Cutting Planes for Convex Objective Nonconvex Optimization

Alexander Michalka

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Cutting Planes for Convex Objective Nonconvex Optimization

Alexander Michalka

This thesis studies methods for tightening relaxations of optimization problems with convex objective values over a nonconvex domain. A class of linear inequalities obtained by lifting easily obtained valid inequalities is introduced, and it is shown that this class of inequalities is sufficient to describe the epigraph of a convex and differentiable function over a general domain. In the special case where the objective is a positive definite quadratic function, polynomial time separation procedures using the new class of lifted inequalities are developed for the cases when the domain is the complement of the interior of a polyhedron, a union of polyhedra, or the complement of the interior of an ellipsoid. Extensions for positive semidefinite and indefinite quadratic objectives are also studied. Applications and computational considerations are discussed, and the results from a series of numerical experiments are presented.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

This thesis studies methods for generating linear inequalities to strengthen convex relaxations of nonconvex optimization problems. We focus on problems with a convex function f in the objective, or constraints on a convex function f, over a nonconvex domain $\mathcal{F}$. We are interested specifically in problems whose simplest convex relaxations are very weak. Problems with this structure arise in the fields of signal processing ([47], [35]), semiconductor lithography ([1], [58]), and mixed-integer nonlinear programming ([20], [15]), among others. Current solution methods rely on heuristics or branching to obtain good feasible solutions, and achieving useful lower bounds may come at the cost of substantial amounts of branching or an "extended" formulation which introduces a large number of extra variables. With this motivation, we seek efficient ways to strengthen convex relaxations and improve bounds on problems of this type.

Our approach focuses on the use of two main ideas: *lifting* and *cutting planes*. These two methodologies share a similar history: both were originally conceived in the context of linear integer programming (with some of the earliest work in both cases being done by Gomory - see [37], [38], and [39]). More recently, both have been applied in continuous and nonlinear optimization. Later in this chapter, we provide an introduction to the concepts of lifting inequalities and cutting planes from their early development, briefly describe some later extensions, and highlight examples of work which is similar to our own.

We devote significant attention to the case of a quadratic objective function. This is due both to the prevalence of quadratics in applications (all of the applications listed above contain a quadratic objective), and also to their tractability: the geometry of quadratically defined sets makes them especially amenable to the analysis required in our constructions of strong valid lifted inequalities.

The remainder of this chapter provides examples of particular problem classes motivating this work, and reviews previous research in the fields of lifting and cutting plane methods. Chapter 2 introduces our lifting constructions and provides general results for convex and differentiable objective functions. Chapters 3 and 4 focus on the special case where the objective is a quadratic function, for different choices of domain. Chapter 5 presents results from numerical experiments in which we apply the techniques we develop to randomly-generated problem instances and compare their performance with similar existing methods. Finally, Chapter 6 provides some final discussion and conclusion.

1.1 Notation

The notation we use is standard, but for completeness we provide a quick review before we begin in earnest.

For a set $S \subseteq \mathbb{R}^n$, $\operatorname{int}(S)$ denotes the interior of S, and $\operatorname{cl}(S)$ denotes the closure. ∂S denotes the boundary of S, which is defined as the set of all points $x \in \mathbb{R}^n$ where, for every $\epsilon > 0$, the ball of radius ϵ centered at x contains both points S and points not in S. $\operatorname{conv}(S)$ represents the convex hull of the set S. For $\mu \in \mathbb{R}^n$ and $r \geq 0$, we use $\mathcal{B}(\mu, r)$ to denote the closed ball of radius r centered at μ. For any two sets X and Y, $X \setminus Y$ denotes the difference between X and Y: the set of all elements of X that are *not* also elements of Y.

For two vectors $x, y \in \mathbb{R}^n$, $x \geq y$ means that $x_i \geq y_i$ for each $i \in \{1, \ldots, n\}$. $x > y$ means that each inequality is strict. $x \not< y$ indicates that there is at least one index i for which $x_i \geq y_i$. If $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^p$ are two vectors in different spaces, (x, y) denotes the $(n+p)$-dimensional vector obtained by concatenating x and y. $\operatorname{diag}(x)$ represents the $(n \times n)$ diagonal matrix whose diagonal entries are the entries of x.

When x and y are elements in a general vector space and $\mathcal{K}$ is a cone, $x \succeq_{\mathcal{K}} y$ means $(x - y) \in \mathcal{K}$. For matrices, when the $\succeq$ symbol appears alone, the cone $\mathcal{K}$ is assumed to be the set of positive semidefinite matrices of the proper dimension: $X \succeq 0$ means X is positive semidefinite, and $X \succ 0$ means X is positive definite. $\operatorname{tr}(X)$ denotes the trace of the square matrix X, and I represents the identity matrix, whose dimension should be evident from the context.

The term "argmin" denotes the set of optimal solutions to an optimization problem. For example, $\operatorname{argmin}\{f(x) \mid x \in \mathcal{F}\}$ is the set of optimal solutions to the problem

$$\begin{aligned} \text{minimize:} \quad & f(x) \\ \text{subject to:} \quad & x \in \mathcal{F}. \end{aligned}$$

1.2 Nonconvex Optimization

Convex optimization studies the minimization of convex objective functions over convex domains. Such problems have the convenient feature that any local minimum is a global minimum. Convex optimization is well-studied and, under reasonable assumptions, interior-point algorithms can solve convex optimization problems in polynomial time (see [51]). Many practical problems from a variety of applications can be formulated as convex optimization problems (see [19]).

Nonconvex, or *global* optimization, studies the minimization of general functions over general domains, which may be nonconvex and can include integrality constraints on some or all of the problem variables. Solution methods typically rely on branching - partitioning the feasible region into smaller components and solving subproblems over these components - and convex approximations of nonconvex functions (see [66] and [13] for more on global optimization techniques). Heuristics, including randomized methods such as simulated annealing, can be used to find good feasible solutions, but do not provide lower bounds on the objective value and therefore do not prove optimality.

Our work was originally motivated by optimization problems with a convex objective function but nonconvex domain. These are problems of the form

$$\begin{aligned} \text{minimize:} \quad & f(x) \\ \text{subject to:} \quad & x \in \mathcal{F} \subseteq \mathbb{R}^n \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. In particular, our goal is to investigate ways to provide strong lower bounds on problems on this form whose simplest convex relaxation

$$\begin{aligned} \text{minimize:} \quad & f(x) \\ \text{subject to:} \quad & x \in \operatorname{conv}(\mathcal{F}) \end{aligned}$$

gives only a very weak bound on the optimal value of the original problem. This is the case, for instance, when $\mathcal{F} = \mathbb{R}^n \setminus P$, where P is a bounded set. In this case, $\operatorname{conv}(\mathcal{F}) = \mathbb{R}^n$ and the convex relaxation is trivial. In the sections that follow, we introduce several classes of problems fitting into this mold, and describe some of the previous work concerning solution methods and strong relaxations to these problems.

1.2.1 Quadratically Constrained Quadratic Programming

The formulation of the general Quadratically Constrained Quadratic Program (QCQP) is

$$\begin{aligned} \text{minimize:} \quad & x^T Q x + q^T x \\ \text{subject to:} \quad & x^T A_i x - 2b_i^T x + c_i \leq 0 \quad i = 1, \ldots, m \end{aligned} \tag{1.1}$$

where Q and each A_i are symmetric $(n \times n)$ matrices. When Q and all of the A_i are positive semidefinite, the problem is convex - it is a special case of Second Order Cone Programming [2] - but the general QCQP problem is NP-hard.

Nonconvex QCQPs of the form

$$\begin{aligned} \text{minimize:} \quad & x^T x \\ \text{subject to:} \quad & x^T A_i x \geq 1 \quad i = 1, \ldots, m \end{aligned} \tag{1.2}$$

where each A_i is positive definite arise in semiconductor lithography and signal processing (see [47], [1], [58]). Figure 1.1 shows an instance of such a problem with three quadratic constraints. Geometrically, the goal of these problems is to find the vector of minimum (squared) Euclidean norm which is *not* contained in the interior of any of the sets

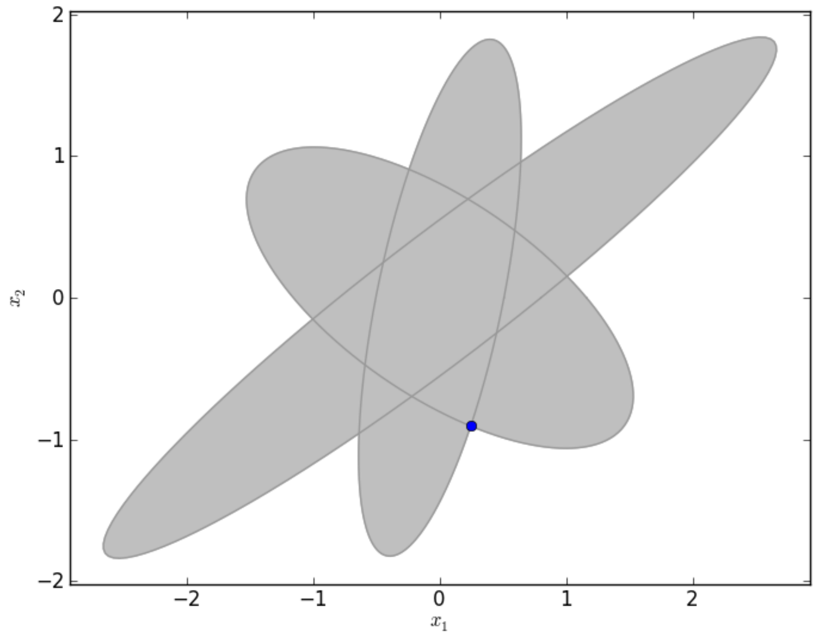
$$E_i = \{x \in \mathbb{R}^n \mid x^T A_i x \leq 1\},$$

each of which is a full-dimensional and bounded ellipsoid centered at the origin. In these problems, the convex hull of the feasible region is $\mathbb{R}^n$, and a naive convex relaxation would give a trivial lower bound of 0.

A commonly used approach (see [68], [47], or [34]) for computing lower bounds for nonconvex QCQPs is a semidefinite relaxation. As we have $x^T C x = \operatorname{tr}(C x x^T)$ for any symmetric matrix C, we can reformulate (1.1) as the following:

$$\begin{aligned} \text{minimize:} \quad & \operatorname{tr}(QX) + q^T x \\ \text{subject to:} \quad & \operatorname{tr}(A_i X) - 2b_i^T x + c_i \leq 0 \quad i = 1, \ldots, m \\ & X = x x^T. \end{aligned} \tag{1.3}$$

Figure 1.1: An instance of a nonconvex QCQP. The infeasible region is shaded in grey and the point closest to the origin is shown in blue.



This formulation is completely equivalent to (1.1). The objective and the constraints $\operatorname{tr}(A_i X) - 2b_i^T x + c_i \leq 0$ are all linear, but the new formulation includes the nonconvex constraint $X = xx^T$. In semidefinite relaxations, this constraint is relaxed to $X \succeq xx^T$, which yields a semidefinite program (SDP). This relaxation is known to be exact when $m = 1$, and is in fact the Lagrange bidual (the dual of the dual) of the original problem, for any n. Semidefinite relaxations have been observed empirically to give strong lower bounds (see [4]), at the cost of increasing the number of variables in the problem significantly and requiring the solution of a semidefinite program. Moreover, the solution of the SDP typically does not immediately yield a feasible solution to the original problem - see [47] for discussion of methods for obtaining feasible solutions from an SDP relaxation. Semidefinite relaxations have been applied to combinatorial optimization problems (see [36]), as the constraint $x_i \in \{0, 1\}$ can be written as the quadratic constraint $x_i^2 = x_i$.

Another relaxation method is the *Reformulation-Linearization Technique* (RLT) (see [61] and [62] for development of the RLT). This relaxation also introduces an $(n \times n)$ matrix X which, ideally, is equal to xx^T. Consider an instance of the problem (1.2) which includes "box" constraints bounding the values of the components of x:

$$\ell_i \leq x_i \leq u_i \quad \text{for all } i, \tag{1.4}$$

where the bounds ℓ_i and u_i are all finite. Constraints relating X and x are formed by multiplying pairs of constraints from (1.4) and then *linearizing* by replacing each bilinear term $x_i x_j$ with the linear term X_{ij}. As an example, take the constraints $x_i - \ell_i \geq 0$ and $u_j - x_j \geq 0$ for two arbitrary choices of indices i and j. Certainly the constraint $(x_i - \ell_i)(u_j - x_j) \geq 0$ is valid; expanding out the product on the left hand side gives the nonconvex constraint

$$u_j x_i - x_i x_j - \ell_i u_j + \ell_i x_j \geq 0.$$

The linearization step relaxes this to the convex constraint

$$X_{ij} - u_j x_i - \ell_i x_j \leq -\ell_i u_j.$$

The final reformulation, which explicitly includes the box constraints, is the following *linear* program:

$$\begin{array}{rl}
\text{minimize:} & \operatorname{tr}(QX) + q^T x \\
\text{subject to:} & \operatorname{tr}(A_k X) - 2b_k^T x + c_k \leq 0 \quad k = 1, \ldots, m \\
& \ell_i \leq x_i \leq u_i \quad \forall\, i \\
& X_{ij} - \ell_j x_i - \ell_i x_j \geq -\ell_i \ell_j \quad \forall\, i, j \\
& X_{ij} - u_i x_j - u_j x_i \geq -u_i u_j \quad \forall\, i, j \\
& X_{ij} - u_j x_i - \ell_i x_j \leq -\ell_i u_j \quad \forall\, i, j \\
& X = X^T
\end{array} \tag{1.5}$$

The bounds obtained from the RLT are typically not as strong as those obtained from the SDP relaxation, and the RLT is not applicable when the entries of x are unbounded. Anstreicher [4] gives a computational comparison of the SDP and RLT relaxations, and studies the effect of combining the two methods.

In the case where $\ell_i = 0$ and $u_i = 1$ for each i, we can take advantage of the following result:

Proposition 1.2.1. *Let Q be any $(n \times n)$ symmetric matrix and $q \in \mathbb{R}^n$. Let $\alpha \in \mathbb{R}^n_+$ be fixed. Then*

$$x^T (Q + \operatorname{diag}(\alpha)) x + (q - \alpha)^T x \quad \leq \quad x^T Q x + q^T x \quad \textit{for all} \quad x \in [0,1]^n.$$

The function $x^T(Q + \operatorname{diag}(\alpha))x + (q - \alpha)^T x$ is called an *αBB-underestimator* of the function $x^T Qx + q^T x$. Consider an instance of (1.1) which includes the constraint $x_i \in [0, 1]$ for each i. If nonnegative vectors $\alpha_0, \alpha_1, \ldots, \alpha_m$ are chosen so that $Q + \operatorname{diag}(\alpha_0) \succeq 0$ and $A_i + \operatorname{diag}(\alpha_i) \succeq 0$ for each i, then the following QCQP is a convex relaxation:

$$\begin{aligned} \text{minimize:} \quad & x^T (Q + \operatorname{diag}(\alpha_0))\, x + (q - \alpha_0)^T x \\ \text{subject to:} \quad & x^T (A_i + \operatorname{diag}(\alpha_i))\, x - (2b_i + \alpha_i)^T x + c_i \leq 0 \quad i = 1, \ldots, m \\ & x_j \in [0, 1] \quad j = 1, \ldots, n \end{aligned} \tag{1.6}$$

The idea of the *αBB-underestimator* was first developed by Androulakis et al. in [3] for general nonlinear functions. Anstreicher [5] shows that the relaxation (1.6) is weaker than one obtained by adding the constraint $\operatorname{diag}(X) \leq x$ to a semidefinite relaxation. The same paper also provides a similar result for generalizations of these methods when the problem contains linear constraints $x \geq 0,\ Ax \leq b$ rather than the box constraint $x \in [0, 1]^n$.

1.2.1.1 The Trust Region Subproblem

A frequently encountered example of a nonconvex QCQP is the *trust region subproblem*:

$$\begin{aligned} \text{minimize:} \quad & x^T Qx + q^T x \\ \text{subject to:} \quad & x^T x \leq 1 \end{aligned} \tag{1.7}$$

This problem arises in nonlinear programming algorithms. The objective is a quadratic approximation of some nonlinear objective function f about a current solution point. The set $\{x \mid x^T x \leq 1\}$ is the *trust region*. This is the set in which the quadratic approximation is "trusted" to be a suitably close approximation of f. Because there is only a single constraint in the trust-region problem, it can be solved exactly using the semidefinite relaxation (see [19]):

$$\begin{aligned} \text{minimize:} \quad & \operatorname{tr}(QX) + q^T x \\ \text{subject to:} \quad & \operatorname{tr}(X) \leq 1 \\ & X \succeq xx^T \end{aligned} \tag{1.8}$$

The problem becomes more difficult with the addition of extra constraints. The authors of [65] and [21] provide an exact semidefinite formulation in the presence of a single linear constraint, with [21] providing a similar fomulation for the case with two parallel linear constraints. Burer and Yang [24] extend these results to derive a semidefinite formulation when a system of in equality constraints $Ax \leq b$ is present, provided that the constraints do not intersect inside the unit ball: that is, there is no point $y \in \mathbb{R}^n$ with $y^T y \leq 1$, $a_j^T y = b_j$, and $a_i^T y = b_i$ for any distinct pair of indices i, j. These formulations are obtained in a manner very similar to the RLT: by multiplying together pairs of constraints $(b_i - a_i^T x) \geq 0$ and $(b_j - a_j^T x) \geq 0$ and linearizing the resulting valid inequalities, and by linearizing the valid inequality $\left|\left|(b_j - a_j^T x)x\right|\right| \leq (b_j - a_j^T x)$.

The addition of a second quadratic constraint (obtained as a sufficient decrease condition) is considered in [25], and [70] provides an algorithm for the solution of this extension.

1.2.1.2 The S-Lemma

The S-Lemma (or S-Procedure) was first described in [67] and provides a means for describing when one quadratic constraint implies another, regardless of the convexity or nonconvexity of either. It is, in a sense, a quadratic analog to Farkas's Lemma. It is often stated in different ways; here we present one.

Lemma 1.2.2 (The S-Lemma)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be two quadratic functions. Assume that there is some $\hat{x}$ with $g(\hat{x}) < 0$. Then*

$$f(x) \geq 0 \textit{ for all } x \textit{ with } g(x) \leq 0$$

if and only if there exists a nonnegative multiplier τ where

$$f(x) + \tau g(x) \geq 0 \textit{ for all } x \in \mathbb{R}^n$$

The original proof of the S-Lemma relies on the result of Dines [32], stating that if f and g are *homogenous* quadratic functions, the joint range $\{(f(x),\ g(x)) \mid x \in \mathbb{R}^n\} \subseteq \mathbb{R}^2$ is convex. The survey [54] provides a thorough background on the S-Lemma, including some extensions and variations, as well as counterexamples in more complicated settings.

1.2.2 Disjunctive Programming

The authors of [58] and [1] study a relaxation of the nonconvex QCQP (1.2), in which the nonconvex quadratic constraints are replaced with nonconvex polyhedral constraints. Letting $A_i = V_i^T V_i$ be the eigenvalue decomposition of each matrix A_i, their relaxation is the following:

$$\begin{aligned} \text{minimize:} \quad & x^T x \\ \text{subject to:} \quad & ||V_i x||_1 \geq 1 \quad i = 1, \ldots, m \end{aligned} \tag{1.9}$$

In this formulation, the ellipsoids E_i are replaced with the "diamond" shaped polyhedra

$$D_i = \{x \mid \ ||V_i x||_1 \leq 1\}.$$

The authors exploit the structure of these constraints to devise a parallelized algorithm which searches the sphere enclosing all of the D_i and solves a series of convex quadratic programs over pieces of the feasible region. They are able to obtain good solutions very quickly using this method, but as the resolution of their search is finite and the only lower bounds obtained are quite weak, they are not able to guarantee global optimality.

This problem can be seen as a special case of a class of problems with a convex objective and a feasible region which is the complement of the interior of a union of polyhedra. The general form of these problems is

$$\begin{aligned} \text{minimize:} \quad & x^T x \\ \text{subject to:} \quad & A^k x \not< b^k \quad k = 1, \ldots, K \end{aligned} \tag{1.10}$$

where $A^k \in \mathbb{R}^{m_k \times n}$ and $b^k \in \mathbb{R}^{m_k}$ for each k. These problems can be formulated and solved as Mixed Integer Quadratic Programs (MIQP), as follows:

$$\begin{aligned} \text{minimize:} \quad & x^T x \\ \text{subject to:} \quad & a_j^{kT} x \geq b_j^k - M(1 - B_j^k) \quad k = 1, \ldots, K, \ j = 1, \ldots, m_k \\ & \textstyle\sum_{j=1}^{m_k} B_j^k = 1 \quad k = 1, \ldots, K \\ & B_j^k \in \{0, 1\} \quad k = 1, \ldots, K, \ j = 1, \ldots, m_k \end{aligned} \tag{1.11}$$

The parameter M is a number large enough so that the problem is feasible and all constraints can be satisfied at the optimal solution. Relaxing the constraints $B_j^k \in \{0, 1\}$ to $B_j^k \in [0, 1]$ gives a convex formulation, but this relaxation tends to result in very weak lower bounds on the true optimal value of the MIQP.

All of these problems are examples of *disjunctive programs*: optimization problems where the feasible region is defined by a disjunction, or union, of individual components. Disjunctive programming was introduced by Balas [8] in linear integer programming. Later work by Ceria and

Soares [26] generalized disjunctive programming to the nonlinear setting, using the perspective transformation (see [19]) to obtain a description of the convex hull of a disjunctive set defined by convex functions, and a means for optimizing over this set. Later, we will consider the formulation of Ceria and Soares as an option for generating valid inequalities for the feasible regions of problems we consider.

In the examples above, the components are polyhedra. In Sections 3.1 and 3.2, we will discuss the use of lifting techniques to obtain lower bounds on problems of the form (1.11).

1.2.3 Cardinality Constraints

A *cardinality constraint* on a vector $x \in \mathbb{R}^n$ is a constraint on the number of nonzero entries a feasible vector x is allowed to have. Such a constraint is represented as $||x||_0 \leq K$, where $||\cdot||_0$ is the "zero-norm" (although it is not actually a norm). The set of vectors $x \in \mathbb{R}^n$ with $||x||_0 \leq K$ for any positive K forms a basis for $\mathbb{R}^n$, so the convex relaxation of any feasible set with a cardinality constraint is typically weak. Bienstock [15] studies cardinality-constrained problems in portfolio optimization, where the constraint limits the number of stocks the investor purchases. Later we will point out a way in which the techniques we develop can be applied to improve lower bounds in these types of problems.

1.3 Lifting

Much of the work in this thesis concerns or utilizes *lifted* inequalities. Generally speaking, lifting is a method of modifying an inequality $\pi^T x \leq \pi_0$ which is valid for a given set $\mathcal{S}$ to produce a new valid inequality $(\pi + \alpha)^T x \leq (\pi_0 + \alpha_0)$. Ideally, the new inequality will be "strong", in that it is not implied by some another valid inequality, or a combination of valid inequalities. Exactly how this is accomplished depends on the structure of the set $\mathcal{S}$: as Nemhauser and Wolsey remark before introducing lifting in [50], "The determination of families of strong valid inequalities is more of an art than a formal methodology." In this section we provide a brief overview of the history of lifting and some of the classes of problems to which lifting has been applied successfully.

1.3.1 Lifting in Integer Programming

Lifting has been studied since at least the late 1960s as a means for constructing valid inequalities in 0/1 linear integer programs. In early work, lifting is performed sequentially: one variable at a time. The initial valid inequality, which we will call the *base* inequality, is

$$\sum_{j\in J} \pi_j x_j \leq \pi_0, \tag{1.12}$$

where $J \subset \{1, \dots n\}$. The *lifted* inequality gives a positive coefficient to a variable x_ℓ where $\ell \notin J$:

$$\sum_{j\in J}^{n} \pi_j x_j + \alpha_\ell x_\ell \leq \pi_0. \tag{1.13}$$

The new coefficient α_ℓ is called the *lifting coefficient.* As the variables x_i must take values in $\{0, 1\}$, any positive value for α_ℓ results in an an inequality that is stronger than the original, with the strongest lifted inequality being the one with the largest value for α_ℓ (subject to validity for the set $\mathcal{S}$). The value of the largest possible lifting coefficient is computed by solving the *lifting problem*:

$$\alpha_\ell = \left\{ \begin{array}{ll} \text{maximize:} & \alpha \\ \text{subject to:} & \sum_{j\in J}^{n} \pi_j x_j + \alpha \leq \pi_0 \quad \forall x \in \mathcal{S} \end{array} \right\}$$

The constraint in this problem ensures that the lifted inequality is valid for $\mathcal{S}$. In this constraint, the value of x_ℓ is fixed at one, as the validity of (1.12) implies (1.13) holds when $x_\ell = 0$. Once the lifting coefficient has been computed, the lifted inequality can be used as the next base inequality, and the lifting coefficient for a new variable x_k (where $k \notin J \cup \{\ell\}$) can be computed. In general, the values for the lifting coefficients will depend on the order in which the new variables are chosen. To compute all valid lifted inequalities, it is necessary to consider all possible orderings of the indices not in the original index set J.

The formulation and solution of the lifting problem is problem specific; early work relied on the combinatorial structure of the feasible set. Padberg [52] provides a method for using sequential lifting to obtain facet-defining inequalities in set-packing problems. Nemhauser and Wolsey provide derivations of lifted inequalities for the set-packing problem as well as the 0/1 knapsack problem in [50]. Wolsey [69] studies lifting for more general integer linear programs, and Zemel [71] introduces procedures for lifting multiple variables simultaneously in general 0/1 programs.

1.3.2 Lifting in Continuous Optimization

Since its origination in integer programming, lifting has been applied in continuous (including nonlinear) problems. Atamtürk and Narayanan [7] develop a generalization where a valid conic inequality

$$C_0 x^0 \succeq_{\mathcal{K}} d$$

defined over a subset of variables is lifted to a new conic inequality

$$\sum_{j=0}^{p} C_j x^j \succeq_{\mathcal{K}} d$$

over the complete set of variables. Here, $\mathcal{K}$ is a closed, convex, and full-dimensional cone with at least one extreme point, the variables x^j are vectors in $\mathbb{R}^{n_j}$, and the lifting "coefficients" C_j are $(m \times n_j)$ matrices. The lifting approach presented in the previous section is a special case of this approach in the case where $n_j = 1$ for each j and the cone $\mathcal{K}$ is $\mathbb{R}_+$,

Richard and Tawarmalani [56] present a method for lifting linear inequalities approximating a nonlinear function f over a subspace to linear inequalities valid over the entire space, and discuss applications to nonlinear mixed-integer knapsack problems. The lifting techniques we develop are similar in that they use lifting to produce tight linear approximations to nonlinear functions. As the functions f we consider are convex, they meet the criterion imposed by these authors of having an affine minorant and their results apply. Our constructions, however, begin with an inequality which is globally valid and produce a lifted inequality which is valid only over a (typically nonconvex) subset of interest.

Belotti et al. [14] use lifted inequalities to obtain the convex envelope of sets of the form $M = \{x \in \mathbb{R}^3 \mid x_3 = x_1 x_2,\ \ell_i \leq x_i \leq u_i, i = 1, 2, 3\}$. They use for the base inequality the *tangent* inequality

$$y_2(x_1 - y_1) + y_1(x_2 - y_2) \geq 0$$

at a point $y \in \mathbb{R}^2$ with $y_1 y_2 = \ell_3$ and derive a lifted inequality

$$y_2(x_1 - y_1) + y_1(x_2 - y_2) - \lambda(x_3 - \ell_3) \geq 0$$

where $\lambda \in \mathbb{R}_+$ is the lifting coefficient. The lifting procedures we develop will also use the tangent inequality for a convex function f as the base inequality.

1.4 Cutting Planes

Cutting plane methods are a means for solving optimization problems through a series of relaxations whose feasible sets are progressively tightened through the addition of valid inequalities. The basis of any cutting plane algorithm is the following result from convex analysis (see [57]):

Proposition 1.4.1. *Any closed convex set $\mathcal{F}$ has a linear inequality description:*

$$\mathcal{F} = \bigcap_{j \in J} \{x \mid \gamma_j^T x \leq \beta_j\}.$$

In this expression, J is an index set which may be infinite. The generic cutting plane algorithm for solving an optimization problem $\min\{f^T x \mid x \in \mathcal{F}\}$ (the objective is assumed to be linear without loss of generality) over a closed convex set $\mathcal{F}$ is as follows:

Algorithm 1.4.2 (Generic Cutting Plane Algorithm)**.**

1. Let $\mathcal{F}_0$ be a closed convex set containing $\mathcal{F}$, and set $i = 0$.
2. Let $y^i \in \text{argmin}\{f^T x \mid x \in \mathcal{F}_i\}$.
3. If y^i meets the stopping criteria, stop. Otherwise, find an inequality $\pi^T x \leq \pi_0$ valid for $\mathcal{F}$, with $\pi^T y^i > \pi_0$.
4. Let $\mathcal{F}_{i+1} = \mathcal{F}_i \cap \{x \mid \pi^T x \leq \pi_0\}$. Set $i = i + 1$ and go to Step 2.

In practice, the "stopping criteria" in Step 3 could be the condition $y^i \in \mathcal{F}$ or something less strict - for instance if y^i is provably "close" to $\mathcal{F}$ or if $f^T y^i$, the objective value of the current relaxation, is close to a known upper bound for the original problem. As each point y^i is the solution to a relaxation, each objective value $f^T y^i$ gives a valid lower bound on the value of the original problem.

The key step of the cutting plane algorithm is in Step 3 - finding the linear inequality $\pi^T x \leq \pi_0$ which separates y^i from $\mathcal{F}$. This is known as the *separation problem.* The closedness and convexity of $\mathcal{F}$ guarantees that if $y^i \notin \mathcal{F}$, such a separating inequality exists, but the method of finding π and π_0 will be specific to structure of the set $\mathcal{F}$. In cases where the separation problem can be solved efficiently, cutting plane algorithms provide an attractive means for solving (or providing bounds for) problems where $\mathcal{F}$ does not have a convenient closed-form expression - for example when the linear inequality description requires an infinite number of inequalities.

1.4.1 Cutting Planes in Linear Integer Programming

As with lifting, the concept of cutting planes arose in linear integer programming. The general formulation of a linear integer program is

$$\begin{aligned} \text{maximize:} \quad & f^T z \\ \text{subject to:} \quad & Az \le b \\ & z \in \mathbb{Z}^n \end{aligned} \tag{1.14}$$

where $A \in \mathbb{Z}^{m \times n}$ and $b \in \mathbb{Z}^n$. Let $P = \{x \in \mathbb{R}^n \mid Ax \le b\}$. The set $P_I = \operatorname{conv}(\mathbb{Z}^n \cap P)$ is known as the *integer hull* of P: it is the convex hull of all integral points in P. As all extreme points of P_I are integral, solving $\max\{f^T x \mid x \in P_I\}$ is equivalent to solving (1.14). In the terminology of the general cutting plane algorithm, $\mathcal{F} = P_I$ and $\mathcal{F}_0 = P$.

Gomory, in [37] and [38], provides a method to find a linear inequality separating a nonintegral point y from P_I, and proves that the integer program (1.14) can be solved in a finite number of iterations using this method, when the polyhedron P is bounded. The cuts Gomory considered, which are now known as *Chvátal-Gomory cuts*, are of the form

$$c^T x \le \lfloor \delta(c, P) \rfloor,$$

where

$$c \in \mathbb{Z}^n, \quad \text{and} \quad \delta(c, P) = \max\{c^T x \mid x \in P\}.$$

These cuts arise from the observation that for any $c \in \mathbb{Z}^n$, the optimal value of

$$\max\{c^T z \mid z \in \mathbb{Z}^n,\ z \in P\}$$

is an integer.

Adding the Chvátal-Gomory cut for *every* $c \in \mathbb{Z}^n$ yields the *Chvátal-Gomory closure* of P, denoted P':

$$P' = \bigcap_{c \in \mathbb{Z}^n} \{x \mid c^T x \le \lfloor \delta(c, P) \rfloor\}.$$

It was proved by Schrijver in [60] that P', despite being defined by an infinite number of inequalities, is in fact a polyhedron. This result was recently extended by Dadush et al. in [29], where the authors show that the Chvátal-Gomory closure of *any* compact convex set (including a polytope defined by irrational data) is a polytope.

