On Interval-subgradient and No-good Cuts

Claudia D'Ambrosio

DEIS, University of Bologna, Italy, c.dambrosio@unibo.it.

Antonio Frangioni

Dipartimento di Informatica, University of Pisa, Italy, frangio@di.unipi.it.

Leo Liberti

LIX, École Polytechnique, France, liberti@lix.polytechnique.fr.

Andrea Lodi

DEIS, University of Bologna, Italy, andrea.lodi@unibo.it.

Abstract

Interval-gradient cuts are (nonlinear) valid inequalities for nonconvex NLPs defined for constraints $g(x) \leq 0$ with g being continuously differentiable in a box $[\underline{x}, \bar{x}]$. In this paper we define interval-subgradient cuts, a generalization to the case of nondifferentiable g, and show that no-good cuts (which have the form $\|x - \hat{x}\| \geq \varepsilon$ for some norm and positive constant ε) are a special case of interval-subgradient cuts whenever the 1-norm is used. We then briefly discuss what happens if other norms are used.

1 Introduction

We consider a general (nonconvex) Nonlinear Program (NLP)

$$(P) \quad \min \quad f(x) \tag{1}$$

$$g_j(x) \leq 0 \qquad j \in C \tag{2}$$

$$\underline{x}_i \leq x_i \leq \overline{x}_i \qquad i \in N \tag{3}$$

where the constraint functions $g_j : \mathbb{R}^n \to \mathbb{R}$ ($n = |N|$) are not necessarily convex. We denote by $X = [\underline{x}, \overline{x}]$ the (finite) box containing the feasible region.

If no further structure is known for problem (1)–(3), the most widely used solution algorithm is spatial Branch-and-Bound (sBB) [28, 18, 6]. This involves finding a lower and an upper bound to the optimal objective function value. Whilst any feasible point of P yields an upper bound, lower bounds are obtained by solving a relaxation of P. If these bounds differ by more than a required solution accuracy $\varepsilon > 0$, then two sets X^ℓ, X^r are determined so that $X^\ell \cup X^r$ contains the feasible region. This procedure is applied recursively to each of the problems (P subject to $x \in X^\ell$) and (P subject to $x \in X^r$). The disjunction given by X^ℓ, X^r is chosen so that it changes the formulation of the relaxation: in particular, convergence is attained if the lower bound is guaranteed to increase monotonically. Common choices for generating the disjunction are to select a branching variable and a branching point in its range, and construct X^ℓ, X^r as the two sub-boxes obtained by splitting X along the branching variable coordinate at the branching point. Iterating this procedure, sBB generates a search tree whose exploration finitely yields a ε-optimal solution of P, which means that, technically speaking, it is an approximation algorithm (for specific problem structures, convergence to an exact optimum is possible [1, 8]). In general, setting $\varepsilon = 0$ might yield a nonterminating procedure. Within the sBB algorithm, if the solution $\hat{x}$ for the the relaxation is feasible for P, then the lower bound is surely larger than or equal to the upper bound and no branching occurs (the node is *fathomed*). If, instead $\hat{x}$ is infeasible for P, it is highly desirable to tighten the current relaxation and improve the bound by adding a *valid cutting plane* (*cut* for short) that cuts off $\hat{x}$.

Although (linear) cutting planes have been an essential part of Branch-and-Bound (BB) algorithms for Mixed-Integer Linear Programming (MILP) for decades now, generic sBB implementations have only recently started to include nontrivial cuts. A good review for existing Mixed-Integer Nonlinear Programming

