

We need to compare now $\bar{\tau}_1$ and $\bar{\tau}_2$ with the critical value $\hat{\tau} = -1/(1-2a_1^2)$. We know from (3.32) that $\hat{\tau}$ is not between the two roots and that $f(\hat{\tau}) > 0$ if $\alpha \neq -\beta$. Hence, to complete the comparison we need to check the value of the derivative $f'(\hat{\tau})$ to verify in which branch of f the value $\hat{\tau}$ lies. We have that

$$f'(\tau) = -\frac{(\alpha-\beta)^2}{2} - (\alpha\beta-(1-2a_1^2)) = -\frac{(\alpha^2+\beta^2)}{2} + (1-2a_1^2).$$

For this comparison first we consider the case $\alpha \neq -\beta$. Then, $f(\hat{\tau}) > 0$ and we have two possibilities:

- If $\beta^2 \leq (1-2a_1^2)$ and $\alpha^2 \leq (1-2a_1^2)$, then $f'(\hat{\tau}) > 0$ and we obtain the inequality

$\bar{\tau}_2 < \hat{\tau}$. As a result, neither $\mathcal{Q}(\bar{\tau}_1)$ nor $\mathcal{Q}(\bar{\tau}_2)$ are cones, which implies that there are no cones in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in this case.

- If $\beta^2 \geq (1 - 2a_1^2)$ and $\alpha^2 \geq (1 - 2a_1^2)$, then $f'(\hat{\tau}) < 0$ and we obtain the inequality $\hat{\tau} < \bar{\tau}_1$. As a result, the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ are two different cones in this case.

To complete the comparison we consider now the case $\alpha = -\beta$, i.e., when the two hyperplanes are symmetric with respect to the origin. Then, the roots $\bar{\tau}_1$ and $\bar{\tau}_1$ of f are given by

$$\begin{aligned} &\frac{-\alpha^2 - (1 - 2a_1^2) \pm \sqrt{(\alpha^2 + (1 - 2a_1^2))^2 - 4(1 - 2a_1^2)\alpha^2}}{2(1 - 2a_1^2)\alpha^2} \\ &= \frac{-\alpha^2 - (1 - 2a_1^2) \pm \sqrt{(\alpha^2 - (1 - 2a_1^2))^2}}{2(1 - 2a_1^2)\alpha^2}. \end{aligned}$$

On one hand, if $\beta^2 \leq (1 - 2a_1^2)$ and $\alpha^2 \leq (1 - 2a_1^2)$, then we obtain the equality $\hat{\tau} = \bar{\tau}_2$. On the other hand, if $\beta^2 \geq (1 - 2a_1^2)$ and $\alpha^2 \geq (1 - 2a_1^2)$, then we obtain the equality $\hat{\tau} = \bar{\tau}_1$. Additionally, recall from Section 3.2.5.1 that $\mathrm{In}(P(\hat{\tau})) = \{1, 1, \ell - 2\}$. Then, we can divide the classification of the quadric $\mathcal{Q}(\hat{\tau})$ in three cases:

- If $\alpha^2 < (1 - 2a_1^2)$, then

$$\rho(\hat{\tau}) = -1 + \hat{\tau}\alpha\beta = -1 + \frac{\alpha^2}{(1 - 2a_1^2)} < 0.$$

Thus, similarly to the proof of Theorem 3.3, one can use the eigenvalue decomposition of $P(\hat{\tau})$ to show that in this case $\mathcal{Q}(\hat{\tau})$ is an affine transformation to the set

$$\{z \in \mathbb{R}^\ell \mid z^\top \hat{J} z \leq 1\}.$$

Thus, because $\mathrm{In}(P(\hat{\tau})) = \mathrm{In}(J)$, $\mathcal{Q}(\hat{\tau})$ is a hyperbolic cylinder of one sheet.

- If $\alpha^2 > (1 - 2a_1^2)$, then

$$\rho(\bar{\tau}_2) = -1 + \bar{\tau}_1 \alpha\beta = -1 + \frac{\alpha^2}{(1 - 2a_1^2)} > 0.$$

  Again, similarly to the proof of Theorem 3.3, one can use the eigenvalue decomposition of $P(\hat{\tau})$ to show that in this case $\mathcal{Q}(\hat{\tau})$ is an affine transformation of the set

$$\{u \in \mathbb{R}^\ell \mid u^\top \hat{J} u \leq -1\}.$$

  Hence, because $\mathrm{In}(P(\hat{\tau})) = \mathrm{In}(J)$, $\mathcal{Q}(\hat{\tau})$ is a hyperbolic cylinder of two sheets.

- If $\alpha^2 = (1 - 2a_1^2)$, then

$$\rho(\hat{\tau}) = -1 + \hat{\tau}\alpha\beta = -1 + \frac{\alpha^2}{(1 - 2a_1^2)} = 0.$$

  Again, similarly to the proof of Theorem 3.3, one can use the eigenvalue decomposition of $P(\hat{\tau})$ to show that in this case $\mathcal{Q}(\hat{\tau})$ is an affine transformation of the set

$$\{z \in \mathbb{R}^\ell \mid z^\top \hat{J} z \leq 0\}.$$

  Hence, because $\mathrm{In}(P(\hat{\tau})) = \mathrm{In}(J)$, $\mathcal{Q}(\hat{\tau})$ is a conic cylinder.

The results a summarize in the following theorem.

Theorem 3.5. *If $\rho = -1$ and $a_1 < 1/2$ in (3.26), then for the shape of the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ we have the following cases:*

- *if $\beta^2 \leq (1 - 2h_k^2)$, $\alpha^2 \leq (1 - 2h_k^2)$, and $\beta \neq -\alpha$, then there are no cones in the family;*
- *if $\beta^2 \geq (1 - 2h_k^2)$, $\alpha^2 \geq (1 - 2h_k^2)$, and $\beta \neq -\alpha$, then both quadrics are cones;*
- *if $\beta = -\alpha$, then we have the following sub-cases:*

- ◇ *if* $\alpha^2 < (1 - 2h_k^2)$, *then* $\hat{\tau} = \bar{\tau}_2$ *and* $\mathcal{Q}(\hat{\tau})$ *is a hyperbolic cylinder of one sheet;*
- ◇ *if* $\alpha^2 > (1 - 2h_k^2)$, *then* $\hat{\tau} = \bar{\tau}_1$ *and* $\mathcal{Q}(\hat{\tau})$ *is a hyperbolic cylinder of two sheets, and* $\mathcal{Q}(\hat{\tau}_2)$ *is a cone;*
- ◇ *if* $\alpha^2 = (1 - 2h_k^2)$, *then* $\hat{\tau} = \bar{\tau}_1 = \hat{\tau}_2$ *and* $\mathcal{Q}(\hat{\tau})$ *is a conic cylinder.*

3.3 Intersections with nonparallel hyperplanes

In this section, we investigate the intersection of the quadric $\mathcal{Q} \in \mathbb{R}^\ell$ with two hyperplanes in general position. For the sake of simplifying the algebra we may assume w.l.o.g. throughout this section that the quadric $\mathcal{Q}$ is unit hypersphere in $\mathbb{R}^\ell$ centered at the origin. Recall that using the inverse transformation of (3.4) the results obtained for the unit hypersphere can be generalized. Let $\mathcal{A}^= = \{x \in \mathbb{R}^\ell \mid a^\top x = \alpha\}$ and $\mathcal{B}^= = \{x \in \mathbb{R}^\ell \mid b^\top x = \beta\}$, for some $a, b \in \mathbb{R}^\ell$ and $\alpha, \beta \in \mathbb{R}$, be two given hyperplanes in general position. Additionally, we may assume w.l.o.g. that $\|a\| = \|b\| = 1$. Here we consider nonparallel hyperplanes, i.e., the vectors (a_1, α_1) and (a_2, α_2) are not scalar multiples of each other. We assume that both the intersections $\mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q} \cap \mathcal{B}^=$ are nonempty. We first present a generalization of Theorem 3.1 to the case of two hyperplanes in general position. Then, in Section 3.3.2 we analyze the behavior of the new family of quadrics when $P \succ 0$, to show that a quadric always exists that satisfies the definition of either a cone or a cylinder as is given in Section 1.1.2.1. Finally, we discuss how these results can be generalized to some cases when $\mathcal{Q}$ is not a unit hypersphere.

3.3.1 The family of quadrics with fixed planar sections

Here we generalize the results presented in Theorem 3.1 to the case when $\mathcal{A}^=$ and $\mathcal{B}^=$ are not parallel. We use the Definition 3.1 of a pencil of quadrics.

Theorem 3.6. *Let a quadric $\mathcal{Q} \in \mathbb{R}^\ell$ represented by (P, p, ρ), where $P \in \mathbb{R}^{\ell\times\ell}$, $p \in \mathbb{R}^\ell$, $\rho \in \mathbb{R}$, and two non-parallel hyperplanes $\mathcal{A}^= = \{x \in \mathbb{R}^\ell \mid a^\top x = \alpha\}$, $\mathcal{B}^= = \{x \in \mathbb{R}^\ell \mid b^\top x = \beta\}$, where $a, b \in \mathbb{R}^\ell$ and $\alpha, \beta \in R$, be given. The uni-parametric family of quadrics having the same intersection with $\mathcal{A}^=$ and $\mathcal{B}^=$ as the quadric $\mathcal{Q}$ is defined by the pencil of quadrics $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$, where $\mathcal{Q}(\tau)$ is represented by $(P(\tau), p(\tau), \rho(\tau)$, and*

$$\begin{aligned} P(\tau) &= P + \tau\frac{ab^\top + ba^\top}{2}, \\ p(\tau) &= p - \tau\frac{\beta a + \alpha b}{2}, \\ \rho(\tau) &= \rho + \tau\alpha\beta. \end{aligned}$$

Proof. Consider the set $\mathcal{A}^= \cup \mathcal{B}^=$, which can be described as

$$\{x \in \mathbb{R}^\ell \mid (a^\top x - \alpha)(b^\top x - \beta) = 0\},$$

and observe that

$$\begin{aligned} (a^\top x - \alpha)(b^\top x - \beta) &= x^\top ab^\top x - (\alpha b^\top + \beta a^\top)x + \alpha\beta \\ &= x^\top \left(\frac{ab^\top + ba^\top}{2}\right) x - (\alpha b^\top + \beta a^\top)x + \alpha\beta = 0. \end{aligned} \tag{3.41}$$

Now, let

$$\tilde{P} = \frac{ab^\top + ba^\top}{2}, \qquad \tilde{p} = -\frac{(\alpha b + \beta a)}{2}, \qquad \tilde{\rho} = \alpha\beta.$$

Then, the set of solutions of the equation (3.41) can be described by the quadric surface $\tilde{\mathcal{Q}}$ represented by $(\tilde{Q}, \tilde{q}, \tilde{\rho})$. Now, consider the pencil $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$, where $\mathcal{Q}(\tau)$ is represented by $P(\tau) = P + \tau\tilde{P}$, $p(\tau) = p + \tau\tilde{p}$, and $\rho(\tau) = \rho + \tau\tilde{\rho}$. Let $\bar{x}$ be a given vector

satisfying $\bar{x}^\top \tilde{P}\bar{x} + 2\tilde{p}^\top \bar{x} + \tilde{\rho} = 0$. Then, for $\tau \in R$ we have $\bar{x} \in \mathcal{Q}(\tau)$ if and only if

$$\bar{x}^\top (P + \tau\tilde{P})\bar{x} + 2(p - \tau\tilde{p})^\top \bar{x} + (\rho + \tau\tilde{\rho}) = \bar{x}^\top P\bar{x} + 2p^\top \bar{x} + \rho \leq 0.$$

Hence, we have $\bar{x} \in \mathcal{Q}(\tau) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$ if and only if $\bar{x} \in \mathcal{Q} \cap (\mathcal{A}^= \cup \mathcal{B}^=)$ for all $\tau \in \mathbb{R}$. □

We call the attention of the reader to the fact that Theorem 3.6 is rather general in that $\mathcal{Q}$ does not need to be constraint to be an ellipsoid. This assumption is made for the sake of simplifying the algebra in the analysis of the subsequent sections. Additionally, if $a = b$, theorem 3.6 simplifies to the result of Theorem 3.1.

3.3.2 Classification of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ when $P \succ 0$

In what follows, and until the end of Section 3.3 we assume that the quadric $\mathcal{Q}$ is an ellipsoid, i.e., $P \succ 0$. In other words, we consider the case when the sets $\mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q} \cap \mathcal{B}^=$ are bounded. If not said otherwise, we assume that the quadric $\mathcal{Q}$ is not a single point. Recall the affine transformation described in Section 3.1. Hence, to simplify the algebra, we may assume w.l.o.g. that $\mathcal{Q}$ is a unit hypersphere centered at the origin, and that $\|a\| = \|b\| = 1$. In this case we have that $P = I$, $p = 0$, and $\rho = -1$, and the representation of $\mathcal{Q}(\tau)$ is defined by

$$P(\tau) = I + \tau\frac{ab^\top + ba^\top}{2}, \quad p(\tau) = -\tau\frac{\beta a + \alpha b}{2}, \quad \rho(\tau) = -1 + \tau\alpha\beta. \tag{3.42}$$

We characterize the behavior of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in (3.42) as a function of parameter τ. First, we discuss the inertia of $P(\tau)$. Then, we analyze the cases: 1) when the matrix $P(\tau)$ is non-singular, and 2) when the matrix $P(\tau)$ is singular. Finally, we present a summary of the shapes of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ in Theorem 3.7.

3.3.2.1 The eigenvalues of $P(\tau)$

One of the characteristics deciding the shape of $\mathcal{Q}(\tau)$ is the number of negative or zero eigenvalues of $P(\tau)$, i.e., its inertia. Since P is modified with a rank-2 matrix in (3.42), $P(\tau)$ may possibly have two negative eigenvalues. The following lemma shows that this cannot happen when $P \succ 0$.

Lemma 3.8. *If $P \succ 0$, then $P(\tau)$ can have at most one non-positive eigenvalue.*

Proof. The eigenvalues of

$$P(\tau) = \left(I + \frac{\tau}{2}\left(ab^\top + ba^\top\right)\right) \tag{3.43}$$

are as follows:

- 1 is an eigenvalue with multiplicity $\ell - 2$, the corresponding eigenvectors are orthogonal to a and b;
- $1 + \frac{\tau}{2}\left(a^\top b + 1\right)$, with the eigenvector $(a + b)$;
- $1 + \frac{\tau}{2}\left(a^\top b - 1\right)$, with the eigenvector $(b - a)$.

Since $\left|a^\top b\right| \leq \|a\| \, \|b\|$, for

$$\hat{\tau}_1 = \frac{-2}{a^\top b + 1} \tag{3.44a}$$

$$\hat{\tau}_2 = \frac{-2}{a^\top b - 1}, \tag{3.44b}$$

we have that $\hat{\tau}_1 < 0 < \hat{\tau}_2$. This implies that $P(\tau)$ is positive definite if $\tau \in (\hat{\tau}_1, \hat{\tau}_2)$. It has a zero eigenvalue if $\tau = \hat{\tau}_1$ or $\tau = \hat{\tau}_2$, and it is indefinite with exactly one negative eigenvalue otherwise. □

From Lemma 3.8 we have that the possible shapes for $\mathcal{Q}(\tau)$ are the ones given in Section 1.1.2.1. We distinguish two cases: $P(\tau)$ is non-singular, and $P(\tau)$ is singular. In the following sections, we analyze these two cases separately.

3.3.2.2 $P(\tau)$ is non-singular

If $\tau \neq \hat{\tau}_1, \hat{\tau}_2$, then it follows from Lemma 3.8 that $P(\tau)$ is non-singular, which restricts the quadrics to the shapes in Table 1.1. Hence, to verify the existence of a cone in the family $\mathcal{Q}(\tau)$, it is necessary to identify a τ for which $p(\tau)^\top P(\tau)^{-1} p(\tau) - \rho(\tau) = 0$.

We use the Sherman-Morrison-Woodbury formula Golub and Van Loan [1996] to compute the inverse of $P(\tau)$:

$$\begin{aligned} P^{-1}(\tau) &= \left(I + \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 0 & \frac{\tau}{2} \\ \frac{\tau}{2} & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right)^{-1} \\ &= I - \frac{\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \tau^2 & -2\tau - \tau^2 a^\top b \\ -2\tau - \tau^2 a^\top b & \tau^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}}{\tau^2 \left(1 - (a^\top b)^2\right) - 4a^\top b\tau - 4} \\ &= I - \frac{(aa^\top + bb^\top)\tau^2 - (a^\top b\tau^2 + 2\tau)(ba^\top + ab^\top)}{\tau^2 \left(1 - (a^\top b)^2\right) - 4a^\top b\tau - 4}. \end{aligned} \tag{3.45}$$

Note that the roots of denominator of the second term in (3.45) are $\hat{\tau}_1$ and $\hat{\tau}_2$, given in (3.44a) and (3.44a). These are the values for which $P(\tau)$ is singular, as it was show by Lemma 3.8.

Now, we evaluate $p(\tau)^\top P^{-1}(\tau) p(\tau) - \rho(\tau)$. Substituting $p(\tau)$, $P^{-1}(\tau)$, and $\rho(\tau)$ from

(3.42) we obtain

$$\begin{aligned}
&p(\tau)^\top P^{-1}(\tau)p(\tau) - \rho(\tau) \\
=&p(\tau)^\top \left(I - \frac{(aa^\top + bb^\top)\tau^2 + (a^\top b\tau^2 + 2\tau)(ba^\top + ab^\top)}{\tau^2\left(1 - (a^\top b)^2\right) - 4a^\top b\tau - 4} \right) p(\tau) - \rho(\tau) \\
=&\frac{(1 - (a^\top b)^2 + 2\alpha\beta a^\top b - \alpha^2 - \beta^2)\tau^2 + 4(\alpha\beta - a^\top b)\tau - 4}{\tau^2\left(1 - (a^\top b)^2\right) - 4a^\top b\tau - 4} \\
=&\frac{\left((1-\alpha^2)(1-\beta^2) - (\alpha\beta - a^\top b)^2\right)\tau^2 + 4(\alpha\beta - a^\top b)\tau - 4}{\tau^2\left(1 - (a^\top b)^2\right) - 4a^\top b\tau - 4} \\
=&\frac{\left((\alpha\beta - a^\top b)^2 - (1-\alpha^2)(1-\beta^2)\right)\tau^2 + 4(a^\top b - \alpha\beta)\tau + 4}{\tau^2\left((a^\top b)^2 - 1\right) + 4a^\top b\tau + 4}.
\end{aligned} \tag{3.46}$$

Recall that the denominator of (3.46) is non-zero if $\tau \neq \hat{\tau}_1, \hat{\tau}_2$, then we need to focus only on its numerator. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be the numerator of (3.46) as function of τ:

$$f(\tau) = \left((\alpha\beta - a^\top b)^2 - (1-\alpha^2)(1-\beta^2)\right)\tau^2 + 4(a^\top b - \alpha\beta)\tau + 4.$$

This is a quadratic function of τ, whose discriminant is

$$16(1-\alpha^2)(1-\beta^2). \tag{3.47}$$

Thus, since $\mathcal{Q}$ is a unit hypersphere, we know that f has real roots if $\mathcal{Q} \cap \mathcal{A}^= \neq \emptyset$ and $\mathcal{Q} \cap \mathcal{B}^= \neq \emptyset$. Let the roots of f be denoted by $\bar{\tau}_1$ and $\bar{\tau}_2$. We may assume w.l.o.g. that $\bar{\tau}_1 \leq \bar{\tau}_2$.

Summary of shapes We need to compare the roots of f with $\hat{\tau}$ and $\hat{\tau}_2$ to characterize the shapes of $\mathcal{Q}(\tau)$. Recall that in this section we consider the case when $P(\tau)$ in non-singular, i.e., $\tau \neq \hat{\tau}_1, \hat{\tau}_2$. We first analyze the case when $\hat{\tau}_1 < \bar{\tau}_i < \hat{\tau}_2$, for some $i = 1, 2$. In

such case it follows from Lemma 3.8 that for the root of f that is between $\hat{\tau}_1$ and $\hat{\tau}_2$ we have that $P(\bar{\tau}_i) \succ 0$. Now, since

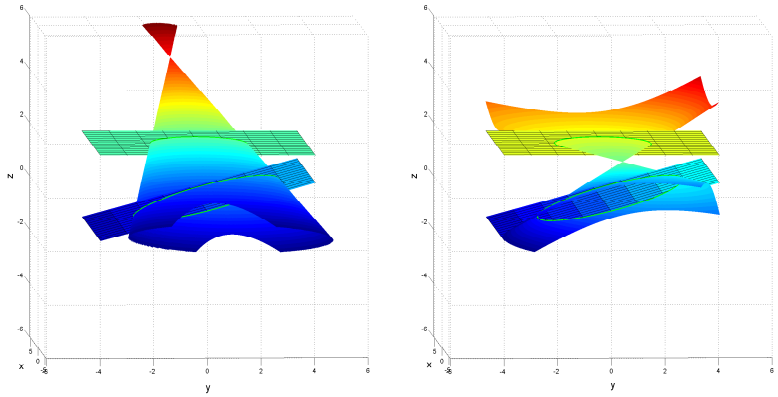
$$p(\bar{\tau}_i)^\top P^{-1}(\bar{\tau}_i)\bar{p}(\bar{\tau}_i) - \rho(\bar{\tau}_i) = 0,$$

from Table 1.1 in Section 1.1.2.1 we know that $\mathcal{Q}(\bar{\tau}_i)$ is a point. This is possible only if $\mathcal{Q}$ is a point and $\mathcal{A}^= \cap \mathcal{B}^= \neq \emptyset$, because $\mathcal{Q} \cap (\mathcal{A}^= \cup \mathcal{B}^=) = \mathcal{Q}(\tau) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$ for $\tau \in \mathbb{R}$. This implies that $p^\top P^{-1} p - \rho = 0$ and $\alpha = \beta = 0$. Hence, $p(\tau) = 0$ and $\rho(\tau) = 0$ for all $\tau \in \mathbb{R}$, which simplifies the characterization of all the shapes of $\mathcal{Q}(\tau)$ for $\tau \in \mathbb{R}$. First, for any $\hat{\tau}_1 < \tau < \hat{\tau}_2$ the quadric $\mathcal{Q}(\tau)$ is a point. Second, the identity $-P(\hat{\tau}_i)0 = p(\hat{\tau}_i)$ holds for $\hat{\tau}_1$ and $\hat{\tau}_2$, where 0 is the all zeros vector in $\mathbb{R}^\ell$. Thus, it follows from **Case 1** in Section 1.1.2.1 that the quadrics $\mathcal{Q}(\hat{\tau}_1)$ and $\mathcal{Q}(\hat{\tau}_2)$ are lines. Finally, for $\tau < \hat{\tau}_1$ and $\tau > \hat{\tau}_2$, the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ are cones.

Now, if neither $\bar{\tau}_1 \notin (\hat{\tau}_1, \hat{\tau}_2)$ nor $\bar{\tau}_2 \notin (\hat{\tau}_1, \hat{\tau}_2)$, then the shapes of the quadrics $\mathcal{Q}(\bar{\tau}_1)$, $\mathcal{Q}(\bar{\tau}_2)$, $\mathcal{Q}(\hat{\tau}_1)$, $\mathcal{Q}(\hat{\tau}_2)$, depend on the value of the discriminant (3.47). We have the following cases:

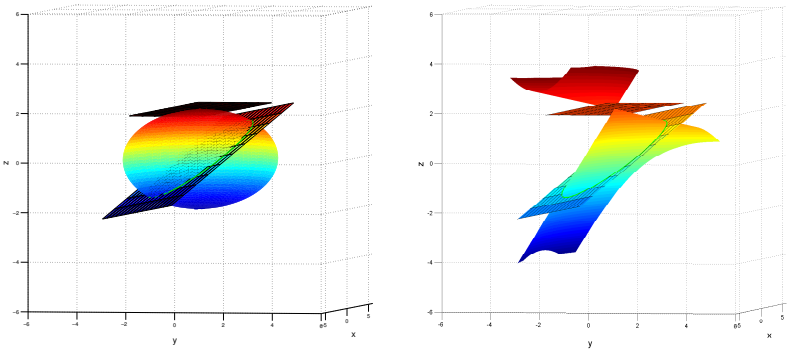
- If the discriminant (3.47) of f is not equal to zero, then $\hat{\tau}_2 < \bar{\tau}_1 < \bar{\tau}_2$, or $\bar{\tau}_1 < \bar{\tau}_2 < \hat{\tau}_1$, or $\bar{\tau}_1 < \hat{\tau}_1 < \hat{\tau}_2 < \bar{\tau}_2$. In these cases we have that $\mathcal{Q}(\hat{\tau}_1)$ and $\mathcal{Q}(\hat{\tau}_2)$ are two paraboloids, and $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ are two different cones. For illustrations see Figure 3.9.

- If the discriminant (3.47) of f is zero, then $\bar{\tau}_1 = \bar{\tau}_2 < \hat{\tau}_1$ or $\hat{\tau}_2 < \bar{\tau}_1 = \bar{\tau}_2$. In these cases $\mathcal{Q}(\hat{\tau}_1)$ and $\mathcal{Q}(\hat{\tau}_2)$ are two paraboloids, and there is a unique cone $\mathcal{Q}(\bar{\tau}_1) = \mathcal{Q}(\bar{\tau}_2)$. Observe that in these cases one of the hyperplanes must be tangent to the hypersphere $\mathcal{Q}$. For illustrations see Figure 3.10.



(a) $\tau = \bar{\tau}_1$ (b) $\tau = \bar{\tau}_2$

Figure 3.9: Function f has two distinct roots $\bar{\tau}_1$ and $\bar{\tau}_1$, which are different from $\hat{\tau}_1$ and $\hat{\tau}_2$.



(a) One of the hyperplanes is tangent to the ellipsoid. (b) $\tau = \bar{\tau}_1 = \bar{\tau}_2$

Figure 3.10: The two roots of f coincide, but are different from $\hat{\tau}_1$ and $\hat{\tau}_2$.

3.3.2.3 $P(\tau)$ is singular

If $\tau = \hat{\tau}_1$ or $\tau = \hat{\tau}_2$, then it follows from Lemma 3.8 that $P(\hat{\tau}_i)$, $i = 1, 2$ is singular. In this case we have $P(\hat{\tau}_i) \succeq 0$ but not $P(\hat{\tau}_i) \succ 0$. Thus, from Section 1.1.2.1 we have that $\mathcal{Q}(\hat{\tau}_i)$ is either a line, a cylinder, or a paraboloid. The shape of $\mathcal{Q}(\hat{\tau}_i)$ can be decided by verifying if $p(\hat{\tau}_i)$ is in the range of $P(\hat{\tau}_i)$. This happens exactly when $p(\hat{\tau}_i)$ is orthogonal

to the eigenvector corresponding to the zero eigenvalue of $P(\hat{\tau}_i)$. Lemma 3.9 provides the eigenvectors corresponding to the zero eigenvalue of $P(\hat{\tau}_1)$ and $P(\hat{\tau}_2)$.

Lemma 3.9. *The eigenvector for the zero eigenvalue of $P(\hat{\tau}_1)$ is $(b+a)$, and for the zero eigenvalue of $P(\hat{\tau}_2)$ is $(b-a)$.*

Proof. For $P(\hat{\tau}_1)$, direct computation yields

$$\begin{aligned} P(\hat{\tau}_1)\,(b+a) &= \left(I - \frac{ab^\top + ba^\top}{a^\top b + 1}\right)(b+a) \\ &= (b+a) - \frac{(b+a)\left(a^\top b + 1\right)}{a^\top b + 1} = 0, \end{aligned}$$

and similarly for $p(\hat{\tau}_2)$, we obtain

$$\begin{aligned} P(\hat{\tau}_2)\,(b-a) &= \left(I - \frac{ab^\top + ba^\top}{a^\top b - 1}\right)(b-a) \\ &= (b-a) - \frac{(b-a)\left(a^\top b - 1\right)}{a^\top b - 1} = 0. \end{aligned}$$

We used here that a and b are linearly independent, otherwise they would be parallel and the analysis would reduce to the case in Section 3.2. Thus the two eigenvectors are not the zero vector. This completes the proof. □

Now we can compute the inner product of these eigenvectors with $p(\hat{\tau}_1)$ and $p(\hat{\tau}_2)$. Consider first $p(\hat{\tau}_1)$, then we obtain:

$$p(\hat{\tau}_1)^\top (a+b) = \frac{(\alpha b^\top + \beta a^\top)(a+b)}{a^\top b + 1} = \alpha + \beta. \tag{3.48}$$

For the case $p(\hat{\tau}_2)$ we obtain:

$$p(\hat{\tau}_2)^\top (b-a) = \frac{(\alpha b^\top + \beta a^\top)(b-a)}{a^\top b - 1} = \beta - \alpha. \tag{3.49}$$

Recall that if (3.48) or (3.49) is not zero, then we have that either $-p(\hat{\tau}_1)$ is not in the range of $P(\hat{\tau}_1)$ or $-p(\hat{\tau}_2)$ is not in the range of $P(\hat{\tau}_2)$. These to cases can occur simultaneously. Hence, from **Case 2** in Section 1.1.2.1 either $\mathcal{Q}(\hat{\tau}_1)$ or $\mathcal{Q}(\hat{\tau}_2)$, or both are paraboloids.

