A p-Cone Sequential Relaxation Procedure for 0-1 Integer Programs*

Samuel Burer† Jieqiu Chen‡

February 11, 2008
Revised: June 4, 2008

Abstract

Given a 0-1 integer programming problem, several authors have introduced sequential relaxation techniques — based on linear and/or semidefinite programming — that generate the convex hull of integer points in at most n steps. In this paper, we introduce a sequential relaxation technique, which is based on p-order cone programming ($1 \leq p \leq \infty$). We prove that our technique generates the convex hull of 0-1 solutions asymptotically. In addition, we show that our method generalizes and subsumes several existing methods. For example, when $p = \infty$, our method corresponds to the well-known procedure of Lovász and Schrijver based on linear programming (so that finite convergence is obtained by our method in special cases). Although the p-order cone programs in general sacrifice some strength compared to the analogous linear and semidefinite programs, we show that for $p = 2$ they enjoy a better theoretical iteration complexity. Computational considerations of our technique are also discussed.

Keywords: Global optimization, integer programming, second-order cone programming, cone programming, relaxation

1 Introduction

Consider solving the 0-1 integer program

$$\begin{aligned} \min \quad & c^T x \\ \text{s. t.} \quad & a_i^T x \leq b_i \qquad \forall \;\; i = 1, \ldots, m \\ & x \in \{0,1\}^n. \end{aligned} \tag{1}$$

*Department of Management Sciences, University of Iowa, Iowa City, IA, 52242-1994, USA. Both authors supported in part by NSF Grant CCF-0545514.

†Corresponding author. Email: samuel-burer@uiowa.edu.

‡Email: jieqiu-chen@uiowa.edu.

Beyond the basic linear programming (LP) relaxation P of the feasible set of (1), many authors have considered general techniques for achieving tighter relaxations (Gomory, 1963; Sherali and Adams, 1990; Lovász and Schrijver, 1991; Balas et al., 1993; Kojima and Tunçel, 2000; Lasserre, 2001; Parrilo, 2003; Bienstock and Zuckerberg, 2004). One recurring theme is to lift the feasible set of (1) into a higher dimensional space, construct a convex relaxation in that space, and project this relaxation back into the original space, thus obtaining a relaxation P^1, which is (hopefully) tighter than P. The choice of lifting and relaxation determines the strength of P^1. In all previous works, LP and semidefinite (SDP) relaxations have been used in the lifted space.

In Lovász and Schrijver (1991), Balas et al. (1993), and Kojima and Tunçel (2000), the idea of sequential relaxation is also introduced. Stated simply, this idea is to repeat the lift-and-project procedure on the tighter relaxation P^1, thus obtaining an even tighter relaxation P^2. Then, if P^k denotes the k-th relaxation obtained inductively, a fundamental question is whether $\{P^k\}$ converges to P^{01}, the convex hull of the feasible set of (1). In each study referenced above, the answer is positive; in fact, $P^k = P^{01}$ for all $k \geq n$. Computationally, one can typically optimize over P^k in polynomial time as long as k is constant with respect to n. (Kojima and Tunçel (2000) actually apply their techniques to a much broader class of problems than just 0-1 programs and show asymptotic convergence to the convex hull of solutions.)

Sherali and Adams (1990), Lasserre (2001), and Parrilo (2003) do not explicitly employ the idea of sequential relaxation. Rather, they lift to ever higher dimensions before projecting, a technique which is analogous to sequential relaxation. Here, too, the authors show that lifting to a finite dimension (dependent in some manner on n) achieves P^{01}. Similar to the work of Kojima and Tunçel (2000), these authors' techniques can be applied to more general problem classes, but one may have to lift "infinitely" to achieve the convex hull of solutions.

Although these lift-and-project procedures are very powerful theoretically, they present significant computational challenges, even after a single iteration. One must deal with more variables in the higher dimensional space as well as additional constraints introduced by the lifting. For example, after one iteration of the LP-based procedure of Balas et al. (1993), the resulting LP contains $2n$ variables and $2m + 1$ constraints (assuming that the constraints $a_i^T x \leq b_i$ already imply the bounds $0 \leq x_j \leq 1$). The procedure of Lovász and Schrijver (1991) is more extreme. After one iteration, their LP-based procedure has $O(n^2)$ variables and $O(nm)$ linear constraints, and their SDP-based procedure contains an additional semidefinite constraint on an order $n \times n$ matrix. In both of these particular cases, computational progress has been achieved by exploiting structure (Balas and Perregaard, 2003; Burer and Vandenbussche, 2006).

The use of LP and SDP relaxations in lift-and-project procedures comes about quite naturally and, of course, is convenient since LPs and SDPs are well understood both theoretically and algorithmically. In principle, however, it may be possible to use other types of relaxations, which may have their own theoretical or algorithmic benefits in the context of lift-and-project.

The purpose of this paper is to explore p-order cone programming (POCP) relaxations, which include in particular second-order cone programming (SOCP) relaxations when $p = 2$. Our interest in POCP arises from the fact that POCP is becoming a well understood tool in convex optimization (Xue and Ye, 2000; Andersen et al., 2002; Glineur and Terlaky, 2004; Krokhmal and Soberanis, 2008). Moreover, there are by now several high quality implementations for SOCP, and in fact POCP can be formulated exactly via SOCP (Ben-Tal and Nemirovski, 2001; Alizadeh and Goldfarb, 2003) (see also Krokhmal and Soberanis (2008)). Our hope is that, by introducing POCP relaxations, we might discover new lift-and-project procedures that have their own theoretical and computational advantages.

In this paper, we introduce an entire family of lift-and-project procedures parameterized by $p \in [1, \infty]$ and prove that each asymptotically yields P^{01}, the convex hull of 0-1 solutions (Theorem 4.0.2). A feature of this family of procedures is the ability to lift and project with respect to different subsets of variables at different iterations. Although we do not achieve finite convergence in general (so that our procedure is weaker than existing methods in this sense), we do observe theoretical advantages. In particular, we show that the theoretical iteration complexity of solving the POCP relaxations via interior-point methods is minimized when $p = 2$ at which the iteration complexity is an order of magnitude less than solving existing LP and SDP relaxations (Corollary 3.2.1). In addition, our family of procedures unifies existing approaches. For example, when $p = \infty$, we recover the LP-based procedure of Lovász and Schrijver (1991).

The paper is organized as follows. In Section 2, we describe the basic p-order cone lift-and-project procedure as well as an alternate derivation of it. (This is just one iteration of the entire sequential relaxation approach, which is described in Section 4.) We compare and contrast our procedure with three existing approaches: Lovász and Schrijver (1991), Kojima and Tunçel (2000), and Balas et al. (1993). In particular, we point out that our method includes the LP based lift-and-project procedure of Lovász and Schrijver (1991) and the relaxation of Balas et al. (1993) as special cases.

In Section 3, we study fundamental properties of the p-order cone procedure. In particular, we examine duality properties and two types of monotonicity. For example, one monotonicity property establishes that the strength of the procedure increases with p, so that the strength is maximized at $p = \infty$. We also study the (iteration) complexity of solving the resultant p-order cone relaxation via interior-point methods, where it is shown that the lowest iteration complexity is obtained for $p = 2$. Following these results is the main technical result of the paper (Theorem 3.4.1): the p-order cone procedure, when applied to a generic compact, convex set P, cuts off all fractional extreme points of P. Theorem 3.4.1 is motivated and proved in three layers. We first show the result holds when P is a polytope, which is the easiest case. Then we establish the result when P is a ball, which finally allows us to prove the theorem for general P.

Continuing in Section 4, we describe the sequential relaxation approach based on repetition of the p-order cone procedure and prove that it generates the convex hull of 0-1 solutions asymptotically (Theorem 4.0.2). An example is provided to show that, in general, the iterated procedure does not converge after n iterations.

In Section 5, we consider computational issues associated with the p-cone sequential relaxation procedure — particularly, with optimizing over the first-iteration relaxation. We compare the SOCP relaxation ($p = 2$, lowest theoretical complexity) to the LP relaxation of Lovász-Schrijver ($p = \infty$, tightest relaxation). As it turns out, even though the SOCP relaxation enjoys a much lower theoretical iteration complexity, the LPs solve more quickly in practice. We believe this is a strong testament to the speed of modern LP solvers.

Finally, in Section 6, we conclude with a few final remarks.

1.1 Notation and terminology

$\Re^n$ refers to n-dimensional Euclidean space, and $\Re^{n \times n}$ is the set of real, $n \times n$ matrices. We let $e_i \in \Re^n$ represent the i-th unit coordinate vector and $e \in \Re^n$ represent the vector of all ones. We denote by $[n]$ the set $\{1, 2, \ldots, n\}$. For $\mathcal{J} \subseteq [n]$ an index set, $x_{\mathcal{J}} \in \Re^{|\mathcal{J}|}$ is defined as the vector composed of entries of x that are indexed by $\mathcal{J}$. Similarly, given a matrix $A \in \Re^{n \times n}$, $A_{\mathcal{J}\cdot}$ represents the $|\mathcal{J}| \times n$ matrix composed of the rows of A indexed by $\mathcal{J}$. $\mathrm{Diag}(x)$ denotes the diagonal matrix with diagonal x, and $A \succeq 0$ means that the matrix A is symmetric positive semidefinite; its dimension will be clear from context. Finally, given a set S defined over variables (x, y), $\mathrm{Proj}_x(S)$ denotes the projection of S onto the coordinates in x.

For $p \geq 1$, the usual p-norm on $\Re^n$ is defined as

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

We also consider when $p = \infty$ and $\|x\|_\infty := \max_{i=1}^{n} |x_i|$. Associated with $p \in [1, \infty]$ is q such that $p^{-1} + q^{-1} = 1$. Both the p-norm and q-norm give rise to closed, convex cones in $\Re^{1+n}$:

$$\mathcal{K}_p := \{(x_0, x) \in \Re^{1+n} : x_0 \geq \|x\|_p\}$$
$$\mathcal{K}_q := \{(y_0, y) \in \Re^{1+n} : y_0 \geq \|y\|_q\}.$$

It is well known that $\mathcal{K}_q$ is dual to $\mathcal{K}_p$, i.e.,

$$\mathcal{K}_q := \{(y_0, y) \in \Re^{1+n} : y_0 x_0 + y^T x \geq 0 \ \ \forall\ (x_0, x) \in \mathcal{K}_p\},$$

which is written as $K_q = K_p^*$. Important special cases occur when $p = 2$ or $p = \infty$. When $p = 2$, $K_p = K_q$, i.e., K_p and K_q are self-dual. When $p = \infty$, $q = 1$, and both K_p and K_q are polyhedral cones.

2 Relaxation Procedure and Comparisons

In this section, we formally state our p-cone lift-and-project procedure and compare and contrast it with the methods of Lovász and Schrijver (1991), Kojima and Tunçel (2000), and Balas et al. (1993).

2.1 Relaxation procedure

For the purposes of generality (particularly with regards to Section 4), we consider a slightly different form of the feasible set of the integer program (1), the only difference being that the linear constraints are indexed by an arbitrary set $\mathcal{I}$ (possibly infinite):

$$F := \left\{x \in \{0,1\}^n : a_i^T x \leq b_i \ \ \forall\ i \in \mathcal{I}\right\}.$$

This semi-infinite representation for F, as opposed to a finite one, does not affect the theoretical exposition of the p-cone procedure (though it may pose computational issues). Associated with F is its basic convex relaxation

$$P := \{x \in \Re^n : a_i^T x \leq b_i\ \forall\ i \in \mathcal{I}\},$$

which we assume is contained in $[0,1]^n$ — for example, by including bounds on x via explicit constraints $a_i^T x \leq b_i$. So P is compact convex.

We wish to generate a compact convex relaxation of F, which is tighter than P. Unless stated otherwise, we assume throughout this section the fixed choice of

- $p \in [1, \infty]$,
- $\emptyset \neq \mathcal{J} \subseteq [n]$.

We will denote the proposed convex relaxation as $N_{(p,\mathcal{J})}(P)$, or more often simply as $N(P)$. Defining

$$P^{01} := \mathrm{Conv}(F),$$

our goal is to produce $N(P)$ such that $P^{01} \subseteq N(P) \subseteq P$.

Our first step is to lift F into a higher dimensional space. We will make use of the following simple (but key) geometric proposition.