(MINLP) cuts is [21, Sect. 7.1]. It includes linearization or outer approximation cuts (tangents at $\hat{x}$ whenever the relaxation is convex), knapsack cuts (which require solving an auxiliary global optimization problem), interval gradient cuts (discussed below), Lagrangian cuts (derived from a "partial dual" relating to some linear constraints in the problem), and level cuts (derived from an upper bound to the optimal objective function value). RLT-type cuts, derived by multiplying constraint factors (e.g. if $g_i(x) \leq 0$ and $g_j(x) \leq 0$, then $g_i(x)g_j(x) \geq 0$ is a valid inequality) are discussed in [26], and a specialization thereof in [19]. In [22], lifting techniques are discussed in the framework of NLP; [25] discusses an extension of the RLT to convex Mixed-Integer Programming (MIP). A certain attention has been devoted to conic MIP [10, 2]; in part, this is due to the fact that Lift&Project techniques (see, e.g., [3]) to compute valid inequalities for the union of two convex sets can easily be extended to the nonlinear setting [11], and this may produce strong conical reformulations of MIPs [27, 15] out of which effective cuts may be obtained [14].

In this paper we consider in particular *Interval-gradient cuts* [7, 21]. Generated from constraints (2), these cuts are based on estimating the range of the gradient of each of the functions g_j over the box X. Our first result is the generalization of the concept of interval-gradient cuts to that of *interval-subgradient cuts*, so as to allow application to nondifferentiable functions. We show by means of an example that this may lead to stronger cuts with respect to those obtained by a smooth reformulation of the nonsmooth constraint.

Moreover, we consider the extension to MINLP of a classical family of MILP cuts mostly known as *No-good cuts* (or *Farkas cuts*) and originally introduced, to the best of our knowledge, in [4] with the name of *canonical cuts*. These cutting planes are generated with respect to a specific solution $\hat{x}$ by imposing a positive distance between $\hat{x}$ and any new solution[1]. Such a distance can be enforced in the MINLP context through any norm while the 1-norm is used in MILP. Our main result is to show that no-good cuts in the 1-norm are a special case of interval-subgradient cuts. Furthermore, we discuss the case of no-good cuts with a p-norm for any $p > 1$, which are stronger than those with the 1-norm, showing that the corresponding interval-subgradient cuts are the same (and, therefore, not stronger than) those obtained by the 1-norm no-good cut.

The paper is organized as follows. In Sections 2 and 3 interval-gradient/subgradient and no-good cuts are presented, respectively. In Section 4 we show how to obtain no-good cuts starting from interval-subgradient cuts. In Section 5 we discuss no-good cuts derived from more general norms and their relationships. Finally, Section 6 concludes the paper.

2 Interval-gradient and Interval-subgradient Cuts

Let g_j be a selected nonconvex constraint in the set (2) above. Because in this section the index j is fixed, for the sake of simplifying the notation we drop it. We assume knowledge of the *interval-gradient* of g over X, i.e., of a finite box $D = [\underline{d}, \overline{d}]$ such that $\nabla g(x) \in D$ for all $x \in X$. Of course, this definition requires g to be differentiable everywhere on X. Then, one can show [7, 21] that the (nonconvex) function

$$\underline{g}(x) := g(\hat{x}) + \min_{d \in D} d(x - \hat{x}) \tag{4}$$

with $\hat{x} \in X$, underestimates g in the feasible region, i.e., $\underline{g}(x) \leq g(x)$ for all $x \in X$. Therefore, the *interval-gradient* (nonconvex) cut

$$\underline{g}(x) \leq 0 \tag{5}$$

is valid.

We now proceed to show that interval-gradient cuts can be defined even for *nondifferentiable* constraint functions g, as long as they are locally Lipschitz at every point in an open set containing X. This requires appropriate tools from nondifferentiable analysis, and in particular *Clarke's subgradient*

$$\partial g(x) \ := \ \left\{ \ \xi \in \mathbb{R}^n \ : \ g^\circ(x; v) \geq \xi v \quad \forall \ v \in \mathbb{R}^n \ \right\}$$

[1] No-good cuts have been recently used in MINLP in [20].

