Cutting-planes for optimization of convex functions over nonconvex sets

Daniel Bienstock and Alexander Michalka, Columbia University

November 2011, revised May 2013, December 2013

Abstract

We derive linear inequality characterizations for sets of the form $\text{conv}\{(x,q) \in \mathbb{R}^d \times \mathbb{R} : q \geq Q(x), \quad x \in \mathbb{R}^d - \text{int}(P)\}$ where Q is convex and differentiable and $P \subset \mathbb{R}^d$. We show that in several cases our characterization leads to polynomial-time separation algorithms that operate in the original space of variables, in particular when Q is a positive-definite quadratic and P is a polyhedron or an ellipsoid.

1 Introduction

The current state-of-the-art for linear mixed-integer programming relies on cutting-planes, a methodology supported by a strong body of theory that has also achieved computational success. Nevertheless, the solution of an optimization problem $\min\{\, Q(x) \;:\; x \in \mathcal{F}\,\}$ with $Q(x)$ convex and $\mathcal{F} \subseteq \mathbb{R}^d$ mixed-integer would present a challenge to the cutting-plane approach. Any algorithm that relies on separation from $\text{conv}(\mathcal{F})$ will in general fail, because an optimal solution x^* to $\min\{\, Q(x) \;:\; x \in \text{conv}(\mathcal{F})\,\}$ may satisfy (i) $x^* \notin \mathcal{F}$, and (ii) x^* is in the relative interior of a face of $\text{conv}(\mathcal{F})$, and so no cutting-plane can separate x^* from $\mathcal{F}$.

This observation suggests a paradigm used in the "lattice-free set" methodology in mixed-integer programming (reviewed below). Given $(x^*, q^*) \in \mathbb{R}^d \times \mathbb{R}$ with $x^* \notin \mathcal{F}$, one computes a set $P \subset \mathbb{R}^d$ with $x^* \in \text{int}(P)$ and $\mathcal{F} \cap \text{int}(P) = \emptyset$ ("int" denotes interior) together with an inequality that separates (x^*, q^*) from the set

$$S \;\doteq\; \{\, (x,q) \in \mathbb{R}^d \times \mathbb{R} \;:\; q \geq Q(x), \quad x \in \mathbb{R}^d - \text{int}(P)\,\}. \tag{1}$$

The focus of this paper is the study of sets of the general form (1) with the goal of characterizing $\text{conv}(S)$ by linear inequalities.[1] Two classes of 'trivial' valid inequalities for S are (a) valid inequalities for $\mathbb{R}^d - P$; and (b) linearization, or first-order, inequalities

$$q \;\geq\; Q(y) + \nabla Q(y)^T (x - y), \tag{2}$$

where $y \in \mathbb{R}^d$. Usually these two families of inequalities are not sufficient to characterize *conv*(S). In this paper, motivated by mixed-integer programming considerations, we consider 'lifted' versions of (2), that is to say inequalities of the form

$$q \;\geq\; Q(y) + \nabla Q(y)^T (x - y) + \alpha\, p^T (x - y), \tag{3}$$

where $\alpha > 0$ and $p \in \mathbb{R}^d$. For (3) to be valid y must lie in the boundary of P; further p cannot be arbitrary, and instead must point "into" P in a sense made precise later. We obtain the following results:

Theorem I. Let $Q(x)$ be convex and differentiable. Any linear inequality $\delta q \geq \beta^T x + \beta_0$ valid for S and such that $\{(x,q) \in \mathbb{R}^d \times \mathbb{R} : \delta q = \beta^T x + \beta_0\}$ is a supporting hyperplane for $\text{conv}(S)$ is dominated by a combination of up to two inequalities of three types: (a) valid inequalities for $\mathbb{R}^d - P$, (b) linearization inequalities obtained at points $y \in \mathbb{R}^d - \text{int}(P)$, and (c) valid lifted inequalities obtained at points y in the boundary of P.

Separation over the three types inequalities listed in Theorem I is closely related, but not precisely equivalent to separation from $\text{conv}(S)$. In this regard, we obtain a sharpening of Theorem I:

[1] Note that for any extreme point (x,q) of $\text{conv}(S)$ we have $x \in S$ and $q = Q(x)$; a folklore observation.

Theorem II. Suppose that $Q(x)/\|x\| \to +\infty$ as $\|x\| \to +\infty$, that there is a polynomial-time separation oracle for $\mathbb{R}^d - \text{int}(P)$, and that $\nabla Q(x)$ is polynomial-time computable at any x. Then polynomial-time separation over $\text{conv}(S)$ is equivalent to polynomial-time separation over the lifted inequalities.

Unlike what happens in the linear mixed-integer setting, one can produce examples where a lifted inequality (3) is binding at just one point – y. We show that such cases can be essentially characterized in terms of the structure of the boundary of P where lifting is attempted and (again) the degree of strong convexity of $Q(x)$:

Theorem III. Let y be a point in the boundary of P such that an open half-ball with center y and positive radius is contained in P. Suppose further that $Q(x)/\|x\| \to +\infty$ as $\|x\| \to +\infty$. Then any lifted inequality (3) obtained at y, and using the maximum valid lifting coefficient α, will be binding at some point $(w, Q(w))$ where $w \neq y$ is in the boundary of P.

Finally, in several cases the characterization provided by Theorem I leads to polynomial-time separation:

Theorem IV. One can separate in polynomial time from:

(i) A set $\text{conv}(S)$ as above, when $Q(x)$ is a positive-definite quadratic and P is a polyhedron or an ellipsoid.

(ii) A set of the form $\{(x, w, q) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \,:\, q \geq x^T H x + h^T x,\ w \leq x^T A x\}$, where $H \succ 0$ and $A \succeq 0$.

Case (i) is important because the exclusion of a polyhedron has been proposed in several of the lattice-free set schemes in the literature. The ellipsoidal case arises, for example, when considering the cardinality-constrained convex quadratic programming problem [18]; also see [20]. In Section 4 we provide motivation for the study of the set in (ii); however, using $d = n + 1$ and $P = \{(x, w) \in \mathbb{R}^n \times \mathbb{R} \,:\, w \geq x^T A x\}$ this set is of the form (1).

1.0.1 Motivation and background

The "lattice-free set" paradigm can be considered one of the single most fundamental ideas underlying the theory of cutting-planes for linear mixed-integer programming. Our work in this paper seeks to extend this methodology to the nonlinear setting. In the linear, pure integer case the methodology can be outlined as follows. Let $\mathcal{F} = \{x \in \mathbb{Z}^d \,:\, Ax \geq b\}$ and consider an integer program $\min\{c^T x \,:\, x \in \mathcal{F}\}$. Suppose that $x^* \notin \mathbb{Z}^d$ is an extreme point optimal solution to some convex relaxation to this problem. Then one attempts to cut-off x^* by applying the following procedure.

First, a set $X = X(x^*) \subset \mathbb{R}^d$ with $x^* \in \text{int}(X)$ and $\text{int}(X) \cap \mathbb{Z}^d = \emptyset$ is identified. Thus, denoting

$$P \ \doteq\ X \cup \{x \in \mathbb{R}^d \,:\, Ax \not\geq b\}$$

we have that $\mathcal{F} \ \subseteq\ \mathbb{R}^d - \text{int}(P)$. A valid inequality $\pi^T x \geq \pi_0$ is then sought such that

$$\begin{aligned} &(1) \qquad \pi^T x = \pi_0 \ \text{ supports } \ \text{conv}(\mathbb{R}^d - \text{int}(P)), \ \text{ and} \\ &(2) \qquad \pi^T x^* < \pi_0. \end{aligned}$$

One of the earliest versions of this idea is embodied by the family of split cuts (see [44]), where the set X is the region bounded by two parallel hyperplanes. Intersection cuts [9] are a closely related methodology. For further material see [3], [25]. [43] studies the use of split and intersection cuts in nonlinear optimization. Also see [15]. The disjunctive method for mixed-integer programming (see [10], [11], [12]), and the mixed-integer rounding procedure [44] rely on a similar paradigm (also see [37], [54]). We note that in the standard form of the above procedure integrality of the variables is only used in the construction of the lattice-free set X; however the cut $\pi^T x \geq \pi_0$ is computed simply using the geometry of the set $\mathbb{R}^d - \text{int}(P)$ (or even, just the set $\mathbb{R}^d - X$) and the point x^*. Another point is that the set X is typically quite simple (e.g. a polyhedron defined by a small number of inequalities). Our approach to nonlinear, mixed-integer programs seeks to adapt the lattice-free approach, leading to the study of sets S as outlined above from a cutting-plane perspective. In Section 5.1 we will describe a specific rendition of our method in the context of cardinality-constrained convex quadratic programming.

We apply the procedure closely related to *lifting* valid inequalities for mixed-integer programs. See [44] for background. It can be summarized as follows. Let $\mathcal{F} \subseteq \mathbb{R}^n_+$ be the feasible region of an integer program, and for $k < n$ let $\sum_{j=1}^k \beta_j x_j \leq \beta_0$ be valid for $\mathcal{F}$ when $x_j = 0$ for all $j > k$. Lifting is the process whereby this inequality is modified so as to yield an inequality $\sum_{j=1}^{k+1} \beta_j x_j \leq \beta_0$ valid for $\mathcal{F}$ when $x_j = 0$ for all $j > k+1$. Geometrically, the hyperplane defined by $\sum_j^k \beta_j x = \beta_0$ is being rotated so as to support $\text{conv}(\mathcal{F})$ at a point $\hat{x}$ with $\hat{x}_{k+1} > 0$. From this perspective, lifting may more aptly be referred to as "tilting" (see the discussion in [28]). Lifting techniques have proved compelling in that they are supported by strong theory and can also provide a computationally practicable way to strengthen valid inequalities. As a result lifting is ubiquitous in mixed-integer programming solvers.

The extension of lifting to the nonlinear or continuous setting is not new: see [17], [27], [34], and [7]. Also see [16], which lifts "tangent" inequalities to approximate multilinear functions. An interesting use of lifting appears in [46], which approximates, using lifted linear inequalities, SDP relaxations of quadratically constrained sets. A comprehensive framework for coordinate-wise lifting is presented in [47], where given an arbitrary function $f(x, y) : \mathbb{R}^{p+n} \to \mathbb{R}$, and a linear underestimator for f which is valid when $y = 0$, i.e. $f(x, 0) \geq \bar{\alpha}^T x - \delta$, lifting is used to modify this inequality so as to make it valid for all (x, y). This results in an inequality of the form $f(x, y) \geq \bar{\alpha}^T x + \nu^T y - \delta$, for an appropriate vector ν. The authors discuss sequence independent lifting (over the variables y) and applications to several problem types, such as bilinear knapsack sets.

The techniques in this paper involve linear approximation to nonlinear functions and further we focus on quadratics. Both subjects have received significant attention in the literature. One of the earliest results (see [40] and [2]) is the characterization of the convex envelope of a box-constrained bilinear form $x_1 x_2$ (also see [48]). These results lead to techniques that have been incorporated in software systems such as BARON [49] and Couenne [14].

The problems we consider fall within the broader scope of global optimization problems. The work in [57] and [58] has resulted in key advances that can be applied in very general settings. An important idea is that of outer approximation. Given a convex function $g : \mathbb{R} \to \mathbb{R}$, [57] shows how to construct an outer approximation to g with arbitrary accuracy ϵ and using a number of lines that grows (asymptotically) as ϵ^{-1}. Other techniques include the development of effective relaxations to typical nonlinear functions (such as exponential, logarithmic, and multilinear functions), the automatic generation of convex underestimators of general functions, and the adaptation of traditional branch-and-bound to continuous domains. These techniques are amenable to implementation in a branch-and-cut setting (see [58]); and have been included in BARON, achieving computational success. Convex extensions of a given function f are considered in [56]; these are convex functions that agree with f on a subset of its domain. This theory is further used in [56] to study the convex envelope of the function x/y of two real variables x, y over a rectangle in $\mathbb{R}^2$; additionally several results are presented concerning convex envelopes of multilinear functions.

Recently, some interesting new results on multilinear forms have been obtained, see for example [38]. A survey is provided in [22]; with additional material of interest in [23], [50], [51] and [41]. A polyhedral approximation scheme for nonconvex quadratically constrained quadratic programs is given in [8]. Also see [36]. A different, frequently-applied construct is the Reformulation-Linearization Technique (RLT) and semidefinite programming extensions; see [52] and [53]. Earlier work [55], [29], [32] (also see [24]) in nonlinear 0-1 mixed-integer programming produced families of cuts arising from first-order information. Finally, the connection with semidefinite programming has yielded a number of deep results focusing on quadratic functions, see for example [4], [21], [5]. This provides an alternative (but related) methodology for addressing some of the problems we consider.

1.0.2 Organization

This paper is organized as follows. In Section 2 we study the set S defined in (1); we formally define our lifted inequalities and prove Theorem I given above. Then we strengthen our result when $Q(x)$ satisfies an appropriate generalization of strong convexity (Section 2.0.3, obtaining Theorems II and III) with additional strengthening when $Q(x)$ is a positive-definite quadratic in Section 2.1. In Sections 3.1 and 3.2 we obtain Theorem IV (i); presenting a polynomial-time separation algorithm for $\text{conv}(S)$ when $Q(x)$ is positive-definite quadratic and the set P in (1) is a polyhedron and an ellipsoid, respectively. Section 4 obtains a polynomial-time separation algorithm for a set $\{(x, w, q) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \, : \, q \geq x^T H x + h^T x, \, w \leq x^T A x\}$, where $H \succ 0$ and $A \succeq 0$, which is also a special case of (1) (Theorem IV (ii)). Finally, in Section 5 we present initial experimental results.

We will use the following terminology:

Definition 1.1

(1) The boundary *of a set $V \subseteq \mathbb{R}^d$ is defined as*

$$\partial V = \{x \in \mathbb{R}^d \, : \, \mathcal{B} \cap V \neq \emptyset \text{ and } \mathcal{B} \cap (\mathbb{R}^d - V) \neq \emptyset \text{ for every open ball } \mathcal{B} \text{ containing } x\}.$$

(2) Given a set $X \subseteq \mathbb{R}^d$, an inequality $\gamma^T x \geq \gamma_0$ which is valid for X will be called a supporting inequality for X *if $\{x \in \mathbb{R}^d \, : \, \gamma^T x = \gamma_0\}$ defines a supporting hyperplane for* $\text{conv}(X)$. *We will also say that $\gamma^T x \geq \gamma_0$* supports X, *and if $y \in X$ is such that $\gamma^T y = \gamma_0$ we will also say that $\gamma^T x \geq \gamma_0$ supports X* at y.

2 Lifted first-order cuts

Here we consider the set S given by (1) where Q is convex and differentiable and $P \subset \mathbb{R}^d$. We will prove a more detailed version of Theorem I, given as Theorem 2.7, below. This theorem provides a characterization of supporting hyperplanes for $\text{conv}(S)$, in particular singling out the lifted inequalities (3). We will first introduce these lifted inequalities and prove a series of results (Lemma 2.6, and Propositions 2.8 - 2.10) leading to Theorem 2.7. Following this material, in Section 2.0.3 we provide two results that hold when $Q(x)$ grows faster than linearly in every direction. First, we obtain Theorem II (which characterizes polynomial-time separation from $\text{conv}(S)$). Second, assuming (additionally) that the boundary of P is appropriately structured, we prove that the lifted inequalities define hyperplanes guaranteed to support S at (at least) two different points (Theorem III in the Introduction). Finally, in Section 2.1 we discuss the case where $Q(x)$ is a positive-definite quadratic, which allows for a geometric characterization of the lifted inequalities.

We first provide a brief motivation for our approach. We are interested in strengthening the linearized inequality (2) at a point $y \in \partial P$ by modifying it in the form

$$q \;\geq\; Q(y) + \nabla Q(y)^T (x - y) + 2p^T (x - y) \tag{4}$$

for some $p \in \mathbb{R}^d$. Note that this constitutes a strengthening only in the half-plane $\{x \in \mathbb{R}^d \, : \, p^T (x - y) > 0\}$. In order for this strengthening to be valid for S we must also have that

$$\{\, x \in \mathbb{R}^d \quad : \quad Q(x) \,<\, Q(y) + \nabla Q(y)^T (x - y) + 2p^T (x - y) \,\} \quad \subseteq \quad \text{int}(P). \tag{5}$$

Since Q is differentiable it follows that for any $r \in \mathbb{R}^d$ with $p^T r > 0$, any x of the form $x = y + \lambda r$ will be such that $(x, Q(x))$ violates (4) provided $\lambda > 0$ is small enough (and how small may depend on r). That is to say, for any r with $p^T r > 0$ there exist points $x = y + \lambda r$ with $\lambda > 0$ with x contained in the set in the left-hand side of (5). We now make these notions precise.