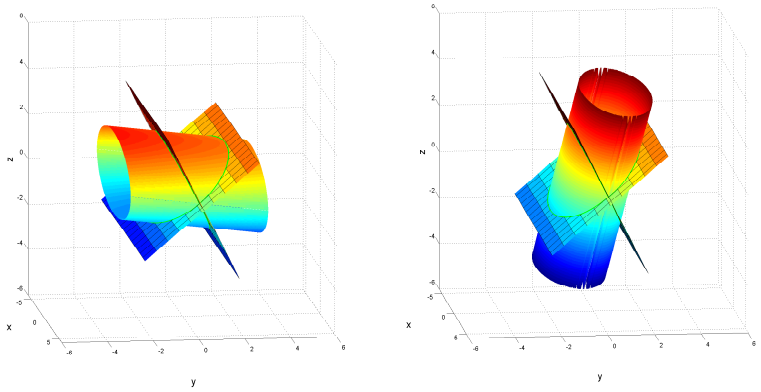
Summary of shapes We use the discriminant of f in (3.47) to complete the classification the shapes of the quadrics $\mathcal{Q}(\bar{\tau}_1)$, $\mathcal{Q}(\bar{\tau}_2)$, $\mathcal{Q}(\hat{\tau}_1)$, $\mathcal{Q}(\hat{\tau}_2)$. Recall that $\bar{\tau}_1 \leq \bar{\tau}_2$ and $\hat{\tau}_1 \leq \hat{\tau}_2$. Then, we have the following cases:

- If the discriminant (3.47) of f is not equal to zero, then we need to consider two possibilities:
  - ◇ Both $\hat{\tau}_1 = \bar{\tau}_1$ and $\hat{\tau}_2 = \bar{\tau}_2$, which is illustrated in Figure 3.11. In this case $f(\hat{\tau}_1) = f(\hat{\tau}_2) = 0$, thus

$$(\alpha + \beta)^2 = 0 \tag{3.50}$$

$$(\alpha - \beta)^2 = 0, \tag{3.51}$$

    which implies that $\alpha = \beta = 0$, i.e., both hyperplanes intersect at the origin, which is the center of $\mathcal{Q}$. Hence, for the vector $w^{\mathrm{c}} = 0$ the identity $-P(\hat{\tau}_i)w^{\mathrm{c}} = \bar{p}(\hat{\tau}_i)$ holds for $\hat{\tau}_1$ and $\hat{\tau}_2$. Furthermore, since $p(\hat{\tau}_i) = 0$ and $\bar{\rho}(\hat{\tau}_i) = -1$, then $p(\hat{\tau}_i)^\top p(\hat{\tau}_i)p(\hat{\tau}_i) - \rho(\hat{\tau}_i) = 1 > 0$ for $i = 1, 2$. Thus, it follows from **Case 1** in Section 1.1.2.1 that the quadrics $\mathcal{Q}(\hat{\tau}_i)$, $i = 1, 2$, are two cylinders.
  - ◇ Exactly one of the roots $\bar{\tau}_1$ or $\bar{\tau}_2$ is equal to either $\hat{\tau}_1$ or $\hat{\tau}_2$, which is illustrated in Figure 3.12. Recall that if the discriminant is not equal to zero, then $|\alpha| < 1$ and $|\beta| < 1$, i.e., neither of the hyperplanes $\mathcal{A}^=$ or $\mathcal{B}^=$ are tangent to $\mathcal{Q}$. Assume that one of the roots is equal to $\hat{\tau}_1$. It follows from equations (3.50) and (3.48)



(a) $\tau = \bar{\tau}_1 = \hat{\tau}_1$ (b) $\tau = \bar{\tau}_2 = \hat{\tau}_2$

Figure 3.11: $\bar{\tau}_1 \neq \bar{\tau}_2$, and $\bar{\tau}_1 = \hat{\tau}_1$, $\bar{\tau}_2 = \hat{\tau}_2$.

that $\alpha = -\beta$, and that $(a + b)$ is orthogonal to $p(\hat{\tau}_1)$. Now, let

$$w^{\mathrm{c}} = \frac{\beta}{2}(b - a). \tag{3.52}$$

Then, we have

$$\begin{aligned} P(\hat{\tau}_1)w^{\mathrm{c}} =& \left(I - \frac{(ab^\top + ba^\top)}{(a^\top b + 1)}\right)\left(\frac{\beta(b-a)}{2}\right) \\ =& \frac{\beta}{2}\left((b-a) - \frac{(a - a^\top ba + a^\top bb - b)}{(a^\top b + 1)}\right) \\ =& \frac{\beta}{2}\left((b-a) - \frac{(a^\top b - 1)(b-a)}{(a^\top b + 1)}\right) \\ =& \frac{\beta(b-a)}{2}\left(\frac{(a^\top b + 1) - (a^\top b - 1)}{(a^\top b + 1)}\right) \\ =& -\frac{\beta(a-b)}{(a^\top b + 1)} = -p(\hat{\tau}_1). \end{aligned} \tag{3.53}$$

Additionally, for the choice of x_c in (3.52) we have that

$$\begin{aligned}(w^{\mathrm{c}})^\top P(\hat{\tau}_1) w^{\mathrm{c}} - \rho(\hat{\tau}_1) =& \frac{\beta^2 (b-a)^\top (b-a)}{2(a^\top b + 1)} - \rho(\hat{\tau}_1) \\ =& \frac{\beta^2 (1 - a^\top b)}{(a^\top b + 1)} + \frac{2\alpha\beta}{(a^\top b + 1)} + 1 \\ =& \frac{\beta^2 (1 - a^\top b)}{(a^\top b + 1)} - \frac{2\beta^2}{(a^\top b + 1)} + 1 \\ =& - \frac{\beta^2 (a^\top b + 1)}{(a^\top b + 1)} + 1 \\ =& 1 - \beta^2 > 0, \end{aligned} \tag{3.54}$$

where the strict inequality holds in the case $|\beta| < 1$. As a result, from **Case 1** in Section 1.1.2.1 we obtain that the quadric $\mathcal{Q}(\hat{\tau}_1)$ is a cylinder.

Similarly, when one of the roots equals to $\hat{\tau}_2$, then we can choose

$$w^{\mathrm{c}} = \frac{\beta}{2} (b + a) . \tag{3.55}$$

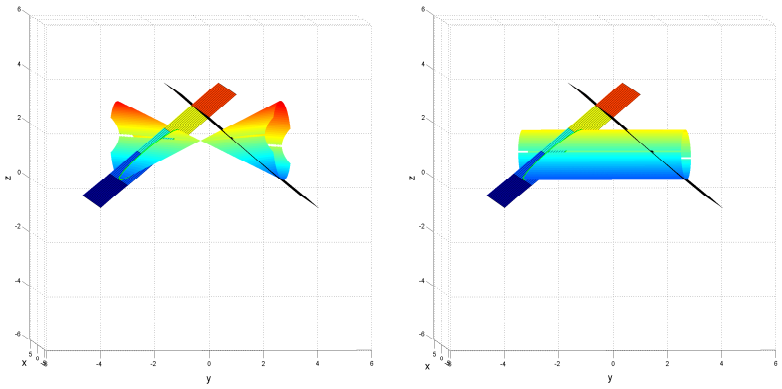
In this case, it follows from equations (3.51) and (3.49) that $\alpha = \beta$, and that $(b - a)$ is orthogonal to $p(\hat{\tau}_2)$. Additionally, we have that

$$\begin{aligned} P(\hat{\tau}_2) x w^{\mathrm{c}} =& \left(I - \frac{(ab^\top + ba^\top)}{(a^\top b - 1)} \right) \left(\frac{\beta(b+a)}{2} \right) \\ =& \frac{\beta(b+a)}{2} - \frac{\beta(a + a^\top b a + a^\top b b + b)}{2(a^\top b - 1)} \\ =& \frac{\beta}{2} \left((b+a) - \frac{(a^\top b + 1)(b+a)}{(a^\top b - 1)} \right) \\ =& \frac{\beta(b+a)}{2} \left(\frac{(a^\top b - 1) - (a^\top b + 1)}{(a^\top b - 1)} \right) \\ =& - \frac{\beta(b+a)}{(a^\top b - 1)} = -p(\hat{\tau}_2). \end{aligned} \tag{3.56}$$

Additionally, for the choice of w^{c} in (3.55) we have that

$$\begin{aligned}(w^{\mathrm{c}})^{\top} p(\hat{\tau}_2) w^{\mathrm{c}} - \rho(\hat{\tau}_2) &= -\frac{\beta^2 (b+a)^{\top}(b+a)}{2(a^{\top}b-1)} - \rho(\hat{\tau}_2) \\ &= -\frac{\beta^2(1+a^{\top}b)}{(a^{\top}b-1)} + \frac{2\alpha\beta}{(a^{\top}b-1)} + 1 \\ &= -\frac{\beta^2(1+a^{\top}b)}{(a^{\top}b-1)} + \frac{2\beta^2}{(a^{\top}b-1)} + 1 \\ &= -\frac{\beta^2(a^{\top}b-1)}{(a^{\top}b-1)} + 1 \\ &= 1-\beta^2 > 0, \end{aligned} \tag{3.57}$$

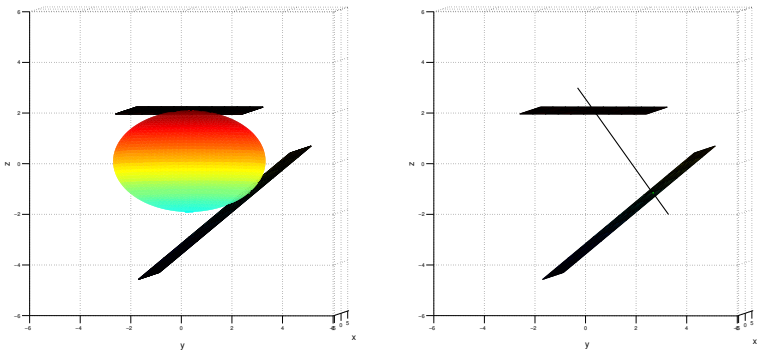
where the strict inequality holds in the case $|\beta| < 1$. As a result, from **Case 1** in Section 1.1.2.1 we obtain that the quadric $\mathcal{Q}(\hat{\tau}_2)$ is a cylinder.



(a) $\tau = \bar{\tau}_1$ (b) $\tau = \bar{\tau}_2 = \hat{\tau}_1$

Figure 3.12: Function $f(\tau)$ has two distinct roots, but one of them coincides with either $\hat{\tau}_1$ or $\hat{\tau}_2$.

- If the discriminant of f in (3.47) is zero, then the two roots of f are equal, i.e., $\bar{\tau} = \bar{\tau}_1 = \bar{\tau}_2$. Let $\bar{\tau} = \hat{\tau}_1$, then from equation (3.50) we obtain the identity $\alpha = -\beta$.



(a) Two hyperplanes are tangent to the ellipsoid.

(b) $\tau = \bar{\tau}_1 = \bar{\tau}_2 = \hat{\tau}_1$

Figure 3.13: The two roots of $f(\tau)$ coincide with either $\hat{\tau}_1$ or $\hat{\tau}_2$.

Now, since the discriminant of f is zero, we have

$$\alpha^2 = \beta^2 = 1, \tag{3.58}$$

and, since $\mathcal{Q}$ is a unit hyper-sphere, it follows that the hyperplanes $\mathcal{A}^=$ and $\mathcal{B}^=$ are both tangent to the ellipsoid $\mathcal{Q}$. Recall that from Equation (3.53) for w^c in (3.52) we have $P(\hat{\tau}_1)w^c = -p(\hat{\tau}_1)$. Furthermore, from (3.54) and (3.58) we have that $(w^c)^\top P(\hat{\tau}_1)w^c - \rho(\hat{\tau}_1) = 0$. Hence, the quadric $\mathcal{Q}(\hat{\tau}_1)$ is a line.

Similarly, if $\bar{\tau} = \hat{\tau}_2$, then from equation (3.51) we obtain $\beta = \alpha$, and the identity (3.58) still holds. Then, from equation (3.56) for w^c in (3.55)) we have $\bar{P}(\hat{\tau}_2)w^c = -p(\hat{\tau}_2)$, and from (3.57) we have that $(w^c)^\top P(\hat{\tau}_2)w^c - \rho(\hat{\tau}_2) = 0$. Then, the quadric $\mathcal{Q}(\hat{\tau}_1)$ is a line in this case as well. For illustrations of these cases see Figure 3.13.

3.3.2.4 Summarizing the shapes of $\mathcal{Q}(\tau)$

We can now summarize the possible shapes of the quadrics in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ at $\hat{\tau}_1$, $\hat{\tau}_2$, $\bar{\tau}_1$, and $\bar{\tau}_2$, where $\hat{\tau}_1 < \hat{\tau}_2$ and $\bar{\tau}_1 < \bar{\tau}_2$.

Theorem 3.7. *The following cases may occur for the shape of $\mathcal{Q}(\tau)$:*

- *$\hat{\tau}_2 < \bar{\tau}_1 < \bar{\tau}_2$, or $\bar{\tau}_1 < \bar{\tau}_2 < \hat{\tau}_1$, or $\bar{\tau}_1 < \hat{\tau}_1 < \hat{\tau}_2 < \bar{\tau}_2$: Then both $\mathcal{Q}(\hat{\tau}_1)$, and $\mathcal{Q}(\hat{\tau}_2)$ are paraboloids, and both $\mathcal{Q}(\bar{\tau}_1)$, and $\mathcal{Q}(\bar{\tau}_2)$ are cones.*
- *$\bar{\tau}_1 = \bar{\tau}_2 < \hat{\tau}_1$ or $\hat{\tau}_2 < \bar{\tau}_1 = \bar{\tau}_2$: Then both $\mathcal{Q}(\hat{\tau}_1)$, and $\mathcal{Q}(\hat{\tau}_2)$ are paraboloids, and $\mathcal{Q}(\bar{\tau}_1) = \mathcal{Q}(\bar{\tau}_2)$ is a cone.*
- *$\bar{\tau}_1 = \hat{\tau}_1$ and $\hat{\tau}_2 = \bar{\tau}_2$: Then both $\hat{\mathcal{Q}}(\hat{\tau}_1) = \mathcal{Q}(\bar{\tau}_1)$, and $\mathcal{Q}(\hat{\tau}_2) = \hat{\mathcal{Q}}(\bar{\tau}_2)$ are cylinders.*
- *$\bar{\tau}_1 \neq \hat{\tau}_2$ and exactly one of $\bar{\tau}_1$ or $\bar{\tau}_2$, is equal to either $\hat{\tau}_1$ or $\hat{\tau}_2$: Then either $\mathcal{Q}(\hat{\tau}_1)$ is a cylinder and $\mathcal{Q}(\hat{\tau}_2)$ is a paraboloid, or $\mathcal{Q}(\hat{\tau}_2)$ is a cylinder and $\mathcal{Q}(\hat{\tau}_1)$ is a paraboloid. In both cases exactly one of $\mathcal{Q}(\bar{\tau}_1)$ or $\mathcal{Q}(\bar{\tau}_2)$ is a cone.*
- *$\bar{\tau}_1 = \bar{\tau}_2 = \hat{\tau}_1$ or $\hat{\tau}_2 = \bar{\tau}_1 = \bar{\tau}_2$: Then either $\mathcal{Q}(\hat{\tau}_1)$ is a line and $\hat{\mathcal{Q}}(\hat{\tau}_2)$ is a paraboloid, or $\mathcal{Q}(\hat{\tau}_2)$ is a line and $\mathcal{Q}(\hat{\tau}_1)$ is a paraboloid.*

This completes the description of the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of quadrics when $P \succ 0$ and $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ are non-parallel.

3.3.3 Generalization

In Section 3.3.2 we assumed that the initial quadric $\mathcal{Q}$ is an ellipsoid. This assumption indeed facilitates the analysis of the family. However, recall that the domain of τ is the whole real line. Hence, the results obtained in Section 3.3.2 cover also the cases where $\mathcal{Q}$ *has an ID1 matrix P and the intersection with the hyperplanes are bounded.* We formalize this result, which follows directly from Theorem 3.6 and Lemma 3.8, as the following corollary.

Corollary 3.1. *If there exists a $\tilde{\tau} \in \mathbb{R}$ such that the quadric $\mathcal{Q}(\tilde{\tau}) \in \{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ has a matrix $P(\tilde{\tau}) \succ 0$, then $P(\tau)$ has at most one non-positive eigenvalue for any $\tau \in \mathbb{R}$.*

Hence, in this case to characterize the family we may use $\mathcal{Q}(\tilde{\tau})$, which is an ellipsoid.

Chapter 4

Disjunctive conic cuts for MISOCO problems

In this chapter, we describe the procedure to obtain DCCs for MISOCO problems. The fundamental problem considered in this section is a MISOCO problem with a single cone, i.e., a problem of the form

$$\begin{aligned} \text{minimize: } & c^\top x \\ \text{subject to: } & Ax = b \\ & x \in \mathbb{L}^n \\ & x \in \mathbb{Z}^d \times \mathbb{R}^{n-d}, \end{aligned} \tag{4.1}$$

where $A \in \mathbb{R}^{m \times n}$ is a matrix with full row rank, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$.

The relaxation of the integer variables in (4.1) to continuous variables yields the SOCO

problem

$$\begin{aligned} \text{minimize: } & c^\top x \\ \text{subject to: } & Ax = b \\ & x \in \mathbb{L}^n. \end{aligned} \tag{4.2}$$

We show in this chapter how a DCC can be derived for (4.1) when the feasible set of (4.2) is intersected with a disjunctive set $\mathcal{U} \cup \mathcal{V}$, where $\mathcal{U} = \{x \in \mathbb{R}^n \mid u^\top x \geq \varphi\}$ and $\mathcal{V} = \{x \in \mathbb{R}^n \mid v^\top x \leq \varpi\}$.

We start in Section 4.1 describing the relation of the feasible set of (4.2) and quadrics. Then, in Section 4.2 we use the results of Chapters 2 and 3 to derive disjunctive conic cuts when the disjunctive set used is defined with parallel hyperplanes. We close this chapter in Section 4.3 discussing the derivation of DCCs when the disjunctive set used is defined with non-parallel hyperplanes.

4.1 Properties of the quadric $\mathcal{Q}$ associated with the feasible set of problem (4.2)

Recall the definition of quadric sets given in Section 1.1.2. In this section we find an equivalent representation of the feasible set of (4.2), given by

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax = b,\ x \in \mathbb{L}\}, \tag{4.3}$$

in terms of a quadric set. Then, provided a quadric for the equivalent representation of $\mathcal{F}$, we show that it is possible to derive cuts for problem (4.2) using that quadric. Finally, we analyze the possible shapes that the quadric in the representation of $\mathcal{F}$ can take.

For this analysis, we use a representation of the set $\{x \in \mathbb{R}^n \mid Ax = b\}$ in terms of an

orthonormal basis of the null space of A. To define this representation, consider a vector $x^0 \in \mathcal{F}$, and let $H_{n\times\ell}$ be a matrix whose columns form an orthonormal basis for the null space of A, where $\ell = n - m$. We may assume that $\ell > 0$, since otherwise $Ax = b$ has a unique solution and the set $\mathcal{F}$ reduces to a point. Then, we have the identity

$$\{x \in \mathbb{R}^n \mid Ax = b\} = \{x \in \mathbb{R}^n \mid \exists w \in \mathbb{R}^\ell,\ x = x^0 + Hw\}. \tag{4.4}$$

We can use (4.4) to rewrite the set $\mathcal{F}$ in (4.3) in terms of a quadric as follows. We first find a quadric $\mathcal{Q}$ represented by a matrix $P \in \mathbb{R}^{\ell\times\ell}$, a vector $p \in \mathbb{R}^\ell$, and a scalar ρ, which are defined in terms of H and x^0. Let $J \in \mathbb{R}^{n\times n}$ be a diagonal matrix defined as

$$J = \begin{bmatrix} -1 & 0 \\ 0 & I \end{bmatrix},$$

and let us relax the constraint $x \in \mathbb{L}^n$ to $x^\top Jx \leq 0$. Substituting $x = x^0 + Hw$ in the relaxed constraint we obtain

$$\begin{aligned} (x^0 + Hw)^\top J(x^0 + Hw) &\leq 0 \\ w^\top H^\top JHw + 2(x^0)^\top JHw + (x^0)^\top Jx^0 &\leq 0. \end{aligned} \tag{4.5}$$

Define $P = H^\top JH$, $p = H^\top Jx^0$, and $\rho = (x^0)^\top Jx^0$. Now, substituting P, p, and ρ in (4.5) we obtain the constraint

$$w^\top Pw + 2p^\top w + \rho \leq 0. \tag{4.6}$$

Let us define the following quadric

$$\mathcal{Q} = \{w \in \mathbb{R}^\ell \mid w^\top Pw + 2p^\top w + \rho \leq 0\}. \tag{4.7}$$

Thus, the set $\mathcal{F}$ admits the following equivalent representation

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid x = x^0 + Hw, \text{ with } w \in \mathcal{Q}, \text{ and } x_1 \geq 0\}. \tag{4.8}$$

Our goal in this chapter is to generate cuts for the set $\mathcal{F}$ using the set $\mathcal{Q}$ in (4.7). This can be achieved by using the affine transformation $x = x^0 + Hw$. To justify this, we need to show first that the cuts derived for $\mathcal{Q}$ are effective on excluding the undesired solutions x in $\mathcal{F}$.

Lemma 4.1. *Given a vector $\hat{x} \in \mathcal{F}$ and a vector $\hat{w} \in \mathcal{Q}$ such that $x = x^0 + H\hat{z}$, a cut cuts off the vector $\hat{x}$ from $\mathcal{F}$ if it cuts off $\hat{w}$ from $\mathcal{Q}$.*

Proof. Recall the alternative representation of $\mathcal{F}$ given in (4.8). Note that any $x \in \mathcal{F}$ is define by linear combination of x^0 and the columns of H. Additionally, recall that the columns of H are linearly independent. Then, the vector $\hat{w}$ defining $\hat{x}$ is unique. Therefore, given a cut that excludes $\hat{w}$ from $\mathcal{Q}$, it excludes $\hat{x}$ from $\mathcal{F}$. □

Now, the cuts presented in this chapter are disjunctive cuts for problem (4.2). In other words, the disjunctive cuts introduced in this chapter result from convexification of the set

$$\{x \in \mathbb{R}^n \mid Ax = b,\ x \in \mathbb{L},\ x \in \mathcal{U} \cup \mathcal{V}\}. \tag{4.9}$$

We show that the set $\mathcal{U} \cup \mathcal{V}$ has an equivalent representation in terms of w. Note that the sets $\mathcal{U}$ and $\mathcal{V}$ admit the following equivalent representation

$$\mathcal{U} = \{x \in \mathbb{R}^n \mid \exists w \in \mathbb{R}^n,\ x = x^0 + Hw,\ u^\top Hw \geq \varphi - u^\top x^0\},$$

and

$$\mathcal{V} = \{x \in \mathbb{R}^n \mid \exists w \in \mathbb{R}^n,\ x = x^0 + Hw,\ v^\top Hw \leq \varpi - v^\top x^0\}.$$

Define $a = u^\top H$, $d = v^\top H$, $\alpha = \varphi - u^\top x^0$ and $\beta = \varpi - v^\top x^0$. Now, let $\mathcal{A} = \{w \in \mathbb{R}^\ell \mid a^\top w = \alpha\}$ and $\mathcal{B} = \{w \in \mathbb{R}^\ell \mid d^\top w = \beta\}$. Therefore, we can rewrite (4.9) in terms of the set $\mathcal{A} \cup \mathcal{B}$ as follows

$$\{x \in \mathbb{R}^n \mid x = x^0 + Hw, \text{ with } w \in \mathcal{Q}, \ x_1 \geq 0, \text{ and } w \in \mathcal{A} \cup \mathcal{B}\}. \tag{4.10}$$

Thus, for the generation of the DCCs introduced in this chapter we will focus on convexifying the set $\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$.

Let us now focus on the shapes of the set $\mathcal{Q}$. Recall from Section 1.1.2 that the inertia of matrix the P in the representation of $\mathcal{Q}$ is one of the elements defining its shape. We have the following result about the inertia of P.

Lemma 4.2. *The matrix P in the representation of the quadric $\mathcal{Q}$ has at most one non-positive eigenvalue, and at least $\ell - 1$ positive eigenvalues.*

Proof. First of all, we have that

$$J = I - 2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Now, recall that the columns of H form an orthonormal basis for the null space of A. Thus

$$P = H^\top J H = I - 2H^\top \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} H = I - 2H_{1:} H_{1:}^\top,$$

where $H_{1:}$ is the first row of H. From Section 3.2.2 we know that 1 is an eigenvalue of P with multiplicity $\ell - 1$, and the last eigenvalue is $1 - 2H_{1:}^\top H_{1:}$. Thus, P has at most one non-positive eigenvalue. □

Hence, from Lemma 4.2 we have that the shapes of the quadric $\mathcal{Q}$ are limited to those

described in Section 1.1.2.1. However, there are some shapes considered in Section 1.1.2.1 that the quadric $\mathcal{Q}$ will never take. For that reason, we need to refine further the analysis of possible shapes for $\mathcal{Q}$. We may assume that $\mathcal{Q}$ is not an empty set, otherwise there is no need for classification. Now, for the analysis of the shapes of $\mathcal{Q}$ we need the following. First, recall that $Ax^0 = b$, then the system $Hw = x^0$ will have a solution if and only if $b = 0$. Second, recall that $P = H^\top JH$, and that $H_{1:}$ is the first row of H. Then we have that

$$PH_{1:} = (H^\top JH)H_{1:} = (I - 2H_{1:}H_{1:}^\top)H_{1:} = (1 - 2H_{1:}^\top H_{1:})H_{1:}.$$

As a result, we obtain that $H_{1:}$ is an eigenvector of P associated with the eigenvalue $(1 - 2H_{1:}^\top H_{1:})$. Third, let us define the set

$$\mathcal{F}^r = \{x \in \mathbb{R}^n \mid Ax = b, x^\top Jx \leq 0\}, \tag{4.11}$$

which is a relaxation of $\mathcal{F}$. Note that due to the constraint $x^\top Jx \leq 0$, the set $\mathcal{F}^r$ contains a whole line if and only if the zero vector is an element of $\mathcal{F}^r$, i.e., if $b = 0$. We divide the classification of the shapes of $\mathcal{Q}$ in two cases: P is singular, and P is non-singular.

Let us begin classifying the shapes of $\mathcal{Q}$ when P is singular. First of all, from Lemma 4.2 we know that if P is singular, then $P \succeq 0$ and $(1 - 2H_{1:}^\top H_{1:}) = 0$. Consequently, $H_{1:}$ is an eigenvector of P associated with its zero eigenvalue. Now, from Section 1.1.2.1 we know that $\mathcal{Q}$ may be a paraboloid or a cylinder. The criteria to decide this is if the system $Pw = -p$ is solvable. On one hand, if $Pw = -p$ has not solution, then we obtain that $\mathcal{Q}$ is a paraboloid. On the other hand, if the system $Pw = -p$ is solvable, then $\mathcal{Q}$ is a cylinder. We show now that if $Pw = -p$, then in our context $\mathcal{Q}$ is always a line, i.e., a cylinder which base is a point.

Let $w^{\mathrm{c}} \in \mathbb{R}^\ell$ be such that $Pw^{\mathrm{c}} = -p$. First, we show that w^{c} can be found by solving the system $Hw = -x^0$. Consider the vector $w^{\mathrm{c}} + \sigma H_{1:}$, where σ is a scalar. Note that

$(w^c + \sigma H_{1:}) \in \mathcal{Q}$ for any $\sigma \in \mathbb{R}$, because

$$(w^c + \sigma H_{1:})^\top P(w^c + \sigma H_{1:}) + 2p^\top(w^c + \sigma H_{1:}) + \rho = (w^c)^\top P w^c + 2p^\top w^c + \rho \leq 0.$$

The last inequality follows from the assumption that $\mathcal{Q}$ is not an empty set. Thus, $H_{1:}$ is the vector defining the direction of the cylinder $\mathcal{Q}$. Now, let us define the set

$$\mathcal{S} = \{x \in \mathbb{R}^n \mid x = x^0 + H(w^c + \sigma H_{1:}),\ \sigma \in \mathbb{R}\}.$$

If $\mathcal{Q}$ is a cylinder, then $\mathcal{S} \subseteq \mathcal{F}^r$. Additionally, $\mathcal{S} \subseteq \mathcal{F}^r$ only if $b = 0$. Now, if $b = 0$, then the system $Hw = -x^0$ is solvable because $-x^0$ is in the null space of A. Let w^c be the solution of $Hw = -x^0$, which is unique since the columns of H are linearly independent. Then, we have that $Pw^c = H^\top JHw^c = -H^\top Jx^0$. Furthermore, we have that $(w^c)^\top Pw^c - \rho = (w^c)^\top H^\top JHw^c - (x^0)^\top Jx^0 = 0$. Hence, if the system $Pw = -p$ is solvable, then from Section 1.1.2.1 we know that $\mathcal{Q}$ is a line.

We now classify the shapes of $\mathcal{Q}$ when P is non-singular. Recall that (4.6) is equivalent to

$$\left(w + P^{-1}p\right)^\top P \left(w + P^{-1}p\right) \leq p^\top P^{-1}p - \rho. \tag{4.12}$$

Additionally, recall from Section 1.1.2.1 that the shape of $\mathcal{Q}$ in this case is determined by the inertia of P and the value of the right hand side of (4.12). First of all, we know that if $P \succ 0$, then $\mathcal{Q}$ is an ellipsoid. Thus, to complete the classification of $\mathcal{Q}$ we need to consider the case when P is an ID1 matrix. Recall from Section 1.1.2.1 that in this case, we have the following cases:

- if $p^\top P^{-1}p - \rho \leq 0$, then $\mathcal{Q}$ is a hyperboloid of two sheets;
- if $p^\top P^{-1}p - \rho = 0$, then $\mathcal{Q}$ is a scaled second order cone;
- if $p^\top P^{-1}p - \rho \geq 0$, then $\mathcal{Q}$ is a hyperboloid of one sheet.

We show here that in our context $p^\top P^{-1}p - \rho \leq 0$. In other words, we show that $\mathcal{Q}$ is never a hyperboloid of one sheet.

The vector $-P^{-1}p$ is either the vertex of a scaled second order cone or the intersection of the asymptotes of a hyperboloid. Even more, if $\mathcal{Q}$ is a cone or a hyperboloid of one sheet, then $-P^{-1}p \in \mathcal{Q}$. On the other hand, if $\mathcal{Q}$ is a hyperboloid of two sheets, then $-P^{-1}p \notin \mathcal{Q}$. Now, from Lemma 4.2 we know that if P is ID1, then $(1 - 2H_{1:}^\top H_{1:}) < 0$. Consequently, $H_{1:}$ is an eigenvector of P associated with its negative eigenvalue. Consider the vector $(-P^{-1}p + \sigma H_{1:})$ for $\sigma \in \mathbb{R}$, then

$$(-P^{-1}p + \sigma H_{1:} + P^{-1}p)^\top P(-P^{-1}p + \sigma H_{1:} + P^{-1}p) = \sigma^2 H_{1:}^\top P H_{1:} \leq 0.$$

Define the set

$$\mathcal{T} = \{x \in \mathbb{R}^n \mid x = x^0 + H(-P^{-1}p + \sigma H_{1:}),\ \sigma \in \mathbb{R}\}$$

Hence, $\mathcal{T} \subset \mathcal{F}^r$ if and only if $p^\top P^{-1}p - \rho \geq 0$. Recall that $\mathcal{T} \subseteq \mathcal{F}^r$ if and only if $b = 0$, and if $b = 0$, then there is a vector $w^{\mathrm{c}} \in \mathbb{R}^\ell$ such that $Hw^{\mathrm{c}} = -x^0$. Further, we have that

$$\begin{aligned}
(w^{\mathrm{c}} + P^{-1}p)^\top P(w^{\mathrm{c}} + P^{-1}p) &= (w^{\mathrm{c}})^\top P w^{\mathrm{c}} + 2p^\top P^{-1} w^{\mathrm{c}} + p^\top P^{-1}p \\
&= (w^{\mathrm{c}})^\top H^\top J H w^{\mathrm{c}} + 2(x^0)^\top J H w^{\mathrm{c}} + p^\top P^{-1}p \\
&= p^\top P^{-1}p + (x^0)^\top J x^0 - 2(x^0)^\top J x^0 \\
&= p^\top P^{-1}p - \rho.
\end{aligned}$$

On the other hand, we have that

$$P(w^{\mathrm{c}} + P^{-1}p) = P w^{\mathrm{c}} + p = H^\top J H w^{\mathrm{c}} + H^\top J x^0 = -H^\top J x^0 + H^\top J x^0 = 0.$$

Hence, if $b = 0$, then $p^\top P^{-1}p - \rho = 0$. As a conclusion, we have that the quadric $\mathcal{Q}$ cannot

be a hyperboloid of one sheet.

In summary, we have the following possible shapes for $\mathcal{Q}$:

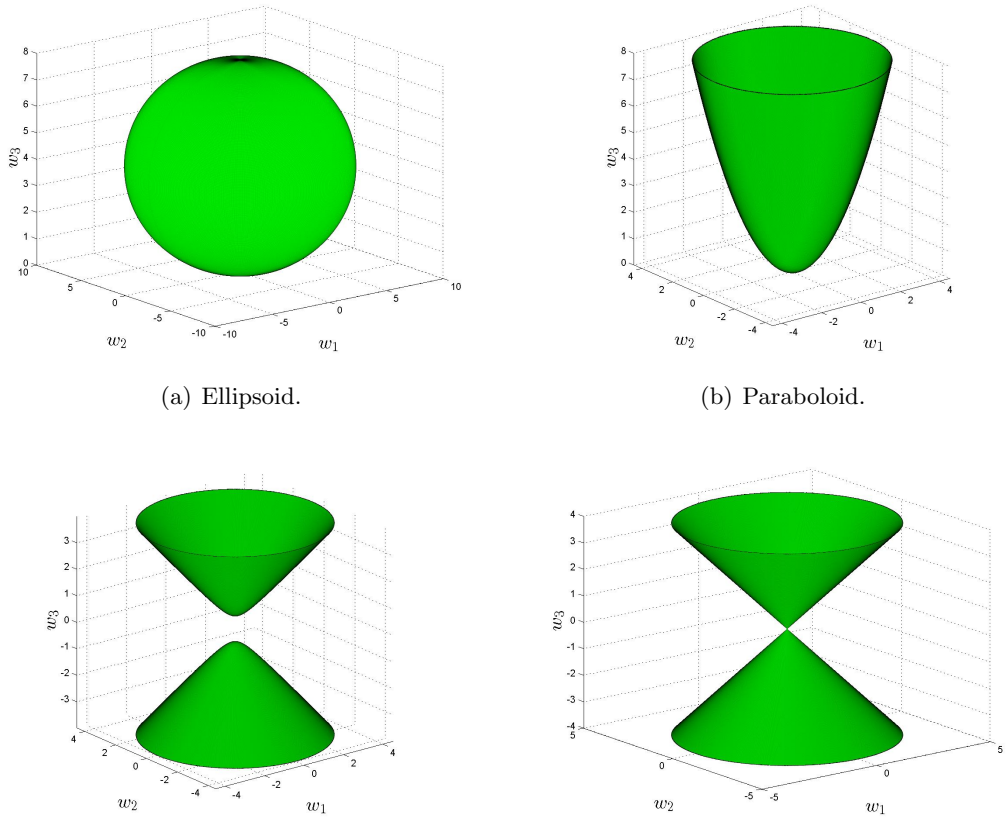
- if $P \succ 0$, then $\mathcal{Q}$ is an ellipsoid, see Figure 4.1(a) for an illustration;
- if $P \succeq 0$ and singular, then $\mathcal{Q}$ is:
  - ◇ a parabolid if there is no vector $w^c \in \mathbb{R}^\ell$ such that $Pw^c = -p$, see Figure 4.1(b) for an illustration;
  - ◇ a line if there is a vector $w^c \in \mathbb{R}^\ell$ such that $Pw^c = -p$;
- if P is ID1, then $\mathcal{Q}$ is:
  - ◇ a hyperboloid of two sheets if $p^\top P^{-1} p - \rho < 0$, see Figure 4.1(c) for an illustration;
  - ◇ a cone if $p^\top P^{-1} p - \rho = 0$, see Figure 4.1(d) for an illustration.