where

$$g^\circ(x;\xi) \;:=\; \limsup_{y\to x, t\downarrow 0} \frac{g(y+t\xi)-g(y)}{t}$$

is *Clarke's generalized directional derivative.* We will loosely refer to the elements $\xi \in \partial g(x)$ as subgradients, mostly in homage to their convex counterparts (see below). It can be shown [12] that ∂g is a sound generalization of the gradient ∇g at least in the case where g is locally Lipschitz at all points of X, because:

- $\partial g(x)$ is nonempty, convex and compact for each $x \in X$;
- whenever g is differentiable at x, $\partial g(x) = \{\ \nabla g(x)\ \}$;
- if g is convex, then $\partial g(x)$ coincides with the standard definition of subdifferential from convex analysis, that is the set of all subgradients $\xi \in \mathbb{R}^n$ satisfying

  $$g(y) \geq g(x) + \xi(y-x) \quad \forall\, y \in \mathbb{R}^n$$

  (known as the subgradient inequality); furthermore, since $\partial(-f)(x) = -\partial f(x)$, the same holds for concave functions (modulo the appropriate change in sign);
- if g is locally Lipschitz at each point of (the compact set) X, then it is globally Lipschitz on the whole of X; therefore, there exists a finite box $D = [\underline{d}, \overline{d}]$ such that $\partial g(x) \subseteq D$ for all $x \in X$, since all subgradients belong to the ball of center 0 and radius K, where $K < \infty$ is the global Lipschitz constant of g over X [12, Proposition 2.1.2(a)].

All this leads to the following proposition:

Proposition 2.1. *Let g be locally Lipschitz at every point in an open set containing X, let D be a finite box such that $\partial g(x) \subseteq D$ for all $x \in X$, and let $\underline{g}(x) = g(\hat{x}) + \min_{d\in D} d(x-\hat{x})$ as in* (4). *Then the inequality $\underline{g}(x) \leq 0$ is valid for P.*

Proof. We simply invoke the Mean-Value Theorem for nondifferentiable functions [12, Theorem 2.3.7], which states that, given x and $\hat{x}$ such that g is Lipschitz in an open set containing the (closed) interval $[\hat{x}, x]$ there exists some u in the (open) interval $(\hat{x}, x)$ and some $\xi \in \partial g(u)$ such that $g(x) = g(\hat{x}) + \xi(x-\hat{x})$. Whence, $g(x) \geq g(\hat{x}) + \min_{d\in D} d(x-\hat{x}) = \underline{g}(x)$ for all $x \in X$, as desired. ☐

Therefore, (5) is also valid in the nondifferentiable case. We refer to these as *interval-subgradient* cuts, as D can be reasonably called the *interval-subgradient* of g over X.

For future reference, we note here that (5) can be reformulated by means of added binary variables and constraints as follows:

$$g(\hat{x}) + \sum_{i\in N}(\underline{d}_i x_i^+ - \overline{d}_i x_i^-) \leq 0 \tag{6}$$

$$x - \hat{x} = x^+ - x^- \tag{7}$$

$$x_i^+ \leq z_i(\overline{x}_i - \underline{x}_i) \qquad i \in N \tag{8}$$

$$x_i^- \leq (1-z_i)(\overline{x}_i - \underline{x}_i) \qquad i \in N \tag{9}$$

$$x^+ \geq 0, x^- \geq 0 \tag{10}$$

$$z \in \{0,1\}^n. \tag{11}$$

This requires introducing $2n$ additional continuous variables, n additional binary variables and $3n+1$ additional constraints.