Defining $P^{(1)} = P'$, and then $P^{(i+1)} = P^{(i)\prime}$ for $i \geq 1$ gives a series $P^{(1)}, P^{(2)}, \ldots$ of progressively tighter approximations of P_I. Later, in [27], Chvátal showed that for any bounded P defined by a rational system $Ax \leq b$, there is a finite integer r (which could be very large) where $P^{(r)} = P_I$.

Chvátal-Gomory cuts are only one of a number of families of valid cuts for integer linear programs. Cornuéjols [28] gives a thorough overview of several different families of cuts, and the relationships between them, for mixed integer sets of the form

$$S = \{(x, y) \in \mathbb{R}^n \times \mathbb{Z}^p \mid Ax + Cy \leq b\}$$

which inludes the pure integer case when $p = 0$. The validity of any cut $\pi_x^T x + \pi_y^T y \leq \pi_0$ is equivalent to the condition

$$\{(x, y) \mid (x, y) \in S,\ \pi_x^T x + \pi_y^T y > \pi_0\} = \varnothing.$$

That is, the subset of S violating the cut does not contain any y-integral points. Such sets are termed *lattice free* sets.

1.4.2 Extensions to Mixed Integer Nonlinear Programming

Cutting plane algorithms have also been used outside of linear integer optimization, especially when the integer variables are 0/1. In these problems, the separation problem concerns cutting off fractional solutions of continuous relaxations. A simplified outline of the general procedure is as follows: Let $C \subseteq \mathbb{R}^n \times [0, 1]^p$ be a bounded convex set. Let $(x^*, y^*) \in C$ be a solution to a continuous relaxation of the following mixed 0/1 program with linear objective:

$$\begin{aligned} \text{maximize:} \quad & f_x^T x + f_y^T y \\ \text{subject to:} \quad & (x, y) \in C \\ & y \in \{0, 1\}^p \end{aligned} \tag{1.15}$$

with $y_j^* \in (0, 1)$. Because the objective in this problem is linear, we can assume that (x^*, y^*) is an extreme point of C. Define

$$C^0 = \{(x, y) \in C \mid y_j = 0\}, \quad \text{and } C^1 = \{(x, y) \in C \mid y_j = 1\}.$$

The union $C^0 \cup C^1$ contains the feasible set of (1.15), and thus $K = \text{conv}\,(C^0 \cup C^1)$ is a valid convex relaxation for (1.15). A point $(\hat{x}, \hat{y})$ is in K if and only if there is a solution $(\lambda_0, x_0, y_0, \lambda_1, x_1, y_1)$

to the system

$$(\hat{x}, \hat{y}) = \lambda_0(x_0, y_0) + \lambda_1(x_1, y_1) \tag{1.16}$$

$$(x_0, y_0) \in C^0, \quad (x_1, y_1) \in C^1 \tag{1.17}$$

$$\lambda_0, \lambda_1 \geq 0, \quad \lambda_0 + \lambda_1 = 1 \tag{1.18}$$

This system is nonconvex, but a convex reformulation can be obtained through a change of variables or transformation. Because (x^*, y^*) is an extreme point of C, we have $(x^*, y^*) \notin K$, which implies that there is no solution to the system (1.16)-(1.18), nor to its convex reformulation, when $(\hat{x}, \hat{y}) = (x^*, y^*)$. A hyperplane separating (x^*, y^*) from K is found using this fact, and is added to the constraints in (1.15).

Balas et al. introduce this approach, called the *disjunctive* approach, for the linearly constrained case in [9]. Stubbs and Mehrotra [63] provide a generalization for the case when C is defined by convex inequalities $g_i(x, y) \leq 0$, and later describe a method for obtaining valid nonlinear inequalities in this setting. Ceria and Soares [26] extend these ideas to more general disjunctive feasible sets. Both [63] and [26] use the perspective transformation to achieve convexity. This presents a problem computationally as the perspective function, while convex, is not everywhere differentiable and therefore not compatible with general nonlinear programming algorithms. Moreover, the formulation used to generate cuts contains multiples of the number of variables in the original problem. Both [18] and [44] seek to avoid these difficulties and propose methods for generating valid cuts using *splits* (see [28] for an introduction to split cuts). The former uses nonlinear programming to find a cut separating a fractional point, while the latter uses a sequence of linear programs.

In [42], Iyengar and Çezik present a method of obtaining linear and quadratic cuts in mixed 0/1 convex quadratically-constrained problems, using SDP duality to formulate the separation problem. The same authors extend their results to mixed 0/1 semidefinite programming in [43]. Although the settings we consider do not necessarily include 0/1 variables, the use of duality to generate cuts closely resembles the approach we take in Section 3.2.

Underlying much of this work is the goal of obtaining the convex hull of the union of two or more convex sets - for instance, the sets C^0 and C^1 from above. Recent papers by Modaresi et al. [48] and Belotti et al. [12] address this directly. [48] focuses mainly on sets obtained through quadratic constraints and split disjunctions, but also includes results describing the convex hull of the difference of two quadratically-defined sets. [12] focuses on the case of a conic set with a linear disjunction. The work in this thesis is similar: we will seek to obtain the convex hull of a set obtained by taking the epigraph of a convex function over a non-convex region of $\mathbb{R}^n$.

An interesting and illustrative application of the cutting plane methodology in a purely continuous nonlinear problem comes from Qualizza et al. in [55], in which cutting planes are used to approximate the cone of $(n \times n)$ positive semidefinite matrices. Consider a semidefinite program of the following form:

$$\begin{aligned} \text{minimize:} \quad & \operatorname{tr}(X) \\ \text{subject to:} \quad & \operatorname{tr}(A_i X) \geq 1 \quad i = 1, \ldots, m \\ & X \succeq 0 \end{aligned} \tag{1.19}$$

Removing the constraint $X \succeq 0$ and instead constraining X only to lie in the set of symmetric $(n \times n)$ matrices yields a linear program. Suppose such a linear relaxation is solved, and let X^* be the optimal solution. If $X^* \succeq 0$, then X^* is feasible (and therefore optimal) for (1.19). Otherwise, X^* has at least one eigenvector-eigenvalue pair (v, λ) with $v^T X^* v = \operatorname{tr}(vv^T X^*) = \lambda < 0$. Adding the valid linear cut $\operatorname{tr}(vv^T X) \geq 0$ to the formulation (1.19) cuts off X^*, and in fact one cut can be added for each negative eigenvalue of X^*. The authors use this idea, with several heuristics to promote sparsity in the coefficients of the cuts, to provide bounds on problems similar to (1.19), which are themselves relaxations of nonconvex quadratically constrained quadratic programs.

Chapter 2

Approach and General Lifting Results

In this chapter, we describe the settings we wish to consider and introduce some terminology which will be used throughout. We also provide some general results on the construction of lifted inequalities in our settings of interest, as well as their ability to describe the convex hull of the epigraph of a convex differentiable function over a general domain.

2.1 Generic Setting and Approach

In our general setting, we consider an optimization problem which includes a function $f(x) : \mathbb{R}^n \to \mathbb{R}$ in either the objective or constraints. We assume that f is convex and differentiable, and that the feasible set for x, which we denote by $\mathcal{F}$, is a nonconvex subset of $\mathbb{R}^n$. We refer to any point $x \in \mathcal{F}$ as *feasible*. In the specific cases for $\mathcal{F}$ that we consider, relaxing $x \in \mathcal{F}$ to $x \in \operatorname{conv}(\mathcal{F})$ would result in a very poor bound on the value of the optimization problem. We aim to find methods to strenghen this bound.

2.1.1 Epigraph Formulation

We introduce a new variable z and study the set

$$\mathcal{S} = \{(x, z) \in \mathbb{R}^{n+1} \mid z \geq f(x),\ x \in \mathcal{F}\},$$

the epigraph of f over the domain $\mathcal{F}$. The use of the epigraph is helpful in cutting plane methods with a nonlinear convex objective function, as it guarantees that the optimal solution to a relaxation will occur at an extreme point. Without the epigraph formulation, the optimal solution to a relaxation may be in the interior of the relaxed feasible set, and it may be impossible to separate this solution from $\mathcal{F}$ with a linear inequality. In our setting, it allows us to move from the weak bound provided by the relaxation

$$z \geq f(x), \quad x \in \operatorname{conv}(\mathcal{F})$$

to the stronger formulation

$$(x, z) \in \operatorname{conv}(\mathcal{S}).$$

We will attempt to obtain, through the use of lifted inequalities, a linear inequality description of $\operatorname{conv}(\mathcal{S})$.

2.1.2 General Lifting Construction

Let y be any point in $\mathbb{R}^n$. As f is convex, the inequality

$$f(x) \geq \nabla f(y)^T (x - y) + f(y)$$

holds for all $x \in \mathbb{R}^n$. With the addition of the extra variable z, we can express this as the *linear* inequality

$$z \geq \nabla f(y)^T (x - y) + f(y) \tag{2.1}$$

which is valid for all (x, z) with $z \geq f(x)$. In particular, this inequality is valid for the set $\mathcal{S}$ and supports $\mathcal{S}$ at the point $(y, f(y))$ when $y \in \mathcal{F}$. This inequality, which will serve as the basis for our lifting procedures, will be referred to as the *tangent* inequality generated at y.

By the differentiability of f, when $y \in \operatorname{int}(\mathcal{F})$ there is a *unique* hyperplane supporting $\mathcal{S}$ at $(y, f(y))$, namely the set of points (x, z) satisfying (2.1) with equality. On the other hand, when y is on the boundary of $\mathcal{F}$, it may be possible to modify the coefficients of the tangent inequality generated at y in such a way that the resulting inequality is still valid for $\mathcal{S}$, but provides a tighter description of $\operatorname{conv}(\mathcal{S})$. We now provide a definition of the type of modifications we consider.

Definition 2.1.1. *Let $y \in \partial\mathcal{F}$, $\lambda \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}_+$. The inequality*

$$z \geq \left(\nabla f(y) + \alpha\lambda\right)^T (x - y) + f(y), \tag{2.2}$$

is called a lifted first-order inequality generated from y, with lifting normal λ and lifting coefficient α.

All of the lifted inequalities we consider are in fact lifted first-order inequalities, and for the most part we will refer to them more simply as lifted inequalities. The appearance of both λ and α together in this definition may seem superfluous, but will prove convenient later. We do allow the possibility that either $\alpha = 0$ or $\lambda = 0$, in which case the lifted inequality is simply the tangent inequality at y, but of course we are most interested in the cases where α and λ are nonzero. Note that for points x with $\alpha\lambda^T(x - y) > 0$, the lifted inequality is a *strengthening* of the tangent inequality, in that it implies a higher value of z, our linearized proxy for the objective value.

For the lifted inequality (2.2) to be valid for $\mathcal{S}$, we need

$$f(x) \geq (\nabla f(y) + \alpha\lambda)^T (x - y) + f(y) \quad \text{for all } x \in \mathcal{F},$$

or equivalently

$$z \geq (\nabla f(y) + \alpha\lambda)^T (x - y) + f(y) \quad \text{for all } (x, z) \in \mathcal{S}.$$

Suppose a point $(x, z) \in \mathcal{S}$ violates the lifted inequality. Then because $z \geq f(x)$ for all $(x, z) \in \mathcal{S}$, the point $(x, f(x))$ violates the lifted inequality as well. If we define $V = V(y, \lambda, \alpha)$ to be the set of all points x such that $(x, f(x))$ violates (2.2), an equivalent characterization of validity is $\operatorname{int}(V) \cap \mathcal{F} = \varnothing$. A recurring theme in this work will be to maximize the strength of the lifted inequality, subject to its validity.

2.1.3 Separation

In general, the linear inequality description of the set $\operatorname{conv}(\mathcal{S})$ will require an infinite number of inequalities and will not have a convenient closed-form description. We therefore focus on finding efficient solutions to the problem of *separating* from $\operatorname{conv}(\mathcal{S})$ by means of lifted inequalities, with the intention that our methods could be used as a component in a cutting plane algorithm. The separation problem is formally stated below.

Problem 2.1.2 (The Separation Problem)**.** *Given a point $(\hat{x}, \hat{z})$ with $\hat{x} \in \operatorname{conv}(\mathcal{F})$, find $\gamma \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ where*

$$z \geq \gamma^T x + \beta \quad \forall\ (x, z) \in \mathcal{S}$$

and

$$\hat{z} < \gamma^T \hat{x} + \beta,$$

or provide proof that such γ and β do not exist.

We point out that the set $\mathcal{F}$ does not necessarily need to be the feasible region for the "original" problem being solved, but could be a *relaxed* feasible region with a simpler structure which allows for easier separation from the set $\mathcal{S}$.

2.2 Lifting in the General Case

In this section we present some of the results from [17] concerning the construction of lifted first order inequalities for general convex and differentiable functions f and feasible sets $\mathcal{F}$. We introduce a necessary condition for when a tangent inequality can be lifted, and show that first-order lifted inequalities, together with tangent inequalities and inequalities valid for $\mathcal{F}$, are sufficient to describe the convex hull of the epigraph set $\mathcal{S}$.

Throughout, we assume that $\mathcal{F}$ and $\operatorname{conv}(\mathcal{F})$ are closed. We define the set $P = \operatorname{cl}(\mathbb{R}^n \setminus \mathcal{F})$, so we have $\mathcal{F} = \mathbb{R}^n \setminus \operatorname{int}(P)$.

We begin with a definition, which we then establish as a necessary condition for constructing a nontrivial lifted inequality from a point y.

Definition 2.2.1. *Let P be a full-dimensional subset of $\mathbb{R}^n$. Let $y \in \partial P$ and $p \in \mathbb{R}^n$, with $||p|| = 1$. P is called locally flat at y with normal p if for every $r \in \mathbb{R}^n$ with $p^T r > 0$, there exists an $\epsilon > 0$ with*

$$y + \delta r \in \operatorname{int}(P) \quad \textit{for all } \delta \in (0, \epsilon).$$

Lemma 2.2.2. *Let $y \in \partial P$, and let*

$$z \geq (\nabla f(y) + \lambda)^T (x - y) + f(y) \tag{2.3}$$

be valid for $\mathcal{S}$, with $\lambda \neq 0$. Then P is locally flat at y, with normal $\frac{\lambda}{||\lambda||}$.

Proof. Let $r \in \mathbb{R}^n$ with $\lambda^T r > 0$. Define

$$g(\delta) = f(y + \delta r), \qquad \text{and} \qquad h(\delta) = f(y) + \delta(\nabla f(y) + \lambda)^T r.$$

Then we have $g(0) = h(0) = f(y)$, and

$$h'(0) \;=\; (\nabla f(y) + \lambda)^T r \;=\; \nabla f(y)^T r + \lambda^T r \;>\; \nabla f(y)^T r = g'(0).$$

This implies $g(\delta) < h(\delta)$, or equivalently

$$f(y + \delta r) < f(y) + \delta(\nabla f(y) + \lambda)^T r$$

for δ positive and approaching 0. Then for any such δ, the inequality (2.3) is violated at $(y + \delta r, f(y + \delta r))$. As this inequality is valid for $\mathcal{S}$, we must have $y + \delta r \in \text{int}(P)$, which concludes the proof. ■

The following examples should help give a better understanding of local flatness.

Example 2.2.3. *(a) Let P be a convex polygon in $\mathbb{R}^2$. Then P is locally flat at every point on its boundary* except *the vertices; using as normals the unit vectors normal to the facets, oriented into P. (b) The non-convex set $P = \{(x_1, x_2) \in \mathbb{R}^2 \ : \ |x_1| \geq x_2\}$ is locally flat at every point on its boundary, even the vertex at $(0, 0)$ (with normal $(0, -1)$).*

As mentioned, local flatness of the set P at a point y is a necessary condition for the tangent inequality at y to be nontrivially lifted. Unfortunately, this condition is *not* sufficient, as the next example shows.

Example 2.2.4. *In $n = 2$, Let P be the ellipsoid $\{x \in \mathbb{R}^2 \mid \frac{1}{9}x_1^2 + x_2^2 \leq 1\}$, and $f(x) = x^T x$. Then P is locally flat at every point on its boundary. However, at the point $y = (3, 0) \in \partial P$, the only inequality of the form (2.3) valid for $\mathcal{S}$ has $\lambda = 0$.*

The intuitive reason for the negative result in this example is that it is impossible for a ball of positive radius to be contained in P and contain the point $(x_1, x_2) = (3, 0)$ in its boundary. We will return to the problem of constructing lifted inequalities in the case where P is an ellipsoid in Section 3.3.

Our next result provides a classification of all linear inequalities that are valid for $\mathcal{S}$ and support $\mathcal{S}$ at some point.

Theorem 2.2.5. *Let*

$$\delta z \geq \gamma^T x + \beta \tag{2.4}$$

be valid for $\mathcal{S}$. Assume that this inequality is tight at a point $(y, f(y))$ with $y \in \mathcal{F}$. Then one of the following three conditions holds.

(1) *$\delta = 0$, and $\gamma^T x + \beta \leq 0$ is valid for $\mathcal{F}$.*

(2) *$\delta > 0$, and $\delta z \geq \gamma^T x + \beta$ is a positive multiple of the tangent inequality at y.*

(3) *$\delta > 0$, and $\delta z \geq \gamma^T x + \beta$ is a positive multiple of a lifted first-order inequality generated at y.*

Proof. By the structure of $\mathcal{S}$, it clearly must be the case that δ is nonnegative. First assume that $\delta = 0$. In this case,

$$\gamma^T x + \beta \leq 0$$

must be valid for $\mathcal{F}$, so case **(1)** holds.

From here on we assume that $\delta > 0$, and without loss of generality that $\delta = 1$, so the inequality in question is now

$$z \geq \gamma^T x + \beta. \tag{2.5}$$

If $y \in \operatorname{int}(\mathcal{F})$, then by the differentiability of f, (2.5) is simply the tangent inequality at y, implying case **(2)** holds.

Now we assume $y \in \partial \mathcal{F}$. Define

$$\lambda = \gamma - \nabla f(y).$$

Then, using the fact that $\beta = \gamma^T y - f(y)$, we can write (2.5) as

$$z \geq (\nabla f(y) + \lambda)^T (x - y) + f(y)$$

If $\lambda = 0$, then (2.5) is simply the tangent inequality at y. We assume $\lambda \neq 0$, and case **(2)** holds. If $\lambda \neq 0$, then (2.5) is a nontrivial lifted first-order inequality and case **(3)** holds. ■

This theorem leads immediately to the following result.

Theorem 2.2.6. *Let $(\hat{x}, \hat{z}) \in \mathbb{R}^{n+1}$ with $\hat{x} \in \operatorname{conv}(\mathcal{F})$, $\hat{z} \geq f(\hat{x})$, and $(\hat{x}, \hat{z}) \notin \operatorname{conv}(\mathcal{S})$. Then there is a lifted first-order inequality separating $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$.*

Proof. Because $\operatorname{conv}(\mathcal{S})$ is a closed convex set and does not contain $(\hat{x}, \hat{z})$, there is some linear inequality

$$\delta z \geq \gamma^T x + \beta$$

which is valid for $\mathcal{S}$ and separates $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$. We can assume without loss of generality that this inequality supports $\mathcal{S}$. We now consider each of the possibilities of Theorem 2.2.5.

If $\delta = 0$, then $\gamma^T x + \beta \leq 0$ is valid for $\mathcal{F}$. But because $\hat{x} \in \operatorname{conv}(\mathcal{F})$, this inequality cannot separate $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$. So we have $\delta > 0$.

As $\hat{z} \geq f(\hat{x})$, we know that $(\hat{x}, \hat{z})$ must satisfy the tangent inequality at every point $y \in \mathcal{F}$. So the inequality separating $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$ cannot be a tangent inequality (or a positive multiple of a tangent inequality).

By Theorem 2.2.5, the only case remaining is that the separating inequality is, up to a scaling, a lifted first-order inequality. ■

The importance of Theorem 2.2.6 is that given a point $(\hat{x}, \hat{z})$ with $\hat{z} \geq f(\hat{x})$ and $\hat{x} \in \operatorname{conv}(\mathcal{F})$, we need only to consider lifted first-order inequalities to solve the separation problem.

Chapter 3

Positive Definite Quadratic Objectives

In this chapter we study the specific case where the $f(x)$ is a positive definite quadratic function. Quadratic objectives arise in a number of applications, as a measure of distance or error ([20]), variance ([15]), or power ([35]). We consider sets of the form

$$\mathcal{S} = \{(x, z) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \mathcal{F},\ z \geq f(x)\}, \tag{3.1}$$

where $\mathcal{F}$ is the feasible region for the x variables. We provide characterizations of lifted inequalities that are valid for $\operatorname{conv}(\mathcal{S})$, and demonstrate how these inequalities can be used to separate from $\operatorname{conv}(\mathcal{S})$ in polynomial time, in the cases where the feasible set $\mathcal{F}$ is

- The complement of the interior of a polyhedron
- A union of polyhedra (which includes the previous case, but is handled differently)
- The complement of the interior of an ellipsoid.

We begin with a simplifying assumption in this case.

Proposition 3.0.7. *If $f(x)$ is a positive definite quadratic function, then without loss of generality, we can assume $f(x) = x^T x$, the squared Euclidean norm.*

Proof. Assume $f(x)$ is given by

$$f(x) = x^T H x - 2h^T x + h_0.$$

Let $H = C^TC$ be the Cholesky decomposition of H. By means of the linear transformation $y = C(x - H^{-1}h)$, or $x = C^{-1}y + H^{-1}h$, we have

$$\begin{aligned} x^THx - 2h^Tx + h_0 &= (C^{-1}y + H^{-1}h)^T H(C^{-1}y + H^{-1}h) - 2h^T(C^{-1}y + H^{-1}h) + h_0 \\ &= y^TC^{-T}HC^{-1}y + 2h^TC^{-1}y - h^TH^{-1}h - 2h^TC^{-1}y + h_0 \\ &= y^TC^{-T}C^TCC^{-1}y + h^TH^{-1}h - 2h^TH^{-1}h + h_0 \\ &= y^Ty - h^TH^{-1}h + h_0 \end{aligned}$$

which is, up to an additive constant, the squared Euclidean norm. ■

Now consider any inequality

$$z \geq \gamma^T x + \beta. \tag{3.2}$$

A point (x, x^Tx) violates this inequality if and only if

$$x^Tx < \gamma^Tx + \beta$$

which can be stated equivalently as

$$\left|\left|x - \frac{1}{2}\gamma\right|\right|^2 < \frac{1}{4}\left|\left|\gamma\right|\right|^2 + \beta.$$

That is, the set of points x for which (x, x^Tx) violates the inequality (3.2) is the interior of the ball with center $\frac{1}{2}\gamma$ and radius $\sqrt{\frac{1}{4}\left|\left|\gamma\right|\right|^2 + \beta}$. This leads to the following characterization for validity of linear inequalities over $\mathcal{S}$.

Proposition 3.0.8. *A linear inequality*

$$z \geq \gamma^T x + \beta \tag{3.3}$$

is valid for $\mathcal{S}$ if and only if the interior of the ball

$$B = \mathcal{B}\left(\frac{1}{2}\gamma,\ \sqrt{\frac{1}{4}\left|\left|\gamma\right|\right|^2 + \beta}\right)$$

does not intersect $\mathcal{F}$.

Proof. Let $\hat{x} \in \mathcal{F}$. If $\hat{x} \in \operatorname{int}(B)$, we have

$$\hat{x}^T\hat{x} < \gamma^T\hat{x} + \beta.$$

Let $\hat{z} = \hat{x}^T\hat{x}$. Then $(\hat{x}, \hat{z}) \in \mathcal{S}$ and violates (3.3), so the inequality is not valid for $\mathcal{S}$.

Conversely, if the inequality (3.3) is not valid for $\mathcal{S}$, then there is some point $(\hat{z}, \hat{x})$ with $\hat{x} \in \mathcal{F}$ and $\hat{x}^T\hat{x} \leq \hat{z}$ which violates (3.3). But then we have

$$\hat{x}^T\hat{x} \leq \hat{z} < \gamma^T\hat{x} + \beta,$$

which implies $x \in \operatorname{int}(B)$. Because $\hat{x} \in \mathcal{F}$, we have that $\operatorname{int}(B)$ intersects $\mathcal{F}$. ■

This result allows us to characterize validity geometrically, and in the space of the original x variables only, which will prove useful later as we devise constraints to ensure the validity of lifted inequalities.

With $f(x) = x^T x$ the tangent inequality at a point y, which is valid for all $(x, z) \in \mathbb{R}^n \times \mathbb{R}$, is

$$z \quad \geq \quad 2y^T(x - y) + y^T y = 2y^T x - y^T y. \tag{3.4}$$

In what follows, we use this inequality as the base inequality in our lifting procedures.

3.1 Excluding a Polyhedron

The first case we consider is when the feasible region is given as the complement of the interior of a single polyhedron. Let

$$P = \{x \in \mathbb{R}^n \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\} = \{x \in \mathbb{R}^n \mid Ax \leq b\}$$

be the polyhedron, and define

$$\mathcal{F} = \mathbb{R}^n \setminus \{x \mid Ax < b\},$$

so

$$\mathcal{S} = \{(x, z) \in \mathbb{R}^{n+1} \mid Ax \not< b,\ z \geq x^T x\}.$$

We assume that $m \geq 2$, that P is full-dimensional and that every inequality in the system $Ax \leq b$ defines a facet of P; that is, for each $i \in \{1, \ldots, m\}$ there is some x with $a_i^T x = b_i$ and $a_j^T x < b_j$ for $j \neq i$.

In this section, we prove constraints on the structure of valid lifted inequalities for the set $\mathcal{S}$, and derive expressions for the lifting coefficients of these inequalities. Using this, we develop a procedure which solves the separation problem for $\mathcal{S}$ by solving a quadratic program in $n + 1$ variables for each facet of P. We contrast this separation procedure with one obtained using a disjunctive programming formulation and discuss some computational considerations with our method.

3.1.1 The Structure of Valid Inequalities

For any $\bar{x} \in \mathbb{R}^n$, the tangent inequality

$$z \geq 2\bar{x}^T(x - \bar{x}) + ||\bar{x}||^2 \tag{3.5}$$

is valid for $\mathcal{S}$ and supports $\mathcal{S}$ at $(\bar{x}, \bar{x}^T\bar{x})$. If $\bar{x} \notin P$, this is the unique (up to multiplication by a nonnegative scalar) inequality with these two properties. We seek lifted cuts of the form

$$z \geq (2\bar{x} - \lambda)^T(x - \bar{x}) + ||\bar{x}||^2 \tag{3.6}$$

where $\bar{x} \in \partial P$ and $\lambda \in \mathbb{R}^n$. The following lemma describes which choices of λ result in valid cuts, in a slightly more general setting.

Lemma 3.1.1. *Let $\mathcal{F}$ be given as the union of a finite number of polyhedra:*

$$\mathcal{F} = \bigcup_{k=1}^{K} \left\{ x \mid C^k x \geq d^k \right\}$$

Let $\bar{x} \in \mathcal{F}$, so $C^k\bar{x} \geq d^k$ for some particular index k. Assume without loss of generality that

$$\begin{array}{rcll} {c_i^k}^T \bar{x} & = & d_i^k & \textit{for} \;\; i = 1, \ldots, p \\ {c_i^k}^T \bar{x} & > & d_i^k & \textit{for} \;\; i > p. \end{array}$$

If the inequality

$$z \geq \gamma^T x + \beta \tag{3.7}$$

is valid for $\mathcal{S}$ and holds with equality at $(\bar{x}, \bar{x}^T\bar{x})$, then

$$\gamma = 2\bar{x} - \sum_{i=1}^{p} \alpha_i c_i^k \tag{3.8}$$

$$\beta = \sum_{i=1}^{p} \alpha_i d_i^k - ||\bar{x}||^2 \tag{3.9}$$

with $\alpha_i \geq 0$ for each i.

Proof. Suppose that the inequality (3.7) is valid but γ *cannot* be expressed in the form (3.8). By Farkas's Lemma, this implies that there is a vector π where $(2\bar{x} - \gamma)^T\pi = -1$ and ${c_i^k}^T\pi \geq 0$ for each $i = 1, \ldots, p$. The latter of these conditions implies that the point $\bar{x} + \epsilon\pi$ is feasible for nonnegative ϵ, and thus the cut must be satisfied at the point $(x, z) = \left(\bar{x} + \epsilon\pi, ||\bar{x} + \epsilon\pi||^2\right)$ for $\epsilon \geq 0$. Substituting into (3.7), we get that this condition is equivalent to

$$\begin{aligned} & \gamma^T(\bar{x} + \epsilon\pi - \bar{x}) + \beta \leq ||\bar{x} + \epsilon\pi||^2 \\ \Leftrightarrow \quad & \epsilon\pi^T\gamma \leq 2\epsilon\pi^T\bar{x} + \epsilon^2\,||\pi||^2 \\ \Leftrightarrow \quad & \epsilon(\gamma - 2\bar{x})^T\pi \leq \epsilon^2\,||\pi||^2 \\ \Leftrightarrow \quad & \epsilon - \epsilon^2\,||\pi||^2 \leq 0. \end{aligned}$$

This fails to hold for ϵ close to 0. Thus we have a contradiction, and the inequality cannot be valid.

The equality (3.9) can now be established using the fact that (3.7) is tight at $\left(\bar{x}, ||\bar{x}||^2\right)$.

This proof also covers the case when $C^k\bar{x} > d^k$, in which case we have $\gamma = 2\bar{x}$ and $\beta = ||\bar{x}||^2$: the inequality (3.7) is the tangent inequality generated from $\bar{x}$. ■

This result shows that, for $\mathcal{F}$ and $\bar{x}$ as described in the lemma, the vector λ in (3.6) must be of the form

$$\lambda = \sum_{k=1}^{K} \alpha_i c_i^k$$

with $\alpha_i \geq 0$ for all i.

As a corollary, we obtain the necessary form of the vector λ when there is only a single excluded polyhedron.

Corollary 3.1.2. *Let*

$$\mathcal{F} = \mathbb{R}^n \setminus \{x \mid Ax < b\}$$

Suppose $A\bar{x} \leq b$ and that ${a_i}^T\bar{x} = b_i$ for some particular index i. If the lifted inequality (3.6) is valid and tight at $(\bar{x}, \bar{x}^T\bar{x})$, then

$$\lambda = \alpha a_i$$

for some $\alpha \geq 0$.

In this case we refer to α as the lifting coefficient. From Lemma 3.1.1, we can see that in this simpler setting, if $\bar{x}$ is not in the relative interior of the i^{th} facet of P, then α must be 0. This is because if $a_j^T\bar{x} = b_j$ for some $j \neq i$, then we would need $\lambda = \alpha a_i = \beta a_j$ for some positive α and nonnegative β. However, this is impossible due to the assumption that each inequality in the definition of P is facet-defining. Recalling Definition 2.2.1, this result states that P is not locally flat (for any choice of normal vector) at any point not in the relative interior of one of its facets. Moreover, if P is locally flat with normal p at a point y with $a_i^T y = b_i$ for some i, then we must have $p = \frac{-a_i}{||a_i||}$.

3.1.2 Deriving the Maximum Lifting Coefficient

Returning to the case where $\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax \not< b\}$, we now focus on deriving the strongest lifted cut of the form

$$z \geq (2\bar{x} - \alpha a_i)^T (x - \bar{x}) + ||\bar{x}||^2 \tag{3.10}$$

where i is a fixed but arbitrary index and $\bar{x}$ is a fixed point with $a_i^T \bar{x} = b_i$ and $a_j^T \bar{x} \leq b_j$ for $j \neq i$. We refer to any inequality of this form as a lifted inequality generated from $\bar{x}$. This section establishes that the lifting coefficient giving the strongest lifted inequality generated from $\bar{x}$ can be obtained in closed form, which will lead to a polynomial-time solvable formulation of the separation problem.

For any $x \in \text{int}\,(P)$, we have $a_i^T x < b_i$ and so the cut value

$$(2\bar{x} - \alpha a_i)^T (x - \bar{x}) + \bar{x}^T \bar{x} \quad = \quad -\bar{x}^T \bar{x} + 2\bar{x}^T x + \alpha(b_i - a_i^T x) \tag{3.11}$$

is increasing in α. Thus in order to get the strongest lifted inequality, we wish to make α as large as possible while preserving validity. This motivates the following definitions.