Proposition 2.1.1. *Define $r := \sqrt[p]{|\mathcal{J}|}/2$ and $d := e/2 \in \Re^{|\mathcal{J}|}$. Then $x_{\mathcal{J}} \in \{0,1\}^{|\mathcal{J}|}$ implies $\|x_{\mathcal{J}} - d\|_p \leq r$.*

It is important to keep in mind that r depends on p and $|\mathcal{J}|$ and that d depends on $|\mathcal{J}|$. This proposition establishes the existence of a family of p-balls circumscribing the integer points $\{0,1\}^{|\mathcal{J}|}$. In fact, $p' \geq p$ implies that the p'-ball is contained in the p-ball (see Proposition 3.3.2), with $p = \infty$ corresponding to the convex hull $[0,1]^{|\mathcal{J}|}$ of the integer points.

Using Proposition 2.1.1, F can be rewritten redundantly as

$$F = \{x \in \Re^n : x = x^2,\ a_i^T x \leq b_i \ \ \forall\, i \in \mathcal{I},\ \|x_{\mathcal{J}} - d\|_p \leq r\}.$$

We note that

$$\left.\begin{array}{l} a_i^T x \leq b_i \\ \|x_{\mathcal{J}} - d\|_p \leq r \end{array}\right\} \quad \Longrightarrow \quad \left\|(b_i - a_i^T x)(x_{\mathcal{J}} - d)\right\|_p \leq r(b_i - a_i^T x), \tag{2}$$

which in turn implies

$$F = \left\{x \in \Re^n : x = x^2,\ \left\|b_i x_{\mathcal{J}} - x_{\mathcal{J}} x^T a_i - (b_i - a_i^T x)d\right\|_p \leq r(b_i - a_i^T x) \ \ \forall\, i \in \mathcal{I}\right\}$$

since $b_i - a_i^T x$ is kept nonnegative. Next, introducing an $n \times n$ matrix variable X satisfying $X = xx^T$ and defining

$$\hat{F} := \left\{(x, X) \in \Re^n \times \Re^{n \times n} : \begin{array}{c} X = xx^T \qquad \mathrm{diag}(X) = x \\ \left\|b_i x_{\mathcal{J}} - X_{\mathcal{J}\cdot} a_i - (b_i - a_i^T x)d\right\|_p \leq r(b_i - a_i^T x) \ \ \forall\, i \in \mathcal{I} \end{array}\right\}$$

we see that $F = \mathrm{Proj}_x(\hat{F})$, i.e., $\hat{F}$ is the lifted version of F. In addition, dropping the nonconvex constraint $X = xx^T$ from $\hat{F}$, we obtain a convex relaxation of $\hat{F}$:

$$\hat{P} := \left\{ (x, X) \in \Re^n \times \Re^{n \times n} : \begin{array}{c} \operatorname{diag}(X) = x \\ \left\| b_i x_{\mathcal{J}} - X_{\mathcal{J}\cdot} a_i - (b_i - a_i^T x) d \right\|_p \leq r(b_i - a_i^T x) \ \ \forall \ i \in \mathcal{I} \end{array} \right\}.$$

Finally, we define $N(P)$ as the projection of $\hat{P}$:

$$N(P) := \operatorname{Proj}_x(\hat{P})$$

The desired property of $N(P)$ is immediate.

Proposition 2.1.2. *$P^{01} \subseteq N(P) \subseteq P$.*

Proof. $P^{01} \subseteq N(P)$ by construction. Moreover, the definition of $\hat{P}$ implies that every $x \in N(P)$ satisfies $r(b_i - a_i^T x) \geq 0$ for all $i \in \mathcal{I}$. Since $r > 0$, this implies $x \in P$. So $N(P) \subseteq P$. □

Note that $\hat{P}$ is closed and convex because it is the intersection of closed, convex sets. Hence, $N(P)$ is closed and convex as well. Furthermore, because P is bounded, so is $N(P)$. Thus, $N(P)$ is compact convex.

Just like P, $N(P)$ has its own semi-infinite outer description, which can be the basis of lift-and-project applied to $N(P)$ itself. This will be the idea behind the iterated procedure of Section 4.

2.1.1 Example

Consider the example feasible set

$$F = \left\{ x \in \{0,1\}^2 : \begin{array}{c} -x_1 \leq 0 \\ -x_2 \leq 0 \\ x_1 + 2x_2 \leq 2.5 \\ 3x_1 + x_2 \leq 2.5 \end{array} \right\} = \{(0,0), (0,1)\} \tag{3}$$

so that

$$P^{01} = \{(0, x_2) \in \Re^2 : 0 \leq x_2 \leq 1\}.$$

Let $\mathcal{J} = \{1, 2\}$. In Figure 1, we illustrate the p-cone procedure described in the previous section by depictng the four sets

$$P \supseteq N_{(1,\mathcal{J})}(P) \supseteq N_{(2,\mathcal{J})}(P) \supseteq N_{(\infty,\mathcal{J})}(P)$$

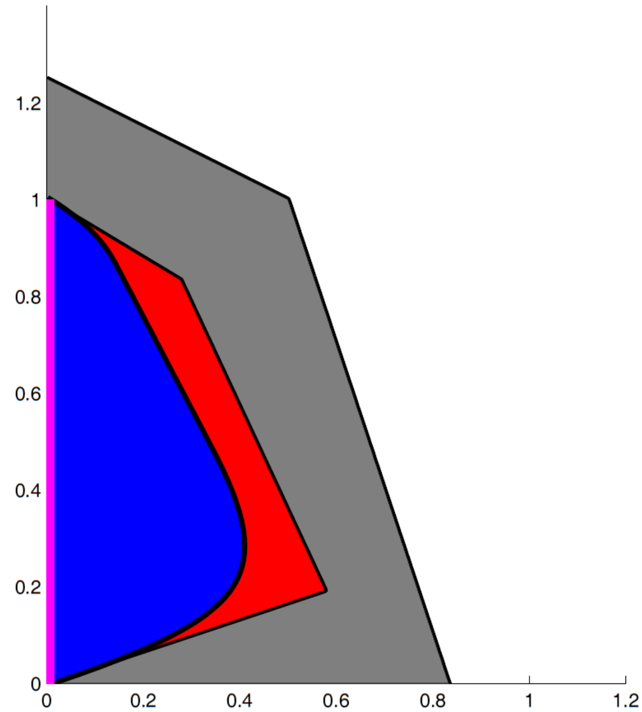

Figure 1: The four sets $P \supseteq N_{(1,\mathcal{J})}(P) \supseteq N_{(2,\mathcal{J})}(P) \supseteq N_{(\infty,\mathcal{J})}(P)$ relative to the example feasible set F in (3), where $\mathcal{J} = \{1, 2\}$. Note that $N_{(\infty,\mathcal{J})}(P) = P^{01}$ in this example.

containing P^{01}. (Figure 1 was drawn by explicitly determining points on the boundary of each set via a collection of linesearch procedures. For each of the four sets, a single linesearch started at $(0, 0)$ and moved into the first quadrant at an angle $\theta \in [0, \pi/2]$. The point on the boundary was precisely the point where the linesearch left the set. The linesearch was repeated for a sufficiently fine grid on $[0, \pi/2]$ for each of the four sets.)

Recall that P is the basic LP relaxation of F; in the figure, it is the largest set, which contains the other three. The next largest is $N_{(1,\mathcal{J})}$, a polyhedral set since $p = 1$, which contains the remaining two sets. $N_{(2,\mathcal{J})}$ is the projection of a second-order cone set and hence has a curved boundary. Finally, the depicted line segment between $(0, 0)$ and $(0, 1)$ is $N_{(\infty,\mathcal{J})}$, which equals P^{01} (in this example).

2.2 A different derivation

In the derivation of $N(P)$, we have strongly used the implication (2), which can be thought of as replacing two constraints by their product (respecting nonnegativity). We now show that one can obtain an alternate representation of the right-hand side of (2) via an alternate representation of $\|x_{\mathcal{J}} - d\|_p \leq r$. Still, multiplying constraints is the key idea.

Consider $\mathcal{K}_p$ and $\mathcal{K}_q$ as described in Section 1.1. Because $\mathcal{K}_q = \mathcal{K}_p^*$, it holds that

$$\begin{aligned} \|x_{\mathcal{J}} - d\|_p \leq r \quad &\iff \quad (r, x_{\mathcal{J}} - d) \in \mathcal{K}_p \\ &\iff \quad vr + u^T(x_{\mathcal{J}} - d) \geq 0 \quad \forall\ (v, u) \in \mathcal{K}_q. \end{aligned} \tag{4}$$

Proposition 2.2.1. *For a given $x \in \Re^n$, the right-hand side of (2) holds if and only if*

$$(b_i - a_i^T x)\left(vr + u^T(x_{\mathcal{J}} - d)\right) \geq 0 \quad \forall\ (v, u) \in \mathcal{K}_q. \tag{5}$$

Proof. ($\Rightarrow$): The right-hand side of (2) implies $b_i - a_i^T x \geq 0$. If $b_i - a_i^T x = 0$, then clearly (5) holds. On the other hand, if $b_i - a_i^T x > 0$, then dividing the right-hand side of (2) by $b_i - a_i^T x$ shows $\|x_{\mathcal{J}} - d\|_q \leq r$, which in turn implies $vr + u^T(x_{\mathcal{J}} - d) \geq 0$ for all $(v, u) \in \mathcal{K}_q$ by (4). Now multiplying with $b_i - a_i^T x$ implies (5).

($\Leftarrow$): If $b_i - a_i^T x = 0$, then the right-hand side of (2) holds trivially. On the other hand, if $b_i - a_i^T x \neq 0$, then for nonzero u the two inequalities

$$\begin{aligned} (b_i - a_i^T x)\left(\|u\|_q r + u^T(x_{\mathcal{J}} - d)\right) \geq 0 \\ (b_i - a_i^T x)\left(\|u\|_q r - u^T(x_{\mathcal{J}} - d)\right) \geq 0 \end{aligned}$$

together imply $b_i - a_i^T x > 0$. As a consequence, $vr + u^T(x_{\mathcal{J}} - d) \geq 0$ for all $(v, u) \in \mathcal{K}_q$, which means $\|x_{\mathcal{J}} - d\|_q \leq r$ by (4), which in turn implies the right-hand side of (2). ☐

An immediate consequence of Proposition 2.2.1 is that $\hat{P}$ defined in the derivation of $N(P)$ can be equivalently expressed using the semi-infinite collection of inequalities

$$(b_i - a_i^T x)(vr - u^T d) + b_i u^T x_{\mathcal{J}} - u^T X_{\mathcal{J}\cdot} a_i \geq 0 \qquad \forall\ \ (i, (v, u)) \in \mathcal{I} \times \mathcal{K}_q,$$

thus providing an equivalent definition of $N(P)$. This representation will be useful in the following subsection.

2.3 Comparison with existing approaches

In the derivation of $N(P)$, we did not use the full strength of the relationship $X = xx^T$ in the relaxation $\hat{P}$. In particular, we could have imposed in $\hat{P}$ the following two convex conditions, which are implied by $X = xx^T$:

$$X = X^T \qquad \text{and} \qquad \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0. \tag{6}$$

If we had imposed these, then $N(P)$ would be even tighter. However, we purposely did not impose them in order to have the weakest theoretical assumptions for the iterated procedure of Section 4. In practice, one would certainly want to impose as many constraints that can be handled efficiently. In particular, imposing symmetry $X = X^T$ can be useful to eliminate variables.

We mention the conditions (6) here because they facilitate comparison with existing lift-and-project methods in the following subsections.

2.3.1 Lovász-Schrijver

The LP-based approach of Lovász and Schrijver (1991) is derived like ours except that the following lifted and relaxed sets serve in the place of our $\hat{F}$ and $\hat{P}$:

$$\hat{F}_{\text{LS}} = \left\{ (x, X) \in \Re^n \times \Re^{n \times n} : \begin{array}{r} X = xx^T \qquad \operatorname{diag}(X) = x \\ (b_i - a_i^T x)x_k \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n] \\ (b_i - a_i^T x)(1 - x_k) \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n] \end{array} \right\}$$

$$\hat{P}_{\text{LS}} = \left\{ (x, X) \in \Re^n \times \Re^{n \times n} : \begin{array}{r} X = X^T \qquad \operatorname{diag}(X) = x \\ b_i x_k - X_{k\cdot} a_i \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n] \\ (b_i - a_i^T x) - (b_i x_k - X_{k\cdot} a_i) \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n] \end{array} \right\}.$$

We have the following proposition relating $\hat{F}_{\text{LS}}$ and $\hat{P}_{\text{LS}}$ to $\hat{F}$ and $\hat{L}$.