Definition 2.1 *Let $y \in \partial P$ and $p \in \mathbb{R}^d$ with $\|p\| = 1$. We say that P is locally flat at y with normal p, if for every $r \in \mathbb{R}^d$ with $p^T r > 0$ there exists $\epsilon(r) > 0$ such that*

$$y \,+\, \delta\, r \;\in\; \mathrm{int}(P) \quad \forall\, \delta \text{ with } 0 < \delta \le \epsilon(r).$$

Intuitively, P being locally flat at y with normal p means that for any vector r with positive inner product with p we can move, starting at y, a positive distance "into" P along r.

Example 2.2 *(a) If every connected component of ∂P is a differentiable manifold homeomorphic to $\mathbb{R}^{d-1}$ then ∂P is locally flat at any $y \in \partial P$, with a unique normal vector: the normal to the tangent space to ∂P at y, oriented into P. (b) 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. (c) The non-convex set $P = \{(x_1, x_2) \in \mathbb{R}^2 \;:\; |x_1| \ge x_2\}$ is locally flat at every point on its boundary, even the vertex at $(0, 0)$ (with normal $(0, -1)$).*

We can now begin our lifting construction.

Definition 2.3 *Let $y \in \partial P$ be locally flat with normal p. For $\alpha \ge 0$ consider the inequality*

$$q \;\;\ge\;\; Q(y) + \nabla Q(y)^T (x - y) \,+\, 2\alpha p^T (x - y). \tag{6}$$

The lifting coefficient at y, with respect to p, is $\hat{\alpha} = \hat{\alpha}(P, p, y) \doteq \sup\{\, \alpha \,:\, (6) \;\text{ is valid for } S\,\}$.

We remark that one can equivalently write $\hat{\alpha} = \sup\{\alpha \,:\, (6)$ is valid for S for $0 \le \alpha \le \bar{\alpha}\}$. Clearly the lifting coefficient is nonnegative, and we are interested in the cases where this quantity is actually positive.

Example 2.4 *Let $d = 2$, $Q(x) = x_1^2 + x_2^4$, and $P = \{x \in \mathbb{R}^2 \,:\, x_2 \le |x_1| + 1$ and $x_2 \ge (x_1 - 1)(x_1 - 2)^2\,\}$. Thus $\nabla Q(y)^T = (2y_1, 4y_2^3)$ for any $y \in \mathbb{R}^2$. For $v \in \mathbb{R}$, write $c(v) = (v - 1)(v - 2)^2$. Then P is locally flat at $y = (1, 0)^T$ with normal*

$$\left(\frac{-1}{\sqrt{1 + [1/c'(1)]^2}} \,,\, \frac{1/c'(1)}{\sqrt{1 + [1/c'(1)]^2}} \right)^T \;=\; \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)^T.$$

Hence (2) takes the form

$$q \;\ge\; 1 \,+\, 2(x_1 - 1) \,+\, \hat{\alpha} \left(\frac{-1}{\sqrt{2}}(x_1 - 1) + \frac{1}{\sqrt{2}} x_2 \right).$$

A calculation shows that $\hat{\alpha} = \sqrt{2}$, and so the inequality can be rewritten as $q \;\ge\; x_1 \,+\, x_2$, which is binding at $(x, Q(x))$ with $x = (1, 0)^T$ and $x = (0, 1)^T$.

Definition 2.5 *Suppose P is locally flat at y with normal p, and that $\hat{\alpha} < +\infty$. We call*

$$q \;\;\ge\;\; Q(y) \,+\, \nabla Q(y)^T (x - y) \,+\, 2\hat{\alpha} p^T (x - y) \tag{7}$$

a lifted first-order (LFO) inequality *generated at y (with respect to p).*

The following result establishes simple properties of the lifting coefficient.

Lemma 2.6 *Let P be locally flat at y with normal p.* **(a)** *If there exists $v \notin P$ such that $p^T(v - y) > 0$ then $\hat{\alpha} < +\infty$.* **(b)** *If $p^T(v - y) \le 0$ for all $v \notin P$ then $\hat{\alpha} = +\infty$.* **(c)** *If $\hat{\alpha} < +\infty$ then (7) is valid for S.*

Proof. **(a)** Let $v \notin P$ satisfy $p^T(v - y) > 0$. Then clearly for α large enough, $(v, Q(v))$ will violate (6), and so $\hat{\alpha} < +\infty$. **(b)** This follows trivially since for any $\alpha > 0$ and any $x \notin P$ the right-hand side of (6) is dominated by that of (2). **(c)** This fact follows by continuity. ■

Remarks. The construction that culminates in Definition 2.5 is a generalization of the classical lifting construction in mixed-integer programming (see [44], [27], [34], [7]). The LFO inequality at y uses the local structure of P to strengthen the linearization inequality (2); the strengthening is only local, however the LFO inequality is (globally) valid.

We next prove the main result in this section.

Theorem 2.7 *Let $\delta q \geq \beta^T x + \beta_0$ be valid for S and binding at $(y, z) \in S$ for some $y \in \mathbb{R}^d - \mathrm{int}(P)$ and $z \geq Q(y)$. Then at least one of the following conditions holds:*

(1) $\delta = 0$ and $\beta^T x + \beta_0 \leq 0$ is valid for $\mathbb{R}^d - P$.

(2) $\delta > 0$, $z = Q(y)$ and $\delta q \geq \beta^T x + \beta_0$ is a positive multiple of the linearization inequality at y.

(3) $\delta > 0$, $z = Q(y)$ and $y \in \partial P$. Moreover, $\delta q \geq \beta^T x + \beta_0$ is a nonnegative linear combination of the linearization inequality at y and a linear inequality supporting $\mathbb{R}^d - \mathrm{int}(P)$ at y.

(4) $\delta > 0$, $z = Q(y)$, $y \in \partial P$ and P is locally flat at y with some normal p. Moreover $\delta q \geq \beta^T x + \beta_0$ is implied by the linearization inequality at y together with the LFO inequality at y with respect to p.

The proof of this theorem will be broken into a sequence of steps. We will consider a fixed inequality

$$\delta q \quad \geq \quad \beta^T x \, + \, \beta_0 \tag{8}$$

that is supporting for S at a point (y, z) with $y \in \mathbb{R}^d - \mathrm{int}(P)$ and $z \geq Q(y)$ and obtain Theorem 2.7 through a sequence of results. We begin with some simple preliminary observations.

1. Since (8) is valid for S, we have $\delta \geq 0$. If $\delta = 0$ then $\beta^T x + \beta_0 \leq 0$ is valid for $\mathbb{R}^d - P$ and we are done (case (1) of Theorem 2.7). Thus we will assume $\delta > 0$ and by scaling if necessary that $\delta = 1$. Hence we must have $z = Q(y)$ and in summary (8) is binding at $(y, Q(y))$.

2. Write

$$\beta^T x + \beta_0 \quad = \quad Q(y) + \nabla Q(y)^T (x - y) \, + \, 2v^T x \, - \, 2v_0,$$

for appropriate $v \in \mathbb{R}^d$ and $v_0 \in \mathbb{R}$. Since (8) holds with equality at $(y, Q(y))$ it follows that $v_0 = v^T y$ and we can rewrite (8) as

$$q \quad \geq \quad Q(y) + \nabla Q(y)^T (x - y) + 2v^T (x - y). \tag{9}$$

3. If $v = 0$ in (9) then (9) is the linearization inequality at y (case (2) of Theorem 2.7). We will therefore assume $v \neq 0$.

We will use form (9) of (8), with $v \neq 0$, in Propositions 2.8, 2.9, 2.10 and Corollary 2.11 given next.

Proposition 2.8 *Let $w \in \mathbb{R}^d$ be arbitrary such that $v^T w > 0$. Then there exists a positive value $\epsilon = \epsilon(w)$ such that for any $0 < \delta < \epsilon$ the right-hand side of (9) evaluated at $x = y + \delta w$ exceeds $Q(y + \delta w)$ since $v^T v > 0$.*

Proof. For $\delta > 0$ write $F(\delta) = Q(y + \delta w)$. The right-hand side of (9) evaluated at $y + \delta w$ equals

$$Q(y) + \delta(\nabla Q(y) + 2v)^T w \ \doteq \ G(\delta).$$

Then $G(0) = Q(y) = F(0)$ and $G'(0) = (\nabla Q(y) + 2v)^T w > \nabla Q(y)^T w = F'(0)$. As a result $F(\delta) < G(\delta)$ for $\delta > 0$ small enough. ■

Proposition 2.9 *We have $y \in \partial P$.*

Proof. Suppose by contradiction that $y \in \text{int}(\mathbb{R}^d - P)$. By Proposition 2.8, for $\delta > 0$ small enough the right-hand side of (9) evaluated at $y + \delta v$ exceeds $Q(y + \delta v)$. This is a contradiction since for $\delta > 0$ small enough $y + \delta v \in \text{int}(\mathbb{R}^d - P) \subseteq \mathbb{R}^d - \text{int}(P)$. ■

Proposition 2.10 *P is locally flat at y, with normal $p \doteq \frac{v}{\|v\|}$.*

Proof. Follows directly from Propositions 2.8 and 2.9. ■

As a result of Proposition 2.10, we can rewrite (9) as

$$q \quad \geq \quad Q(y) + \nabla Q(y)^T (x - y) + 2\|v\| p^T (x - y). \tag{10}$$

where p is as in Proposition 2.10.

Corollary 2.11 *Let $\hat{\alpha} \doteq \hat{\alpha}(P, p, y)$. Then $\hat{\alpha} \geq \|v\| > 0$, and furthermore*

(a) If $\hat{\alpha} = +\infty$ then $p^T(x - y) \leq 0$ is valid for $\mathbb{R}^d - \text{int}(P)$ and (10) is a nonnegative linear combination of the linearization inequality at y, and $p^T(x - y) \leq 0$.

(b) If $\hat{\alpha} = \|v\|$ the LFO inequality at y with respect to p

$$q \quad \geq \quad Q(y) + \nabla Q(y)^T (x - y) + 2\hat{\alpha} p^T (x - y) \tag{11}$$

and (10) are identical constraints.

(c) Suppose $\|v\| < \hat{\alpha} < +\infty$. Then at any x with $p^T(x - y) > 0$ the right-hand side of (11) is strictly larger than that of (10). Further, at any x with $p^T(x - y) < 0$ the right-hand side of the linearization inequality at y is strictly larger than that of (10).

(d) Suppose $\|v\| < \hat{\alpha} < +\infty$. For any $x \in \mathbb{R}^d$ inequality (10) is weaker than the combination of the linearization inequality at y and inequality (11).

Proof. The definition of lifting coefficient and validity of (10) implies $\hat{\alpha} \geq \|v\|$. (a) Part (a) of Lemma 2.6 implies that $0 \geq p^T(x - y)$ is valid for $\mathbb{R}^d - \text{int}(P)$; the rest of the statement follows since clearly (10) is obtained by adding $\|v\|$ times $0 \geq p^T(x - y)$ to the linearization inequality at y. Part (b) is clear, and (c) follows since $\hat{\alpha} > 0$. Finally, (d) is a corollary of (c): given $x \in \mathbb{R}^d$, if $p^T(x - y) \leq 0$ the right-hand side of (10) evaluated at x, is at most that of the the linearization inequality at y. And if $p^T(x - y) > 0$ then the right-hand side of (10) evaluated at x is less than that of (11). ■

Note that (a) of Corollary 2.11 amounts to case (3) of Theorem 2.7, and (b) and (c) to case (4). Thus we have completed the proof of Theorem 2.7.

Remark 2.12 *There are a number of conditions under which the separation problem for $\text{conv}(S)$ is equivalent to separation by LFO inequalities, linearization inequalities, and valid inequalities for $\mathbb{R}^d - P$. We will return to this issue in Theorem 2.14, below.*

Example 2.13 *As an illustration of the use of LFO inequalities, consider Example 2.4. Suppose we apply the following heuristic for the problem $\min\{Q(x) \, : \, x \in \mathbb{R}^2 - \text{int}(P)\}$. We start with the relaxation*

$$\min\{q \, : \, q \geq x_1 + x_2, \; q \geq 0, \; x \in \mathbb{R}^2\}$$

consisting of the LFO inequality at $(1, 0)^T$ and the linearization inequality at $(0, 0)$. We initialize $\breve{x}_1 = 1$. Then we perform the following steps.
1. *Solve the relaxation, obtaining solution x^*.*
2. If *$x^* \in \mathbb{R}^2 - \text{int}(P)$, add to the relaxation the linearization inequality at x^*, and go to 1.* **Otherwise:**
3. *Update $\breve{x}_1 = \frac{1}{2}(x_1^* + \breve{x}_1)$.*

4. *Add to the relaxation the LFO inequality and the linearization inequality at* $(\breve{x}_1, (\breve{x}_1 - 1)(\breve{x}_1 - 2)^2)^T$.
5. *Go to 1.*
This heuristic will produce the following sequence of values $\breve{x}$ *(truncated to three digits):* $1.000, 0.500, 0.545, 0.596, 0.648, 0.702, 0.687, 0.691, 0.694, 0.696$. *After nine iterations, the formulation proves a lower bound of* 0.55577 *on the value of the optimization problem. Moreover, setting* $\hat{x}_1 = 0.696$ *and* $\hat{x}_2 = (\hat{x}_1 - 1)(\hat{x}_1 - 2)^2$*, we have that* $\hat{x}$ *is feasible while* $Q(\hat{x}) \approx 0.55582$.

2.0.3 Strong convexity implications on LFO inequalities

Here we address two issues that arise from the above analysis and which are resolved when $Q(x)$ satisfies a strong convexity assumption: the relationship between separation from LFO inequalities and separation from $\text{conv}(S)$ (Theorem II in the Introduction), and whether LFO inequalities are binding at more than one point (Theorem III). To formalize our approach we first review some standard concepts.

A function $f : \mathbb{R}^d \to \mathbb{R}$ is called strongly convex (with modulus 2) [19] if

$$f(x) \ \geq \ f(y) + \nabla f(y)^T (x - y) + \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^d.$$

Strongly convex functions are of interest because they include positive-definite quadratics, which we will focus on in some of our results, below. A generalization of strong convexity is ψ-strong convexity, where for $\psi : \mathbb{R}_+ \to \mathbb{R}$, the following condition is satisfied:

$$f(x) \ \geq \ f(y) + \nabla f(y)^T (x - y) + \psi(\|x - y\|), \quad \forall x, y \in \mathbb{R}^d.$$

This is a generalization because a strongly convex function is ψ-strongly convex with $\psi(t) = t^2$, but for example, for $d = 1$, $f(x) = x^4$ is not strongly convex (near $x = 0$) but is ψ-strongly convex with $\psi(t) = t^4$. See [39] for a discussion of generalized strong convexity. Here, we will be relying on ψ-strongly convexity where ψ is a function satisfying the following conditions:

$$\psi \text{ is strictly increasing}, \qquad \psi(0) = 0, \qquad \text{and } \lim_{t \to +\infty} \frac{\psi(t)}{t} = +\infty, \tag{12}$$

Under this criterion, ψ-strongly convex functions grow faster than linearly in every direction.

We now turn to the first issue raised above. Whereas Theorem 2.7 classifies supporting inequalities for the set S it does not directly address separation from $\text{conv}(S)$. The next result and corollary address this issue.

Theorem 2.14 *Suppose* $Q(x)$ *is* ψ*-strongly convex where* ψ *satisfies (12), and that* $(\bar{x}, \bar{q}) \in \mathbb{R}^d \times \mathbb{R}$ *satisfies the following conditions:*

(a) $(\bar{x}, \bar{q}) \notin \text{cl}(\text{conv}(S))$.

(b) $\bar{x} \in \text{conv}(\mathbb{R}^d - \text{int}(P))$.

(c) $(\bar{x}, \bar{q})$ *satisfies the linearization inequality at* $\bar{x}$.

Then there is an LFO inequality that separates $(\bar{x}, \bar{q})$ *from* $\text{conv}(S)$.