2.1 Computing interval subgradient ranges

In the literature, the computation of an outer approximation of the interval vector D is proposed for the set $\mathbb{F}$ of *closed form representable differentiable functions*, whose elements can be written recursively in terms of arithmetic and algebraic operators of other functions in $\mathbb{F}$. That is, given constant and identity functions (variables) as "leafs", each element of $\mathbb{F}$ is associated to a syntactic tree whose inner nodes corresponds to differentiable n-ary functions $h : \mathbb{R}^n \to \mathbb{R}$ such that all partial derivatives are computable and the computation algorithm is provided explicitly [18]. Contracting leaf vertices with equal labels yields a Directed Acyclic Graph (DAG), and the gradient f' of a function $f \in \mathbb{F}$ can be constructed recursively by exploiting its DAG [24, 6, 18]. Enclosing approximations to the minimum and maximum values attained by $f'(x)$ whenever x ranges in X can be obtained using well-established techniques such Optimization-Based Bounds Tightening (OBBT) [18, 6, 9], which exploits a convex relaxation of f' constructed using the DAG representation, or Feasibility-Based Bounds Tightening (FBBT) [18, 6, 5], a forward-backward interval arithmetic recursive algorithm on the DAG of f'.

Similar techniques can be used to construct outer approximations of the Clarke subdifferential of nondifferentiable functions, thus extending $\mathbb{F}$ to a larger set functions. Indeed, for a univariate function $f : \mathbb{R} \to \mathbb{R}$ (with the properties assumed above) the set $\nabla f(x)$ is a real interval, so basically the same interval arithmetic techniques can be easily adapted. This allows to extend the treatment to several useful n-ary functions $h : \mathbb{R}^n \to \mathbb{R}$ that cannot be ordinarily dealt with, one of the most relevant being the "max" function (which, by standard techniques, implies other useful functions such as "min" and "$|\cdot|$"). Indeed, for $h(x) = \max(h_1(x), h_2(x))$ one has

$$\partial h(x) = \begin{cases} \partial h_1(x) & \text{if } h_1(x) > h_2(x) \\ \partial h_2(x) & \text{if } h_1(x) < h_2(x) \\ co(\{\partial h_1(x), \partial h_2(x)\}) & \text{if } h_1(x) = h_2(x) \end{cases}$$

[12, Proposition 2.3.12]. Therefore, interval analysis allows to derive an estimate over D for h given the ranges $X_1 \subseteq X$ and $X_2 \subseteq X$ such that $h_1(x) \geq h_2(x)$ and $h_1(x) \leq h_2(x)$, respectively, and estimates D_1 and D_2 for h_1 over X_1 and h_2 over D_2, respectively.

2.2 Example

We now show, by means of an example, that interval-subgradient cuts may be stronger than interval-gradient ones for equivalent smooth formulations, precisely because the ranges D for the former are tighter (smaller) than those for the latter. Consider the NLP formulation:

$$\min \; y \tag{12}$$
$$-y + x - 3 \leq 0 \tag{13}$$
$$-y - x - 3 \leq 0 \tag{14}$$
$$-y + \min(x(x-2), x(x+2)) \leq 0 \tag{15}$$
$$(x, y) \in [-2, 2] \times [-3, 0], \tag{16}$$

whose difficult part is the nonlinear, nonconvex and nondifferentiable constraint (15). A practical way to handle (12)–(16) is to drop (15) and solve the resulting LP relaxation; this yields $(x, y) = (0, -3)$, which is infeasible with respect to (15). We thus derive the interval-subgradient cut corresponding to (15) at $(0, -3)$. It is easy to verify that $D = [-2, 2] \times \{-1\}$, whence

$$\begin{aligned} (3+0) + \min_{d \in [-2,2]} d(x-0) + \min_{d \in [-1,-1]} d(y+3) = 3 - 2|x| - y - 3 &\leq 0 \Rightarrow \\ -y - 2|x| &\leq 0. \end{aligned} \tag{17}$$

By comparison, in order to derive an interval-gradient cut we reformulate the original problem to a differentiable MINLP as follows:

$$
\begin{aligned}
\min\ & y \\
& (13), (14), (16) \\
& -y + z(x(x-2)) + (1-z)(x(x+2)) \leq 0 \qquad (18) \\
& z \in \{0, 1\}
\end{aligned}
$$