Proposition 2.3.1. *Let $p = \infty$ and $\mathcal{J} = [n]$. If the p-cone lift-and-project procedure also enforces the symmetry condition of (6), then $\hat{F} = \hat{F}_{\text{LS}}$ and $\hat{P} = \hat{P}_{\text{LS}}$.*

Proof. Note that $r = 1/2$ with $p = \infty$ and $\mathcal{J} = [n]$. It suffices to show that the conditions

$$\left\| b_i x - xx^T a_i - (b_i - a_i^T x)d \right\|_p \leq r(b_i - a_i^T x) \ \ \forall\ i \in \mathcal{I}$$

of $\hat{F}$ are equivalent to the conditions

$$\begin{array}{r} (b_i - a_i^T x)x_k \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n] \\ (b_i - a_i^T x)(1 - x_k) \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n] \end{array} \tag{7}$$

of $\hat{F}_{\text{LS}}$. By Proposition 2.2.1, the conditions of $\hat{F}$ can be replaced by

$$(b_i - a_i^T x)\left(vr + u^T(x - d)\right) \geq 0 \quad \forall\ (i, (v,u)) \in \mathcal{I} \times \mathcal{K}_1.$$

Since $\mathcal{K}_1$ is finitely generated by $\{(1, \pm e_1), \ldots, (1, \pm e_n)\}$, we can reduce these to

$$(b_i - a_i^T x)\left(r + e_k^T(x-d)\right) \geq 0 \quad \forall\ (i,k) \in \mathcal{I} \times [n]$$
$$(b_i - a_i^T x)\left(r - e_k^T(x-d)\right) \geq 0 \quad \forall\ (i,k) \in \mathcal{I} \times [n],$$

which simplify to (7), as desired. □

The following theorem is an immediate result.

Theorem 2.3.1. *Let $p = \infty$ and $\mathcal{J} = [n]$, and suppose the p-cone lift-and-project procedure enforces the symmetry condition of (6). Then the p-cone procedure replicates the LP-based Lovász-Schrijver lift-and-project procedure exactly.*

Lovász and Schrijver (1991) also proposed an SDP-based procedure, which also enforces the semidefiniteness condition of (6) in $\hat{P}_{\text{LS}}$. Identical arguments as above show that, if our procedure enforces both symmetry and semidefiniteness, then it replicates the SDP-based Lovász-Schrijver procedure.

Theorem 2.3.2. *Let $p = \infty$ and $\mathcal{J} = [n]$, and suppose the p-cone lift-and-project procedure enforces the symmetry and semidefiniteness conditions of (6). Then the p-cone procedure replicates the SDP-based Lovász-Schrijver lift-and-project procedure exactly.*

2.3.2 Kojima-Tunçel

Kojima and Tunçel (2000) present their method as a direct extension of the approach of Lovász and Schrijver (1991) to a much broader class of problems. So, in fact, their approach essentially reduces to that of Lovász-Schrijver in the case of 0-1 integer programming — with one important difference, which we explain next. This difference, in particular, will have relevance to our discussion and proofs in Section 4.

As discussed in the previous subsection, the Lovász-Schrijver approach is based on lifting with respect to the collection of constraints

$$(b_i - a_i^T x)x_k \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n]$$
$$(b_i - a_i^T x)(1 - x_k) \geq 0 \ \ \forall\ (i,k) \in \mathcal{I} \times [n]$$

In contrast, Kojima and Tunçel (2000) (see section 6, page 767, third full paragraph) lift with respect to the larger, extended collection

$$
\begin{aligned}
(b_i - a_i^T x)x_k &\geq 0 \;\; \forall\, (i,k) \in \mathcal{I} \times [n] \\
(b_i - a_i^T x)(1 - x_k) &\geq 0 \;\; \forall\, (i,k) \in \mathcal{I} \times [n] \\
(b_i - a_i^T x)(b_h - a_h^T x) &\geq 0 \;\; \forall\, (i,h) \in \mathcal{I} \times \mathcal{I} \\
x_j x_k &\geq 0 \;\; \forall\, (j,k) \in [n] \times [n] \\
x_j(1 - x_k) &\geq 0 \;\; \forall\, (j,k) \in [n] \times [n] \\
(1 - x_j)(1 - x_k) &\geq 0 \;\; \forall\, (j,k) \in [n] \times [n],
\end{aligned}
$$

which actually reduces to lifting with respect to

$$
(b_i - a_i^T x)(b_h - a_h^T x) \geq 0 \;\; \forall\, (i,h) \in \mathcal{I} \times \mathcal{I},
$$

since P implies the constraints $0 \leq x_k \leq 1$ by assumption. (For more insight on this point, please refer to Section 3.3 for a discussion on the monotonicity properties of lift-and-project procedures.)

This broader lifting makes the Kojima-Tunçel approach at least as strong as the Lovász-Schrijver approach (and at least as strong as our approach for $p = \infty$ and $\mathcal{J} = [n]$).

We remark that Lovász and Schrijver (1991) did indeed consider the broader lifting of Kojima and Tunçel (2000) but chose not to focus on it for algorithmic reasons. (This point is explained very well in detail by Kojima and Tunçel (2000).)

2.3.3 Balas-Ceria-Cornuéjols

The approach of Balas et al. (1993) can also be viewed as a special case of our approach. They choose a single index j and then apply the lift-and-project procedure outlined above in Section 2.1, replacing $\hat{F}$ and $\hat{P}$ by the following:

$$
\hat{F}_{\text{BCC}} = \left\{ (x,X) \in \Re^n \times \Re^{n \times n} : \begin{array}{c} X = xx^T \qquad \operatorname{diag}(X) = x \\ (b_i - a_i^T x)x_j \geq 0 \;\; \forall\, i \in \mathcal{I} \\ (b_i - a_i^T x)(1 - x_j) \geq 0 \;\; \forall\, i \in \mathcal{I} \end{array} \right\}
$$

$$
\hat{P}_{\text{BCC}} = \left\{ (x,X) \in \Re^n \times \Re^{n \times n} : \begin{array}{r} \operatorname{diag}(X) = x \\ b_i x_j - X_{j\cdot} a_i \geq 0 \;\; \forall\, i \in \mathcal{I} \\ (b_i - a_i^T x) - (b_i x_j - X_{j\cdot} a_i) \geq 0 \;\; \forall\, i \in \mathcal{I} \end{array} \right\}.
$$

We point out two important details. First, $\hat{P}_{\text{BCC}}$ does not enforce the symmetry condition $X = X^T$ of (6). Second, because all rows $X_{k\cdot}$ for $k \neq j$ are unconstrained except for the equation $X_{kk} = x_k$, $\hat{P}_{\text{BCC}}$ may be reduced to

$$\hat{P}_{\text{BCC}} = \left\{ (x, y) \in \Re^n \times \Re^n : \begin{array}{r} y_j = x_j \\ b_i x_j - a_i^T y \geq 0 \ \ \forall\, i \in \mathcal{I} \\ (b_i - a_i^T x) - (b_i x_j - a_i^T y) \geq 0 \ \ \forall\, i \in \mathcal{I} \end{array} \right\}$$

without affecting the resulting $N(P)$.

We claim that the Balas-Ceria-Cornuéjols approach is a special case of our method with $\mathcal{J} = \{j\}$ and arbitrary p.

Proposition 2.3.2. *Let $p \in [1, \infty]$ and $\mathcal{J} = \{j\}$ for some fixed index j. It holds that $\hat{F}_{\text{BCC}} = \hat{F}$ and $\hat{P}_{\text{BCC}} = \hat{P}$.*

Proof. It suffices to show that the conditions

$$\left\| b_i x_j - x_j x^T a_i - (b_i - a_i^T x) d \right\|_p \leq r(b_i - a_i^T x) \ \ \forall\, i \in \mathcal{I} \tag{8}$$

of $\hat{F}$ are equivalent to the conditions

$$\begin{aligned} (b_i - a_i^T x) x_j \geq 0 \ \ \forall\, i \in \mathcal{I} \\ (b_i - a_i^T x)(1 - x_j) \geq 0 \ \ \forall\, i \in \mathcal{I} \end{aligned}$$

of $\hat{F}_{\text{BCC}}$. Noting that $r = 1/2$ and $d = 1/2$ and that the p-norm is applied to a scalar in this case, (8) can be rewritten as

$$\left| (b_i - a_i^T x) \left(x_j - \frac{1}{2} \right) \right| \leq \frac{1}{2} (b_i - a_i^T x) \ \ \forall\, i \in \mathcal{I},$$

which is clearly equivalent to the conditions of $\hat{F}_{\text{BCC}}$. □

We thus have the following theorem.

Theorem 2.3.3. *Suppose $p \in [1, \infty]$ and $\mathcal{J} = \{j\}$ for some fixed index j. Then the p-cone lift-and-project procedure replicates the LP-based Balas-Ceria-Cornuéjols lift-and-project procedure exactly.*

3 Duality, Complexity, Monotonicity, and Fractional Extreme Points

In this section, we examine fundamental properties of the p-cone lift-and-project procedure outlined in Section 2.1. The first main result, proved in Section 3.2, establishes the theoretical iteration complexity for optimizing over $N_{(p,\mathcal{J})}(P)$. The second main result, proved below in Section 3.4.3, is that $N(P)$ contains no extreme points of P having fractional entries in $\mathcal{J}$, a result which will prove critical in Section 4.

Unless stated otherwise, we assume throughout this section that the pair $(p, \mathcal{J})$ are fixed, and we use the short notation $N(P)$ in place of $N_{(p,\mathcal{J})}(P)$. We also assume throughout that $\mathcal{J} = \{1, \ldots, |\mathcal{J}|\}$, i.e., $\mathcal{J}$ specifies the first $|\mathcal{J}|$ variables in x; this is for notational simplicity only.

3.1 Duality

For the discussion in this subsection, we assume that $|\mathcal{I}| < \infty$, i.e., P is a polytope.

Consider the relaxation $\min\{c^T x : x \in N(P)\}$ of the 0-1 integer program $\min\{c^T x : x \in F\}$. Its explicit p-cone representation is

$$\begin{aligned} \min \quad & c^T x \\ \text{s. t.} \quad & \operatorname{diag}(X) = x \\ & \big(r(b_i - a_i^T x), b_i x_{\mathcal{J}} - X_{\mathcal{J}.} a_i - (b_i - a_i^T x) d\big) \in \mathcal{K}_p \qquad \forall \;\; i \in \mathcal{I}, \end{aligned} \tag{9}$$

where $(x, X) \in \Re^n \times \Re^{n \times n}$. It can be derived that the associated dual is

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{I}} b_i (d^T u^i - r v_i) \\ \text{s. t.} \quad & \sum_{i \in \mathcal{I}} \left((d^T u^i - r v_i) a_i + b_i \begin{pmatrix} u^i \\ 0 \end{pmatrix} \right) + \lambda = c \\ & \sum_{i \in \mathcal{I}} \begin{pmatrix} u^i \\ 0 \end{pmatrix} a_i^T + \operatorname{Diag}(\lambda) = 0 \\ & (v_i, u^i) \in \mathcal{K}_q \qquad \forall \;\; i \in \mathcal{I}, \end{aligned} \tag{10}$$

where the dual variables are $\lambda \in \Re^n$ and $(v_i, u^i) \in \Re^{1+|\mathcal{J}|}$ for all $i \in \mathcal{I}$. As a means to illustrate the dual derivation without going deep into the details, we prove weak duality between (9) and (10) in the following proposition.