4.2 Building a disjunctive conic cut with parallel disjunctions

Before describing the procedure let us illustrate the concept of building a DCC cut with the following MISOCO problem

$$\begin{array}{rrrrrl} \text{minimize:} & 3x_1 & +2x_2 & +2x_3 & +x_4 & \\ \text{subject to:} & 9x_1 & +x_2 & +x_3 & +x_4 & = 10 \\ & & (x_1, x_2, x_3, x_4) & & & \in \mathbb{L}^4 \\ & & & & x_4 & \in \mathbb{Z}. \end{array} \tag{4.13}$$

Let $\mathcal{F}$ denote the feasible set of this problem. The quadric $\mathcal{Q} \in \mathbb{R}^3$ associated with the feasible set $\mathcal{F}$ is an ellipsoid. In this case $\mathcal{Q}$ can be written in terms of the variables x_2,



(a) Ellipsoid.

(b) Paraboloid.

(c) Hyperboloid.

(d) Cone.

Figure 4.1: Illustration of the shapes of $\mathcal{Q}$.

x_3, x_4, and the problem can be reformulated as follows

$$\begin{aligned} \text{minimize:} \quad & \tfrac{1}{3}\left(10 + 5x_2 + 5x_3 + 2x_4\right) \\ \text{subject to:} \quad & \begin{bmatrix} x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} 8 & -\frac{1}{10} & -\frac{1}{10} \\ -\frac{1}{10} & 8 & -\frac{1}{10} \\ -\frac{1}{10} & -\frac{1}{10} & 8 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} + 2 \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} - 10 \quad \leq 0 \\ & x_4 \in \mathbb{Z}. \end{aligned} \tag{4.14}$$

The feasible set of the continuous relaxation of the reformulation is shown in Figure 4.2.

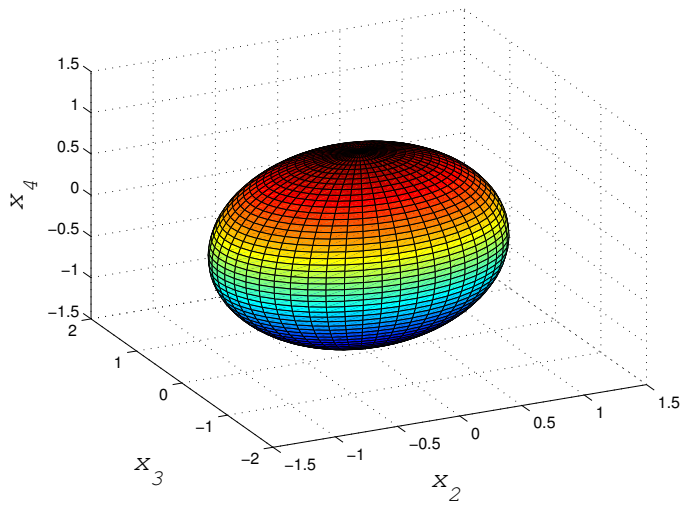


Figure 4.2: Feasible region of the reformulation of Problem (4.13).

We can build a disjunctive cut for Problem (4.14) in three steps. First, we relax the constraint $x_4 \in \mathbb{Z}$ and solve the SOCO problem associated with (4.13). The optimal solution of this relaxation is

$$x^*_{\text{soco}} = \begin{bmatrix} 1.36 & -0.91 & -0.91 & -0.45 \end{bmatrix},$$

with and optimal objective value $\zeta^*_{\text{soco}} = 0.00$. Clearly, $x^*_{\text{soco}} \notin \mathcal{F}$, since the variable x_4 takes a fractional value.

Second, we identify a disjunction that is violated by x^*_{soco}. The solution of (4.14) must be contained either in the set $\mathcal{A} = \{x \in \mathbb{R}^4 \mid x_4 \geq 0\}$ or in the set $\mathcal{B} = \{x \in \mathbb{R}^4 \mid x_4 \leq -1\}$. Hence, the disjunction $\mathcal{A} \vee \mathcal{B}$ is violated by the solution of the continuous relaxation x^*_{soco}, which is illustrated in Figure 4.3.

Third, we convexify the non convex set $\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$. In particular, we want to find the set $\text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$. Recall from Theorem 1.4 that this is the smallest convex set containing $\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$. Additionally, we have that $\mathcal{F} \subset \text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$. In this case, this

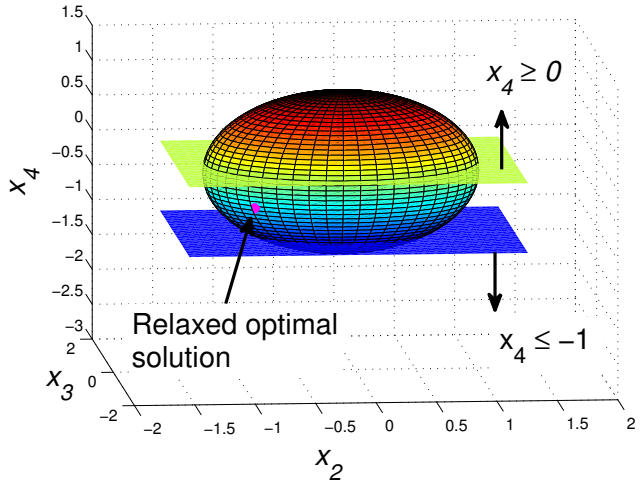


Figure 4.3: Disjunction violated by the solution x^*_{soco} to the continuous relaxation of (4.14).

is the tightest convex formulation we can obtain for the continuous relaxation of Problem (4.14). To find $\text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$, in this example it is enough to add a conic constrain to Problem (4.14). Figure 4.4 illustrates the addition of a conic constraint to the formulation of Problem (4.14).

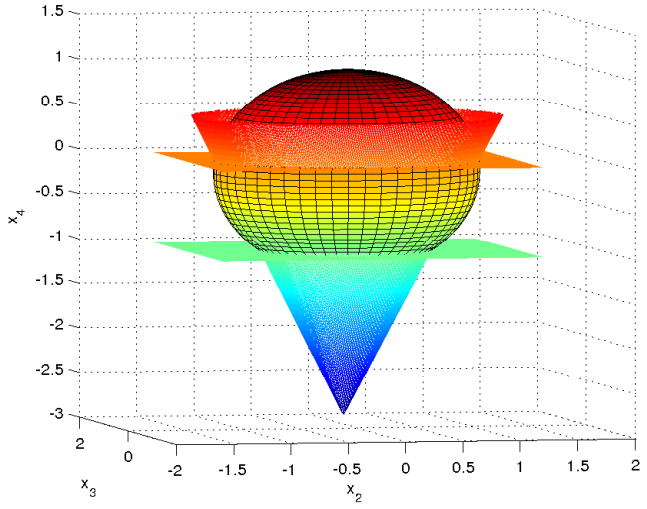


Figure 4.4: Scaled second order cone as a constrain of Problem (4.15)

The new formulation of Problem (4.14) after adding the conic cut is given by

$$
\begin{array}{lrrrrrrrl}
\text{min:} & 3x_1 & +2x_2 & +2x_3 & +x_4 & & & & \\
\text{s.t:} & 9x_1 & +x_2 & +x_3 & +x_4 & & & & = 10 \\
 & & -0.04x_2 & -0.04x_3 & -3.56x_4 & +x_5 & & & = 10.14 \\
 & & -6.28x_2 & -6.28x_3 & +0.14x_4 & & +x_6 & & = 1.65 \\
 & & 6.36x_2 & -6.36x_3 & & & & +x_7 & = 0 \\
 & & & & (x_1, x_2, x_3, x_4) \in \mathbb{L}^4 & & & & \\
 & & & & (x_5, x_6, x_7) \in \mathbb{L}^3 & & & & \\
 & & & & x_4 \in \mathbb{Z} & & & & .
\end{array}
\tag{4.15}
$$

In particular the conic constraint illustrated in Figure 4.4 is described by the constraints

$$
\begin{array}{rrrrrrr}
-0.04x_2 & -0.04x_3 & -3.56x_4 & +x_5 & & & = 10.14 \\
-6.28x_2 & -6.28x_3 & +0.14x_4 & & +x_6 & & = 1.65 \\
6.36x_2 & -6.36x_3 & & & & +x_7 & = 0 \\
 & & (x_5, x_6, x_7) \in \mathbb{L}^3 & & & & .
\end{array}
$$

These constraints define a translated and scaled second order cone in the space of variables x_1, x_2, and x_3, which is a DCC. The feasible set of the continuous relaxation of Problem (4.15) is illustrated in Figure 4.5, which is $\text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$.

Now, if we solve the continuous relaxation of Problem (4.15) we obtain the solution

$$
x^*_{\text{soco}} = \begin{bmatrix} 1.32 & -0.93 & -0.93 & 0.00 & 10.06 & -10.06 & 0.00 \end{bmatrix}
$$

with an optimal objective value $\zeta^*_{\text{soco}} = 0.24$. Note that in this case $x_4 = 0.00$, which is integer. Hence, since $\mathcal{F} \subset \text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$, we have that this is in fact optimal for Problem (4.13).

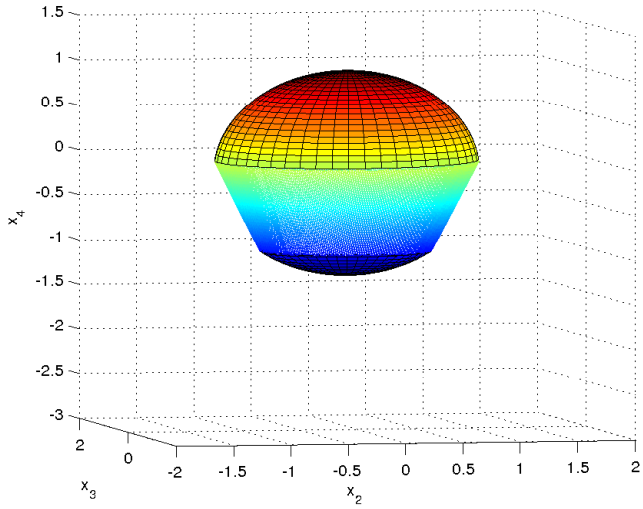


Figure 4.5: Feasible set of Problem (4.15)

Our goal in this section is to show that the procedure described for Problem (4.13) can be generalized. With this on mind, let us review two important characteristics of the cone illustrated in Figure 4.4 and defined by (4.2). First, the constraints in (4.2) are equivalent to the constraint

$$\begin{bmatrix} x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} 8 & -\frac{1}{10} & -\frac{1}{10} \\ -\frac{1}{10} & 8 & -\frac{1}{10} \\ -\frac{1}{10} & -\frac{1}{10} & -1.27 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} + 2 \begin{bmatrix} 1 & 1 & -3.63 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} - 10 \leq 0, \qquad (4.16)$$

which define a quadric in the space of x_2, x_3, and x_4. Second, the quadric in the feasible set of Problem (4.14) and the quadric defined by (4.16) have the same intersection with the hyperplane $\mathcal{A}^{=} = \{[x_2\ x_3\ x_4] \in \mathbb{R}^3 \mid x_4 = 0\}$, which is a quadric in $\mathbb{R}^2$ defined by

$$\begin{bmatrix} x_2 & x_3 \end{bmatrix} \begin{bmatrix} 8 & -\frac{1}{10} \\ -\frac{1}{10} & 8 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + 2 \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - 10 \leq 0.$$

Similarly, both the feasible set and the quadric have the same intersection with the hyperplane $\mathcal{B}^{=} = \{[x_2\ x_3\ x_4] \in \mathbb{R}^3 \mid x_4 = -1\}$, which is a quadric in $\mathbb{R}^2$ defined by

$$\begin{bmatrix} x_2 & x_3 \end{bmatrix} \begin{bmatrix} 8 & -\frac{1}{10} \\ -\frac{1}{10} & 8 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + 2 \begin{bmatrix} \frac{8}{10} & \frac{8}{10} \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - 4 \leq 0.$$

These two characterizations are key for the derivation of the DCCs proposed in this section.

Let us consider now a given quadric $\mathcal{Q}$ associated with Problem (4.2) and a disjunctive set $\mathcal{A} \cup \mathcal{B}$, where $\mathcal{A} = \{x \in \mathbb{R}^n \mid a^\top x \geq \alpha\}$ and $\mathcal{B} = \{x \in \mathbb{R}^n \mid a^\top x \leq \beta\}$. From Chapter 3 we know that it is always possible to find a quadric $\tilde{\mathcal{Q}}$, that is a cone or a cylinder, that has the same intersection with $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$ than $\mathcal{Q}$. Additionally, using the general results of Chapter 2 and some additional analysis for particular cases, we can show that $\mathcal{Q} \cap \tilde{\mathcal{Q}} = \text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$. Hence, it is possible to generalize the derivation of DCCs for MISOCO problems and parallel disjunctions. To simplify the algebra, we may assume w.l.o.g. that the quadric $\mathcal{Q}$ has been transformed using the affine transformation described in Section 3.1. Additionally, we may assume w.l.o.g. that $\|a\| = 1$. We separate the analysis in two cases. In Section 4.2.1 we study the cylinder case, and in Section 4.2.2 the cone case.

4.2.1 Cylinders

We begin with the analysis of the cases in which we can derive DCyCs. We need to consider four cases, given the possible shapes of the quadric $\mathcal{Q}$ as listed in Section 4.1. Hence, we have that $\mathcal{Q}$ can be an ellipsoid, a paraboloid, a scaled second order cone, or a hyperboloid of two sheets. Recall from Theorem 3.1 that given a disjunctive set $\mathcal{A} \cup \mathcal{B}$, we can define a family of quadrics $\mathcal{Q}(\tau)$, $\tau \in \mathbb{R}$, having the same intersection with the hyperplanes $\mathcal{A}^{=}$ and $\mathcal{B}^{=}$, as $\mathcal{Q} \in \mathcal{Q}(\tau)$ has. In particular, in Chapter 3 we can find a characterization of a family where the DCyC is a cylinder for each of the cases considered here. Hence, we

have that:

- if $\mathcal{Q}$ is an ellipsoid, then we need to consider the third and fourth cases in Theorem 3.2;
- if $\mathcal{Q}$ is a paraboloid, then we need to consider the family in Lemma 3.4;
- if $\mathcal{Q}$ is a hyperboloid of two sheets, we need to consider the second case in Theorem 3.4;
- if $\mathcal{Q}$ is a cone, then we need to consider the second case in Theorem 3.3.

An important observation in all these cases is that in the associated family of quadrics $\mathcal{Q}(\tau)$ there exist only one cylinder. These cylinders are fully identified in each of the cases mentioned. Hence, the next step is to verify that they are DCyCs for Problem (4.1). The verification divides the cases in two groups. First we consider the cases in Theorems 3.2 and 3.4, which are convex cylinders. Second, we consider the cases in Theorems 3.4 and 3.3, which are hyperbolic cylinders of two sheets.

4.2.1.1 Case when the DCyC is an ellipsoidal or parabolic cylinder

We start with the verification that the second and fourth cases in Theorem 3.2, and the case in Lemma 3.4 are DCyCs for Problem (4.1) when $\mathcal{Q}$ is either an ellipsoid or a paraboloid, respectively. In these cases the quadric given by $\mathcal{Q}(-1)$ is a cylinder. Furthermore, we know that $P(-1) \succeq 0$, thus they are convex cylinders. On the other hand, in the cases in Theorem 3.2 the vector a defines the direction of the cylinder. Additionally, in the case of Lemma 3.4 we know that $a_1 = 0$ and that the direction of the cylinder is $[\gamma \; a_{2:n}^\top]^\top$. Hence, in both cases the product of the normal vector of the hyperplanes a with the direction of the cylinder is different from 0. Hence, it follows from Theorem 2.2 and Lemma 4.1 that $\mathcal{Q}(-1)$ is a DCyC for Problem (4.1). Finally, since $P(-1) \succeq 0$ this cylinders are second order cone representable, see for example Ben-Tal and Nemirovski [2001a].

4.2.1.2 Case when the DCyC is a branch of a hyperbolic cylinder

We verify now that the second cases of Theorems 3.4 and 3.3 are DCyCs for Problem (4.1) when $\mathcal{Q}$ is either a cone or a hyperboloid of two sheets, respectively. In these cases the quadric $\mathcal{Q}(\hat{\tau})$, where $\hat{\tau} = -\frac{1}{(1-2a_1^2)}$, is a hyperbolic cylinder of two sheets, which is a non-convex quadric. Even more, recall that both cases of Theorems 3.3 and 3.4 happen only when $\alpha = -\beta$, then we have that $p(\hat{\tau}) = 0$, $\rho(\hat{\tau}) > 0$ and

$$\mathcal{Q}(\hat{\tau}) = \{x \in \mathbb{R}^\ell \mid x^\top P(\hat{\tau}) x \leq -\rho(\hat{\tau})\}.$$

Consider the eigenvalue decomposition $P(\hat{\tau}) = V(\hat{\tau})D(\hat{\tau})V(\hat{\tau})^\top$, where $D(\hat{\tau}) \in \mathbb{R}^{\ell\times\ell}$, and $V(\hat{\tau}) \in \mathbb{R}^{\ell\times\ell}$ is non-sigular. We may assume w.l.o.g. that $D_{1,1} = -1$, $D_{2,2} = 0$, and $D_{i,i} > 0$, $i \in \{3, \ldots, n\}$. Now, let $W(\hat{\tau}) = V(\hat{\tau})\bar{D}(\hat{\tau})^{\frac{1}{2}}$, where $\bar{D}(\hat{\tau})$ is a diagonal matrix such that $\bar{D}(\hat{\tau})_{i,i} = |D(\hat{\tau})_{i,i}|$. Let $W(\hat{\tau})_{3:n}$ be the matrix that has the last $n-2$ columns of $W(\hat{\tau})$, and $W(\hat{\tau})_1$ the firs column of $W(\hat{\tau})$. Then,

$$\mathcal{Q}(\hat{\tau}) = \left\{ x \in \mathbb{R}^\ell \;\middle|\; \left\| W(\hat{\tau})_{3:n}^\top x \right\|^2 \leq -\rho(\hat{\tau}) + \left(W(\hat{\tau})_1^\top x \right)^2 \right\}.$$

Let us define the following two sets

$$\mathcal{Q}^+(\hat{\tau}) = \left\{ x \in \mathbb{R}^n \;\middle|\; \left\| W(\hat{\tau})_{3:n}^\top x \right\| \leq \xi,\ \left\| \begin{bmatrix} \xi & \sqrt{\rho(\hat{\tau})} \end{bmatrix}^\top \right\| \leq W(\hat{\tau})_1^\top x \right\},$$
$$\mathcal{Q}^-(\hat{\tau}) = \left\{ x \in \mathbb{R}^n \;\middle|\; \left\| W(\hat{\tau})_{3:n}^\top x \right\| \leq \xi,\ \left\| \begin{bmatrix} \xi & \sqrt{\rho(\hat{\tau})} \end{bmatrix}^\top \right\| \leq -W(\hat{\tau})_1^\top x \right\}.$$

Thus, $\mathcal{Q}(\hat{\tau}) = \mathcal{Q}^+(\hat{\tau}) \cup \mathcal{Q}^-(\hat{\tau})$, and each of these branches of $\mathcal{Q}(\hat{\tau})$ are a convex cylinders in the direction $V(\hat{\tau})_2$, which is the 2nd column of $V(\hat{\tau})$. Also, note that $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{Q}^-(\hat{\tau}) = \emptyset$. Recall that in the cases we are analyzing the set $\mathcal{Q}$ is either a cone or a hyperboloid of two sheets. Let $\mathcal{Q}^+$ be the set of vectors in $\mathcal{Q}$ such that for $w \in \mathcal{Q}^+$ we have $x_0 + Hw \in \mathcal{F}$.

Similarly, let $\mathcal{Q}^-$ be the set of vectors in $\mathcal{Q}$ such that for $w \in \mathcal{Q}^-$ we have $x_0 + Hw \notin \mathcal{F}$. Hence, from equation (4.8) we know that $\mathcal{Q}^+$ and $\mathcal{Q}^-$ are the two branches of $\mathcal{Q}$. The last statement follows from the fact that $H_{1:}$ is an eigenvector associated with the negative eigenvalue of $\mathcal{Q}$. Hence, $\mathcal{Q}^+$ and $\mathcal{Q}^-$ are convex sets and $\mathcal{Q}^+ \cup \mathcal{Q}^- = \mathcal{Q}$. We have the following result.

Lemma 4.3. *In the second case in Theorem 3.3 and the second case in Theorem 3.4 the set* $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}^+$ *is a subset of a single branch of* $\mathcal{Q}(\hat{\tau})$.

Proof. We show that if the set $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}^+$ is not a subset of a single branch of $\mathcal{Q}(\hat{\tau})$, then $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q} = (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$ is contradicted.

Let $u, v \in (A^= \cup \mathcal{B}^=) \cap \mathcal{Q}^+$ be two vectors such that $u \in \mathcal{Q}^+(\hat{\tau})$ and $v \in \mathcal{Q}^-(\hat{\tau})$. Note that if $a^\top u = \alpha$ and $a^\top v = \alpha$, or $a^\top u = \beta$ and $a^\top v = \beta$, then in that case there must exists a $0 \leq \tilde{\lambda} \leq 1$ such that $w = \tilde{\lambda}v + (1-\tilde{\lambda})u \in (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}^+$ but $w \notin (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$. This statement is true because $\mathcal{Q}^+$, $\mathcal{Q}^+(\hat{\tau})$, and $\mathcal{Q}^-(\hat{\tau})$ are convex, and $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{Q}^-(\hat{\tau}) = \emptyset$.

Now, assume that $a^\top u = \alpha$ and $a^\top v = \beta$ and let $\tilde{a} = [-a_1 \; a_{2:n}^\top]^\top$. Recall from Section 3.2.5 that $P(\hat{\tau}) = \tilde{J} - \frac{aa^\top}{(1-2a_1^2)}$, then for any $\theta \in \mathbb{R}$

$$(v + \theta\tilde{a})^\top P(\hat{\tau})(v + \theta\tilde{a}) + \rho(\hat{\tau}) = v^\top P(\hat{\tau})v + \rho(\hat{\tau}) \leq 0.$$

Similarly,

$$(u + \theta\tilde{a})^\top P(\hat{\tau})(u + \theta\tilde{a}) + \rho(\hat{\tau}) = u^\top P(\hat{\tau})u + \rho(\hat{\tau}) \leq 0,$$

Additionally, since $a^\top \tilde{a} \neq 0$, then $\exists\hat{\theta}$ such that $a^\top(u + \hat{\theta}\tilde{a}) = \beta$, and $\exists\tilde{\theta}$ such that $a^\top(v + \tilde{\theta}\tilde{a}) = \alpha$. Hence, $\mathcal{Q}^-(\hat{\tau}) \cap \mathcal{A}^= \neq \emptyset$ and $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{B}^= \neq \emptyset$.

Now, we show that $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{B}^= \cap \mathcal{Q}^+ = \emptyset$ and $\mathcal{Q}^-(\hat{\tau}) \cap \mathcal{A}^= \cap \mathcal{Q}^+ = \emptyset$. Assume to the contrary that $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{B}^= \cap \mathcal{Q}^+ \neq \emptyset$. Then, for any $s \in \mathcal{Q}^+(\hat{\tau}) \cap \mathcal{B}^= \cap \mathcal{Q}^+$ there must exists a $0 \leq \tilde{\lambda} \leq 1$ such that $w = \tilde{\lambda}s + (1 - \tilde{\lambda})v \in (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}^+$ but $w \notin (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$. This is true because $\mathcal{Q}^+$ is convex, $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{Q}^-(\hat{\tau}) = \emptyset$, $v \in \mathcal{Q}^-(\hat{\tau})$, and $a^\top v = \beta$.

A similar contradiction would be obtain if $\mathcal{Q}^-(\hat{\tau}) \cap \mathcal{A}^= \cap \mathcal{Q}^+ \neq \emptyset$. Hence, we have that $\mathcal{Q}^+(\hat{\tau}) \cap \mathcal{B}^= \cap \mathcal{Q}^- \neq \emptyset$ and $\mathcal{Q}^-(\hat{\tau}) \cap \mathcal{A}^= \cap \mathcal{Q}^- \neq \emptyset$, because $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q} = (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$.

Let $w \in \mathcal{Q}^+(\hat{\tau}) \cap \mathcal{B}^= \cap \mathcal{Q}^-$. Then, $\lambda w + (1-\lambda)u \in \mathcal{Q}^+(\hat{\tau})$ for $0 \leq \lambda \leq 1$, since $\mathcal{Q}^+(\hat{\tau})$ is convex. Now, if $\mathcal{Q}$ is a hyperboloid, then there exist a $0 \leq \tilde{\lambda} \leq 1$ such that $\tilde{\lambda} w + (1-\tilde{\lambda})u \notin \mathcal{Q}$, because $u \in \mathcal{Q}^+$ and $w \in \mathcal{Q}^-$. This contradicts $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q} = (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$. On the other hand, if $\mathcal{Q}$ is a cone, then there exist a $\tilde{\lambda}$ such that either $\tilde{\lambda} w + (1-\tilde{\lambda})u \notin \mathcal{Q}$ or $\tilde{\lambda} w + (1-\tilde{\lambda})u$ is the zero vector. In the first case, we find a contradiction to $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q} = (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$ again. In the second case, let us consider a vector $s \in \mathcal{Q}^-(\hat{\tau}) \cap \mathcal{A}^= \cap \mathcal{Q}^-$. Then, $\lambda s + (1-\lambda)v \in \mathcal{Q}^-(\hat{\tau})$ for $0 \leq \lambda \leq 1$, since $\mathcal{Q}^-(\hat{\tau})$ is convex. In this case, then there exist a $\bar{\lambda}$ such that $\bar{\lambda} s + (1-\bar{\lambda})v \notin \mathcal{Q}$. The last statement is true because $v \in \mathcal{Q}^+$ and $s \in \mathcal{Q}^-$, the zero vector is in $\mathcal{Q}^+(\hat{\tau})$ and $\mathcal{Q}^-(\hat{\tau}) \cap \mathcal{Q}^+(\hat{\tau}) = \emptyset$. This contradicts $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q} = (\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}(\hat{\tau})$ again. □

From Proposition 2.2 and Lemma 4.3, we know that the branch of $\mathcal{Q}(\hat{\tau})$ containing the set $(\mathcal{A}^= \cup \mathcal{B}^=) \cap \mathcal{Q}^+$ is a DCyC for Problem (4.1). Finally, we need to define a criteria to identify the branch of $\mathcal{Q}(\hat{\tau})$ that defines the cylindrical cut. First, consider the case when $\mathcal{Q}^+ = \{x \in \mathbb{R}^\ell \mid x \in \mathcal{Q}, x_1 \geq 0\}$. Then, if $W(\hat{\tau})_1^\top e_1 \geq 0$, then the cylindrical cut is given by $\mathcal{Q}^+(\hat{\tau})$. On the other hand, if $W(\hat{\tau})_1^\top e_1 \leq 0$, then the cylindrical cut is given by $\mathcal{Q}^-(\hat{\tau})$. Now, consider the case when $\mathcal{Q}^+ = \{x \in \mathbb{R}^\ell \mid x \in \mathcal{Q}, x_1 \leq 0\}$. Then, if $-W(\hat{\tau})_1^\top e_k \geq 0$, then the cylindrical cut is given by $\mathcal{Q}^+(\hat{\tau})$. On the other hand, if $-W(\hat{\tau})_1^\top e_1 \leq 0$, then the cylindrical cut is given by $\mathcal{Q}^-(\hat{\tau})$.

4.2.2 Cones

We analyze the cases for which we can derive DCCs. Given the possible shapes of the quadric $\mathcal{Q}$ as listed in Section 4.1, we need to consider the cases when $\mathcal{Q}$ is an ellipsoid, a cone, or a hyperboloid of two sheets. Recall from Theorem 3.1 that given the disjunctive set $\mathcal{A} \cup \mathcal{B}$, there is a family of quadrics $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$, having the same intersection with

the hyperplanes $\mathcal{A}^=$ and $\mathcal{B}^=$, as $\mathcal{Q} \in \mathcal{Q}(\tau)$ has. In particular, in Chapter 3 we can find a characterization of the families that can be used to derive DCC for each of the cases considered here. Hence, we have that:

- if $\mathcal{Q}$ is an ellipsoid, then we need to consider the first and second cases in Theorem 3.2;
- if $\mathcal{Q}$ is a hyperboloid of two sheets, we need to consider the first case in Theorem 3.4;
- if $\mathcal{Q}$ is a cone, then we need to consider the first case in Theorem 3.3.

An important observation in these cases is that in the associated family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of quadrics there exists more than one candidate cone. For that reason we need to analyze the three cases separately.

For the derivation of the cuts, we show the existence of a convex cone in the first cases of Theorems 3.4 and 3.3 that satisfy Proposition 2.1. Recall from Sections 3.2.3 and 3.2.5 that the first cases in Theorems 3.4 and 3.3 consider the quadrics found at the roots of the function $f(\tau)$, which is defined in (3.20) and (3.31), respectively. Also, recall that $\bar{\tau}_1$ and $\bar{\tau}_2$ denote the roots of $f(\tau)$, and that we assume $\bar{\tau}_1 \leq \bar{\tau}_2$. Before analyzing each of the cases considered here, we show that the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of each theorem can be written as the union of two scaled second order cones.

The quadric $\mathcal{Q}(\bar{\tau}_i)$, $i = 1, 2$, is defined by the triplet $(P(\bar{\tau}_i), p(\bar{\tau}_i), \rho(\bar{\tau}_i))$, where $P(\bar{\tau}_i) \in R^{\ell \times \ell}$, $p(\bar{\tau}_i) \in \mathbb{R}^\ell$, and $\rho(\bar{\tau}_i)$ is a scalar. From Sections 3.2.3 and 3.2.5 we know that $\mathcal{Q}(\bar{\tau}_i)$ is a cone, with vertex $x(\bar{\tau}_i) = -Q(\bar{\tau}_i)^{-1} q(\bar{\tau}_i)$. Hence, $P(\bar{\tau}_i)$ is a symmetric non-singular matrix that has exactly one negative eigenvalue and $\ell - 1$ positive eigenvalues. Now, the eigenvalue decomposition of $P(\bar{\tau}_i)$ is given by $V(\bar{\tau}_i) D(\bar{\tau}_i) (V(\bar{\tau}_i)^\top$, where $V(\bar{\tau}_i)$ is an orthonormal matrix and $D(\bar{\tau}_i)$ is a diagonal matrix having the eigenvalues of $P(\bar{\tau}_i)$ in its diagonal. We may assume w.l.o.g. that $D(\bar{\tau}_i)_{1,1} < 0$, and let $W(\bar{\tau}_i) = V(\bar{\tau}_i) \bar{D}(\bar{\tau}_i)^{1/2}$,

where $\bar{D}(\bar{\tau}_i)_{j,k} = |D(\bar{\tau}_i)_{j,k}|$, $j = 1, \dots, \ell$, $k = 1, \dots, \ell$. Thus, we may write $\mathcal{Q}(\bar{\tau}_i)$ in terms of $W(\bar{\tau}_i)$ as follows

$$\left\{ x \in \mathbb{R}^n \mid (x - x(\bar{\tau}_i))^\top W(\bar{\tau}_i)_{2:n} W(\bar{\tau}_i)_{2:n}^\top (x - x(\bar{\tau}_i)) \leq \left(W(\bar{\tau}_i)_1^\top (x - x(\bar{\tau}_i)) \right)^2 \right\},$$

where $W(\bar{\tau}_i)_{2:n}$ has the columns $2, \dots, n$ of $W(\bar{\tau}_i)$ and $W(\bar{\tau}_i)_1$ is the first column of $W(\bar{\tau}_i)$.