Definition 3.1.3. *Let $\bar{x}$ be a point in the i^{th} facet of P, and let $\mathcal{C} \subseteq \mathcal{S}$. Assume $\mathcal{C}$ contains a point (x', z') with $a_i^T x' < b_i$. Then let*

$$\alpha^*(\bar{x}, \mathcal{C}) = \sup \{\alpha \mid \textit{inequality (3.10) is valid for } \mathcal{C}\}.$$

Definition 3.1.4. *Let $\bar{x}$ be a point in the i^{th} facet of P. The lifted inequality*

$$z \geq (2\bar{x} - \alpha^*(\bar{x}, \mathcal{S}) a_i)^T (x - \bar{x}) + ||\bar{x}||^2$$

is called the nondominated lifted inequality generated from $\bar{x}$.

By the assumption about the point $(x', z') \in \mathcal{C}$ in Definition 3.1.3, the value of $\alpha^*(\bar{x}, \mathcal{C})$ is finite. If it is the case that $a_i^T x \geq b_i$ for all $x \in \mathcal{F}$, then the inequality (3.10) is valid for any nonnegative

α. However, this possibility is precluded by our assumption that P has at least two facets, meaning $\alpha^*(\bar{x}, \mathcal{S})$ is finite. Note that we have not (yet) proved that the nondominated lifted inequality is actually valid.

A key insight in deriving the maximum lifting coefficient will be the ability to write the excluded polyhedron P as the intersection of $(m-1)$ polyhedra defined by only two inequalities:

$$P = \bigcap_{j \neq i} \{x \mid a_i^T x \leq b_i,\ a_j^T x \leq b_j\}.$$

With this in mind, let $j \neq i$ be another fixed index, and let $\bar{x}$ be a point with $a_i^T \bar{x} = b_i$ and $a_j^T \bar{x} \leq b_j$. Define

$$W_{ij} = \{x \mid a_i^T x \leq b_i,\ a_j^T x \leq b_j\},$$

which is nonempty by assumption. We wish to find a lifted inequality of the form (3.10) which is valid for the set

$$\mathcal{S}_{ij} = \{(x, z) \mid x \notin \operatorname{int}(W_{ij}),\ z \geq x^T x\},$$

a subset of $\mathcal{S}$.

The first result of this section narrows down the set of points (x, z) we need to consider to ensure validity of the lifted cut (3.10) over $\mathcal{S}_{ij}$.

Lemma 3.1.5. *Suppose a point $(x', z') \in \mathcal{S}_{ij}$ violates (3.10). Then $a_i^T x' < b_i$.*

Proof. As $(x', z') \in \mathcal{S}_{ij}$, we have ${x'}^T x' \leq z'$ and thus

$$\begin{aligned} {x'}^T x' \leq z' < (2\bar{x} - \alpha a_i)^T (x' - \bar{x}) + \bar{x}^T \bar{x} \quad &\Rightarrow \quad \left|\left|\bar{x} - x'\right|\right|^2 < -\alpha a_i^T (x' - \bar{x}) \\ &\Rightarrow \quad 0 < \alpha a_i^T (\bar{x} - x') \\ &\Rightarrow \quad a_i^T x' < a_i^T \bar{x} \\ &\Rightarrow \quad a_i^T x' < b_i \end{aligned}$$

■

Lemma 3.1.5 tells us that the validity of (3.10) over the set

$$\left\{(x, z) \mid a_j^T x \geq b_j,\ z \geq x^T x\right\}$$

is sufficient to guarantee its validity over $\mathcal{S}_{ij}$. This leads to the following characterization of validity of the lifted cut over $\mathcal{S}_{ij}$.

Lemma 3.1.6. *The lifted cut (3.10) is valid over $\mathcal{S}_{ij}$ if and only if the following inequality holds*

$$\left\{ \begin{array}{ll} \textit{minimize:} & ||x||^2 - (2\bar{x} - \alpha a_i)^T x + ||\bar{x}||^2 - \alpha b_i \\ \textit{subject to:} & a_j^T x \geq b_j \end{array} \right\} \geq 0. \tag{3.12}$$

Proof. Suppose (3.12) holds. If any point (x', z') violates the lifted cut (3.10), then $(x', {x'}^T x')$ violates it as well, and we have

$$ {x'}^T x' < (2\bar{x} - \alpha a_i)^T (x' - \bar{x}) + \bar{x}^T \bar{x}. $$

If $a_j^T x' \geq b_j$, then we have a contradiction with (3.12), so it must be the case that $a_j^T x' < b_j$. By Lemma 3.1.5 we have $a_i^T x' < b_i$. So $x' \in \operatorname{int}(W_{ij})$, which means $(x', z') \notin \mathcal{S}_{ij}$. This implies that the cut is valid over $\mathcal{S}_{ij}$.

Now suppose (3.12) does not hold: there exists some x' with $a_j^T x' \geq b_j$ and

$$ {x'}^T x' < (2\bar{x} - \alpha a_i)^T (x' - \bar{x}) + \bar{x}^T \bar{x}. $$

Then $(x', {x'}^T x') \in \mathcal{S}_{ij}$, and this point violates (3.10), implying that the lifted cut is not valid over $\mathcal{S}_{ij}$. ■

For any α, we define V_α to be the set of points x for which $(x, x^T x)$ (weakly) violates the lifted cut. As shown previously, V_α is a ball: specifically we have

$$\begin{aligned} x \in V_\alpha \quad &\Leftrightarrow \quad x^T x \leq (2\bar{x} - \alpha a_i)^T (x - \bar{x}) + \bar{x}^T \bar{x} \\ &\Leftrightarrow \quad x^T x - 2\left(\bar{x} - \frac{\alpha}{2} a_i\right)^T x + \bar{x}^T \bar{x} - \alpha b_i \leq 0 \\ &\Leftrightarrow \quad \left|\left| x - \left(\bar{x} - \frac{\alpha}{2} a_i\right)\right|\right|^2 - \left|\left| \bar{x} - \frac{\alpha}{2} a_i \right|\right|^2 + ||\bar{x}||^2 - \alpha b_i \leq 0 \\ &\Leftrightarrow \quad \left|\left| x - \left(\bar{x} - \frac{\alpha}{2} a_i\right)\right|\right|^2 \leq \frac{\alpha^2}{4} ||a_i||^2 . \end{aligned} \tag{3.13}$$

Note that we can state the condition $x \in V_\alpha$ equivalently as

$$ ||x - \bar{x}||^2 + \alpha(a_i^T x - b_i) \leq 0. $$

When x is fixed and $a_i^T x < b_i$, the left hand side of this expression is decreasing in α. This means that V_α grows with α, in the sense that $V_\alpha \subset V_{\alpha'}$ for $\alpha < \alpha'$. This implies that if the lifted cut (3.10) is valid with $\alpha = \alpha'$ for some particular value α', then it is valid for any $\alpha \in [0, \alpha']$. Using this fact, we can show that the value of the supremum in Definition 3.1.3 provides a valid inequality.

Lemma 3.1.7. *Let $\bar{x}$ and $\mathcal{C}$ be as in Definition 3.1.3. Let $\alpha^* = \alpha^*(\bar{x}, \mathcal{C})$. Then the lifted cut*

$$z \geq (2\bar{x} - \alpha^* a_i)^T (x - \bar{x}) + ||\bar{x}||^2$$

is valid.

Proof. Suppose not. Then there exists some $y \in \mathcal{C}$ with $y \in \text{int}\,(V_{\alpha^*})$. This implies $\alpha^* > 0$. By the definition of $\alpha^*(\bar{x}, \mathcal{C})$, the lifted cut is valid for any $\alpha \in [0, \alpha^*)$. However, for ϵ positive and small enough, we have $y \in \text{int}\,(V_{\alpha^*-\epsilon})$, implying that the lifted cut with $\alpha = \alpha^* - \epsilon$ is not valid, a contradiction. ■

An immediate consequence of this result is that the nondominated lifted inequality as defined in Definition 3.1.4 is in fact valid.

We now continue the derivation of the quantity $\alpha^*(\bar{x}, \mathcal{S}_{ij})$. The convexity of V_α allows the following tightening of Lemma 3.1.6.

Lemma 3.1.8. *The lifted cut (3.10) is valid over $\mathcal{S}_{ij}$ if and only if the following inequality holds*

$$\left\{ \begin{array}{ll} \textit{minimize:} & ||x||^2 - (2\bar{x} - \alpha a_i)^T x + ||\bar{x}||^2 - \alpha b_i \\ \textit{subject to:} & a_j^T x = b_j \end{array} \right\} \geq 0. \tag{3.14}$$

Proof. Suppose (3.14) holds, but the cut is not valid over $\mathcal{S}_{ij}$. Then there is some point $(x', x'^T x') \in \mathcal{S}_{ij}$ which violates the cut. By Lemma 3.1.5, we have $a_i^T x' < b_i$, so if $(x', x'^T x') \in \mathcal{S}_{ij}$, we must have $a_j^T x' \geq b_j$. Then by (3.14) we have $a_j^T x' > b_j$.

Now, $\bar{x}$ is on the boundary of the ball V_α, and $a_j^T \bar{x} < b_j$. x' is in the interior of V_α, and $a_j^T x' > b_j$. By the convexity of V_α, every point x on the line segment between $\bar{x}$ and x' is in the interior of V_α, and therefore the point $(x, x^T x)$ violates the cut for any such x. But because $a_j^T \bar{x} < b_j$ and $a_j^T x' > b_j$ this line segment must intersect the hyperplane $\{x \mid a_j^T x = b_j\}$ at some point, which we will denote $\hat{x}$. At this point we have $||\hat{x}||^2 - (2\bar{x} - \alpha a_i)^T \hat{x} + ||\bar{x}||^2 - \alpha b_i < 0$ and $a_j^T \hat{x}$, contradicting (3.14).

The proof of the converse is essentially the same as in the previous Lemma and we omit it. ■

Next we show that the inequality in (3.14) can be assumed to be tight, which leads to a necessary condition which must be satisfied by the lifting coefficent $\alpha^*(\bar{x}, \mathcal{S}_{ij})$.

Lemma 3.1.9. *Let $\alpha^* = \alpha^*(\bar{x}, \mathcal{S}_{ij})$ be the largest lifting coefficient giving a lifted inequality valid for $\mathcal{S}_{ij}$. Then α^* satisfies*

$$\left\{ \begin{array}{ll} \textit{minimize:} & ||x||^2 - (2\bar{x} - \alpha^* a_i)^T x + ||\bar{x}||^2 - \alpha^* b_i \\ \textit{subject to:} & a_j^T x = b_j \end{array} \right\} = 0 \tag{3.15}$$

Proof. We have already shown that any valid lifted cut must satisfy (3.14). Suppose $\alpha \geq 0$ is such that the inequality (3.14) is strict. This implies that the ball V_α does not intersect the halfspace $\{x \mid a_j^T x \geq b_j\}$. As both of these sets are closed and convex, there is a hyperplane strictly separating them. For any $\epsilon \in \mathbb{R}$, the center and radius of the ball $V_{\alpha+\epsilon}$ vary continuously with ϵ. So for ϵ positive but small enough, the same hyperplane separates $V_{\alpha+\epsilon}$ from $\{x \mid a_j^T x \geq b_j\}$. Moreover, by Lemma 3.1.5, $\operatorname{int}(V_{\alpha+\epsilon}) \cap \{x \mid a_i^T x = b_i\} = \varnothing$ for any $\epsilon > 0$. Together, these facts imply that the lifted inequality with lifting coefficient $\alpha + \epsilon$ is valid. This shows that α is not as large as possible, and does not give the strongest lifted inequality. ■

As a consequence of Lemma 3.1.9 we see that the lifted inequality generated from $\bar{x}$ with lifting coefficient α^* will support $\mathcal{S}_{ij}$ at *two* points: $(\bar{x}, ||\bar{x}||^2)$ and $(x^*, ||x^*||^2)$, where x^* is the optimal solution to the quadratic program in (3.15).

Given the condition (3.15), we can now derive closed-form solutions for the optimal lifting coefficient α^* as well as the second support point x^*.

Lemma 3.1.10. *The value of $\alpha^* = \alpha^*(\bar{x}, \mathcal{S}_{ij})$ is given by the affine expression*

$$\alpha^* = \frac{-2(b_j - a_j^T \bar{x})}{a_i^T a_j - ||a_i||\,||a_j||}. \tag{3.16}$$

Proof. From Corollary A.1.2, the optimal solution to the quadratic program in (3.15) is given by

$$x^* = \bar{x} + \left(\frac{b_j - a_j^T \bar{x} + \frac{\alpha^*}{2} a_i^T a_j}{||a_j||^2} \right) a_j - \frac{\alpha}{2} a_i \tag{3.17}$$

and results in an objective value of

$$
\begin{aligned}
& ||x^*||^2 - (2\bar{x} - \alpha^* a_i)^T x^* + ||\bar{x}||^2 - \alpha^* b_i \\
= \quad & ||\bar{x}||^2 + \frac{(b_j - a_j^T\bar{x} + \frac{\alpha^*}{2}a_i^T a_j)^2}{||a_j||^2} + \frac{\alpha^{*2}}{4}\,||a_i||^2 + 2\left(\frac{b_j - a_j^T\bar{x} + \frac{\alpha^*}{2}a_i^T a_j}{||a_j||^2}\right) a_j^T\bar{x} - \alpha^* a_i^T\bar{x} \\
& - \alpha\left(\frac{b_j - a_j^T\bar{x} + \frac{\alpha^*}{2}a_i^T a_j}{||a_j||^2}\right) a_i^T a_j - 2\,||\bar{x}||^2 - 2\left(\frac{b_j - a_j^T\bar{x} + \frac{\alpha^*}{2}a_i^T a_j}{||a_j||^2}\right) a_j^T\bar{x} + \alpha^* a_i^T\bar{x} \\
& + \alpha^* a_i^T\bar{x} + \alpha^*\left(\frac{b_j - a_j^T\bar{x} + \frac{\alpha^*}{2}a_i^T a_j}{||a_j||^2}\right) a_i^T a_j - \frac{\alpha^{*2}}{2}\,||a_i||^2 + ||\bar{x}||^2 - \alpha^* b_i \\
= \quad & \frac{1}{4}\left(\frac{(a_i^T a_j)^2 - ||a_i||^2\,||a_j||^2}{||a_j||^2}\right)\alpha^{*2} + \frac{a_i^T a_j}{||a_j||^2}(b_j - a_j^T\bar{x})\alpha^* + \frac{(b_j - a_j^T\bar{x})^2}{||a_j||^2}
\end{aligned}
\tag{3.18}
$$

We now set the resulting objective value equal to 0 as in (3.15) and multiply by $||a_j||^2$ to obtain a quadratic equation which α^* must satisfy:

$$
\frac{1}{4}\left((a_i^T a_j)^2 - ||a_i||^2\,||a_j||^2\right)\alpha^{*2} + a_i^T a_j(b_j - a_j^T\bar{x})\alpha^* + (b_j - a_j^T\bar{x})^2 = 0 \tag{3.19}
$$

The coefficient corresponding to α^{*2} in this equation is nonpositive. In the case that it is strictly negative, the quadratic equation has one positive and one negative root. The positive root, which gives us the strongest valid lifted cut, is given by:

$$
\alpha^* = \frac{-2(b_j - a_j^T\bar{x})}{a_i^T a_j - ||a_i||\,||a_j||}. \tag{3.20}
$$

In the case that a_i and a_j are parallel, the coefficient in (3.19) corresponding to α^{*2} is 0, and the sole root to (3.19) is given by

$$
\alpha^* = \frac{-(b_j - a_j^T\bar{x})}{a_i^T a_j}
$$

By the assumption that each inequality in $Ax \leq b$ is facet-defining, we must have that a_i is a *negative* multiple of a_j in the case that the two vectors are parallel. Assume $a_i = -\beta a_j$ for some positive scalar β. Then $a_i^T a_j - ||a_i||\,||a_j|| = -2\beta\,||a_j||^2$. Also $a_i^T a_j = -\beta\,||a_j||^2$, and so we obtain

$$
\alpha^* = \frac{-(b_j - a_j^T\bar{x})}{a_i^T a_j} = \frac{-(b_j - a_j^T\bar{x})}{-\beta\,||a_j||^2} = \frac{-2(b_j - a_j^T\bar{x})}{a_i^T a_j - ||a_i||\,||a_j||},
$$

the same value as (3.20). ■

The lifting coefficient we have derived here gives a cut which is valid for the set $\mathcal{S}_{ij}$, but this cut will not in general be valid for $\mathcal{S}$. There may be feasible points on other facets of P, or outside of

P, where the lifted inequality with lifting coefficient α^* is violated. However, we have shown that any $\alpha \in [0, \alpha^*]$ gives a lifted inequality valid for $\mathcal{S}_{ij}$. To get a lifted cut valid over the entirety of $\mathcal{S}$, we can take the *smallest* lifting coefficient obtained by considering each of the other facets of P individually. Specifically, we obtain the following.

Theorem 3.1.11. *The lifting coefficient giving the strongest lifted cut from a point $\bar{x}$ in the i^{th} facet of P which is valid over $\mathcal{S}$ is given by*

$$\alpha^*(\bar{x}, \mathcal{S}) = \min_{j \neq i} \left\{ \frac{-2(b_j - a_j^T \bar{x})}{a_i^T a_j - ||a_i||\,||a_j||} \right\}. \tag{3.21}$$

The following example illustrates a very simple case.

Example 3.1.12. *Let $n = 1$ and $\mathcal{F} = \mathbb{R} \setminus (-1, 2)$. Figure 3.1 shows the epigraph of $f(x) = x^2$ shaded in grey, with $\mathcal{S}$ overlaid in dark blue. The tangent inequality $z \geq -2x - 1$, which supports $\mathcal{S}$ at $(-1, 1)$ is shown as well.*

Figure 3.1: Graphical depiction of the set $\mathcal{S}$ with a tangent inequality.

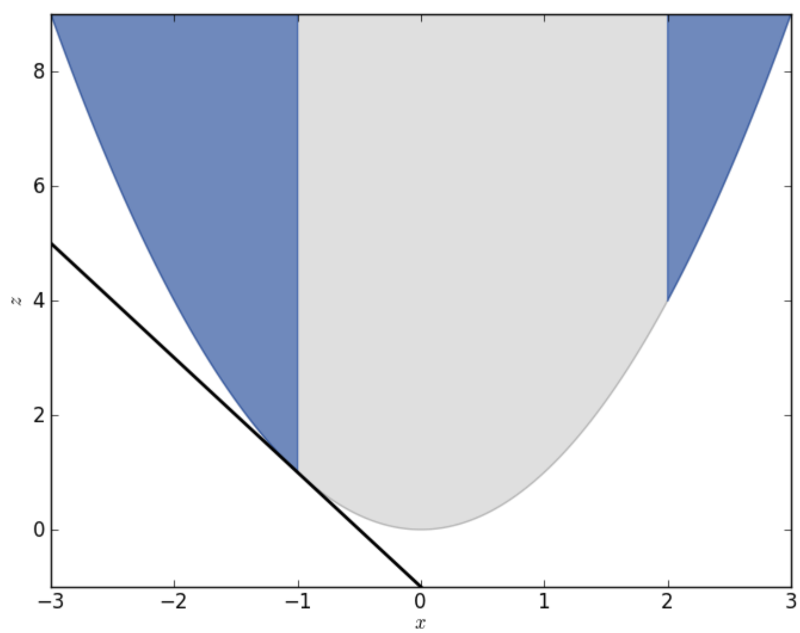
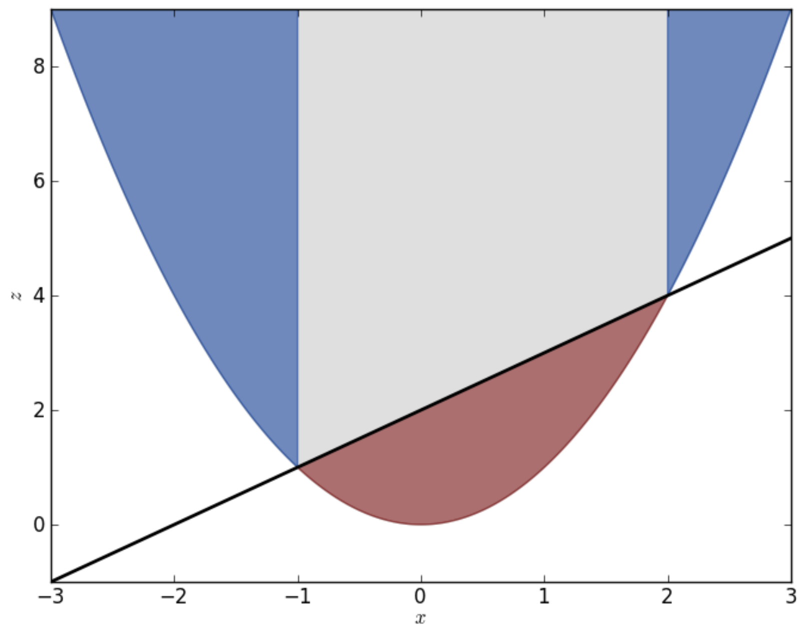


Figure 3.2 depicts the strongest lifted inequality generated from -1, which is given by $z \geq x + 2$. The region of the epigraph that this inequality cuts off is shaded in red. We see that the lifted cut supports $\mathcal{S}$ at two points: (-1, 1) and (2, 4). In this example,

$$\text{conv}\,(\mathcal{S}) = \{(x, z) \mid z \geq x^2,\ z \geq x + 2\},$$

and obtaining the convex hull requires only a single cut (in addition to the $z \geq x^2$ constraint). In general this will not be the case - infinitely many cuts will be required.

Figure 3.2: Illustrating the lifted cut.



3.1.3 Solving the Separation Problem

We now show how to use this expression for the lifting coefficient to derive a polynomial time procedure for separating from the set $\mathcal{S}$. Assume that we have a point $(\hat{x}, \hat{z})$ which we have obtained as the optimal solution to a relaxed problem, with $A\hat{x} < b$. Let $i \in \{1, \ldots, m\}$ be a fixed index. Our goal is to find the valid lifted inequality originating from a point on the i^{th} facet of P which gives the largest cut value. This can be obtained by solving the following quadratic program with variables $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}_+$:

$$
\begin{aligned}
\text{maximize:} \quad & -||x||^2 + 2\hat{x}^T x + \alpha(b_i - a_i^T \hat{x}) \\
\text{subject to:} \quad & a_i^T x = b_i \\
& a_j^T x \leq b_j \quad \forall j \neq i \\
& \alpha \leq \tfrac{-2(b_j - a_j^T x)}{a_i^T a_j - ||a_i|| ||a_j||} \quad \forall j \neq i \\
& \alpha \geq 0
\end{aligned}
$$

This problem can be simplified further by noticing that the constraints $a_j^T x \leq b_j$ are redundant. At a point with $a_j^T x > b_j$ for some $j \neq i$, the constraint

$$\alpha \leq \frac{-2(b_j - a_j^T x)}{a_i^T a_j - ||a_i||\,||a_j||}$$

would imply $\alpha < 0$, which contradicts the nonnegativity constraint on α. Therefore we can eliminate the constraints $a_j^T x \leq b_j$, resulting in the simpler problem:

$$\begin{aligned} \text{maximize:} \quad & -||x||^2 + 2\hat{x}^T x + \alpha(b_i - a_i^T \hat{x}) \\ \text{subject to:} \quad & a_i^T x = b_i \\ & \alpha \leq \tfrac{-2(b_j - a_j^T x)}{a_i^T a_j - ||a_i||||a_j||} \quad \forall j \neq i \\ & \alpha \geq 0 \end{aligned} \qquad (\textbf{SEP(i)})$$

Solving this problem for each $i \in \{1, \ldots, m\}$ gives valid cuts originating from points on each facet of P, some of which will in fact be identical. The strongest cut overall is simply the one with the largest objective value. Note that $\hat{z}$, the function value from the relaxation, does not appear in the formulation. Separation of $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$ is achieved if and only if the strongest cut results in a value greater than $\hat{z}$, which we now state formally:

Theorem 3.1.13. *Let $(\hat{x}, \hat{z}) \in \mathbb{R}^{n+1}$ with $\hat{x} \in \operatorname{int}(P)$. Let z_i^* be the optimal objective value for the problem* SEP(i)*. Then $(\hat{x}, \hat{z}) \in conv(\mathcal{S})$ if and only if*

$$\max_{1 \leq i \leq m} \{z_i^*\} \leq \hat{z}.$$

3.1.4 A Geometric Derivation of the Lifting Coefficient

This section presents an alternative derivation of the affine expression for the maximum lifting coefficient α^*. The alternative derivation exposes more of the geometry underlying the validity of lifted inequalities. Moreover, it was the original method used by the author to derive the maximum lifting coefficient.

As in Section 3.1.2, we consider two fixed indices $i, j \in \{1, \ldots, m\}$, and define the sets W_{ij} and $\mathcal{S}_{ij}$ in the same way. We again form the lifted inequality from a point $\bar{x}$ with $a_i^T \bar{x} = b_i$ and $a_j^T \bar{x} < b_j$: an inequality of the form

$$z \geq (2\bar{x} - \alpha a_i)^T (x - \bar{x}) + ||\bar{x}||^2 . \qquad (3.22)$$

In this derivation we assume that a_i and a_j are not parallel.

Define the two-dimensional plane

$$\begin{aligned}\Pi &= \{x \mid x = \bar{x} + v,\ v^T u = 0 \text{ for all } u \text{ where } a_i^T u = 0 \text{ and } a_j^T u = 0\} \\ &= \{x \mid x = \bar{x} + \eta a_i + \theta a_j,\ \text{where } \eta, \theta \in \mathbb{R}\}\end{aligned}$$

Let

$$\alpha^* = \max\{\alpha \in \mathbb{R} \mid \text{the lifted inequality (3.22) is valid for } \mathcal{S}_{ij}\} \tag{3.23}$$

and

$$\alpha' = \max\{\alpha \in \mathbb{R} \mid \text{the lifted inequality (3.22) is valid for } \mathcal{S}_{ij} \cap \Pi\} \tag{3.24}$$

Lemma 3.1.14. *$\alpha^* = \alpha'$.*

Proof. Clearly we have $\alpha' \geq \alpha^*$, as the maximum in (3.23) is taken over a smaller set than in (3.24). So suppose $\alpha' > \alpha^*$, and there is therefore some point w with $a_j^T w \geq b_j$ and

$$||w||^2 < (2\bar{x} + \alpha' a_i)^T (w - \bar{x}) + ||\bar{x}||^2 .$$

Let $\hat{w}$ be the projection of w onto Π. Then we have

$$||\bar{x} - w||^2 = ||\bar{x} - \hat{w}||^2 + ||w - \hat{w}||^2$$

which gives

$$\begin{aligned}||w||^2 < (2\bar{x} + \alpha' a_i)^T (w - \bar{x}) + ||\bar{x}||^2 \quad &\Leftrightarrow \quad ||\bar{x} - w||^2 < \alpha' a_i^T (w - \bar{x}) \\ &\Leftrightarrow \quad ||\bar{x} - \hat{w}||^2 + ||w - \hat{w}||^2 < \alpha' a_i^T (w - \bar{x}) \\ &\Rightarrow \quad ||\bar{x} - \hat{w}||^2 < \alpha' a_i^T (w - \bar{x}) \\ &\Leftrightarrow \quad ||\hat{w}||^2 < (2\bar{x} + \alpha' a_i)^T (\hat{w} - \bar{x}) + ||\bar{x}||^2\end{aligned}$$

which shows that $(\hat{w}, \hat{w}^T \hat{w})$ violates (3.22) with $\alpha = \alpha'$. This contradicts the definition of α'. ∎

Now suppose some point $(w, w^T w)$ with $a_j^T w > b_j$ violates (3.22). Define V_α as in the previous section. The point $\bar{x}$ is on the boundary of V_α, and $a_j^T \bar{x} < b_j$. Because V_α is convex and contains both w and $\bar{x}$, every point x on the line segment between $\bar{x}$ and w is in V_α as well. This means that there is some point $\hat{x}$ with $a_j^T \hat{x} = b_j$ and for which the point $(\hat{x}, \hat{x}^T \hat{x})$ violates the lifted inequality.

This, together with the previous lemma, shows that we only need to ensure validity on the ray $R = \{x \mid a_j^T x = b_j, a_i^T x \leq b_i\} \cap \Pi$ to ensure validity over $\mathcal{S}_{ij}$.

Continuing, let the point p be the projection of $\bar{x}$ onto $\{x \mid a_i^T x = b_i,\ a_j^T x = b_j\}$. p is at the end of the ray R. We can write any point in $x \in R$ as $x = p + \beta\delta$ where δ is the *unique* unit vector satisfying

$$\delta^T a_j = 0,\ \delta^T a_i \leq 0,\ \text{and } \delta^T u = 0 \text{ for all } u \text{ where } a_i^T u = 0 \text{ and } a_j^T u = 0.$$

The right hand side value of the cut (3.22) at the point x is

$$(2\bar{x} - \alpha a_i)^T (x - \bar{x}) + ||\bar{x}||^2 ,$$

and we want to find a point $x^* = p + \beta^* \delta$ where the hyperplane defined by the cut supports $\mathcal{S}_{ij}$. Such a point x^* will solve the quadratic equation

$$(2\bar{x} - \alpha a_i)^T (x - \bar{x}) + ||\bar{x}||^2 = ||x^*||^2 .$$

Substituting $p + \beta^*\delta$ for x^* gives a quadratic equation with roots given by

$$\beta^* = \frac{1}{2}\left((2(p - \bar{x}) + \alpha a_i)^T \delta \pm \sqrt{\left(\delta^T \left(2(p - \bar{x}) + \alpha a_i\right)\right)^2 - 4\,||\bar{x} - p||^2} \right) \tag{3.25}$$

If the two roots are distinct, then there are points x on the ray R for which the cut value $(2\bar{x} - \alpha a_i)^T (x - \bar{x}) + ||\bar{x}||^2$ is strictly greater than $||x||^2$, meaning the cut is not valid for $\mathcal{S}_{ij}$. So, in order for the cut to be valid, there must only be a single root, or equivalently

$$\left(\delta^T \left(2(p - \bar{x}) + \alpha a_i\right)\right)^2 - 4\,||\bar{x} - p||^2 = 0.$$

This gives a new quadratic equation in α. The roots are given by

$$\alpha = 2\frac{-\delta^T \left(p - \bar{x}\right) \left(a_i^T \delta\right) \pm \sqrt{\left(a_i^T \delta\right)^2 ||p - \bar{x}||^2}}{\left(a_i^T \delta\right)^2}.$$

Because $a_i^T \delta < 0$, the positive root is

$$\alpha = 2\frac{-\delta^T \left(p - \bar{x}\right) \left(a_i^T \delta\right) - \sqrt{\left(a_i^T \delta\right)^2 ||p - \bar{x}||^2}}{\left(a_i^T \delta\right)^2} = 2\frac{-\delta^T (p - \bar{x}) - ||x - p||}{a_i^T \delta}$$

Subsituting this value into (3.25) gives

$$\begin{aligned}\beta^* &= -\frac{1}{2}\big(2(p-\bar{x})+\alpha a_i\big)^T\delta \\ &= -\frac{1}{2}\left[2(p-\bar{x})^T\delta + 2\left(-\delta^T(p-\bar{x}) - ||\bar{x}-p||\right)\right] \\ &= ||\bar{x}-p||\end{aligned}$$

This gives the intuitively clear fact that the origination point $\bar{x}$ and the point x^* are equidistant from the intersection of the two hyperplanes $\{x \mid a_i^T x = b_i\}$ and $\{x \mid a_j^T x = b_j\}$. With this fact, we can derive the lifting coefficient α. As the lifted inequality (3.22) holds with equality at $(x^*, ||x^*||^2)$, we have

$$\begin{aligned}&||x^*||^2 = (2\bar{x}-\alpha a_i)^T(x^*-\bar{x}) + ||\bar{x}||^2 \\ \Rightarrow\quad &\alpha = \frac{||\bar{x}-x^*||^2}{b_i - a_i^T x^*}.\end{aligned} \tag{3.26}$$

Using Corollary A.1.2, we can obtain a solution to the quadratic program which finds p, the projection of $\bar{x}$ onto the intersection of the two hyperplanes, in closed form to get

$$||\bar{x}-p||^2 = \frac{(b_j - a_j^T\bar{x})^2\,||a_i||^2}{||a_i||^2\,||a_j||^2 - (a_i^T a_j)^2}. \tag{3.27}$$

The points $\bar{x}$, x^*, and p all lie in the plane Π. Also, the vectors a_i and a_j are parallel to the surface of Π, so we can work entirely in this two-dimensional plane.