Proof. By (a) $(\bar{x}, \bar{q})$ violates an inequality $\delta q \geq \beta^T x + \beta_0$ which is valid for $conv(S)$. By (b), $\delta > 0$ and without loss of generality $\delta = 1$. Choose $z \in \mathbb{R}^d - \text{int}(P)$; thus $Q(z) \geq \beta^T x + \beta_0 + \epsilon_0$ for some $\epsilon_0 \geq 0$. Furthermore, since ψ satisfies (12), we have that there exists $R \geq \|z\|$ such that

$$Q(x) \ \geq \ \beta^T x + \beta_0 + \epsilon_0, \qquad \forall \ x \text{ with } \|x\| \geq R.$$

Consequently, there exists $y \in \mathbb{R}^d - \text{int}(P)$ with $\|y\| \leq R$ and $0 \leq \epsilon_1 \leq \epsilon_0$ such that

$$\begin{aligned} Q(y) \ &= \ \beta^T y + \beta_0 + \epsilon_1 \quad \text{and} \\ q \ &\geq \ \beta^T x + \beta_0 + \epsilon_1 \quad \text{is valid for } S. \end{aligned} \tag{13}$$

We will now classify (13) as per Theorem 2.7. Clearly case (1) of Theorem 2.7 does not apply. Further, by convexity of $Q(x)$, if the linearization inequality at any $z \in \mathbb{R}^d - \text{int}(P)$ is violated by $(\bar{x}, \bar{q})$, then so is the linearization inequality at $\bar{x}$ itself, a contradiction by assumption (c). Thus cases (2) and (3) of Theorem 2.7 do not apply. We conclude as desired. ■

Corollary 2.15 *Suppose $Q(x)$ is ψ-strongly convex where ψ satisfies (12), that there is a polynomial-time separation oracle for $\mathbb{R}^d - \text{int}(P)$, and that $\nabla Q(x)$ is polynomial-time computable at any x. Then polynomial-time separation over $\text{conv}(S)$ is equivalent to polynomial-time separation over the LFO inequalities.*

Proof. Given point $(\bar{x}, \bar{q}) \in \mathbb{R}^d \times \mathbb{R}$ we can check in polynomial time whether it satisfies the linearization inequality at $\bar{x}$ as well as $\bar{x} \in \text{conv}(\mathbb{R}^d - P)$. ■

We now address the second issue raised above, where a ψ-strong convexity assumption again has a significant implication. A question left open in Section 2 is whether, given an LFO inequality obtained at a point y, there exists $w \neq y$ and feasible such that the lifted inequality is also binding at $(w, Q(w))$. In Section 3.1 we will see a specific case where the fact that this condition holds is used to construct a polynomial-time separation algorithm for $\text{conv}(S)$. However, and in contrast to what happens in the linear mixed-integer programming setting, the following two examples show that the condition does not always hold, for two possible reasons illustrated in the following examples.

Example 2.16 *Let $P = \{x \in \mathbb{R}^2 : -x_1^2 \leq x_2 \leq 1 + e^{-x_1}\}$, and $Q(x) = x_2 + e^{-x_2} - 1$. Then $0 \in \partial P$, P is locally flat at 0 with normal $(0,1)^T$, $Q(0) = 0$ and $\nabla Q(0) = 0$. Thus the LFO inequality at 0 has the form $q \geq \hat{\alpha} x_2$. Furthermore when $x_2 > 1$, $Q(x) > e^{-1} x_2$. It follows that $\hat{\alpha} = e^{-1}$ since with this choice $q \geq \hat{\alpha} x_2$ is valid, but any larger value will exclude points in S. However any point $x \neq 0$ with $Q(x) = e^{-1} x_2$ satisfies $x_2 = 1$ and thus $x \in \text{int}(P)$.*

Example 2.17 *Let $P = \{x \in \mathbb{R}^2 : x_1 \geq 1\} \cup \{x \in \mathbb{R}^2 : 0 \leq x_1 \leq 1 \text{ and } |x_2| \leq (2x_1 - x_1^2)^{1/2} + x_1\}$. Suppose $Q(x) = \|x\|^2$. By construction, the ball with center $(1,0)^T$ and unit radius is contained in P, and so P is locally flat at 0, with unique normal $(1,0)^T$, and we obtain the LFO inequality $q \geq \hat{\alpha} x_1$.*

Now, for any $R > 0$, the inequality $q \geq 2Rx_1$ is valid for S if and only if $x \in \text{int}(P)$ whenever $\|x\|^2 < 2Rx_1$, i.e. whenever the ball with center $(R,0)^T$ and radius R is contained in P. Therefore $\hat{\alpha} \geq 2$. However, for any value $R > 1$ we can find $0 < x_1 < 1$ such that

$$(2Rx_1 - x_1^2)^{1/2} > (2x_1 - x_1^2)^{1/2} + x_1.$$

It follows that for any $R > 1$ the ball with radius R and center at $(R,0)^T$ is not contained in P. As a result $\hat{\alpha} = 2$ and yet the only point $(x, Q(x))$ with $x \in \mathbb{R}^d - \text{int}(P)$ where $q \geq 2x_1$ is binding is $(0,0)$.

In Example 2.16, the function $Q(x)$ effectively grows at a linear rate in x_2, while in Example 2.17 the boundary of P, near 0, is curved too steeply in the direction of the lifting (so that for $R = 1 + \epsilon$ with $\epsilon > 0$ and small, the inequality $q > 2Rx_1$ is violated by points whose norm tends to zero as $\epsilon \to 0$).

We will show next that if (1) $Q(x)$ grows faster than linearly, in every direction, and (2) $\text{int}(P)$ contains a half-ball with positive radius and center at each point where lifting is performed, then any LFO inequality is binding at (at least) two different points, obtaining Theorem IV of the Introduction. This will be done in Theorem 2.19 below. We first prove a simple technical result concerning the lifting construction. Given a function ψ satisfying (12), for $k > 0$ we define

$$\chi^{\psi}(k) \doteq \sup\{t \geq 0 : \psi(t) < 2kt\}$$

which is finite by (12).

Lemma 2.18 *Assume $Q(x)$ is ψ-strongly convex with ψ satisfying (12). Let $y \in \mathbb{R}^d$ and $0 \neq p \in \mathbb{R}^d$. Suppose $x \in \mathbb{R}^d$ satisfies*

$$Q(x) \quad < \quad Q(y) + \nabla Q(y)^T(x-y) \ + \ 2p^T(x-y).$$

Then $\|x - y\| \leq \chi^{\psi}(\|p\|)$.

Proof. Using ψ-strong convexity yields

$$\begin{aligned} Q(y) + \nabla Q(y)^T(x-y) + \psi(\|x-y\|) \quad &< \quad Q(y) + \nabla Q(y)^T(x-y) \ + \ 2p^T(x-y), \quad \text{i.e.} \\ \psi(\|x-y\|) \quad &< \quad 2p^T(x-y) \ \leq \ 2\|p\|\|x-y\|. \end{aligned}$$

■

Below we will use the following notation: given vectors v and nonzero $p \in \mathbb{R}^d$, and a real $\gamma > 0$, write

$$\mathcal{H}(v, \gamma) \doteq \{x \in \mathbb{R}^d \, : \, p^T(x-v) > 0, \ \|x-v\| < \gamma\},$$

which is an (open) half-ball with center v and radius γ.

Theorem 2.19 *Suppose $Q(x)$ is ψ-strongly convex where ψ satisfies (12). Let $y \in \partial P$ and $p \in \mathbb{R}^d$ satisfy:*

(i) There exists $v \notin P$ such that $p^T(v-y) > 0$.

(ii) There exists a real $\gamma > 0$ such that $\mathcal{H}(y, \gamma) \subseteq P$.

Then P is locally flat at y with normal p, and

(1) there exists $w \notin \text{int}(P)$ with $w \neq y$ and such that the LFO inequality at y is binding at $(w, Q(w))$, i.e.

$$Q(w) \ = \ Q(y) + \nabla Q(y)^T(w-y) \ + \ 2\hat{\alpha}p^T(w-y).$$

(2) Furthermore, $\hat{\alpha} > 0$ and $(w-y)^T p > 0$.

Pproof. Condition (ii) implies that P is locally flat at y. Furthermore, Lemma 2.6 (b) implies that $\hat{\alpha} < +\infty$. But by definition of $\hat{\alpha}$, for any $\epsilon > 0$ there exists $x^{\epsilon} \notin \text{int}(P)$ such that

$$Q(x^{\epsilon}) \quad < \quad Q(y) + \nabla Q(y)^T(x^{\epsilon}-y) \ + \ 2(\hat{\alpha}+\epsilon)p^T(x^{\epsilon}-y). \tag{14}$$

Using Lemma 2.18 we obtain $\|x^{\epsilon} - y\| \leq \chi^{\psi}(\hat{\alpha}+\epsilon)$. Note that for any pair of values $0 < \epsilon < \delta$ we have $\chi^{\psi}(\hat{\alpha}+\epsilon) \leq \chi^{\psi}(\hat{\alpha}+\delta)$. We conclude that since $\mathbb{R}^d - \text{int}(P)$ is closed, as $\epsilon \to 0$ there is an accumulation point $w \in \mathbb{R}^d - \text{int}(P)$ of the points x^{ϵ}. Since $Q(x)$ is continuous, from (14) we obtain

$$Q(w) \ \leq \ Q(y) + \nabla Q(y)^T(w-y) \ + \ 2\hat{\alpha}p^T(w-y). \tag{15}$$

Since (6) is valid at $\alpha = \hat{\alpha}$, and we have $w \in \mathbb{R}^d - \text{int}(P)$, we conclude that (15) holds as an equality. Further, by assumption (ii), $\|w - y\| \geq \gamma > 0$, which implies $w \neq y$, and since ψ is strictly increasing we have (2) as well. ■

Remark 2.20 *Condition (b) is a strengthening on the locally flat requirement in Definition 2.1, which required that for each vector r with with $p^T r > 0$ there exists $\epsilon(r) > 0$ such that $y + \delta r \in \text{int}(P)$ for all $0 < \delta < \epsilon(r)$. In condition (b) we have $\epsilon(r) = \gamma$ for all appropriate r. It is possible to relax condition (b) somewhat, to better account for the relationship between the values $\psi(\epsilon(r))$ and $p^T r$, for all r.*

2.1 Specialization when Q(x) is a positive-definite quadratic

When $Q(x)$ is a positive-definite quadratic the constructions above can be simplified and strengthened. By changing coordinates if necessary we assume $Q(x) = \|x\|^2$, and thus the LFO inequality at a point y with respect to a unit vector p has the form $q \geq \|y\|^2 + 2(y + \hat{\alpha}p)^T(x - y)$. In this section we will show that $\hat{\alpha} > 0$ if and only if there exists a ball $\mathcal{B}$ such that $y \in \partial\mathcal{B}$ and $\mathcal{B} \subseteq P$ (Corollaries 2.22 and 2.23, below). We will also obtain other structural results, in particular a geometrical interpretation of LFO inequalities. In Section 3 we will use these improvements to obtain polynomial-time separation procedures for $\mathrm{conv}(S)$ when P is either a polyhedron or an ellipsoid and $Q(x)$ is positive-definite quadratic.

We will use the following notation: given $\mu \in \mathbb{R}^d$ and $R \geq 0$, we write $\mathcal{B}(\mu, R) = \{x \in \mathbb{R}^d : \|x - \mu\| \leq R\}$. The following property will be used in the sequel.

Remark 2.21 *Let $y \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$. Then $x \in \mathbb{R}^d$ satisfies*

$$\|x\|^2 \quad \leq \quad \|y\|^2 + 2y^T(x - y) \; + \; 2v^T(x - y). \tag{16}$$

if and only if $x \in \mathcal{B}(y + v, \|v\|)$ with equality in (16) iff $\|x - (y + v)\| = \|v\|$.

Proof. We can restate (16) as $\|x\|^2 \leq 2(y + v)^T x - \|y\|^2 - 2y^T v$, from which the result follows. ∎

From this observation we obtain two corollaries:

Corollary 2.22 *Suppose P is locally flat at $y \in \partial P$ with normal p, and that $\hat{\alpha} = \hat{\alpha}(P, p, y) > 0$ is finite. Then:*

(a) $\mathcal{B}(y + \hat{\alpha}p, \hat{\alpha}) \subseteq P$.

Further, suppose that there is a real $\gamma > 0$ such that $\mathcal{H}(y, \gamma) \subseteq P$. Then there exists $z \neq y$, $z \notin \mathrm{int}(P)$ satisfying conditions (b) and (c) given next.

(b) $\|z - (y + \hat{\alpha}p)\| = \hat{\alpha}$, and

$$\|z\|^2 \quad = \quad \|y\|^2 \; + \; 2(y + \hat{\alpha}p)^T(z - y). \tag{17}$$

Further, $z \in \partial P$, and writing $q = z - (y + \hat{\alpha}p)$, P is locally flat at z, with normal $v = q/\|q\|$.

(c) $\hat{\alpha}(P, v, z) = \hat{\alpha}$.

Proof. (a) This follows from Remark 2.21. (b) Let $z \in \mathbb{R}^d - \mathrm{int}(P)$ be a vector that satisfies the LFO inequality at y with respect to p with equality, with $z \neq y$, which is guaranteed to exist by Theorem 2.19. Then by construction z satisfies (17), and by Remark 2.21, $\|z - (y + \hat{\alpha}p)\| = \hat{\alpha}$. Hence by part (a), $z \in \partial\mathcal{B}(y + \hat{\alpha}p, \hat{\alpha})$ and so P is locally flat at z with normal v. (c) By (a) and (b), $\hat{\alpha}(P, v, z) \geq \hat{\alpha}$. But any larger value of $\hat{\alpha}(P, v, z)$ would cut off $(y, \|y\|^2)$. ∎

Corollary 2.23 *Suppose $y \in \partial P$ and let $v \in \mathbb{R}^d - \{0\}$ be such that $\mathcal{B}(y + v, \|v\|) \subseteq P$. Then P is locally flat at y with normal $p = v/\|v\|$ and $\hat{\alpha}(P, p, y) > 0$.*

Proof. Since $\mathcal{B}(y + v, \|v\|) \subseteq P$ it clearly follows that P is locally flat at y with normal p. Moreover, by Remark 2.21, $\|x\|^2 \geq \|y\|^2 + 2y^T(x - y) \; + \; 2v^T(x - y)$ holds for any $x \in \mathbb{R}^d - \mathrm{int}(P)$. Hence $\hat{\alpha}(P, p, y) \geq 0$. ∎

The significance of Corollaries 2.22 and 2.23 is the following: in the case that $Q(x) = \|x\|^2$ an LFO inequality obtained at a point $y \in \partial P$ has positive lifting coefficient if and only if there is a ball $\mathcal{B}(\mu, \sqrt{\rho}) \subset P$, with $\rho > 0$ and such that $\|y - \mu\|^2 = \rho$; furthermore the existence of such a ball implies that P is locally flat at y with normal $(\mu - y)/\|\mu - y\|$. This is a sharpening of Theorem 2.7 in that it simplifies the separation problem for LFO inequalities. For example, if P is a polyhedron then we only need to consider LFO inequalities defined at points y which are in the relative interior of some facet.

To conclude this section we point out a geometric characterization of a set of the form $\mathbb{R}^d - \text{int}(P)$ which can be used to derive valid inequalities for the corresponding set S when $Q(x) = \|x\|^2$. Let P be given. Then for any $x \in \mathbb{R}^d$,

$$x \in \mathbb{R}^d - \text{int}(P) \text{ if and only if } \|x - \mu\|^2 \geq \rho, \text{ for each ball } \mathcal{B}(\mu, \sqrt{\rho}) \text{ contained in } P. \quad (18)$$

As a result

$$q \geq 2\mu^T x + \rho - \|\mu\|^2, \text{ for each ball } \mathcal{B}(\mu, \sqrt{\rho}) \text{ contained in } P \quad (19)$$

is a family of inequalities valid for S. We will term these the *ball inequalities*. In fact, LFO inequalities are ball inequalities: consider the LFO inequality at a point $y \in \partial P$ with normal p and lifting coefficient $\hat{\alpha}$. Writing $\mu = y + \hat{\alpha} p$ and $\rho = \|\mu - y\|^2$, we have by Remark 2.21 that the inequality is violated precisely by those points $(x, \|x\|^2)$ with x in the interior $B(\mu, \sqrt{\rho})$. Thus the LFO inequality can be written

$$q \;\geq\; 2\mu^T x \,+\, \rho \,-\, \|\mu\|^2, \quad (20)$$

which is a ball inequality. Hence, even though (18) provides a geometric characterization of membership in $\mathbb{R}^d - \text{int}(P)$, Theorem 2.7 implies that only a subset of the ball inequalities (19) are needed to characterize $\text{conv}(S)$: the LFO inequalities.

3 Polynomial-time separable cases

In this section we consider two cases where the characterization in Theorem 2.7 of a set $\text{conv}(S)$ with S as in (1) leads to polynomial-time separation algorithms: $Q(x)$ is positive-definite quadratic and P is either a polyhedron or an ellipsoid (handled in Sections 3.1 and 3.2, respectively).