Dropping (18), the obtained MILP has two equivalent optimal solutions $(x, y, z) = (0, -3, 0)$ and $(x, y, z) = (0, -3, 1)$, neither of which is feasible in the original MINLP. It is easy to verify that $D = [-6, 6] \times \{-1\} \times [-8, 8]$; thus, the interval-gradient cuts derived from (18) at $(0, -3, 0)$ and $(0, -3, 1)$ are, respectively,

$$
\begin{aligned}
3 + \min_{d \in [-6,6]} d(x-0) + \min_{d \in [-1,-1]} d(y+3) + \min_{d \in [-8,8]} d(z-0) &= -6|x| - y - 8z \leq 0 \\
3 + \min_{d \in [-6,6]} d(x-0) + \min_{d \in [-1,-1]} d(y+3) + \min_{d \in [-8,8]} d(z-1) &= -6|x| - y + 8(z-1) \leq 0 \quad .
\end{aligned}
$$

For the feasible values z can take in $\{0, 1\}$, these yield

$$
\begin{aligned}
-6|x| - y &\leq 0 \qquad (19) \\
-6|x| - y - 8 &\leq 0, \qquad (20)
\end{aligned}
$$

(20) being clearly weaker than (19) and therefore redundant. In turn, (19) is weaker than (17), despite the fact that both require 2 continuous variables, 1 binary variable and 7 constraints in order to be linearized. This shows that interval-subgradient cuts may prove to be stronger than interval-gradient ones.

3 No-good Cuts

A no-good cut is an inequality which cuts off a specific solution $\hat{x}$ of a problem P. One possible general formulation for this cut is

$$\|x - \hat{x}\| \geq \varepsilon, \qquad (21)$$

with $\varepsilon > 0$ chosen in such a way that no feasible solution of P lies in the ball of center $\hat{x}$ and radius ε. An appropriate ε ensuring that (21) does not cut off any other feasible point can only be found if $\hat{x}$ is an isolated point (in the topology induced by $\|\cdot\|$) of the feasible region of P.

An issue with constraint (21) is that it is nonconvex (reverse convex, more precisely). However, there are different ways to reformulate (21) as a linear constraint. In general they are quite inefficient, but for some special cases, like the (important) case in which $x \in \{0, 1\}^n$, (21) using the $\|\cdot\|_1$ norm becomes

$$\sum_{i \in N : \hat{x}_i = 0} x_i + \sum_{i \in N : \hat{x}_i = 1} (1 - x_i) \geq 1. \qquad (22)$$

We remark that this reformulation does not require additional variables or constraints. Defining the norm of constraint (21) as $\|\cdot\|_1$ and because $\hat{x}_i$ is a binary variable, $\|x_i - \hat{x}_i\| = x_i$ when $\hat{x}_i = 0$ and $\|x_i - \hat{x}_i\| = 1 - x_i$ when $\hat{x}_i = 1$, and we have, for $\varepsilon = 1$, inequality (22). Exploiting this idea one can generalize the no-good cut to continuous variables

$$\sum_{i \in N : \hat{x}_i = \underline{x}_i} (x_i - \underline{x}_i) + \sum_{i \in N : \hat{x}_i = \overline{x}_i} (\overline{x}_i - x_i) + \sum_{i \in N : \underline{x}_i < \hat{x}_i < \overline{x}_i} (x_i^+ + x_i^-) \geq \varepsilon \qquad (23)$$

(and to general integer variables by setting $\varepsilon = 1$) where, for all $i \in \hat{N} := \{\hat{i} \in N : \underline{x}_i < \hat{x}_i < \overline{x}_i\}$, we need the following additional constraints and variables:

$$\begin{aligned} x_i &= \hat{x}_i + x_i^+ - x_i^- && (24) \\ x_i^+ &\leq z_i(\overline{x}_i - \underline{x}_i) && (25) \\ x_i^- &\leq (1 - z_i)(\overline{x}_i - \underline{x}_i) && (26) \\ x_i^+ &\geq 0,\ x_i^- \geq 0 && (27) \\ z_i &\in \{0, 1\}. && (28) \end{aligned}$$