Let θ be the angle between a_i and a_j, and $\phi = \pi - \theta$ be the angle between $\bar{x}-p$ and x^*-p. Using the law of cosines and (3.27) along with the fact that $\bar{x}$ and x^* are equidistant from p, we get

$$\begin{aligned}||\bar{x}-x^*||^2 &= 2\,||p-\bar{x}||^2\,(1-\cos(\phi)) \\ &= 2\,||p-\bar{x}||^2\,(1-\cos(\pi-\theta)) \\ &= 2\frac{(b_j - a_j^T\bar{x})^2\,||a_i||^2}{||a_i||^2\,||a_j||^2 - (a_i^T a_j)^2}(1+\cos(\theta)) \\ &= 2\frac{(b_j - a_j^T\bar{x})^2\,||a_i||^2}{||a_i||^2\,||a_j||^2 - (a_j^T a_i)^2}\left(1 + \frac{a_i^T a_j}{||a_i||\,||a_j||}\right) \\ &= 2\frac{(b_j - a_j^T\bar{x})^2\,||a_i||}{||a_j||\,(||a_i||\,||a_j|| - a_i^T a_j)}.\end{aligned} \tag{3.28}$$

Next, we can see that $\frac{\pi}{2} - \frac{\theta}{2}$ is the angle between a_i and $\bar{x} - x^*$. Then we have

$$\begin{aligned} a_i^T(\bar{x} - x^*) &= ||a_i||\,||\bar{x} - x^*||\cos\left(\frac{\pi}{2} - \frac{\theta}{2}\right) \\ &= ||a_i||\,||\bar{x} - x^*||\sin\left(\frac{\theta}{2}\right). \end{aligned} \tag{3.29}$$

Finally we combine (3.26), (3.27), and (3.29) to get

$$\begin{aligned} \alpha &= \frac{||\bar{x} - x^*||^2}{a_i^T(\bar{x} - x^*)} \\ &= \frac{||\bar{x} - \hat{x}_j||^2}{||a_i||\,||\bar{x} - \hat{x}_j||\sin\left(\frac{\theta_j}{2}\right)} \\ &= \frac{||\bar{x} - x^*||}{||a_i||\sin\left(\frac{\theta}{2}\right)} \\ &= \frac{1}{||a_i||\sin\left(\frac{\theta}{2}\right)}\left(\frac{2(b_j - a_j^T\bar{x})^2\,||a_i||}{||a_j||\,(||a_i||\,||a_j|| - a_i^T a_j)}\right)^{\frac{1}{2}} \\ &= \frac{(b_j - a_j^T\bar{x})}{\sin\left(\frac{\theta}{2}\right)}\left(\frac{2}{||a_i||\,||a_j||\,(||a_i||\,||a_j|| - a_i^T a_j)}\right)^{\frac{1}{2}} \end{aligned}$$

Use of the trigonometric half-angle identity

$$2\sin^2\left(\frac{\theta}{2}\right) = 1 - \cos(\theta)$$

confirms that this expression agrees with (3.20).

3.1.5 The Disjunctive Approach

In this section we discuss the application of the disjunctive programming formulation of Ceria and Soares (see [26]) in the case where the feasible region is the complement of a polyhedron. We demonstrate how this approach can be used to obtain linear inequalities for separating from $\operatorname{conv}(\mathcal{S})$, and discuss the theoretical complexity advantage of the method derived in Section 3.1.3 over this disjunctive approach.

We assume again that we have a point $(\hat{x}, \hat{z})$ with $\hat{x} \in \operatorname{int}(P)$, and we wish to separate $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$, if possible. Writing

$$\mathcal{S} = \bigcup_{i=1}^{m}\left\{(x, z) \in \mathbb{R}^{n+1} \mid a_i^T x \geq b_i, z \geq x^T x\right\},$$

we have $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ if and only if there exists a set of vectors

$$\{(x_i, z_i, \theta_i)\}_{i=1}^m \subset \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$$

satisfying the following system:

$$(\hat{x}, \hat{z}) = \sum_{i=1}^m \theta_i (x_i, z_i) \tag{3.30a}$$

$$a_i^T x_i - b_i \geq 0 \quad i \in \{1, \ldots, m\} \tag{3.30b}$$

$$||x_i||^2 \leq z_i, \quad i \in \{1, \ldots, m\} \tag{3.30c}$$

$$\sum_{i=1}^m \theta_i = 1, \quad \theta_i \geq 0, \quad i \in \{1, \ldots, m\} \tag{3.30d}$$

Unfortunately this system contains nonlinear equalities, due to the the products $\theta_i(x_i, z_i)$ in (3.30a), and could not be used in a convex optimization formulation. Our next result shows that, through a change of variables, we can answer the question of whether or not $(\hat{x}, \hat{z})$ is in $\operatorname{conv}(\mathcal{S})$ by solving a convex feasibility system.

Lemma 3.1.15. *$(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ if and only if there is a set of vectors $\{(y_i, w_i, \lambda_i)\}_{i=1}^m$ following system:*

$$(\hat{x}, \hat{z}) = \sum_{i=1}^m (y_i, w_i) \tag{3.31a}$$

$$a_i^T y_i - \lambda_i b_i \geq 0 \quad i \in \{1, \ldots, m\} \tag{3.31b}$$

$$||(2y_i, w_i - \lambda_i)|| \leq \lambda_i + w_i \quad i \in \{1, \ldots, m\} \tag{3.31c}$$

$$\sum_{i=1}^m \lambda_i = 1, \quad \lambda_i \geq 0, \quad w_i \geq 0, \quad i \in \{1, \ldots, m\} \tag{3.31d}$$

Proof. First suppose $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$. Then there is a set of vectors $\{(x_i, z_i, \theta_i)\}_{i=1}^m$ satisfying (3.30). Let $\lambda_i = \theta_i$, $y_i = \lambda_i x_i$, and $w_i = \lambda_i z_i$ for each i. Then $\{(y_i, w_i, \lambda_i)\}_{i=1}^m$ clearly satisfy (3.31a), (3.31b), and (3.31d). We now only need to show that these vectors satisfy (3.31c). Let i be fixed. If $\lambda_i = 0$, then $y_i = 0$ and $w_i = 0$ and (3.31c) holds for this i. So suppose $\lambda_i > 0$. Notice that because $w_i \geq 0$ and $\lambda_i \geq 0$, we have

$$\begin{aligned} ||(2y_i, w_i - \lambda_i)|| \leq \lambda_i + w_i &\Leftrightarrow ||(2y_i, w_i - \lambda_i)||^2 \leq (\lambda_i + w_i)^2 \\ &\Leftrightarrow 4\,||y_i||^2 + w_i^2 - 2\lambda_i w_i + \lambda_i^2 \leq w_i^2 + 2\lambda_i w_i + \lambda_i^2 \\ &\Leftrightarrow 4\,||y_i||^2 \leq 4\lambda_i w_i \\ &\Leftrightarrow \lambda_i^2\,||x_i||^2 \leq \lambda_i^2 z_i \\ &\Leftrightarrow ||x_i||^2 \leq z_i \end{aligned}$$

and therefore (3.31c) is satisfied.

Now suppose that $\{(y_i, w_i, \lambda_i)\}_{i=1}^m$ satisfy (3.31). If $\lambda_i = 0$, then by squaring both sides of (3.31c) (which is valid as both sides are nonnegative), we have

$$4\,||y_i||^2 + w_i^2 \leq w_i^2$$

which implies $y_i = 0$. For i with $\lambda_i = 0$, let $(x_i, z_i, \theta_i) = (0, 0, 0) \in \mathbb{R}^{n+2}$. For i with $\lambda_i > 0$, define $x_i = y_i/\lambda_i$, $z_i = w_i/\lambda_i$ and $\theta_i = \lambda_i$. Then $\{(x_i, z_i, \theta_i)\}_{i=1}^m$ satisfy (3.30b)-(3.30d), and

$$\begin{aligned} (\hat{x}, \hat{z}) &= \sum_{i=1}^m (y_i, w_i) \\ &= \sum_{\{i|\lambda_i>0\}} \theta_i(x_i, z_i) + \sum_{\{i|\lambda_i=0\}} (0, w_i). \end{aligned} \tag{3.32}$$

The first expression on the right hand side of (3.32) is in $\operatorname{conv}(\mathcal{S})$, and as $w_i \geq 0$ for all i, the second term is in the recession cone of $\mathcal{S}$. Thus we have $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ as well. ■

If $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$, then by the definition of $\mathcal{S}$ we will have $(\hat{x}, \hat{z} + t) \in \operatorname{conv}(\mathcal{S})$ for any $t \geq 0$. However, we do not know for which $t \geq 0$ we will have $(\hat{x}, \hat{z} - t) \in \operatorname{conv}(\mathcal{S})$. We now show how we can answer this question and provide a separation procedure by imposing the system (3.31) as constraints in an optimization problem.

Consider the primal problem below, which we denote $\mathrm{P}(\hat{x})$:

$$\begin{aligned} \text{minimize:} \quad & w_1 + \ldots + w_m \\ \text{subject to:} \quad & y_1 + \ldots + y_m = \hat{x} \\ & a_i^T y_i - \lambda_i b_i \geq 0, \quad i \in \{1, \ldots, m\} \\ & ||(2y_i, w_i - \lambda_i)|| \leq \lambda_i + w_i \quad i \in \{1, \ldots, m\} \\ & \lambda_1 + \ldots + \lambda_m = 1 \\ & \lambda \geq 0,\ w \geq 0 \end{aligned} \tag{P($\hat{x}$)}$$

The constraints of P($\hat{x}$) are identical to the system (3.31), except for the omission of the constraint

$$\sum_{i=1}^{m} w_i = \hat{z}.$$

Instead of simply seeking a representation of $(\hat{x}, \hat{z})$ as a convex combination of points in $\mathcal{S}$, this problem seeks (after reversing the change of variables) the *smallest* value z where $(\hat{x}, z) \in \operatorname{conv}(\mathcal{S})$. Furthermore, as we now show, it will allow us to separate $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$ when possible by means of a linear inequality obtained from the dual problem.

Lemma 3.1.16. *Let z^* be the optimal value of P($\hat{x}$). Then $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ if and only if $z^* \leq \hat{z}$.*

Proof. First suppose that $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$. Then by Lemma 3.1.15, there exists a set of vectors $\{(y_i, w_i, \lambda_i)\}_{i=1}^m$ satisfying the system (3.31). These vectors provide a feasible solution for P($\hat{x}$), with an objective value of $\hat{z}$. Because this solution is feasible but not necessarily optimal, we have $z^* \leq \hat{z}$.

Conversely, suppose that $\{(y_i^*, w_i^*, \lambda_i^*)\}_{i=1}^m$ is an optimal solution for P($\hat{x}$), with an objective value of $z^* \leq \hat{z}$.

For i where $\lambda_i^* = 0$, we must have $y_i^* = 0$. Then the constraint

$$||(2y_i, w_i - \lambda_i)|| \leq \lambda_i + w_i$$

would be satisfied by any nonnegative w_i, but due to the form of the objective, we must have $w_i^* = 0$ in an optimal solution. Thus we can ignore all indices i where $\lambda_i^* = 0$.

For indices i where $\lambda_i^* > 0$, define $x_i = y_i^*/\lambda_i^*$ and $z_i = w_i^*/\lambda_i^*$. We then have $(x_i, z_i) \in \mathcal{S}$ for each i,

$$\sum_{\{i|\lambda_i^*>0\}} \lambda_i^* x_i = \hat{x}, \quad \text{and} \quad \sum_{\{i|\lambda_i^*>0\}} \lambda_i^* z_i = z^*,$$

and so $(\hat{x}, z^*) \in \operatorname{conv}(\mathcal{S})$. Then, because $z^* \leq \hat{z}$ and $(x, z + \epsilon) \in \mathcal{S}$ for any $(x, z) \in \mathcal{S}$ and $\epsilon \geq 0$, we have that $(\hat{x}, \hat{z})$ is indeed contained in $\operatorname{conv}(\mathcal{S})$. ■

Note that in the primal problem, the conic constraints will always hold with equality. This can be seen from the equivalence

$$||(2y_i, w_i - \lambda_i)|| \leq \lambda_i + w_i \quad \Leftrightarrow \quad ||y_i||^2 \leq \lambda_i w_i.$$

For fixed values of y_i and λ_i, w_i will be as small as possible in an optimal solution. Thus we will have the intuitively obvious fact $z_i = ||x_i||^2$ for each i.

Next we show how the dual problem of the disjunctive formulation can be used to obtain a linear equality separating $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$, if one exists. First observe that as long as $\hat{x} \in \operatorname{conv}(\mathcal{F})$, the primal problem $\mathrm{P}(\hat{x})$ is feasible and satisfies the Slater condition for strong duality, as the w variables can be made arbitrarily large while preserving feasibility. The objective of the dual problem to $\mathrm{P}(\hat{x})$ (see [46]) is to maximize $\gamma^T \hat{x} + \beta$. γ and β are only a subset of the variables in the dual problem, but they are the only ones appearing in the objective. Additionally, $\hat{x}$ does not appear in the constraints: the feasible set for the dual of $\mathrm{P}(x)$ is the same for any x. If z^* is the optimal objective value for $\mathrm{P}(\hat{x})$ and (γ^*, β^*) is part of an optimal solution to the dual, then by strong duality we have

$$\gamma^{*T} \hat{x} + \beta^* = z^*.$$

Lemma 3.1.17. *Let $\{(y_i^*, w_i^*, \lambda_i^*)\}_{i=1}^m$ be optimal for $P(\hat{x})$, with an objective value of z^*, and (γ^*, β^*) be part of an optimal solution to the dual problem to $P(\hat{x})$. Then the linear inequality*

$$z \geq \gamma^{*T} x + \beta^*$$

is valid for $\operatorname{conv}(\mathcal{S})$ *and holds with equality at the points $(y_i^*/\lambda_i^*, w_i^*/\lambda_i^*)$ for i with $\lambda_i^* > 0$.*

Proof. Let $(x', z') \in \operatorname{conv}(\mathcal{S})$. Let the primal problem $\mathrm{P}(x')$ have an optimal objective value of $v' \leq z'$. γ^*, β^*, and the other variables in the optimal solution to the dual of $\mathrm{P}(\hat{x})$ are feasible for the dual of $\mathrm{P}(x')$, but of course not necessarily optimal. So we have

$$\gamma^{*T} x' + \beta^* \leq v' \leq z'$$

which shows that the cut holds for (x', z'). As (x', z') was arbitrary this implies that the cut is valid for $\operatorname{conv}(\mathcal{S})$.

Now suppose, without loss of generality, that $\lambda_i^* > 0$ for $i = 1, \ldots, p$ and $\lambda_i^* = 0$ for $i > p$. This implies $w_i^* = 0$ and $y_i^* = 0$ for $i > p$. For $i \leq p$ define $x_i = y_i^*/\lambda_i^*$ and $z_i = w_i^*/\lambda_i^*$. Then we have

$$\begin{aligned}
\gamma^{*T} \hat{x} + \beta^* = \sum_{i=1}^p w_i^* \quad &\Rightarrow \quad \gamma^{*T} \left(\sum_{i=1}^p y_i^* \right) + \beta^* = \sum_{i=1}^p w_i^* \\
&\Rightarrow \quad \sum_{i=1}^p \left(\gamma^{*T} y_i^* - w_i^* \right) = -\beta^* \\
&\Rightarrow \quad \sum_{i=1}^p \lambda_i^* \left(\gamma^{*T} x_i - z_i \right) = -\beta^*
\end{aligned}$$

Then, because the λ_i^*'s are nonnegative and sum to one, and ${\gamma^*}^T x_i - z_i \leq \beta^*$ for each i, we have

$${\gamma^*}^T x_i - z_i = \beta^*$$

for each i: the cut is tight at each point (x_i, z_i). ∎

The disjunctive approach provides a straightforward and easily implementable means of solving the separation problem in this setting. One possible disadvantage is the computational burden for large problems. The disjunctive formulation requires solving a second-order cone program with $\mathrm{O}(mn)$ variables and $\mathrm{O}(m+n)$ constraints, including m conic constraints. Alternatively, our approach requires solving at most m quadratic programs in $\mathrm{O}(n)$ variables and $\mathrm{O}(m)$ linear constraints. For larger problem sizes, our method could provide a significant decrease in computational cost. Morever, for very large problems the disjunctive formulation may be impractical for commercial solvers.

The development of the separation procedure using duality in this section is quite similar to that in [33] and [44], with the difference being that both the problem $\mathrm{P}(\hat{x})$ and its dual are conic (rather than linear) programs, which obviates the need for any normalization constraint in the dual problem.

3.1.6 Computational Considerations

We have shown that the problem of separating a point $(\hat{x}, \hat{z})$ from $\mathrm{conv}(\mathcal{S})$ using lifted inequalities can be achieved by solving m convex quadratic programs. Although each QP is small, this could represent a substantial amount of computation when m is large. As such, we would like to identify some ways in which the computational load may be decreased. In this section, we will provide a theoretical result giving a sufficient condition for optimality in the separation problem, and point out a heuristic for choosing the order of searching through the facets of P for the separating cut.

Let y be a point in the i^{th} facet of P, and define

$$\alpha^*(y) = \alpha^*(y, \mathcal{S}) = \min_{j \neq i} \left\{ \frac{2(b_j - a_j^T y)}{||a_i||\,||a_j|| - a_i^T a_j} \right\}. \tag{3.33}$$

Then the inequality

$$z \geq (2y - \alpha^*(y) a_i)^T (x - y) + ||y||^2 \tag{3.34}$$

is the nondominated lifted inequality generated from y. If y is not in the relative interior of the i^{th} facet, then $(y, ||y||^2)$ is the only point in $\mathcal{S}$ for which this inequality is tight. Otherwise, by Lemma 3.1.9 there is at least one other feasible point y' so that the inequality is tight at $(y', ||y'||^2)$. Any such point y' must lie in the relative interior of some facet of P, and due to the strict convexity of the function $x^T x$, it cannot lie if the i^{th}. In fact, from the derivation of the lifting coefficient in Section 3.1.2 it can be seen that for each j attaining the minimum in (3.33), there is a point y^j in the j^{th} facet of P for which the lifted inequality (3.34) holds with equality at $(y^j, ||y^j||^2)$. This motivates the following definition.

Definition 3.1.18. *For a point $y \in \mathcal{F}$, define $T(y)$ to be the set of all points $y' \in \mathcal{F}$ for which the nondominated lifted inequality generated from y holds with equality at $(y', ||y'||^2)$.*

The set $T(y)$ is defined for any feasible point y, but will only contain more than one point when y is in the relative interior of a facet of P. For any y, we have $1 \leq |T(y)| \leq m$, as $y \in T(y)$ for all y and $T(y)$ can contain at most one point from each facet of P.

The next result establishes a relationship between the sets $T(y)$ and $T(y')$ for distinct points y and y'.

Lemma 3.1.19. *Let y and y' be distinct feasible points. If $y' \in T(y)$, then $T(y) = T(y')$. Otherwise $T(y)$ and $T(y')$ do not intersect.*

Proof. Assume that y and y' are in the i^{th} and j^{th} facets of P, respectively. Let

$$z \geq \left(2y - \alpha^*(y)a_i\right)^T (x - y) + ||y||^2 \tag{3.35}$$

be the nondominated lifted inequality generated at y. As $y' \in T(y)$, we have

$$||y'||^2 = \left(2y - \alpha^*(y)a_i\right)^T (y' - y) + ||y||^2$$

and therefore

$$\alpha^*(y) = \frac{||y - y'||^2}{a_i^T (y - y')}. \tag{3.36}$$

The inequality (3.35) is valid and holds with equality at $(y', ||y'||^2)$, and so by Lemma (3.1.1), we have

$$\begin{aligned} 2y - \alpha^*(y)a_i &= 2y' - \theta a_j && (3.37) \\ \alpha^*(y)b_i - ||y||^2 &= \theta b_j - ||y'||^2 && (3.38) \end{aligned}$$

for some nonnegative scalar θ. By (3.37) we have

$$\left(2y - \alpha^*(y)a_i\right)^T (y' - y) = (2y' - \theta a_j)^T (y' - y)$$

which gives

$$\begin{aligned} \theta a_j^T (y' - y) &= -\left(2y - \alpha^*(y)a_i\right)^T (y' - y) + 2y'^T (y' - y) \\ &= 2\left|\left|y - y'\right|\right|^2 + \alpha^*(y)a_i^T (y' - y) \\ &= 2\left|\left|y - y'\right|\right|^2 - \left|\left|y - y'\right|\right|^2 && \text{by (3.36)} \\ &= \left|\left|y - y'\right|\right|^2 \end{aligned}$$

and so

$$\theta = \frac{||y - y'||^2}{a_j^T (y' - y)}.$$

Let the nondominated lifted inequality generated at y' be given by

$$z \geq \left(2y' - \alpha^*\left(y'\right) a_j\right)(x - y') + \left|\left|y'\right|\right|^2.$$

From the preceding analysis we know that the cut

$$z \geq \left(2y' - \theta a_j\right)^T (x - y') + \left|\left|y'\right|\right|^2.$$

is valid, so $\alpha^*(y') \geq \theta$. Moreover, we know that this cut holds with equality $(y, ||y||^2)$. So if $\alpha^*(y') > \theta$, it would be violated at $(y, ||y||^2)$, implying $\theta = \alpha^*(y')$. This means that the nondominated inequality generated at y' is exactly the same as the nondominated inequality generated at y. Applying this result to every point in $T(y)$ proves the first part of claim, and the second part follows clearly from this. ■

We obtain the following as a corollary:

Corollary 3.1.20. *The relation "$x \in T(y)$" is an equivalence relation whose equivalence classes are* $\{T(x) \mid x \in \mathcal{F}\}$.

Proof. Reflexivity of the relation is clear. Lemma (3.1.19) shows symmetry. For transitivity, suppose $x \in T(y)$ and $y \in T(w)$. Then we have $T(x) = T(y) = T(w)$ by Lemma (3.1.19), and therefore $x \in T(w)$. ■

The last theoretical result in this section gives a sufficient condition for a lifted cut to be the strongest possible at a given point. This gives a justifiable criterion for stopping the search through all m facets to find a separating cut.

Theorem 3.1.21. *Let $\hat{x} \in \operatorname{int}(P)$, and let $\bar{x}$ be a feasible point with $\hat{x} \in conv(T(\bar{x}))$. Then the nondominated lifted cut generated from $\bar{x}$ is the strongest possible lifted cut at $\hat{x}$.*

Proof. Let $T(\bar{x}) = \{v_1, \ldots, v_p\}$, and let

$$z \geq \gamma^T x + \beta$$

be the nondominated lifted cut generated from $\bar{x}$. Then let y be any other feasible point, and let

$$z \geq \delta^T x + \theta$$

be the nondominated lifted cut generated from y. If $T(\bar{x})$ and $T(y)$ intersect, then by Lemma 3.1.19 the two sets are identical and the lifted cuts generated at the points in this set are the same. So we assume that $T(\bar{x})$ and $T(y)$ are disjoint. This implies

$$||v_j||^2 > \delta^T v_j + \theta \quad \text{for all } v_j \in T(\bar{x}). \tag{3.39}$$

The strict inequality comes from the fact that the lifted cut generated from y is valid and $v_j \notin T(y)$.

Now let $\lambda_1, \ldots, \lambda_p$ be the coefficients such that

$$\hat{x} = \sum_{j=1}^{p} \lambda_j v_j, \qquad \sum_{j=1}^{p} \lambda_j = 1, \qquad \text{and } \lambda_j \geq 0 \quad \forall j.$$

Then we have

$$\begin{aligned}
\gamma^T \hat{x} + \beta &= \gamma^T \left(\sum_{j=1}^{p} \lambda_j v_j \right) + \beta \\
&= \sum_{j=1}^{p} \lambda_j \left(\gamma^T v_j + \beta \right) \\
&= \sum_{j=1}^{p} \lambda_j \, ||v_j||^2 && \text{because } v_j \in T(\bar{x}) \\
&> \sum_{j=1}^{p} \lambda_j \left(\delta^T v_j + \theta \right) && \text{by (3.39)} \\
&= \delta^T \hat{x} + \theta.
\end{aligned}$$

Thus we have shown that at $\hat{x}$, the cut generated from $\bar{x}$ is strictly stronger than the cut generated from y. ■

Now suppose we are searching for a lifted cut separating some given point $(\hat{x}, \hat{z})$ with $\hat{x}$ infeasible. We can accomplish this by solving the quadratic program (SEP(i)) from Section 3.1.2 for each $i = 1, \dots, m$. Suppose we solve this problem for some fixed i and the optimal solution is $(\bar{x}, \bar{\alpha})$. Let J be the set of all indices j for which $(\bar{x}, \bar{\alpha})$ satisfy

$$\bar{\alpha} = \frac{-2(b_j - a_j^T \bar{x})}{a_i^T a_j - ||a_i||\,||a_j||}.$$

The facets of P corresponding to the indices in J are the ones containing points in $T(\bar{x})$. Using Equation (3.17), we can compute the points in $T(\bar{x})$ and check if $\hat{x}$ is in their convex hull, which can be accomplished by solving a linear program. If it is, then the cut from the i^{th} facet is the strongest possible, and we can terminate the search.

This result guarantees that we have found the strongest cut when we stop early, but doesn't give any guidance on what order to search through the indices $\{1, \dots, m\}$. One heuristic that we have found useful is to order the facets according to their distance to the point $\hat{x}$. This makes sense intuitively: the first two terms in the objective of (SEP(i)) are equivalent to the negative squared distance between $\hat{x}$ and the point x in the i^{th} facet of P. Moreover, suppose that the i^{th} facet of P is the one closest to $\hat{x}$, and let p be the projection onto this facet. Then the ball with center $\hat{x}$ and radius $\frac{b_i - a_i^T \hat{x}}{||a_i||}$ is contained in P and touches the i^{th} facet of P at p. This implies that the nondominated lifted cut generated from p will be violated by $(\hat{x}, ||\hat{x}||^2)$, although possibly not by $(\hat{x}, \hat{z})$. In practice, we have found that finding the lifted cut from *only* the nearest facet gives good cuts quickly.

3.2 Multiple Polyhedra

We now consider the case where $\mathcal{F}$ is the complement of the interior of the union of multiple polyhedra:

$$\mathcal{F} = \mathbb{R}^n \setminus \left(\bigcup_{k=1}^{K} \{x \mid A^k x < b^k\} \right)$$

where $A^k \in \mathbb{R}^{m_k \times n}$ and $b^k \in \mathbb{R}^{m_k}$ for each k. First we note that we can also express $\mathcal{F}$ as a disjunction. Choosing an index $i_k \in \{1, \dots, m_k\}$ for each $k \in \{1, \dots, K\}$ gives a K-tuple $(i_1, \dots, i_K)$

and a corresponding (possibly empty) polyhedron

$$\mathcal{D}_{(i_1,\ldots,i_K)} = \{x \mid {a^k_{i_k}}^T x \geq b^k_{i_k},\ k \in 1,\ldots,K\}.$$

Here, $a^k_{i_k}$ is the ${i_k}^{th}$ row of the matrix A^k. $\mathcal{F}$ is the union of all such polyhedra. This type of feasible set is encountered in applications: lattice-free sets in mixed integer programming often take the form of a union of polyhedra (see [45], [31], or [20]), and the feasible set in [1] is obtained by excluding the interiors of multiple polytopes from $\mathbb{R}^n$.

As before, we define the set

$$\mathcal{S} = \{(x,z) \in \mathbb{R}^{n+1} \mid z \geq x^T x,\ x \in \mathcal{F}\}.$$

This section introduces a new parameterization of lifted inequalities, then develops validity constraints for this new parameterization and a method for solving the separation problem for $\mathcal{S}$ by means of a single QCQP. We also discuss the relationship between our separation procedure and the disjunctive programming approach.

3.2.1 Applying the Disjunctive Method

As the feasible region $\mathcal{F}$ can be expressed as a finite union of polyhedra, the disjunctive formulation can be applied. Following the construction above, we write

$$\begin{aligned} \mathcal{F} &= \bigcup \{\mathcal{D}_{(i_1,\ldots,i_K)} \mid i_k \in \{1,\ldots,m_k\} \text{ for } i \in \{1,\ldots,K\}\} \\ &\triangleq \bigcup_{j=1}^{D} \{x \mid C_j x \geq d_j\} \end{aligned} \tag{3.40}$$

Here, D is the total number of polyhedra defining the disjunction.

Given a point $(\hat{x},\hat{z})$ with $\hat{x} \notin \mathcal{F}$, the primal problem for the disjunctive formulation is:

$$\begin{aligned} \text{minimize:} \quad & w_1 + \ldots + w_D \\ \text{subject to:} \quad & y_1 + \ldots + y_D = \hat{x} \\ & C_j y_j - \lambda_j d_j \geq 0 \quad \text{for } j \in \{1,\ldots,D\} \\ & ||(2y_j, w_j - \lambda_j)|| \leq \lambda_j + w_j \quad \text{for } j \in \{1,\ldots,D\} \\ & \lambda_1 + \ldots + \lambda_D = 1 \\ & \lambda \geq 0, w \geq 0 \end{aligned} \tag{P($\hat{x}$)}$$

The analysis of the use of this method to separate $(\hat{x}, \hat{z})$ from $\text{conv}(\mathcal{S})$ is identical to that in Section (3.1.5), so we state the following result without proof.

Lemma 3.2.1. *Let z^* be the optimal value of $P(\hat{x})$. Then $(\hat{x}, \hat{z}) \in \text{conv}(\mathcal{S})$ if and only if $z^* \leq \bar{z}$. Moreover, let (γ^*, β^*) be optimal variables for the dual problem of $P(\hat{x})$. Then the linear inequality*

$$\gamma^{*T} x - z \leq -\beta^*$$

is valid for $conv(\mathcal{S})$ and holds with equality at $(\hat{x}, z^*)$.

The computational burden imposed by the disjunctive approach becomes even more severe than in the case of a single excluded polyhedron. The problem $\mathrm{P}(\hat{x})$ is a second-order cone problem with $\mathrm{O}\left(n \prod_{k=1}^{K} m_k\right)$ variables and constraints, which will become impractical for even modest problem sizes. This motivates the search for more efficient methods in this setting.

3.2.2 A Simple Case

We first consider a case with two polyhedra, where the first is defined by two inequalities and the second is a halfspace. This is an extremely simple example, but in trying to devise a separation procedure similar to that in Section 3.1.3 we will encounter an amount of complexity which will motivate alternative approaches. For this case, let the inequalities $a_1^T x \leq b_1$ and $a_2^T x \leq b_2$ define the first polyhedron, and $c^T x \leq d$ define the second. We then have

$$\mathcal{F} = \{x \mid a_1^T x \geq b_1,\ c^T x \geq d\} \bigcup \{x \mid a_2^T x \geq b_2,\ c^T x \geq d\}$$

Starting from a point $\bar{x}$ with $a_1^T \bar{x} = b_1,\ a_2^T \bar{x} \leq b_2$ and $c^T \bar{x} \geq d$, we want to find the strongest valid cut of the form

$$z \geq (2\bar{x} - \alpha a_1)^T (x - \bar{x}) + ||\bar{x}||^2 . \tag{3.41}$$

Specifically, we are in search of the largest value α so that this cut is valid. If we fix any x where $a_1^T x < b_1$, then for α large enough, the inequality (3.41) will be violated at $(x, x^T x)$. So as long as there is some point x' where $a_1^T x' < b_1$, $a_2^T x' \geq b_2$, and $c^T x' \geq d$, the lifting coefficient will be finite. If there is no such x', then $\mathcal{F} = \{x \mid a_1^T x \geq b_1,\ c^T x \geq d\}$, and the cut is valid for any $\alpha \geq 0$. We assume that such a point x' does exist, so the lifting coefficient is finite.

In what follows, we will frequently refer to the set

$$V_\alpha = \{x \mid \ ||x||^2 - (2\bar{x} - \alpha a_1)^T x + ||\bar{x}||^2 - \alpha b_1 \leq 0\},$$

the ball of points x for which $(x, x^T x)$ weakly violates (3.41). Recall from Proposition 3.3 that the inequality (3.41) is valid for $\mathcal{S}$ if and only if $\mathrm{int}\,(V_\alpha) \cap \mathcal{F} = \varnothing$.