Here we recall the implications of Theorem 2.14 and Corollary 2.15 concerning the separation problem. Given $(x^*, q^*) \in \mathbb{R}^d \times \mathbb{R}$ we can trivially check whether this point satisfies all linearization inequalities simply by checking if it satisfies the linearization inequality at x^*. Assuming that x^* also satisfies all valid inequalities for $\mathbb{R}^d - P$, the only nontrivial case of the separation problem is that where $x^* \in \text{int}(P)$, and we need to verify that (x^*, q^*) satisfies all LFO inequalities.

3.1 Polynomial-time separation of LFOs when P is a polyhedron and Q(x) is positive-definite quadratic

As before we can assume $Q(x) = \|x\|^2$. Without loss of generality $\text{int}(P) \neq \emptyset$ and so P is full-dimensional, so that $P = \{\, x \in \mathbb{R}^d \,:\, a_i^T x \geq b_i,\ 1 \leq i \leq m\}$, where each inequality is facet-defining and nonredundant. We assume $d \geq 2$ and $m \geq 2$.
For $1 \leq i \leq m$ let $\bar{P}^i = \{\, x \in \mathbb{R}^d \,:\, a_i^T x \leq b_i\}$; thus $\mathbb{R}^d - P = \bigcup_i \bar{P}^i$. Further, for $1 \leq i \leq m$ write:

$$\bar{Q}^i = \{\, (x, q) \in \mathbb{R}^d \times \mathbb{R} \,:\, a_i^T x \leq b_i,\ q \geq \|x\|^2\}.$$

Thus, $(x^*, q^*) \in \text{conv}(S)$ if and only if (x^*, q^*) can be written as a convex combination of points in the sets $\bar{Q}^i$. This is the disjunctive approach pioneered in Ceria and Soares [24]. Also see [55], [29], [32]. The resulting separation problem is carried out by solving a second-order cone program with $m(d + 2)$ variables, m conic constraints and $d + 1$ linear constraints, and then using second-order cone duality in order to obtain a linear inequality (details in [42]). In the specific context of this section, the work in this paper can be seen as a generalization of that in [55], [29], [32], all of which use the geometry of $0 - 1$ disjunctions and first-order estimates in order to derive cuts.

Here we will present an algorithm that, given $x^* \in \text{int}(P)$, finds an LFO inequality which proves the strongest lower bound on q at x^*, that is to say, an LFO inequality $q \geq \beta^T x + \beta_0$ whose right-hand side is maximized at x^*. This requires solving $m - 1$ convex quadratic programs, each with

$d+1$ variables and $2m-1$ linear constraints. The quadratic programs are given below in formulation SEP(i), for $1 \leq i < m$. The potential advantage of this approach relative to the use of the disjunctive formulation is twofold: computations are done in the original space of variables, and the separation problem is of a simpler nature. A numerical comparison between the two methods will be provided in Section 5.2.

Our main construction is given in Lemma 3.2 below. In order to motivate our approach we first present some introductory remarks. First, any LFO inequality is generated at some point $y \in \partial P$ where P is locally flat (with some normal p). This property holds iff y is in the relative interior of one of the facets defining P, say the facet corresponding to inequality $a_i^T x \geq b_i$, in which case $p = a_i/\|a_i\|$. Moreover, by Corollary 2.22 there is a ball contained in P, with radius equal to the lifting coefficient, which contains in its boundary both y and another point $z \in \mathbb{R}^d - \text{int}(P)$. Using Corollaries 2.22 and 2.23, necessarily we must then have that z is in the relative interior of another facet of P, say the facet defined by $a_j^T x \geq b_j$, for some $j \neq i$. And, moreover, there is a symmetric relationship between y and z, in the sense that lifting from z will produce the same ball (as per Corollary 2.22) and the same lifting coefficient.

These observations suggest that we explore the interaction between pairs of facets. To that effect, writing for any pair of distinct indices $1 \leq i \leq m$, $1 \leq j \leq m$

$$P^{i,j} \doteq \{x \in \mathbb{R}^d \,:\, a_i^T x \geq b_i,\ a_j^T x \geq b_j\}, \tag{21}$$

we then have:

Proposition 3.1 *Let $1 \leq i \leq m$ and let $y \in P$ be in the relative interior of the facet defined by $a_i^T x \geq b_i$. Then $\hat{\alpha}(P, a_i/\|a_i\|, y) = \min_{j \neq i} \hat{\alpha}(P^{i,j}, a_i/\|a_i\|, y)$.*

Proof sketch. Any $0 \leq \alpha \leq \min_{j \neq i} \hat{\alpha}(P^{i,j}, a_i/\|a_i\|, y)$ is a valid lifting coefficient at y (with respect to the set P); and any larger coefficient will exclude at least one feasible point. ■

We now use these observations to obtain a characterization of the LFO inequalities that leads to polynomial-time separability. In particular, Lemma 3.2 below will allow us to compute the minimum in Proposition 3.1 (for a given $1 \leq i \leq m$) by solving a convex quadratic program with $d+1$ variables and m linear constraints. For $1 \leq i \leq m$ let $H_i \doteq \{ x \in \mathbb{R}^d \,:\, a_i^T x = b_i\}$. For $i \neq j$ let $H_{\{i,j\}} \doteq \{ x \in \mathbb{R}^d \,:\, a_i^T x = b_i,\ a_j^T x = b_j\}$.

Lemma 3.2 *Let $1 \leq i \leq m$, $1 \leq j \leq m$ be distinct and let y be in the relative interior of the facet of P defined by $a_i^T x \geq b_i$. Then there exists polynomially computable $p_{ij} \in \mathbb{R}^d$ and $q_{ij} \in \mathbb{R}$ such that*

$$\hat{\alpha}_{ij} \doteq \hat{\alpha}(P^{i,j}, a_i/\|a_i\|, y) = p_{ij}^T y + q_{ij}.$$

Further, for any $v \in H_{\{i,j\}}$, $p_{ij}^T v + q_{ij} = 0$.

Proof. By Corollary 2.22 $\hat{\alpha}_{ij}$ is equal to the largest radius of a ball that can be inscribed in $P^{i,j}$, with y in its boundary. Thus, if $H_{\{i,j\}} = \emptyset$, i.e. H_i and H_j are parallel, $\hat{\alpha}_{ij}$ equals half the distance between H_i and H_j, and is therefore independent of y, i.e. it is trivially an affine function of y. In what follows we assume $H_{\{i,j\}} \neq \emptyset$. Set

$$\mu \doteq y + \hat{\alpha}_{ij} \frac{a_i}{\|a_i\|}, \quad \text{and} \quad \rho \doteq \hat{\alpha}_{ij}^2 = \|\mu - y\|^2. \tag{22}$$

Then by Corollary 2.22

$$\mathcal{B}(\mu, \sqrt{\rho}) \subseteq P^{i,j}, \quad \mathcal{B}(\mu, \sqrt{\rho}) \cap H_i = y, \quad \text{and} \quad \mathcal{B}(\mu, \sqrt{\rho}) \cap H_j = z, \text{ for some } z \in H_j. \tag{23}$$

We have that $H_{\{i,j\}}$ is $(d-2)-$dimensional (because H_i and H_j are not parallel). Denote by ω_{ij} the unique unit norm vector orthogonal to both $H_{\{i,j\}}$ and a_i (it is unique up to reversal), and by Ω_{ij} be the 2-dimensional hyperplane through μ generated by a_i and ω_{ij}. By construction Ω_{ij} is orthogonal to $H_{\{i,j\}}$ and is thus the orthogonal complement to $H_{\{i,j\}}$ through μ. It follows that $\Omega_{ij} = \Omega_{ji}$ and by (23) that this hyperplane contains the orthogonal projection of μ onto H_i (which is y) and the orthogonal projection of μ onto H_j (which is z). Further, $\Omega_{ij} \cap H_{\{i,j\}}$ consists of a single point $k_{\{i,j\}}$ satisfying

$$\begin{aligned} \|\mu - k_{\{i,j\}}\|^2 &= \|\mu - y\|^2 + \|y - k_{\{i,j\}}\|^2 \\ &= \|\mu - z\|^2 + \|z - k_{\{i,j\}}\|^2. \end{aligned} \tag{24}$$

Moreover

$$\begin{aligned} & y - k_{\{i,j\}} \text{ is parallel to } \omega_{ij} \text{ and } z - k_{\{i,j\}} \text{ is parallel to } \omega_{ji}, \\ & \|\mu - y\|^2 = \|\mu - z\|^2 = \rho, \quad \text{and by (24)}, \\ & \|y - k_{\{i,j\}}\| = \|z - k_{\{i,j\}}\|, \quad \text{and} \quad \|\mu - y\| = \tan\phi\, \|y - k_{\{i,j\}}\|, \end{aligned} \tag{25}$$

where 2ϕ is the angle formed by ω_{ij} and ω_{ji}. Let $h^g_{\{i,j\}}$ $(1 \le g \le d-2)$ be a basis for $\{\, x \in \mathbb{R}^d : a_i^T x = a_j^T x = 0\}$. Then a_i, together with ω_{ij} and the $h^g_{\{i,j\}}$ form a basis for $\mathbb{R}^d$. Let

- O_i be the orthogonal projection of the origin onto H_i – hence O_i is a multiple of a_i,
- N_i be the orthogonal projection of O_i onto $H_{\{i,j\}}$.

We have

$$y \;=\; O_i \;+\; (N_i - O_i) \;+\; (k_{\{i,j\}} - N_i) \;+\; (y - k_{\{i,j\}}) \tag{26}$$

and thus, since $N_i - O_i$ and $y - k_{\{i,j\}}$ are parallel to ω_{ij}, and $k_{\{i,j\}} - N_i$ and O_i are orthogonal to ω_{ij},

$$\omega_{ij}^T y \;=\; \omega_{ij}^T(N_i - O_i) + \omega_{ij}^T(y - k_{\{i,j\}}) \;=\; \omega_{ij}^T(N_i - O_i) + \|\omega_{ij}\|\|y - k_{\{i,j\}}\|, \tag{27}$$

or

$$\|y - k_{\{i,j\}}\| \;=\; \|\omega_{ij}\|^{-1}\omega_{ij}^T\,(y - N_i + O_i). \tag{28}$$

Consequently, by (25) and (22)

$$\begin{aligned} \hat{\alpha}(P^{i,j}, y) &= \tan\phi\, \|y - k_{\{i,j\}}\| \\ &= \tan\phi\, \|\omega_{ij}\|^{-1}\omega_{ij}^T\,(y - N_i + O_i), \end{aligned} \tag{29}$$

which is affine as desired. Note that for any $v \in H_{\{i,j\}}$ the lifting coefficient at v as per (29) is zero, since ω_{ij} is orthogonal to both a_i and $H_{\{i,j\}}$. Finally, since $\tan\phi$, ω_{ij}, N_i and O_i are all polynomially computable the proof is now complete. ∎

Now let $x^* \in \text{int}(P)$. The problem of finding the strongest possible lifted first-order inequality at x^* chosen from among those obtained by lifting from a point on the facet defined by the i^{th} inequality can thus be written as follows:

$$\begin{aligned} \textbf{SEP(i):} \quad \min \quad & -2y^T x^* \;+\; \|y\|^2 \;-\; 2\alpha(a_i^T x^* - b_i) \\ s.t. \quad & y \in P \\ & a_i^T y = b_i \\ & 0 \le \alpha \le p_{ij}^T y + q_{ij} \quad \forall\; j \ne i. \end{aligned}$$

[Here, the last constraint is valid by the last part of the statement of Lemma 3.2.] This is a linearly constrained, convex quadratic program with $d+1$ variables and $2m-1$ constraints. As a result we have:

Theorem 3.3 *Given $x^* \in \text{int}(P)$, in polynomial-time we can compute an LFO inequality (7) that attains the largest right-hand side value at $x = x^*$.*

Proof. We can attain the desired goal by solving SEP(i) for each choice of $1 \le i < m$. ∎

3.2 Polynomial-time separation of LFOs when P is an ellipsoid and Q(x) is positive-definite quadratic

In this section we will discuss a polynomial-time separation procedure for LFO inequalities in the case that P is an ellipsoid with nonempty interior. As before we assume without loss of generality that $Q(x) = \|x\|^2$. Write

$$P \quad = \quad \{x \in \mathbb{R}^d \,:\, x^T A x \,-\, 2b^T x \,+\, c \,\leq\, 0\} \tag{30}$$

for appropriate $A \succ 0$, c and b.[2] To address the separation of LFO inequalities, consider a given a point $\bar{x} \in \mathrm{int}(P)$. Using equation (20), the problem of finding an LFO inequality $q \geq \beta^T x + \beta_0$ whose right-hand side is maximized at $\bar{x}$ can be written as:

$$\bar{V}^{LFO} \doteq \min_{\mu,\rho} \quad \|\mu\|^2 - \rho - 2\bar{x}^T\mu \tag{31}$$

$$\text{s.t.} \quad \{x \in \mathbb{R}^d : \|x - \mu\|^2 \leq \rho\} \subseteq P \tag{32}$$

$$\mu \in \mathbb{R}^d, \quad \rho \geq 0. \tag{33}$$

[**Remark.** $-\bar{V}^{LFO}$ is the maximum right-hand side of an LFO inequality at $\bar{x}$.]

Since P is compact, problem (31)-(33) is well-posed, that is to say the objective is a "min" rather than simply an "inf". Denoting the maximum eigenvalue of A by λ_{max}, we will show in Theorem 3.9 given below that an optimal solution to this problem is given by

$$\begin{aligned} \hat{\mu} &= \lambda_{max}^{-1} b + (I - \lambda_{max}^{-1} A)\bar{x}, \\ \hat{\rho} &= \|\hat{\mu} - \bar{x}\|^2 - \lambda_{max}^{-1}(\bar{x}^T A \bar{x} - 2b^T\bar{x} + c). \end{aligned}$$

A numerical application of this result to cardinality-constrained convex quadratic programs will be provided in Section 5.1.

For simplicity, in what follows we will refer to problem (31)-(33) as *the LFO separation problem at* $\bar{x}$, or the LFO separation problem for short. Below we will provide a series of steps, culminating in Theorem 3.9, that yield the description of $\hat{\mu}$ and $\hat{\rho}$ given above. First we present two technical results that will be used in the sequel.

Lemma 3.4 *Let $\tilde{\mu} \in \mathbb{R}^d$ and let $\tilde{A}$ a symmetric $d \times d$ matrix. The following two statements are equivalent:*

(i) $\quad \min_{x \in \mathbb{R}^d} \{\, x^T \tilde{A} x - 2\tilde{\mu}^T x \,\} > -\infty.$

(ii) $\quad \tilde{A} \succeq 0$, *and there exists* $\pi \in \mathbb{R}^d$ *satisfying* $\tilde{A}\pi = \tilde{\mu}$.

Furthermore, when (i) holds, a vector π is an optimal solution to the minimization in (i) if and only if $\tilde{A}\pi = \tilde{\mu}$, in which case

$$\min_{x \in \mathbb{R}^d} \{\, x^T \tilde{A} x - 2\tilde{\mu}^T x \,\} \;=\; -\pi^T \tilde{A} \pi.$$

Proof. Suppose (i) holds. Then clearly $\tilde{A} \succeq 0$ and for any $\delta \in \mathbb{R}^d$,

$$\tilde{\mu}^T \delta = 0 \quad \text{whenever} \quad \tilde{A}\delta = 0.$$

Farkas's Lemma then implies that there exists $\pi \in \mathbb{R}^d$ such that $\tilde{A}\pi = \tilde{\mu}$. Thus (ii) holds; the proof that (ii) implies (i) is similar and will be omitted. Moreover, if (i) holds then (ii) does, and since the quadratic minimized in (i) is convex, $x \in \mathbb{R}^d$ is an optimal solution to the minimization in (i) iff $\tilde{A}x = \tilde{\mu}$. ■

[2] A generalization to the case $A \succeq 0$ is possible, but omitted for brevity.