Proposition 3.1.1. *Suppose $|\mathcal{I}| < \infty$, i.e., P is a polytope. The dual of the p-cone relaxation (9) is the q-cone optimization (10). In particular, weak duality holds. If, in addition, both (9) and (10) have interior feasible solutions, then strong duality holds.*

Proof. We prove weak duality to illustrate the dual nature of (9) and (10). (The strong duality result is standard.) Let (x, X) be feasible for (9) and let $(\lambda, (v_i, u^i))$ be feasible for (10). Also, let $s_i := b_i - a_i^T x$. Then

$$
\begin{aligned}
c^T x - &\sum_{i \in \mathcal{I}} b_i (d^T u^i - r v_i) \\
&= \left(\sum_{i \in \mathcal{I}} \left((d^T u^i - r v_i) a_i + b_i \begin{pmatrix} u^i \\ 0 \end{pmatrix} \right) + \lambda \right)^T x - \sum_{i \in \mathcal{I}} b_i (d^T u^i - r v_i) \\
&= \sum_{i \in \mathcal{I}} (r v_i - d^T u^i) s_i + \sum_{i \in \mathcal{I}} b_i \, x^T \begin{pmatrix} u^i \\ 0 \end{pmatrix} + \lambda^T x \\
&= \sum_{i \in \mathcal{I}} \left(r s_i v_i + [b_i x_{\mathcal{J}} - s_i d]^T u^i \right) + \lambda^T x \\
&= \sum_{i \in \mathcal{I}} \left(r s_i v_i + [b_i x_{\mathcal{J}} - s_i d]^T u^i \right) - \left(\sum_{i \in \mathcal{I}} \begin{pmatrix} u^i \\ 0 \end{pmatrix} a_i^T \right) \bullet X \\
&= \sum_{i \in \mathcal{I}} \left(r s_i v_i + [b_i x_{\mathcal{J}} - s_i d]^T u^i \right) - \left(\sum_{i \in \mathcal{I}} u^i a_i^T \right) \bullet X_{\mathcal{J}\cdot} \\
&= \sum_{i \in \mathcal{I}} \left(r s_i v_i + [b_i x_{\mathcal{J}} - s_i d - X_{\mathcal{J}\cdot} a_i]^T u^i \right) \\
&\geq \sum_{i \in \mathcal{I}} 0 = 0.
\end{aligned}
$$

□

Related to the primal and dual problems (9) and (10), we consider the following question and derive a duality result: given $\bar{x} \in P \subseteq [0,1]^n$, is $\bar{x} \in N(P)$? To answer this question, we must determine whether or not the set

$$
\left\{ X \in \Re^{n \times n} : \begin{array}{c} \operatorname{diag}(X) = \bar{x} \\ (r \bar{s}_i, b_i \bar{x}_{\mathcal{J}} - X_{\mathcal{J}\cdot} a_i - \bar{s}_i d) \in \mathcal{K}_p \quad \forall \, i \in \mathcal{I} \end{array} \right\} \tag{11}
$$

is empty, where $\bar{s}_i := b_i - a_i^T \bar{x}$. This question is in turn related to the following set by

Proposition 3.1.2 below:

$$\left\{ \big(\lambda, (v_i, u^i)\big) \in \Re^n \times \mathcal{K}_q^{|\mathcal{I}|} : \begin{array}{c} \sum_{i\in\mathcal{I}} \begin{pmatrix} u^i \\ 0 \end{pmatrix} a_i^T + \mathrm{Diag}(\lambda) = 0 \\ \bar{x}^T\lambda + \sum_{i\in\mathcal{I}} \big(r\bar{s}_i\, v_i + (b_i\bar{x}_{\mathcal{J}} - \bar{s}_i d)^T u^i\big) < 0 \end{array} \right\} \tag{12}$$

Proposition 3.1.2. *Suppose $|\mathcal{I}| < \infty$, i.e., P is a polytope. Let $\bar{x} \in P$, and define $\bar{s}_i := b_i - a_i^T\bar{x}$ for all $i \in \mathcal{I}$. Then (11) is empty, i.e., $\bar{x} \notin N(P)$, if and only if (12) is nonempty.*

Proof. We first argue that, when feasible, the set

$$\left\{ \big(\lambda, (v_i, u^i)\big) : \sum_{i\in\mathcal{I}} \begin{pmatrix} u^i \\ 0 \end{pmatrix} a_i^T + \mathrm{Diag}(\lambda) = 0 \right\}$$

has nonempty interior with respect to the cones $\mathcal{K}_q \ni (v_i, u^i)$. This follows because we may arbitrarily increase each v_i without affecting the matrix equation. The proposition is now a straightforward application of the conic version of Farkas' lemma (Anderson and Nash, 1987). □

A comment on the assumption $|\mathcal{I}| < \infty$ in Proposition 3.1.2 is in order. As it will turn out, if one were able to relax this assumption, then the proof of Theorem 4.0.2 in Section 4 could be simplified somewhat. However, we are not sure if the result holds without this assumption.

3.2 Iteration complexity

For the discussion in this subsection, we assume that $|\mathcal{I}| < \infty$, i.e., P is a polytope.

The general interior-point methodology of Nesterov and Nemirovskii (1994) can be used to derive iteration complexity results for solving the p-cone relaxation (9) and/or its dual (10). Stated with respect to (9), the key result is as follows:

Theorem 3.2.1 (Nesterov and Nemirovskii (1994))**.** *Suppose that (9) is interior feasible with finite optimal value v^*. Let a polynomial-time self-concordant barrier for $\mathcal{K}_p$ with barrier parameter θ_p, an interior feasible solution (x^0, X^0), and a tolerance $\varepsilon > 0$ be given. Then there exists an algorithm ("short-step primal-only interior-point algorithm"), which delivers a solution (x^*, X^*) satisfying $c^Tx^* - v^* < \varepsilon$ within $O(\sqrt{\theta_p|\mathcal{I}|}\log(\varepsilon^{-1}(c^Tx^0 - v^*)))$ iterations, each of which takes polynomial time.*

As is evident from the theorem, the key ingredient determining the iteration complexity of the interior-point algorithm is the barrier parameter θ_p. It is well known that there exists a self-concordant barrier for $\mathcal{K}_2$ with $\theta_2 = 2 = O(1)$, and when $p \neq 2$, Andersen et al. (2002) show the existence of a self-concordant barrier for $\mathcal{K}_p$ such that $\theta_p = 4|\mathcal{J}| = O(|\mathcal{J}|)$. This implies the following corollary.

Corollary 3.2.1. *With respect to Theorem 3.2.1, (9) can be solved in* $O(\sqrt{|\mathcal{I}|})$ *iterations when* $p = 2$ *and* $O(\sqrt{|\mathcal{J}||\mathcal{I}|})$ *iterations otherwise.*

It is interesting that the iteration complexity does not depend on $|\mathcal{J}|$ when $p = 2$.

This corollary illustrates that, among all relaxations as p varies in $[1, \infty]$, the second-order cone relaxation has the lowest overall theoretical iteration complexity. In addition, when $p = q = 2$, one can also apply the stronger algorithmic framework of Nesterov and Todd (1997) for homogeneous self-dual cones to obtain long-step primal-dual path-following algorithms, which are the basis of high quality practical implementations.

From a theoretical point of view, then, one may be interested only in the relaxations when $p = 2$ (lowest iteration complexity) and $p = \infty$ (strongest relaxation and same iteration complexity as all other $p \neq 2$). Of course, what happens in practice may differ from theory, as we will see in Section 5.

To close this subsection, we remark that, for $p = 1$ and $p = \infty$, the iteration complexity given by Corollary 3.2.1 matches the iteration complexity obtained if one first formulates (9) as its standard LP representation and then applies an LP interior-point method to that representation.

3.3 Two types of monotonicity

Monotonicity is a relatively simple, but important, property of the p-cone procedure outlined in Section 2.1. In fact, there are two types of monotonicity (though both are derived from the same principle). The first involves the effect of the p-cone procedure on P and its subsets for fixed $(p, \mathcal{J})$, while the second involves the effect on P under different p and p'.

The monotonicity properties that we wish to prove for $N_{(p,\mathcal{J})}(P)$ stem directly from the derivation of the p-cone procedure and particularly from the fact that F, $\hat{F}$, and $\hat{P}$ are defined with respect to the inequalities

$$(b_i - a_i^T x) \left(r - \|x_{\mathcal{J}} - d\|_p \right) \geq 0 \qquad \forall\ i \in \mathcal{I}; \tag{13}$$

see also (2). It is evident that any strengthening of these inequalities in the representations of F, $\hat{F}$, and $\hat{P}$ can affect a corresponding strengthening of $N_{(p,\mathcal{J})}(P)$ around P^{01}. This is the key observation for the following two monotonicity properties.

The first monotonicity property involves strengthening the portion $b_i - a_i^T x$ of (13):

Proposition 3.3.1. *Let $p \in [1, \infty]$ and $\emptyset \neq \mathcal{J} \subseteq [n]$ be fixed. Suppose Q is a compact convex set such that $F \subseteq Q \subseteq P$. Then $P^{01} \subseteq N_{(p,\mathcal{J})}(Q) \subseteq N_{(p,\mathcal{J})}(P) \subseteq P$.*

Proof. The inclusions $P^{01} \subseteq N(Q)$ and $N(P) \subseteq P$ are derived directly from Proposition 2.1.2. To prove $N(Q) \subseteq N(P)$, we simply note that, with respect to $N(Q)$, the sets F, $\hat{F}$, and $\hat{P}$ are based on the inequalities

$$(g_\ell - f_\ell^T x)\left(r - \|x_{\mathcal{J}} - d\|_p\right) \geq 0 \qquad \forall\, \ell \in \mathcal{L},$$

where $Q = \left\{x \in \Re^n : f_\ell^T x \leq g_\ell \;\; \forall\, \ell \in \mathcal{L}\right\}$. Since $Q \subseteq P$, these inequalities are clearly a strengthening of (13), and so $N(Q) \subseteq N(P)$. □

The second monotonicity property involves strengthening the portion $r - \|x_{\mathcal{J}} - d\|_p$ of (13) and requires the following lemma:

Lemma 3.3.1. *Let $p' \geq 1$, and suppose $v \in \Re^s$ satisfies $\|v\|_{p'}^{p'} \leq s$. Then $\|v\|_p^p \leq s$ for all $p \in [1, p']$.*

Proof. Without loss of generality, we replace v by its component-wise absolute value, i.e., we assume $v_j \geq 0$ for all j.

As a function of p (v fixed), $f(p) := \|v\|_p^p = \sum_{j=1}^s v_j^p$ is convex, and so its maximum over $[1, p']$ occurs at 1 or p'. So to prove the lemma it suffices to show $f(1) \leq s$, i.e.,

$$\sum_{j=1}^s v_j \leq s \qquad \Longleftrightarrow \qquad \left(\sum_{j=1}^s v_j\right)^{p'} \leq s^{p'}.$$

Next, $g(a) := a^{p'}$ is convex for nonnegative a since $p' \geq 1$. In particular,

$$\begin{aligned}
\left(\sum_{j=1}^s v_j\right)^{p'} &= g\left(\sum_{j=1}^s v_j\right) = g\left(\sum_{j=1}^s \frac{1}{s} \cdot s v_j\right) \\
&\leq \sum_{j=1}^s \frac{1}{s}\, g(s v_j) = \sum_{j=1}^s s^{p'-1}\, v_j^{p'} = s^{p'-1} \sum_{j=1}^s v_j^{p'} \\
&= s^{p'-1}\, \|v\|_{p'}^{p'} \leq s^{p'},
\end{aligned}$$

as desired. □

Proposition 3.3.2. *Let $\emptyset \neq \mathcal{J} \subseteq [n]$ and $1 \leq p \leq p' \leq \infty$ be given. Define $r := \sqrt[p]{|\mathcal{J}|}/2$, $r' := \sqrt[p']{|\mathcal{J}|}/2$, and $d = e/2 \in \Re^{|\mathcal{J}|}$ in accordance with Proposition 2.1.1. If $x \in \Re^n$ satisfies $\|x_{\mathcal{J}} - d\|_{p'} \leq r'$, then $\|x_{\mathcal{J}} - d\|_p \leq r$. As a consequence, $N_{(p',\mathcal{J})}(P) \subseteq N_{(p,\mathcal{J})}(P)$.*

Proof. Regarding the first statement of the proposition, we can rearrange the desired implication as

$$\|2x_{\mathcal{J}} - e\|_{p'}^{p'} \le |\mathcal{J}| \qquad \Longrightarrow \qquad \|2x_{\mathcal{J}} - e\|_p^p \le |\mathcal{J}|,$$

which is true by Lemma 3.3.1. Next, the inclusion $N_{(p',\mathcal{J})}(P) \subseteq N_{(p,\mathcal{J})}(P)$ follows because (13) is strengthened when p and r are replaced by p' and r'. □

3.4 Elimination of fractional extreme points

We introduce the following definition:

> Let $\emptyset \ne \mathcal{J} \subseteq [n]$ be given. For any $\bar{x} \in \Re^n$, we say that $\bar{x}$ is $\mathcal{J}$-fractional if the subvector $x_{\mathcal{J}}$ is contained in $[0,1]^{|\mathcal{J}|}$ and has one or more fractional entries.