Let

$$\alpha' \ = \ \frac{2(b_2 - a_2^T \bar{x})}{||a_1||\,||a_2|| - a_1^T a_2} \tag{3.42}$$

and

$$\begin{aligned} x^* \ &= \ \bar{x} + \left(\frac{b_2 - a_2^T \bar{x} + \frac{\alpha'}{2} a_1^T a_2}{||a_2||^2} \right) a_2 - \frac{\alpha'}{2} a_1 \\ &= \ \bar{x} + \frac{(b_2 - a_2^T \bar{x})}{||a_1||\,||a_2|| - a_1^T a_2} \left(\frac{||a_1||}{||a_2||} a_2 - a_1 \right) \end{aligned} \tag{3.43}$$

be the lifting coefficent and support point derived in Section 3.1.2. Recall that α' was the largest possible lifting coefficient resulting in a lifted inequality valid for

$$\{(x,z) \mid z \geq x^T x,\ a_1^T x \geq b_1\} \cup \{(x,z) \mid z \geq x^T x,\ a_2^T x \geq b_2\},$$

and that this inequality was tight at $(x^*, ||x^*||^2)$. We may find that x^* satisfies $c^T x^* \geq d$ and is therefore feasible. If this is the case, then α' is the largest possible lifting coefficient and we are done. We therefore assume that $c^T x^* < d$. As a consequence, we obtain the following.

Proposition 3.2.2. *Assume that x^* as defined in (3.43) satisfies $c^T x^* < d$, and define*

$$\alpha^* = \sup\{\alpha \mid \textit{inequality } (3.41) \textit{ is valid}\}.$$

Then $\alpha^* > \alpha'$.

Proof. The ball $V_{\alpha'}$ intersects

$$\{x \mid a_1^T x \geq b_1\} \cup \{x \mid a_2^T x \geq b_2\}$$

only at the points x^* and $\bar{x}$. Because $c^T x^* < d$, there is a ball $\mathcal{B}(x^*, r)$ centered at x^* with positive radius r such that $c^T x < d$ for all $x \in \mathcal{B}(x^*, r)$. For $\epsilon > 0$ and small enough $V_{\alpha'+\epsilon} \cap \{x \mid a_2^T x \geq b_2\}$ is contained in $\mathcal{B}(x^*, r)$. Moreover, by Lemma 3.1.5,

$$\mathrm{int}\,(V_\alpha) \cap \{x \mid a_1^T x \geq b_1\} = \varnothing$$

for any $\alpha \geq 0$. Therefore, for small but positive ϵ, $\operatorname{int}(V_{\alpha'+\epsilon}) \cap \mathcal{F} = \varnothing$, and therefore the cut (3.41) with $\alpha = \alpha' + \epsilon$ is valid, and is stronger than the cut with $\alpha = \alpha'$. This finishes the proof. ■

The next example should help to illuminate the current setting.

Example 3.2.3. *Let*

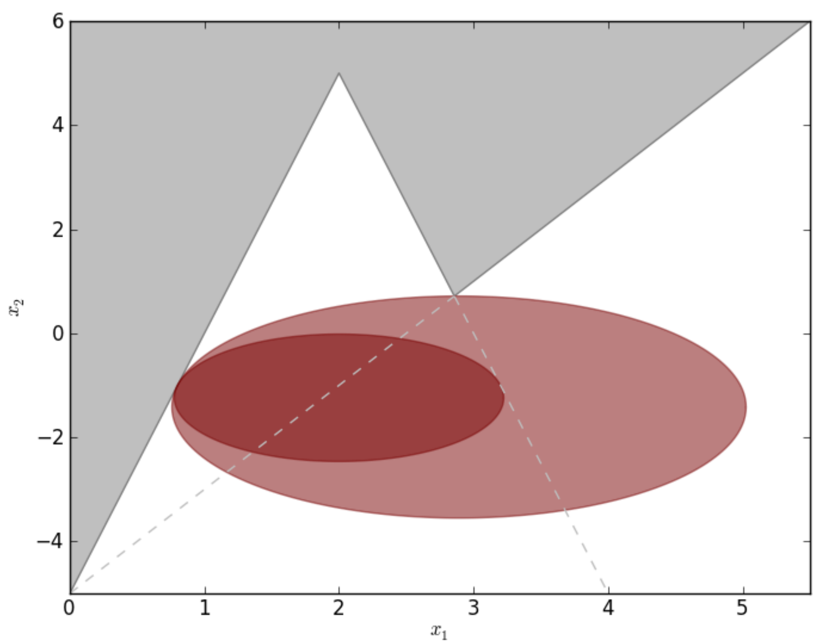
$$A = \begin{bmatrix} -5 & 1 \\ 5 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 15 \\ -5 \end{bmatrix}, \quad c = \begin{bmatrix} -2 & 1 \end{bmatrix}, \quad \textit{and} \quad d = -5.$$

Suppose we want to compute the nondominated lifted inequality from $\bar{x} = (0.8, -1.0)$. Ignoring the new constraint $c^T x \geq d$ and using equations (3.42) and (3.43), we obtain

$$\alpha' = 0.48, \quad \textit{and} \quad x^* = \left(\frac{16}{5}, -1.0\right),$$

and we have $c^T x^* = -7.4 < d$. Thus the lifting coefficient for the nondominated lifted cut generated from $\bar{x}$ is strictly larger than α'. It can be shown that the largest possible lifting coefficient is $\alpha^* \approx .8366$. Figure 3.3 illustrates this example.

Figure 3.3: A case where the lifting coefficient can be increased due to a new linear inequality.



In this figure, $\mathcal{F}$ is shaded in grey. The projected cut-off regions from two different lifted cuts generated from $\bar{x}$ are shaded in red. The cut with lifting coefficient α' cuts off the darker inner *region, which intersects $\mathcal{F}$ at only one point. The cut with lifting coefficient α^* cuts off the lighter outer region, which contains the darker region and intersects $\mathcal{F}$ at two points: $\bar{x}$ and $\left(\frac{20}{7}, \frac{5}{7}\right)$.*

We point out that we can express the condition $c^T x^* < d$ as a linear inequality:

$$c^T x^* < d \;\Leftrightarrow\; c^T \bar{x} + \frac{(b_2 - a_2^T \bar{x})}{||a_1||\,||a_2|| - a_1^T a_2} c^T \left(\frac{||a_1||}{||a_2||} a_2 - a_1 \right) < d. \tag{3.44}$$

Following the derivation in Section 3.1, we can characterize the lifting coefficient α which gives the strongest lifted cut through the optimal value of a quadratic program.

Lemma 3.2.4. *The strongest valid lifted cut (3.41) is given by the largest α satisfying*

$$\left\{ \begin{array}{ll} \textit{minimize:} & ||x||^2 - (2\bar{x} - \alpha a_1)^T x + ||\bar{x}||^2 - \alpha b_1 \\ \textit{subject to:} & a_2^T x \geq b_2 \\ & c^T x \geq d \end{array} \right\} = 0. \tag{3.45}$$

Proof. Let v^* be the optimal value of the minimization problem in (3.45). $v^* \geq 0$ is necessary and sufficient to ensure the lifted cut (3.41) is valid. Suppose $v^* > 0$; we have

$$||\bar{x} - x||^2 + \alpha(a_1^T x - b_1) > 0 \quad \text{for all } x \text{ with } a_2^T x \geq b_2 \text{ and } c^T x \geq d.$$

This statement is equivalent to

$$V_\alpha \cap \{x \mid a_2^T x \geq b_2,\ c^T x \geq d\} = \varnothing.$$

Both of these sets are closed and convex, so this implies that there is a hyperplane strictly separating them. Then, because the center and radius of V_α vary continuously with α, the same hyperplane strictly separates $V_{\alpha+\epsilon}$ for small but positive ϵ. The cut (3.41) is valid for all $(x, z) \in \mathcal{S}$ with $a_1^T x \geq b_1$, for any nonnegative choice of the lifting coefficient, so this implies that the cut is valid with lifting coefficent $\alpha + \epsilon$, and therefore α does not give the strongest possible cut. ■

As in Section 3.1, we will show that we can tighten the constraints in (3.45) to equalities, which will allow us the solve the quadratic program in closed form and obtain an expression for the optimal α.

Lemma 3.2.5. *The strongest valid lifted cut (3.41) is given by the α satisfying*

$$\left\{ \begin{array}{ll} \textit{minimize:} & ||x||^2 - (2\bar{x} - \alpha a_1)^T x + ||\bar{x}||^2 - \alpha b_1 \\ \textit{subject to:} & a_2^T x = b_2 \\ & c^T x = d \end{array} \right\} = 0. \tag{3.46}$$

Proof. Suppose we have found an α satisfying (3.45). Let x' be optimal for the minimization problem. First suppose that $c^T x' > d$. This means that for this α, we could solve the optimization problem without the $c^T x \geq d$ constraint and still get the same optimal solution. But this is the same problem as in Lemma 3.1.6. Thus the α satisfying (3.45) would be exactly the same as in (3.20), meaning $x' = x^*$. As we assumed $c^T x^* < d$, this is a contradiction, implying $c^T x' = d$.

Next suppose that $a_2^T x' > b_2$. The ball V_α has $\bar{x}$ and x' on its boundary, and contains x^*. If $a_2^T x' > b_2$, then the set $\{x \mid c^T x = d\}$ must be a supporting hyperplane for V_α at x'. If not, then there is some point x' with

$$||y||^2 - (2\bar{x} - \alpha a_1)^T x' + ||\bar{x}||^2 - \alpha b_1 < 0$$
$$a_2^T y > b_2$$
$$c^T y > d$$

which contradicts the optimality of x'. This implies that $c^T x \geq d$ for all $x \in V_\alpha$. However, this leads to a contradiction: as we know $\alpha > \alpha'$, we know V_α contains x^* which would imply $c^T x^* \geq d$. This contradicts our prior assumption that $c^T x^* < d$. ■

So we now face the problem of finding $\alpha \geq 0$ satisfying (3.46). As in Section 3.1, we will derive a closed-form solution to the quadratic program in (3.46) and use the resulting optimal objective value to get an expression for α. To save space, we will represent the two equality constraints in the quadratic program as $Gx = g$. We first derive a quadratic equation which the optimal α must satisfy.

Lemma 3.2.6. *Let α satisfy (3.46). Then α must satisfy the quadratic equation*

$$\begin{aligned}\frac{-\alpha^2}{4} a_1^T \left(I - G^T \left(GG^T\right)^{-1} G\right) a_1 + \alpha \left(a_1^T G^T \left(GG^T\right)^{-1} (g - G\bar{x})\right) \\ + \left|\left|G^T \left(GG^T\right)^{-1} (g - G\bar{x})\right|\right|^2 = 0.\end{aligned} \tag{3.47}$$

Proof. By Corollary (A.1.2), the optimal solution to the quadratic program in (3.46), which we denote by x', is given by:

$$x' = \bar{x} - \frac{\alpha}{2} a_1 + \frac{1}{2} G^T \left(GG^T\right)^{-1} (2g - 2G\bar{x} + \alpha G a_1). \tag{3.48}$$

Now we use this expression for x', piece by piece, to determine the optimal objective value:

$$\begin{aligned}
\left|\left|x'\right|\right|^2 &= \frac{1}{4}(2g^T - 2\bar{x}^T G^T + \alpha a_1^T G^T)\left(GG^T\right)^{-1} GG^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&\quad + \bar{x}^T\bar{x} + \frac{\alpha^2}{4}\left|\left|a_1\right|\right|^2 - \alpha\bar{x}^T a_1 + \bar{x}^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&\quad - \frac{\alpha}{2} a_1^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&= \frac{1}{2} g^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) - \frac{1}{2}\bar{x}^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&\quad + \frac{\alpha}{4} a_1^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) + \bar{x}^T\bar{x} + \frac{\alpha^2}{4}\left|\left|a_1\right|\right|^2 - \alpha b_1 \\
&\quad + \bar{x}^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) - \frac{\alpha}{2} a_1^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&= \frac{1}{2} g^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) + \frac{1}{2}\bar{x}^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&\quad - \frac{\alpha}{4} a_1^T G^T \left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) + \bar{x}^T\bar{x} + \frac{\alpha^2}{4}\left|\left|a_1\right|\right|^2 - \alpha b_1 \\
&= g^T \left(GG^T\right)^{-1} g - \bar{x}^T G^T \left(GG^T\right)^{-1} G\bar{x} + \alpha a_1^T G^T \left(GG^T\right)^{-1} G\bar{x} \\
&\quad - \frac{\alpha^2}{4} a_1^T G^T \left(GG^T\right)^{-1} G a_1 + \bar{x}^T\bar{x} + \frac{\alpha^2}{4}\left|\left|a_1\right|\right|^2 - \alpha b_1
\end{aligned}$$

Next,

$$\begin{aligned}
(2\bar{x} - \alpha a_1)^T x' &= 2\bar{x}^T\left(\bar{x} - \frac{\alpha}{2}a_1 + \frac{1}{2}G^T\left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1)\right) \\
&\quad - \alpha a_1^T\left(\bar{x} - \frac{\alpha}{2}a_1 + \frac{1}{2}G^T\left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1)\right) \\
&= 2\bar{x}^T\bar{x} - 2\alpha b_1 + \bar{x}^T G^T\left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) + \frac{\alpha^2}{2}\left|\left|a_1\right|\right|^2 \\
&\quad - \frac{\alpha}{2} a_1^T G^T\left(GG^T\right)^{-1}(2g - 2G\bar{x} + \alpha G a_1) \\
&= 2\bar{x}^T\bar{x} - 2\alpha b_1 + \frac{\alpha^2}{2}\left|\left|a_1\right|\right|^2 + 2\bar{x}^T G^T\left(GG^T\right)^{-1} g - 2\bar{x}^T G^T\left(GG^T\right)^{-1} G\bar{x} \\
&\quad + \alpha\bar{x}^T G^T\left(GG^T\right)^{-1} G a_1 - \alpha a_1^T G^T\left(GG^T\right)^{-1} g + \alpha a_1^T G^T\left(GG^T\right)^{-1} G\bar{x} \\
&\quad - \frac{\alpha^2}{2} a_1^T G^T\left(GG^T\right)^{-1} G a_1 \\
&= 2\bar{x}^T\bar{x} - 2\alpha b_1 + \frac{\alpha^2}{2}\left|\left|a_1\right|\right|^2 + 2\bar{x}^T G^T\left(GG^T\right)^{-1} g + 2\alpha a_1^T G^T\left(GG^T\right)^{-1} G\bar{x} \\
&\quad - \alpha a_1^T G^T\left(GG^T\right)^{-1} g - \frac{\alpha^2}{2} a_1^T G^T\left(GG^T\right)^{-1} G a_1 - 2\bar{x}^T G^T\left(GG^T\right)^{-1} G\bar{x}
\end{aligned}$$

Combining these two results gives

$$\begin{aligned}\left|\left|x'\right|\right|^2 - (2\bar{x} - \alpha a_1)^T x' &= g^T \left(GG^T\right)^{-1} g + \bar{x}^T G^T \left(GG^T\right)^{-1} G\bar{x} - \alpha a_1^T G^T \left(GG^T\right)^{-1} G\bar{x} \\ &+ \frac{\alpha^2}{4} a_1^T G^T \left(GG^T\right)^{-1} G a_1 - 2\bar{x} G^T \left(GG^T\right)^{-1} g \\ &+ \alpha a_1^T G^T \left(GG^T\right)^{-1} g - ||\bar{x}||^2 + \alpha b_1 - \frac{\alpha^2}{4} ||a_1||^2 .\end{aligned}$$

Finally we get that the condition

$$\left|\left|x'\right|\right|^2 - (2\bar{x} - \alpha a_1)^T x' + ||\bar{x}||^2 - \alpha b_1 = 0$$

is equivalent to

$$\begin{aligned}\frac{-\alpha^2}{4} a_1^T \left(I - G^T \left(GG^T\right)^{-1} G\right) a_1 + \alpha \left(a_1^T G^T \left(GG^T\right)^{-1} (g - G\bar{x})\right) \\ + \left|\left|G^T \left(GG^T\right)^{-1} (g - G\bar{x})\right|\right|^2 = 0\end{aligned} \tag{3.49}$$

which is precisely (3.47). ■

As in the case of a single polyhedron, we have found that the optimal α must satisfy a quadratic equation, although the coefficients of the particular quadratic equation are much more complicated in this setting. We can, however, still find its solution, as we now show.

Lemma 3.2.7. *The quadratic equation (3.49) has real roots.*

Proof. Let

$$K_1 = a_1^T \left(I - G^T \left(GG^T\right)^{-1} G\right) a_1,$$

which appears in the coefficient for α^2. Because the matrix $G^T \left(GG^T\right)^{-1} G$ is an orthogonal projection matrix, its eigenvectors are either 0 or 1 and we have

$$a_1^T G^T \left(GG^T\right)^{-1} G a_1 \leq a_1^T a_1$$

and so $K_1 \geq 0$. Define

$$D = \frac{1}{\sqrt{||a_2||^2 \, ||c||^2 - (a_2^T c)^2}} \begin{bmatrix} ||c|| & 0 \\ \frac{-a_2^T c}{||c||} & \sqrt{||a_2||^2 - \frac{(a_2^T c)^2}{||c||^2}} \end{bmatrix}$$

so $DD^T = \left(GG^T\right)^{-1}$, then we can write the discriminant of (3.49) as

$$16 \left|\left| \begin{bmatrix} \sqrt{K_1} D^T G \\ a_1^T G^T \left(GG^T\right)^{-1} \end{bmatrix} \bar{x} - \begin{bmatrix} \sqrt{K_1} D^T g \\ a_1^T G^T \left(GG^T\right)^{-1} g \end{bmatrix} \right|\right|^2$$

which is clearly nonnegative, implying that (3.49) does indeed have real roots. ■

Finally, we can obtain the solution to (3.49) and the expression for the lifting coefficent α.

Theorem 3.2.8. *In the case where $K_1 > 0$, the positive root of the quadratic equation (3.49) is given by*

$$\alpha = \frac{2\left(a_1^T G^T \left(GG^T\right)^{-1}(g - G\bar{x}) + \left|\left|\begin{bmatrix}\sqrt{K_1}D^T G \\ a_1^T G^T \left(GG^T\right)^{-1}\end{bmatrix}\bar{x} - \begin{bmatrix}\sqrt{K_1}D^T g \\ a_1^T G^T \left(GG^T\right)^{-1} g\end{bmatrix}\right|\right|\right)}{a_1^T \left(I - G^T \left(GG^T\right)^{-1} G\right) a_1} \tag{3.50}$$

When $K_1 = 0$, (3.49) becomes a linear equation and we get

$$\alpha = -\frac{\left|\left|G^T \left(GG^T\right)^{-1}(g - G\bar{x})\right|\right|^2}{a_1^T G^T \left(GG^T\right)^{-1}(g - G\bar{x})} \tag{3.51}$$

In both cases we observe that α is a strictly convex function of the point $\bar{x}$. This is unfortunate; we would hope to be able to solve the separation problem using a procedure similar to that of Section 3.1.3. However, the objective function in this formulation of the objective problem would be

$$\text{maximize:} - ||x||^2 + 2\hat{x}^T x + \alpha(b_1 - a_1^T \hat{x})$$

where $(\hat{x}, \hat{z})$ is the point we wish to separate. As α is a strictly convex function of x and $a_1^T \hat{x} < b_1$, this formulation of the separation problem is no longer convex.

3.2.3 A New Parameterization

Until now, we have attempted to construct lifted inequalities following the algebraic definition: starting with a tangent inequality and lifting by explicitly modifying its coefficients. With this approach, lifted inequalities were parameterized by their origination point, the lifting normal vector, and the lifting coefficient. This worked well in the case where $\mathcal{F}$ was the complement of a polyhedron, but as we saw in the previous section, breaks down in an even slightly more complex setting. This section introduces a new parameterization for lifted cuts which emerges naturally from the geometry of the sets of points at which they are violated.

Recall that the set of points x for which $(x, x^T x)$ violating the linear inequality

$$z \geq \gamma^T x + \beta, \tag{3.52}$$

is a ball in $\mathbb{R}^n$ with center $\frac{1}{2}\gamma$ and radius $\sqrt{\frac{1}{4}\left|\left|\gamma\right|\right|^2+\beta}$. Defining

$$\mu = \frac{1}{2}\gamma$$

and

$$\begin{aligned} \rho &= \frac{1}{4}\left|\left|\gamma\right|\right|^2+\beta \\ &= \left|\left|\mu\right|\right|^2+\beta \end{aligned} \tag{3.53}$$

we have

$$\gamma^T x + \beta = 2\mu^T x + \rho - \mu^T \mu.$$

We will now parameterize cuts directly by μ and ρ, the center and *squared* radius of the ball of points x for which $(x, x^T x)$ violates the cut.

Definition 3.2.9. *For $(\mu, \rho) \in \mathbb{R}^n \times \mathbb{R}_+$, we call the inequality*

$$z \geq 2\mu^T x - \left|\left|\mu\right|\right|^2 + \rho \tag{3.54}$$

the ball inequality *defined by (μ, ρ).*

We note that ball inequalities are only defined for $\rho \geq 0$, but as the next result shows, this is not a restriction.

Lemma 3.2.10. *Any inequality*

$$z \geq \gamma^T x + \beta \tag{3.55}$$

which holds with equality at at least one point $(y, y^T y)$ is a ball inequality.

Proof. We need only to show that ρ as defined in (3.53) is nonnegative. Let $(y, y^T y)$ be the point where the inequality (3.55) is tight. Then we have

$$y^T y = \gamma^T y + \beta, \quad \text{or} \quad \beta = y^T y - \gamma^T y.$$

Then

$$\rho = \frac{1}{4}\left|\left|\gamma\right|\right|^2+\beta = \frac{1}{4}\left|\left|\gamma\right|\right|^2 + y^T y - \gamma^T y = \left|\left|y - \frac{1}{2}\gamma\right|\right|^2 \geq 0.$$

■

Conversely, we have the following.

Lemma 3.2.11. *Consider a ball inequality defined by the parameters* (μ, ρ), *which supports* $\mathcal{S}$ *at some point* $(y, y^T y)$. *If* $\rho = 0$, *then the ball inequality is the tangent inequality at* y. *If* $\rho > 0$, *then the ball inequality is a nontrivial first-order lifted inequality.*

Proof. If $\rho = 0$, then the ball inequality is valid for all (x, z) with $z \geq x^T x$. The differentiability of $f(x) = x^T x$ and the fact that the ball inequality is tight at $(y, y^T y)$ together imply that the ball inequality must be the tangent inequality at y.

If $\rho > 0$, then by Theorem 2.2.5, the ball inequality must be a lifted inequality. ■

With this new parameterization, the generic formulation for finding the strongest valid cut at a given point $\hat{x}$ is:

$$\begin{aligned} \text{maximize:} \quad & 2\mu^T \hat{x} + \rho - \mu^T \mu \\ \text{subject to:} \quad & \operatorname{int}(\mathcal{B}(\mu, \rho)) \cap \mathcal{F} = \varnothing \end{aligned}$$

The constraint in this problem follows from Proposition 3.0.8. Note that the objective remains a concave quadratic function of the parameters of the inequality.

3.2.4 Separation for Multiple Excluded Polyhedra

In this section we utilize the new geometric characterization of lifted inequalities to devise a separation algorithm for the case when $\mathcal{F}$ is the complement of the union of the interiors of several polyhedra. As before, we choose to express $\mathcal{F}$ as a union of polyhedra. We begin by deriving a necessary and sufficient condition for the validity of a ball inequality

$$z \geq 2\mu^T x - \mu^T \mu + \rho \tag{3.56}$$

in the case of a single feasible polyhedron. The extension of this validity condition to the case of a union of polyhedra follows easily, and we conclude this section by using this validity condition to develop a formulation of the separation problem.

Our first result characterizes the validity of the ball inequality (3.56) when the feasible region is a polyhedraon as a lower bound on the optimal value of a quadratic program.

Lemma 3.2.12. *The ball inequality (3.56) is valid over*

$$\{(x, z) \mid Cx \geq d, \ z \geq x^T x\}$$

if and only if

$$\left\{\begin{array}{ll} \textit{minimize:} & \frac{1}{2}\,||x-\mu||^2 - \frac{1}{2}\rho \\ \textit{subject to:} & Cx \geq d \end{array}\right\} \geq 0 \tag{3.57}$$

Proof. By Proposition 3.0.8, the inequality (3.56) is valid if and only if there is no point x with $Cx \geq d$ and $||x-\mu||^2 < \rho$, which is equivalent to (3.57). ∎

The quadratic program in (3.57) is convex and all constraints are linear. Therefore strong duality holds for this problem (for results on strong duality and the dual of a quadratic program, see [19]). Using this, we can derive a new equivalent condition for the validity of the ball inequality (3.56).

Lemma 3.2.13. *The ball inequality (3.56) defined by μ and ρ is valid for*

$$\{(x,z) \mid Cx \geq d,\ z \geq x^T x\}$$

if and only if there exists a vector of nonnegative multipliers λ where

$$-\frac{1}{2}\left|\left|C^T\lambda\right|\right|^2 - \lambda^T(C\mu - d) - \frac{1}{2}\rho \geq 0. \tag{3.58}$$

Proof. The dual of the quadratic program in (3.57) is

$$\begin{array}{ll} \text{maximize:} & -\frac{1}{2}\left|\left|C^T\lambda\right|\right|^2 - \lambda^T(C\mu - d) - \frac{1}{2}\rho \\ \text{subject to:} & \lambda \geq 0 \end{array}$$

By weak duality, the objective value of any dual feasible solution gives a lower bound on the primal value. So the existence of $\lambda \geq 0$ satisfying (3.58) implies (3.57) and therefore that the cut is valid.

If no such λ exists, then the optimal dual value is strictly negative. By strong duality, primal and dual optimal values are equal, and so the primal value is negative as well. This implies that there is some point x with $Cx \geq d$ and $||x-\mu||^2 < \rho$ and therefore the ball inequality is not valid. ∎

By multiplying (3.58) by 2 and then adding and subtracting $\mu^T\mu$, we get the equivalent condition:

$$-\left|\left|C^T\lambda + \mu\right|\right|^2 + 2\lambda^T d + \mu^T\mu - \rho \geq 0. \tag{3.59}$$

When $\mathcal{F}$ is the union of multiple polyhedra, we can apply this condition to each one to ensure that the lifted inequality is valid over $\mathcal{S}$.

Write

$$\mathcal{F} \;=\; \bigcup_{k=1}^{K} \{x \mid C_k x \geq d_k\}. \tag{3.60}$$

and

$$\begin{aligned} \mathcal{S} &= \{(x,z) \mid z \geq x^T x,\ x \in \mathcal{F}\} \\ &= \bigcup_{k=1}^{K} \{(x,z) \mid z \geq x^T x,\ C_k x \geq d_k\}. \end{aligned} \tag{3.61}$$

The ball inequality defined by (μ, ρ) is valid for $\mathcal{S}$ if and only if the interior of the ball $\mathcal{B}(\mu, \sqrt{\rho})$ does not intersect *any* of the polyhedra comprising $\mathcal{F}$. Thus we have the following characterization of validity:

Theorem 3.2.14. *The ball inequality defined by μ and ρ is valid for $\mathcal{S}$ if and only if there exist nonnegative vectors $\lambda_1, \ldots, \lambda_K$ with*

$$-\left|\left|C_k^T \lambda_k + \mu\right|\right|^2 + 2\lambda_k^T d + \mu^T \mu - \rho \geq 0 \;\; \textit{for each} \;\; k \in \{1, \ldots, K\}.$$

Proof. Apply Lemma 3.2.13 to each set

$$\{(x,z) \mid z \geq x^T x,\ C_k x \geq d_k\}$$

in the definition of $\mathcal{S}$. ■

Given a point $\hat{x}$ with $\hat{x} \notin \mathcal{F}$ but $\hat{x} \in \operatorname{conv}(\mathcal{F})$, we can find the strongest valid ball inequality at $\hat{x}$ using the following formulation:

$$\begin{aligned} \text{maximize:} \quad & 2\mu^T \hat{x} - (\mu^T \mu - \rho) \\ \text{subject to:} \quad & -\left|\left|C_k^T \lambda_k + \mu\right|\right|^2 + 2\lambda_k^T d_k + \mu^T \mu - \rho \geq 0 \quad k = 1, \ldots, K \\ & \lambda_k \geq 0 \quad k = 1, \ldots, K \end{aligned}$$

This constraints in this formulation are nonconvex. However, notice that the term $\mu^T \mu - \rho$ appears in the objective and in each of the quadratic constraints. Suppose we replace $\mu^T \mu - \rho$ in the objective by a new variable τ, and add the valid convex constraint $\mu^T \mu - \rho \leq \tau$ to the formulation. We would clearly have $\tau = \mu^T \mu - \rho$ in any optimal solution: if $\mu^T \mu - \rho < \tau$, then τ could be decreased to yield a strictly better objective value. Thus we replace $\mu^T \mu - \rho$ with τ in both the objective and the constraints, obtaining the convex formulation:

$$\begin{aligned} \text{maximize:} \quad & 2\mu^T \hat{x} - \tau \\ \text{subject to:} \quad & -\left|\left|C_k^T \lambda_k + \mu\right|\right|^2 + 2\lambda_k^T d_k + \tau \geq 0 \quad k = 1, \ldots, K \\ & \lambda_k \geq 0 \quad k = 1, \ldots, K \end{aligned} \tag{3.62}$$

With this formulation, we have the following separation result.

Theorem 3.2.15. *Let $\mathcal{F}$ and $\mathcal{S}$ be as defined in (3.60) and (3.61), and let $(\hat{x}, \hat{z}) \in \mathbb{R}^{n+1}$ with $\hat{x} \in \operatorname{conv}(\mathcal{F})$. Let z^* be the optimal value to the optimization problem in (3.62). Then $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ if and only if $z^* \leq \hat{z}$.*

3.2.5 A Closer Look

In our analysis of the disjunctive approach, we observed that the dual problem to $P(\hat{x})$ had the linear objective of $\gamma^T \hat{x} + \beta$. Moreover, we showed that any feasible solution to the dual, composed of the variables γ and β along with others which do not appear in the objective, provided an inequality

$$z \geq \gamma^T x + \beta$$

which is valid for $\mathcal{S}$, and that the optimal γ^* and β^* gave the strongest valid lifted inequality at $\mathcal{S}$. Now compare this to the formulation (3.62) from the last section. This problem also has a linear objective $(2\mu^T \hat{x} - \tau)$, and any feasible solution provides an inequality valid for $\mathcal{S}$, with the optimal solution providing the strongest valid inequality at $\hat{x}$. Given these facts, it seems that our formulation for solving the separation problem is at least very similar to the dual of the problem $P(\hat{x})$ from the disjunctive formulation, and as such we will refer to it as the "dual" formulation.

At first glance, it is unclear what we have gained through this new formulation. As with the primal formulation, it also has a large number of variables, and a quadratic constraint for every component of the disjunction defining $\mathcal{F}$. The sheer number of quadratic constraints may make solution of (3.62) difficult. However, notice the independence in the constraints: each vector of variables λ_k appears only in one constraint. For fixed values of μ and τ, the left hand side of each constraint

$$-\left|\left|C_k^T \lambda_k + \mu\right|\right|^2 + 2\lambda_k^T d_k + \tau \geq 0 \tag{3.63}$$

can be maximized independently (possibly in parallel) to see if the cut defined by μ and τ (which is a ball inequality whose corresponding ball has center μ and radius $\sqrt{\mu^T\mu - \tau}$) is valid for the k^{th} component of the disjunction defining $\mathcal{S}$. That is, we solve the problem:

$$\begin{aligned} \text{maximize:} \quad & -\left|\left|C_k^T\lambda_k + \mu\right|\right|^2 + 2\lambda_k^T d_k \\ \text{subject to:} \quad & \lambda_k \geq 0 \end{aligned} \tag{3.64}$$

with μ fixed. The cut defined by μ and τ is valid for

$$\{(x, z) \mid C_k x \geq d,\ z \geq x^T x\}.$$

if and only if the optimal λ_k satisfies (3.63). We have experimented in solving the problem (3.62) using the following procedure:

Algorithm 3.2.16 (Iterative Algorithm for the Dual Formulation)**.**

1. Choose an initial μ. Initialize the variables λ_k by solving (3.64) for each k. Initialize a small subset of the constraints to be "active".

2. Solve a relaxed version of (3.62), enforcing only the current subset of active constraints, to obtain a new μ and τ. If the relaxation is unbounded, add constraints as necessary until it is bounded.