Lemma 3.5 *Let $\tilde{v} \in \mathbb{R}^d$ and let $\tilde{A} \succeq 0$ be a $d \times d$ matrix with maximum eigenvalue $\tilde{\lambda}_{max} > 0$. Suppose $0 \leq \theta \leq \tilde{\lambda}_{max}^{-1}$. Then $x = \tilde{v}$ is an optimal solution to the problem*

$$\min_{x \in \mathbb{R}^d} \left\{ x^T (I - \theta \tilde{A})^T x \ - 2\tilde{v}^T (I - \theta \tilde{A}) x \right\}. \tag{34}$$

Proof. Since $\theta \leq \tilde{\lambda}_{max}^{-1}$ we have $I - \theta\tilde{A} \succeq 0$. The result now follows from Lemma 3.4 with $\tilde{\mu} = (I - \theta\tilde{A})\tilde{v}$. ■

We now return to the LFO separation problem at a given point $\bar{x} \in \text{int}(P)$; we will show that it is equivalent to:

$$\hat{V} \doteq \min_{\mu, \theta} \left[-2\bar{x}^T \mu - \min_x \{ x^T (I - \theta A) x - 2(\mu - \theta b)^T x \} - \theta c \right] \tag{35a}$$

$$\text{s.t.} \quad \mu \in \mathbb{R}^d, \quad 0 \leq \theta \leq \lambda_{max}^{-1}. \tag{35b}$$

Our equivalence proof will proceed in several steps, given by Lemmas 3.6 and 3.7.

Lemma 3.6 *We have that $\hat{V} \leq \bar{V}^{LFO}$. Further, suppose problem (35) has an optimal solution $(\hat{\mu}, \hat{\theta})$ such that $\hat{\rho} \geq 0$, where*

$$\hat{\rho} \doteq \|\hat{\mu}\|^2 + \min_x \{ x^T (I - \hat{\theta} A) x - 2(\hat{\mu} - \hat{\theta} b)^T x \} - \hat{\theta} c. \tag{36}$$

Then $\hat{V} = \bar{V}^{LFO}$ and $(\hat{\mu}, \hat{\rho})$ is optimal for the LFO separation problem.

Proof. Since we assume $\text{int}(P) \neq \emptyset$, the S-Lemma (see [59], [45], [13]) implies that (μ, ρ) satisfies (32) if and only if there is some nonnegative real $\theta = \theta(\mu, \rho)$ such that

$$\|x - \mu\|^2 - \rho - \theta(x^T A x - 2b^T x + c) \geq 0 \quad \forall x \in \mathbb{R}^d.$$

This is equivalent to saying that there is $\theta \geq 0$ with

$$\min_x \{ \|x - \mu\|^2 - \rho - \theta(x^T A x - 2b^T x + c) \} \ \geq \ 0,$$

or equivalently

$$\min_x \{ x^T (I - \theta A) x - 2(\mu - \theta b)^T x + \|\mu\|^2 - \rho - \theta c \} \ \geq \ 0. \tag{37}$$

We clearly must have $\theta \leq \lambda_{max}^{-1}$ for (37) to hold. We can now write the LFO separation problem as:

$$\min_{\mu, \rho, \theta} \quad \|\mu\|^2 - \rho - 2\bar{x}^T \mu \tag{38a}$$

$$\text{s.t.} \quad \|\mu\|^2 - \rho + \min_x \{ x^T (I - \theta A) x - 2(\mu - \theta b)^T x \} - \theta c \ \geq \ 0 \tag{38b}$$

$$\mu \in \mathbb{R}^d, \quad \rho \geq 0, \quad 0 \leq \theta \leq \lambda_{max}^{-1}. \tag{38c}$$

We now use this formulation to argue that $\hat{V} \leq \bar{V}^{LFO}$. Note that in any feasible solution to (38) the minimum in (38b) must be finite and is therefore attained (as the quantity being minimized is a quadratic). Hence, without loss of generality, (38b) will hold with equality (or we could increase ρ); using this fact we can eliminate $\|\mu\|^2 - \rho$ from the objective. Thus $\hat{V} \leq \bar{V}^{LFO}$, as desired. Moreover if an optimal solution $(\hat{\mu}, \hat{\theta})$ for problem (38) is such that $\hat{\rho}$ defined as in (36) is nonnegative, then clearly $(\hat{\mu}, \hat{\theta}, \hat{\rho})$ is feasible for (38), with value at most $\hat{V}$. ■

Our next task is to further simplify problem (35). Toward this goal we seek an alternative description of the inner minimization in (35a); in particular a characterization of those cases when it has finite value. For $0 \leq \theta \leq \lambda_{max}^{-1}$ define

$$F(\theta) \ \doteq \ \min_{\pi \in \mathbb{R}^d} \left\{ -2\bar{x}^T \left[\theta b + (I - \theta A)\pi \right] + \pi^T (I - \theta A)^T \pi \right\} \ + \ \theta c. \tag{39}$$

We have:

Lemma 3.7 *(a) Problem (35) can be equivalently rewritten as*

$$\min_{0\le\theta\le\lambda_{max}^{-1}} F(\theta). \tag{40}$$

(b) Further, suppose $\hat{\theta}$ is an optimal solution to (40), and let $\hat{\pi}$ attain the minimum in (39) when $\theta = \hat{\theta}$. If

$$\tilde{\rho} \ \doteq \ \|\hat{\mu} - \hat{\pi}\|^2 - \hat{\theta}(\hat{\pi}^T A\hat{\pi} - 2b^T\hat{\pi} + c) \ \ge \ 0 \tag{41}$$

then $(\hat{\mu}, \tilde{\rho})$ is an optimal solution for the LFO separation problem, where

$$\hat{\mu} = \hat{\theta}b + (I - \hat{\theta}A)\hat{\pi}. \tag{42}$$

Proof. We have $\hat{V} \le 0$, since a feasible solution for the LFO separation problem is $\mu = \bar{x}$ and $\rho = 0$ with objective value $-\|\bar{x}\|^2$. It follows that the inner minimum in (35a) can be assumed to be finite. Thus, by Lemma 3.4 we have that

$$\operatorname*{argmin}_x\{x^T(I - \theta A)x - 2(\mu - \theta b)^T x\} \ = \ \{\pi \in \mathbb{R}^d \ : \ (I - \theta A)\pi = \mu - \theta b\}. \tag{43}$$

Any π in the set in (43) clearly satisfies

$$-\pi^T(I - \theta A)\pi - \theta c \ = \ \min_x\{x^T(I - \theta A)x - 2(\mu - \theta b)^T x\} - \theta c.$$

Consequently, problem (35) can be equivalently rewritten as

$$\min_{0\le\theta\le\lambda_{max}^{-1}} G(\theta), \tag{44}$$

where

$$G(\theta) \quad \doteq \quad \min_{\mu,\pi} \quad -2\bar{x}^T\mu \ + \ \pi^T(I - \theta A)\pi + \theta c \tag{45a}$$

$$\text{s.t.} \quad \mu - \theta b = (I - \theta A)\pi \tag{45b}$$

$$\mu, \, \pi \in \mathbb{R}^d. \tag{45c}$$

Substituting (45b) into (45a) we obtain that problem (40) is equivalent to problem (44). This proves part (a). To prove (b), we have that at $\theta = \hat{\theta}$ the optimum solution to (45) is $(\hat{\mu}, \hat{\pi})$ which satisfy equation (45b), i.e. expression (42). Further, by Lemma 3.6, if we can argue that

$$\hat{\rho} \ \doteq \ \|\hat{\mu}\|^2 + \min_x\{x^T(I - \hat{\theta}A)x - 2(\hat{\mu} - \hat{\theta}b)^T x\} - \hat{\theta}c \ \ge \ 0,$$

then $(\hat{\mu}, \hat{\rho})$ is optimal for the LFO problem. But using (43) and (42), $\hat{\pi}$ solves $\min_x\{x^T(I - \hat{\theta}A)x - 2(\hat{\mu} - \hat{\theta}b)^T x\}$, and so

$$\hat{\rho} = \|\hat{\mu}\|^2 + \hat{\pi}^T(I - \hat{\theta}A)\hat{\pi} - 2(\hat{\mu} - \hat{\theta}b)^T\hat{\pi} \ - \ \hat{\theta}c \ = \tilde{\rho}$$

as defined in (41), which is nonnegative by assumption. ■

The following result characterizes $F(\theta)$.

Lemma 3.8 *Let $0 \le \theta \le \lambda_{max}^{-1}$. Then $F(\theta)$ is obtained by choosing $\pi = \bar{x}$ in (39). Thus*

$$F(\theta) = -\|\bar{x}\|^2 + \theta(\bar{x}^T A\bar{x} - 2b^T\bar{x} + c). \tag{46}$$

Proof. The result follows from Lemma 3.5, with $\tilde{v} = \bar{x}$. ■

Theorem 3.9 *The optimizer for (40) is* $\hat{\theta} = \lambda_{max}^{-1}$*, and an optimal solution to the LFO separation problem at* $\bar{x}$ *is*

$$\hat{\mu} = \lambda_{max}^{-1} b + (I - \lambda_{max}^{-1} A)\bar{x}, \tag{47}$$
$$\hat{\rho} = \|\hat{\mu} - \bar{x}\|^2 - \lambda_{max}^{-1}(\bar{x}^T A \bar{x} - 2b^T \bar{x} + c). \tag{48}$$

Proof. By assumption $\bar{x} \in \text{int}(P)$. Hence the multiplier of θ in (46) is negative; consequently $F(\theta)$ is minimized at $\hat{\theta} = \lambda_{max}^{-1}$. The result now follows from Lemmas 3.8 and 3.7, since the right-hand side of (48) is nonnegative because $\bar{x} \in \text{int}(P)$. ■

Substituting (47) and (48) into (31) we obtain that $-\bar{V}^{LFO}$, which we defined as the maximum right-hand side of an LFO inequality at $\bar{x}$, satisfies

$$-\bar{V}^{LFO} = \bar{x}^T \bar{x} - \lambda_{max}^{-1} \left(\bar{x}^T A \bar{x} - 2b^T \bar{x} + c \right).$$

This closed-form expression for $-\bar{V}^{LFO}$ yields a compact description for the convex hull of the set S in (1)

Corollary 3.10 *Let* $S = \{(x, q) \in \mathbb{R}^d \times \mathbb{R} : q \geq Q(x),\ x \in \mathbb{R}^d - \text{int}(P)\}$ *where* $Q(x)$ *is a positive-definite quadratic and* $P = \{x \in \mathbb{R}^d : x^T A x - 2b^T x + c \leq 0\}$ *with* $A \succ 0$ *and maximum eigenvalue* $\lambda_{\max}$*. Then*

$$\text{conv}\,(S) = \{(x, z) \in \mathbb{R}^{d+1} \mid z \geq Q(x),\ z \geq Q(x) - \lambda_{max}^{-1} \left(x^T A x - 2b^T x + c \right)\}.$$

A related result has been recently obtained in [43].

Remark. As discussed in Section 1.0.1 we are interested in iterative tightening procedures for optimization problems that construct, at each iteration, a set P that contain the solution x^* to the current relaxation but does not intersect the feasible region, and use geometry to obtain valid inequalities that cut-off x^*. In the context of this Section, the sets P would be of the form (30), with different A, b and c encountered at different iterations. In that situation, relying on Corollary 3.10 would produce relaxations with possibly a large number of quadratic constraints. Such relaxations may prove very difficult for current solvers (see Section 5). In contrast, a reliance on (linear) cutting-planes may produce formulations that are more practicable.

4 Tightening a general quadratic expression

Consider a set of the form

$$\Pi \doteq \{ (x, w, q) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \ : \ q \geq x^T H x + h^T x, \ \ w \leq x^T A x \} \tag{49}$$

where $H \succ 0$ and $A \succeq 0$ are $n \times n$ matrices. With $P = \{(x, w) \in \mathbb{R}^n \times \mathbb{R} : x^T A x \leq w\}$, the set Π is an example of our general set S as in (1). We will assume that A has positive largest eigenvalue λ_{max}. Here we show how the specialization of the LFO inequalities to this case leads to a polynomial-time separable characterization of $\text{conv}(\Pi)$.

As motivation for this study, consider an optimization problem of the form

$$\min\{ f(x) \ : \ x^T M x \ + \ h^T x \leq b_0,\ x \in F \} \tag{50}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $M \in \mathbb{R}^n \times \mathbb{R}^n$ and $F \subseteq \mathbb{R}^n$, and $b_0 \in \mathbb{R}$. We now apply a "d.c." (difference between convex) step (see [6]): we can always find matrices $H \succ 0$ and $A \succ 0$ such that $x^T M x = x^T H x - x^T A x$ for all $x \in \mathbb{R}^n$. Thus, (50) can be restated as

$$\min\{ f(x) \ : \ q \geq x^T H x + h^T x,\ w \leq x^T A x,\ q - w \leq b_0,\ x \in F\},$$

which can be relaxed to

$$\min\{ f(x) \ : \ (x, w, q) \in \text{conv}(\Pi),\ q - w \leq b_0,\ x \in F, q \in \mathbb{R}, w \in \mathbb{R} \}, \tag{51}$$

where Π is as in (49). A polynomial-time separation procedure for conv(Π) can thus be used as a component in an algorithm for solving the relaxation (51).

In the remainder of this section we address the separation problem for a set conv(Π) with Π as in (49). We stress that we only require $A \succeq 0$, with positive maximum eigenvalue λ_{max}, rather than $A \succ 0$ as in the above paragraph. In Section 4.0.1 we first describe LFO inequalities as they pertain to the set Π given above. In Section 4.0.2 we then introduce another set of valid inequalities, the *paraboloid* inequalities, which are valid for conv(Π). We then show that, effectively, the strongest paraboloid inequality at any given point is an LFO inequality. Finally, in Section 4.0.3 we show how to separate over paraboloid inequalities in polynomial time.

4.0.1 LFO inequalities for conv(Π)

Let

$$P \doteq \{(x,w) \in \mathbb{R}^n \times \mathbb{R}_+ \,:\, x^T A x \leq w\},$$

and

$$\Pi = \{(x,w,q) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \,:\, q \,\geq\, x^T H x + h^T x,\ (x,w) \in \mathbb{R}^n \times \mathbb{R} - \operatorname{int}(P)\}.$$

By Theorem 2.7 we have that conv(Π) is described by linearization and LFO inequalities. Since ∂P is a differentiable manifold homeomorphic to $\mathbb{R}^n$, it follows that P is locally flat at every point $(\bar{x}, \bar{x}^T A \bar{x}) \in \mathbb{R}^n \times \mathbb{R}$, using as normal the unit vector in the direction of

$$\begin{pmatrix} -2A\bar{x} \\ 1 \end{pmatrix}.$$

Thus in the construction of the LFO inequality at $(\bar{x}, \bar{x}^T A \bar{x})$ we consider inequalities of the form

$$q \;\geq\; \bar{x}^T H \bar{x} + h^T \bar{x} \;+\; \left(\begin{bmatrix} 2H\bar{x} + h \\ 0 \end{bmatrix} - \alpha \begin{bmatrix} 2A\bar{x} \\ -1 \end{bmatrix} \right)^T \begin{bmatrix} x - \bar{x} \\ w - \bar{x}^T A \bar{x} \end{bmatrix}, \tag{52}$$

or in other words

$$\begin{aligned} q \;&\geq\; \bar{x}^T H \bar{x} + h^T \bar{x} + (2H\bar{x} + h)^T (x - \bar{x}) \;+\; \alpha(-2\bar{x}^T A (x - \bar{x}) \,+\, w - \bar{x}^T A \bar{x}) \\ &=\; -\bar{x}^T H \bar{x} \,+\, (2H\bar{x} + h - 2\alpha A \bar{x})^T x + \alpha(\bar{x}^T A \bar{x} + w); \end{aligned} \tag{53}$$

when $\alpha = \hat{\alpha}$ we obtain the LFO inequality. As can be seen, (53) strengthens the linearized inequality $q \geq \bar{x}^T H \bar{x} + h^T \bar{x} + (2H\bar{x} + h)^T (x - \bar{x})$ in the (excluded) region where $w > x^T A x$.