This leads to an inefficient way to handle no-good cuts, because $2|\hat{N}|$ additional continuous variables, $|\hat{N}|$ additional binary variables and $3|\hat{N}| + 1$ additional equations are needed. As will be pointed out in the next section, this MILP formulation of the no-good cut for general integer variables is the interval-subgradient cut of constraint (21) at $\hat{x}$ by using the $\|\cdot\|_1$ norm.

4 Interval-subgradient and No-good Cuts

In the following we prove that the interval-subgradient cut is a generalization of the no-good cut (23)-(28).

Theorem 4.1. *The no-good cut* (23)-(28) *can be derived by generating the linearization of the interval-subgradient cut* (6)-(11) *from constraint* (21) *using* $\|\cdot\|_1$.

Proof. Let us consider the nonconvex inequality (21) with $\|\cdot\|$ being $\|\cdot\|_1$. We try to generate an interval-subgradient cut with respect to point $\hat{x}$. Since $g(\hat{x}) = 0$, we have

$$\underline{g}(x) = \min_{d \in D} d(x - \hat{x}) = \min_{d \in [-e, e]} d(x - \hat{x}) \qquad (29)$$

with $e = (1, 1, ..., 1)$ because the subgradient of $|x_i - \hat{x}_i|$ stays in the range $[-1, 1]$ $\forall i \in N$. This can be rewritten as

$$\underline{g}(x) = \sum_{i \in N} \min_{d_i \in [-1,1]} d_i(x_i - \hat{x}_i) = \sum_{i \in N} \min((x_i - \hat{x}_i), -(x_i - \hat{x}_i)) =$$
$$\sum_{i \in N} - \max(-(x_i - \hat{x}_i), (x_i - \hat{x}_i)) = - \sum_{i \in N} |x_i - \hat{x}_i| \qquad (30)$$

whence

$$- \sum_{i \in N} |x_i - \hat{x}_i| \leq -\varepsilon \qquad (31)$$

is our interval-subgradient cut which is equivalent to (21), thus can be linearized with (23)-(28).

□

No-good cuts have been extensively used both in MILP and Constraint Programming in a number of sophisticated algorithmic frameworks. For example, they have been used in [13] to tighten linear relaxations of MILPs involving logical implications modeled through big-M coefficients, in [16] with the name of "conflicts" to guide the search and for propagation in [17]. The fact of no-good cuts are in turn a special case of interval-subgradient cuts could lead to the extension of some of the above techniques to the MINLP context.

5 No-good Cuts of p-norms

We now extend the previous treatment to the general case of p-norms

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

where $1 \le p < \infty$. It is well-known that p-norms are convex and non-increasing in p, i.e., $\|\cdot\|_q \le \|\cdot\|_p$ for $p < q$. Of course, the most common case is the standard Euclidean norm $p = 2$. It is also well-known that one can also take $p \to \infty$, resulting in the ∞-norm (or Tchebycheff norm)

$$\|x\|_\infty = \max \left\{ \, |x_i| \; : \; i = 1, \ldots, n \, \right\} .$$

Since balls in the q-norm are larger than balls in the p-norm when $q > p$, the generic no-good constraint in the p-norm:

$$\|x - \hat{x}\|_p \ge \varepsilon \tag{32}$$

(which requires to be outside one such ball) gets stronger as p increases. In other words, the constraint in the 1-norm of the previous sections is the weakest possible. Therefore, assuming one derives a valid no-good constraint for some $p > 1$, it might be reasonable to derive the corresponding interval-subgradient cut, in the hope that it also turns out to be stronger. We now prove that this is not the case.