3. Check validity: for each k, test if λ_k satisfies (3.63) with the current μ and τ. If not, update λ_k by solving (3.64), and test again if the updated λ_k satisfies (3.63).

4. If all constraints are satisfied, stop. The current μ and τ are feasible and optimal. If not, update the set of active constraints and go to Step 2.

The task in Step 4 of updating the set of active constraints may be performed in any number of ways. We may simply choose the new set to be the set of all constraints which were not satisfiable with the previous μ and τ, or perhaps choose only a subset of the constraints which were most violated. In our experiments, we kept all constraints which were satisfied with equality in the current solution in the active set, and added a small number of the constraints which were most violated. We also ensured that the size of the active set would always increase with each update, to give a simple guarantee that the algorithm would terminate (although we hope not to reach the point where all constraints are active). We point out that in Step 3, the constraints in the current active set will be satisfied, and it is not necessary to check these constraints. Also, once a valid cut has been found for a particular $\hat{x}$, the variables λ_k can be stored and reused if another cut is needed. This allows the initialization of the λ_k's in Step 1 to be bypassed, which can lead to a substantial decrease in computational cost.

This approach has theoretical justification. In the primal form we obtain, after reversing the change of variables, a description of $(\hat{x}, z^*)$ (with z^* defined as in Lemma 3.2.1) as a convex combination of vectors in $\mathcal{S}$. By Carathéodory's Theorem, expressing any point in $\operatorname{conv}(\mathcal{S})$ in this manner requires at most $n+2$ elements from $\mathcal{S}$. Therefore, we should expect that at most $n+2$ constraints are truly needed in (3.62), even if K is much larger than n.

Our approach has been successful in finding cuts in problems with a small number of variables but a large number of component polyhedra defining $\mathcal{F}$. We provide the details of our experiments and results in Section 5.3.

3.3 Excluding an Ellipsoid

We now extend our analysis of the set $\operatorname{conv}(\mathcal{S})$ to the case where the feasible region is the complement of an ellipsoid. This is an extremely simple case of a nonconvex quadratically constrained quadratic program as described in Section 1.2.1, and bears strong resemblance to the approach of "no-good" cuts in nonlinear programming (see [10], [30], or [49]), where a nonconvex constraint on the minimum distance from a current infeasible relaxed solution point is imposed using known information about the structure of the feasible set. The methods used in the case of a single excluded polyhedron, including the disjunctive formulation, relied upon the excluded region having a representation as a union of a finite number of components and no longer apply in this setting. Fortunately, the geometric insights obtained in the polyhedral setting do prove valuable in deriving a polynomial-time separation procedure.

Let $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$, where A is symmetric and positive definite. The set

$$P = \{x \in \mathbb{R}^n \mid x^T Ax - 2b^T x + c \leq 0\}$$

is an ellipsoid which we assume is nonempty and full-dimensional. We define the feasible region as

$$\begin{aligned} \mathcal{F} &= \mathbb{R}^n \setminus \operatorname{int}(P) \\ &= \{x \in \mathbb{R}^n \mid x^T Ax - 2b^T x + c \geq 0\} \end{aligned}$$

and as usual we define the set

$$\mathcal{S} = \{(x, z) \mid x \in \mathcal{F},\ z \geq x^T x\}. \tag{3.65}$$

In this section, we show that the parameters for separating inequalities for $\mathcal{S}$ can be obtained in closed form, at the cost of computing the largest eigenvalue of the matrix A. We also point out some applications of this setting and consider an extension in which a linear inequality is added to the definition of the feasible region.

3.3.1 Valid Lifted Inequalities

Before continuing, we state a corollary to Lemma 3.1.1 which defines the choice of lifting vector when the feasible set is defined by a convex function.

Corollary 3.3.1. *Let $\bar{x} \in \partial F$, let $\epsilon > 0$ and suppose that for all x with $||x - \bar{x}|| \leq \epsilon$, we have*

$$x \in \mathcal{F} \quad \text{if and only if} \quad g(x) \geq 0.$$

where g is convex and differentiable. Let

$$z \geq \gamma^T x + \beta$$

be valid for $\mathcal{S}$, and suppose that this inequality is tight at $(\bar{x}, ||\bar{x}||^2)$. Then we must have

$$\begin{aligned} \gamma &= 2\bar{x} - \alpha \nabla g(\bar{x}) \\ \beta &= \alpha \nabla g(\bar{x})^T \bar{x} - ||\bar{x}||^2 \end{aligned}$$

for some $\alpha \geq 0$.

Proof. By the convexity of g, the condition

$$\nabla g(\bar{x})^T x \geq \nabla g(\bar{x})^T \bar{x}$$

is sufficient to guarantee $x \in \mathcal{F}$, for x inside the ball $\mathcal{B}(\bar{x}, \epsilon)$. This inequality is tight at $\bar{x}$. The result follows from the arguments in the proof of Lemma 3.1.1. ■

Returning to our current setting, consider a point $\bar{x} \in \partial P$. Applying Corollary 3.3.1 with $g(x) = x^T Ax - 2b^T x + c$, we see that any valid lifted inequality generated from $\bar{x}$ must be of the form

$$z \geq \left(2\bar{x} - \alpha(2A\bar{x} - 2b)\right)^T (x - \bar{x}) + \bar{x}^T \bar{x}. \tag{3.66}$$

This inequality is valid for α nonnegative and small enough - in some cases only for $\alpha = 0$. Of course, constructing lifted cuts in this form will not be helpful in solving the separation problem, as it would require optimizing over the boundary of the ellipsoid P. We instead choose to parameterize cuts as ball inequalities, in the form

$$z \geq 2\mu^T x - \mu^T \mu + \rho. \tag{3.67}$$

As we will see, this choice of parameterization proves natural when P is a region defined by a quadratic constraint.

3.3.2 Applying the S-Lemma

As the first step towards showing how we can solve the separation problem in this setting, we now show how the S-Lemma can be applied to obtain an equivalent condition for the validity of a ball inequality. By Proposition 3.0.8, the validity of a lifted cut defined by the ball with center μ and radius $\sqrt{\rho}$ is equivalent to the condition

$$\operatorname{int}(\mathcal{B}(\mu, \sqrt{\rho})) \cap \mathcal{F} = \varnothing.$$

Geometrically, this is equivalent to $\mathcal{B}(\mu, \sqrt{\rho})$ being contained in the ellipsoid P. That is, we need

$$-(x^T Ax - 2b^T x + c) \geq 0 \text{ for all } x \text{ with } ||x - \mu||^2 \leq \rho.$$

Applying the S-Lemma (1.2.2) with

$$f(x) = -(x^T Ax - 2b^T x + c) \quad \text{and} \quad g(x) = ||x - \mu||^2 - \rho,$$

this is equivalent to the existence of $\tau \geq 0$ where

$$-\left(x^T Ax - 2b^T x + c\right) + \tau \left(||x - \mu||^2 - \rho\right) \geq 0 \quad \forall x \in \mathbb{R}^n. \tag{3.68}$$

Note that the condition of the existence of $\hat{x}$ with $g(\hat{x}) < 0$ is satisfied by the assumption that P is full-dimensional. Also, because A has positive eigenvalues, it cannot be the case that $\tau = 0$. Thus, dividing (3.68) by τ and defining $\theta = \tau^{-1}$ the validity of the cut defined by μ and ρ is equivalent to the existence of $\theta > 0$ where

$$-\theta(x^T Ax - 2b^T x + c) + \left(||x - \mu||^2 - \rho\right) \geq 0 \quad \forall x \in \mathbb{R}^n. \tag{3.69}$$

This condition will prove instrumental in formulating and solving the separation problem.

3.3.3 The Separation Problem

We now show how to solve the separation problem in the ellipsoidal case. Assume we have a point $(\hat{x}, \hat{z})$ with $\hat{x} \in \operatorname{int}(P)$. We wish to find a valid linear inequality separating $(\hat{x}, \hat{z})$ from $\operatorname{conv}(\mathcal{S})$, if one exists. If the lifted cut is parameterized by its center μ and squared radius ρ, the value of the cut at $\hat{x}$ is

$$2\mu^T \hat{x} - \mu^T \mu + \rho,$$

and this is therefore the quantity to be maximized. Also note that the validity condition (3.69) can be expressed as the existence of a positive multiplier θ such that

$$\min_{x \in \mathbb{R}^n} \left\{ -\theta(x^T A x - 2b^T x + c) + \left(||x - \mu||^2 - \rho \right) \right\} \geq 0. \tag{3.70}$$

Once again we have characterized the validity of a lifted cut as a bound on the optimal value of a minimization problem.

The following theorem gives closed-form expressions for the optimal μ and ρ.

Theorem 3.3.2. *Let $\bar{\lambda}$ be the largest eigenvalue of A, and define $\theta^* = \bar{\lambda}^{-1}$. Given a point $\hat{x} \in \operatorname{int}(P)$, the μ and ρ giving the strongest valid inequality at $\hat{x}$ are*

$$\mu^* = \theta^* b + (I - \theta^* A)\, \hat{x} \tag{3.71}$$

$$\rho^* = ||\mu^* - \hat{x}||^2 - \theta^*(\hat{x}^T A \hat{x} - 2b^T \hat{x} + c) \tag{3.72}$$

Proof. Using the validity condition derived in Section 3.3.2, we can formulate the separation problem as:

$$\begin{aligned}
\text{maximize:} \quad & 2\mu^T \hat{x} - \mu^T \mu + \rho \\
\text{subject to:} \quad & \min_{x \in \mathbb{R}^n} \left\{ -\theta(x^T A x - 2b^T x + c) + \left(||x - \mu||^2 - \rho \right) \right\} \geq 0 \\
& \theta > 0
\end{aligned}$$

The constraint $\rho \geq 0$ is omitted, but we will show that we always have $\rho \geq 0$ at optimality. Also, the strict inequality constraint on θ might seem to be a cause for concern, but we will show that θ is strictly positive at optimality as well.

Expanding some terms in the first constraint and moving terms that do not depend on x from the scope of the minimization gives the equivalent constraint:

$$\min_{x \in \mathbb{R}^n} \left\{ x^T (I - \theta A) x - 2(\mu - \theta b)^T x) \right\} - \theta c + \mu^T \mu - \rho \geq 0 \tag{3.73}$$

For this constraint to hold, the value of the minimization problem must be finite. By Lemma (A.2.1), this is true if and only if $(I - \theta A)$ is positive semidefinite and $\mu = \theta b + (I - \theta A)\pi$ for some $\pi \in \mathbb{R}^n$. Requiring $\theta \leq \bar{\lambda}^{-1}$ is necessary and sufficient to ensure $(I - \theta A) \succeq 0$. With these conditions, the optimal solution to the minimization problem is any x^* satisfying

$$(I - \theta A)x^* = \mu - \theta b = (I - \theta A)\pi,$$

implying that π is an optimal solution, with an optimal value of $-\pi^T(I - \theta A)\pi$. We now have the following formulation of the separation problem:

$$\begin{aligned} \text{maximize:} \quad & 2\mu^T\hat{x} - \mu^T\mu + \rho \\ \text{subject to:} \quad & -\pi^T(I - \theta A)\pi - \theta c + \mu^T\mu - \rho \geq 0 \\ & 0 < \theta \leq \bar{\lambda}^{-1} \\ & \mu = \theta b + (I - \theta A)\pi \end{aligned}$$

The problem now includes the variable vector π in addition to μ, θ, and ρ. Notice that ρ appears in the objective and the first constraint, but nowhere else. Clearly the first constraint will hold with equality in any optimal solution. If not, it would be possible to increase ρ by a small but positive amount, which would yield a better objective value while preserving feasibility. We can then substitute

$$\rho = \mu^T\mu - \pi^T(I - \theta A)\pi - \theta c \tag{3.74}$$

in the objective, removing the variable ρ and the first constraint from the formulation. At the same time, we substitute

$$\mu = \theta b + (I - \theta A)\pi \tag{3.75}$$

into the objective to eliminate μ and the last constraint. This results in the new formulation:

$$\begin{aligned} \text{maximize:} \quad & 2(\theta b + (I - \theta A)\pi)^T\hat{x} - \pi^T(I - \theta A)\pi - \theta c \\ \text{subject to:} \quad & 0 < \theta \leq \bar{\lambda}^{-1} \end{aligned}$$

For any fixed feasible value of θ, the optimal solution to this problem is given by $\pi^* = \hat{x}$. The objective value obtained is

$$\hat{x}^T(I - \theta A)\hat{x} + 2\theta b^T\hat{x} - \theta c \quad = \quad \hat{x}^T\hat{x} - \theta(\hat{x}^TA\hat{x} - 2b^T\hat{x} + c). \tag{3.76}$$

Because $\hat{x} \in \text{int}(P)$ by assumption, we have $\hat{x}^T A\hat{x} - 2b^T\hat{x} + c < 0$ and the objective value is increasing in θ. Thus the optimal θ^* will be at the upper bound of $\bar{\lambda}^{-1}$. Recalling the substitutions (3.74) and (3.75), we obtain expressions for the optimal μ^* and ρ^*:

$$\begin{aligned}\mu^* &= \theta^* b + (I - \theta^* A)\pi^* \\ &= \theta^* b + (I - \theta^* A)\hat{x}\end{aligned}$$

and

$$\begin{aligned}\rho^* &= {\mu^*}^T\mu^* - {\pi^*}^T(I - \theta A)\pi^* - \theta^* c \\ &= {\mu^*}^T\mu^* - \hat{x}^T(I - \theta^* A)\hat{x} - \theta^* c \\ &= {\mu^*}^T\mu^* - 2\hat{x}^T(I - \theta^* A)\hat{x} + \hat{x}^T(I - \theta^* A)\hat{x} - \theta^* c \\ &= {\mu^*}^T\mu^* - 2\hat{x}^T(\mu^* - \theta^* b) + \hat{x}^T(I - \theta^* A)\hat{x} - \theta^* c \\ &= ||\mu^* - \hat{x}||^2 - \theta^*(\hat{x}^T A\hat{x} - 2b^T\hat{x} + c)\end{aligned}$$

which completes the proof. ■

Note that in this case, we do not know for which points $(x, x^T x)$ with x feasible the cut

$$z \geq 2{\mu^*}^T x - {\mu^*}^T\mu^* + \rho^*$$

holds with equality. We are, however, guaranteed that there will be at least one such feasible x. This is due to the fact that the value of the cut increases with the squared radius ρ. Any valid cut with a fixed center $\mu \in \text{int}(P)$ is made as strong as possible by increasing ρ until the ball $\mathcal{B}(\mu, \sqrt{\rho})$ touches the boundary of the infeasible ellipsoid P at some point $\bar{x}$. This implies that the ball inequality supports $\mathcal{S}$ at $(\bar{x}, \bar{x}^T\bar{x})$. Certainly the ball inequality defined by (μ^*, ρ^*) must support $\mathcal{S}$, for if not ρ^* could be increased while preserving validity. By Lemma 3.2.11, this shows that the ball inequality defined by μ^* and ρ^* is a lifted inequality.

Using Theorem 2.2.6, we obtain as a corollary a means for solving the separation problem.

Corollary 3.3.3. *Let $(\hat{x}, \hat{z}) \in R^{n+1}$ with $\hat{x} \in \text{int}(P)$. Define μ^* and ρ^* as in (3.71) and (3.72). Then $(\hat{x}, \hat{z}) \in \text{conv}(\mathcal{S})$ if and only if*

$$2{\mu^*}^T\hat{x} - {\mu^*}^T\mu^* + \rho^* \leq \hat{z}.$$

We note in passing that that result in this section can be generalized slightly.

Remark 3.3.4. *We assumed that the matrix A is positive definite, but this is not necessary. All of the analysis above holds when A is positive semidefinite with at least one positive eigenvalue.*

In [48] it was pointed out that conv $(\mathcal{S})$ actually has a finite representation involving a second convex quadratic constraint. Here we present a proof of this fact which follows from our derivation of the solution to the separation problem.

Corollary 3.3.5. *Let $\bar{\lambda}$ be the largest eigenvalue of A, and define $\theta = \bar{\lambda}^{-1}$. Then*

$$conv(\mathcal{S}) = \{(x, z) \in \mathbb{R}^{n+1} \mid z \geq x^T x,\ z \geq x^T(I - \theta A)x + 2\theta b^T x - \theta c\}.$$

Proof. Let $(x, z) \in \mathcal{S}$. Then $x^T Ax - 2b^T x + c \geq 0$, which implies

$$x^T x \geq x^T(I - \theta A)x + 2\theta b^T x - \theta c.$$

Also, because $(x, z) \in \mathcal{S}$, we have $z \geq x^T x$, and therefore

$$z \geq x^T(I - \theta A)x + 2\theta b^T x - \theta c.$$

Any point $(\hat{x}, \hat{z}) \in \text{conv}(\mathcal{S})$ can be written as

$$(\hat{x}, \hat{z}) \ = \ \sum_{i=1}^{p} \lambda_i(x_i, z_i)$$

where $\sum_i \lambda_i = 1$, $\lambda_i \geq 0$ and $(x_i, z_i) \in \mathcal{S}$ for each i. As $(x_i, z_i) \in \mathcal{S}$, we have

$$z_i \ \geq \ x_i^T x_i \ \geq \ x_i^T(I - \theta A)x_i + 2\theta b^T x_i - \theta c \quad \text{for each} \quad i,$$

and so by convexity we have

$$\hat{z} \ \geq \ \hat{x}^T\hat{x} \quad \text{and} \quad \hat{z} \ \geq \ \hat{x}^T(I - \theta A)\hat{x} + 2\theta b^T\hat{x} - \theta c.$$

This shows

$$\text{conv}(\mathcal{S}) \subseteq \{(x, z) \mid z \geq x^T x,\ z \geq x^T(I - \theta A)x + 2\theta b^T x - \theta c\}.$$

To show the opposite inclusion, consider a point $(\hat{x}, \hat{z})$ satisfying

$$\hat{z} \ \geq \ \hat{x}^T\hat{x} \tag{3.77}$$

$$\hat{z} \ \geq \ \hat{x}^T(I - \theta A)\hat{x} + 2\theta b^T\hat{x} - \theta c \tag{3.78}$$

First, assume $\hat{x} \notin \text{int}(P)$. In this case the condition $\hat{z} \geq \hat{x}^T\hat{x}$ is sufficient to conclude $(\hat{x}, \hat{z}) \in \text{conv}(\mathcal{S})$. Next we assume $\hat{x} \in \text{int}(P)$. Suppose $(\hat{x}, \hat{z}) \notin \text{conv}(\mathcal{S})$. If this is the case, then the linear inequality

$$z \geq 2\mu^T x - \mu^T\mu + \rho$$

with μ and ρ defined as in (3.71) and (3.72) separates $(\hat{x}, \hat{z})$ from $\text{conv}(\mathcal{S})$. That is, we have

$$\hat{z} < 2\mu^T\hat{x} - \mu^T\mu + \rho. \tag{3.79}$$

However, making the substitutions from (3.71) and (3.72) into (3.79) gives

$$\hat{z} < \hat{x}^T(I - \theta A)\hat{x} + 2\theta b^T\hat{x} - \theta c,$$

which contradicts (3.78). ■

3.3.4 An Application

Consider an optimization problem of the form

$$\begin{aligned} \text{minimize:} \quad & z \\ \text{subject to:} \quad & z \geq x^T Q x + q^T x \\ & x \in \mathcal{C} \end{aligned} \tag{3.80}$$

where $\mathcal{C} \subset \mathbb{R}^n$ is a nonconvex set and Q is positive definite. Let $(\hat{x}, \hat{z})$ be the optimal solution to a relaxation of this problem, and suppose that it is known that there is a ball $\mathcal{B}(\sigma, \sqrt{\phi})$ with $\hat{x}$ in its interior, but whose interior does not intersect $\mathcal{C}$. This would be true, for instance, if a lower bound on the distance from $\hat{x}$ to the set $\mathcal{C}$ was known, in which case we would have $\sigma = \hat{x}$.

Because any feasible solution to 3.80 must lie outside of $\text{int}\left(\mathcal{B}(\sigma, \sqrt{\phi})\right)$, any lower bound on the value of the problem

$$\begin{aligned} \text{minimize:} \quad & z \\ \text{subject to:} \quad & z \geq x^T Q x - 2q^T x \\ & ||x - \sigma||^2 \geq \phi \end{aligned} \tag{3.81}$$

gives a valid lower bound for the value of (3.80). Let $Q = C^T C$ be the Cholesky decomposition of Q. Introducing the change of variables $y = Cx - C^{-T}q$ and letting $w = z + q^T Q^{-1} q$ gives an equivalent problem:

$$\begin{aligned}\text{minimize:} \quad & w - q^T Q^{-1} q \\ \text{subject to:} \quad & w \geq y^T y \\ & y^T C^{-T} C^{-1} y - 2(\sigma - Q^{-1}q)^T C^{-1} y + \left|\left|Q^{-1}q - \sigma\right|\right|^2 - \phi \geq 0\end{aligned} \tag{3.82}$$

the constraint set of this problem is exactly in the form of the set $\mathcal{S}$ defined in (3.65), with

$$A = C^{-T}C^{-1}, \quad b = C^{-T}(\sigma - Q^{-1}q), \quad \text{and} \quad c = \left|\left|Q^{-1}q - \sigma\right|\right|^2 - \phi.$$

From (3.71) and (3.72), we have that the parameters for the strongest ball inequality at $\hat{y} = C\hat{x} - C^{-T}q$ are

$$\mu^* = \theta^* C^{-T}(\sigma - Q^{-1}q) + (I - \theta^* C^{-T}C^{-1})\hat{y} \tag{3.83}$$

$$\rho^* = ||\mu^* - \hat{y}||^2 - \theta^* \left(\hat{y}^T C^{-T} C^{-1} \hat{y} - 2(\sigma - Q^{-1}q)^T C^{-1}\hat{y} + \left|\left|Q^{-1}q - \sigma\right|\right|^2 - \phi\right), \tag{3.84}$$

where θ^* is the inverse of the largest eigenvalue of $C^{-T}C^{-1}$. These parameters give the valid cut

$$w \geq 2\mu^{*T} y + \rho^* - ||\mu^*||^2 . \tag{3.85}$$

Using (3.83) and (3.84) and the transformation $\hat{y} = C\hat{x} - C^{-T}q$, we have

$$\begin{aligned}\rho^* - ||\mu^*||^2 \quad &= \quad -2\mu^{*T}\hat{y} + \hat{y}^T\hat{y} - \theta^* \, ||\hat{x} - \sigma||^2 \\ &= \quad -2\mu^{*T}\hat{y} + \hat{x}^T Q\hat{x} - 2q^T\hat{x} + q^T Q^{-1} q - \theta^*(||\hat{x} - \sigma||^2 - \phi).\end{aligned}$$

The cut (3.85) is then equivalent to

$$z \geq 2\mu^{*T}(y - \hat{y}) + \hat{x}^T Q\hat{x} - 2q^T\hat{x} - \theta^*(||\hat{x} - \sigma||^2 - \phi),$$

which, after transforming back into the original variables, is

$$z \geq 2\mu^{*T} C(x - \hat{x}) + \hat{x}^T Q\hat{x} - 2q^T\hat{x} - \theta^*(||\hat{x} - \sigma||^2 - \phi).$$

This can be simplified even further to give the following:

$$z \geq 2\left(\theta^*(\sigma - \hat{x}) + Q\hat{x} - q\right)^T x + \hat{x}^T\left(\theta I - Q\right)\hat{x} - \theta(\sigma^T\sigma - \phi) \tag{3.86}$$

The set of points x for which $(x, x^T Qx - 2q^T x)$ violates this cut is an ellipsoid which is contained in the ball $\mathcal{B}(\sigma, \sqrt{\phi})$, implying that the cut is valid.

The expression (3.86) no longer contains the matrix C. As the maximum eigenvector of $C^{-T}C^{-1}$ is the inverse of the minimum eigenvector of Q, computing the Cholesky decomposition of Q is actually not necessary. Then, as the eigenvalue θ^* can be computed once and stored, this represents a very inexpensive way to potentially cut off a series of points $(\hat{x}, \hat{z})$ to improve the bound on the optimal solution to the original problem.

3.3.4.1 Two Examples

The situation described above occurs in (at least) two situations. The first is in cardinality constrained quadratic programming. In this problem, the objective is to minimize $x^TQx - 2q^Tx$ (where $Q \succ 0$) and the feasible set includes the constraints $a^Tx = b$ and $||x||_0 \leq K$. Suppose we are given a vector $\hat{x}$ with $||x||_0 > K$. In [16] it is shown that it is possible to obtain a strong lower bound (within a factor of $(1+\epsilon)$) on the distance from $\hat{x}$ to $\{x \mid a^Tx = b,\ ||x||_0 \leq K\}$, which can then be used to form the ball $\mathcal{B}(\sigma, \sqrt{\phi})$ as described above.

The second situation is mixed-integer quadratic programming. In this problem, the objective is the same, and the feasible set includes the constraints

$$x_i \in \mathbb{Z} \quad \text{for} \quad i \in \mathcal{I} \subseteq \{1, \ldots, n\}.$$

Given a point $\hat{x}$ with $\hat{x}_i \notin \mathbb{Z}$ for at least one $i \in \mathcal{I}$, we construct the center point σ as follows:

$$\sigma_i = \begin{cases} \hat{x}_i & \text{if } i \notin \mathcal{I} \text{ or } \hat{x}_i \in \mathbb{Z} \\ \lfloor \hat{x}_i \rfloor + \frac{1}{2} & \text{if } i \in \mathcal{I} \text{ and } \hat{x}_i \notin \mathbb{Z}. \end{cases}$$

Then with

$$\phi = \frac{1}{4}\big|\{i \mid \hat{x}_i \notin \mathbb{Z},\ i \in \mathcal{I}\}\big|,$$

the interior of the ball $\mathcal{B}(\sigma, \sqrt{\phi})$ does not contain any feasible points, and a valid cut can be obtained using the procedure above.

3.3.5 Adding a Linear Inequality

We now consider an extension where the feasible region $\mathcal{F}$ is the intersection of a halfspace and the complement of the interior of an ellipsoid. The addition of the linear inequality complicates the characterization of valid ball inequalities, and in this case we are not able to obtain a provably polynomial-time separation procedure, although we do obtain a method which we have observed to work well empirically.

To begin, assume $\mathcal{F}$ is given by

$$\mathcal{F} = \{x \mid x^T Ax - 2b^T x + c \geq 0,\ h^T x \leq h_0\}.$$

where A is positive definite. Our first result establishes a set of conditions for the validity of the ball inequality

$$z \geq 2\mu^T x - \mu^T \mu + \rho \tag{3.87}$$

defined by (μ, ρ) over the set

$$\mathcal{S} = \{(x, z) \mid z \geq x^T x,\ x \in \mathcal{F}\}.$$

Lemma 3.3.6. *The ball inequality (3.87) is valid over $\mathcal{S}$ if and only if there exist a multiplier $\beta > 0$, a vector $q \in \mathbb{R}^n$, and a scalar $q_0 \in \mathbb{R}$ satisfying the following:*

$$\beta\left(-x^T Ax + 2b^T x - c\right) + ||x - \mu||^2 - \rho + (q^T x - q_0)(h^T x - h_0) \ \geq\ 0 \quad \forall x \in \mathbb{R}^n \tag{3.88a}$$

$$(q^T x - q_0) \ \geq\ 0 \quad \forall x : ||x - \mu||^2 \leq \rho, -2\mu^T x + \mu^T \mu \leq \rho \tag{3.88b}$$

Proof. It is easily shown that the ball inequality (3.87) is valid if and only if there is no point y with

$$\begin{aligned} -y^T Ay + 2b^T y - c &< 0 \\ ||y - \mu||^2 - \rho &\leq 0 \\ h^T y &\leq h_0 \end{aligned}$$

By a variant of the S-Lemma given in [54], this is equivalent to the existence of a nonnegative multiplier θ, a vector $v \in \mathbb{R}^n$, and a scalar v_0 satisfying the following conditions:

$$-x^T Ax + 2b^T x - c + \theta\left(||x - \mu||^2 - \rho\right) + (v^T x - v_0)(h^T x - h_0) \ \geq\ 0 \quad \forall x \in \mathbb{R}^n \tag{3.89a}$$

$$v^T x \ \geq\ 0 \quad \forall x : x^T x \leq 0, -2\mu^T x \leq 0 \tag{3.89b}$$

$$v^T x - v_0 \ \geq\ 0 \quad \forall x : ||x - \mu||^2 \leq \rho, -2\mu^T x + \mu^T \mu \leq \rho \tag{3.89c}$$

First, we notice that the condition (3.89b) can be ignored: the only x with $x^T x \leq 0$ is $x = 0$, which satisfies $v^T x \geq 0$ for any choice of the vector v. Next, by the assumption that $A \succ 0$, we see that it cannot be the case that $\theta = 0$. Thus we can define $\beta = \theta^{-1}$ and divide (3.89a) and (3.89c) by θ to obtain that an equivalent condition for validity is the existence of a vector v, scalar v_0, and positive multiplier β satisfying:

$$\beta\left(-x^T Ax + 2b^T x - c\right) + ||x-\mu||^2 - \rho + \beta(v^T x - v_0)(h^T x - h_0) \geq 0 \quad \forall x \in \mathbb{R}^n \tag{3.90a}$$

$$\beta(v^T x - v_0) \geq 0 \quad \forall x : ||x-\mu||^2 \leq \rho, -2\mu^T x + \mu^T \mu \leq \rho \tag{3.90b}$$

Next, we observe that v and v_0 only appear in (3.90) with a multiplicative factor of β. Introducing the changes of variables $q = \beta v$ and $q_0 = \beta v_0$ establishes the equivalence of (3.88) and (3.90), and proves the claim. ■

Next we analyze each of the conditions in (3.88) separately to obtain a characterization of validity more suitable for use in a formulation of the separation problem.

Lemma 3.3.7. *The condition (3.88a) is equivalent to the following:*

$$\mu^T\mu - \rho \;\geq\; y^T\left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right)y - q_0 h_0 + \beta c \tag{3.91a}$$

$$\left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right)y \;=\; \mu - \beta b + \frac{1}{2}\left(q_0 h + h_0 q\right) \tag{3.91b}$$

$$\left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right) \;\succeq\; 0 \tag{3.91c}$$

for some $y \in \mathbb{R}^n$.

Proof. First, observe that (3.88a) is equivalent to

$$\min_{x\in\mathbb{R}^n}\left\{x^T\left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right)x - \left(2\mu - 2\beta b + q_0 h + h_0 q\right)^T x\right\} + \mu^T\mu - \rho + q_0 h_0 - \beta c \geq 0. \tag{3.92}$$

For this to hold, the value of the minimization problem must be bounded. By Lemma A.2.1, this is true if and only if (3.91b) and (3.91c) hold, for some $y \in \mathbb{R}^n$. Assuming these conditions are satisfied, the optimal solution to the minimization problem is y, and the corresponding optimal value is

$$-\,y^T\left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right)y. \tag{3.93}$$

Substituting (3.93) for the value of the minimization problem in (3.92) and rearranging gives (3.91a).