Now, let us define the sets $\mathcal{Q}^+(\bar{\tau}_i)$, $\mathcal{Q}^-(\bar{\tau}_i)$ as follows

$$\mathcal{Q}^+(\bar{\tau}_i) = \left\{ x \in \mathbb{R}^n \mid \left\| W(\bar{\tau}_i)_{2:n}^\top (x - x(\bar{\tau}_i)) \right\| \leq W(\bar{\tau}_i)_1^\top (x - x(\bar{\tau}_i)) \right\}, \tag{4.17}$$

$$\mathcal{Q}^-(\bar{\tau}_i) = \left\{ x \in \mathbb{R}^n \mid \left\| W(\bar{\tau}_i)_{2:n}^\top (x - x(\bar{\tau}_i)) \right\| \leq -W(\bar{\tau}_i)_1^\top (x - x(\bar{\tau}_i)) \right\}. \tag{4.18}$$

These sets are two scaled and translated second order cones with vertex $x(\bar{\tau}_i)$, which satisfy Definition 2.1. It is easy to verify that $\mathcal{Q}(\bar{\tau}_i)$ is equal to $\mathcal{Q}^+(\bar{\tau}_i) \cup \mathcal{Q}^-(\bar{\tau}_i)$. Also, it is clear from (4.17) and (4.18) that $\mathcal{Q}^+(\bar{\tau}_i)$, $\mathcal{Q}^-(\bar{\tau}_i)$ are two convex cones. This shows that the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ can be written as the union of two convex cones. Our next step is to define a criteria to identify which of the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ provides a DCC.

First, we show that in the cases of Theorem 3.2, the DCCs are found at the largest root of the polynomial (3.20). Second, we show separately that in the first cases of Theorems 3.4 and 3.3, the DCCs are found at smallest root of the polynomial (3.31).

4.2.2.1 DCC when $\mathcal{Q} \cap \mathcal{A}$ and $\mathcal{Q} \cap \mathcal{B}$ are bounded

We focus here on the DCC that are derived from the first and second cases of Theorem 3.2. In this case we may assume that $\mathcal{Q}$ is an ellipsoid and thus we can use the results on Section 3.2.3. Recall that $P = I$, $p = 0$, $\rho = -1$, $\|a\| = 1$, and that the polynomial in the numerator of (3.19) is

$$f(\tau) = \tau^2 \frac{(\alpha - \beta)^2}{4} + \tau(1 - \alpha\beta) + 1.$$

Given the convex cones $\mathcal{Q}^+(\bar{\tau}_1)$, $\mathcal{Q}^-(\bar{\tau}_1)$, $\mathcal{Q}^+(\bar{\tau}_2)$, and $\mathcal{Q}^-(\bar{\tau}_2)$, we need a criteria to identify which cone gives the convex hull of $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$. First, we decide between the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_1)$. For the proof of the following result we need to show that the vertex $x(\bar{\tau}_2)$ is either in $\mathcal{A}$ or in $\mathcal{B}$. This step is omitted here for the sake of readability, and the details are presented in Lemma A.6 in Appendix A.

Lemma 4.4. *The quadric $\mathcal{Q}(\bar{\tau}_2)$ found at the larger root of $f(\tau)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of the first and second cases of Theorem 3.2 contains a cone that satisfies Definition 2.2.*

Proof. From Theorem 3.1 we know that $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{A}^= = \mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{B}^= = \mathcal{Q} \cap \mathcal{B}^=$. Additionally, we have that $\mathcal{Q}(\bar{\tau}_2) = \mathcal{Q}^+(\bar{\tau}_2) \cup \mathcal{Q}^-(\bar{\tau}_2)$, where $\mathcal{Q}^+(\bar{\tau}_2)$, $\mathcal{Q}^-(\bar{\tau}_2)$ are two convex cones with vertex $x(\bar{\tau}_2)$. From Lemma A.1 we know that the vertex $x(\bar{\tau}_2)$ is either in $\mathcal{A}$ or in $\mathcal{B}$. Thus, since the intersections $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{A}^=$ and $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{B}^=$ are bounded, then one of the following two cases holds:

- Case 1: $\mathcal{Q}^+(\bar{\tau}_2) \cap \mathcal{A}^= = \mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{Q}^+(\bar{\tau}_2) \cap \mathcal{B}^= = \mathcal{E} \cap \mathcal{B}^=$;
- Case 2: $\mathcal{Q}^-(\bar{\tau}_2) \cap \mathcal{A}^= = \mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{Q}^-(\bar{\tau}_2) \cap \mathcal{B}^= = \mathcal{E} \cap \mathcal{B}^=$.

Consequently, we have that one of the cones $\mathcal{Q}^+(\bar{\tau}_2)$ and $\mathcal{Q}^-(\bar{\tau}_2)$ found at the root $\bar{\tau}_2$ satisfy Proposition 2.1. □

This result reduces the choices to the cones $\mathcal{Q}^+(\bar{\tau}_2)$ and $\mathcal{Q}^-(\bar{\tau}_2)$. We need to decide now between the two cones using the sign of $-W(\bar{\tau}_2)_1^\top x(\bar{\tau}_2)$. Thus, we choose $\mathcal{Q}^+(\bar{\tau}_2)$ if $-W(\bar{\tau}_2)_1^\top x(\bar{\tau}_2) > 0$, and we choose $\mathcal{Q}^-(\bar{\tau}_2)$ when $-W(\bar{\tau}_2)_1^\top x(\bar{\tau}_2) < 0$. Finally, it follows from Proposition 2.1 that the selected cone gives a DCC for Problem (4.1). Note that if $x(\bar{\tau}_2) = 0$, then the center of the ellipsoid $\mathcal{Q}$ coincides with the vertex of the selected cone. In this case the feasible set is a single point. Thus, by identifying this unique solution, the problem is solved. This completes the procedure.

4.2.2.2 DCC when $\mathcal{Q} \cap \mathcal{A}$ and $\mathcal{Q} \cap \mathcal{B}$ are unbounded

We focus here on the DCCs that are derived from the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ associate with the first cases of Theorems 3.3 and 3.4.

In the case of Theorem 3.3, we have from Section 3.2.5.5 that $\mathcal{Q} = \{y \in \mathbb{R}^n \mid \|y_{2:n}\|^2 \leq y_1^2\}$. In this case we define $\mathcal{Q}^+ = \{y \in \mathbb{R}^n \mid \|y_{2:n}\| \leq y_1\}$ and $\mathcal{Q}^- = \{y \in \mathbb{R}^n \mid \|y_{2:n}\| \leq -y_1\}$, then $\mathcal{Q} = \mathcal{Q}^+ \cup \mathcal{Q}^-$ and $\mathcal{Q}^+ \cap \mathcal{Q}^- = 0$. Also, note that $\mathcal{Q}^+$ and $\mathcal{Q}^-$ are two second order cones.

In the case of Theorem 3.4, we have from Section 3.2.5.6 that $\mathcal{Q} = \{y \in \mathbb{R}^n \mid \|y_{2:n}\|^2 \leq y_1^2 - 1\}$. In this case, we define $\mathcal{Q}^+ = \{y \in \mathbb{R}^n \mid y^\top y \leq w,\ \|(w, 1)\| \leq y_1\}$ and $\mathcal{Q}^- = \{y \in \mathbb{R}^n \mid y^\top y \leq w,\ \|(w, 1)\| \leq -y_1\}$, then $\mathcal{Q} = \mathcal{Q}^+ \cup \mathcal{Q}^-$ and $\mathcal{Q}^+ \cap \mathcal{Q}^- = \emptyset$. Also, note that $\mathcal{Q}^+$ and $\mathcal{Q}^-$ are two convex sets.

Since the result for cones and hyperboloids of two sheets is the same, we will use $\mathcal{Q}^+$ and $\mathcal{Q}^-$ without making any difference for which set it is defined. Whenever the specification is needed, we make explicit whether the definition corresponds to a cone or a hyperboloid of two sheets. We may assume w.l.o.g. that the feasible set $\mathcal{F}$ of (4.2) is contained in the branch Q^+. Here, we can use the results in Sections 3.2.5.5 and 3.2.5.6 about the characterization of the shapes in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$. Recall that $\bar{\tau}_1$ and $\bar{\tau}_2$ denote the roots of the function $f(\tau)$ in (3.31), and assume $\bar{\tau}_1 \leq \bar{\tau}_2$. Hence, we need a criteria to identify which cone $\mathcal{Q}^+(\bar{\tau}_1)$, $\mathcal{Q}^-(\bar{\tau}_1)$, $\mathcal{Q}^+(\bar{\tau}_2)$, and $\mathcal{Q}^-(\bar{\tau}_2)$, characterizes the convex hull of $\mathcal{E} \cap (\mathcal{A} \cup \mathcal{B})$. For the proof of the next result, we use some intermediate results that are omitted here for the sake of readability. For the reader interested on the details of this steps, they are presented in the appendix in Lemmas A.2, A.3, A.4, A.5.

Theorem 4.1. *The quadric found at the smallest root of $f(\tau)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of the first case of Theorems 3.3 and 3.4 contains a cone that satisfies Definition 2.2.*

Proof. We divide the proof on two parts. First, we show that theorem is true for the first

case of Theorem 3.3 when $0 \in \mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$. Second, we show that the theorem is true when $\mathcal{Q}$ is a hyperboloid of two sheets or $\mathcal{Q}$ is a cone and $0 \notin \mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$.

DCC when $\mathcal{Q}$ is a cone and the vector zero is an element of $\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$: This occurs when α and β have the same sign. Then, the smallest root of $f(\tau)$ in this case is $\bar{\tau}_1 = 0$. Hence, it is enough to show that $\mathcal{Q}^+ = \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$ in this case. First of all, since $\mathcal{Q}^+$ is a convex set, we have that $\text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B})) \subseteq \mathcal{Q}^+$. Thus, to complete the proof of the first part we need to show that $\mathcal{Q}^+ \subseteq \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$. From Definition 1.7 of convex hull it is clear that $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \in \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$. Now, let $\hat{x} \in \mathcal{Q}^+$ be such that $\hat{x} \notin \mathcal{A} \cup \mathcal{B}$. Then, we have that $\beta \leq a^\top \hat{x} = \sigma \leq \alpha$. Assume first that $0 \leq \beta \leq \alpha$, then the vector zero is contained in $\mathcal{B}$. Since $\mathcal{Q}^+$ is a cone, then $\gamma \hat{x} \in \mathcal{Q}^+$ for $\gamma \geq 0$. Now, we have that $a^\top(\gamma \hat{x}) = \gamma\sigma$. Then, for $\gamma^1 = \frac{\alpha}{\sigma}$ we obtain $a^\top(\gamma^1 \hat{x}) = \alpha$, and for $\gamma^2 = \frac{\beta}{\sigma}$ we obtain $a^\top(\gamma^2 \hat{x}) = \beta$. Now, consider the convex combination $\lambda(\gamma^1 \hat{x}) + (1-\lambda)(\gamma^2 \hat{x})$, $0 \leq \lambda \leq 1$. For $\hat{\lambda} = \frac{1-\gamma^2}{\gamma^1 - \gamma^2}$ we obtain that $0 \leq \hat{\lambda} \leq 1$, and $\lambda(\gamma^1 \hat{x}) + (1-\lambda)(\gamma^2 \hat{x}) = \hat{x}$. Since $\gamma^2 \hat{x} \in \mathcal{Q}^+ \cap \mathcal{B}$ and $\gamma^1 \hat{x} \in \mathcal{Q}^+ \cap \mathcal{A}$, then $\hat{x} \in \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$. Now, if $\beta \leq \alpha \leq 0$, it can be shown with a similar argument that $\hat{x} \in \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$. Hence $\mathcal{Q}^+ \subseteq \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$, and it satisfies Definition 2.2, i.e., it is a DCC for $\mathcal{Q}^+$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$.

DCC when $\mathcal{Q}$ is a hyperboloid of two sheets or $\mathcal{Q}$ is a cone and the vector zero is not an element of $\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$: In this case we have from Lemma A.5 that $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \in \mathcal{Q}^+(\bar{\tau}_1)$ or $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \in \mathcal{Q}^-(\bar{\tau}_1)$. Assume w.l.o.g. that $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{Q}^+(\bar{\tau}_1)$. Since $\mathcal{Q}^+(\bar{\tau}_1)$ is a convex set we have that $\text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B})) \subseteq (\mathcal{Q}^+ \cap \mathcal{Q}^+(\bar{\tau}_1)$.

To complete the proof we need to show that $\mathcal{Q}^+ \cap \mathcal{Q}^+(\bar{\tau}) \subseteq \text{conv}((\mathcal{A} \cup \mathcal{B}) \cap \mathcal{Q}^+)$. For this purpose, we need to show first that $\mathcal{Q}^+ \cap \mathcal{A}^= = \mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{A}^=$ and $\mathcal{Q}^+ \cap \mathcal{B}^= = \mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{B}^=$. Observe that $\mathcal{Q}^+ \cap \mathcal{A}^= \subseteq \mathcal{Q}^+(\bar{\tau}_1)$, then $\mathcal{Q}^+ \cap \mathcal{A}^= \subseteq \mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{A}^=$. Thus, it is enough to show that $\mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{A}^= \subseteq \mathcal{Q}^+ \cap \mathcal{A}^=$. Let $u \in Q^+ \cap \mathcal{A}^=$. Recall that $\mathcal{Q} \cap (\mathcal{A}^= \cup \mathcal{B}^=) = \mathcal{Q}(\bar{\tau}_1) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$, hence if $\mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{A}^= \not\subseteq \mathcal{Q}^+ \cap \mathcal{A}^=$, then there exist a

vector $v \in Q^- \cap \mathcal{A}^= \cap \mathcal{Q}^+(\bar{\tau}_1)$. We know that $\mathcal{Q}^+ \cap \mathcal{Q}^- = 0$ if $\mathcal{Q}$ is a cone, and $\mathcal{Q}^+ \cap \mathcal{Q}^- = \emptyset$ if $\mathcal{Q}$ is a hyperboloid of two sheets. Even more, in this case if $\mathcal{Q}$ is a cone, we know that $0 \notin \mathcal{Q} \cap \mathcal{A}^=$. Hence, from Theorem 1.5 there exist a hyperplane $\mathcal{H} = \{x \in \mathbb{R}^\ell \mid h^\top x = \eta\}$ separating $\mathcal{Q}^+$ and $\mathcal{Q}^-$, such that $0 \in \mathcal{H}$. Then, there exist a $0 \leq \lambda \leq 1$ such that $\lambda u + (1-\lambda)v \in \mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{A}^=$ and $h^\top(\lambda u + (1-\lambda)v) = \eta$, i.e., $(\lambda u + (1-\lambda)v) \notin \mathcal{Q}$. This contradicts $\mathcal{Q} \cap (\mathcal{A}^= \cup \mathcal{B}^=) = \mathcal{Q}(\bar{\tau}_1) \cap (\mathcal{A}^= \cup \mathcal{B}^=)$. Hence, $\mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{A}^= \subseteq \mathcal{Q}^+ \cap \mathcal{A}^=$. Similarly, we can show that $\mathcal{Q}^+ \cap \mathcal{B}^= = \mathcal{Q}^-(\bar{\tau}_1) \cap \mathcal{B}^=$.

Now, for any $x \in \mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B})$, we have that $x \in \mathcal{Q}^+ \cap \mathcal{Q}^+(\bar{\tau})$ and $x \in \text{conv}(\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}))$. Next, we need to consider a vector $\tilde{x} \in \mathbb{R}^n$ such that $\tilde{x} \in \mathcal{Q}^+(\bar{\tau}) \cap \mathcal{Q}^+ \cap \bar{\mathcal{A}} \cap \bar{\mathcal{B}}$, where $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ are the complements of $\mathcal{A}$ and $\mathcal{B}$ respectively. From Lemma A.2 we have that $x(\bar{\tau}_1) \in \mathcal{A}$ or $x(\bar{\tau}_1) \in \mathcal{B}$. Assume w.l.o.g. that $x(\bar{\tau}_1) \in \mathcal{B}$. Since $\mathcal{Q}^+(\bar{\tau})$ is a translated cone, then $\{x \in \mathbb{R}^n \mid x = x(\bar{\tau}_1) + \theta(\tilde{x} - x(\bar{\tau}_1)), \theta \geq 0\} \subseteq \mathcal{Q}^+(\bar{\tau})$. Thus, there exist a scalar $0 < \theta_1 < 1$ such that $a^\top(x(\bar{\tau}_1) + \theta_1(x - x(\bar{\tau}_1))) = \beta$ and a scalar $1 < \theta_2$ such that $a^\top(x(\bar{\tau}_1) + \theta_2(x - x(\bar{\tau}_1))) = \alpha$. Let $\lambda = (1-\theta_1)/(\theta_2 - \theta_1)$, then $\tilde{x} = (1-\lambda)(x(\bar{\tau}_1) + \theta_1(x - x(\bar{\tau}_1))) + \lambda(x(\bar{\tau}_1) + \theta_2(x - x(\bar{\tau}_1)))$. Therefore, $\tilde{x} \in \text{conv}((\mathcal{A} \cup \mathcal{B}) \cap \mathcal{Q}^+)$. The same conclusion is found if we assume that $x(\bar{\tau}_1) \in \mathcal{A}$. This proves that $\mathcal{Q}^+ \cap \mathcal{Q}^+(\bar{\tau}_1) \subseteq \text{conv}((\mathcal{A} \cup \mathcal{B}) \cap \mathcal{Q}^+)$. Henceforth, the cone $\mathcal{Q}^+(\bar{\tau}_1)$ is a DCC for $\mathcal{Q}^+$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$. Finally, if $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{Q}^-(\bar{\tau}_1)$, then we can use a similar argument to prove that $\mathcal{Q}^-(\bar{\tau}_1)$ is a DCC for $\mathcal{Q}^+$ and the disjunctive set $\mathcal{A} \cup \mathcal{B}$. □

Now we define a criteria to identify which branch of $\mathcal{Q}(\bar{\tau}_1)$ in Theorem 4.1 defines a DCC. First, consider the case when the feasible set of 4.2 is contained in $\mathcal{Q}^+$. Then, if $W(\hat{\tau})_1^\top e_1 \geq 0$, then the conic cut is given by $\mathcal{Q}^+(\hat{\tau}_1)$. On the other hand, if $W(\hat{\tau})_1^\top e_1 \leq 0$, then the conic cut is given by $\mathcal{Q}^-(\hat{\tau})$. Second, consider the case when the feasible set of (4.2) is contained in $\mathcal{Q}^-$. Then, if $-W(\hat{\tau})_1^\top e_1 \geq 0$, then the conic cut is given by $\mathcal{Q}^+(\hat{\tau})$. On the other hand, if $-W(\hat{\tau})_1^\top e_1 \leq 0$, then the conic cut is given by $\mathcal{Q}^-(\hat{\tau})$.

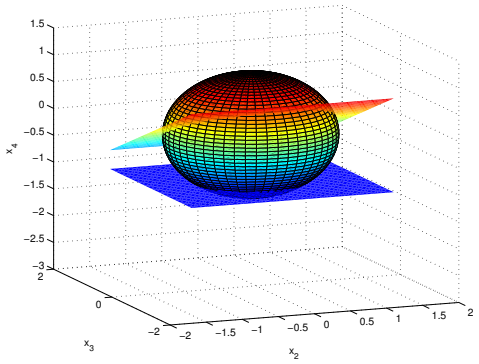
4.3 Building a disjunctive conic cut for nonparallel disjunctions

Some of the results in Section 4.2 can be extended to general disjunctions $\mathcal{A} \cup \mathcal{B}$. In this case, we have that $\mathcal{A} = \{x \in \mathbb{R}^n \mid a^\top x \geq \alpha\}$ and $\mathcal{B} = \{x \in \mathbb{R}^n \mid b^\top x \leq \beta\}$, and there is no $\kappa \in \mathbb{R}$ such that $b = \kappa a$. However, we assume that $\mathcal{Q}$ is in a family of quadrics $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ that contains an ellipsoid. From Corollary 3.1 we know that it is possible to apply the results of the analysis presented in Section 3.3 in this case. This assumption implies that both $\mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q} \cap \mathcal{B}^=$ are bounded. These type of disjunctions are illustrated in Figure 4.6(a) for Problem (4.13) using $\mathcal{A} = \{x \in \mathbb{R}^4 \mid 0.45x_3 + 0.89x_4 \geq 0\}$ and $\mathcal{B} = \{x \in \mathbb{R}^4 \mid x_4 \leq -1\}$ to define $\mathcal{A} \cup \mathcal{B}$.

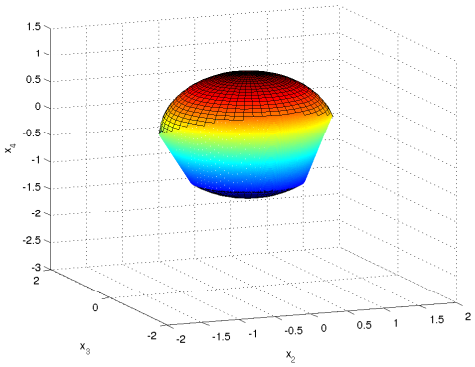
We want to characterize the set $\text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$. For this purpose, we use the results of the geometrical analysis of Section 3.3 to derive DCCs from the cases in Theorem 3.7. Additionally, using the general results in Chapter 2, we show that one of the quadrics $\mathcal{Q}(\bar{\tau})$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ satisfies the condition $\mathcal{Q} \cap \mathcal{Q}(\bar{\tau}) = \text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$. Observe that given Assumption 2.1, the third case in Theorem 3.7 cannot occur. Hence, this case is not considered for building a cut for general disjunctions. To simplify the algebra, we may assume w.l.o.g. that the quadric $\mathcal{Q}$ has been transformed using the affine transformations described on Section 3.1. Additionally, we may assume w.l.o.g. that $\|a\| = 1$ and $\|b\| = 1$. We separate the analysis in two cases. In Section 4.3.1 we study the cylinder case, and in Section 4.3.2 and the cone case.

4.3.1 Cylinders

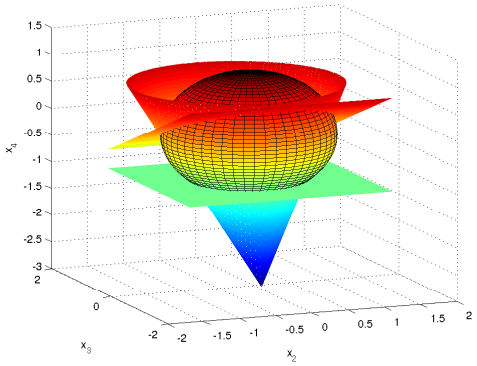
We start with the verification that the fourth and fifth cases in Theorem 3.7 allow the derivation of DCyCs for Problem (4.1). Recall that the classification given in Theorem 3.7 is derived using the roots of the numerator and denominator in (3.46). Let $\bar{\tau}_1$ and $\bar{\tau}_2$ be



(a) $\mathcal{A}^{=}$, $\mathcal{B}^{=}$, and $\mathcal{Q}$.



(b) conv$(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$.



(c) The cone yielding conv$(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}))$.

Figure 4.6: Convex hull of the intersection between a non-parallel disjunction and an ellipsoid.

the roots of the numerator, and $\hat{\tau}_2$ or $\hat{\tau}_1$ be the roots of the denominator. In the fourth and fifth cases in Theorem 3.7, the cylinder is found at one of the quadrics $\mathcal{Q}(\hat{\tau}_1)$ or $\mathcal{Q}(\hat{\tau}_2)$. The cylinder can be identified by comparing $\hat{\tau}_2$ or $\hat{\tau}_1$ with the roots $\bar{\tau}_1$ and $\bar{\tau}_2$ using the criteria described in Theorem 3.7. Let $\hat{\tau}$ be a value such that $\mathcal{Q}(\hat{\tau})$ is a cylinder. From Lemma 3.8, we know that $P(\hat{\tau})$ is a positive semi-definite matrix. From equation (1.3), it is easy to verify that $\mathcal{Q}(\hat{\tau})$ is a convex set. Consequently, from Proposition 2.2, we obtain that $\mathcal{Q}(\hat{\tau}) \cap \mathcal{Q} = \mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$. Finally, note that the cylinder $\mathcal{Q}(\hat{\tau})$ can be represented in terms of a second order cone, see for example Ben-Tal and Nemirovski [2001a]. This

shows that $\mathcal{Q}(\hat{\tau})$ is a DDc.

4.3.2 Cones

We focus now on the cones described in the first and fourth cases of Theorem 3.7. Let $\bar{\tau}_i \neq \hat{\tau}_1, \hat{\tau}_2$, $i = 1, 2$. In these two cases, $Q(\bar{\tau}_i)$ is a symmetric and non-singular matrix with exactly one negative eigenvalue. This is a similar situation as the first and third cases of Theorem 3.2. From the analysis in Section 4.2 follows that $\mathcal{Q}(\bar{\tau}_i) = \mathcal{Q}(\bar{\tau}_i)^+ \cup \mathcal{Q}(\bar{\tau}_i)^-$, where $\mathcal{Q}(\bar{\tau}_i)^+, \mathcal{Q}(\bar{\tau}_i)^-$ are the second order cones (4.17) and (4.18). Observe that $x(\bar{\tau}_i) = -Q(\bar{\tau}_i)q(\bar{\tau}_i)$ is the vertex of $\mathcal{Q}(\bar{\tau}_i)^+$ and $\mathcal{Q}(\bar{\tau}_i)^-$. Then, using Lemma 2.3, we can verify if there is a cone in $\mathcal{Q}(\bar{\tau}_i)^+, \mathcal{Q}(\bar{\tau}_i)^-$, $i = 1, 2$, that satisfies Proposition 2.1. In particular, we need to prove that there is one $x(\bar{\tau}_i)$, $i = 1, 2$, that is either in $\mathcal{A}$ or $\mathcal{B}$. This step is omitted here for the sake of readability. The interested reader can find the details of this step in Lemma A.6 in Appendix A.

Lemma 4.5. *In the first and fourth cases of Theorem 3.7, the cone $\mathcal{Q}(\bar{\tau}_2)$ contains a cone that satisfies Definition 2.2.*

Proof. From Theorem 3.6 we know that $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{A}^= = \mathcal{Q} \cap \mathcal{A}^=$ and $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{B}^= = \mathcal{Q} \cap \mathcal{B}^=$. Additionally, we have that $\mathcal{Q}(\bar{\tau}_2) = \mathcal{Q}^+(\bar{\tau}_2) \cup \mathcal{Q}^-(\bar{\tau}_2)$, where $\mathcal{Q}^+(\bar{\tau}_2)$, $\mathcal{Q}^-(\bar{\tau}_2)$ are two convex cones with their vertices at $x(\bar{\tau}_2)$. From Lemma A.6 we know that the vertex $x(\bar{\tau}_2)$ is either in $\mathcal{A}$ or $\mathcal{B}$. Thus, since the intersections $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{A}^=$ and $\mathcal{Q}(\bar{\tau}_2) \cap \mathcal{B}^=$ are bounded, then one of the following two cases is true:

- Case 1: $\mathcal{Q}^+(\bar{\tau}_2) \cap \mathcal{A}^= = \mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{Q}^+(\bar{\tau}_2) \cap \mathcal{B}^= = \mathcal{E} \cap \mathcal{B}^=$.
- Case 2: $\mathcal{Q}^-(\bar{\tau}_2) \cap \mathcal{A}^= = \mathcal{E} \cap \mathcal{A}^=$ and $\mathcal{Q}^-(\bar{\tau}_2) \cap \mathcal{B}^= = \mathcal{E} \cap \mathcal{B}^=$.

Consequently, we have that one of the cones $\mathcal{Q}^+(\bar{\tau}_2)$, $\mathcal{Q}^-(\bar{\tau}_2)$ found at the root $\bar{\tau}_2$ satisfies Proposition 2.1. □

Now we can define a procedure to identify a conic cut. We need to identify which of the cones $\mathcal{Q}(\bar{\tau}_2)^+, \mathcal{Q}(\bar{\tau}_2)^-$ gives the conic cut. For this purpose we use the sign of $W(\bar{\tau}_2)_1^\top(-Q^{-1}q - x(\bar{\tau}_2))$. Hence, we choose $\mathcal{Q}(\bar{\tau}_2)^+$ if $W(\bar{\tau}_2)_1^\top(-Q^{-1}q - x(\bar{\tau}_2)) > 0$, and we choose $\mathcal{Q}(\bar{\tau}_2)^-$ when $W(\bar{\tau}_2)_1^\top(-Q^{-1}q - x(\bar{\tau}_2)) < 0$. This completes the procedure.

4.4 Disjunctive conic cut vs Nonlinear conic mixed-integer rounding inequality

Atamtürk and Narayanan Atamtürk and Narayanan [2010] present a procedure for generating a *nonlinear conic mixed-integer rounding inequality*. Since this is a conic cut, we examine how it compares to the DCC introduced here. For this purpose, let us consider the following example

$$\begin{aligned} \text{minimize:} \quad & -x_1 \quad -x_2 \\ \text{subject to:} \quad & \left\| \begin{matrix} x_1 - \frac{4}{3} \\ x_2 - 1 \end{matrix} \right\| \leq \tfrac{4}{3} - \tfrac{x_1}{2} - \tfrac{x_2}{2} \\ & x_1 \in \mathbb{Z},\ x_2 \in \mathbb{R} \end{aligned} \tag{4.19}$$

First, notice that the example in (4.19) is in the form used in Atamtürk and Narayanan [2010], which is different from the one in (MISOCO). The main difference is the way we write the conic constraint. Despite this difference we can still construct a DCC, because the feasible region of this problem is an ellipsoid in the (x_1, x_2) space.

Using a branch and bound procedure one can easily solve the integer problem in (4.19), and get that the optimal solution is $(x_1^*, x_2^*) = (1, 1)$ with the optimal cost of -2.

We can rewrite problem (4.19) in the following form:

$$\begin{array}{lll} \text{minimize:} & -x_1 \quad -x_2 & \\ \text{subject to:} & x_1 \quad +x_2 \quad +2x_3 & = \frac{8}{3} \\ & \sqrt{(x_1 - \frac{4}{3})^2 + (x_2 - 1)^2} \leq x_3 & \\ & x_1 \in \mathbb{Z},\ x_2, x_3 \in \mathbb{R}. & \end{array} \tag{4.20}$$

Figure 4.7 presents the feasible region of this equivalent problem.

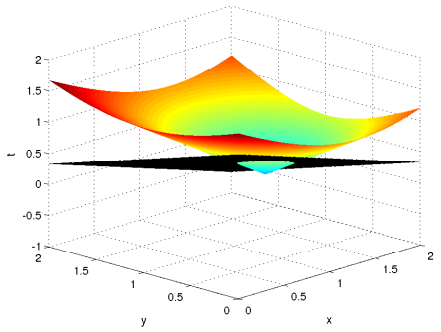


Figure 4.7: Feasible region of the sample problem (4.19).