To simplify the discussion below we will assume, without loss of generality, that $H = I$. Thus the LFO inequality at $(\bar{x}, \bar{x}^T A \bar{x})$ becomes

$$q \;\;\geq\;\; (2\bar{x} + h - 2\hat{\alpha} A \bar{x})^T x + \hat{\alpha} w + \hat{\alpha} \bar{x}^T A \bar{x} - \|\bar{x}\|^2. \tag{54}$$

4.0.2 Paraboloid inequalities

In order to obtain a separation algorithm for inequalities (54) we will first derive a geometrical characterization of the set P, similar to that involving the ball inequalities developed in Section 2.1, equation (18). The intuitive reason that paraboloid inequalities supersede ball inequalities is that the latter arose in the context of our generic framework (1) with $Q(x)$ positive-definite. Given $\mu \in \mathbb{R}^n$, $\nu \in \mathbb{R}_+$ and $\alpha > 0$, let

$$\Gamma^{\mu,\nu,\alpha} \;\;\doteq\;\; \{\, (x,w) \,\in\, \mathbb{R}^n \times \mathbb{R} \;:\; \|x - \mu\|^2 + \alpha\nu \;\leq\; \alpha w \,\}, \tag{55}$$

which defines a paraboloid. For technical reasons, for $\mu \in \mathbb{R}^n$, $\nu \in \mathbb{R}_+$ we also set

$$\Gamma^{\mu,\nu,0} \;\;\doteq\;\; \{\, (x,w) \,\in\, \mathbb{R}^n \times \mathbb{R}_+ \;:\; x = \mu,\ \nu \leq w \} \tag{56}$$

[This is different from what the behavior of (55) would achieve with $\alpha = 0$, which is the set $\{(x, w) \in \mathbb{R}^n \times \mathbb{R}_+ \,:\, x = \mu\}$. On the other hand (56) better reflects the limiting behavior of (55) at $x = \mu$ as $\alpha \to 0^+$.] Then it can be seen (proof omitted for brevity) that for any point $(x, w) \in \mathbb{R}^n \times \mathbb{R}$,

$$(x, w) \in \mathbb{R}^n \times \mathbb{R} - \text{int}(P) \quad \text{iff} \quad (x, w) \in \mathbb{R}^n \times \mathbb{R} - \text{int}(\Gamma^{\mu,\nu,\alpha}), \quad \text{for all } (\mu, \nu, \alpha) \text{ such that } \Gamma^{\mu,\nu,\alpha} \subseteq P.$$

Using this characterization we have that for each triple $(\mu, \nu, \alpha) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+$ with $\Gamma^{\mu,\nu,\alpha} \subseteq P$ the following inequality is valid for Π:

$$\mathcal{P}(\mu, \nu, \alpha) : \qquad q \quad \geq \quad (2\mu + h)^T x + \alpha w \,-\, \|\mu\|^2 \,-\, \alpha\nu,$$

which cuts-off $\text{int}(\Gamma^{\mu,\nu,\alpha})$ as given by the following result.

Proposition 4.1 *Let* $(\mu, \nu, \alpha) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+$, *and* $(\bar{x}, \bar{w}) \in \mathbb{R}^n \times \mathbb{R}$. *(a) Suppose* $(\bar{x}, \bar{w}) \in \text{int}(\Gamma^{\mu,\nu,\alpha})$. *If* $\bar{q} \leq \|\bar{x}\|^2 + h^T \bar{x}$ *then* $(\bar{x}, \bar{w}, \bar{q})$ *violates* $\mathcal{P}(\mu, \nu, \alpha)$. *(b)* $(\bar{x}, \bar{w}, \|\bar{x}\|^2 + h^T \bar{x})$ *violates* $\mathcal{P}(\mu, \nu, \alpha)$ *iff* $(\bar{x}, \bar{w}) \in \text{int}(\Gamma^{\mu,\nu,\alpha})$.

Proof. (a) Since $(\bar{x}, \bar{w}) \in \text{int}(\Gamma^{\mu,\nu,\alpha})$, $-\alpha\bar{w} < -\|\bar{x} - \mu\|^2 - \alpha\nu$, and so

$$0 \,>\, \|\bar{x}\|^2 - 2\mu^T \bar{x} + \|\mu\|^2 + \alpha(\nu - \bar{w}) \;\geq\; \bar{q} \,+\, \|\mu\|^2 - (2\mu + h)^T \bar{x} + \alpha(\nu - \bar{w})$$

since $\bar{q} \leq \|\bar{x}\|^2 + h^T \bar{x}$. (b) The proof of this fact is similar to that of (a) and will be omitted for brevity. ■

We term $\mathcal{P}(\mu, \nu, \alpha)$ a *paraboloid* inequality, as an extension of the ball inequalities (20). We stress that paraboloid inequalities are only defined for triples $(\mu, \nu, \alpha) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+$ such that $\Gamma^{\mu,\nu,\alpha} \subseteq P$. In particular, we must have $\alpha \leq \lambda_{max}^{-1}$, where as stated before λ_{max} is the largest eigenvalue of A.

For future reference, we state the following result which follows directly from the definition of Π and the paraboloid inequalities:

Remark 4.2 $\mathcal{P}(\mu, \nu, \alpha)$ *supports* Π *if and only if there exists* $\bar{x} \in \mathbb{R}^n$ *such that*

$$\|\bar{x} - \mu\|^2 = \alpha(\bar{x}^T A \bar{x} - \nu),$$

in which case $\mathcal{P}(\mu, \nu, \alpha)$ *is valid for* Π *and binding at* $(\bar{x}, \bar{x}^T A \bar{x}, \|\bar{x}\|^2 + h^T \bar{x})$.

In the next sequence of results we show that LFO and paraboloid inequalities are essentially equivalent. We will first analyze those paraboloid inequalities that are 'strongest' at a given point, and then consider paraboloid inequalities that are supporting for Π. The final proof will be provided in Theorem 4.9 below.

We will use the following notation:

$$U \doteq \{\, (\mu, \nu, \alpha) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ \,:\, \Gamma^{\mu,\nu,\alpha} \subseteq P \,\}.$$

Proposition 4.3 U *is closed.*

Proof. Let $(\breve{\mu}, \breve{\nu}, \breve{\alpha}) \in \text{cl}(U) - U$. Suppose first that $\breve{\alpha} > 0$. Since $\Gamma^{\breve{\mu},\breve{\nu},\breve{\alpha}} \not\subseteq P$, there exists (x, w) with

$$w \,<\, x^T A x \;\text{ and }\; \|x - \breve{\mu}\|^2 + \breve{\alpha}\breve{\nu} \,\leq\, \breve{\alpha} w.$$

It follows that we can find $\epsilon > 0$ and $\delta > 0$ such that for any (μ, ν) with $\|\mu - \breve{\mu}\| \leq \delta$ and $|\nu - \breve{\nu}| \leq \delta$ we nevertheless still have

$$w + \epsilon \,<\, x^T A x \;\text{ and }\; \|x - \mu\|^2 + \alpha\nu \,\leq\, \alpha(w + \epsilon),$$

a contradiction. The case $\breve{\alpha} = 0$ is similar and will be omitted. ■

Proposition 4.4 *Let $(\tilde{x}, \tilde{w}) \in \mathbb{R}^n \times \mathbb{R}$. Then there is a paraboloid inequality whose right-hand side evaluated at $(\tilde{x}, \tilde{w})$ is maximum among all paraboloid inequalities.*

Proof. The right-hand side of $\mathcal{P}(\mu, \nu, \alpha)$ at $(\tilde{x}, \tilde{w})$ equals $-\|\tilde{x} - \mu\|^2 + \|\tilde{x}\|^2 + h^T \tilde{x} + \alpha \tilde{w} - \alpha\nu$; we want to maximize this expression subject to $(\mu, \nu, \alpha) \in U$. Removing constants from the expression, this is equivalent to maximizing

$$-\|\tilde{x} - \mu\|^2 + \alpha\tilde{w} - \alpha\nu \tag{57}$$

subject to $(\mu, \nu, \alpha) \in U$. A feasible choice for the maximization is $(\mu, \nu, \alpha) = (\tilde{x}, \tilde{x}^T A \tilde{x}, 0)$ (which is in U, by construction) for which (57) attains value 0. Suppose

$$\kappa \ \doteq\ \sup\{ -\|\tilde{x} - \mu\|^2 + \alpha\tilde{w} - \alpha\nu \ : \ (\mu, \nu, \alpha) \in U \} \ > \ 0. \tag{58}$$

We will argue that in this case when computing the supremum in (58) we can assume that (μ, ν, α) can be constrained to lie in a bounded set; since by Proposition 4.3, U is closed, it will follow that there the supremum is achieved, as desired.

So assume (58) holds. Now whenever $(\mu, \nu, \alpha) \in U$ we have $0 \leq \alpha \leq \lambda_{max}^{-1}$ and $\nu \geq 0$. Since $\kappa > 0$, nonnegativity of α implies that $\tilde{w} > 0$, and since $\kappa > 0$ we have $\nu \leq \bar{w}$. As a consequence $\alpha\bar{w} - \alpha\nu$ is bounded (above and below) and so $\|\mu\|$ must be bounded as well. This concludes the proof. ∎

Proposition 4.4 motivates the following definition.

Definition 4.5 *Let $(\tilde{x}, \tilde{w}) \in \mathbb{R}^n \times \mathbb{R}$. Then a paraboloid inequality whose right-hand side evaluated at $(\tilde{x}, \tilde{w})$ is largest from among all paraboloid inequalities is called a strongest paraboloid inequality at $(\tilde{x}, \tilde{w})$.*

Lemma 4.6 *Suppose the inequality $\mathcal{P}(\mu, \nu, \alpha)$ does not support Π. Then $\alpha > 0$, and there exists $\epsilon > 0$ such that $\mathcal{P}(\mu, \nu - \epsilon, \alpha)$ is a paraboloid inequality supporting Π.*

Proof. If $\alpha = 0$ then by definition $\Gamma^{\mu,\nu,\alpha} = \{(x, w) \in \mathbb{R}^n \times \mathbb{R} \, : \, x = \mu, \, w \geq \nu\}$ and trivially $\mathcal{P}(\mu, \nu, \alpha)$ holds as an equality at any point of the form $(\mu, w, \|\mu\|^2 + h^T \mu)$. Thus $\alpha > 0$. Since $\mathcal{P}(\mu, \nu, \alpha)$ does not support Π, by Remark 4.2 it follows that

$$\|x - \mu\|^2 - \alpha x^T A x \ > \ -\alpha\nu, \ \ \forall \, x \in \mathbb{R}^n, \tag{59}$$

from which we obtain $\nu > 0$. The left-hand side of (59) is a quadratic which is lower bounded by the quantity on the right-hand side. Thus, the minimum of the quadratic is attained at some $x^0 \in \mathbb{R}^n$, and since at $x = \mu$ the quadratic takes nonpositive value, it does so at x^0 as well. Therefore there exists $\epsilon > 0$ such that

$$\|x - \mu\|^2 - \alpha x^T A x \ \geq \ -\alpha(\nu - \epsilon), \ \ \forall \, x \in \mathbb{R}^n,$$

with equality at x^0. Since at x^0 the quadratic is nonpositive and $\alpha > 0$, we have $\nu - \epsilon \geq 0$. It follows that $P(\mu, \nu - \epsilon, \alpha)$ is a paraboloid inequality supporting Π. ∎

Corollary 4.7 *Let $(\bar{x}, \bar{w}) \in \mathbb{R}^n \times \mathbb{R}$. Then there is a paraboloid inequality supporting Π that is a strongest paraboloid inequality at $(\bar{x}, \bar{w})$.*

Proof. A strongest paraboloid inequality $\mathcal{P}(\mu, \nu, \alpha)$ at $(\bar{x}, \bar{w})$ exists by Proposition 4.4. If this inequality does not support Π, then by Proposition 4.6 $\alpha > 0$ and $\mathcal{P}(\mu, \nu - \epsilon, \alpha)$ supports Π for some $\epsilon > 0$. But since $\alpha > 0$, $\mathcal{P}(\mu, \nu - \epsilon, \alpha)$ is stronger at $(\bar{x}, \bar{w})$ than $\mathcal{P}(\mu, \nu, \alpha)$, a contradiction. ∎

Proposition 4.8 *Any LFO inequality is a paraboloid inequality.*

Proof. Inequality (54) is the same as $\mathcal{P}(\mu, \nu, \alpha)$ where $\alpha = \hat{\alpha}$, $\mu = \bar{x} - \alpha A \bar{x}$ and $\nu = -\bar{x}^T A \bar{x} + \alpha \|A\bar{x}\|^2$.
∎

We can now prove the main result in this section. We first remind the reader that the linearization inequality for the function $\|x\|^2 + h^T x$ at any point $(\bar{x}, \bar{w})$ is given by

$$q \;\geq\; (2\bar{x} + h)^T x - \|\bar{x}\|^2.$$

Theorem 4.9 *Let $(\bar{x}, \bar{w}, \bar{q})$ be such that $(\bar{x}, \bar{w}) \in \mathrm{int}(P)$ and $(\bar{x}, \bar{w}, \bar{q})$ satisfies the linearization inequality at $(\bar{x}, \bar{w})$. If $(\bar{x}, \bar{w}, \bar{q})$ violates a paraboloid inequality, then a paraboloid inequality that is maximally violated by $(\bar{x}, \bar{w}, \bar{q})$ is an LFO inequality. Conversely, any LFO inequality violated at $(\bar{x}, \bar{w}, \bar{q})$ is a paraboloid inequality.*

Proof. Suppose that $(\bar{x}, \bar{w}, \bar{q})$ violates some paraboloid inequality $\mathcal{P}(\mu, \nu, \alpha)$. By Corollary 4.6 without loss of generality $\mathcal{P}(\mu, \nu, \alpha)$ supports Π. Thus by Theorem 2.7 there is an LFO inequality whose violation at $(\bar{x}, \bar{w}, \bar{q})$ is at least as large as that of $\mathcal{P}(\mu, \nu, \alpha)$. But then by Proposition 4.8 that LFO inequality and $\mathcal{P}(\mu, \nu, \alpha)$ are one and the same inequality. The converse is similar and will be omitted. ■

The principal consequence of Theorem 4.9 is that separation over the LFO inequalities is equivalent to separation over the paraboloid inequalities. In the next section we show how to do this in polynomial time.

4.0.3 Polynomial-time separation of paraboloid inequalities

In this section we provide a polynomial-time algorithm that, given a point $(\bar{x}, \bar{w}) \in \mathbb{R}^n \times \mathbb{R}$ with $\bar{x}^T A \bar{x} < \bar{w}$ computes a paraboloid inequality $\mathcal{P}(\mu^*, \nu^*, \alpha^*)$ that is strongest at $(\bar{x}, \bar{w})$. By Proposition 4.4 such an inequality exists. We will refer to this task as *the paraboloid separation problem.*

In what follows we assume that the pair $(\bar{x}, \bar{w})$ is given. Recall that inequality $\mathcal{P}(\mu, \nu, \alpha)$ requires $q \geq (2\mu + h)^T x + \alpha w - \|\mu\|^2 - \alpha\nu$. Therefore, the paraboloid separation problem at $(\bar{x}, \bar{w})$ can be stated as:

$$\begin{aligned}
\min \quad & -(2\mu + h)^T \bar{x} - \alpha\bar{w} + \alpha\nu + \|\mu\|^2 && (60)\\
\text{s.t.} \quad & \Gamma^{\mu,\nu,\alpha} \quad \subseteq \quad \left\{(x, w) \in \mathbb{R}^n \times \mathbb{R} : x^T A x - w \leq 0\right\} && (61)\\
& \mu \in \mathbb{R}^n,\ \nu \in \mathbb{R}_+,\ \alpha \geq 0. && (62)
\end{aligned}$$

We will show below (Theorem 4.13) that an optimal solution to this problem is:

$$\mu^* = (I - \lambda_{max}^{-1} A)\bar{x}, \quad \nu^* = -\lambda_{max}\|\mu\|^2 \;+\; \bar{x}^T(\lambda_{max} I - A)\bar{x}, \quad \alpha^* = \lambda_{max}^{-1}.$$

To obtain this result we will need a number of technical steps.

Lemma 4.10 *Let $(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ be an optimal solution to:*

$$\begin{aligned}
\min \quad & -(2\mu + h)^T \bar{x} - \alpha\bar{w} + \alpha\nu + \|\mu\|^2 && (63)\\
s.t. \quad & \alpha\nu + \|\mu\|^2 + \min_x \left\{x^T(I - \alpha A)x - 2\mu^T x\right\} \geq 0 && (64)\\
& \mu \in \mathbb{R}^n,\ \nu \in \mathbb{R}_+,\ \alpha \geq 0. && (65)
\end{aligned}$$

Suppose $\hat{\alpha} > 0$. Then $(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ is an optimal solution to problem (60)-(62).

Proof. First note that the two objective functions are identical. We claim that for $\alpha > 0$ constraints (61) and (64) are identical. This holds because (61) holds if and only if for all $x \in \mathbb{R}^n$

$$\|x - \mu\|^2 + \alpha\nu \;\geq\; \alpha x^T A x,$$

which is (64). Thus $(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ is feasible for problem (60)-(62), and by the preceding remarks, if it is not optimal for (60)-(62) then it is improved upon by a solution of the form $(\mu, \nu, 0)$. But any such vector is also feasible for (63)-(65). Since we assumed $(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ was optimal for (63)-(65) the proof is complete. ■

It seems intuitively clear that at optimality constraint (64) will be binding. A formal proof is as follows:

Proposition 4.11 *Without loss of generality, at optimality for problem (63)-(65), constraint (64) will hold as an equality.*

Proof. Let $(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ be optimal for (63)-(65). If $\hat{\alpha} = 0$ the result is clear (minimum in (64) attained at $x = \hat{\mu}$). And if $\hat{\alpha} > 0$, by Lemma 4.10, $\mathcal{P}(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ is a strongest paraboloid inequality at $(\bar{x}, \bar{w})$. However, if (64) is not binding at $(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ then by Remark 4.2, $\mathcal{P}(\hat{\mu}, \hat{\nu}, \hat{\alpha})$ is not supporting. This is a contradiction by Lemma 4.6. ■

This result allows us to further simplify the paraboloid separation problem.