In this subsection, we prove that $N(P)$ contains no $\mathcal{J}$-fractional points, which are extreme in P. Said differently, we show that $N(P)$ cuts off all $\mathcal{J}$-fractional extreme points of P.

We start with the case that $|\mathcal{I}| < \infty$, i.e., P is a polytope; this is essentially a warm-up exercise. Then we extend the ideas for polytopes to consider when P is a ball. Finally, we use the analysis with balls to show our main result that, no matter the geometric structure of P, all $\mathcal{J}$-fractional points of P are cut off by $N(P)$.

3.4.1 Polytopes

When $|I| < \infty$, P is a polytope, and since P is bounded in $[0,1]^n$, we know that $|\mathcal{I}| > n$ and that P has extreme points. We prove the following proposition:

Proposition 3.4.1. *Suppose $|I| < \infty$ such that P is a polytope. Suppose $\bar{x}$ is a $\mathcal{J}$-fractional extreme point of P. Then $\bar{x} \notin N(P)$.*

We actually give two proofs since we feel that both are instructive. The first is a direct proof that the set (11) is empty, which implies $\bar{x} \notin N(P)$ (see the discussion in Section 3.1). The second is a direct proof that the set (12) is nonempty, which implies $\bar{x} \notin N(P)$ by Proposition 3.1.2.

Proof. We must show (11) is empty, where $\bar{s}_i := b_i - a_i^T \bar{x}$. Because $\bar{x}$ is an extreme point of P, there exists $\mathcal{T} \subseteq \mathcal{I}$ of size n such that $\bar{s}_i = 0$ for all $i \in \mathcal{T}$ and the set $\{a_i : i \in \mathcal{T}\}$ is linearly independent. Hence, any X in (11) must satisfy, for all $i \in \mathcal{T}$,

$$\begin{aligned}(0, b_i\bar{x}_{\mathcal{J}} - X_{\mathcal{J}\cdot}a_i) \in \mathcal{K}_p \quad &\Longleftrightarrow \quad X_{\mathcal{J}\cdot}a_i = b_i\bar{x}_{\mathcal{J}} \\ &\Longleftrightarrow \quad X_{\mathcal{J}\cdot}a_i = (a_i^T\bar{x})\bar{x}_{\mathcal{J}} \\ &\Longleftrightarrow \quad (X_{\mathcal{J}\cdot} - \bar{x}_{\mathcal{J}}\bar{x}^T)a_i = 0.\end{aligned}$$

By the linear independence of $\{a_i : i \in \mathcal{T}\}$, it follows that X must satisfy $X_{\mathcal{J}\cdot} = \bar{x}_{\mathcal{J}}\bar{x}^T$ with $\text{diag}(X) = \bar{x}$. However, since $\bar{x}$ is $\mathcal{J}$-fractional, these conditions are inconsistent. So in fact (11) is empty. □

Proof. This proof assumes without loss of generality that $\mathcal{J} = \{1, \ldots, |\mathcal{J}|\}$ in line with (12). We demonstrate that (12) is nonempty by constructing an explicit solution. Let $\mathcal{T}$ be defined as in the previous proof; we assume for simplicity that $\mathcal{T} = [n]$.

For all $i > n$, set $(v_i, u^i) = (0, 0)$. Then the matrix equation of (12) simplifies to

$$\begin{pmatrix} U \\ 0 \end{pmatrix} B^T + \text{Diag}(\lambda) = 0, \tag{14}$$

where $U := [u^1 \cdots u^n] \in \Re^{|\mathcal{J}|\times n}$ and $B := [a_1 \cdots a_n] \in \Re^{n\times n}$. We set the values of $u^1, \ldots, u^n$ via the equation

$$U := \left[B^{-T}\right]_{\mathcal{J}\cdot}.$$

Since $\{a_i : i \in \mathcal{T}\}$ are linearly independent, B is invertible, and so U is well-defined. We also set

$$\lambda := \begin{pmatrix} -e \\ 0 \end{pmatrix},$$

where e is the all-ones vector of length $|\mathcal{J}|$, so that (14) is satisfied. Finally, we set v_i to any value no less than $\|u^i\|_q$ so that $(v_i, u^i) \in \mathcal{K}_q$.

For our specification of λ and (v_i, u^i), it remains only to show

$$\begin{aligned}
0 &> \bar{x}^T\lambda + \sum_{i\in\mathcal{I}} \left(r\bar{s}_i\, v_i + (b_i\bar{x}_{\mathcal{J}} - \bar{s}_i d)^T u^i\right) \\
&= \bar{x}^T \begin{pmatrix} -e \\ 0 \end{pmatrix} + \sum_{i=1}^n \left(r\cdot 0\cdot v_i + (b_i\bar{x}_{\mathcal{J}} - 0\cdot d)^T u^i\right) \\
&= -e^T\bar{x}_{\mathcal{J}} + \bar{x}_{\mathcal{J}}^T \left(\sum_{i=1}^n b_i u^i\right) = -e^T\bar{x}_{\mathcal{J}} + \bar{x}_{\mathcal{J}}^T \left(\sum_{i=1}^n u^i a_i^T\right)\bar{x} \\
&= -e^T\bar{x}_{\mathcal{J}} + \bar{x}_{\mathcal{J}}^T U B^T\bar{x} = -e^T\bar{x}_{\mathcal{J}} + \bar{x}_{\mathcal{J}}^T \begin{pmatrix} \text{Diag}(e) & 0 \end{pmatrix}\bar{x} \\
&= -e^T\bar{x}_{\mathcal{J}} + \bar{x}_{\mathcal{J}}^T\bar{x}_{\mathcal{J}},
\end{aligned}$$

which follows because $\bar{x}_{\mathcal{J}}$ is fractional in $[0,1]^{|\mathcal{J}|}$. □

3.4.2 Balls

In the previous subsection, we showed that, if P is a polytope, then $N(P)$ cuts off all $\mathcal{J}$-fractional extreme points from P. The proof strongly used that every extreme point in a polytope corresponds to n linearly independent active constraints. For general P, however, extreme points do not necessarily correspond to n active constraints. For example, if P is a ball in the interior of $[0,1]^n$, then all extreme points of P have exactly one active constraint in the semi-infinite LP representation of P. In this subsection, we study balls to establish that $N(P)$ does in fact cut off all $\mathcal{J}$-fractional extreme points in this case as well. To avoid notational confusion with the P defined in Section 2.1, however, we will use B to denote the ball under investigation.

Let B be a ball centered at $h \in \Re^n$ with radius $R > 0$, i.e.,

$$\begin{aligned} B &:= \{x : \|x - h\|_2 \le R\} \\ &= \{x : w^T(x - h) \le R \ \ \forall\, w \ \text{ s.t. } \ \|w\|_2 = 1\}. \end{aligned} \tag{15}$$

In keeping with the development of the p-cone procedure, we could just as well assume that B is the intersection of a ball and $[0,1]^n$, but this is actually not necessary for the result that we present (and the analysis is a bit simpler without the assumption). The result is as follows:

Proposition 3.4.2. *Suppose B is a ball given by (15) for some center $h \in \Re^n$ and radius $R > 0$. Suppose $\bar{x}$ is a $\mathcal{J}$-fractional extreme point of B. Then $\bar{x} \notin N(B)$.*

The proof of Proposition 3.4.2, although related to the proof of Proposition 3.4.1 for polytopes, is technically quite different. The fundamental difference is that, for balls, we have only one active constraint at $\bar{x}$, whereas for polytopes, we have n linearly independent ones. Nevertheless, the idea of the proof below is to carefully select n linearly independent constraints, which are *nearly* active at $\bar{x}$. By analyzing those constraints, we see that they have the effect of cutting off $\bar{x}$ (similar to Proposition 3.4.1).

Proof. Since $\bar{x}$ is extreme, there exists some $\bar{w}$ with $\|\bar{w}\|_2 = 1$ such that $\bar{w}^T(\bar{x} - h) = R$. In fact, $\bar{w} = R^{-1}(\bar{x} - h)$ since $\|\bar{x} - h\|_2 = R$. This vector $\bar{w}$ will play an important role in our analysis. Related to $\bar{w}$, we also setup two additional vectors $\eta, \beta \in \Re^n$. First, we let η be any vector having all nonzero entries such that $\eta^T\bar{w} \ne 0$. For example, η could be taken as a small perturbation of $\bar{w}$ itself. (The condition $\eta^T\bar{w} \ne 0$ is critical later in the proof; see the appendix.) Second, we define $\beta := \eta^{-1}$.

For $\theta > 0$, define the following collection of n vectors, each of which is a unit-length perturbation of $\bar{w}$:

$$w^j(\theta) := \ell_j(\theta)^{-1} \left((1-\theta)\bar{w} + \theta\beta_j\, e_j\right) \qquad \forall \;\; j = 1, \ldots, n,$$

where

$$\ell_j(\theta) := \|(1-\theta)\bar{w} + \theta\beta_j\, e_j\|_2 = \sqrt{(1-\theta)^2 + \theta^2\beta_j^2 + 2(1-\theta)\theta\beta_j\, \bar{w}_j}.$$

Note that $\ell_j(\theta) > 0$ for θ small and so $w^j(\theta)$ is well-defined. We will consider the polyhedron

$$C(\theta) := \left\{x : w^j(\theta)^T(x-h) \le R \;\; \forall\, j = 1, \ldots, n\right\},$$

which contains B since its defining inequalities are a subset of those defining B. In particular, we will show that $\bar{x} \notin N(C(\theta))$ for θ sufficiently close to 0. Since $N(B) \subseteq N(C(\theta))$ due to monotonicity (see Proposition 3.3.1), this will imply $\bar{x} \notin N(B)$. We prove $\bar{x} \notin N(C(\theta))$ in the appendix. □

We will actually use a feature of the above proof again for the proof of Theorem 4.0.2 in Section 4. So we catalog this result for easier reference.

Corollary 3.4.1. *Let B and $\bar{x}$ be as in Proposition 3.4.2. Then there exists a polyhedron $C \supseteq B$ such that $\bar{x} \notin N(C) \supseteq N(B)$.*

Proof. The desired polyhedron C is simply $C(\theta)$ in the proof of Proposition 3.4.2. □

3.4.3 The general case

We now show that $N(P)$ cuts off all $\mathcal{J}$-fractional extreme points of P. The basic idea is that, given a $\mathcal{J}$-fractional extreme point $\bar{x} \in P$, there exists a ball $B \supseteq P$ such that $\bar{x}$ is also a $\mathcal{J}$-fractional extreme point of B. Thus, by Proposition 3.4.2 and the monotonicity property of Proposition 3.3.1, $\bar{x} \notin N(B) \supseteq N(P)$.

We first establish the existence of the ball B just described. (We note also that a similar result has been used in Kojima and Tunçel (2000).)

Proposition 3.4.3. *Let $\bar{x}$ be a $\mathcal{J}$-fractional extreme point of P. Then there exists a ball B such that $P \subseteq B$ and $\bar{x}$ is a $\mathcal{J}$-fractional extreme point of B.*

Proof. This proposition is just a simple application of standard convex analysis. Recall that P is compact convex. Hence, there exists a hyperplane $H := \{x : \alpha^T x = \beta\}$ supporting P at $\bar{x}$, i.e., $\bar{x} \in H$ and $P \setminus \{\bar{x}\} \subseteq H_{++} := \{x : \alpha^T x > \beta\}$. We also define $H_+ := \{x : \alpha^T x \ge \beta\}$.