Relaxing the integrality constraint, the resulting relaxation from problem (4.20) can be solved analytically using calculus. First, notice that this relaxation is just a problem of maximizing a linear function over an ellipsoid. In particular, we can rewrite the relaxation of problem (4.20) as

$$\begin{array}{ll} \text{minimize:} & -x_1 - x_2 \\ \text{subject to:} & \frac{3}{4}x_1^2 + \frac{3}{4}x_2^2 - \frac{1}{2}x_1x_2 - \frac{4}{3}x_1 - \frac{2}{3}x_2 + 1 \leq 0 \\ & x,\ x_2 \in \mathbb{R}. \end{array} \tag{4.21}$$

The feasible set of this problem in terms of the variables x_1, x_2 is presented in Figure 4.8. The feasible set is an ellipsoid, the optimal objective function value is -2.471, and the relaxed optimal solution for the example in problem (4.19) is $(x_1^*, x_2^*, x_3^*) = (1.402, 1.069, 0.098)$.

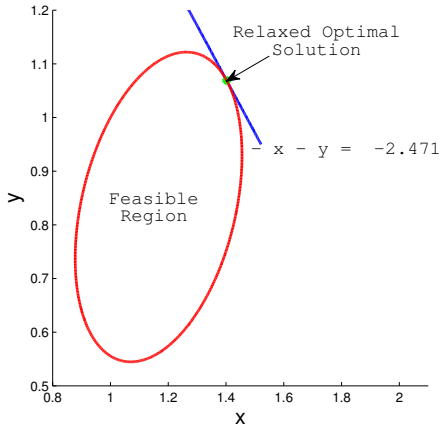


Figure 4.8: Optimal solution of the sample problem (4.19).

The problem reformulation (4.20) presents a case similar to the one studied in Example 1 in Atamtürk and Narayanan [2010], which shows how to obtain a nonlinear conic mixed-integer rounding inequality for the set

$$T_0 = \left\{ (x_1, x_2, x_3) \in \mathbb{Z} \times \mathbb{R} \times \mathbb{R} : \sqrt{(x_1 - \frac{4}{3})^2 + (x_2 - 1)^2} \leq x_3 \right\}, \tag{4.22}$$

which relates closely to the last constraint in (4.20). In general, the procedure discussed by Atamtürk and Narayanan Atamtürk and Narayanan [2010] focuses on generating the convex hull for each *polyhedral second-order conic constraint* in the problem. Then, by adding those new cuts, the original formulation is tightened. In particular, applying that procedure to the set in (4.22) they obtain the cut

$$\sqrt{\left(\frac{x_1}{3}\right)^2 + (x_2 - 1)^2} \leq x_3, \tag{4.23}$$

which is a valid cut for the problem in (4.20).

Analyzing the relaxed solution showed in Figure 4.8, we can see that the solution is not feasible for the integer problem. First, observe that if we use the disjunction $x_1 \leq 1 \vee x_1 \geq 2$

it is not possible to apply the DCC here, because the line $x = 2$ does not intersect the set of feasible solutions that is an ellipsoid, violating one of the assumptions in Chapter 2. However, we can still use the nonlinear conic mixed-integer rounding inequality procedure. Figure 4.9 shows the result of applying the nonlinear conic cut (4.22) to the problem in (4.20). The point $(x_1^*, x_2^*, x_3^*) = (1, 1, 1/3)$ is the new optimal solution for the continuous relaxation of the resulting problem with the cut added, which turns out to be optimal for the integer problem. The optimal objective value is -2.

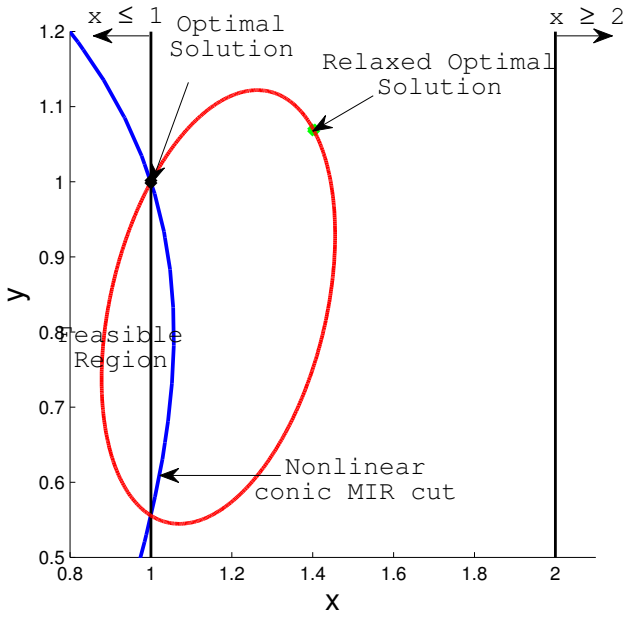


Figure 4.9: Nonlinear conic mixed-integer rounding inequality.

Now, let us modify the first constraint in (4.20) as follows

$$x_1 + x_2 + 2x_3 = \frac{14}{3}.$$

Figure 4.10 shows the new feasible region. With this modification the relaxed optimal solution is $(x_1^*, x_2^*, x_3^*,) = (1.81, 1.48, 0.68)$, which is not feasible for the integer problem.

Now, for this example, we can use the disjunction $x_1 \leq 1 \vee x_1 \geq 2$ and obtain a DCC that can be represented in the (x_1, x_2) space as follows:

$$\sqrt{(x_2 - 0.33x_1 + 0.22)^2} \leq 2.67 - 0.93x_1. \tag{4.24}$$

Observe that the nonlinear conic mixed-integer rounding inequality (4.23) stays the same, since we have not modified the conic constraint. Figure 4.10 shows these two cuts and highlights the difference between applying the nonlinear conic mixed-integer rounding inequality and our DCC to the modified problem. As expected, the DCC gives the convex hull of the intersection between the disjunction $x_1 \leq 1 \vee x_1 \geq 2$ and the feasible set of problem (4.20). This is not the case for the nonlinear conic mixed-integer rounding inequality (4.23). The new optimal solution for the relaxed problem when either of the cuts is applied is $(x_1^*, x_2^*, x_3^*) = (2.0, 1.25, 0.71)$. In particular, we can see that any of the cuts is enough to find the optimal solution. The optimal value for the objective function is -3.25.

Finally, we perform an additional test modifying the first constraint in (4.20) as follows:

$$x_1 + x_2 + 2x_3 = 8.$$

In this case we use the disjunction $x_1 \leq 2 \vee x_1 \geq 3$. Then, we can obtain a DCC that can be represented in the (x_1, x_2) space as follows:

$$\sqrt{(x_2 - 0.33x_1 + 1.33)^2} \leq 6.04 - 1.21x_1.$$

For this example the nonlinear conic mixed-integer rounding inequality (4.23) fails to eliminate the continuous optimal solution of the relaxed problem, as illustrated in Figure 4.11. Thus, there is no gain in adding this cut to the problem. However, the DCC is

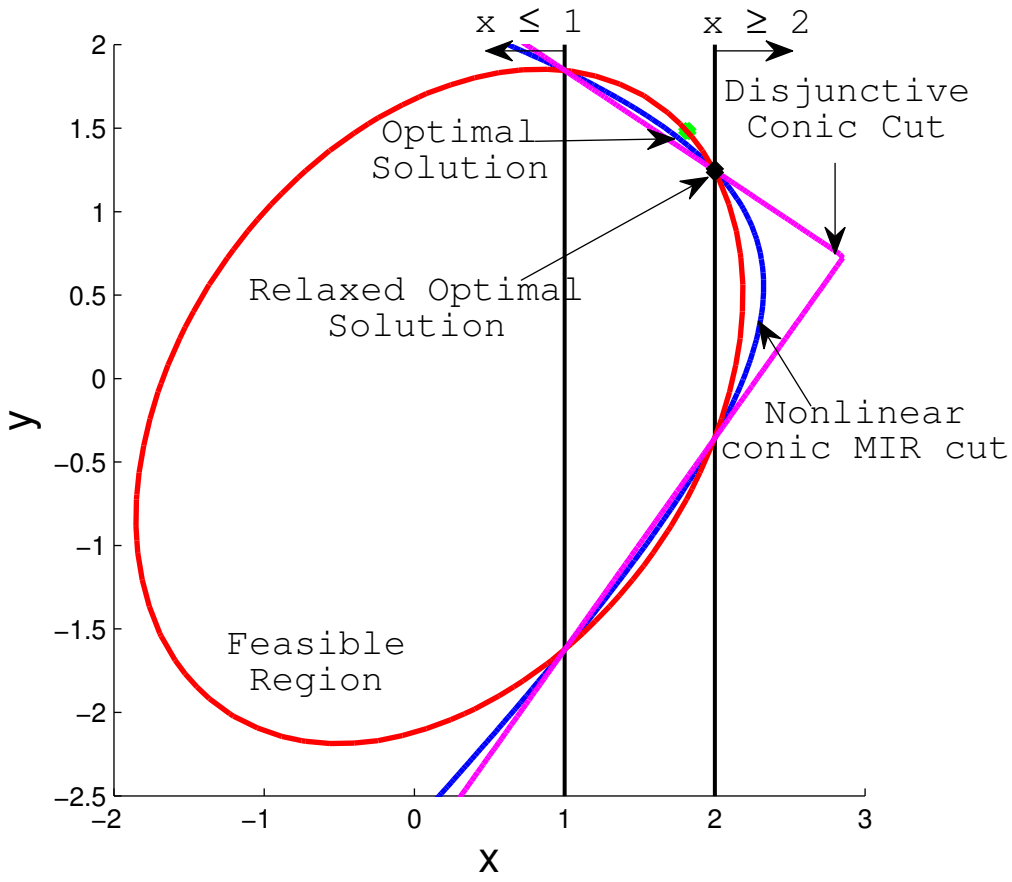


Figure 4.10: The DCC and the nonlinear conic mixed-integer rounding inequality cutting off the relaxed optimal solution.

violated by the current fractional solution, and the addition of the DCC is enough to find the integer solution of the problem.

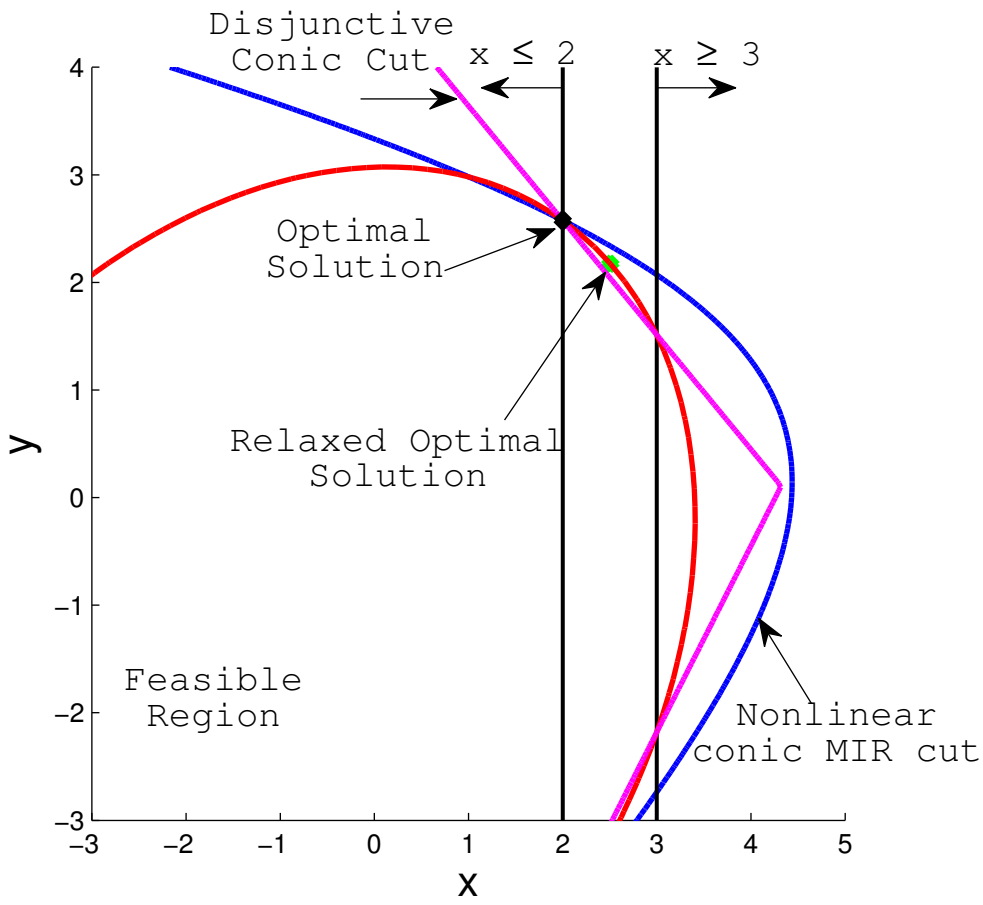


Figure 4.11: The nonlinear conic mixed-integer rounding inequality fails to cut off the optimal solution of the relaxed problem.

Chapter 5

Implementation

In this chapter we discuss the implementation of the DCCs in a branch and cut framework. The implementation presented here was developed in C++ using the CHiPPS framework Xu et al. [2009, 2005]. We begin in Section 5.1 describing the branch-and-cut algorithm. Then, in Section 5.2 we present how the procedure to generate conic cuts can be applied when a MISOCO problem has multiple cones. Finally, in Section 5.3 we give a brief description of the elements of our implementation.

5.1 Branch-and-cut Algorithm

In this section we describe the branch-and-cut algorithm that is used in our implementation to solve MISOCO problems. This algorithm is based in the BB algorithm 1 described in Section 1.1.4, the modified algorithm is presented in 2. Branch-and-cut algorithms are widely and successfully used for solving MILO problems Balas [1979], Cornuéjols [2008], Xpress, GUROBI [2013], CPLEX [2011], Mitchell [2002], MOSEK [2011], Nemhauser and Wolsey [1999], Ralphs et al. [2011], Schrijver [1986], and currently some commercial solvers used it to solve MISOCO problems as well GUROBI [2013], CPLEX [2011], MOSEK [2011]. From the description of the BB algorithm in Section 1.1.4, we can observe that the speed

Algorithm 2 Branch-and-Cut

Data: $\mathcal{M} = \{\Pi^0\}$, $\zeta^* = \infty$, the index set of integer variables $\mathcal{J}$, and the maximum number of rounds for adding DCCs maxrounds.

```
while M ≠ ∅ do
    Select an active problem Π^a from M, which has a feasible set F^a
    M ← M \ Π^a
    addcuts ← true
    rounds ← 0
    branch ← true
    while addcuts = true and rounds ≤ maxrounds do
        Solve the continuous relaxation Π^r of Π^a
        if Π^r is feasible then                          ▷ (prune by infeasibility)
            ζ^r ← optimal objective value of Π^r
            x^r ← optimal solution of Π^r
            if ζ^r ≤ ζ* then                             ▷ (prune by value dominance)
                if x^r ∈ F^a then                        ▷ (prune by integrality)
                    ζ* ← ζ^r,                            ▷ (update upper bound)
                    x* ← x^r,                            ▷ (update incumbent)
                    branch ← false
                else if addcuts then
                    Search for DCCs that are violated by x^r,
                    if any found add them to Π^r, else addcuts ← false
                end if
            else
                branch ← false
            end if
        else
            branch ← false
        end if
        rounds ← rounds + 1
    end while
    if branch then
        Select a branching variable x_j ∉ Z, j ∈ J
        M ← {M, min{c^⊤ x | x ∈ F^a, x_j ≥ ⌈x_j⌉}, min{c^⊤ x | x ∈ F^a, x_j ≤ ⌊x_j⌋}}
    end if
end while
```

and success of a BB algorithm relies on three aspects:

- early detection of infeasible nodes;
- early findings of integer solutions that enable the fathoming of the search tree branches;
- improvement of the lower bounds.

In MILO, the use of pre-processing and heuristics plays a crucial role in speeding up the solution time in a mixed integer solver Fischetti et al. [2005], Mahajan [2010], Martin [2001], Savelsbergh [1994]. Pre-processing helps to simplify the mathematical formulation of the problem as well as to verify its correctness and detect infeasibility. Heuristics help in the early detection of integer solutions, which allow to early fathom branches of the tree. The detection of an integer solution may improve the upper bound in the algorithm, which helps the early detection of sub-optimal solutions in the search tree. Although these are crucial elements in the development of an effective implementation of a branch-and-cut algorithm, these are not the focus of our experiments.

The last aspect mentioned in our list is the improvement of the lower bound during the execution of a branch-and-cut algorithm. A common technique used to achieve improvements in the lower bound is the addition of linear cuts and nonlinear cuts Drewes [2009], Xpress, GUROBI [2013], Grossmann [2002], CPLEX [2011], MOSEK [2011], Ralphs et al. [2011]. In MILO, this has shown to be an effective and successful technology used in the solution of large scale problems. A sharp lower bound is crucial for a quick termination of the algorithm. On one hand, observe that in Algorithm 2, at each node, the improvement of its lower bound can lead to the early pruning of a tree branch by objective value dominance. Additionally, it is possible that the addition of valid cuts may result in finding an integer solution, which enables the early pruning of a tree branch by integrality. Our goal in this section is to investigate the effect of DCCs in the tree search and solution time

when used in a branch and cut algorithm.

Finally, it is well-know in the literature from empirical experience that the performance of Algorithm 2 is affected by three critical decisions. First, in Section 5.1.1 we discuss the selection of the branching variable. This decides which new problems will be added to the set $\mathcal{M}$. Second, in Section 5.1.2 we discuss the criteria for selecting the seed to create a DCC. This will affect the effectiveness of the cuts on the relaxed problem during the evaluation of each node. Third, in Section 5.1.3 we discuss the selection rules for next node to be explored.

5.1.1 Strategies for branching

On of the algorithmic choices in Algorithm 5.1 is associate with the selection of the disjunctive set that would be used to do the partition process. This defines the branching step. This selection is govern by the branching rules in a BB-based algorithm, which have been extensively studied for MILO and for some mixed integer nonlinear optimization problems, see, e.g., for example Bonami et al. [2011], Martin [2001], Drewes [2009], Achterberg et al. [2005]. Based on previous work, we choose to implement three rules. Two of them, trong branching and Reliability Branching, have been shown to work well in general Bonami et al. [2011], Achterberg et al. [2005]. We implement the most fractional rule too, which is a simple rule to keep a base for comparison.

Most fractional: Although it is a naive approach, given its simplicity, this strategy has been always considered as an option in computational testing Bonami et al. [2011], Gupta and Ravindra [1985], Drewes [2009]. The results in Bonami et al. [2011] for this strategy show that in principle it is not a good strategy, although it outperforms a random branching strategy. We will consider this strategy in the tests as a base measure.

Strong branching Bonami et al. [2011]: This rule usually shows the best performance in terms of the number of nodes explored, see for example Bonami et al. [2011], Achterberg

et al. [2005], for the cases of MILO and convex mixed integer nonlinear problems. Let x^* be an optimal relaxed solution and ζ^* its optimal objective value. Then, in each node the rule perform the following procedure. First, identify the index set $\mathcal{J}$ of all the fractional variables in the relaxed solution. Then, it computes the relaxations for the two child node relaxations of all the fractional variables and assign the relaxed objective function values to ζ_j^+ and ζ_j^-, $j \in \mathcal{J}$. Using ζ_j^+, ζ_j^-, and a parameter $0 \leq \gamma \leq 1$ it assigns a score ξ_j to each fractional variable computed as follows

$$\xi_j := (1-\gamma)\min(\zeta_j^- - \zeta^*, \zeta_j^+ - \zeta^*) + \gamma \max(\zeta_j^- - \zeta^*, \zeta_j^+ - \zeta^*).$$

Finally, it chooses a variable with maximal branching score ξ_j as the branching variable. The main concern with this rule is that it requires to solve a large number of problems before making the choice. For this reason, this rule usually do not come first in solution time.

Reliability Branching Bonami et al. [2011]: The key of this procedure is to avoid solving to many problems to make a branching decision. Then, for each variable x_j, it keeps estimates Ψ_j^- and Ψ_j^+ of the potential change per unit in the objective functions if we add the inequality $x_j \leq \lfloor x_j \rfloor$ and $\lceil x_j \rceil \leq x_j$, respectively. For each direction the estimate is computed as the average of the gain over all the objective per unit change in variable j in all the problems in which it has been used for branching. Using this estimates, the predicted objective function values are computed as

$$\zeta_j^- = \zeta^* + \Psi_j^-(x_j^* - \lfloor x_j \rfloor) \text{ and } \zeta_j^+ = \zeta^* + \Psi_j^+(\lceil x_j \rceil - x_j^*).$$

These values are used for computing the branching score ξ_j. The problem with the estimates Ψ_j^- and Ψ_j^+ is that they cannot be computed at the beginning of the tree. In order to initialize the estimates and improve their veracity strong branching is executed

a fixed number of times k. This rule is reported in Bonami et al. [2011] to outperforms strong branching in solution time while keeping a good improvement in the number of nodes explored.

One of the major issues with these strategies is related to the solver choice. The tests in Bonami et al. [2011] show that using a Non-Linear Optimization Solver can reduce the number of nodes required for solving the problem. However, it will be more expensive in terms of cpu time.

In this case, we will use IPM to solve the second order cones problems at each node. We note that the lack of effective warm-start methods for IPM may affect the performance of the algorithm. Specially during using branching. One of the goals with this experimentation is to observe how this characteristic of IPMs affects the solution strategy for MISOCO problems.

5.1.2 Strategies for selecting the seed to formulate a DCC cut

In general, we will have more than one cone in the formulation of the problem to solve. In this case, we need to find the cones for which the relaxed solution is in the boundary. This is necessary because adding a DCC based on a cone that is not active will not cut off the relaxed solution. Once the cones have been identified, the next decision to make is which violated disjunction to use. Here, we only consider disjunctions of the form $x_i \leq \lfloor x_i \rfloor \bigvee x_i \geq \lceil x_i \rceil$. Then, for each cone we can potentially have more than one violated disjunction. In our implementation we use the fractional part $f_0 = x - \lfloor x_i \rfloor$ of a variable as a measure of the strength for the candidate cuts. Then, we choose the most fractional variable, i.e., the variable in each cone with a value closest to 0.5.

5.1.3 Strategies for selecting the next node to explore

In addition to the choice of the branching variable, we need to provide criteria to decide what node to process. In our implementation, we provide following rules for this decision:

- Best first chooses the node with the best objective function.
- Depth first chooses the node furthest away from the root node of the search tree.

The first rule aim to chose the node that will improve the lower bound the most. In other words, best first search focus on ensuring that no solution better than the current one exists. However, it requires a lot of memory, since the list of unprocessed nodes can grow quite fast. The second one has the advantage that it requires less memory. Additionally, the experience with MILO has shown that an integer solution is more likely to be found deep in the tree. However, the lower bound usually does not improve significantly while exhausting one branch of the tree. This can affect the termination of the algorithm. Hence this rule can result in very large search trees. For more on node selection rules see Linderoth and Savelsbergh [1999], Nemhauser and Wolsey [1999], Ralphs [2006].

5.2 Multiple cones in the MISOCO

In the disjunctive conic cut theory and in the procedure presented in Chapters 3 and 4, we assume that the problems had a single second order. This is not the general case; thus it is necessary to explore how to extend this procedure to include problems with more than one cone.

Recall the MISOCO problem

$$\text{minimize: } c^T x \tag{5.1}$$

$$\text{subject to: } Ax = b \qquad \text{(MISOCO)} \tag{5.2}$$

$$x \in \mathcal{K} \tag{5.3}$$

$$x \in \mathbb{Z}^d \times \mathbb{R}^{n-d}, \tag{5.4}$$

where $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $x = \left((x^1)^\top, (x^2)^\top, \ldots, (x^k)^\top\right)^\top$, $\mathbb{L}^{n_i} = \{x^i | x^i_1 \geq \|x^i_{2:n_i}\|\}$ are Lorentz cones, $\mathcal{K} = \mathbb{L}_1^{n_1} \times \cdots \times \mathbb{L}_k^{n_k}$, and the rows of A are linearly independent.

We propose a reformulation of the problem using the null space of the matrix A. The objective is to build blocks of variables that can be separated for using the cones in the model independently. Then, let $H_{n\times(m-n)}$ be a matrix with its columns being orthogonal and forming a basis for the null space of A. We can rewrite the MISOCO problems in term of H as follows,

$$\begin{aligned}
\text{minimize: } & c^\top \hat{x} + c^\top H w \\
\text{subject to: } & x = \hat{x} + Hw \\
& x \in \mathcal{K} \\
& x \in \mathbb{Z}^d \times \mathbb{R}^{n-d} \\
& w \in \mathbb{R}^{m-n},
\end{aligned}$$

where $A\hat{x} = b$. In this reformulation we can brake H and $\hat{x}$ using the blocks of variables

in $x = \left((x^1)^\top, (x^2)^\top, \ldots, (x^k)^\top\right)^\top$ as follows

$$\begin{aligned} \text{minimize: } & c^\top \hat{x} + c^\top H w \\ \text{subject to: } & x^i = \hat{x}^i + H^i w, i = 1, \ldots, k \\ & x^i \in \mathbb{L}^{n_i}, i = 1, \ldots, k \\ & x \in \mathbb{Z}^d \times \mathbb{R}^{n-d} \\ & w \in \mathbb{R}^{m-n}, \end{aligned}$$

where H^i is the set of rows in H corresponding to the block of variables x^i. Similarly, $\hat{x}^i$ is the block of $\hat{x}$ corresponding to the block of variables x^i. If the integer constraints are relaxed, we get the relaxed problem formulation

$$\begin{aligned} \text{minimize: } & c^\top \hat{x} + c^\top H w \\ \text{s.t: } & w^\top (H^i)^\top J H^i w + 2(\hat{x}^i)^\top J H^i w + (\hat{x}^i)^\top J \hat{x}^i \leq 0 \quad \forall\, i = 1, \ldots k \\ & w^\top (H^i_{1:})^\top + \hat{x}^i_1 \geq 0 \quad \forall\, i = 1, \ldots k \\ & w \in \mathbb{R}^{m-n}, \end{aligned} \tag{5.5}$$

where $(H^i_{1:})^\top$ is the first row of matrix H^i, and $\hat{x}^i_1$ is the first component of vector $\hat{x}^i$. Then, we could potentially use each of the k quadrics in the first set of constraints of (5.5) to build disjunctive conic cuts.

We illustrate the procedure with the following example

$$
\begin{array}{rrrrrrrrrr}
\text{minimize:} & 3x_1 & +x_2 & +2x_3 & +x_4 & & & +x_7 & +x_8 & \\
\text{subject to:} & 9x_1 & +x_2 & +x_3 & +x_4 & +x_5 & & & & = 10 \\
& x_1 & & & & +9x_5 & +x_6 & +x_7 & +x_8 & = 10 \\
& x_1 & & +x_3 & & +x_5 & & & +x_8 & = 2 \\
& x_1 & +x_2 & & & +x_5 & +x_6 & +x_7 & & = 1 \\
& & +x_2 & & & x_5 & +x_6 & & +3x_8 & = 1 \\
& & & & (x_1, x_2, x_3, x_4) \in \mathbb{L}^4 & & & & & \\
& & & & (x_5, x_6, x_7, x_8) \in \mathbb{L}^4 & & & & & \\
& & & & x_2, x_6 \in \mathbb{Z}. & & & & &
\end{array}
$$

If we solve the relaxation of this problem, we get the solution

$$x^*_{\text{soco}} = (1.22, -0.88, -0.34, -0.77, 0.10, 0.51, -0.85, 0.12),$$

with an optimal objective value of $\zeta^*_{\text{soco}} = 0.72$. This solution is not feasible since x_2, and

x_6 are fractional. We can rewrite this problem in the w space as follows:

$$
\begin{aligned}
\text{minimize:} \quad & 0.72 + 0^\top w \\
\text{subject to:} \quad & 0.15w_1 + 0.04w_2 + 0.07w_3 + 1.06 \geq 0 \\
& -0.04w_1 - 0.07w_2 - 0.09w_3 + 1.0596 \geq 0 \\
& w^\top \begin{pmatrix} 0.93 & -0.13 & -0.01 \\ -0.13 & 0.18 & 0.23 \\ -0.01 & 0.23 & 0.52 \end{pmatrix} w + 2 \begin{pmatrix} -0.09 \\ 0.14 \\ 0.17 \end{pmatrix}^\top w - 0.91 \leq 0 \\
& w^\top \begin{pmatrix} 0.02 & 0.12 & -0.01 \\ 0.12 & 0.80 & -0.25 \\ -0.01 & -0.25 & 0.45 \end{pmatrix} w + 2 \begin{pmatrix} 0.01 \\ -0.09 \\ -0.13 \end{pmatrix}^\top w - 0.76 \leq 0 \\
& (w_1, w_2, w_3) \in \mathbb{R}^3.
\end{aligned}
$$

The feasible region of the problem in the w space is shown in Figure 5.1.

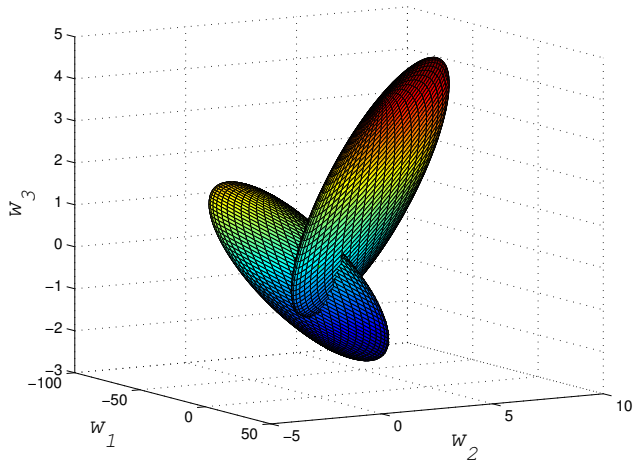
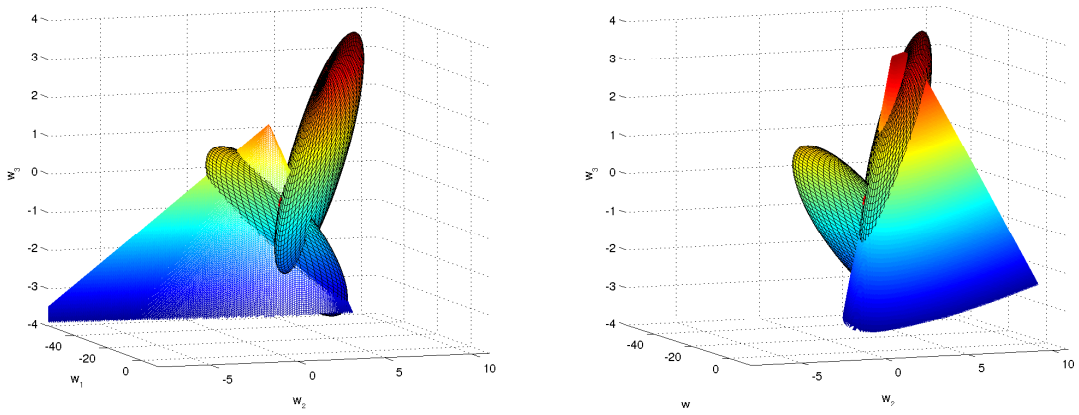


Figure 5.1: Feasible set of problem in the w space.

Then, since x_2 and x_6 are in two different cones, we could get a disjunctive conic cut from each of the variables. For x_2, we use the disjunction $\{x \in \mathbb{R}^8 | x_2 \geq 0\} \cup \{x \in \mathbb{R}^8 | x_2 \leq$

$-1\}$ and for x_6 we use the disjunction $\{x \in \mathbb{R}^8 | x_6 \geq 1\} \cup \{x \in \mathbb{R}^8 | x_6 \leq 0\}$. With the reformulation in the null space of A we can take each cone independently to build the cuts. In this example we have that for the two cones, the intersection is an ellipsoid. Then it is easy to verify that all the assumptions in Chapter 2 are satisfied. We can follow the procedure in Chapter 4 to construct the DCCs for each ellipsoid. These cuts are illustrated in Figure 5.2.



(a) (b)

Figure 5.2: Adding cuts in the presence of multiple cones.