Proposition 4.12 *Problem (63)-(65) can be equivalently restated as:*

$$\min \quad -2\bar{x}^T(I - \alpha A)\pi \ + \ \pi^T(I - \alpha A)\pi \ - \ \alpha\bar{w} \ - \ h^T\bar{x} \tag{66}$$

$$s.t. \quad \pi \in \mathbb{R}^n,\ \alpha \geq 0. \tag{67}$$

Further, if π^*, α^* are optimal to this problem, then (μ^*, ν^*, α^*) are optimal for problem (63)-(65), where

$$\begin{aligned} \mu^* &= (I - \alpha^* A)\pi^*, \\ \alpha^*\nu^* &= -\|\mu\|^2 \ + \ \pi^{*T}(I - \alpha^* A)\pi^*. \end{aligned}$$

Proof. By Proposition 4.11 the paraboloid separation problem is equivalent to:

$$\begin{aligned} \min \quad & -(2\mu + h)^T\bar{x} - \alpha\bar{w} \ - \ \min_x \left\{ x^T(I - \alpha A)x - 2\mu^T x \right\} \\ \text{s.t.} \quad & \mu \in \mathbb{R}^n. \end{aligned}$$

Since at optimality the inner minimum must be finite, we can now apply Lemma 3.4, with $\tilde{A} = I - \alpha A$ and $\tilde{\mu} = \mu$. Lemma 3.4 implies the already noticed fact that $\alpha \leq \lambda_{max}^{-1}$, and it implies that

$$\min_x \left\{ x^T(I - \alpha A)x - 2\mu^T x \right\} = -\pi^T(I - \alpha A)\pi,$$

for some π with $(I - \alpha A)\pi \ = \ \mu$. Furthermore, we can equivalently rewrite the paraboloid separation problem as

$$\begin{aligned} \min \quad & -(2\mu + h)^T\bar{x} - \alpha\bar{w} \ + \ \pi^T(I - \alpha A)\pi \\ \text{s.t.} \quad & (I - \alpha A)\pi \ = \ \mu \\ & \pi \in \mathbb{R}^n,\ \mu \in \mathbb{R}^n,\ \alpha \geq 0, \end{aligned}$$

which together with Proposition 4.11 completes the proof. ■

Theorem 4.13 *An optimal solution to the paraboloid separation problem is*

$$\mu^* = (I - \lambda_{max}^{-1} A)\bar{x}, \quad \nu^* = -\lambda_{max}\|\mu\|^2 \ + \ \bar{x}^T(\lambda_{max} I - A)\bar{x}, \quad \alpha^* = \lambda_{max}^{-1}.$$

Proof. Applying Lemma 3.5, we get that if (π^*, α^*) is an optimal solution to problem (66)-(67), we have $\pi^* = \bar{x}$. Thus the optimal objective value of (66)-(67) is

$$-\bar{x}^T(I - \alpha^* A)\bar{x} - \alpha^*\bar{w} - h^T\bar{x}.$$

Since $\bar{x}^T A\bar{x} - \bar{w} < 0$, it follows that $\alpha^* = \lambda_{max}^{-1} > 0$ is the optimal choice for α, and using Propositions 4.10 and 4.12 we obtain the desired result. ■

5 Numerical experiments

In this section we present initial numerical experiments involving LFO cuts. Our implementations are straightforward and in particular do not include any cut management strategies. Nevertheless the experiments are promising and highlight interesting behavior of two problem classes. In Section 5.1 we consider cardinality-constrained convex quadratic programs; in Section 5.2 we compare LFO cuts with the disjunctive approach used in [24] in the context used Section 3.1, i.e. $Q(x) = \|x\|^2$ and P is a polyhedron.

5.1 Cardinality-constrained convex quadratic programs

In this section we present preliminary experiments involving problems of the form

$$\min \{ \, M(x) \; : \; x \in \Delta, \; \|x\|_0 \leq K \, \} \tag{68}$$

where $M(x)$ is a convex quadratic, $\Delta = \{x \in \mathbb{R}^d_+ \, : \, \sum_j x_j = 1\}$ is the unit simplex, $0 \leq K \leq d$, and for $x \in \mathbb{R}^d$, $\|x\|_0$ is the number of nonzero entries in x. This problem class has been studied before, see e.g. [17], [18], [32] and arises in several applications. When M is positive-definite, d is large and K much smaller than d, problems of this type can be quite difficult. This is in particular the case if the solution to the relaxation to (68) obtained by ignoring the cardinality constraint is contained in the (relative) interior of Δ. The goal of our experiments is to study the effect of using LFO cuts to obtain lower bounds on the value of problem (68).

Problem (68) can be formulated as a nonlinear mixed-integer program:

$$\begin{aligned}
\min \quad & M(x) && (69)\\
\text{s.t.} \quad & \sum_j x_j = 1 \\
& x_j - y_j \leq 1 \quad \text{and} \quad y_j \in \{0,1\}, \qquad 1 \leq j \leq d \\
& \sum_j y_j \leq K, \quad x \geq 0. && (70)
\end{aligned}$$

However this formulation can prove weak in difficult cases. A much stronger relaxation, the *perspective* relaxation, was used in [32]; it is also related to the disjunctive method in [24] (also see [29]). However, the perspective relaxation can also prove computationally expensive; see [18].

In our implementation of the LFO cuts we rely on the following result proved in [18] in a more general context. Here and below we denote $\mathcal{F} \doteq \{x \in \Delta \, : \, \|x\|_0 \leq K \, \}$, and for $\omega \in \mathbb{R}^d$ we write

$$\rho(\omega) \;\doteq\; \frac{(1 - \sum_{j \notin X} \omega_j)^2}{K} \;+\; \sum_{j \in X} \omega_j^2,$$

where $X \subseteq \{1, \ldots, d\}$ is the set of indices of the $d - K$ smallest values ω_j.

Lemma 5.1 *Suppose $w \in \mathbb{R}^d$ satisfies $\sum_{j=1}^d \omega_j = 1$. Then* $\min\{\|y - \omega\|^2 \, : \, y \in \mathcal{F}\} = \rho(\omega)$.

Using this result we can derive LFO cuts for problem (68). Given any point $\omega \in \Delta$, Lemma 5.1 guarantees that $\mathcal{F} \subseteq \mathbb{R}^d - \text{int}(\mathcal{B}(\omega, \sqrt{\rho(\omega)}))$ and the results in Section 3.2 can be used to generate cuts. We have implemented these ideas in the following straightforward cutting-plane procedure.

0. We initialize our formulation as $\min\{q \, : \, q \geq M(x), \; x \in \Delta\}$.

1. Solve the current formulation, with solution $(\bar{x}, \bar{q})$. Let $P^{\bar{x}} \doteq \mathcal{B}\left(\bar{x}, \sqrt{\rho(\bar{x})}\right)$.

2. Compute an LFO cut that separates $(\bar{x}, \bar{q})$ from the set

$$\text{conv}\left(\; \{ \, (x, q) \in \mathbb{R}^d \times R \, : \, q \geq M(x), \; x \in \mathbb{R}^d - \text{int}(P^{\bar{x}}) \, \} \; \right).$$

3. If no such cut is found, or if the violation of this cut by $(\bar{x}, \bar{q})$ is smaller than a tolerance $\epsilon > 0$, or if the number of iterations exceeds a limit T, exit. Otherwise add the cut to the formulation and return to step 1.

The cut in step 2 is obtained precisely as in Section 3.2, i.e. it is a strongest LFO cut at $\bar{x}$. In this implementation only one cut is obtained from each given set $P^{\bar{x}}$. The cutting procedure can be improved (see the discussion following the numerical results).

In the results reported below, we compare the strength of the lower bound obtained by our cutting-plane algorithm to two alternatives. The first one is the bound obtained by running the mixed-integer formulation (69)-(70) using a commercial solver, with a very long time limit. The second one is the bound obtained by an application of the S-Lemma to the cardinality-constrained problem (68). This second approach was used in [18] and it can be summarized as follows: let

$$x^* \ = \text{argmin}\{M(x) \,:\, e^T x = 1\},$$

where $e = (1, \ldots, 1)^T$. Then, as argued above

$$\min\{M(x) \,:\, x \notin \text{int}(\mathcal{B}(x^*, \sqrt{\rho(x^*)}))\} \tag{71}$$

provides a valid lower bound to the value of problem (68). Moreover, by optimality of x^* we have

$$e^T \nabla M(x^*) = 0.$$

Consequently, by convexity, for any $y \in \mathbb{R}^d$

$$M(x^* + y) \ \geq \ \lambda_{min} \|y\|^2,$$

where λ_{min} is the minimum eigenvalue of the Hessian of $M(x)$. Together with (71) we obtain that

$$L^* \doteq \ M(x^*) \, + \, \lambda_{min} \rho(x^*) \tag{72}$$

is a lower bound to the value of problem (68). This approach can be viewed as an application of the S-Lemma. As shown in [18], L^* improves not only on what the mixed-integer programming formulation yields in practicable time, but, usually, on the value of the perspective relaxation as well.

In the experiments below, our cutting-plane algorithm using LFO inequalities was run with tolerance $\epsilon = 10^{-3}$ and iteration limit $T = 10$. The problem instances considered in Tables 1 and 2 were generated as follows. In each case, the quadratic $M(x)$ is positive-definite and separable. Thus without loss of generality we can write $M(x) = (x - x^0)^T \Lambda (x - x^0)$ where $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$ and the λ_i are positive, and $x^0 \in \mathbb{R}^d$. Each λ_i was chosen randomly, by drawing from the uniform distribution on the interval $[1, 1 + \theta]$, where $\theta > 0$ is a fixed parameter. Note that $x^0 = \text{argmin}\{M(x) \,:\, x \in \mathbb{R}^d\}$. In our experiments we used $x^0 = 0$ and $x^0 = d^{-1} e$.

In the tables below, the columns headed "LFO-L" and "LFO-t" describe the lower bound on problem (68) and running time produced the cutting-plane algorithm, respectively. The mixed-integer programming formulation (69)-(70) was run until either a limit of 1000 CPU seconds was reached or one million branch-and-cut nodes were enumerated (whichever came first); columns headed "MIP-L", "MIP-t" and "MIP nodes" indicate, respectively, the resulting lower bound, (wall-clock) running time on sixteen threads, and number of branch-and-cut nodes. "MIP-U" provides the *upper* bound on problem (68) obtained by the mixed-integer programming approach. Finally, the column headed "S-L" is the lower bound provided by the S-Lemma approach as in (72). On the problem instances with $d > 100$ the objective functions were scaled up by a factor of 1000.

All computations (here and in the next section) were performed on an 8-core i7 computer, with 48 GB of physical memory. The cutting-plane algorithm used Gurobi 5.50 [33] for step 1. To run the mixed-integer programs we used both Gurobi 5.50 and CPLEX 12.2 [26] (and report the better of the two).

Table 1: ***Cardinality-constrained problems with $x^0 = d^{-1}e$***

d	K	θ	LFO-L	S-L	MIP-L	MIP-U	LFO-t (sec)	MIP-t (sec)	MIP nodes
100	20	2.00	0.0411	0.0412	0.0005	0.0587	0.127	227	1011704
100	50	5.00	0.0108	0.0108	0.0006	0.0314	0.102	222	1004975
100	20	10.00	0.0465	0.0465	0.0009	0.1284	0.120	288	1008679
1000	100	10.00	9.1009	9.1010	0.0010	18.2534	0.883	1012	246063
1000	100	100.00	10.0109	10.0125	0.0048	87.8492	0.848	1004	208633
1000	70	20.00	13.5842	13.5844	0.0011	32.0741	0.879	1000	176152
2000	100	40.00	9.5178	9.5178	0.0003	26.8787	3.014	1086	34699
2000	90	50.00	10.6348	10.6358	0.0003	32.2729	2.563	1019	14298
2000	80	50.00	12.0266	12.0280	0.0003	33.8795	3.186	1015	152638

Table 1 reports on results using $x^0 = d^{-1}e$. Problems of this type are especially hard for the mixed-integer programming formulation, which is unsuccessful at moving the lower bound significantly away from zero. This also holds for the smaller problem instances, even though over a million branch-and-cut nodes are enumerated. The LFO-based approach quickly (within ten iterations) attains a bound that is essentially equal to that provided by the S-Lemma, and several orders of magnitude larger than the mixed-integer lower bound, and thereafter tails off, sharply. The bound proved by the LFO-based approach, in many cases, does not completely close the gap relative to the best upper bound obtained by the mixed-integer programming formulation in the provided time/node limit.

To further explore this behavior consider Table 2 which displays results in cases where $x^0 = 0$. When this is the case one can prove that the optimal value of problem (68) is obtained as follows: where $I \subseteq \{1, \ldots, d\}$ is the set of indices corresponding to the K smallest λ_i, the optimal value of problem (68) equals $\left(\sum_{j \in I} \lambda_j^{-1}\right)^{-1}$. Table 2 displays this value in the column headed "OPT". We can see that the mixed-integer solver obtains this value as an upper bound (but does not prove so) nearly all the time (in one case, highlighted with an asterisk, round-off error by the solver resulted in a better-than-optimum upper bound). In this case, again, we see that the lower bound obtained by using the mixed-integer formulation is nearly always greatly improved by the LFO-driven lower bound (which as before essentially ties the S-Lemma lower bound). Generally, the LFO lower bound reduces the duality gap by at least 50%.

Table 2: ***Cardinality-constrained problems with $x^0 = 0$***

d	K	θ	LFO-L	S-L	MIP-L	MIP-U	OPT	LFO-t (sec)	MIP-t (sec)	MIP nodes
100	50	0.50	0.0204	0.0204	0.0127	0.0224	0.0224	0.11	260	1016899
100	20	0.50	0.0489	0.0490	0.0129	0.0524	0.0491	0.10	258	1009229
100	10	0.50	0.0984	0.0984	0.0103	0.1026	0.1026	0.09	250	1011881
100	50	1.00	0.0213	0.0213	0.0157	0.0246	0.0214	0.09	309	1022007
100	20	1.00	0.0484	0.0485	0.0151	0.0548	0.0487	0.09	293	1013290
100	10	1.00	0.0972	0.0972	0.0158	0.1053	0.1053	0.12	242	1004596
100	50	4.00	0.0288	0.0288	0.0254	0.0362	0.0362	0.13	330	1001651
1000	100	10.00	10.0295	10.0322	4.2512	14.7429	14.7429	1.01	1013	171452
1000	90	10.00	10.9764	10.9763	4.2736	15.8986	15.8986	0.86	1015	170395
1000	80	10.00	12.1727	12.1760	4.2541	17.3035	17.3012	0.979	1005	169706
1000	70	20.00	14.7284	14.7285	6.7351	23.2463*	23.2464	0.638	1005	170092
2000	100	40.00	11.0015	11.0038	5.5414	18.9675	18.9675	2.738	1011	111417
2000	90	50.00	12.5227	12.5229	6.5581	22.1117	22.1117	2.168	1025	103016
2000	80	50.00	13.5913	13.5914	6.5594	23.6386	23.6386	2.267	1056	107836

The above cutting-plane scheme would likely be improved in a number of ways. Principal among these is the concept of *sampling* the infeasible region so as to generate strong cutting-planes in advance of the formal algorithm, possibly used in preprocessing form. A number of such sampling techniques (related to the so-called Sandwich algorithm) are described in [8]. The dampened method in Step 3 of Example 2.13 can also be viewed as an example of this idea. Preprocessing by sampling so as to generate good cuts in advance of the formal algorithm is usually a very effective idea, especially in the context of first-order algorithms used to approximate a very nonlinear function.
In forthcoming work we plan to address the following more substantial enhancements to the preliminary work described here:

(i) The approach given by steps **0 - 4** above relies on Lemma 5.1 to exclude a ball $\mathcal{B}$ from the feasible region, with the current iterate $\bar{x}$ at its center. Typically, the optimizer z of the quadratic $M(x)$ on the boundary of this ball will be such that $z - \bar{x}$ is parallel to an eigenvector corresponding to the minimum eigenvalue for $M(x)$. However, the ball $\mathcal{B}$, computed as per Lemma 5.1, has as radius the minimum distance from $\bar{x}$ to the feasible region – and this minimum distance is attained by some point $y \in \partial\mathcal{B}$ such that $y - \bar{x}$ is, often, *far from* parallel with any eigenvector arising from a small eigenvalue for $M(x)$. In fact $y - \bar{x}$ may more likely be aligned with eigenvectors corresponding to much larger eigenvalues. This phenomenon suggests the use a polyhedral, rather than a spherical, relaxation. Rather than excluding a ball from the feasible region, we would exclude a polyhedron which "pushes" into the corners of the unit simplex. Ideally, the excluded polyhedron would have larger diameter along the eigenspace corresponding to the larger eigenvalues.