Next, given $\gamma > 0$, we define a ball $B(\gamma)$ dependent on $\bar{x}$ and α:

$$B(\gamma) := \{x : \|x - (\bar{x} + \gamma\alpha)\|_2 \le \gamma\|\alpha\|_2\}.$$

It is easy to check that $B(\gamma) \subseteq H_+$ and that $\bar{x}$ is an extreme point of $B(\gamma)$. Furthermore, for every $x \in H_{++}$, there exists sufficiently large γ such that $x \in B(\gamma)$. Hence, because $P \setminus \{\bar{x}\} \subseteq H_{++}$ is bounded, there exists sufficiently large γ such that $P \setminus \{\bar{x}\} \subseteq B(\gamma)$, and so $P \subseteq B(\gamma)$. For any such large γ, we can take $B := B(\gamma)$ to achieve the proposition. □

With the above proposition in hand, we can prove the key theorem.

Theorem 3.4.1. *Let $\bar{x}$ be a $\mathcal{J}$-fractional extreme point of P. Then $\bar{x} \notin N(P)$.*

Proof. Let B be the ball of Proposition 3.4.3. Then, by Proposition 3.4.2, $\bar{x} \notin N(B)$. Since $N(B) \supseteq N(P)$ by monotonicity of Proposition 3.3.1, $\bar{x} \notin N(P)$. □

4 Iterated Procedure and Convergence

So far we have discussed how the p-cone procedure produces $N_{(p,\mathcal{J})}(P)$ from P for a given $(p, \mathcal{J})$. Because $N(P)$ is compact convex with its own semi-infinite outer description (which may or may not be known explicitly), we may conceptually apply the p-cone procedure — perhaps for a different choice of $(p, \mathcal{J})$ — to $N(P)$ itself. In fact, we may repeat the p-cone procedure *ad infinitum*. A key question is whether the resultant sequence of compact convex sets converges to P^{01}.

More formally, let $\{(p_k, \mathcal{J}^k)\}_{k\ge1}$ be a sequence of choices $p_k \in [1,\infty]$ and $\emptyset \ne \mathcal{J}^k \subseteq [n]$, and define $N^1(P) := N_{(p_1,\mathcal{J}^1)}(P)$ and $N^k(P) := N_{(p_k,\mathcal{J}^k)}(N^{k-1}(P))$ for all $k > 1$. We then ask whether $\lim_{k\to\infty} N^k(P)$ equals P^{01}.

Lovász and Schrijver (1991), Kojima and Tunçel (2000), and Balas et al. (1993) have all considered the same question for their own procedures. In particular, one may interpret Lovász and Schrijver (1991) as taking $p_k = \infty$ and $\mathcal{J}^k = [n]$ for all k; they show finite convergence after n iterations, i.e., $N^n(P) = P^{01}$. Recall that the method of Kojima and Tunçel (2000), when applied to 0-1 programs, essentially reduces to that of Lovász and Schrijver (1991); so they take the same p_k and $\mathcal{J}^k$. However, their method actually applies to a much broader class of quadratically constrained problems for which they show asymptotic (not finite) convergence. Finally, one may interpret Balas et al. (1993) as taking p_k arbitrary and $\mathcal{J}^k$ a single element in $[n]$. They prove that, if $\mathcal{J}^1 \cup \cdots \cup \mathcal{J}^n = [n]$, then $N^n(P) = P^{01}$.

We show in Theorem 4.0.2 below that the iterated p-cone procedure converges asymptotically for arbitrary $\{p_k\}_{k=1}^{\infty}$ as long as each index $j \in [n]$ appears infinitely often in the sequence $\{\mathcal{J}^k\}_{k=1}^{\infty}$. Before stating and proving the theorem, we discuss a few items.

First, it seems difficult to obtain a finite convergence proof for the general p-cone iterated procedure. In fact, we suspect — but are unable to give an explicit example at this time — that an infinite number of iterations is required in general. (Of course, for specific sequences $\{(p_k, \mathcal{J}^k)\}_{k=1}^{\infty}$, it may be possible to prove finiteness as with Lovász and Schrijver (1991) and Balas et al. (1993).) In addition, we give a computational example below in Section 4.1 to show that our method does not converge in n iterations in general.

Second, we suspect that obtaining a rate of asymptotic convergence is difficult as well. This relates to the difficulties with finite convergence but also stems from the methodology used to prove Theorem 3.4.1, which establishes that $\mathcal{J}$-fractional extreme points are cut off by the p-cone procedure. To establish a rate of convergence, it seems necessary to establish how "deep" these cuts are with respect to P^{01}, but the methodology of Theorem 3.4.1 uses the existence of cuts with little to no quantitative knowledge of their strength.

Third, we have mentioned that the more general approach of Kojima and Tunçel (2000) obtains asymptotic convergence in general. It is reasonable to ask if their approach or proof techniques may somehow subsume ours and hence prove convergence for us. However, this is not the case since their asymptotic analysis uses *all* valid "rank-2" quadratic inequalities, i.e., valid inequalities $(b_i - a_i^T x)(b_h - a_h^T x) \geq 0$ obtained by multiplying *any* pair of valid linear inequalities $b_i - a_i^T x \geq 0$ and $b_h - a_h^T x \geq 0$ for F. In contrast, our approach and analysis require only a partial subset of such inequalities (in particular those gotten by multiplying valid linear inequalities for the p-cone constraint $\|x_{\mathcal{J}} - d\|_p \leq r$ with the valid inequalities $b_i - a_i^T x \geq 0$ defining P). See Section 2.3.2 for more discussion. In this sense, one can think of our approach as proving asymptotic convergence under weaker conditions than those used by Kojima and Tunçel (2000) (although of course we are considering a special case compared to their general case).

We are now ready to state and prove the theorem.

Theorem 4.0.2. *Let $\{(p_k, \mathcal{J}^k)\}_{k\geq 1}$ be a sequence of choices $p_k \in [1, \infty]$ and $\emptyset \neq \mathcal{J}^k \subseteq [n]$, which give rise to compact, convex sets $N^k(P) \supseteq P^{01}$ via the definitions $N^1(P) := N_{(p_1,\mathcal{J}^1)}(P)$ and $N^k(P) := N_{(p_k,\mathcal{J}^k)}(N^{k-1}(P))$ for all $k > 1$. Then $N^k(P) \supseteq N^{k+1}(P)$ so that $\lim_{k\to\infty} N^k(P)$ exists and equals $\cap_{k\geq 1} N^k(P)$. In addition, if it holds that $\cup_{k\geq \bar{k}} \mathcal{J}^k = [n]$ for all $\bar{k}$, then $\lim_{k\to\infty} N^k(P) = P^{01}$.*

Proof. Since each $N^k(P)$ is compact and convex and contained in $N^{k-1}(P)$, $\lim_{k\to\infty} N^k(P)$ exists and equals $Z := \cap_{k=1}^{\infty} N^k(P)$. This proves the first part of the theorem.

To prove the second part, we first claim that every extreme point of Z is integer. Suppose for contradiction that $\bar{z}$ is a fractional extreme point of Z, and let j be any index where $\bar{z}_j$

is fractional. Next, let $\mathcal{S} := \{\mathcal{J} \subseteq [n] : j \in \mathcal{J}\}$. Theorem 3.4.1 implies $\bar{z} \notin N_{(1,\mathcal{J})}(Z)$ for all $\mathcal{J} \in \mathcal{S}$.

A continuity argument (which we prove two paragraphs below) implies that, for each $\mathcal{J} \in \mathcal{S}$, there exists $k_{\mathcal{J}}$ large enough so that $\bar{z} \notin N_{(1,\mathcal{J})}(N^{k-1}(P))$ for all $k \geq k_{\mathcal{J}}$. Define $\hat{k} := \max\{k_{\mathcal{J}} : \mathcal{J} \in \mathcal{S}\}$. In particular, consider $k \geq \hat{k}$ such that $j \in \mathcal{J}^k$. Since $\mathcal{J}^k \in \mathcal{S}$, it holds that $\bar{z} \notin N_{(1,\mathcal{J}^k)}(N^{k-1}(P))$. By the monotonicity property of Proposition 3.3.2, it also holds that $\bar{z} \notin N_{(1,\mathcal{J}^k)}(N^{k-1}(P)) \supseteq N_{(p_k,\mathcal{J}^k)}(N^{k-1}(P)) = N^k(P)$, which contradicts the statement $\bar{z} \in Z$. Hence, we conclude that every extreme point of Z is integer.

Since $P^{01} \subseteq Z$ by construction, it thus follows that $Z = P^{01}$.

Now we prove the continuity argument from above, i.e., for each $\mathcal{J} \in \mathcal{S}$, we prove the existence of $k_{\mathcal{J}}$ large enough so that $\bar{z} \notin N_{(1,\mathcal{J})}(N^{k-1}(P))$ for all $k \geq k_{\mathcal{J}}$. So let $\mathcal{J} \in \mathcal{S}$ be fixed. Since $\bar{z}$ is a $\mathcal{J}$-fractional extreme point of Z, by Proposition 3.4.3 there exists a ball $B \supseteq Z$ such that $\bar{z}$ is a $\mathcal{J}$-fractional extreme point of B. Furthermore, by Corollary 3.4.1, there exists a polyhedron

$$C = \{x \in \Re^n \ : \ f_\ell^T x \leq g_\ell \ \ \forall\ \ell \in \mathcal{L}\} \supseteq B$$

with $|\mathcal{L}|$ finite such that $\bar{x} \notin N_{(1,\mathcal{J})}(C)$. Hence, the following set is nonempty by Proposition 3.1.2:

$$\left\{ \big(\lambda, (v_\ell, u^\ell)\big) \in \Re^n \times \mathcal{K}_\infty^{|\mathcal{L}|} : \begin{array}{c} \sum_{\ell\in\mathcal{L}} \begin{pmatrix} u^\ell \\ 0 \end{pmatrix} f_\ell^T + \mathrm{Diag}(\lambda) = 0 \\ \bar{x}^T\lambda + \sum_{\ell\in\mathcal{L}} \left(\frac{|\mathcal{J}|}{2} \cdot \bar{s}_\ell\, v_\ell + (g_\ell \bar{z}_{\mathcal{J}} - \bar{s}_\ell d)^T u^\ell \right) < 0 \end{array} \right\}, \tag{16}$$

where $\bar{s}_\ell := g_\ell - f_\ell^T \bar{z}$. Since $N^{k-1}(P)$ converges to $Z \subseteq C$, for any $\epsilon > 0$, there exists k_ϵ large enough so that the constraints $\{f_\ell^T x \leq g_\ell + \epsilon : \ell \in \mathcal{L}\}$ are redundant for $N^{k-1}(P)$ for all $k \geq k_\epsilon$. In other words,

$$N^{k-1}(P) \subseteq C_\epsilon := \big\{x \in \Re^n \ : \ f_\ell^T x \leq g_\ell + \epsilon \ \ \forall\ \ell \in \mathcal{L}\big\} \qquad \forall \ \ k \geq k_\epsilon.$$

Moreover, by Proposition 3.1.2, $\bar{z} \notin N_{(1,\mathcal{J})}(C_\epsilon)$ if and only if

$$\left\{ \big(\lambda, (v_\ell, u^\ell)\big) : \begin{array}{c} \sum_{\ell\in\mathcal{L}} \begin{pmatrix} u^\ell \\ 0 \end{pmatrix} f_\ell^T + \mathrm{Diag}(\lambda) = 0 \\ \bar{x}^T\lambda + \sum_{\ell\in\mathcal{L}} \left(\frac{|\mathcal{J}|}{2} \cdot (\bar{s}_\ell + \epsilon) v_\ell + ((g_\ell + \epsilon)\bar{z}_{\mathcal{J}} - (\bar{s}_\ell + \epsilon) d)^T u^\ell \right) < 0 \end{array} \right\} \tag{17}$$

is nonempty. Note that (17) differs from (16) only in the appearance of ϵ in the strict inequality constraint. Clearly, for sufficiently small ϵ, (17) is nonempty because (16) is nonempty. For such small ϵ, it follows that $\bar{z} \notin N_{(1,\mathcal{J})}(C_\epsilon) \supseteq N_{(1,\mathcal{J})}(N^{k-1})(P)$ for all $k \geq k_\epsilon$. So the desired $k_{\mathcal{J}}$ is k_ϵ. $\square$

4.1 Counter-example for convergence in n iterations

Consider the example feasible set

$$F = \left\{ x \in \{0,1\}^2 : \begin{array}{c} -x_1 \leq 0 \\ -x_2 \leq 0 \\ x_1 + 2x_2 \leq 2.5 \\ 2x_1 + x_2 \leq 2.5 \end{array} \right\} = \{(0,0),(1,0),(0,1)\} \tag{18}$$

so that

$$P^{01} = \{(x_1, x_2) \geq 0 : x_1 + x_2 \leq 1\}.$$

Note that (18) is slightly different from example (3). Let $\mathcal{J} = \{1, 2\}$. We will show that $N_{(1,\mathcal{J})}(N_{(1,\mathcal{J})}(P)) \neq P^{01}$ for this example, which proves that the p-cone procedure does not converge to P^{01} after n iterations in general.