5.3 Computational Framework

Here, we briefly describe the implementation of the branch-and-cut algorithm with disjunctive conic cuts, which is used for our experiments. For this implementation, we used the COIN-OR High Performance Parallel Search Framework (CHiPPS) framework. In particular, we extended the Branch, Constrain, and Price Software (BiCePS) for our implementation. This library is the data-handling layer needed in addition to the Abstract Library for Parallel Search (ALPS) to support relaxation-based branch-and-bound algo-

rithms. The ALPS library provides the fundamental classes that can be extended to implement the algorithmic components required to specify a tree search. The interested reader interested in a detailed description of the frameworks can see Xu et al. [2009, 2005].

Our implementation is built based on the BiCePS Linear Integer Solver (BLIS) and uses most of its features. We kept the BLIS knowledge management structures given in the classes `BlisConstraint`, `BlisVariable`, `BlisSolution`, and `BlisNodeDesc` with some minor modifications. The modifications were needed to include the link between variables and constrains with the second order cones present in a MISOCO problem. This modifications are included in `IclopsConstraint`, `IclopsVariable`, `IclopsSolution`, and `IclopsNodeDesc`. On the other hand, we use the branching methods provided by BLIS: strong branching, reliability branching, and most fractional branching.

For the knowledge structures of BLIS given in the classes `BlisModel` and `BlisTreeNode` we made substantial changes. Additionally, we added some extra components for the derivation of conic cuts and the interaction with the conic solver. We describe briefly our major modifications and enhancements in the following sections.

5.3.1 Class `IclopsModel`

This class provides all the data structures needed to describes a MISOCO model. It is based on the class `BlisModel` with some modifications. First, it includes some extra structure to handle the second order cones in the model description. Additionally, we added some extra methods to manage all the possible status of the interior point method solver. Finally, we redesigned the method needed for the initialization of the cut generators. In the BLIS case, these are designed for the COIN-OR Cut Generator Library, while in our implementation. these are adjusted to use our conic cut generator.

5.3.2 Class `IclopsTreenode`

This class includes the data structures and the methods needed to store and evaluate each node in the tree. The major difference with `BlisTreeNode` is contained in the method used to do the bounding and decide the branching of each node. Essentially, this method implements the inner `while` cycle and the branching part in Algorithm 2. Here we replace the section in `BlisTreeNode` that adds linear cuts with a new section based on our conic cuts generation tools. Additionally, we eliminated the call to heuristics that where present there.

5.3.3 Class `IclopsSolver`

This class is an interface of our branch-and-cut implementation with the conic solver. It provides a shell listing all the methods that are essential for our implementation to interact with the solver. Particularly, how to call the solves to solve the continuous relaxation of the MISOCO problem, and the query the status of the solution and the problem. Additionally, we need to be able to handle all the modifications in the problem performed in each node. Specifically, the changes in the bounds of some variables and the addition of the conic cuts. In our current development, we have implemented this interface to interact with MOSEK for processing the relaxation in each node.

5.3.4 Class `IclopsConicCutGenerator`

This class provides the data structure to store a quadric associated with the MISOCO problem that can be used to executed the procedure described in Chapter 4. In this description we store the triplet (P, p, ρ) that describes the quadric, where $P \in \mathbb{R}^{\ell} \times \ell$, $p \in \mathbb{R}^{\ell}$ and ρ is a scalar. Additionally, we identify the classification of the quadric. This facilitates the identification of which of the procedures presented in Section 4.2 should be applied to create the cut.

There are two options for the creation of these objects. In the first, place the user can specify a cone and a subset of the linear constrains in the MISOCO problem. In this case, the quadric is constructed using the procedure described in Section 4.1. This method considers the fact that in general a MISOCO problem may have multiple cones in its formulation. For that reason the implementation is designed to build the quadrics for each cone using the procedure described in Section 5.2. On the other hand, some times it is easy to formulate the quadric in terms of the original variables, as is the case in the example used in Section 4.2. For that reason, in this class we implemented the option to create a conic cut generator using the explicit description of the quadric provided by the user.

5.3.5 Class `IclopsConicCut`

This class provides the methods to build DCCs, and the structures to store them. Note that we can generate several cuts starting from the same quadric. For that reason, this class does not store information about the quadric, which avoid storing duplicate information. Instead, this objects points to one of the conic cut generators initialized and stored with the model. Using this information, based on the classification of the quadric of the conic cut generator, we apply here one of the methods described in Section 4.2. Finally, in this class we provide a method to add the resulting conic constraint to the model.

5.3.6 Input format

Currently, our implementation admits MISOCO problems in the extended mps format of MOSEK, which allows the modeling of second order cones. Additionally, the user must provide an additional input file describing the parameters to be used in the generation of the conic cuts. In this file the user must provide in the same order the following set of parameters:

- `numConicCutGenerators:` Defines the number of conic cut generators to be created.
- `MaxPoolSize:` Defines the maximum size of the pool of cuts to be available at any time during the exploration of the tree.
- `MaxNumCuts:` Defines the maximum number of cuts to add to the problem when processing a node.
- `MaxCutRounds:` Limits the number of cut generation rounds that can be executed when processing a node.
- `BoundaryDistance:` Limits how far from the boundary of a cone a solution can be to be considered good to define a DCC.
- `CutsParams:` This is an array that defines the parameters of the cone. The first component of the array defines the number of rows to be used in the creation of the cut. The second parameters is the index of a variable that belongs to the cone used in the creation of the cut. The last parameter is 1 if the cut is created using a primal form, i.e., using the procedure in Section 5.2. On the other hand, it is -1 if the cut is created using a dual form, or providing the description of the quadric in the original variables.
- `CutsRows:` A list of the rows of the matrix A that is used in the creation of the cut.

5.4 Implementation considerations

We close this chapter with some comments about several consideration that are needed in the implementation of the DCC for solving MISOCO problems.

5.4.1 Building the quadrics to derive DCCs

One of the challenges in the implementation of the DCCs is the construction of the quadrics needed to derive them. In our implementation, we rely on the knowledge of the user about the structure of the problem. In particular, in our input format, we require the user to identify the cone and the set of constraints that will be used for the derivation of the DCCs. However, it is not necessary true that this knowledge is available to the user. For that reason, it is sill necessary to develop procedures that can be run in a pre-processing phase that identifies the quadrics needed for deriving DCCs. Ideally, this should be as transparent to the user as it is in MILO solvers.

Assuming that the quadrics are available, there is still an additional consideration. It is important to identify whether the quadric is classified as a cone or as a cylinder. In these cases, there is a limitation in the possibility of generating tight cuts for the problem. First, if the quadric is a cone, we can create tight cuts only if the disjunctive set $\mathcal{A} \cup \mathcal{B}$ does not contain the vertex of the cone. It is shown in the first case of the proof of Theorem 4.1 that if the vertex of the cone is in $\mathcal{A} \cup \mathcal{B}$, then the DCC is the original cone. A similar situation is faced when the quadric is a cylinder. In this case, we can only use disjunctive sets, where the normal vector a defining the hyperplanes $\mathcal{A}^=$ and $\mathcal{B}^=$ is orthogonal to the direction of the cylinder. It that is not the case, again then the DCC is the original cylinder.

Finally, recall that an important goal associated with adding a cut in a branch an bound algorithm is to improve the lower bound of the problem. This can be achieve if the DCC successfully excludes a solution x^* that is not feasible for the integer problem. If we use an Interior Point Method solver it is possible to get an optimal solution x^* of the relaxed problem in the relative interior of the optimal set. This is also reported in Bonami et al. [2011], Drewes [2009]. In this case it is not possible to generate a cut that separates the relaxed solution, thus it is better to branch. However, if this is not verified it

is still possible to generate a DCC that is not tight for x^*. To avoid this we use a tolerance parameter $\epsilon \geq 0$. Then, given a quadric $\mathcal{Q} = \{x \in \mathbb{R}^n \mid x^\top Q x + 2q^\top x + \rho \leq 0\}$, the goal is to check how tight is $\mathcal{Q}$ for the solution x^*. Thus, if $(x^*)^\top Q x^* + 2q^\top x^* + \rho \leq \epsilon$, then we use $\mathcal{Q}$ to generate a cut, otherwise we branch or choose another quadric in the problem. By default this parameter is set to $10e-8$, but it may be defined by the user in the input file with the value `BoundaryDistance`.

5.4.2 Managing the addition of DCCs

Consider the following extreme case, illustrated by the problem

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.:} \quad & d^\top x = \delta \\ & \|x_{2:n}\| \leq x_1, \end{aligned} \tag{5.6}$$

where $d \in \mathbb{R}^n$, $\delta \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and recall that $x_{2:n}$ are the vector with the components $2, \ldots, n$ of x. We illustrate the effect on the problem size when adding a DCC with this problem. For convenience, we do all the operation in the original variables space. The feasible set $\mathcal{F}$ of Problem (5.6) can be rewritten as $\mathcal{F} = \{x \in \mathbb{R}^n \mid x_1 = \frac{d_{2:n}^\top x_{2:n}}{d_1},\ x_1 \geq 0,\ x_{2:n}^\top Q x_{2:n} + 2q x_{2:n} + \rho \leq 0\}$, where

$$Q = I - \frac{1}{d_1^2} d_{2:n} d_{2:n}^\top, \quad q = \frac{\delta}{d_1^2} d_{2:n}, \quad \rho = -\left(\frac{\delta}{d_1}\right)^2.$$

Now if we use the disjunction $x_2 \leq \sigma$ and $x_2 \geq \sigma + 1$, we know from Section 4.2 in Chapter 4 that $\exists \bar{\tau} \in \mathbb{R}$ for which the quadric $\mathcal{Q}(\bar{\tau}) = \{x_{2:n}^\top \in \mathbb{R}^{n-1} \mid x_{2:n}^\top Q(\bar{\tau}) x_{2:n} + 2q(\bar{\tau}) x_{2:n} + \rho(\bar{\tau}) \leq 0\}$ allows us to derive a DCC for the Problem (5.6), where

$$Q(\bar{\tau}) = Q + \bar{\tau} e_1 e_1^\top, \quad q(\bar{\tau}) = q - \bar{\tau} \frac{2\sigma + 1}{2}\, e_1, \quad \rho(\bar{\tau}) = \rho + \bar{\tau}(\sigma^2 + \sigma).$$

In our implementation, we represent the DCC as a second order cone. For the sake of this example, let us assume that $Q(\bar{\tau})$ is non-singular and ID1. Now, consider its eigenvalue decomposition $Q(\bar{\tau}) = VDV^\top$, and assume that the DCC derived from $\mathcal{Q}(\bar{\tau})$ is

$$\mathcal{Q}(\bar{\tau})^+ = \left\{ x_{2:n}^\top \in \mathbb{R}^{n-1} \;\middle|\; \left\| V_{2:n}^\top \left(x_{2:n}^\top + Q(\bar{\tau})^{-1} q(\bar{\tau}) \right) \right\| \leq V_1^\top \left(x_{2:n}^\top + Q(\bar{\tau})^{-1} q(\bar{\tau}) \right) \right\}.$$

If we add this cut in Problem (5.6) we obtain

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.:} \quad & d^\top x = \delta \\ & V^\top x_{2:n}^\top - w = V^\top Q(\bar{\tau})^{-1} q(\bar{\tau})) \\ & \|x_{2:n}\| \leq x_1 \\ & \|w_{2:\ell}\| \leq w_1, \end{aligned}$$

where $\ell = n - 1$ and $w \in \mathbb{R}^\ell$. Hence, the addition of the DCC implies in this case the addition of $n-1$ new variables and constraints. Now, typically we have a set of constraints $Ax = b$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $n \leq m$, and $\text{rank}(A) = m$. Then, from Section 4.1 we know that dimension of the DCC in this case is $\ell = n - m$, and this will imply the addition of ℓ new variables and constraints.

Although this is a challenge, this is a problem that has to be faced in general when cuts are used in a branch-and-cut algorithm. For managing this increase, we currently have a cut manager that limits the number of cuts that can be added to the problem. This number can be defined in the input file using the parameter `MaxNumCuts`. Additionally to this limit, we keep a measures of the "age" of the cuts, and they are deactivated after they reach certain age. We use the concept of age described in Martin [2001]. The age of the cuts is set to zero when the cut is created. Then, before using the cut in a node we verify that it is indeed violated. Every time the cut is not tight we increase the age of the cut

by one in the current branch. Once the cut reaches certain age limit, it is deactivated in that branch.

Finally, to control the use of memory to store cuts, every time a cut is created it is placed in a pool of cuts. Then, in each node of the tree we use the cuts that were used in its parent. If these cuts do not exceed the number of cuts allowed, we then generate additional cuts in that node. Each node owns the cuts created when it is being evaluated. Then, those cuts are eliminated when the node is pruned.

5.4.3 Numerical challenges when building DCCs

During the derivation of a DCC we need to decide if the result is a cylinder or a cone. For the implementation of this decision we use a tolerance parameter $\epsilon > 0$. Then, given the computed $\bar{\tau}$ defining the DCC $\mathcal{Q}(\bar{\tau})$, we compare it with the value of $\hat{\tau}$, in which case we classify the DCC as a cylinder. This help us to avoid the computation of eigenvalues for the classification of the DCC. Then if $|\bar{\tau} - \hat{\tau}| < \epsilon$ we classify the DCC as a cylinder. On the other hand, if $|\bar{\tau} - \hat{\tau}| > \epsilon 10^4$, then we classify the DCC as a cone. Now, consider the case where $\epsilon < |\bar{\tau} - \hat{\tau}| < \epsilon 10^4$. In this case the DCC can be a narrow cone if the intersections $\mathcal{A}^{=} \cap \mathcal{Q}$ and $\mathcal{B}^{=} \cap \mathcal{Q}$ are bounded. On the other hand, it can be a wide cone if the intersections $\mathcal{A}^{=} \cap \mathcal{Q}$ and $\mathcal{B}^{=} \cap \mathcal{Q}$ are unbounded. In this case, we have observe that the addition of DCC can lead to numerical problems with the continuous solver. This can be explained by the fact that in this cases, there may be a big difference between the eigenvalue close to zero and the rest of the eigenvalues. Currently, in our implementation we discard this case as candidate for DCCs. This has helped to avoid problems with the continuous solver, but the definition of the rejection tolerances is still the subject of more analysis.

Chapter 6

Computational experiments

In this chapter we describe the test sets that were used for performing some preliminary experimentation with the DCCs presented in Chapter 4 for parallel disjunctions. This test sets are also described in Çay et al. [2013]. With each test set we performed some computational experiments, and here we present and analyze the main findings. These experiments are provided in the spirit of a proof of concept, and more extensive experimentation is needed before we can draw any strong conclusion about the performance of the DCCs and DCyCs.

6.1 Random problems

The first set used during the experimentation consists of randomly generated problems. This sets were generated so that the problems are feasible and bounded. The main characteristics of these problems are:

- The cuts can be derived from quadrics that are ellipsoids.
- The problems in this set all have multiple cones.

- The set of variables is divided into integer and continuous, but all the cones in the set have integer variables.
- The integrality constraint defines general integer variables.

Finally, we have the following naming convention for this problem set: R#.C#.Con#.Int#, where the first component gives the number of rows, the second gives the total number of columns, the third gives the number of cones, and the fourth one gives the number of integer variables.

6.1.1 Experiments with randomly generated MISOCO problems

In this case, we have collected 30 problems, which are listed in Appendix B. We present here two different computational experiments. First, we use two different branching rules, combined with different criteria to select the seed to create the DCCs. Second, using the branching pseudo cost branching rule we test two different criteria to select the seed to create the DCCs.

6.1.1.1 Experiments with branching rules

In this experiments we tested our 30 problems using two branching rules: reliable branching and strong branching. With each rule we ran three experiments. First we solve the problems with pure branch and bound. Second, we solve the problems using the most fractional criteria for selecting the seed to create the DCCs. Finally, we solve the problems using the pseudo costs for selecting the seed to create the DCCs. In each of these two last experiments we added tree DCCs at the root node, and then one DCC is added every 10 nodes with a limit of 10 DCCs in total.

The results for the experiments using pseudo costs branching are listed in Appendix B.1.1. This results are summarized in performance profiles as follows. In Figure 6.1, we

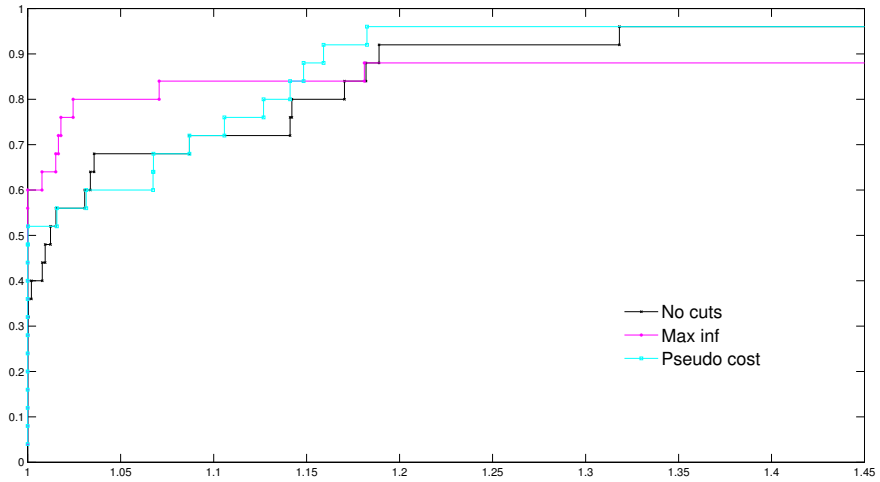


Figure 6.1: Performance profile for pseudo cost branching using the size of the tree as performance measure.

have the performance profile using pseudo cost branching and using the size of the search tree as a performance measure. In this profile we compare the size of the final search tree of pure branch and bound with the size of the final tree for branch and cut using two rules to select the seed to create a DCC: most fractional and the pseudo cost. In Figure 6.2, we use the solution time as a performance measure. In this case we compare the solution time of pure branch and bound with the solution time for branch and cut using two rules to select the seed of the DCC: most fractional and the pseudo cost.

The results for the experiments using strong branching are listed in Appendix B.1.2. These results are summarized in performance profiles as follows. In Figure 6.3 we have the performance profile using strong branching and using the size of the search tree as a performance measure. In this profile we compare the size of the final search tree of pure branch and bound with the size of the final tree for branch and cut using two rules to select the seed to derive a DCC: most fractional and the pseudo cost. In Figure 6.4 we use the solution time as a performance measure. In this case we compare the solution time of

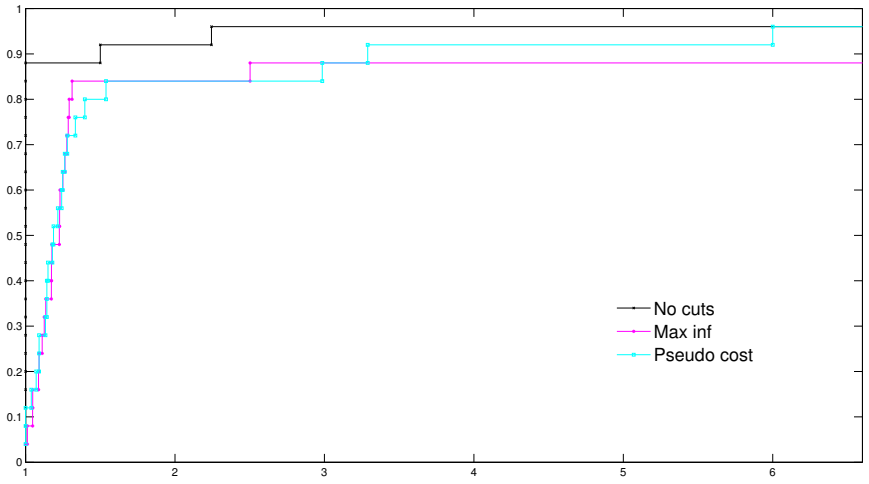


Figure 6.2: Performance profile for pseudo cost branching using the solution time as performance measure.

pure branch and bound with the solution time for branch and cut using two rules to select the seed of the DCC: most fractional and pseudo cost.

6.1.1.2 Experiments with cut manager

In this experiments we tested our 30 problems using pseudo costs branching and our implementation of a cut manager. We ran three experiments. First we solve the problems with pure branch and bound. Second, we solve the problems using the most fractional criteria for selecting the seed to create the DCCs. Finally, we solve the problems using the pseudo costs for selecting the seed to create the DCCs. In each of these two last experiments we created one DCC in every node for each cone as long as a limit of 15 DCCs is not reached. In this experiments we set the limit age for a DCC to be 10.

The results for the experiments are listed in Appendix B.2. These results are summarized in performance profiles as follows. In Figure 6.5 we have the performance profile using the size of the search tree as performance measure. In this profile we compare the

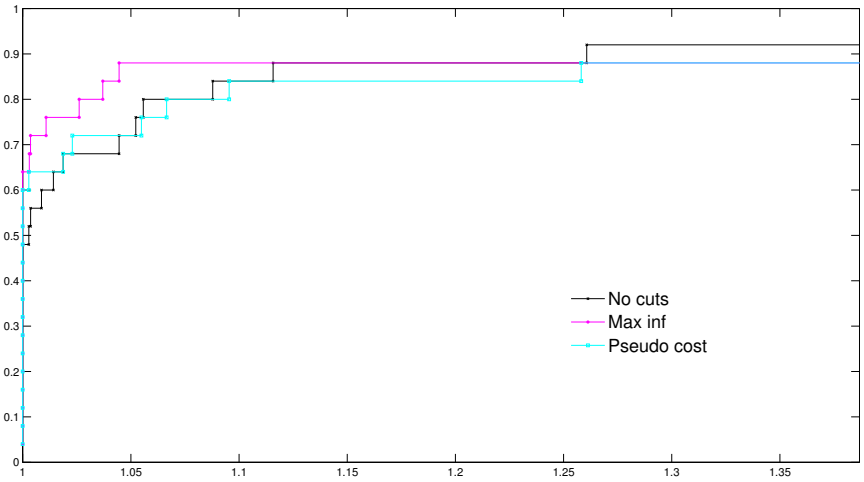


Figure 6.3: Performance profile for strong branching using the size of the tree as performance measure.

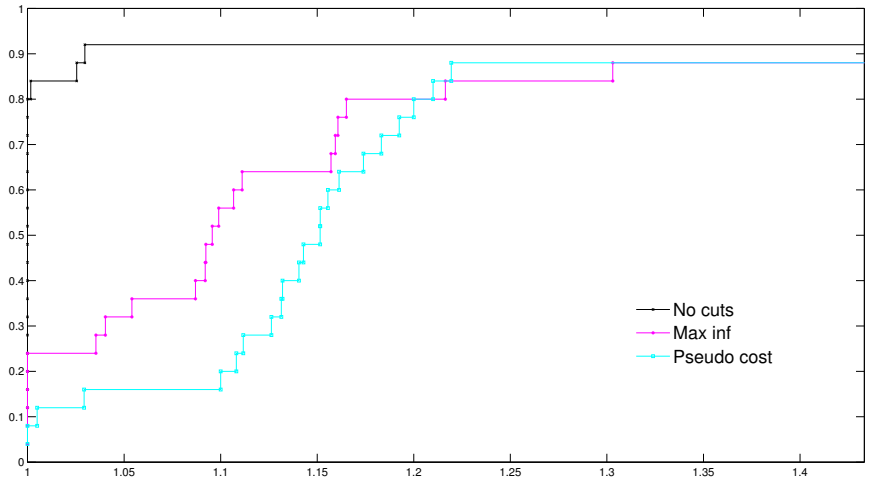


Figure 6.4: Performance profile for strong branching using the solution time as performance measure.

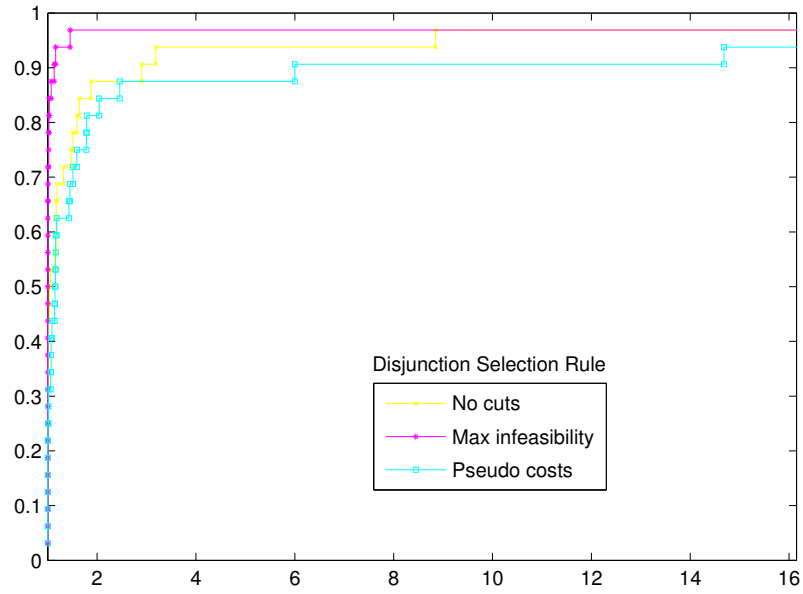


Figure 6.5: Performance profile with cut manager using the size of the tree as performance measure.

size of the final search tree of a pure branch and bound with the size of the final tree for a branch and cut using two rules to select the seed of the DCC: most fractional and the pseudo cost. In Figure 6.6 we use the solution time as performance measure. In this case we compare the solution time of a pure branch and bound with the solution time for a branch and cut using two rules to select the seed of the DCC: most fractional and the pseudo cost.

6.1.1.3 Main findings

The main conclusion from these experiments is that the use of DCCs help to decrease the size of the tree significantly. This conclusion can be drawn from the results presented in the profiles in Figures 6.1, 6.3, and 6.5. In each of these cases we can see that the addition of cuts helped the branch and cut to outperform pure branch and bound when the tree

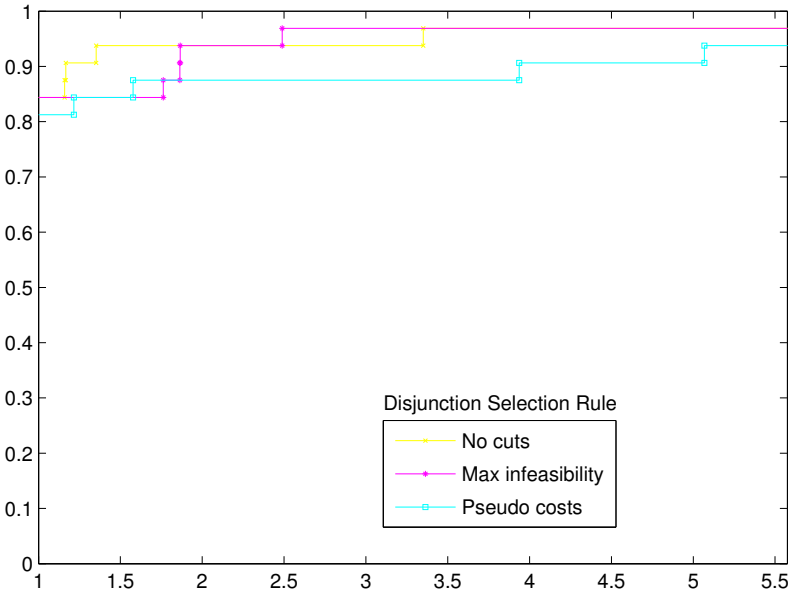


Figure 6.6: Performance profile for strong branching using the solution time as performance measure.

size is use as the performance measure. Additionally, we can see from these results that the criteria for selecting the disjunction affects the performance of branch and cut using DCCs. More specifically, the most fractional rule seems to perform better than using the pseudo cost to select the DCCs seed. Finally, from Figure 6.5 we can see that the usage of the cut manager helped to improve the performance of the branch and cut using the most fractional criteria to select the DCCs seed.

On the other hand, it is clear from the results in Figures 6.2, 6.4, and 6.6, that the use of DCCs may affect the solution time significantly. The reason behind this behavior is that the addition of DCCs to the problem may increase the solution time of the relaxations in each node significantly. This is one of the reasons why the number of DCCs added at each node has to be limited. However, in Figure 6.6 we can see that the usage of the cut manager helped to improve the performance of the branch and cut when using the most fractional criteria to select the DCCs seed. This improvement is explain by the significant reduction in the tree size in many of the problems in this case. For example, in problem R17.C30.Cones3.Int18 in Appendix B.2 we finished with a tree less than a quarter the size of the tree of pure branch and bound. Another example is problem R27.C50.Cones5.Int40, where we obtained a final tree that is almost halt the size of the tree of pure branch and bound. This is important to notice because the solution time for the relaxations still increases with the addition of cuts in this case.

Finally, in the addition of the DCCs, we have to consider the numerical difficulties that arise when adding too many DCCs. For example, after certain number of cuts the solver may complain about constraints that are close to be linearly dependent, this forces us to limit the number of DCCs to add. In general, the generation of DCCs faces numerical issues similar to the ones seen when adding disjunctive cuts in MILO Cook et al. [2009], Cornuéjols et al. [2012], Margot [2009]. We still need to do a more extensive research in this difficulties in order to tune our implementation and improve the effectiveness of the

DCCs.

6.2 Problems from public libraries

The second source we have considered is public test set libraries of problems which have convex quadratic problems. Currently, we have a small set of problems taken from the open source MINLP project source MINLP Project. These are Constrained Layout (CLay) problems, which were presented in Bonami et al. [2008]. The main characteristic of these problems is the presence of convex quadratic constraints that are three dimensional paraboloids. For example, one of the constraints in problem CLay0304M is

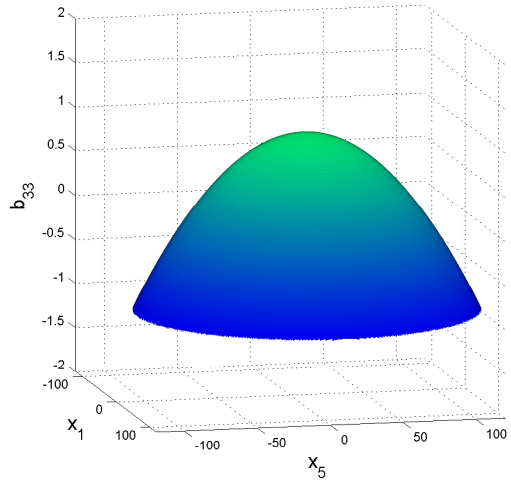
$$(x_1 - 17.50)^2 + (x_5 - 7.00)^2 + 6814b_{33} \leq 6850. \tag{6.1}$$

This constraint is illustrated in Figure 6.7(a).

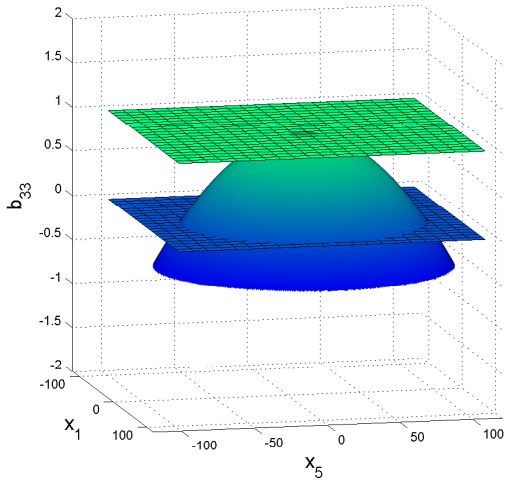
An important observation in this constraint is that b_{33} is a binary variable. Hence, we can create a DCC for this variable using the disjunction $b_{33} \leq 0$, $b_{33} \geq 1$, which is shown in Figure 6.7(b). The DCC in this case is derived using the result in Lemma 4.4. The resulting DCC for (6.1) is given by

$$\sqrt{(x_1 - 15.50)^2 + (x_5 - 7.00)^2} \leq 82.77 - 76.76b_{33}, \tag{6.2}$$

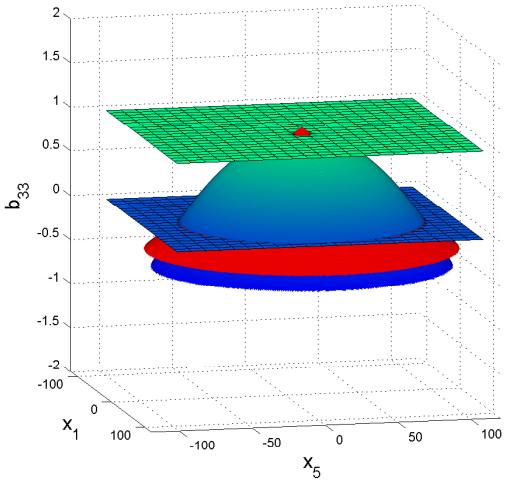
which is illustrated in Figure 6.7(c). Finally, the intersection between the quadric defined by the quadratic constraint (6.1) and the DCC defined by (6.2) is shown in Figure 6.7(d), where $0 \leq b_{33} \leq 1$. Note that since b_{33} is binary, in this case the DCC (6.2) dominates the constrain (6.1), hence we can replace it with the DCC to obtain a leaner formulation of the problem. Thus, for these problems DCCs may be use to tighten the problem formulation in the preprocessing phase.