(ii) Couple the cutting-plane scheme with a branching approach, parameterized by a value $\theta > 0$, as follows. Let v be an eigenvector of the quadratic part of $M(x)$ corresponding to the maximum eigenvalue. If $\hat{x}$ is the solution to the current relaxation, we would branch by considering those points x such that $v^T(x - \hat{x}) > \theta$, those such that $v^T(x - \hat{x}) < -\theta$, and those such that $-\theta \leq v^T(x - \hat{x}) \leq \theta$ – and on the latter set use a relaxation that enforces a minimum distance to the feasible region which would presumably be larger than the minimum distance from $\hat{x}$ to the feasible region.

(iii) Reformulate the problem so as to use a coordinate system contained in the $(d-1)$-dimensional hyperplane $\{x \in \mathbb{R}^d \, : \, e^T x = 0\}$. As shown in [18], an appropriate representation of the quadratic $M(x)$ valid in this restricted space (the so-called projected quadratic) results in more effective bounds, because the minimum eigenvalue usually increases. This point is related to (i), above.

5.2 A comparison with the disjunctive method

In this section we consider the setup in section 3.1. We are given a polyhedron $P = \{x \in \mathbb{R}^d \, : \, Ax \leq b\}$ containing the origin in its interior and we are interested in the set $S = \{(x, q) \in \mathbb{R}^d \times \mathbb{R} \, : \, q \geq \|x\|^2, \ x \in \mathbb{R}^d - \text{int}(P)\}$.

The purpose of the experiments in this section is to compare the performance of the separation algorithm for LFO inequalities given in Section 3.1 to the performance of the disjunctive method, both as generators of cutting-plane used to separate from $\text{conv}(S)$. Here we remind the reader that a set S of the form considered here would not arise as "the" problem being solved. Rather, it would be a *relaxation* of a problem of interest (as was the case with the cardinality-constrained problem considered above). Thus, our algorithm for separating LFO inequalities on the one hand, and the disjunctive formulation (plus SOCP duality) on the other hand, would constitute competing methods for separating from $\text{conv}(S)$, and here we investigate their performance in this context.

In the case of either separation routine we run a cutting-plane algorithm similar to the one given in the Section above:

0. We initialize our formulation as $\min\{q \, : \, q \geq M(x)\}$.
1. Solve the current formulation, with solution $(\bar{x}, \bar{q})$.

2. Compute a cut that separates $(\bar{x}, \bar{q})$ from the set

$$\text{conv}\left(\{ (x, q) \in \mathbb{R}^d \times R \, : \, q \geq M(x), \; x \in \mathbb{R}^d - \text{int}(P) \} \right).$$

3. If no such cut is found, or if the violation of this cut by $(\bar{x}, \bar{q})$ is smaller than a tolerance $\epsilon > 0$, or if the number of iterations exceeds a limit T, exit. Otherwise add the cut to the formulation and return to step 1.

This algorithm solves the problem $N_2 \doteq \min\{\|x\|^2 \, : \, x \notin \text{int}(P)\}$. Of course the value of this problem is known, however here we are concerned with the number of iterations needed by the two algorithms and with any other interesting performance attributes that may arise.

In our experiments, the systems $Ax \leq b$ were generated as follows. The entries of the coefficient 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 normalized to have unit norm and the entries were rounded to 3 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, we set b_i 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 $0.5\sqrt{n}$. In either case b_i was then rounded to 3 digits.

In Table 3, the columns labeled d, m, and val give the dimension, number of rows in A, and true problem value N_2. Lo_l, $Time_l$, and $Cuts_l$ give the best lower bound, time taken (in seconds), and number of cuts generated by the lifting method, with similar information for the disjunctive method given by Lo_d, $Time_d$, and $Cuts_d$, respectively. Each method, using the template provided by steps **0-4** above, was allowed a total of 500 cuts, and was only allowed to generate a new cut if fewer than 600 seconds had passed since the initial setup. An asterisk next to a time in the $time_d$ column indicates that that instance was stopped because the solver was unable to find the dual variables needed to generate a cut. These tests were terminated when the relative gap between the lower bound and the true value was less than a tolerance of 10^{-5}.

Table 3: ***Comparison with disjunctive method***

d	**m**	$\boldsymbol{N_2}$	$\boldsymbol{Lo_l}$	$\boldsymbol{Lo_d}$	$\boldsymbol{Time_l}$	$\boldsymbol{Time_d}$	$\boldsymbol{Cuts_l}$	$\boldsymbol{Cuts_d}$
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	11
200	400	2207.060	2207.038	0.000	8.7	91.0*	92	1
300	500	4583.748	4583.733	0.000	2.4	155.3*	20	1
800	1200	38142.592	38142.243	0.000	32.1	1879.1*	42	1
1000	2000	61726.150	61725.542	0.000	227.8	3304.3*	134	1

We can see from this table that the disjunctive method tends to fail as problem size becomes large. Part of the reason is that the SOCPs to be solved simply prove too difficult. Note that the disjunctive method already seems to encounter numerical difficulties on medium size problems; even on the smallest instances, the lower bounds on N_2 obtained by the disjunctive method can be poor.

Table 4 presents results on various variants of the cutting-plane algorithm. In addition to the LFO cuts, the "Basic" version does not use the constraint $q \geq \|x\|^2$ and instead uses linearization cuts to approximate this constraint. Its initial formulation is $\min\{q \, : \, q \geq 0\}$. We also consider enhancements to the Basic version, using three heuristics to help make faster progress:

1. Before starting, the linearization cut was added at each unit vector e_i as well as $-e_i$.
2. Before starting, the LFO inequality at the point closest to the origin on each facet was added, if possible.
3. The constraint $Ax \leq b$ was added in the relaxation.

The first two heuristics are versions of "sampling" as described at the end of the last section. The "Full" method includes the conic constraint $q \geq \|x\|^2$ and does not use linearization cuts. 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 Table 4, columns labeled d, m, and N_2 are as in the previous table. In the next three sets of columns ("Basic", "Heuristics", and "Full"), q_{lo} gives the best proven lower bound, lin gives the number of linearization cuts added, $lifted$ gives the number of lifted cuts added, and t is the time spent, in seconds. In this test, each method was limited to 30 minutes to add cuts. During each iteration, the relaxation was solved and both the linearization cut and the lifted cut were added, if possible. A maximum of $10,000$ iterations were reached.

We can see from Table 4 that on all instances but the last one, the "Full" version is the clear winner. The use of the conic constraint helps guide the algorithm toward points where cutting (using LFO inequalities) is most effective. At the same time, having a single conic constraint helps control computational cost and numerical instabilities.

Table 4: ***Comparison of cutting-plane strategies***

			Basic				Heuristics				Full		
d	m	N_2	q_{lo}	lin	$lifted$	t	q_{lo}	lin	$lifted$	t	q_{lo}	$lifted$	t
10	50	5.19	5.19	286	120	0.8	5.19	280	212	1.1	5.19	133	1.1
20	100	14.54	14.54	2104	824	19.5	14.54	890	744	8.6	14.54	33	0.3
20	300	17.83	17.83	1903	577	28.7	17.83	2831	13	34.0	17.83	26	0.6
50	200	79.89	0.00	10000	1	688.0	79.89	10300	38	646.2	79.89	28	0.6
75	250	343.90	0.00	10000	1	768.6	343.87	8198	7841	1803.9	339.46	1370	1802.3
100	300	324.49	0.00	10000	1	1127.2	320.57	6780	6361	1804.1	324.49	25	1.1
200	400	2207.06	0.00	269	1	1807.1	2207.06	6168	151	1803.3	2207.03	87	9.6
300	500	4583.75	0.00	301	1	13.3	597.60	2938	2003	1803.3	4583.74	24	3.7
800	1200	38142.59	0.00	302	1	66.7	569.48	2940	651	1811.4	38142.54	55	53.9
1000	2000	61726.15	0.00	246	1	108.7	61726.15	4001	893	2330.6	57689.25	18	40.0

Acknowledgments. We would like to thank Samuel Burer, Adam Letchford and Gábor Pataki for helpful suggestions.

References

[1] F. Alizadeh and D. Goldfarb, Second-order cone programming, *Mathematical Programming* **95** (2001), 3 – 51.

[2] F. Al-Khayyal and J. Falk, Jointly constrained biconvex programming, *Mathematics of Operations Research* **8** (1983), 273-286.

[3] K. Andersen, Q. Louveaux, R. Weismantel and L. Wolsey, Cutting planes from two rows of a simplex tableau, in : *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science* **4513** (2007), pp 1 - 15.

[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 and S. Burer, Computable representations for convex hulls of low-dimensional quadratic forms, *Mathematical Programming B* **124** (2010), 33–43.

[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] X. Bao, N.V. Sahinidis, and M. Tawarmalani, Multiterm polyhedral relaxations for nonconvex, quadratically constrained quadratic programs, *Optimization Methods and Software* **24** (2009), 485–504.

[9] E. Balas, Intersection cuts - a new type of cutting planes for integer programming, *Operations Research* **19** (1971), 19 - 39.

[10] E. Balas, Disjunctive programs: cutting planes from logical conditions, in O.L. Mangasarian et al., eds., *Nonlinear Programming* **2** (1975), Academic Press, 279 – 312.

[11] E. Balas, Disjunctive programming. *Annals of Discrete Mathematics* **5** (1979), 3 – 51.

[12] 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.

[13] 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.

[14] 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.

[15] P. Belotti, J. C. Goez, I. Pólik, T. K. Ralphs and T. Terlaky, On Families of Quadratic Surfaces Having Fixed Intersection with Two Hyperplanes, *Discrete Applied Mathematics* **161** (2013), 2778 –2793.

[16] 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.

[17] D. Bienstock, Computational study of a family of mixed-integer quadratic programming problems, *Mathematical Programming* **74** (1996), 121– 140.

[18] D. Bienstock, Eigenvalue techniques for proving bounds for convex objective, nonconvex programs, in: *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science* **6080** (2010), pp 29–42.

[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 programming, *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science* **6080** (2010), 285 – 298.

[21] S. Burer and A. N. Letchford, On non-convex quadratic programming with box constraints, *SIAM Journal on Optimization* **20** (2009), 1073–1089.

[22] S. Burer and A. N. Letchford, Non-convex mixed-integer nonlinear programming: a survey, *Optimization Online*, February 2012.

[23] S. Cafieri, J. Lee, and L. Liberti, On convex relaxations of quadrilinear terms, *Journal of Global Optimization* **47** (2010), 661-685.

[24] S. Ceria and J. Soares, Convex programming for disjunctive convex optimization, *Mathematical Programming* **86** (1999), 595 - 614.

[25] G. Cornuéjols and F. Margot, On the facets of mixed integer programs with two integer variables and two constraints, *Mathematical Programming* **120** (2009), 429-456.

[26] IBM ILOG CPLEX Optimizer. http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

[27] I.R. de Farias JR., E.L. Johnson and G.L. Nemhauser, Facets of the complementarity knapsack polytope, *Mathematics of Operations Research* **27** (2002), 210 –226.

[28] D.G. Espinoza, R. Fukasawa and M. Goycoolea, Lifting, tilting and fractional programming revisited, *Operations Research Letters* **38** (2010), 559 – 563.

[29] A. Frangioni and C. Gentile, Perspective cuts for a class of convex 0-1 mixed integer programs, *Mathematical Programming* **106** (2006), 225 - 236.

[30] D. Goldfarb and G. Iyengar, Robust portfolio selection problems, *Mathematics of Operations Research* **28** (2002), 1 – 38.

[31] G.H. Golub, Some modified matrix eigenvalue problems, *SIAM Review* **15** (1973), 318 – 334.

[32] O. Günlük and J. T. Linderoth, Perspective reformulations of mixed integer nonlinear programs with indicator variables, *Mathematical Programming* **124** (2010), 183 –205.

[33] Gurobi Optimizer. http://www.gurobi.com/.

[34] A.B. Keha, I.R. de Farias JR. and G.L. Nemhauser, A branch-and-cut algorithm without binary variables for nonconvex piecewise linear optimization," *Operations Research* **54** (2006), 847 – 858.

[35] M. Kilinc, J. Linderoth and J. Luedtke, Effective separation of disjunctive cuts for convex mixed integer nonlinear programs, *Optimization Online* (2010).

[36] M. Kojima and L. Tuncel, Discretization and localization in successive convex relaxation methods for nonconvex quadratic optimization problems, *Mathematical Programming* **89** (2000), 79 – 111.

[37] L. Lovász and A. Schrijver, Cones of matrices and set-functions and 0-1 optimization, *SIAM Journal on Optimization* **1** (1991), 166-190.

[38] J. Luedtke, M. Namazifar and J. Linderoth, Some results on the strength of relaxations of multilinear functions, Optimization Online, August 2010.

[39] J. Makó, K. Nikodem and Z. Páles, On strong $(\alpha, \mathbb{F})$-convexity, *Mathematical Inequalities and Applications* **2** (2012), 289 – 299.

[40] G.P. McCormick, Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems. *Mathematical Programming* **10** (1976), 147 – 175.

[41] J.E. Mitchell, J.-S. Pang and B. Yu, Convex quadratic relaxations of nonconvex quadratically constrained quadratic programs, *Optimization Methods and Software*, (published online) 2012.

[42] A. Michalka, *Cutting Planes for Convex Objective Nonconvex Optimization*, Ph.D. Dissertation, Columbia University (2013).

[43] S. Modaresi, M. R. Kılınç, and J. P. Vielma, Intersection Cuts for Nonlinear Integer Programming: Convexification Techniques for Structured Sets, manuscript, (2013).

[44] G.L. Nemhauser and L.A. Wolsey, *Integer and Combinatorial Optimization*, Wiley, New York (1988).

[45] I. Pólik and T. Terlaky, A survey of the S-lemma, *SIAM Review* **49** (2007), 371 – 418.

[46] A. Qualizza, P. Belotti, F. Margot, Linear programming relaxations of quadratically constrained quadratic programs, manuscript, 2011.

[47] J.-P. Richard and M. Tawarmalani, Lifting inequalities: a framework for generating strong cuts for nonlinear programs, *Mathematical Programming* **121** (2010), 61 – 104.

[48] A.D. Rikun, A convex envelope formula for multilinear functions. *Journal of Global Optimization* **10** (1997), 425 - 437.

[49] N.V. Sahinidis, BARON: A general purpose global optimization software package, *Journal of Global Optimization* **8** (996), 201 - 205.

[50] A. Saxena, P. Bonami and J. Lee, Convex relaxations of non-convex mixed integer quadratically constrained programs: Extended formulations. *Mathematical Programming* **124** (2010), 383 – 411.

[51] A. Saxena, P. Bonami and J. Lee, Convex relaxations of non-convex mixed integer quadratically constrained programs: Projected formulations. To appear, *Mathematical Programming.*

[52] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*, Kluwer, Dordrecht (1998).

[53] S. Sherali and W. 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.

[54] S. Sherali and W. Adams, A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems, *SIAM Journal on Discrete Mathematics* **3** (1990), 411-430.

[55] R.A. Stubbs and S. Mehrotra, A branch-and-cut method for 0-1 mixed convex programming, *Mathematical Programming* **86** (1999), 515 - 532.

[56] M. Tawarmalani and N.V. Sahinidis, Convex extensions and envelopes of lower semi-continuous functions, *Mathematical Programming* **93** (2002), 247 – 263.

[57] M. Tawarmalani and N.V. Sahinidis, Global optimization of mixed-integer nonlinear programs: a theoretical and computational study, *Mathematical Programming* **99** (2004), 563 – 591.

[58] M. Tawarmalani and N.V. Sahinidis, A polyhedral branch-and-cut approach to global optimization, *Mathematical Programming* **103** (2005), 225 - 249.

[59] V. A. Yakubovich, S-procedure in nonlinear control theory, Vestnik Leningrad University, **1** (1971), 62–77.