By enumerating extreme points of $N_{(1,\mathcal{J})}(P)$ and then determining a facet representation, it can be shown that

$$N_{(1,\mathcal{J})}(P) = \left\{ x \in \Re^2 : \begin{array}{c} -x_1 \leq 0 \\ -x_2 \leq 0 \\ 5x_1 + x_2 \leq 5 \\ x_1 + 5x_2 \leq 5 \\ 7x_1 + 3x_2 \leq 7.5 \\ 3x_1 + 7x_2 \leq 7.5 \end{array} \right\}$$

With this explicit representation of $N_{(1,\mathcal{J})}(P)$, it can then be shown computationally that the point $(\frac{5}{8}, \frac{5}{8})$ is in $N_{(1,\mathcal{J})}(N_{(1,\mathcal{J})}(P))$. However, $(\frac{5}{8}, \frac{5}{8})$ is clearly not in P^{01}.

We provide a picture of $P \supseteq N_{(1,\mathcal{J})}(P) \supseteq N_{(1,\mathcal{J})}(N_{(1,\mathcal{J})}(P)) \supseteq P^{01}$ in Figure 2. This figure was drawn via the same procedure as for Figure 1.

5 Computational Considerations

For fixed $\mathcal{J}$, the p-cone lift-and-project procedure gives rise to a family of relaxations of P^{01} parameterized by p. From the monotonicity property of Proposition 3.3.2, we know that the

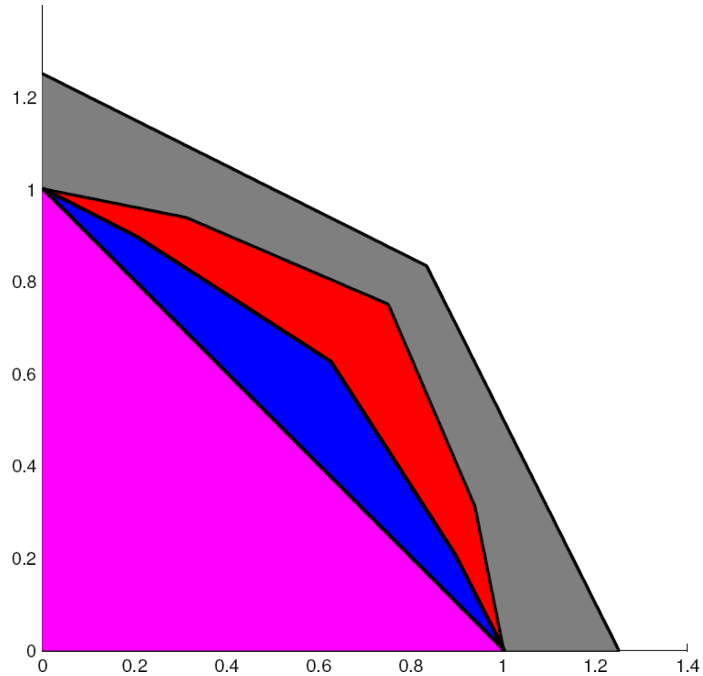

Figure 2: The four sets $P \supseteq N_{(1,\mathcal{J})}(P) \supseteq N_{(1,\mathcal{J})}(N_{(1,\mathcal{J})}(P)) \supseteq P^{01}$ relative to the example feasible set F in (18), where $\mathcal{J} = \{1, 2\}$. This figure demonstrates that the p-cone procedure does not converge to P^{01} after n iterations in general.

larger p is, the tighter the corresponding relaxation will be. So $p = \infty$ is the tightest. On the other hand, we have shown in Corollary 3.2.1 that optimizing over the $p = 2$ relaxation induces the lowest theoretical iteration complexity (in fact, an order of magnitude less than all other p, which themselves share the same iteration complexity). Thus, one may be particularly interested in the cases $p = 2$ (second-order cone programming) and $p = \infty$ (linear programming).

In this section, we computationally test these cases using state-of-the-art SOCP and LP software. We had hypothesized that the lower iteration complexity combined with the high quality of modern SOCP software could make solving $p = 2$ quicker than solving $p = \infty$ — perhaps much quicker so as to justify the loss in relaxation quality. However, in our computational experiments (described next), we have observed that $p = \infty$ solves faster than $p = 2$. Ultimately, we believe this computational performance is a testament to the quality of today's LP solvers.

For a graph G with vertex set V and edge set $E \subseteq V \times V$, the (unweighted) maximum stable set problem is

$$\alpha := \max\{e^T x \mid x_i + x_j \leq 1,\ (i,j) \in E,\ x \in \{0,1\}^n\}.$$

We test 8 instances of the maximum stable set problem obtained from the Center for Discrete Mathematics and Theoretical Computer Science (Johnson and Trick, 1996). Table 1 contains a basic description of the 8 graphs, where $|V|$, $|E|$, and α are the number of vertices, number of edges and the size of the maximum stable set, respectively. Note that the problems are roughly ordered by size.

Name	$\|V\|$	$\|E\|$	α
MANN-a9	45	72	16
johnson8-2-4	28	168	4
hamming6-2	64	192	32
keller4	171	5100	11
brock200_1	200	5066	21
san200_07_1	200	5970	30
sanr200_07	200	6032	18
c-fat200-1	200	18366	12

Table 1: Description of the 8 test problems

We solve both the $p = 2$ and $p = \infty$ relaxations with $\mathcal{J} = [n]$ and enforce the symmetry condition $X = X^T$ of (6) so as to eliminate about half of the variables in X. The SOCPs were solved using MOSEK 5.0, and the LPs were solved using both CPLEX 9.0 and Mosek 5.0. Pre-solving was turned off for all solvers, and computations were performed under the Linux operating system with a 2.8 GHz AMD Opteron processor and 4 GB of RAM.

Regarding the solution of the LPs, we used CPLEX to solve the dual form (10) using the dual simplex method, which gave better results than, for example, solving (9) with the dual simplex method. On the other hand, Mosek's LP solver optimizes (9) and (10) simultaneously using a primal-dual interior-point method.

Table 2 presents the results of our tests, comparing the bounds on α and solution times (in seconds). The values for those cells containing "$*$" were unavailable due to the solvers running out of memory.

Table 2 clearly shows the overall superiority of the LP relaxation (as solved by Mosek) in our tests. Still, it is worth noting that the SOCP relaxations solve more quickly than the LP relaxations via the dual simplex method. As mentioned above, in total we view Table 2 as convincing evidence of the strength of modern LP solvers.

6 Conclusions

In this paper, we have introduced lift-and-project procedures for 0-1 integer programming based on p-order cone programming. From the theoretical point of view, our approach gener-

Name	Bounds LP	Bounds SOCP	Times LP (CPLEX)	Times LP (MOSEK)	Times SOCP
MANN-a9	18.00	20.53	1.76	0.24	1.15
johnson8-2-4	9.33	12.19	0.09	0.13	0.24
hamming6-2	32.00	32.00	23.9	1.91	10.04
keller4	57.00	80.91	2140.25	1439.07	4671.50
brock200_1	66.67	95.05	33554.53	8087.98	11798.74
san200_07_1	66.67	95.05	100000.00	900.18	30471.23
sanr200_07	66.67	95.04	57573.88	6927.47	12139.65
c-fat200-1	*	*	*	*	*

Table 2: The bounds and times (in seconds) for solving the LP ($p = \infty$) and SOCP ($p = 2$) relaxations of the stable set instances from Table 1. Each LP is solved using two methods: the dual simplex method (CPLEX) and the primal-dual interior-point method (Mosek). An asterisk ($*$) indicates that the corresponding solver ran out of RAM. A time limit of 100,000 seconds is enforced for each run.

alizes and unifies several existing methods, which have been based on linear and semidefinite programming. Asymptotic convergence of the repeated application of our procedure has also been established, and for $p = 2$, when applying one iteration of the p-cone procedure, our method enjoys a theoretical iteration complexity, which is an order of magnitude faster than existing lift-and-project techniques. From the computational point of view, solving the SOCP corresponding to $p = 2$ is not competitive with solving the LP for $p = \infty$. Overall, we feel that the p-cone procedure makes a solid theoretical contribution to the literature on lift-and-project procedures, with possible computational improvements in the future as SOCP solvers become more and more efficient.

We conclude with a final observation. Given any SDP relaxation that, say, enforces $Y \succeq 0$, one can derive an SOCP relaxation from the SDP by enforcing positive semidefiniteness only on the 2×2 principal submatrices of Y (Kim and Kojima, 2003) since 2×2 semidefinite matrices can be modeled with a second-order cone of size 3. In addition, Kim and Kojima (2001) and Kim et al. (2003) show how to use the special structure of lift-and-project SDP relaxations to generate valid convex quadratic constraints (equivalent to SOC constraints), which are then enforced in place of $Y \succeq 0$. When $p = 2$, the approach in this paper is different from either just mentioned. Ours does not depend in any way on semidefiniteness and is more tied to the geometry of the feasible set of (1). In fact, irrespective of p, semidefiniteness can be applied to our procedure to further enhance its strength, and so the above SOCP ideas can also be applied to our procedure as well for any p.

References

F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program.*, 95(1, Ser. B):3–51, 2003. ISMP 2000, Part 3 (Atlanta, GA).

E. D. Andersen, C. Roos, and T. Terlaky. Notes on duality in second order and p-order cone optimization. *Optimization*, 51(4):627–643, 2002.

E. J. Anderson and P. Nash. *Linear programming in infinite-dimensional spaces.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Ltd., Chichester, 1987. Theory and applications, A Wiley-Interscience Publication.

E. Balas and M. Perregaard. A precise correspondence between lift-and-project cuts, simple disjunctive cuts, and mixed integer Gomory cuts for 0-1 programming. *Math. Program.*, 94 (2-3, Ser. B):221–245, 2003. The Aussois 2000 Workshop in Combinatorial Optimization.

E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58:295–324, 1993.

A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. ISBN 0-89871-491-5. Analysis, algorithms, and engineering applications.

D. Bienstock and M. Zuckerberg. Subset algebra lift operators for 0-1 integer programming. *SIAM J. Optim.*, 15(1):63–95 (electronic), 2004.

S. Burer and D. Vandenbussche. Solving lift-and-project relaxations of binary integer programs. *SIAM Journal on Optimization*, 16(3):493–512, 2006.

F. Glineur and T. Terlaky. Conic formulation for l_p-norm optimization. *J. Optim. Theory Appl.*, 122(2):285–307, 2004.

R. E. Gomory. An algorithm for integer solutions to linear programs. In R. Graves and P. Wolfe, editors, *Recent Advances in Mathematical Programming*, pages 269–302. McGraw-Hill, 1963.

D. Johnson and M. Trick. *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge.* American Mathematical Society, 1996.

S. Kim and M. Kojima. Second order cone programming relaxation of nonconvex quadratic optimization problems. *Optim. Methods Softw.*, 15(3-4):201–224, 2001.

S. Kim and M. Kojima. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. *Comput. Optim. Appl.*, 26(2):143–154, 2003.

S. Kim, M. Kojima, and M. Yamashita. Second order cone programming relaxation of a positive semidefinite constraint. *Optim. Methods Softw.*, 18(5):535–541, 2003.

M. Kojima and L. Tunçel. Cones of matrices and successive convex relaxations of nonconvex sets. *SIAM J. Optim.*, 10(3):750–778, 2000.

P. Krokhmal and P. Soberanis. Risk optimization with p-order conic constraints: A linear programming approach. Working paper, University of Iowa, Iowa City, IA, USA, 2008.

J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2001.

L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

Y. E. Nesterov and A. S. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, Philadelphia, 1994.

Y. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, 1997.

P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003. Algebraic and geometric methods in discrete optimization.

H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990.

G. Xue and Y. Ye. An efficient algorithm for minimizing a sum of p-norms. *SIAM J. Optim.*, 10(2):551–579 (electronic), 2000.