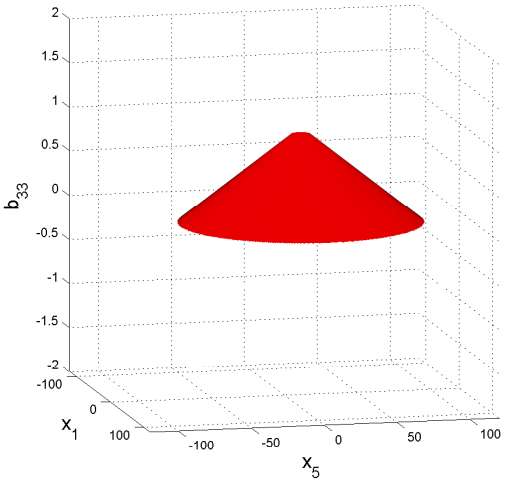
(a) CLay Quadratic Constraint.



(b) Disjunction $b_{33} \leq 0$, $b_{33} \geq 1$.



(c) Resulting DCC for the CLay constraint.



(d) Feasible set using the DCC formulation.

Figure 6.7: Preprocessing the CLay problems by DCCs.

6.2.1 Experiments with CLay problems

In our experiments, we replace all the quadratic constraints with the corresponding DCCs before solving the problem. Additionally, note that once this is done, we have exhausted all the DCCs that can be derived for this problem. For that reason, our goal in this case

is to compare the original formulation with the formulation that is the result of using the DCCs.

Table 6.1 presents some of the main characteristics of the CLay problems. There we show how many variables are in the problem and how many of those variables are binary. Additionally, we give the number of constraints present in the problem, and how many of those constraints a quadratic constraints in the form (6.1).

	0203M	0204M	0205M	0303M	0304M	0305M
Var	31	52	81	34	57	86
Binary	18	21	50	21	36	55
Constraints	55	91	136	67	107	156
Quad	24	32	40	36	48	60

Table 6.1: Description of the CLay problems

Using CPLEX 12.4 we solved the original problems and also the preprocessed problems where all the quadratic constraints are replaced by the DCCs. The results are shown in Tables 6.2 and 6.3, and the comparison is shown in Table 6.4.

	0203M	0204M	0205M	0303M	0304M	0305M
Time	0.84	1.12	2.03	1.001	2.02	4.04
Nodes	167	738	8212	453	2549	11188
Iter	1677	3601	47125	6483	23560	65174
Obj	41572.98	6545.00	8092.5	26668.75	40261.08	8092.5

Table 6.2: Results for the original formulation when solved by CPLEX 12.4.

	0203M	0204M	0205M	0303M	0304M	0305M
Time	0.44	0.41	1.56	0.467	1.19	1.80
Nodes	165	656	6244	481	1336	8957
Iter	1285	3302	37118	3190	11336	62290
Obj	41565.61	6545.00	8092.5	26662.49	40241.57	8092.5

Table 6.3: Results for the DCC formulation when solved by CPLEX 12.4.

We repeated the previous experiment with MOSEK 6.0. The results are shown in Tables 6.5 and 6.6, and the comparison is shown in Table 6.7.

	0203M	0204M	0205M	0303M	0304M	0305M
Time	48%	63%	23%	53%	41%	55%
Nodes	1%	11%	24%	-6%	47%	20%
Iter	23%	8%	21%	51%	52%	4%

Table 6.4: Comparison of the original and the DCC formulation when solved by CPLEX 12.4.

	0203M	0204M	0205M	0303M	0304M	0305M
Time	3.06	16.91	339.40	7.15	101.98	621.41
Nodes	484	1974	25400	868	8467	38184
Iter	6981	28450	377914	12674	130714	570935
Obj	41573.26	6545.00	8092.5	26669.10	40262.38	8092.50

Table 6.5: Results for the original formulation when solved by MOSEK 6.0.

	0203M	0204M	0205M	0303M	0304M	0305M
Time	2.29	15.10	207.90	5.84	76.74	487.46
Nodes	400	2194	20528	838	7013	32875
Iter	5272	27714	271433	10944	104978	455239
Obj	41565.75	6545.00	8092.50	26652.50	40241.57	8092.50

Table 6.6: Results for the DCC formulation when solved by MOSEK 6.0.

	0203M	0204M	0205M	0303M	0304M	0305M
Time	25%	11%	39%	18%	25%	22%
Nodes	17%	-11%	19%	3%	17%	14%
Iter	24%	3%	28%	14%	20%	20%

Table 6.7: Comparison of the original and DCC formulation when solved by MOSEK 6.0.

6.2.1.1 Main findings

We can see from Table 6.4 that the addition of the DCCs in all the problems resulted in a reduction of the solution time when solved by 12.4. The average of this reduction is 47%. In general, we can observe a reduction in the search tree as well, except for problem 0304M. Hence, in this experiments it is apparent that the use of the DCCs helped to speed up the solution process.

From table Table 6.7 we can see also a consistent result, i.e., the addition of the DCCs in all the problems resulted in a reduction of the solution time when solved by MOSEK 6.0.

The average of this reduction in this case is 23.3%. In general, we can observe a reduction in the search tree as well, except for problem 0205M. Hence, from these experiments it is apparent that the use of the DCCs helped to speed up the solution process significantly.

Chapter 7

Conclusions and Future Research

In this thesis, we investigated the derivation of disjunctive conic cuts (DCCs) and Disjunctive Cylindrical Cuts (DCyC) for MISOCO problems. This was achieved by extending the ideas of disjunctive programming that have been applied successfully for obtaining linear cuts for MILO problems. We introduced first the concept of DCCs and DCyCs, which are an extension of the disjunctive cuts that have been well studied for MILO problems. We were able to show under some mild assumptions that the intersection of this cuts with a closed convex set is the convex hull of the intersection of the same set with a linear disjunction. This property makes this cuts the tightest cuts possible for the last intersection. We have also provided the characterization of the family of quadrics having the same intersection with two given hyperplanes. We show the existence of a cylinder or a cone in that family in two cases:

- when the hyperplanes are parallel and there is a quadric in the family that is defined by a matrix with at most one non-positive eigenvalue;
- when the hyperplanes are non-parallel and there a quadric in the family that is defined by a positive definite matrix.

The two aforementioned analysis are then put together to provide a procedure for the derivation of DCCs and DCyCs separating a given point from the feasible set of a MISOCO problem. Some preliminary experiments performed with our test sets have shown encouraging results about the usage of the DCCs. However, these are still to limited to draw any conclusive results about their performance.

Giving the encouraging results in our preliminary experiments, we believe that it would be interesting to pursue a more extensive study about the performance of this cuts. Specially, considering the computational challenges that are mentioned in Chapter 5. Another issue is related with the lack of warm starting and purification algorithms in interior point methods based solvers. This two issues have shown to be a challenge during our experiments. On one hand, the lack of warm start is reflected with the increase in the solution time of the SOCO relaxations in each node of the search tree. On the other hand, the lack of purification algorithms affected the performance of the algorithm depending on the branching rule and the criteria for selecting the seed to derive the DCC cuts. Overall, we believe that a more extensive experimentation would provide more insight to move towards a practical use of DCCs and DCyCs for solving problems in real engineering applications. An important effort in this direction is the construction of a MISOCO test set library with different problem structures, and applications.

Finally, we think that another interesting research direction would be associated with the extensions of the idea of DCCs and DCyCs to p-cone and positive-semidefinite optimization. Given the generality of the result provided in Chapter 2, we believe is worth to explore how to extend the ideas developed for the MISOCO problems to re realm of these two problems.

Bibliography

T. Achterberg, T. Koch, and A. Martin. Branching rules revisited. *Operations Research Letters*, 33(1):42–54, 2005.

M.S. Aktürk, A. Atamtürk, and S. Gürel. A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Operations Research Letters*, 37(3):187–191, 2009.

F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

E.D. Andersen, C. Roos, and T. Terlaky. On implementing a primal-dual interior-point method for conic quadraticoptimization. *Mathematical Programming*, 95:249–277, 2003.

A. Atamtürk and V. Narayanan. Conic mixed-integer rounding cuts. *Mathematical Programming*, 122(1):1–20, 2010.

A. Atamtürk and V. Narayanan. Lifting for conic mixed-integer programming. *Mathematical Programming A*, 126:351–363, 2011.

A. Atamtürk, G. Berenguer, and Z.J. Shen. A conic integer programming approach to stochastic joint location-inventory problems. *Operations Researchs*, 60(2):366–381, 2012.

E. Balas. Disjunctive programming. In P.L. Hammer, E.L. Johnson, and B.H. Korte,

editors, *Annals of Discrete Mathematics 5: Discrete Optimization*, pages 3–51. North Holland, 1979.

E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58:295–324, 1993.

A. Barvinok. *A course in Convexity*. American Mathematical Society, Providence, Rhode Island, USA, 2002.

P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter. Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods and Software*, 24(4-5):597–634, 2009.

P. Belotti, J.C. Góez, I. Pólik, T.K. Ralphs, and T. Terlaky. A conic representation of the convex hull of disjunctive sets and conic cuts for integer second order cone optimization. Technical Report 12T-009, Lehigh University, Department of Industrial and Systems Engineering, 2012. Submmited to Mathematical Programming.

P. Belotti, J.C. Góez, I. Pólik, T.K. Ralphs, and T. Terlaky. Disjunctive conic cuts for mixed integer second order cone optimization. In preparation for publication, 2013a.

P. Belotti, J.C. Góez, I. Pólik, T.K. Ralphs, and T. Terlaky. On families of quadratic surfaces having fixed intersections with two hyperplanes. *Discrete Applied Mathematics*, 2013b. doi: http://dx.doi.org/10.1016/j.dam.2013.05.017. Available online.

A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, Pennsylvania, USA, 2001a.

A. Ben-Tal and A. Nemirovski. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001b.

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

P. Bonami, L.T. Biegler, A.R. Conn, G. Cornuéjols, I.E. Grossmann, C.D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, and A. Wächter. An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization*, 5(2):186–204, 2008. In Memory of George B. Dantzig.

P. Bonami, J. Lee, S. Leyffer, and A. Wächter. More branch-and-bound experiments in convex nonlinear integer programming. Technical Report ANL/MCS-P1949-0911, Argonne National Laboratory, Mathematics and Computer Science Division, 2011.

S. Boyd and L. Vandenberghe. *Convex Optimization*. University Press, Cambridge, UK, 2006.

S.B. Çay, Y. Fu, J.C. Góez, and T. Terlaky. Library of test sets for mixed interger second order cone optimization. Working paper, 2013.

M.T. Çezik and G. Iyengar. Cuts for mixed 0-1 conic programming. *Mathematical Programming*, 104(1):179–202, 2005.

Y. Cheng, S. Drewes, A. Philipp, and M. Pesavento. Joint network optimization and beamforming for coordinated multi-point transmission using mixed integer programming. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3217–3220, 2012.

W. Cook, S. Dash, R. Fukasawa, and M. Goycoolea. Numercally safe gomory mixed-integer cuts. *INFORMS Journal on Computing*, 21(4):641–649, 2009.

G. Cornuéjols. Valid inequalities for mixed integer linear programs. *Mathematical Programming*, 112(1):3–44, 2008.

G. Cornuéjols, F. Margot, and G. Nannicini. On the safety of gomory cug generators. Technical Report 1412, Tepper School of Bussines, 2012.

D.A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer-Verlag, Secaucus, NJ, USA, 1997.

CPLEX. *IBM ILOG CPLEX Optimization Studio V12.5*. IBM, 2011.

D. Dadush, S.S. Dey, and J.P. Vielma. The split closure of a strictly convex body. *Operations Research Letters*, 39(2):121–126, 2011.

S. Drewes. *Mixed Integer Second Order Cone Programming*. PhD thesis, Technische Universität Darmstadt, Germany, 2009.

M. Fampa and N. Maculan. Using a conic formulation for finding Steiner minimal trees. *Numerical Algorithms*, 35(2):315–330, 2004.

M. Fischetti, F. Glover, and A. Lodi. The feasibility pump. *Mathematical Programming*, 104(1):91–104, 2005.

G.H. Golub. Some modified matrix eigenvalues problems. *SIAM Review*, 15(2):318–334, 1973.

G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996.

R.E. Gomory. Outline of an algorithm for integer solutions to linear programs. *Bulletin of American Mathematical Society*, 64:275–278, 1958.

I.E. Grossmann. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optimization and Engineering*, 3:227–252, 2002.

O.K. Gupta and A. Ravindra. Branch and bound experiments in convex nonlinear integer programming. *Management Science*, 31(12):1533–1546, 1985.

GUROBI. *Gurobi Optimizer Reference Manual, Version 5.5*. GUROBI optimization, 2013.

J. Harris. *Algebraic Geometry: A First Course*. Springer-Verlag, New York, NY, USA, 1992.

P.A. Krokhmal and P. Soberanis. Risk optimization with p-order conic constraints: A linear programming approach. *European Journal of Operational Research*, 201(3):653–671, 2010.

M.P. Kumar, P.H.S. Torr, and A. Zisserman. Solving Markov random fields using second order cone programming relaxations. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1045–1052, 2006.

Y. Kuo and H.D. Mittelmann. Interior point methods for second-order cone programming and or applications. *Computational Optimization and Applications*, 28:255–285, 2004.

P. Lancaster and M. Tismenetsky. *The Theory of Matrices, With Applications*. Academic Press, INC., London, Uk, 2nd edition, 1985.

E.L. Lawler and D.E. Wood. Branch-and-bound methods: A survey. *Operations Research*, 14(4):699–719, 1966.

J.T. Linderoth and M.W.P. Savelsbergh. A computational study of search strategies for mixed integer programming. *INFORMS Journal on Computing*, 11(2):173–187, 1999.

M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Its Applications*, 284(3):193 – 228, 1998.

A. Mahajan. *On Selecting Disjunctions for Solving Mixed Integer Programming Problems*. PhD thesis, Lehigh University, 2009.

A. Mahajan. Preprocessing techniques for integer programming. In J. Cochran, editor, *Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2010.

F. Margot. Testing cut generators for mixed-integer linear programming. *Mathematical Programming Computation*, 1(1):69–95, 2009.

A. Martin. General mixed integer programming: Computational issues for branch-and-cut algorithms. In Michael Jnger and Denis Naddef, editors, *Computational Combinatorial Optimization*, volume 2241 of *Lecture Notes in Computer Science*, pages 1–25. Springer Berlin, 2001.

C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA, USA, 2000.

J.E. Mitchell. Branch-and-cut algorithms for combinatorial optimization problems. In P.M. Pardalos and M.G.C. Resende, editors, *Handbook of Applied Optimization*, chapter 3, pages 65–77. Oxford University Press, Oxford, UK, 2002.

MOSEK. *The MOSEK Optimization Tools Manual, Version 6.0*. MOSEK, 2011.

G.L. Nemhauser and L.A. Wolsey. A recursive procedure to generate all cuts for 01 mixed integer programs. *Mathematical Programming*, 46(1-3):379–390, 1990.

G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization*. Wiley-Interscience, New York, NY, USA, 1999.

T.K. Ralphs. Parallel branch and cut. In E. Talbi, editor, *Parallel Combinatorial Optimization*, pages 53–101. Wiley, New York, NY, USA, 2006.

T.K. Ralphs, M. Güzelsoy, and A. Mahajan. SYMPHONY 5.4 user's manual. Technical report, COR@L Laboratory, Lehigh University, 2011.

R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, USA, 1970.

S. Sanjeevi. *Mixed n-step MIR Inequalities, n-step Conic MIR Inequalities and a Polyhedral Study of Single Row Facility Layout Problem*. PhD thesis, Texas A&M University, 2012.

M.W.P. Savelsbergh. Preprocessing and probing techniques for mixed integer programming problems. *INFORMS Journal on Computing*, 6(4):445–454, 1994.

A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1986.

S.R. Searle. *Matrix Algebra Useful for Statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience, New York, NY, USA, 1982.

V. Snyder and C.H. Sisam. *Analytic Geometry of Space*. H. Holt and Company, New York, NY, USA, 1914.

CMU-IBM Open source MINLP Project. URL http://egon.cheme.cmu.edu/ibm/page.htm.

R.A. Stubbs and S. Mehrotra. A branch-and-cut method for 0-1 mixed convex programming. *Mathematical Programming*, 86(3):515–532, 1999.

J.F. Sturm. Implementation of interior point methods for mixed semidefinite and second order cone optimization problems. *Optimization Methods and Software*, 17(6):1105 – 1154, 2002.

K.C. Toh, M.J. Todd, and R.H. Tütüncü. SDPT3 - A MATLAB software package for semidefinite programming, Version 1.3. *Optimization Methods and Software*, 11(1-4): 545 – 581, 1999.

J.P. Vielma. *Mixed Integer Programming Approaches for Nonlinear and Stochastic Programming*. PhD thesis, Georgia Institute of Technology, 2009.

J.P. Vielma, S. Ahmed, and G.L. Nemhauser. A lifted linear programming branch-and-bound algorithm for mixed-integer conic quadratic programs. *INFORMS Journal on Computing*, 20(3):438–450, 2008.

D.J. White. A Lagrangean relaxation approach for a turbine design quadratic assignment problem. *Journal of the Operational Research Society*, 47:766–775, 1996.

Xpress. *FICO Xpress Optimization Suite, Xpress-Optimizer reference manual.* FICO.

Y. Xu, T.K. Ralphs, L. Ladányi, and M.J. Saltzman. ALPS: A framework for implementing parallel search algorithms. In *The Proceedings of the Ninth INFORMS Computing Society Conference*, pages 319–334, 2005.

Y. Xu, T.K. Ralphs, L. Ladányi, and M.J. Saltzman. Computational experience with a software framework for parallel integer programming. *INFORMS Journal on Computing*, 21(3):383–397, 2009.

Appendix A

Additional lemmas for Chapter 4

Lemma A.1. *In the first and second cases of Theorem 3.2 the vertex $x(\bar{\tau}_2)$ of the quadric $\mathcal{Q}(\bar{\tau}_2)$ is either in $\mathcal{A}$ or $\mathcal{B}$.*

Proof. Recall from Section 3.2.3 that the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_1)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$, are computed using the roots of the function (3.20), which is

$$f(\tau) = \tau^2 \frac{(\alpha-\beta)^2}{4} + \tau(1-\alpha\beta) + 1.$$

The roots of $f(\tau)$ are

$$\bar{\tau}_1 = 2\left(\frac{\alpha\beta - 1 - \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha-\beta)^2}\right),$$
$$\bar{\tau}_2 = 2\left(\frac{\alpha\beta - 1 + \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha-\beta)^2}\right),$$

where $\bar{\tau}_1 \leq \bar{\tau}_2$. Note that if $\alpha = \beta$, then $f(\tau)$ does not have two roots. In this case we would have that $\mathcal{A} = \mathcal{B}$ and is easy to verify that $\text{conv}(\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})) = \mathcal{Q}$. However, recall that our assumption is $\beta \neq \alpha$. Hence, for the rest of this proof we assume that $\alpha \neq \beta$.

The vertices of the cones $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_2)$ are $x(\bar{\tau}_i) = -P(\bar{\tau}_i)^{-1}p(\bar{\tau}_i)$, $i = 1, 2$. We can

express $x(\bar{\tau}_i)$ in terms of a, α, and β as follows

$$\begin{aligned} x(\bar{\tau}_i) = -P(\bar{\tau}_i)^{-1}p(\bar{\tau}_i) &= -\left(I - \frac{\bar{\tau}_i}{(1+\bar{\tau}_i)}aa^\top\right)\left(-\bar{\tau}_i\frac{(\alpha+\beta)}{2}a\right) \\ &= \bar{\tau}_i\frac{(\alpha+\beta)}{2}\left(1 - \frac{\bar{\tau}_i}{(1+\bar{\tau}_i)}\right)a \\ &= \bar{\tau}_i\frac{(\alpha+\beta)}{2(1+\bar{\tau}_i)}a. \end{aligned}$$

Consider the inner product

$$a^\top x(\bar{\tau}_i) = -a^\top P(\bar{\tau}_i)^{-1}p(\bar{\tau}_i) = \bar{\tau}_i\frac{(\alpha+\beta)}{2(1+\bar{\tau}_i)}a^\top a = \bar{\tau}_i\frac{(\alpha+\beta)}{2(1+\bar{\tau}_i)}.$$

Note that if $\alpha = -\beta$ then $a^\top x(\bar{\tau}_i) = 0$. Recall from Section 3.2.3.2 that in that case $Q(\bar{\tau}_1)$ is a cylinder, and its analysis is presented in Section 4.2.1. For that reason, we assume that $\alpha \neq -\beta$ for the rest of this proof.

Next, we analyze where the vertices $x(\bar{\tau}_1)$ and $x(\bar{\tau}_2)$ are located with respect to the half-spaces $\mathcal{A}$ and $\mathcal{B}$. Recall from Section 4.2 the assumption that $\beta \leq \alpha$. First of all, from Section 3.2 we know that $\bar{\tau}_1 \leq \bar{\tau}_2 < -1$, $|\alpha| \leq 1$, and $|\beta| \leq 1$. Now, we have that $1 + \frac{1}{\bar{\tau}_i} > 0$,

$$a^\top x(\bar{\tau}_i) = \bar{\tau}_i\frac{(\alpha+\beta)}{2(1+\bar{\tau}_i)} = \frac{(\alpha+\beta)}{2(1+\frac{1}{\bar{\tau}_i})},$$

and

$$\lim_{\bar{\tau}_i \to -\infty} a^\top x(\bar{\tau}_i) = \frac{(\alpha+\beta)}{2}.$$

To complete the proof we need to consider two cases:

- First assume that $|\beta| < |\alpha|$, which implies that $\alpha > 0$. Then, since $\alpha + \beta > 0$ we obtain that $a^\top x(\bar{\tau}_i) > 0$ and

$$\lim_{\bar{\tau}_i \nearrow -1} a^\top x(\bar{\tau}_i) = \infty.$$

Additionally, since $\frac{(\alpha+\beta)}{2} > \beta$ we obtain that $a^\top x(\bar{\tau}_i) > \beta$. Let us now evaluate when $a^\top x(\bar{\tau}_i) \geq \alpha$, then we have that

$$\bar{\tau}_i \frac{(\alpha+\beta)}{2(1+\bar{\tau}_i)} \geq \alpha \Rightarrow \bar{\tau}_i \geq \frac{2\alpha}{(\beta-\alpha)}.$$

On one hand, for $\bar{\tau}_1$ we obtain that

$$\bar{\tau}_1 = 2\left(\frac{\alpha\beta - 1 - \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha-\beta)^2}\right) \leq 2\left(\frac{\alpha\beta-1}{(\alpha-\beta)^2}\right) \leq 2\left(\frac{\alpha\beta-\alpha^2}{(\alpha-\beta)^2}\right) = \frac{2\alpha}{(\beta-\alpha)}.$$

On the other hand, for $\bar{\tau}_2$ we obtain that

$$\begin{aligned}\bar{\tau}_2 = 2\left(\frac{\alpha\beta - 1 + \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha-\beta)^2}\right) &\geq 2\left(\frac{\alpha\beta-1+\sqrt{(1-\alpha^2)^2}}{(\alpha-\beta)^2}\right) \\ &= 2\left(\frac{\alpha\beta-1+(1-\alpha^2)}{(\alpha-\beta)^2}\right) = \frac{2\alpha}{(\beta-\alpha)}.\end{aligned}$$

Hence, if $|\beta| < |\alpha|$, then $x(\bar{\tau}_1) \notin \mathcal{A} \cup \mathcal{B}$ and $x(\bar{\tau}_2) \in \mathcal{A}$.

- Second, assume that $|\alpha| < |\beta|$, which implies that $\beta < 0$. Then, since $\alpha + \beta < 0$ we obtain that $a^\top x(\bar{\tau}_i) < 0$ and

$$\lim_{\bar{\tau}_i \nearrow -1} a^\top x(\bar{\tau}_i) = -\infty.$$

Additionally, since $\frac{(\alpha+\beta)}{2} < \alpha$ we obtain that $a^\top x(\bar{\tau}_i) < \alpha$. Let us now evaluate when $a^\top x(\bar{\tau}_i) \leq \beta$, then we have that

$$\bar{\tau}_i \frac{(\alpha+\beta)}{2(1+\bar{\tau}_i)} \leq \beta \Rightarrow \bar{\tau}_i \geq \frac{2\beta}{(\alpha-\beta)}.$$

For $\bar{\tau}_1$ we obtain that

$$\bar{\tau}_1 = 2\left(\frac{\alpha\beta - 1 - \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha-\beta)^2}\right) \leq 2\left(\frac{\alpha\beta - 1}{(\alpha-\beta)^2}\right) \leq 2\left(\frac{\alpha\beta - \beta^2}{(\alpha-\beta)^2}\right) = \frac{2\beta}{(\alpha-\beta)}.$$

On the other hand, for $\bar{\tau}_2$ we obtain that

$$\begin{aligned}\bar{\tau}_2 = 2\left(\frac{\alpha\beta - 1 + \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha-\beta)^2}\right) &\geq 2\left(\frac{\alpha\beta - 1 + \sqrt{(1-\beta^2)^2}}{(\alpha-\beta)^2}\right) \\ &= 2\left(\frac{\alpha\beta - 1 + (1-\beta^2)}{(\alpha-\beta)^2}\right) = \frac{2\beta}{(\alpha-\beta)}.\end{aligned}$$

Hence, if $|\alpha| < |\beta|$, then $x(\bar{\tau}_1) \notin \mathcal{A} \cup \mathcal{B}$ and $x(\bar{\tau}_2) \in \mathcal{B}$.

□

Lemma A.2. *In the first cases of Theorems 3.3 and 3.4 the vertex $x(\bar{\tau}_1)$ of the quadric $\mathcal{Q}(\bar{\tau}_1)$ is either in $\mathcal{A}$ or $\mathcal{B}$.*

Proof. From Section 3.2.5.5 we have that

$$a^\top x(\bar{\tau}_1) = \bar{\tau}_1 \frac{(\alpha+\beta)(1-2a_1^2)}{2(1+\bar{\tau}_1(1-2a_1^2))}.$$

Recall also that $-\frac{1}{(1-2a_1^2)} \leq \bar{\tau}_1$. Hence,

$$\lim_{\bar{\tau}_i \to \infty} a^\top x(\bar{\tau}_1) = \frac{(\alpha+\beta)}{2}.$$

On the other hand, we have

$$\lim_{\bar{\tau}_i \searrow -\frac{1}{(1-2a_1^2)}} a^\top x(\bar{\tau}_1) = \begin{cases} -\infty & \text{if } \alpha+\beta > 0 \\ +\infty & \text{if } \alpha+\beta < 0. \end{cases}$$

Thus, if $\alpha+\beta>0$, then $a^\top x(\bar{\tau}_1)<\alpha$. Now, if $a^\top x(\bar{\tau}_1)\leq\beta$ is true, then we obtain that

$$\bar{\tau}_1\frac{(\alpha+\beta)(1-2a_1^2)}{2(1+\bar{\tau}_1(1-2a_1^2))}\leq\beta \text{ that implies } \bar{\tau}_1\leq\frac{2\beta}{(\alpha-\beta)(1-2a_1^2)}.$$

On the other hand, if $\alpha+\beta<0$, then $a^\top x(\bar{\tau}_1)>\beta$. Now, if $a^\top x(\bar{\tau}_1)\geq\alpha$ is true, then we obtain that

$$\bar{\tau}_1\frac{(\alpha+\beta)(1-2a_1^2)}{2(1+\bar{\tau}_1(1-2a_1^2))}\geq\alpha \text{ that implies } \bar{\tau}_1\leq\frac{-2\alpha}{(\alpha-\beta)(1-2a_1^2)}.$$

Recall that $\beta<\alpha$. Then, $\alpha+\beta>0$ implies that $\alpha>0$ and $\alpha>|\beta|$. Additionally, $\alpha+\beta<0$ implies that $\beta<0$ and $\beta<-|\alpha|$.

For the first case of Theorem 3.3 we need to consider two cases. On one hand if $\alpha\beta\geq0$, then $\bar{\tau}_1=0$. In this case if $\alpha+\beta>0$, then $\frac{2\beta}{(\alpha-\beta)(1-2a_1^2)}\geq0$, and $x(\bar{\tau}_1)\in\mathcal{B}$. Additionally, if $\alpha+\beta<0$, then $\frac{-2\alpha}{(\alpha-\beta)(1-2a_1^2)}\geq0$, and $x(\bar{\tau}_1)\in\mathcal{A}$. On the other hand, if $\alpha\beta\leq0$, then $\bar{\tau}_1=\frac{4\alpha\beta}{(1-2a_1^2)(\alpha-\beta)^2}\leq0$. Hence, if $\alpha+\beta>0$, then

$$\frac{4\alpha\beta}{(1-2a_1^2)(\alpha-\beta)^2}=\left(\frac{2\beta}{(1-2a_1^2)(\alpha-\beta)}\right)\left(\frac{2\alpha}{(\alpha-\beta)}\right)\leq\frac{2\beta}{(\alpha-\beta)(1-2a_1^2)},$$

and the vertex $x(\bar{\tau}_1)\in\mathcal{B}$. Additionally, if $\alpha+\beta<0$, then

$$\frac{4\alpha\beta}{(1-2a_1^2)(\alpha-\beta)^2}=\left(\frac{2\alpha}{(1-2a_1^2)(\alpha-\beta)}\right)\left(\frac{2\beta}{(\alpha-\beta)}\right)\leq\frac{-2\alpha}{(\alpha-\beta)(1-2a_1^2)},$$

and the vertex $x(\bar{\tau}_1)\in\mathcal{A}$.

For the first case of Theorem 3.4 recall that

$$\begin{aligned}\bar{\tau}_1 &= \frac{2\left(1-2a_1^2+\alpha\beta-\sqrt{(1-2a_1^2+\alpha\beta)^2+(1-2a_1^2)(\alpha-\beta)^2}\right)}{(1-2a_1^2)(\alpha-\beta)^2} \\ &= \frac{2\left(1-2a_1^2+\alpha\beta-\sqrt{(1-2a_1^2+\alpha^2)(1-2a_1^2+\beta^2)}\right)}{(1-2a_1^2)(\alpha-\beta)^2}.\end{aligned}$$

Hence, if $\alpha+\beta>0$, then

$$\begin{aligned}\frac{2\left(1-2a_1^2+\alpha\beta-\sqrt{(1-2a_1^2+\alpha^2)(1-2a_1^2+\beta^2)}\right)}{(1-2a_1^2)(\alpha-\beta)^2} &\leq \frac{2\left(1-2a_1^2+\alpha\beta-(1-2a_1^2+\beta^2)\right)}{(1-2a_1^2)(\alpha-\beta)^2} \\ &= \frac{2\left(\alpha\beta-\beta^2\right)}{(1-2a_1^2)(\alpha-\beta)^2} \\ &= \frac{2\beta}{(\alpha-\beta)(1-2a_1^2)},\end{aligned}$$

and the vertex $x(\bar{\tau}_1)\in\mathcal{B}$. Additionally, if $\alpha+\beta<0$, then

$$\begin{aligned}\frac{2\left(1-2a_1^2+\alpha\beta-\sqrt{(1-2a_1^2+\alpha^2)(1-2a_1^2+\beta^2)}\right)}{(1-2a_1^2)(\alpha-\beta)^2} &\leq \frac{2\left(1-2a_1^2+\alpha\beta-(1-2a_1^2+\alpha^2)\right)}{(1-2a_1^2)(\alpha-\beta)^2} \\ &= \frac{2\left(\alpha\beta-\alpha^2\right)}{(1-2a_1^2)(\alpha-\beta)^2} \\ &= \frac{-2\alpha}{(\alpha-\beta)(1-2a_1^2)},\end{aligned}$$

and the vertex $x(\bar{\tau}_1)\in\mathcal{A}$. This shows that $x(\bar{\tau}_1)$ is contained in one of the sets $\mathcal{A}$ or $\mathcal{B}$. □

Lemma A.3. *In the first cases of Theorems 3.3 and 3.4 we have that* $\mathcal{Q}\cap(\mathcal{A}\cup\mathcal{B})\subseteq\mathcal{Q}(\bar{\tau}_1)$.