■

Lemma 3.3.8. *The condition (3.88b) is equivalent to the existence of $\sigma \geq 0$ with*

$$\mu^T\mu - \rho \geq ||\sigma q - \mu||^2 + 2\sigma q_0 \tag{3.94}$$

Proof. We begin by expressing (3.88b) as a bound on an optimization problem: (3.88b) is equivalent to

$$\left\{\begin{array}{ll} \text{minimize:} & q^T x - q_0 \\ \text{subject to:} & ||x-\mu||^2 \leq \rho \\ & -2\mu^T x + \mu^T\mu \leq \rho \end{array}\right\} \geq 0$$

Notice that if the left hand side of the first constraint in this formulation is always at least as large as the left hand side of the second constraint. This implies that the second constraint is redundant and can be removed, giving the equivalent condition:

$$\left\{\begin{array}{ll} \text{minimize:} & q^T x - q_0 \\ \text{subject to:} & ||x-\mu||^2 \leq \rho \end{array}\right\} \geq 0 \tag{3.95}$$

It is easily shown that strong duality holds for the QCQP in (3.95), as long as ρ is positive. Therefore the inequality (3.95) holds if and only if there is a dual feasible λ with nonnegative objective value. Thus (3.95) is equivalent to the existence of a positive scalar λ with

$$-\frac{q^T q}{\lambda} + 2q^T\mu - \lambda\rho - 2q_0 \geq 0.$$

Dividing this inequality by λ, adding and subtracting $\mu^T\mu$, and introducing the change of variables $\sigma = \lambda^{-1}$ gives the equivalent condition:

$$\mu^T\mu - \rho \geq ||\sigma q - \mu||^2 + 2\sigma q_0 \quad \text{for some} \quad \sigma \geq 0,$$

which is exactly (3.94). ■

Suppose we are given a point $\hat{x}$ with $\hat{x}^T A\hat{x} - 2b^T\hat{x} + c < 0$ and $h^T\hat{x} \leq h_0$. Combining the results of Lemmas 3.3.7 and 3.3.8 gives the following formulation for finding the strongest valid lifted inequality at $\hat{x}$:

$$\begin{array}{ll} \text{maximize:} & 2\hat{x}^T\mu - \mu^T\mu + \rho \\ \text{subject to:} & \mu^T\mu - \rho \geq y^T\left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right)y - q_0h_0 + \beta c \\ & \left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right)y = \mu - \beta b + \frac{1}{2}\left(q_0 h + h_0 q\right) \\ & \left(I - \beta A + \frac{1}{2}\left(hq^T + qh^T\right)\right) \succeq 0 \\ & \mu^T\mu - \rho \geq ||\sigma q - \mu||^2 + 2\sigma q_0 \\ & \sigma \geq 0,\ \beta > 0 \end{array}$$

As in Section 3.2.4, we replace all occurrences of $\mu^T\mu - \rho$ in this formulation with a new variable τ to obtain a slightly simpler formulation:

$$\begin{aligned} \text{maximize:} \quad & 2\hat{x}^T\mu - \tau \\ \text{subject to:} \quad & \tau \geq y^T\left(I - \beta A + \tfrac{1}{2}\left(hq^T + qh^T\right)\right)y - q_0h_0 + \beta c \\ & \left(I - \beta A + \tfrac{1}{2}\left(hq^T + qh^T\right)\right)y = \mu - \beta b + \tfrac{1}{2}\left(q_0h + h_0q\right) \\ & \left(I - \beta A + \tfrac{1}{2}\left(hq^T + qh^T\right)\right) \succeq 0 \\ & \tau \geq \|\sigma q - \mu\|^2 + 2\sigma q_0 \\ & \sigma \geq 0,\ \beta > 0 \end{aligned} \tag{3.96}$$

The variables in this formulation are $(\mu, \tau, \beta, q, q_0, y, \sigma) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}$. This problem is nonconvex, due to q being multiplied by x and σ, q_0 being multiplied by σ, and x being multiplied by β. However, it is *biconvex* [40]: convex in μ, τ, β, q, and q_0 when y and σ are fixed, and convex in μ, τ, σ, and y when q, q_0 and β are fixed. We have experimented with solving this problem using an alternating scheme, optimizing over subsets of variables while holding others fixed, and have been successful in generating much stronger cuts than those obtained using Theorem 3.3.2. There is only a small amount of literature regarding biconvex programming, but we suspect that this particular problem may be aided by the fact that the obective is linear and the variables appearing in the objective never need to be fixed in the alternating scheme. In any case, we have the following result regarding the resolution of the separation problem in this setting:

Theorem 3.3.9. *Given a point $(\hat{x}, \hat{z})$ with $\hat{x}^TA\hat{x} - 2b^T\hat{x} + c < 0$ and $h^T\hat{x} \leq h_0$, then $(\hat{x}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ if and only if the optimal value to problem (3.96) is no greater than $\hat{z}$.*

Chapter 4

General Quadratics

In this chapter we relax the restriction that the function f is a positive definite quadratic, considering first a setting where f is semidefinite, and then showing how lifted inequalities can be used to tighten bounds in problems with indefinite quadratic constraints.

4.1 Semidefinite Objective

Here we consider the case where the f is a positive definite quadratic function of all variables except one. We introduce an extra variable w to denote the one left out by f, so the natural dimension is now $n+1$. We assume $\mathcal{F} \subset \mathbb{R}^{n+1}$ is the complement of the interior of a polyhedron which is explicitly unbounded in the direction w. The analysis here is similar to that of Section 3.1, but presents some differences which we will later see paralleled when we consider the case where f is an indefinite quadratic. Because the analysis is so similar, we skip over some of the details in the derivations in this section.

To begin, let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be given by

$$f(x, w) = x^T H x - 2h^T x$$

where H is an $(n \times n)$ positive definite matrix. Through a linear transformation, we can assume without loss of generality that $f(x, w) = x^T x$. Let $A \in \mathbb{R}^{m \times n}$ and define

$$P = \left\{(x, w) \in \mathbb{R}^{n+1} \mid a_i^T x - w \leq b_i,\ i = 1, \ldots, m\right\}.$$

The set $\mathcal{F}$ is defined as

$$\mathcal{F} = \mathbb{R}^{n+1} \setminus \operatorname{int}(P) \tag{4.1}$$

and the set $\mathcal{S}$ is

$$\mathcal{S} = \{(x, w, z) \in \mathbb{R}^{n+2} \mid z \geq x^T x,\ (x, w) \in \mathcal{F}\}. \tag{4.2}$$

In the remainder of this section we will show how strong lifted inequalities can be derived in this setting, exposing along the way the key differences between this and the positive definite case.

Take a point $(\bar{x}, \bar{w}) \in P$, with $a_i^T \bar{x} - \bar{w} = b_i$ for some fixed index $i \in \{1, \ldots, m\}$. The tangent inequality at $(\bar{x}, \bar{w})$, which holds for all $(x, w, z) \in \mathbb{R}^{n+2}$ with $z \geq x^T x$, is

$$z \geq \begin{bmatrix} 2\bar{x} \\ \bar{w} \end{bmatrix}^T \begin{bmatrix} x - \bar{x} \\ w - \bar{w} \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{w} \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{w} \end{bmatrix}. \tag{4.3}$$

It is easily shown that this inequality can only be lifted by adding a nonpositive multiple of the normal vector $(a_i, -1)$ to the gradient $2\bar{x}$ in (4.3). That is, lifted cuts in this setting must be of the form

$$z \geq \left(\begin{bmatrix} 2\bar{x} \\ \bar{w} \end{bmatrix} - \alpha \begin{bmatrix} a_i \\ -1 \end{bmatrix} \right)^T \begin{bmatrix} x - \bar{x} \\ w - \bar{w} \end{bmatrix} + \begin{bmatrix} \bar{x} \\ \bar{w} \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{w} \end{bmatrix}$$

with $\alpha \geq 0$ being the lifting coefficient, or more simply

$$z \geq (2\bar{x} - \alpha a_i)^T x + \alpha w + \alpha b_i - \bar{x}^T \bar{x}. \tag{4.4}$$

Consider a point (x, w, z) which (weakly) violates (4.4). Then the point $(x, w, x^T x)$ violates (4.4) as well, and we have

$$\begin{aligned} & x^T x \leq (2\bar{x} - \alpha a_i)^T x + \alpha(a_i^T \bar{x} - \bar{w}) + \alpha w \\ \Leftrightarrow \quad & x^T x - 2x^T \bar{x} + \bar{x}^T \bar{x} \leq \alpha b_i + \alpha w \\ \Leftrightarrow \quad & ||x - \bar{x}||^2 \leq \alpha b_i + \alpha w. \end{aligned} \tag{4.5}$$

The projection of the cut-off region to the (x, w) space is the interior of a *paraboloid* with a recession direction of $(0, 1) \in \mathbb{R}^{n+1}$. This is the expected analog to the cut off ball in previous settings where $\mathcal{S}$ was the epigraph of a strictly positive definite function.

Using the same approach as in Section 3.1.2, we can derive an expression for the lifting coefficient α giving the strongest lifted cut of the form (4.4). Let $j \neq i$ be another index in $\{1, \ldots, m\}$. The lifting coefficient for the j^{th} facet is the $\alpha \geq 0$ satisfying

$$\left\{ \begin{array}{rl} \text{minimize:} & x^T x - (2\bar{x} - \alpha a_i)^T x - \alpha w \\ \text{subject to:} & a_j^T x - w = b_j \end{array} \right\} = \alpha b_i - \bar{x}^T \bar{x} \tag{4.6}$$

Let (x^*, w^*) be the optimal solution to the minimization problem in (4.6), which is easily shown to be given by:

$$\begin{aligned} x^* &= \bar{x} - \frac{\alpha}{2} a_i + \frac{\alpha}{2} a_j \\ w^* &= a_j^T \bar{x} - b_j - \frac{\alpha}{2} a_j^T a_i + \frac{\alpha}{2} a_j^T a_j. \end{aligned}$$

Substituting these expressions into Equation (4.6) gives a quadratic equation the optimal α must satisfy:

$$\alpha \left(a_i^T \bar{x} - b_i - (a_j^T \bar{x} - b_j) \right) - \frac{1}{4} \alpha^2 \left(a_i^T a_i - 2 a_i^T a_j + a_j^T a_j \right) = 0. \tag{4.7}$$

$\alpha = 0$ is a trivial root of this equation, which gives $x^* = \bar{x}$ and $w^* = a_j^T \bar{x} - b_j$. With this choice of α, the set of violated points is simply the line $x = \bar{x}$, and no point (x, w) satisfies (4.5) with strict inequality - the lifted cut is no stronger than the tangent cut (4.3).

The second and nontrivial root is

$$\begin{aligned} \alpha &= \frac{4 \left(a_i^T \bar{x} - b_i - (a_j^T \bar{x} - b_j) \right)}{a_i^T a_i - 2 a_i^T a_j + a_j^T a_j} \\ &= \frac{4 \left(a_i^T \bar{x} - b_i - (a_j^T \bar{x} - b_j) \right)}{||a_i - a_j||^2}. \end{aligned} \tag{4.8}$$

Because $a_i^T \bar{x} - \bar{w} = b_i$ and $a_j^T \bar{x} - \bar{w} \leq b_j$, we have $a_i^T \bar{x} - b_i - (a_j^T \bar{x} - b_j) \geq 0$ and therefore $\alpha \geq 0$. Equation (4.8) gives the lifting coefficient for the j^{th} facet of P: the cut (4.4) using this choice of α is only guaranteed to be valid for

$$\{(x, w, z) \mid a_i^T x - w \geq b_i, z \geq x^T x\} \bigcup \{(x, w, z) \mid a_j^T x - w \geq b_j, z \geq x^T x\}$$

and not necessarily for all of $\mathcal{S}$. In order to find a valid cut, we must take the smallest of the lifting coefficients obtained by considering each facet separately:

$$\alpha = \min_{j \neq i} \left\{ \frac{4 \left(a_i^T \bar{x} - b_i - (a_j^T \bar{x} - b_j) \right)}{||a_i - a_j||^2} \right\}.$$

Formulating and solving the separation problem in this setting is essentially identical to the method presented in Section 3.1.3 so we do not present it here.

4.2 Indefinite Quadratics

Consider an optimization problem with an objective of

$$\text{minimize:} \quad x^T Qx + 2b^T x - c$$

where Q is a symmetric but indefinite matrix. Even with simple convex constraint sets, such problems are in general NP-hard; we would like to employ our cutting plane methodologies to strengthen bounds obtained using convex relaxations.

A common technique when working with indefinite quadratic forms is to express $x^T Qx$ as a difference of two convex quadratic forms (see [6] and [53] for more on difference-of-convex methods in nonconvex quadratic programming). By writing $Q = B - A$, where A and B are both positive definite, we can rewrite the problem as

$$\begin{aligned} \text{minimize:} \quad & z - w \\ \text{subject to:} \quad & z \geq x^T Bx \\ & w \leq x^T Ax - 2b^T x + c \end{aligned}$$

Similarly, we may consider a constraint of the form

$$x^T Qx + 2b^T x - c \geq 0.$$

A point x satisfies this constraint if and only if there exist $z, w \in \mathbb{R}$ with

$$\begin{gathered} z - w \geq 0 \\ z \geq x^T Bx \\ w \leq x^T Ax - 2b^T x + c. \end{gathered}$$

Either case leads to an interest in the set

$$\{(x, w, z) \in \mathbb{R}^{n+2} \mid z \geq x^T Bx,\ w \leq x^T Ax - 2b^T x + c\}.$$

By a change of coordinates we may assume that B is the identity matrix, and we therefore study the set

$$\mathcal{S} = \{(x, w, z) \in \mathbb{R}^{n+2} \mid z \geq x^T x,\ (x, w) \in \mathcal{F}\},$$

with $\mathcal{F}$ defined as

$$\mathcal{F} = \{(x, w) \in \mathbb{R}^{n+1} \mid w \leq x^T Ax - 2b^T x + c\}, \tag{4.9}$$

the complement of the interior of a paraboloid in the (x, w) space. This is similar to the case of an excluded ellipsoid studied in Section 3.3. After introducing a new parameterization of cuts, our main result in this section is the derivation of closed-form solutions to the separation problem. Again, the greatest computational task is computing the largest eigenvalue of A.

4.2.1 Paraboloid Inequalities

Consider any inequality

$$\delta z \geq \theta w + \gamma^T x + \beta \tag{4.10}$$

that is valid for $\mathcal{S}$. From the structure of $\mathcal{S}$, we must have $\delta \geq 0$. If $\delta = 0$, the inequality only involves the x and w variables and is valid for $\mathcal{F}$. We assume $\delta > 0$, and without loss of generality that $\delta = 1$. Next, suppose $\theta < 0$. For any fixed x, the point

$$(x, x^T Ax - 2b^T x + c - \epsilon, x^T x)$$

is in $\mathcal{S}$ for any $\epsilon \geq 0$. But then if $\theta < 0$, the inequality (4.10) would be violated for ϵ large enough. So we can assume $\theta \geq 0$.

Now consider a point $(x, w, x^T x)$ violating the inequality

$$z \geq \theta w + \gamma^T x + \beta \tag{4.11}$$

For this point, we have

$$\begin{aligned} x^T x < \theta^T w + \gamma^T x + \beta \quad &\Leftrightarrow \quad x^T x - \gamma^T x < \theta w + \beta \\ &\Leftrightarrow \quad \left|\left| x - \frac{1}{2}\gamma \right|\right|^2 < \theta w + \beta + \frac{1}{4}\left|\left|\gamma\right|\right|^2 \end{aligned}$$

The set of points (x, w) for which $(x, w, x^T x)$ violates (4.11) is a paraboloid. This suggests the following parameterization of cuts in this setting similar the that of Definition 3.2.9.

Definition 4.2.1. *For $(\mu, \alpha, \nu) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}$, we call the inequality*

$$z \geq 2\mu^T x + \alpha w - ||\mu||^2 - \alpha\nu \tag{4.12}$$

the paraboloid inequality *defined by* (μ, α, ν).

We let $\mathcal{P}(\mu, \alpha, \nu)$ denote the paraboloid

$$\{(x, w) \mid x^T x \leq 2\mu^T x + \alpha w - ||\mu||^2 - \alpha\nu\} = \{(x, w) \mid \ ||x - \mu||^2 \leq \alpha(w - \nu)\}$$

which gives the following characterization of valid paraboloid inequalities.

Proposition 4.2.2. *The paraboloid inequality defined by (μ, α, ν) is valid for $\mathcal{S}$ if and only if*

$$\operatorname{int}(\mathcal{P}(\mu, \alpha, \nu)) \cap \mathcal{F} = \varnothing.$$

The proof of this proposition is almost identical to Proposition 3.0.8 and is omitted. Notice that if $\alpha = 0$, then $\mathcal{P}(\mu, \alpha, \nu)$ is the line $\{(x, w) \mid x = \mu\}$, and has no interior. Therefore the paraboloid inequality defined by $(\mu, 0, \nu)$ is valid for any choice of μ and ν.

The following two lemmas demonstrate the equivalence between valid lifted inequalities for $\mathcal{S}$ and paraboloid inequalities.

Lemma 4.2.3. *Let the inequality*

$$z \geq \theta w + \gamma^T x + \beta \tag{4.13}$$

be valid for $\mathcal{S}$, and assume that this inequality is tight at some point $(\bar{x}, \bar{w}, \bar{x}^T \bar{x})$. Then (4.13) is a paraboloid inequality.

Proof. We previously established that if (4.13) is valid, θ must be nonnegative. First assume $\theta = 0$. The inequality (4.13) is violated at a point $(x, w, x^T x)$ if and only if

$$\left|\left|x - \frac{1}{2}\gamma\right|\right|^2 < \beta + \left|\left|\frac{1}{2}\gamma\right|\right|^2.$$

For any fixed x, the point $(x, x^T Ax - 2b^T x + c - \epsilon)$ is in $\mathcal{F}$ for any $\epsilon \geq 0$. So if (4.13) is valid, it must be the case that $\beta + \left|\left|\frac{1}{2}\gamma\right|\right|^2 \leq 0$. By the assumption about the point $(\bar{x}, \bar{w}, \bar{x}^T \bar{x})$, we have

$$\left|\left|\bar{x} - \frac{1}{2}\gamma\right|\right|^2 = \beta + \left|\left|\frac{1}{2}\gamma\right|\right|^2,$$

which, because $\left|\left|\bar{x} - \frac{1}{2}\gamma\right|\right|^2 \geq 0$ and (4.13) is valid, implies $\beta + \left|\left|\frac{1}{2}\gamma\right|\right|^2 = 0$. Then letting $\mu = \frac{1}{2}\gamma$, $\alpha = 0$, and $\nu = 0$ gives a paraboloid inequality equivalent to (4.13).

Next assume $\theta > 0$. In this case letting $\mu = \frac{1}{2}\gamma$, $\alpha = \theta$, and $\nu = -\frac{1}{\theta}\left(\beta + \left|\left|\frac{1}{2}\gamma\right|\right|^2\right)$ gives (4.13) as a paraboloid inequality. ■

Lemma 4.2.4. *Let (μ, α, ν) define a paraboloid inequality which is valid for $\mathcal{S}$ and supports $\mathcal{S}$ at some point $(\bar{x}, \bar{w}, \bar{x}^T\bar{x})$. Then this paraboloid inequality is either a tangent inequality or a lifted first-order inequality.*

Proof. This follows directly from Theorem 2.2.5. The paraboloid inequality is a nontrivial lifted inequality if and only if $\alpha > 0$. ■

Together with Theorem 2.2.6, Lemmas 4.2.3 and 4.2.4 imply that paraboloid inequalities are sufficient to solve the separation problem for a point $(\hat{x}, \hat{w}, \hat{z})$ with $(\hat{x}, \hat{w}) \in \operatorname{conv}(\mathcal{F})$, which is the focus of the next section.

4.2.2 The Separation Problem

We now show how to solve the separation problem in this setting. As in the case of the excluded ellipsoid, we will utilize the S-Lemma and derive closed-form expressions for the optimal parameters (μ^*, α^*, ν^*). The most costly operation required again will be the computation of the largest eigenvalue of the matrix A.

Let us suppose that we have a point $(\hat{x}, \hat{w}) \notin \mathcal{F}$ and that we wish to find the strongest lifted inequality at this point. Our objective is therefore

$$\text{maximize:} \quad 2\mu^T\hat{x} + \alpha\hat{w} - ||\mu||^2 - \alpha\nu,$$

subject to the parameters (μ, α, ν) defining a paraboloid inequality that is valid for $\mathcal{S}$.

We first derive a characterization of validity which will aid in the formulation of the separation problem.

Lemma 4.2.5. *The paraboloid inequality defined by (μ, α, ν) is valid for $\mathcal{S}$ if and only if*

$$\min_{x \in \mathbb{R}^n} \{x^T(I - \alpha A)x - 2(\mu - \alpha b)^T x\} + \mu^T\mu + \alpha(\nu - c) \geq 0.$$

Proof. By Proposition 4.2.2, the paraboloid inequality is valid if and only if

$$||x - \mu||^2 + \alpha\nu - \alpha w \geq 0 \text{ for all } (x, w) \text{ with } w - \left(x^T Ax - 2b^T x + c\right) \leq 0.$$

By the S-Lemma, this is equivalent to the existence of a nonnegative multiplier θ such that

$$||x - \mu||^2 + \alpha\nu - \alpha w + \theta w - \theta\left(x^T Ax - 2b^T x + c\right) \geq 0 \quad \forall (x, w) \in \mathbb{R}^{n+1}.$$

If this is to hold, it must be the case that $\theta = \alpha$. If not, then letting w tend to $\pm\infty$ (depending on the sign of $(\theta - \alpha)$) would violate the condition. Thus validity can be expressed as

$$||x - \mu||^2 + \alpha\nu - \alpha\left(x^T Ax - 2b^T x + c\right) \geq 0 \quad \forall x \in \mathbb{R}^n.$$

or equivalently

$$\min_{x \in \mathbb{R}^n} \{x^T(I - \alpha A)x - 2(\mu - \alpha b)^T x\} + \mu^T\mu + \alpha(\nu - c) \geq 0. \tag{4.14}$$

■

Next we use conditions for the boundedness of the minimization problem in (4.14) to obtain a condition for validity which does not contain any auxiliary optimization problems, and a formulation of the separation problem.

Lemma 4.2.6. *The separation problem can be formulated as*

$$\begin{aligned}
\textit{maximize:} \quad & 2\mu^T\hat{x} + \alpha\hat{w} - ||\mu||^2 - \alpha\nu \\
\textit{subject to:} \quad & -\pi^T(I - \alpha A)\pi + \mu^T\mu + \alpha(\nu - c) \geq 0 \\
& \mu = \alpha b + (I - \alpha A)\pi \\
& 0 \leq \alpha \leq \bar{\lambda}^{-1}
\end{aligned} \tag{4.15}$$

Proof. For the inequality (4.14) to hold, the value of the minimization problem on the left hand side must be bounded. Define $\bar{\lambda}$ to be the largest eigenvalue of A. By Lemma A.2.1, the optimal value of the minimization problem is finite if and only if the following two conditions are satisfied:

$$\alpha \leq \bar{\lambda}^{-1} \tag{4.16}$$

$$\mu = \alpha b + (I - \alpha A)\pi \;\; \text{for some } \pi \in \mathbb{R}^n \tag{4.17}$$

With these conditions imposed, the optimal solution to the minimization problem is any x^* satisfying

$$(I - \alpha A)x^* = \mu - \alpha b = (I - \alpha A)\pi. \tag{4.18}$$

Thus π is an optimal solution, with a resulting optimal value of $-\pi^T(I - \alpha A)\pi$. Making this substitution into (4.14) and adding the constraints (4.16) and (4.17) gives the formulation (4.15).

■

Notice that the formulation (4.15) is not convex, due to α multiplying other variables in the objective as well as in the first two constraints. The formulation *is* convex, however, for any *fixed* choice of α. We use this fact to show that the problem (4.15) can actually be solved in closed form, which gives expressions for the parameters for the optimal paraboloid inequality.

Theorem 4.2.7. *Given a point $(\hat{x}, \hat{w}, \hat{z})$ with $(\hat{x}, \hat{z}) \notin \mathcal{F}$, the parameters for the strongest paraboloid inequality at $(\hat{x}, \hat{z})$ are*

$$\alpha^* = \bar{\lambda}^{-1}$$
$$\mu^* = \alpha^* b + (I - \alpha^* A)\hat{x}$$
$$\nu^* = \hat{x}^T A \hat{x} - 2b^T \hat{x} + c - \frac{1}{\alpha^*} ||\mu^* - \hat{x}||^2$$

where $\bar{\lambda}$ is the largest eigenvalue of A.

Proof. First we will show that we can assume that the first constraint in (4.15) will hold with equality. Assume that α is fixed at any value in the interval $[0, \bar{\lambda}]$. Because $\alpha \geq 0$, the objective value in (4.15) is nonincreasing with ν. Also, as this is the only constraint in which ν appears, it is always possible to decrease ν until the constraint is tight without decreasing the objective value. We can therefore eliminate the first constraint by substituting

$$\mu^T \mu + \alpha \nu = \alpha c + \pi^T (I - \alpha A)\pi \tag{4.19}$$

into the objective. Then, with the additional substitution

$$\mu = \alpha b + (I - \alpha A)\pi \tag{4.20}$$

we can eliminate μ and the second constraint, and arrive at the following reformulation:

$$\begin{aligned} \text{maximize:} \quad & -\pi^T (I - \alpha A)\pi + 2\hat{x}^T (I - \alpha A)\pi + 2\alpha b^T \hat{x} - \alpha c + \alpha \hat{w} \\ \text{subject to:} \quad & 0 \leq \alpha \leq \bar{\lambda}^{-1} \end{aligned}$$

By Lemma A.2.1, for any fixed feasible α, the optimal solution to this problem is any π^* satisfying

$$-2(I - \alpha A)\pi^* + 2(I - \alpha A)\hat{x} = 0.$$

So $\pi^* = \hat{x}$ is optimal and results in an objective value of

$$\hat{x}^T \hat{x} - \alpha(\hat{x}^T A \hat{x} - 2b^T \hat{x} + c - \hat{w}).$$

Because $(\hat{x}, \hat{w}) \notin \mathcal{F}$ by assumption, we have $\hat{x}^T A\hat{x} - 2b^T\hat{x} + c < \hat{w}$, and this objective value is increasing in α. Therefore α^* will be at the upper bound of $\bar{\lambda}^{-1}$. The equality (4.20) gives the expression for μ^*, and the equality (4.19) gives the expression for ν^*. ■

It is easily shown that the optimal parameters derived in the proof of Theorem 4.2.7 produce a cut which supports $\mathcal{S}$ at some point, implying that this cut is a lifted inequality. The optimal parameters also provide a solution to the separation problem.

Corollary 4.2.8. *Let $(\hat{x}, \hat{w}, \hat{z}) \in \mathbb{R}^{n+2}$ with $(\hat{x}, \hat{w}) \notin \mathcal{F}$. Define α^*, μ^*, and ν^* as in the statement of theorem of Theorem 4.2.7. Then $(\hat{x}, \hat{w}, \hat{z}) \in \operatorname{conv}(\mathcal{S})$ if and only if*

$$2\mu^{*T}\hat{x} + \alpha^*\hat{w} - ||\mu^*||^2 - \alpha^*\nu^* \leq \hat{z}.$$

Also, analogous to Corollary 3.3.5, we obtain a finite description of $\operatorname{conv}(\mathcal{S})$.

Corollary 4.2.9. *Let $\bar{\lambda}$ be the largest eigenvalue of A, and define $\alpha = \bar{\lambda}^{-1}$. Then*

$$conv(\mathcal{S}) = \{(x, w, z) \in \mathbb{R}^{n+2} \mid z \geq x^T x,\ z \geq x^T x - \alpha(x^T Ax - 2b^T x + c - w)\}.$$

The proof of this statement is essentially identical to that of Corollary 3.3.5 and is omitted.

Chapter 5

Numerical Experiments

To validate the computational practicality of our separation procedures, we ran several numerical experiments on randomly generated test problem instances. All computations were performed on an 8-core i7 computer, with 48 GB of physical memory. Unless otherwise noted, all optimization problems were solved using Gurobi 5.50 [41] through the GurobiPy interface.

5.1 A Comparison With the Disjunctive Method

In Section 3.1.5 we discussed some potential advantages of the separation method derived in Section 3.1.3 over the disjunctive approach. In this section we present the results of a direct comparison of the two methods. In our comparison, we focus on the use of cutting planes to derive lower bounds on problems of the form $\min\{x^T x \mid Ax \not< b\}$. We point out that solving such problems is trivial: the optimal solution is found by computing the projection of the origin onto each hyperplane $\{x \mid a_j^T x = b_j\}$ and comparing the norms of the resulting vectors. Our interest is not in using the lifting techniques to solve this problem, but instead to compare the efficacy of generating lifted cuts versus using the disjunctive approach.

Each experiment was conducted as follows. First, a random polyhedron $\{x \mid Ax \leq b\}$ was generated, using the following method: The entries of the current normal vector a_i were set to random uniform values between -1 and 1, and then each was set to 0 with probability 0.5. The vector was rejected if it was a positive multiple of any of the previous vectors $a_1, \ldots, a_{i-1}$. a_i was then scaled to have unit norm and the entries were rounded to three digits. Next, the value $\bar{b}_i$ was calculated as

$$\bar{b}_i = \max\{a_i^T x \mid a_j^T x \leq b_j,\ j = 1, \ldots, i-1\}.$$

If $\bar{b}_i$ was finite, b_i was set to a value randomly distributed between $0.5\bar{b}_i$ and $0.95\bar{b}_i$. Otherwise b_i was set to $1+\Gamma$, where Γ was a generated randomly from a gamma distribution with shape $\sqrt{n}$ and scale $0.5\sqrt{n}$. This somewhat complicated procedure was intended to ensure that the inequalities in the system $Ax \leq b$ were all facet-defining. In either case b_i was then rounded to three digits. Next, lower bounds were computed using this algorithm:

Algorithm 5.1.1.

1. Start with the initial feasible set $\mathcal{F} = \{(x, z) \mid z \geq x^T x\}$.
2. Solve the current problem: $\min\{z \mid (x, z) \in \mathcal{F}\}$, and let $(\hat{x}, \hat{z})$ be the optimal solution.
3. Compute an inequality $z \geq \gamma^T x + \beta$ separating $(\hat{x}, \hat{z})$ from $\mathcal{S}$.
4. If the stopping conditions are met, stop. Otherwise, set

$$\mathcal{F} = \mathcal{F} \cap \{(x, z) \mid z \geq \gamma^T x + \beta\}$$

and go to Step 2.

The algorithm was stopped when 500 cuts had been added, when the relative gap between the lower bound and the true value of the problem was less than a tolerance of 10^{-5}, or when more than 600 seconds had passed since the cutting plane procedure started. When generating lifted inequalities, we used the heuristic of only computing the lifted inequality from the facet closest to the current solution $\hat{x}$.

Table 5.1 displays the results from this comparison. The columns marked n and m give the dimension and number of inequalities in the system $Ax \leq b$. val gives the true solution value to the problem. Lo_l, $Time_l$, and $Cuts_l$ give the lower bound proved, the total amount of time spent in the cutting plane algorithm, and the number of cuts added when using lifted inequalities to solve the separation problem. Lo_d, $Time_d$, and $Cuts_d$ give the corresponding information when the disjunctive method was used to generate cuts. An asterisk next to an entry in the $Time_d$ column indicates that the algorithm was stopped early because at some point the disjunctive method was unable to compute the dual variables giving the cut coefficients. From this table we see that lifted

Table 5.1: Comparison with disjunctive method

n	m	val	Lo_l	Lo_d	$Time_l$	$Time_d$	$Cuts_l$	$Cuts_d$
10	50	5.191	5.190	2.121	0.2	74.7	51	500
20	100	14.537	14.536	0.366	0.1	88.1	15	500
20	300	17.831	17.830	0.771	0.3	311.4	15	500
50	200	79.888	79.887	0.154	0.3	381.5	20	500
75	250	343.897	336.353	36.872	126.7	317.1*	500	95
100	300	324.486	324.485	16.139	0.6	126.7*	14	10
200	400	2207.060	2207.038	0.000	8.7	91.0*	92	0
300	500	4583.748	4583.733	0.000	2.4	155.3*	20	0
800	1200	38142.592	38142.243	0.000	32.1	1879.1*	42	0
1000	2000	61726.150	61725.542	0.000	227.8	3304.3*	134	0

inequalities enjoy a clear advantage over the disjunctive method. In six of the ten test cases, the disjunctive method was unable to compute dual variables and was forced to stop early. Even in the smaller cases, the disjunctive method ran into numerical difficulties and reached the maximum number of cuts with a much weaker bound than the one obtained using lifted inequalities.

5.2 Comparing Cutting Plane Variants

We also experimented with variants of Algorithm 5.1.1 in which the quadratic constraint $z \geq x^T x$ was not included in the convex relaxations. Instead, the algorithm was initialized with $\mathcal{F} = \{(x, z) \mid z \geq 0\}$ and tangent inequalities were used to approximate the constraint $z \geq x^T x$. In these tests, a polyhedron $\{x \mid Ax \leq b\}$ was generated randomly using the same method as before, and the objective was $x^T Q x$, with Q being a randomly generated positive definite matrix. The initial feasible set was simply

$$\{(x, z) \in \mathbb{R}^{n+1} \mid z \geq 0\}.$$

When a relaxed solution $(\hat{x}, \hat{z})$ was obtained, the tangent inequality

$$z \geq 2\,(Q\hat{x})^T\,(x - \hat{x}) + \hat{x}^T Q \hat{x}$$

was added to the formulation, in addition to the strongest lifted inequality at $\hat{x}$, in the case where $A\hat{x} < b$. Without the quadratic constraint, every relaxation solved is a linear program.