Appendix

We continue the proof of Proposition 3.4.2.

Proof. We introduce some notation to simplify our handling of $C(\theta)$. Define $\ell(\theta) := [\ell_1(\theta), \ldots, \ell_n(\theta)]^T$ and $L(\theta) := \mathrm{Diag}(\ell(\theta))$. Also define

$$W(\theta) := \left((1-\theta)\bar{w}e^T + \theta\,\mathrm{Diag}(\beta)\right) L(\theta)^{-1}$$

so that $W(\theta)_{\cdot j} = w^j(\theta)$. Then

$$x \in C(\theta) \qquad \Longleftrightarrow \qquad W(\theta)^T x \le \sigma(\theta),$$

where $\sigma(\theta) := Re + W(\theta)^T h$. According to Proposition 3.1.2, $\bar{x} \notin N(C(\theta))$ if and only if the set

$$\left\{ \left(\lambda, (v_j, u^j)\right) \in \Re^n \times \mathcal{K}_q : \begin{array}{c} \sum_{j=1}^n \begin{pmatrix} u^j \\ 0 \end{pmatrix} w^j(\theta)^T + \mathrm{Diag}(\lambda) = 0 \\ \bar{x}^T\lambda + \sum_{j=1}^n \left(r\bar{s}_j(\theta)\, v_j + (\sigma_j(\theta)\bar{x}_{\mathcal{J}} - \bar{s}_j(\theta)d)^T u^j \right) < 0 \end{array} \right\} \tag{19}$$

is nonempty, where $\bar{s}(\theta) := \sigma(\theta) - W(\theta)^T \bar{x}$. Letting U denote the matrix whose columns are the vectors $u^1, \ldots, u^n$ and making use of $W(\theta)$, the matrix equation of (19) can be expressed alternatively as

$$\begin{pmatrix} U \\ 0 \end{pmatrix} W(\theta)^T + \operatorname{Diag}(\lambda) = 0$$

For $\theta > 0$ sufficiently close to 0, we have $\theta + (1-\theta)\beta^{-T}\bar{w} = \theta + (1-\theta)\eta^T\bar{w} \neq 0$. Recall $\eta = \beta^{-1}$ and $\eta^T\bar{w} \neq 0$. Hence, by the standard Shermann-Morrison-Woodbury formula,

$$W(\theta)^{-1} = \theta^{-1} L(\theta) \operatorname{Diag}(\eta) \left(I - \left(\frac{1-\theta}{\theta + (1-\theta)\eta^T\bar{w}} \right) \bar{w}\eta^T \right).$$

Defining

$$U := \left[W(\theta)^{-T} \right]_{\mathcal{J}\cdot} \qquad \text{and} \qquad \lambda := \begin{pmatrix} -e \\ 0 \end{pmatrix},$$

where e is all-ones of length $|\mathcal{J}|$, then λ and U satisfy the equation of (19). We can also define $v_j := \|u^j\|_q$ so that $(v_j, u^j) \in \mathcal{K}_q$. Now, to show $\bar{x} \notin N(C(\theta))$ it remains only to show that the inequality of (19) holds for our selected λ and (v_j, u^j).

As a matter of fact, we will show the following (which suffices):

$$\lim_{\theta \to 0^+} \left[\bar{x}^T\lambda + \sum_{j=1}^n \left(r\bar{s}_j(\theta)\, v_j + (\sigma_j(\theta)\bar{x}_{\mathcal{J}} - \bar{s}_j(\theta)d)^T u^j \right) \right] = -e^T\bar{x}_{\mathcal{J}} + \bar{x}_{\mathcal{J}}^T\bar{x}_{\mathcal{J}} < 0.$$

The inequality holds because $\bar{x}_{\mathcal{J}}$ is fractional in $[0,1]^{|\mathcal{J}|}$ by assumption. Note that, in the expression of the limit, (v_j, u^j) depend on θ even though the notation does not reflect this. Actually, from here on out, we drop the bar and θ notations in hopes of simplifying the presentation.

So consider the expression whose limit in terms of θ we would like to analyze:

$$\begin{aligned} & x^T\lambda + \sum_{j=1}^n \left(rs_j\, v_j + (\sigma_j x_{\mathcal{J}} - s_j d)^T u^j \right) \\ & \qquad = -e^T x_{\mathcal{J}} + \sum_{j=1}^n \left(rs_j \, \|u^j\|_q + (\sigma_j x_{\mathcal{J}} - s_j d)^T u^j \right). \qquad (20) \end{aligned}$$

We make the following claim:

$$\text{Claim 1:} \quad \sum_{j=1}^n (\sigma_j x_{\mathcal{J}} - s_j d)^T u^j = R(x_{\mathcal{J}} - d)^T U e + (x_{\mathcal{J}} - d)^T h_{\mathcal{J}} + d^T x_{\mathcal{J}}.$$

Under this claim, (20) becomes

$$-e^T x_{\mathcal{J}} + \sum_{j=1}^{n} r s_j \|u^j\|_q + R(x_{\mathcal{J}} - d)^T U e + (x_{\mathcal{J}} - d)^T h_{\mathcal{J}} + d^T x_{\mathcal{J}} \tag{21}$$

We then make the following two additional claims:

$$\text{Claim 2:} \qquad \lim_{\theta \to 0^+} s_j u^j = 0$$
$$\text{Claim 3:} \qquad \lim_{\theta \to 0^+} Ue = R^{-1}(x_{\mathcal{J}} - h_{\mathcal{J}}).$$

Under these two claims, the limit as $\theta \to 0^+$ of (21) is

$$\begin{aligned} &- e^T x_{\mathcal{J}} + \sum_{j=1}^{n} r \cdot 0 + R(x_{\mathcal{J}} - d)^T \left(R^{-1}(x_{\mathcal{J}} - h_{\mathcal{J}})\right) + (x_{\mathcal{J}} - d)^T h_{\mathcal{J}} + d^T x_{\mathcal{J}} \\ &\quad = -e^T x_{\mathcal{J}} + x_{\mathcal{J}}^T x_{\mathcal{J}}, \end{aligned}$$

as desired. So it remains to prove the three claims.

Proof of Claim 1. Consider the following chain of equalities, which proves the claim:

$$\begin{aligned} \sum_{j=1}^{n} (\sigma_j x_{\mathcal{J}} - s_j d)^T u^j &= \operatorname{trace}\left(\left[\sigma\, x_{\mathcal{J}}^T - s\, d^T\right] U\right) \\ &= \operatorname{trace}\left(\left[\sigma\, x_{\mathcal{J}}^T - (\sigma - W^T x) d^T\right] U\right) \\ &= \operatorname{trace}\left(\left[\sigma (x_{\mathcal{J}} - d)^T + W^T x d^T\right] U\right) \\ &= \operatorname{trace}\left(\left[(Re + W^T h)(x_{\mathcal{J}} - d)^T + W^T x d^T\right] U\right) \\ &= \operatorname{trace}\left(U \left[(Re + W^T h)(x_{\mathcal{J}} - d)^T + W^T x d^T\right]\right) \\ &= \operatorname{trace}\left(RUe(x_{\mathcal{J}} - d)^T + UW^T h (x_{\mathcal{J}} - d)^T + UW^T x d^T\right) \\ &= \operatorname{trace}\left(RUe(x_{\mathcal{J}} - d)^T + \begin{pmatrix} I & 0 \end{pmatrix} h (x_{\mathcal{J}} - d)^T + \begin{pmatrix} I & 0 \end{pmatrix} x d^T\right) \\ &= \operatorname{trace}\left(RUe(x_{\mathcal{J}} - d)^T + h_{\mathcal{J}} (x_{\mathcal{J}} - d)^T + x_{\mathcal{J}} d^T\right) \\ &= R(x_{\mathcal{J}} - d)^T U e + (x_{\mathcal{J}} - d)^T h_{\mathcal{J}} + d^T x_{\mathcal{J}}. \end{aligned}$$

Proof of Claim 2. For $k \in \mathcal{J}$,

$$\begin{aligned} u_k^j &= \left[W^{-T}\right]_{kj} = \left[W^{-1}\right]_{jk} \\ &= \theta^{-1} \ell_j \eta_j \left(I_{jk} - \left(\frac{1 - \theta}{\theta + (1 - \theta)\eta^T w} \right) w_j \eta_k \right). \end{aligned}$$

Also,

$$
\begin{aligned}
s_j &= R - [W^T]_{j\cdot}(x-h) = R - W^T_{\cdot j}(x-h) \\
&= R - R\,W^T_{\cdot j} w = R\left(1 - W^T_{\cdot j} w\right) \\
&= R\left(1 - \left[\ell_j^{-1}\left((1-\theta)w + \theta\beta_j e_j\right)\right]^T w\right) \\
&= R\left(1 - \ell_j^{-1}(1-\theta) - \ell_j^{-1}\,\theta\,\beta_j\,w_j\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
s_j u^j_k &= R\left(1 - \ell_j^{-1}(1-\theta) - \ell_j^{-1}\,\theta\,\beta_j\,w_j\right)\cdot \theta^{-1}\ell_j\eta_j\left(I_{jk} - \left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right) w_j\eta_k\right) \\
&= R\,\theta^{-1}\eta_j\left(\ell_j - (1-\theta) - \theta\,\beta_j\,w_j\right)\left(I_{jk} - \left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right) w_j\eta_k\right).
\end{aligned}
$$

Since $\eta^T w \neq 0$,

$$
\lim_{\theta\to 0^+} R\eta_j\left(I_{jk} - \left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right) w_j\eta_k\right) = R\eta_j\left(I_{jk} - (\eta^T w)^{-1} w_j\eta_k\right).
$$

On the other hand, a Taylor-series expansion reveals that $\theta^{-1}\left(\ell_j - (1-\theta) - \theta\,\beta_j\,w_j\right)$ is $o(\theta)$ so that

$$
\lim_{\theta\to 0^+} \theta^{-1}\left(\ell_j - (1-\theta) - \theta\,\beta_j w_j\right) = 0.
$$

It follows that $\lim_{\theta\to 0^+} s_j u^j_k = 0$, as desired.

Proof of Claim 3. For $k \in \mathcal{J}$, we have

$$
\begin{aligned}
[Ue]_k &= \sum_{j=1}^n U_{kj} = \sum_{j=1}^n [W^{-T}]_{kj} = \sum_{j=1}^n W^{-1}_{jk} \\
&= \sum_{j=1}^n \theta^{-1}\ell_j\eta_j\left(I_{jk} - \left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right) w_j\eta_k\right) \\
&= \theta^{-1}\ell_k\eta_k - \theta^{-1}\sum_{j=1}^n \ell_j\eta_j\left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right) w_j\eta_k.
\end{aligned}
$$

A Taylor-series expansions shows that

$$
\ell_k = 1 + (\beta_k w_k - 1)\theta + o(\theta^2)
$$

$$
\ell_j\left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right) = (\eta^T w)^{-1} + (\eta^T w)^{-1}(\beta_j w_j - 1)\theta - (\eta^T w)^{-2}\theta + o(\theta^2)
$$

so that

$$
\begin{aligned}
&\sum_{j=1}^{n}\left[\ell_j\left(\frac{1-\theta}{\theta+(1-\theta)\eta^T w}\right)w_j\eta_j\right]\\
&=\sum_{j=1}^{n}\left[\big((\eta^T w)^{-1}+(\eta^T w)^{-1}(\beta_j w_j-1)\theta-(\eta^T w)^{-2}\theta+o(\theta^2)\big)\,w_j\eta_j\right]\\
&=1+\theta\left((\eta^T w)^{-1}w^T w-1\right)-\theta(\eta^T w)^{-1}+o(\theta^2)\\
&=1-\theta+o(\theta^2)
\end{aligned}
$$

Therefore, plugging in the above expression for ℓ_k, we have

$$
\begin{aligned}
[Ue]_k &= \theta^{-1}\eta_k + w_k - \eta_k - \theta^{-1}\eta_k\left(1-\theta+o(\theta^2)\right)+o(\theta)\\
&= \theta^{-1}\eta_k + w_k - \eta_k - \theta^{-1}\eta_k + \eta_k + o(\theta)\\
&= w_k + o(\theta).
\end{aligned}
$$

This proves $\lim_{\theta\to\theta^+} Ue = w_k = R^{-1}(x_k - h_k)$ for all $k \in \mathcal{J}$, as claimed. □