Proof. Recall that $\mathcal{Q}(\tau_1)=\{x\in\mathbb{R}^\ell \mid x^\top P(\bar{\tau}_1)x+2p(\bar{\tau}_1)^\top x+\rho(\bar{\tau}_1)\leq 0\}$, then from Section 3.2.5.5 we have

$$x^\top P(\bar{\tau})x+2p(\bar{\tau})^\top x+\rho(\bar{\tau})=x^\top Jx+\bar{\tau}_1\left((a^\top x)^2-\alpha a^\top x-\beta a^\top x+\alpha\beta\right),$$

and from Section 3.2.5.6 we have

$$x^\top P(\bar{\tau})x + 2p(\bar{\tau})^\top x + \rho(\bar{\tau}) = x^\top Jx + 1 + \bar{\tau}_1 \left((a^\top x)^2 - \alpha a^\top x - \beta a^\top x + \alpha\beta\right).$$

From (3.35) and (3.39) we know that $\bar{\tau}_1 \leq 0$ and for $\tilde{x} \in \mathcal{Q}$ we have either $\tilde{x}^\top J\tilde{x} \leq 0$ or $\tilde{x}^\top J\tilde{x} + 1 \leq 0$. Now, observe that $(a^\top x)^2 - \alpha a^\top x - \beta a^\top x + \alpha\beta = (a^\top x - \alpha)(a^\top x - \beta)$. On one hand, if $\tilde{x} \in \mathcal{B} \cap \mathcal{Q}$, then $(a^\top \tilde{x} - \alpha) \leq 0$ and $(a^\top \tilde{x} - \beta) \leq 0$. On the other hand, if $\tilde{x} \in \mathcal{A} \cap \mathcal{Q}$, then $(a^\top \tilde{x} - \alpha) \geq 0$ and $(a^\top \tilde{x} - \beta) \geq 0$. Thus, if $\tilde{x} \in \mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B})$, we have that

$$(a^\top \tilde{x})^2 - \alpha(a^\top \tilde{x}) - \beta(a^\top \tilde{x}) + \alpha\beta \geq 0,$$

and we obtain that $\tilde{x}^\top P(\bar{\tau})\tilde{x} + 2p(\bar{\tau})^\top \tilde{x} + \rho(\bar{\tau}) \leq 0$ for $\tilde{x} \in \mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B})$. Thus, $\mathcal{Q} \cap (\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{Q}(\bar{\tau}_1)$. □

Lemma A.4. *In the first case of Theorems 3.3 and 3.4 we have that each of the subsets $\mathcal{Q}^+ \cap \mathcal{A}$, $\mathcal{Q}^+ \cap \mathcal{B}$, $\mathcal{Q}^- \cap \mathcal{A}$, $\mathcal{Q}^- \cap \mathcal{B}$, is a subset of one of the branches $\mathcal{Q}^+(\tau_1)$ or $\mathcal{Q}^-(\tau_1)$.*

Proof. First, we show that either $\mathcal{Q}^+ \cap \mathcal{A} \subseteq \mathcal{Q}^+(\bar{\tau}_1)$ or $\mathcal{Q}^+ \cap \mathcal{A} \subseteq \mathcal{Q}^-(\bar{\tau}_1)$. We know from the definition of the sets in Section 4.2.2 that $\mathcal{Q}^+ \cap \mathcal{A}$, $\mathcal{Q}^+(\bar{\tau}_1)$, $\mathcal{Q}^-(\bar{\tau}_1)$ are convex sets and from Lemma A.3 we have that $\mathcal{Q}^+ \cap \mathcal{A} \subseteq \mathcal{Q}(\bar{\tau}_1)$. Recall from Chapter 3 that $\mathcal{Q}(\bar{\tau}_1)$ is a cone, which vertex is denoted by $x(\bar{\tau}_1)$, and recall also that $\mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{Q}^-(\bar{\tau}_1) = x(\bar{\tau}_1)$. Then, observe that if $\mathcal{Q}^+ \cap \mathcal{A} \cap \mathcal{Q}^+(\bar{\tau}_1) \neq \emptyset$, and $\mathcal{Q}^+ \cap \mathcal{A} \cap \mathcal{Q}^-(\bar{\tau}_1) \neq \emptyset$, then $x(\bar{\tau}_1) \in \mathcal{Q}^+ \cap \mathcal{A}$, otherwise $\mathcal{Q}^+ \cap \mathcal{A} \not\subseteq \mathcal{Q}(\bar{\tau}_1)$. We have

$$\begin{aligned} x(\bar{\tau}_1) = -P(\bar{\tau}_1)^{-1}p(\bar{\tau}_1) &= -\left(J - \bar{\tau}_1 \frac{Jaa^\top J}{1 + \bar{\tau}_1(1 - 2a_1^2)}\right)\left(-\bar{\tau}_1 \frac{\alpha + \beta}{2} a\right) \\ &= \bar{\tau}_1 \frac{\alpha + \beta}{2}\left(1 - \bar{\tau}_1 \frac{(1 - 2a_1^2)}{1 + \bar{\tau}_1(1 - 2a_1^2)}\right) Ja \\ &= \bar{\tau}_1 \frac{\alpha + \beta}{2(1 + \bar{\tau}_1(1 - 2a_1^2))} Ja. \end{aligned}$$

Then, we obtain that

$$x(\bar{\tau}_1)^\top J x(\bar{\tau}_1) = \bar{\tau}_1^2 \frac{(\alpha+\beta)^2(1-2a_1^2)}{4(1+\bar{\tau}_1(1-2a_1^2))^2} \geq 0.$$

Now, if $\bar{\tau}_1 = 0$, then $\mathcal{Q}(\bar{\tau}_1) = \mathcal{Q}$, and it is clear that $\mathcal{Q}^+$ is a subset of $\mathcal{Q}^+(\bar{\tau}_1)$. On the other hand, if $\bar{\tau}_1 \neq 0$, then $x(\bar{\tau}_1) \notin \mathcal{Q}$. For that reason, $x(\bar{\tau}_1) \notin \mathcal{Q}^+ \cap \mathcal{A}$, and either $\mathcal{Q}^+ \cap \mathcal{A} \cap \mathcal{Q}^+(\bar{\tau}_1) = \emptyset$ or $\mathcal{Q}^+ \cap \mathcal{A} \cap \mathcal{Q}^-(\bar{\tau}_1) = \emptyset$. Hence, $\mathcal{Q}^+ \cap \mathcal{A}$ must be a subset of either $Q^+(\bar{\tau}_1)$ or $\mathcal{Q}^-(\bar{\tau}_1)$. A similar argument can be build to show that each subset $\mathcal{Q}^+ \cap \mathcal{B}$, $\mathcal{Q}^- \cap \mathcal{A}$, $\mathcal{Q}^- \cap \mathcal{B}$, must be a subset of either $\mathcal{Q}^+(\bar{\tau}_1)$ or $\mathcal{Q}^-(\bar{\tau}_1)$.

To complete the proof, note that if one of the sets $\mathcal{Q}^+ \cap \mathcal{B}$, $\mathcal{Q}^+ \cap \mathcal{B}$, $\mathcal{Q}^- \cap \mathcal{A}$, or $\mathcal{Q}^- \cap \mathcal{B}$ is empty, then the result is trivially true for that set. □

Lemma A.5. *In the first case of Theorems 3.3 and 3.4 if $\mathcal{Q}^+ \cap \mathcal{A} \neq \emptyset$ and $\mathcal{Q}^+ \cap \mathcal{B} \neq \emptyset$, then we have either $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{Q}^+(\bar{\tau}_1)$ or $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{Q}^-(\bar{\tau}_1)$.*

Proof. From Lemma A.4 we know that $\mathcal{Q}^+ \cap \mathcal{A}$ and $\mathcal{Q}^+ \cap \mathcal{B}$ are subsets of one of the branches $\mathcal{Q}^+(\tau_1)$ or $\mathcal{Q}^-(\tau_1)$. Recall that $\mathcal{Q}^+$, $\mathcal{Q}^-$, $\mathcal{Q}^+(\tau_1)$, and $\mathcal{Q}^-(\tau_1)$ are convex sets.

Assume to the contrary that $\mathcal{Q}^+ \cap \mathcal{A} \subseteq \mathcal{Q}^+(\bar{\tau}_1)$ and $\mathcal{Q}^+ \cap \mathcal{B} \subseteq \mathcal{Q}^-(\bar{\tau}_1)$. We need to consider two cases. First, if $\mathcal{Q}$ is a cone and $0 \in \mathcal{A} \cup \mathcal{B}$, then we know from Section 3.2.5.5 that $\bar{\tau}_1 = 0$, i.e., $\mathcal{Q} = \mathcal{Q}(\bar{\tau}_1)$. Hence it is clear that $\mathcal{Q}^+ \cap (\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{Q}^+(\bar{\tau}_1)$, which contradicts the assumption.

Second, if $\mathcal{Q}$ is a hyperboloid of two sheets, or $\mathcal{Q}$ is a cone and $0 \notin \mathcal{A} \cup \mathcal{B}$, then from the proof of Lemma A.4 we know that $x(\bar{\tau}_1) \notin \mathcal{Q}$. Recall that $\mathcal{Q}^+(\bar{\tau}_1) \cap \mathcal{Q}^-(\bar{\tau}_1) = x(\bar{\tau}_1)$. Hence, from Theorem 1.5 we know that there exist a hyperplane $\mathcal{H} = \{x \in \mathbb{R}^\ell \mid h^\top x = \eta\}$ separating $\mathcal{Q}^+(\bar{\tau}_1)$ and $Q^-(\bar{\tau}_1)$, such that $x(\bar{\tau}_1) \in \mathcal{H}$. Given the assumption $\mathcal{Q}^+ \cap \mathcal{A} \subset \mathcal{Q}^+(\bar{\tau}_1)$ and $\mathcal{Q}^+ \cap \mathcal{B} \subset \mathcal{Q}^-(\bar{\tau}_1)$, we have that $\mathcal{H}$ must separate $\mathcal{Q}^+ \cap \mathcal{A}$ and $\mathcal{Q}^+ \cap \mathcal{B}$ as well. Hence, $\mathcal{H}$ must be parallel to $\mathcal{A}$ and $\mathcal{B}$, and $\beta \leq \eta \leq \alpha$. Now, if $\beta < \eta < \alpha$, then we obtain that $x(\bar{\tau}_1) \notin \mathcal{A} \cup \mathcal{B}$, which contradicts Lemma A.2. On the other hand, if $\eta = \alpha$ or $\eta = \beta$,

we obtain that $x(\bar{\tau}_1) \in \mathcal{Q}$, which is also a contradiction. This proofs the lemma. □

Lemma A.6. *In the first and fourth cases of Theorem 3.7 the cone $\mathcal{Q}(\bar{\tau}_2)$ has its vertex $x(\bar{\tau}_2)$ exclusively in either $\mathcal{A}$ or $\mathcal{B}$.*

Proof. Recall from Section 3.3.2 that the quadrics $\mathcal{Q}(\bar{\tau}_1)$ and $\mathcal{Q}(\bar{\tau}_1)$ in the family $\{\mathcal{Q}(\tau) \mid \tau \in \mathbb{R}\}$ of Theorem 3.7, are computed using the roots of the function

$$f(\tau) = \left((\alpha\beta - a^\top b)^2 - (1-\alpha^2)(1-\beta^2)\right)\tau^2 + 4(a^\top b - \alpha\beta)\tau + 4.$$

The roots of $f(\tau)$ are

$$\bar{\tau}_1 = 2\left(\frac{\alpha\beta - a^\top b - \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha\beta - a^\top b)^2 - (1-\alpha^2)(1-\beta^2)}\right) = \frac{2}{\alpha\beta - a^\top b + \sqrt{(1-\alpha^2)(1-\beta^2)}},$$
$$\bar{\tau}_2 = 2\left(\frac{\alpha\beta - a^\top b + \sqrt{(1-\alpha^2)(1-\beta^2)}}{(\alpha\beta - a^\top b)^2 - (1-\alpha^2)(1-\beta^2)}\right) = \frac{2}{\alpha\beta - a^\top b - \sqrt{(1-\alpha^2)(1-\beta^2)}},$$

where $\bar{\tau}_1 \leq \bar{\tau}_2$.

The vertex of the cone $\mathcal{Q}(\bar{\tau}_2)$ is $x(\bar{\tau}_2) = -P(\bar{\tau}_2)^{-1}p(\bar{\tau}_2)$. We can express $x(\bar{\tau}_2)$ in terms of a, b, α, and β as follows

$$\begin{aligned} x(\bar{\tau}_2) &= -P(\bar{\tau}_2)^{-1}p(\bar{\tau}_2) \\ &= -\left(I - \frac{(aa^\top + bb^\top)\tau_2^2 - (a^\top b\tau_2^2 + 2\tau_2)(ba^\top + ab^\top)}{(1-(a^\top b)^2)\tau_2^2 - 4a^\top b\tau_2 - 4}\right)\left(-\tau_2\frac{\beta a + \alpha b}{2}\right) \\ &= \frac{\tau_2\left(\left((\alpha - a^\top b\beta)\tau_2 - 2\beta\right)a + \left((\beta - a^\top b\alpha)\tau_2 - 2\alpha\right)b\right)}{(1-(a^\top b)^2)\tau_2^2 - 4a^\top b\tau_2 - 4}. \end{aligned}$$

Consider the inner products

$$a^\top x(\bar{\tau}_2) = \frac{\tau_2\left((1-(a^\top b)^2)\alpha\tau_2 - 2(a^\top b\alpha + \beta)\right)}{(1-(a^\top b)^2)\tau_2^2 - 4a^\top b\tau_2 - 4},$$

and

$$b^\top x(\bar{\tau}_2) = \frac{\tau_2\left((1-(a^\top b)^2)\beta\tau_2 - 2(a^\top b\beta+\alpha)\right)}{(1-(a^\top b)^2)\tau_2^2 - 4a^\top b\tau_2 - 4},$$

Next, we show that in the first and second cases of Theorem 3.7 the vertex $x(\bar{\tau}_2)$ cannot be in the set $\bar{\mathcal{A}}\cap\bar{\mathcal{B}}$. Assume to the contrary that $x(\bar{\tau}_2)\in\bar{\mathcal{A}}\cap\bar{\mathcal{B}}$. Let $\hat{\tau}_1\leq 0$ and $0\leq\hat{\tau}_2$ be the roots of $(1-(a^\top b)^2)\tau^2 - 4a^\top b\tau - 4$. Now, since we are analyzing the first and second cases of Theorem 3.7 we know that $\hat{\tau}_2<\bar{\tau}_1$, or $\bar{\tau}_2<\hat{\tau}_1$, or $\bar{\tau}_1<\hat{\tau}_1<\hat{\tau}_2<\bar{\tau}_2$. Hence, since $1-(a^\top b)^2\geq 0$ we have that $(1-(a^\top b)^2)\tau_2^2 - 4a^\top b\tau_2 - 4\geq 0$. Thus, if $a^\top x(\bar{\tau}_2)\leq\alpha$ and $b^\top x(\bar{\tau}_2)\geq\beta$, then

$$(a^\top b\alpha-\beta)\tau_2\leq -2\alpha \quad \text{and} \quad (a^\top b\beta-\alpha)\tau_2\geq -2\beta. \tag{A.1}$$

Substituting $\bar{\tau}_2$ in (A.1) we obtain that $\frac{\alpha}{\sqrt{1-\alpha^2}} = -\frac{\beta}{\sqrt{1-\beta^2}}$, which implies that $\alpha=-\beta$. This is possible if $\bar{\tau}_2=\hat{\tau}_2$, which is not in the cases being considered. Hence, in the first and second cases of Theorem 3.7 $x(\bar{\tau}_2)$ cannot be in the set $\bar{\mathcal{A}}\cap\bar{\mathcal{B}}$.

Similarly, we can show that in the first and second cases of Theorem 3.7 the vertex $x(\bar{\tau}_2)$ cannot be in the set $\mathcal{A}\cap\mathcal{B}$. Thus, if $a^\top x(\bar{\tau}_2)\geq\alpha$ and $b^\top x(\bar{\tau}_2)\leq\beta$, then

$$(a^\top b\alpha-\beta)\tau_2\geq -2\alpha \quad \text{and} \quad (a^\top b\beta-\alpha)\tau_2\leq -2\beta. \tag{A.2}$$

Substituting $\bar{\tau}_2$ in (A.2) we obtain that $\frac{\alpha}{\sqrt{1-\alpha^2}} = -\frac{\beta}{\sqrt{1-\beta^2}}$. This implies that $\bar{\tau}_2=\hat{\tau}_2$, which is not in the cases being considered. Hence, the vertex $x(\bar{\tau}_2)$ cannot be in the set $\mathcal{A}\cap\mathcal{B}$. □

Appendix B

Tables of computational results for Chapter 6

This appendix include the tables with the results of the experiments described in Chapter 6.

B.1 Experiments comparing branching rules

B.1.1 Pseudo-costs

Experiments using pseudo-costs branching rule and no cut manager.

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R12.C15.Cones5.Int10	# of Nodes	69	69	69
	CPU time (s)	0.08	0.09	0.08
R12.C15.Cones5.Int15	# of Nodes	27	27	27
	CPU time (s)	0.04	0.04	0.05

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R14.C18.Cones3.Int12	# of Nodes	439	487	437
	CPU time (s)	0.48	0.62	0.57
R14.C18.Cones3.Int15	# of Nodes	377	319	375
	CPU time (s)	0.42	0.44	0.48
R14c18.Cones3.Int18	# of Nodes	193	165	193
	CPU time (s)	0.22	0.24	0.24
R14.C18.Cones3.Int9	# of Nodes	3035	NaN	3039
	CPU time (s)	2.84	NaN	3.35
R17.C20.Cones5.Int15	# of Nodes	35	35	35
	CPU time (s)	0.06	0.05	0.04
R17.C20.Cones5.Int20	# of Nodes	35	35	35
	CPU time (s)	0.06	0.06	0.08
R17.C30.Cones3.Int12	# of Nodes	85	83	81
	CPU time (s)	0.13	0.16	0.20
R17.C30.Cones3.Int15	# of Nodes	845	845	1035
	CPU time (s)	1.21	1.42	1.69
R17.C30.Cones3.Int18	# of Nodes	544859	544641	895759
	CPU time (s)	6228.15	6921.47	20488.52
R17.C30.Cones3.Int21	# of Nodes	540405	540039	393405
	CPU time (s)	4736.22	5282.25	2110.55
R17.C30.Cones3.Int24	# of Nodes	530541	513407	1115158
	CPU time (s)	4500.23	4890.14	27000.65

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R17.C30.Cones3.Int27	# of Nodes	382885	411409	628865
	CPU time (s)	2558.90	3142.05	7639.67
R22.C30.Cones10.Int20	# of Nodes	3687	3687	3687
	CPU time (s)	4.18	4.23	4.56
R22.C40.Cones10.Int20	# of Nodes	77	75	75
	CPU time (s)	0.14	0.18	0.16
R22.C40.Cones10.Int30	# of Nodes	5165	5795	5165
	CPU time (s)	8.25	10.82	9.49
R22.C40.Cones10.Int40	# of Nodes	NaN	NaN	NaN
	CPU time (s)	NaN	NaN	NaN
R23.C45.Cones3.Int21	# of Nodes	358	346	370
	CPU time (s)	0.68	0.80	0.87
R23.C45.Cones3.Int24	# of Nodes	1121	1113	1115
	CPU time (s)	1.99	2.44	2.42
R23.C45.Cones3.Int27	# of Nodes	646406	661936	656103
	CPU time (s)	10837.79	13833.93	13681.17
R27.C50.Cones5.Int25	# of Nodes	959	949	1019
	CPU time (s)	1.85	2.10	2.31
R27.C50.Cones5.Int30	# of Nodes	239844	247740	238856
	CPU time (s)	1359.73	1867.03	1541.55
R27.C50.Cones5.Int35	# of Nodes	2226749	2227761	2186683
	CPU time (s)	67741.79	85598.07	84121.83

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R27.C50.Cones5.Int40	# of Nodes	2795427	2793409	3021927
	CPU time (s)	116732.87	129797.32	146040.20
R27.C50.Cones5.Int45	# of Nodes	2795427	2793091	3021925
	CPU time (s)	125739.27	119388.25	160034.31
R27.C50.Cones5.Int50	# of Nodes	2795427	NaN	3021913
	CPU time (s)	135873.60	NaN	145516.09
R32.C60.Cones15.Int30	# of Nodes	270	250	266
	CPU time (s)	0.52	0.61	0.54
R32.C60.Cones15.Int45	# of Nodes	217115	216787	214887
	CPU time (s)	893.78	936.39	1012.89
R52.C75.Cones5.Int60	# of Nodes	359195	418927	418865
	CPU time (s)	2140.95	3179.29	3253.12

B.1.2 Strong Branching

Experiments using pseudo-costs branching rule and no cut manager.

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R12.C15.Cones5.Int10	# of Nodes	33	33	33
	CPU time (s)	0.10	0.10	0.10
R12.C15.Cones5.Int15	# of Nodes	17	17	17
	CPU time (s)	0.04	0.05	0.05
R14.C18.Cones3.Int12	# of Nodes	139	139	139
	CPU time (s)	0.63	0.70	0.72
R14.C18.Cones3.Int15	# of Nodes	109	107	109
	CPU time (s)	0.70	0.81	0.77
R14.C18.Cones3.Int18	# of Nodes	95	95	95
	CPU time (s)	0.74	0.78	0.82
R14.C18.Cones3.Int9	# of Nodes	2217	2107	2247
	CPU time (s)	5.00	5.46	6.05
R17.C20.Cones5.Int15	# of Nodes	37	37	37
	CPU time (s)	0.10	0.10	0.12
R17.C20.Cones5.Int20	# of Nodes	35	35	35
	CPU time (s)	0.10	0.10	0.10
R17.C30.Cones3.Int12	# of Nodes	32	32	32
	CPU time (s)	0.46	0.50	0.54
R17.C30.Cones3.Int15	# of Nodes	536	504	516
	CPU time (s)	6.97	7.66	7.85

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R17.C30.Cones3.Int18	# of Nodes	225983	227637	226117
	CPU time (s)	2427.99	2652.17	2795.93
R17.C30.Cones3.Int21	# of Nodes	314773	314763	313707
	CPU time (s)	4707.02	5463.58	5439.03
R17.C30.Cones3.Int24	# of Nodes	144123	113837	143985
	CPU time (s)	1816.35	1764.01	2151.07
R17.C30.Cones3.Int27	# of Nodes	42109	42525	41515
	CPU time (s)	578.36	670.54	671.66
R22.C30.Cones10.Int20	# of Nodes	1657	1657	1657
	CPU time (s)	6.02	6.01	6.04
R22.C40.Cones10.Int20	# of Nodes	41	41	41
	CPU time (s)	0.33	0.43	0.38
R22.C40.Cones10.Int30	# of Nodes	1659	1659	1659
	CPU time (s)	20.24	22.40	22.90
R22.C40.Cones10.Int40	# of Nodes	1881	1881	1797
	CPU time (s)	19.87	23.15	22.09
R23.C45.Cones3.Int21	# of Nodes	229	237	226
	CPU time (s)	7.95	9.67	9.48
R23.C45.Cones3.Int24	# of Nodes	6154	5530	6042
	CPU time (s)	209.59	217.01	247.99
R23.C45.Cones3.Int27	# of Nodes	NaN	NaN	NaN
	CPU time (s)	NaN	NaN	NaN

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R27.C50.Cones5.Int25	# of Nodes	579	575	579
	CPU time (s)	17.04	18.67	19.29
R27.C50.Cones5.Int30	# of Nodes	918268	1069228	918266
	CPU time (s)	40554.40	61384.48	50489.82
R27.C50.Cones5.Int35	# of Nodes	NaN	NaN	NaN
	CPU time (s)	NaN	NaN	NaN
R27.C50.Cones5.Int40	# of Nodes	NaN	NaN	NaN
	CPU time (s)	NaN	NaN	NaN
R27.C50.Cones5.Int45	# of Nodes	4367595	NaN	NaN
	CPU time (s)	318614.20	NaN	NaN
R27.C50.Cones5.Int50	# of Nodes	3796593	3492829	NaN
	CPU time (s)	211121.87	205883.77	NaN
R32.C60.Cones15.Int30	# of Nodes	117	119	117
	CPU time (s)	2.73	2.84	2.81
R32.C60.Cones15.Int45	# of Nodes	89307	NaN	94583
	CPU time (s)	1445.38	NaN	1648.51
R52.C75.Cones5.Int60	# of Nodes	NaN	NaN	NaN
	CPU time (s)	NaN	NaN	NaN

B.2 Experiments using cut manager

In all experiments of this section we use the pseudo cost for branching rule an two different rules to select the disjunction.

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R12.C15.Cones5.Int10	# of Nodes	68	45	68
	CPU time (s)	0.08	0.08	0.08
R12.C15.Cones5.Int15	# of Nodes	27	17	27
	CPU time (s)	0.06	0.05	0.05
R14.C18.Cones3.Int12	# of Nodes	427	365	424
	CPU time (s)	0.96	3.20	0.57
R14.C18.Cones3.Int15	# of Nodes	365	307	363
	CPU time (s)	0.42	0.44	0.48
R14.C18.Cones3.Int18	# of Nodes	186	163	186
	CPU time (s)	0.22	0.24	0.24
R14.C18.Cones3.Int9	# of Nodes	3031	2868	3035
	CPU time (s)	7.48	12.21	3.35
R17.C20.Cones5.Int15	# of Nodes	34	34	34
	CPU time (s)	0.06	0.05	0.04
R17.C20.Cones5.Int20	# of Nodes	34	34	34
	CPU time (s)	0.06	0.06	0.08
R17.C30.Cones3.Int12	# of Nodes	58	44	51
	CPU time (s)	0.13	0.16	0.20

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R17.C30.Cones3.Int15	# of Nodes	512	437	624
	CPU time (s)	2.54	5.73	1.69
R17.C30.Cones3.Int18	# of Nodes	380651	43014	631544
	CPU time (s)	6228.15	610.74	20488.52
R17.C30.Cones3.Int21	# of Nodes	431459	148625	365051
	CPU time (s)	4736.22	3002.25	2110.55
R17.C30.Cones3.Int24	# of Nodes	447048	140184	841094
	CPU time (s)	4500.23	1762.77	27000.65
R17.C30.Cones3.Int27	# of Nodes	293040	313762	524041
	CPU time (s)	2558.90	3142.05	7639.67
R22.C30.Cones10.Int20	# of Nodes	3556	3556	3556
	CPU time (s)	4.18	4.23	4.56
R22.C40.Cones10.Int20	# of Nodes	58	57	56
	CPU time (s)	0.14	0.18	0.16
R22.C40.Cones10.Int30	# of Nodes	4713	5336	4713
	CPU time (s)	8.25	10.82	9.49
R22.C40.Cones10.Int40	# of Nodes	22600	32874	NaN
	CPU time (s)	91.06	331.39	NaN
R23.C45.Cones3.Int21	# of Nodes	184	178	190
	CPU time (s)	0.68	0.80	0.87
R23.C45.Cones3.Int24	# of Nodes	770	764	776
	CPU time (s)	1.99	2.44	2.42

		Selection of Disjunctive Conic Cut		
Rows.Cols.Cones.IntV		No cuts added	Max Inf	Pseudo Cost
R23.C45.Cones3.Int27	# of Nodes	393265	402801	399196
	CPU time (s)	10837.79	13833.93	13681.17
R27.C50.Cones5.Int25	# of Nodes	494	488	521
	CPU time (s)	1.85	2.10	2.31
R27.C50.Cones5.Int30	# of Nodes	153067	157791	153654
	CPU time (s)	1359.73	1867.03	1541.55
R27.C50.Cones5.Int35	# of Nodes	1910190	1296983	1881604
	CPU time (s)	67741.79	65263.15	84121.83
R27.C50.Cones5.Int40	# of Nodes	2400487	1277814	2607452
	CPU time (s)	116732.87	62859.13	146040.20
R27.C50.Cones5.Int45	# of Nodes	2400487	2442754	2607451
	CPU time (s)	125739.27	119388.25	160034.31
R27.C50.Cones5.Int50	# of Nodes	2400487	1462856	2607458
	CPU time (s)	135873.60	81770.02	145516.09
R32.C60.Cones15.Int30	# of Nodes	213	182	209
	CPU time (s)	0.52	0.61	0.54
R32.C60.Cones15.Int45	# of Nodes	184948	184642	183230
	CPU time (s)	893.78	936.39	1012.89
R52.C75.Cones5.Int60	# of Nodes	328676	381938	381858
	CPU time (s)	2140.95	3179.29	3253.12

Index

Biography

Name:	Julio César Góez Gutiérrez.
Place of birth:	Medellín, Colombia.
Date of birth:	June 3, 1976.
Parents:	Mrs. Magdalena de Jesús Gutiérrez López and Mr. Roger Góez Gúzman.

Education

- Ph.D. in Industrial Engineering, Lehigh University, Bethlehem, PA, August2007 – September 2013.
- M.Sc. in Industrial Engineering, University of Los Andes, Bogotá, Colombia, August 2002 – September 2004.
- B.S. in Industrial Engineering, University of Antioquia, Medellín, Colombia, January 2005 – June 2002.

Professional experience

- System administrator, CORL@ Lab, Industrial and Systems Engineering Department, Lehigh University (August 2008 – July 2013).

- Summer Intern, Mathematics and Computer Science Division, Argonne National Laboratory (summer 2009).
- Summer Intern, Mathematical Science and Business Analytics Department, Watson Research Center, IBM Research (Summer 2008).
- Teaching Assistant, Department of Industrial and Systems Engineering, Lehigh University (August 2007 – May 2008).
- Instructor, Industrial Engineering Department, University of Los Andes (August 2004 – July 2007).

Publications

- Belotti, P., Góez, J.C., Pólik, I., Ralphs, T., Terlaky, T., On Families of Quadratic Surfaces Having Fixed Intersections with two Hyperplanes. Accepted at *Discrete Applied Mathematics*, 2013.
- Belotti, P., Góez, J.C., Pólik, I., Ralphs, T., Terlaky, T., A Conic Representation of the Convex Hull of Disjunctive Sets and Conic Cuts for Integer Second Order Cone Optimization. Under revision after first round review for *Mathematical Programming*, 2012.
- Góez, J.C., Luedtke, J., Rajan, D., Kalagnanam, J., Stochastic Unit Commitment Problem, *IBM research report RC24713*, December 23, 2008, New York, United States.
- Góez, J.C., Riaño, G., jMarkov: An Object Oriented Framework for Modeling and Analyzing Markov Chains and QBDs, *ACM International Conference Proceeding Series*, Vol. 201. Proceeding from the 2006 workshop on tools for solving structured Markov chains, October 10, 2006, Pisa, Italy.

Awards

- Ph.D. Student of the year, Department of Industrial and Systems Engineering, Lehigh University, 2012.
- Ph.D. Student of the year, Department of Industrial and Systems Engineering, Lehigh University, 2010.
- Rossin Doctoral Fellow, Faculty of Engineering, Lehigh University, 2009.

Mememberships and Professional Activities

- Institute for Operations Research and Management Sciences (INFORMS).
- INFORMS Computing Society.
- INFORMS Optimization Society.
- INFORMS students chapter, Lehigh University, 2007 – 2013.
- Society for Industrial and Applied Mathematics (SIAM).
- Sigma Xi, The Scientific Research Society.