Table 5.2 presents results on the variants we tested. The most basic version (whose results are in the columns marked "Basic") uses the method described above. The second version (in the columns marked "Heuristics") incoroprates the following three heuristics:

1. Before starting, the tangent cut was added at each unit vector e_i as well as $-e_i$.

2. Before starting, the lifted inequality at the point closest to the origin on each facet of $\{x \mid Ax \leq b\}$ was added, if possible.

3. The constraint $Ax \leq b$ was added in the relaxation.

Finally, for comparison purposes, we tested a version of the algorithm which included the quadratic constraint $z \geq x^T Qx$ and used neither tangent cuts nor the heuristics mentioned above. Table 5.3 shows the results from this variant, in the columns marked "Full".

These tests were terminated if, between subsequent iterations, the objective value z and all entries of the solution x were within a tolerance of 10^{-3} of the previous values. In this test, each method was limited to 30 minutes to add cuts, and maximum of $10,000$ iterations were performed.

Table 5.2: Comparison of cutting plane variants

			Basic				**Heuristics**			
n	**m**	**val**	z_{lo}	**lin**	**lfo**	**t**	z_{lo}	**lin**	**lfo**	**t**
10	50	5.2	5.2	286	120	0.8	5.2	280	212	1.1
20	100	14.5	14.5	2104	824	19.5	14.5	890	744	8.6
20	300	17.8	17.8	1903	577	28.7	17.8	2831	13	34.0
50	200	79.9	0.0	10000	1	688.0	79.9	10300	38	646.2
75	250	343.9	0.0	10000	1	768.6	343.9	8198	7841	1803.9
100	300	324.5	0.0	10000	1	1127.2	320.6	6780	6361	1804.1
200	400	2207.1	0.0	269	1	1807.1	2207.1	6168	151	1803.3
300	500	4583.74	0.0	301	1	13.3	597.6	2938	2003	1803.3
800	1200	38142.6	0.0	302	1	66.7	569.5	2940	651	1811.4
1000	2000	61726.1	0.0	246	1	108.7	61726.1	4001	893	2330.6

In these tables, n and m give the dimension of number of facets in the excluded polyhedron, and val gives the true optimal value of the problem (obtained by solving the problem $\min\{x^T Qx \mid a_j^T x \geq b_j\}$ for each j). For each of the three methods $z_l o$ gives the final lower bound, lfo and lin give the number of lifted and tangent inequalities generated during the course of the

Table 5.3: Including the quadratic constraint

			Full		
n	**m**	**val**	z_{lo}	**lfo**	**t**
10	50	5.2	5.2	133	1.1
20	100	14.5	14.5	33	0.3
20	300	17.8	17.8	26	0.6
50	200	79.9	79.9	28	0.6
75	250	343.9	339.5	1370	1802.3
100	300	324.5	324.5	25	1.1
200	400	2207.1	2207.0	87	9.6
300	500	4583.74	4583.74	24	3.7
800	1200	38142.6	38142.5	55	53.9
1000	2000	61726.1	57689.3	18	40.0

procedure, and t gives the total time taken. We see that the Basic method performs poorly for larger instances. In this implementation, we observed that the magnitude of the entries of the linear program relaxations tend to become quite large, creating numerical difficulties. This was the rationale for adding the third heuristic. The implementation which includes the quadratic constraint gives significantly better performance in all instances except the last.

5.3 Testing the Dual Formulation

In order to evaluate the performance of the "dual" separation formulation of Section 3.2.4, we performed an experiment similar to that in the previous section. Instead of a single polyhedron, either two or three polyhedra $\{x \mid A_i x \leq b_i\}$ were randomly generated using the method described in the previous section. The feasible region $\mathcal{F}$ was defined as $\mathcal{F} = \{x \mid A_i x \not< b_i\}$, and polyhedra $\{x \mid C_k x \geq d_k\}$ were computed as described in the beginning of Section 3.2 so that

$$\mathcal{F} = \bigcup_{k=1}^{K} \{x \mid C_k x \geq d_k\}.$$

A cutting plane algorithm was then used to prove bounds on the problem $\min\{x^T x \mid x \in \mathcal{F}\}$. We followed the same cutting plane procedure as in the previous section, with the exception of the time limit, which was changed from ten minutes to 30 minutes. Algorithm 3.2.16 from Section 3.2.5 was used to compute the separating inequalities at the current relaxed solution $(\hat{x}, \hat{z})$. The initial μ

was chosen to be $\hat{x}$, and initially $n+1$ constraints were enforced. At each iteration, the number of enforced constraints was increased by 10. The constraints which had been enforced and were tight in the previous solution were kept active, and the rest of the active set was formed by choosing the constraints most violated with the current values of μ, τ and λ_k. In the first iteration of the cutting plane algorithm, the variables $\{\lambda_k\}_{k=1}^K$ were initialized from scratch by solving the problem (3.64) for each k. If this problem was unbounded for a certain k, the corresponding polyhedron (which is infeasible) was removed from the formulation. After the first iteration, the variables $\{\lambda_k\}_{k=1}^K$ were stored and used as initial values in the computation of the next cut. We found that this makes a significant difference. Figure 5.1 shows the time taken for the first 100 cuts for an instance in $n = 20$ dimensions with $K = 62{,}500$ component polyhedra. The first cut takes about 108 seconds to compute, while subsequent cuts take an average of 8 seconds.

Figure 5.1: Time per cut decreases after the first.

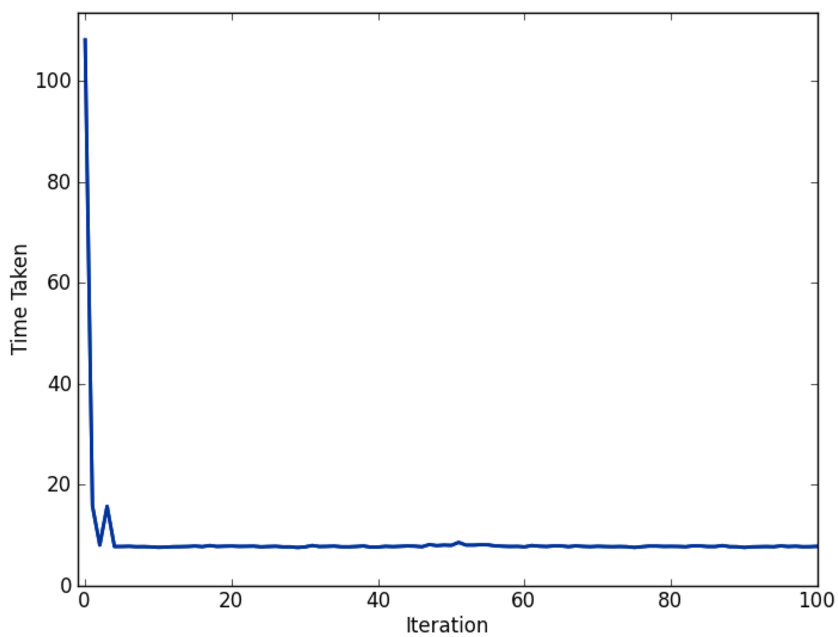


Table 5.4 displays the results from this experiment. n gives the dimension of the problem. The tuple in the column labled m gives the number of rows in the matrices A_i. Depending on this test instance, there are either two or three values for m. K is the total number of component polyhedra defining $\mathcal{F}$, which is the product of the values in m. *val* gives the true value of the problem, obtained by enumerating all of the component polyhedra. *Bound*, *Cuts*, and *Time* give

the lower bound achieved by the cutting plane algorithm, the number of cuts required to achieve this bound, and the total time spent by the algorithm.

Table 5.4: Results from the dualized method

n	m	K	val	Bound	Cuts	Time
10	(40, 50)	2000	7.00	7.00	14	7.97
15	(100, 200)	20000	22.06	22.06	41	141.72
20	(250, 250)	62500	29.51	29.49	217	1806.87
20	(300, 300)	90000	23.85	23.85	131	1619.73
20	(30, 40, 50)	60000	52.38	51.07	121	1805.32
50	(200, 300)	60000	134.17	134.17	47	1017.26
50	(150, 500)	75000	123.36	123.36	80	1810.98
75	(30, 30, 50)	45000	755.13	680.49	16	1808.21
100	(200, 400)	80000	460.12	423.22	7	1878.55
200	(250, 250)	62500	2889.63	2019.59	1	2196.29

As the table shows, Algorithm 3.2.16 is able to generate cuts reasonably quickly for smaller instances, but becomes much slower as the dimension increases. For the largest instance, the algorithm is only able to compute a single cut. We point out that for these test instances (with the exception of the first), the "primal" disjunctive formulation is completely unable to run - the number of variables and constraints is simply too large. We also point out, again, that the problem $\min\{x^T x \mid x \in \mathcal{F}\}$ is easily solved by solving a quadratic program over each polyhedron $\{x \mid C_k x \geq d_k\}$. Our goal in this experiment is to show that the iterative algorithm and the dual formulation make the computation of cutting planes tractable in cases where the disjunctive method fails.

5.3.1 Heuristic Improvements

We experimented with several modifications to help speed up the progress of Algorithm 3.2.16. First was the number of initial active constraints, and the way the set of active constraints was updated in each iteration. Noticing that a large number of active constraints made solving the relaxed problem in Step 2 very slow, and that far fewer than n constraints ever held with equality in the relaxed problems, we decreased the size of the initial active set to the larger of 10 and $0.1n$. We also modified the update rule in Step 4. If we found a relaxed solution violated any inactive constraints, we added the larger of 5 and $0.05n$ constraints to the active set, unless the number of violated constraints was smaller than this value, in which case this number of constraints were added.

The second modification was in the initialization and updates of the variables $\{\lambda_k\}_{k=1}^K$. Solving the problem

$$\begin{aligned} \text{maximize:} \quad & -\left|\left|C_k^T \lambda_k + \mu\right|\right|^2 + 2\lambda_k^T d_k \\ \text{subject to:} \quad & \lambda_k \geq 0 \end{aligned} \tag{5.1}$$

for many k was time consuming. We implemented the following heuristic update rule: given a current vector λ_k^{old} and the current values for μ and τ, we maximized the function

$$g(\sigma) = -\left|\left|C_k^T \left(\sigma \lambda_k^{old}\right) + \mu\right|\right|^2 + 2\left(\sigma \lambda_k^{old}\right)^T d_k$$

over σ (equivalent to maximizing the objective of (5.1) over the line defined by the origin and λ_k^{old}). Because $g(\sigma)$ is quadratic in σ, this could be done easily in closed form. Letting σ^* denote the optimal choice of σ, then if the vector $\sigma^* \lambda_k^{old}$ was feasible and satisfied the k^{th} constraint:

$$-\left|\left|C_k^T \left(\sigma^* \lambda_k^{old}\right) + \mu\right|\right|^2 + 2\left(\sigma^* \lambda_k^{old}\right)^T d_k + \tau \geq 0, \tag{5.2}$$

we set $\lambda_k^{new} = \sigma^* \lambda_k^{old}$. Otherwise, λ_k^{new} was set to the optimal solution of (5.1). In the intialization step, before λ_k^{old} was available, we used the vector of all 1's as a substitute for λ_k^{old}.

The final modification was in the size of the initial set of active constraints. Instead of always starting from the same value in the computation of each cut, the size of the active set was maintained for all cuts after the first, meaning it could only increase as more and more cuts were computed.

Table 5.5 shows the results after the addition of these modifications. We found that the performance was similar (and sometimes slightly worse) than the original version for smaller problems, but markedly better for larger instances.

Table 5.5: Results after adding heuristics

n	m	K	val	Bound	Cuts	Time
10	(40, 50)	2000	7.00	7.00	13	5.56
15	(100, 200)	20000	22.06	22.06	34	127.53
20	(250, 250)	62500	29.51	29.43	173	1808.89
20	(300, 300)	90000	23.85	23.85	118	1807.88
20	(30, 40, 50)	60000	52.38	50.18	100	1827.76
50	(200, 300)	60000	134.17	134.17	27	360.45
50	(150, 500)	75000	123.36	123.36	36	686.10
75	(30, 30, 50)	45000	755.13	719.83	33	1829.78
100	(200, 400)	80000	460.12	460.12	77	1477.92
200	(250, 250)	62500	2889.63	2702.21	7	2016.18

Chapter 6

Conclusion and Future Work

In this work, we have studied the use of lifted inequalities to tighten relaxations on nonconvex optimization problems. We have shown that the class of lifted first-order inequalities of the form

$$z \geq (\nabla f(y) + \alpha\lambda)^T (x - y) + f(y)$$

is sufficient to obtain the convex hull of sets of the form

$$\mathcal{S} = \{(x, z) \in \mathbb{R}^{n+1} \mid z \geq f(x),\ x \in \mathcal{F}\}$$

when f is convex and differentiable. In the case where f is a positive definite quadratic function, we have proven polynomial time methods for finding separating from $\operatorname{conv}(\mathcal{S})$ when $\mathcal{F}$ is the union of polyhedra or the complement of the interior of an ellipsoid by means of lifted inequalities. We have also shown how our methods may be applied in some cases when f is a positive semidefinite quadratic, and how they may be used to tighten constraints on indefinite quadratic functions. We have presented the results of numerical experiments demonstrating the computational tractability and efficacy of our methods, including solving large problem instances completely out of the reach of the current state of the art.

Our development of separation procedures and characterizations of valid lifted inequalities focused on quadratic objective functions. As mentioned in the introduction, this was partly due to the analytical tractability that the geometry of quadratic objectives exhibited. While there is certainly no shortage of applications with quadratic objectives, research into separation procedures and characterizations of strong valid lifted inequalities for other classes of objective functions would be of great interest.

Our computational experiments were successful, if somewhat limited, and the implementations used in these experiments were not especially sophisticated. Although the theory presented in Chapters 2, 3, and 4 is elegant and interesting in its own right, this work was motivated by problems encountered in applications, and its true merit may be best judged by its utility in practice. We welcome further testing and validation of our methods through application to specific "real" problems.

Bibliography

[1] A.J. Ahmadia, Parallel Strategies for Nonlinear Mask Optimization in Semiconductor Lithography, PhD Thesis, Columbia University (2010).

[2] F. Alizadeh and D. Goldfarb, Second-Order Cone Programming, *Mathematical Programming* **95** (2001), 3 – 51.

[3] I. P. Androulakis, C. D. Maranas, C. A. Floudas, αBB: A Global Optimization Method for General Constrained Nonconvex Problems, *Journal of Global Optimization* **7** (1995), 337 – 363.

[4] K. M. Anstreicher, Semidefinite Programming Versus the Reformulation Linearization Technique for Nonconvex Quadratically Constrained Quadratic Programming, *Journal of Global Optimization* **43** (2009), 471 – 484.

[5] K. M. Anstreicher, On Convex Relaxations for Quadratically Constrained Quadratic Programming, *Mathematical Programming B* **136** (2012), 233 – 251.

[6] K. M. Anstreicher and S. Burer, D.C. Versus Copositive Bounds for Standard QP, *Journal of Global Optimization* **33** (2005), 299 – 312.

[7] A. Atamtürk and V. Narayanan, Lifting for Conic Mixed-Integer Programming, *Mathematical Programming* **126** (2011), 351 – 363.

[8] E. Balas, Disjunctive Programming: Properties of the Convex Hull of Feasible Points, MSRR No. 348, Carnegie Mellon University, Pittsburgh, PA (1974).

[9] E. Balas, S. Ceria, and G. Cornuéjols, A Lift-and-Project Cutting Plane Algorithm for Mixed 0-1 Programs, *Mathematical Programming* **58** (1993), 295 – 324.

[10] E. Balas and R. Jeroslow, Canonical Cuts on the Unit Hypercube, *SIAM Journal on Applied Mathematics* **23** (1972), 61 – 69.

[11] A. Ben-Tal and A. Nemirovsky, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (2001) MPS-SIAM Series on Optimization, SIAM, Philadelphia, PA.

[12] 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, *Optimization Online*, (2012).

[13] P. Belotti, J. Lee, L. Liberti, F. Margot and A. Wachter, Branching and Bounds Tightening Techniques for Non-convex MINLP, *Optimization Methods and Software* **24** (2009), 597 - 634.

[14] P. Belotti, A.J. Miller and M. Namazifar, Valid Inequalities and Convex Hulls for Multilinear Functions, *Electronic Notes in Discrete Mathematics* **36** (2010), 805 - 812.

[15] D. Bienstock, Computational Study of a Family of Mixed-Integer Quadratic Programming Problems, *Mathematical Programming* **74** (1996), 121 – 140.

[16] D. Bienstock, Eigenvalue Techniques for Proving Bounds for Convex Objective, Nonconvex Programs, *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science* **6080** (2010), 29 – 42.

[17] D. Bienstock and A. Michalka, Strong Formulations for Convex Functions over Nonconvex Sets, manuscript, (2011).

[18] P. Bonami, Lift-and-Project Cuts for Mixed Integer Convex Programs, *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science* **6655** (2011), 52 – 64.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004).

[20] C. Buchheim, A. Caprara and A. Lodi, An Effective Branch-and-Bound Algorithm for Convex Quadratic Integer Programming, *Lecture Notes in Computer Science* **6080** (2010), 285 – 298.

[21] S. Burer and K. M. Anstreicher, Second-Order Cone Constraints for Extended Trust-Region Subproblems *SIAM Journal on Optimization* (to appear).

[22] S. Burer and A. N. Letchford, On Non-convex Quadratic Programming with Box Constraints, *SIAM Journal on Optimization*, **20** (2009), 1073 – 1089.

[23] S. Burer and A. N. Letchford, Non-Convex Mixed-Integer Nonlinear Programming: A Survey, *Optimization Online*, (2012).

[24] S. Burer and B. Yang, The Trust Region Subproblem with Non-Intersecting Linear Constraints, manuscript, (2013).

[25] M. R. Celis, J. E. Dennis, and R. A. Tapia, A Trust Region Strategy for Nonlinear Equality Constrained Optimization, *Numerical Optimization*, SIAM, Philadelphia, (1985), 71 – 82.

[26] S. Ceria and J. Soares, Convex Programming for Disjunctive Convex Optimization, *Mathematical Programming* **86** (1999), 595 - 614.

[27] V. Chvátal, Edmonds Polytopes and a Hierarchy of Combinatorial Problems, *Discrete Mathematics* **4** (1999), 305 - 337.

[28] G. Cornuéjols, Valid Inequalities for Mixed Integer Linear Programs, *Mathematical Programming* **112** (2007), 3 – 44.

[29] D. Dadush, S. Dey, and J. P. Vielma On the Chvátal-Gomory Closure of a Compact Convex Set, *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science* **6655** (2011), 130 – 142.

[30] C. D'Ambrosio, A. Frangioni, L. Liberti, A. Lodi, On Interval-subgradient and No-good Cuts, *Operations Research Letters* **38** (2012) 341 – 345.

[31] S. Dash, S. S. Dey, and O. Günlük Two Dimensional Lattice-free Cuts and Asymmetric Disjunctions for Mixed-integer Polyhedra, *Mathematical Programming A* **135** (2012), 221 – 254.

[32] L. L. Dines, On the Mapping of Quadratic Forms, *Bulletin of the American Mathematical Society* **47** (1941), 494 – 498.

[33] M. Fischetti, A. Lodi, and A. Tramontani, On the Separation of Disjunctive Cuts, *Mathematical Programming A* **128** (2011), 205 – 230.

[34] T. Fujie and M. Kojima, Semidefinite Programming Relaxation for Nonconvex Quadratic Programs, *Journal of Global Optimization* **10** (1997), 367 – 380.

[35] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, Convex Optimization-based Beamforming: From Receive to Transmit and Network Designs, *IEEE Signal Processing Magazine* (2010), 62 – 75.

[36] M. X. Goemans, Semidefinite Programming in Combinatorial Optimization, *Mathematical Programming* **79** (1997), 143 – 161.

[37] R. E. Gomory, Outline of an Algorithm for Integer Solutions to Linear Programs, *Bulletin of the American Mathematical Society* **64** (1958), 275 – 278.

[38] R. E. Gomory, An Algorithm for Integer Solutions to Linear Programs, *Recent Advances in Mathematical Programming*, McGraw-Hill, New York (1963), 269 – 302.

[39] R. E. Gomory, Some Polyhedra Related to Combinatorial Problems, *Linear Algebra and its Applications* **2** (1979), 451 – 558.

[40] J. Gorski, F. Pfeuffer, and K. Klamroth, Biconvex Sets and Optimization with Biconvex Functions: a Survey and Extensions, *Mathematical Methods of Operations Research* **66(3)** (2007), 373 – 407.

[41] Gurobi Optimizer. http://www.gurobi.com/.

[42] G. Iyengar and M. Çezik Cut Generation for Mixed 0-1 Quadratic Programming, *Columbia University Technical Report TR-2001-02* (2001).

[43] G. Iyengar and M. Çezik Lift-and-Project Cutting Planes for Mixed 0-1 Semidefinite Programming, *Columbia University Technical Report TR-2001-03* (2001).

[44] M. Kilinc, J. Linderoth and J. Luedtke, Effective Separation of Disjunctive Cuts for Convex Mixed Integer Nonlinear Programs, *Optimization Online* (2010).

[45] Y. Li and J.-P. P. Richard, Cook, Kannan and Schrijvers Example Revisited, *Discrete Optimization* **5** (2008) 724 – 734.

[46] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, Applications of Second-Order Cone Programming, *Linear Algebra and its Applications* **284** (1998) 193 – 228.

[47] Z.-C. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S.Zhang, Semidefinite Relaxation of Quadratic Optimization Problems, *IEEE Signal Processing Magazine*, **27** (2010) 20 – 34.

[48] S. Modaresi, M. R. Kilinç, and J. P. Vielma, Intersection Cuts for Nonlinear Integer Programming: Convexification Techniques for Structured Sets, manuscript, (2013).

[49] G. Nannicini and P. Belottti, Rounding-based Heuristics for Nonconvex MINLPs, *Mathematical Programming Computation* **4** (2012), 1 – 31.

[50] G. L. Nemhauser and L.A. Wolsey, *Integer and Combinatorial Optimization*, Wiley, New York (1988).

[51] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Methods in Convex*, Programming. Society for Industrial and Applied Mathematics, 1994.

[52] M. W. Padberg, On the Facial Structure of Set Packing Polyhedra, *Mathematical Programming* **5** (1973), 199 – 215.

[53] P. D. Tao and L. T. H. An, A D.C. Optimization Algorithm for Solving the Trust-Region Subproblem, *SIAM Journal on Optimization* **8** (1998), 476 – 505.

[54] I. Pólik and T. Terlaky, A Survey of the S-lemma, *SIAM Review* **49** (2007), 371 – 418.

[55] A. Qualizza, P. Belotti, and F. Margot, Linear Programming Relaxations of Quadratically Constrained Quadratic Programs, *Mixed Integer Nonlinear Programming: The IMA Volumes in Mathematics and its Applications* **154** (2012) 407 – 426.

[56] J.-P. P. Richard and M. Tawarmalani, Lifting Inequalities: a Framework for Generating Strong Cuts for Nonlinear Programs, *Mathematical Programming* **121** (2010) 61 – 104.

[57] R. T. Rockafellar, *Convex Analysis*, Princeton University Press (1970).

[58] A. E. Rosenbluth, D. Melville, K. Tian, K. Lai, N. Seong, D. Pfeiffer, and M. Colburn, Global Optimization of Masks, including Film Stack Design to Restore TM Contrast in High NA TCC's. *Proceedings of SPIE*, 6520:65200P, (2007).

[59] A. Saxena, P. Bonami and J. Lee, Convex Relaxations of Non-convex Mixed Integer Quadratically Constrained Programs: Projected Formulations. To appear, *Mathematical Programming.*

[60] A. Schrijver, On Cutting Planes, *Annals of Discrete Mathematics* **9** (1980) 291 – 296.

[61] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*, Kluwer, Dordrecht (1998).

[62] H. D. Sherali and W. P. Adams, Reformulation-Linearization Technique (RLT) for Semi-Infinite and Convex Programs under Mixed 0-1 and General Discrete Restrictions, *Discrete Applied Mathematics*, **157**, (2009) 1319 – 1333.

[63] R. A. Stubbs and S. Mehrotra, A Branch-and-Cut Method for 0-1 Mixed Convex Programming, *Mathematical Programming* **86** (1999), 515 - 532.

[64] R. A. Stubbs and S. Mehrotra, Generating Convex Polynomial Inequalities for Mixed 01 Programs, *Journal of Global Optimization* **24** (2002), 311 – 332.

[65] J. F. Sturm and S. Zhang, On Cones of Nonnegative Quadratic Functions, *Mathematics of Operations Research* **2** (2003) 246 – 267.

[66] M. Tawarmalani and N.V. Sahinindis, A Polyhedral Branch-and-cut Approach to Global Optimization, *Mathematical Programming B* **103** (2005), 225 – 249.

[67] V. A. Yakubovich, S-procedure in Nonlinear Control Theory, Vestnik Leningrad University, **1** (1971) 62 – 777.

[68] L. Vandenberghe and S. Boyd, Semidefinite Programming. *SIAM Review* **38** (1996), 49 – 95.

[69] L. A. Wolsey, Facets and Strong Valid Inequalities for Integer Programs, *Operations Research* **24** (1976), 367 - 372.

[70] Y. Ye and S. Zhang, New Results in Quadratic Minimization, *SIAM Journal on Optimization* **14:1** (2003), 245 – 267.

[71] E. Zemel, Lifting the Facets of Zero-One Polytopes, *Mathematical Programming* **11** (1978), 268 - 277.

Appendix A

Quadratic Programming

Several of the results in this thesis rely on closed-form solutions for quadratic programs. In this appendix we provide derivations of the forms of these solutions, as well as their accompanying optimal objective values. We restrict our attention to convex objective, linearly constrained quadratic programs. These are convex optimization problems, and can be solved using the the method of Lagrange multipliers.

A.1 Positive Definite Objective

In this section we give results for the case when the objective function is a positive definite quadratic.

Lemma A.1.1. *Let Q be positive definite, and let $A \in \mathbb{R}^{m \times n}$ be a matrix of full row rank. Then the optimal solution to the quadratic program*

$$\begin{aligned} \textit{minimize:} \quad & x^T Q x + q^T x \\ \textit{subject to:} \quad & Ax = b \end{aligned}$$

is given by

$$x^* = -\frac{1}{2} Q^{-1} q + Q^{-1} A^T \left(A Q^{-1} A^T \right)^{-1} \left(b + \frac{1}{2} A Q^{-1} q \right) \tag{A.1}$$

Proof. The Lagrangian of this quadratic program, with Lagrange multipliers $\theta \in \mathbb{R}^m$, is

$$\mathcal{L}(x, \theta) = x^T Q x + q^T x - \theta^T (Ax - b).$$

Equating the gradient of $\mathcal{L}$ in x to 0 gives

$$\begin{aligned}\nabla_x \mathcal{L} = 0 \quad &\Leftrightarrow \quad 2Qx + q - A^T\theta = 0 \\ &\Leftrightarrow \quad Qx = -\frac{1}{2}q + \frac{1}{2}A^T\theta \\ &\Leftrightarrow \quad x = -\frac{1}{2}Q^{-1}q + \frac{1}{2}Q^{-1}A^T\theta. \end{aligned} \tag{A.2}$$

Next we use the $Ax = b$ constraint to determine θ:

$$\begin{aligned} Ax = b \quad &\Leftrightarrow \quad A\left(-\frac{1}{2}Q^{-1}q + \frac{1}{2}Q^{-1}A^T\theta\right) = b \\ &\Leftrightarrow \quad \frac{1}{2}AQ^{-1}A^T\theta = b + \frac{1}{2}AQ^{-1}q \\ &\Leftrightarrow \quad \theta = \left(AQ^{-1}A^T\right)^{-1}\left(2b + AQ^{-1}q\right). \end{aligned} \tag{A.3}$$

Substituting this expression for θ into (A.2) gives (A.1). ■

Corollary A.1.2. *Let $A \in \mathbb{R}^{m\times n}$ be a matrix of full row rank. Then the optimal solution to the quadratic program*

$$\begin{aligned} \textit{minimize:} \quad & x^Tx + q^Tx \\ \textit{subject to:} \quad & Ax = b \end{aligned}$$

is given by

$$x^* = -\frac{1}{2}q + A^T\left(AA^T\right)^{-1}\left(b + \frac{1}{2}q\right)$$

and has an optimal objective value of

$$-\frac{1}{4}q^Tq + \left(b + \frac{1}{2}Aq\right)^T (AA^T)^{-1}\left(b + \frac{1}{2}Aq\right)$$

Proof. Replace Q with the identity matrix in (A.1) and substitute the result into $x^Tx + q^Tx$. ■

Finally we note that in the case where A has either one or two rows, the matrix $(AA^T)^{-1}$ can be obtained easily in closed form. Specifically, we have

$$(AA^T)^{-1} = \frac{1}{||a_1||^2} \quad \text{when} \quad A = \begin{bmatrix} a_1^T \end{bmatrix} \in \mathbb{R}^{1\times n} \tag{A.4}$$

and

$$\begin{aligned}(AA^T)^{-1} \quad &= \quad \begin{bmatrix} ||a_1||^2 & a_1^Ta_2 \\ a_1^Ta_2 & ||a_2||^2 \end{bmatrix}^{-1} \\ &= \quad \frac{\begin{bmatrix} ||a_2||^2 & -a_1^Ta_2 \\ -a_1^Ta_2 & ||a_1||^2 \end{bmatrix}}{||a_1||^2\,||a_2||^2 - (a_1^Ta_2)^2} \end{aligned} \tag{A.5}$$

when

$$A = \begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} \in \mathbb{R}^{2\times n}.$$

A.2 Semidefinite Objectives

In this section we allow the matrix Q to be positive semidefinite. We give necessary and sufficient conditions for the minimum value of $x^TQx + q^Tx$ to be finite, and an expression for the optimal value when these conditions hold.

Lemma A.2.1. *Let Q be an $n \times n$ symmetric matrix. Then*

$$\textit{minimize: } \{x^TQx + q^Tx\} > -\infty \tag{A.6}$$

if and only if

$$Q \succeq 0 \quad \textit{and} \quad q = Q\pi \quad \textit{for some } \pi \in \mathbb{R}^n \tag{A.7}$$

Proof. Suppose (A.7) does not hold. Then either Q is not positive semidefinite, or there is no solution to $Q\pi = q$. First assume the former is true. Then Q has an eigenvector v with an associated eigenvalue $\lambda < 0$. For any $\sigma \in \mathbb{R}$, we have

$$(\sigma v)^TQ(\sigma v) + q^T(\sigma v) = \sigma^2 v^TQv + \sigma q^Tv = \sigma^2\lambda v^Tv + \sigma q^Tv = \sigma^2\lambda + \sigma q^Tv$$

which approaches $-\infty$ as σ tends to ∞. Next, assume $Q \succeq 0$ but there is no solution to $Q\pi = q$. By Farkas's Lemma, this is equivalent to the existence of a vector δ with $Q\delta = 0$ and $q^T\delta = -1$. Then for $\sigma \in \mathbb{R}$, we have

$$(\sigma\delta)^TQ(\sigma\delta) + q^T(\sigma\delta) = 0 + \sigma q^T\delta = -\sigma$$

which approaches $-\infty$ as σ tends to ∞.

Now assume that (A.7) does hold. Then the minimization problem in (A.6) is equivalent to

$$\text{minimize: } \{x^TQx + \pi^TQx\}. \tag{A.8}$$

Define $f(x) = x^TQx + \pi^TQx$, and $x^* = -\frac{1}{2}\pi$. Then $\nabla f(x^*) = 0$, and $\nabla^2 f(x) = Q$ for all x. Moreover, because the function f is quadratic, the second-order Taylor expansion for f is exact and so for any $y \in \mathbb{R}^n$, we have

$$\begin{aligned} f(y) &= f(x^*) + (y - x^*)^T \nabla f(x^*) + \frac{1}{2}(y - x^*)^T \nabla^2 f(x^*)(y - x^*) \\ &= f(x^*) + \frac{1}{2}(y - x^*)^T Q (y - x^*) \\ &\geq f(x^*) \qquad \text{(because } Q \succeq 0\text{)}. \end{aligned}$$

Therefore x^* is a global minimizer of f, with an objective value of

$$-\frac{1}{4}\pi^T Q \pi > -\infty.$$

■