Theorem 5.1. *The linearization of the interval-subgradient cut derived from the no-good cut* (32) *for any $p > 1$ is equivalent to the one derived from the no-good cut in the 1-norm.*

Proof. We start evaluating the interval-subgradient of the p-norm. From ordinary chain rules of derivation for $\|x\|_p = (\sum_{i=1}^n f(x_i)^p)^{1/p}$ with $f(z) = |z|$, one has that in all points where $\|\cdot\|_p$ is differentiable (that is, none of the x_i is null) the i-th component of the gradient is

$$\frac{f'(x_i) f(x_i)^{p-1}}{\left(\sum_{i=1}^n f(x_i)^p \right)^{(p-1)/p}} \;=\; \frac{\operatorname{sign}(x_i) |x_i|^{p-1}}{\left(\sum_{i=1}^n |x_i|^p \right)^{(p-1)/p}} \;\; . \tag{33}$$

Now, by [23, Theorem 25.6] the subdifferential of any convex function at $\bar{x}$ is the closed convex hull of all vectors g that are limits of sequences of gradients at $\bar{x}^i$ for all possible sequences $\{\bar{x}^i\} \to \bar{x}$ such that the function is differentiable at each $\bar{x}^i$ (plus the normal cone of the domain of at $\bar{x}$, which is $\{0\}$ here since the domain of $\|\cdot\|_p$ is the whole of $\mathbb{R}^n$). Therefore, $\partial\|x\|_p$ for $x \ne 0$ is the set of all vectors of the form (33), provided that one interprets $\operatorname{sign}(x_i)$ as $\partial|x_i|$ (that is, $\operatorname{sign}(0) = [-1, 1]$). Hence, $\partial\|x\|_p \subseteq [-e, e]$, as in (33) the absolute value of the numerator is always smaller than the denominator. The interval-subgradient D cannot be made smaller, as can be clearly seen by considering all the points of the form αe_i, where the ratio evaluates to $\operatorname{sign}(\alpha)$ (with e_i being the i-th component of the canonical basis of $\mathbb{R}^n$). Hence, D contains $[-e, e]$, and since $\partial\|0\|_p \subseteq [-e, e]$ as well for the above-mentioned property, $D = [-e, e]$. The case of $p = \infty$ is even more obvious, although the result has to be obtained along different lines, using rules for the subdifferential of the maximum of convex functions. However, it is well-known [23, comments to Theorem 23.1] that

$$\partial\|x\|_\infty \;=\; \operatorname{conv}\left(\, \operatorname{sign}(x_i) e_i \; : \; i \in I_x \, \right)$$

where $I_x = \{ \, i \; : \; |x_i| = \|x\|_\infty \, \}$, and again $\partial\|0\|_\infty = [-e, e]$. It is therefore clear that $D = [-e, e]$ as well.

This implies that, deriving the interval-subgradient cut from the general no-good cut in the p-norm, gives:

$$\underline{g}(x) := \|\hat{x}\|_p + \min_{d \in [-e, e]} d(x - \hat{x}) := \min_{d \in [-e, e]} d(x - \hat{x}) \tag{34}$$

for any $p > 1$. The result follows by comparing (34) and the interval-subgradient cut obtained using the no-good cut in the 1-norm (29) of Section 4. □

In the example of Sect. 2.2, adjoining a no-good cut to make $(x, y) = (0, -3)$ infeasible would be less effective than the use of interval-gradient/subgradient cuts. Since the variables involved in the formulation are continuous ϵ is small. Thus, the proportion of relaxed feasible region excluded by the resulting no-good cut would be rather small.

6 Conclusions

In this paper we presented a generalization of interval-gradient cuts to the case of nondifferentiable functions, which we called interval-subgradient cuts. We showed that no-good cuts are a special case of interval-gradient cuts when they are generated from the 1-norm function. Finally, we have shown that writing the linearized version of the interval-subgradient cut associated with a no-good cut with p-norm for any $p > 1$ does not help in making the cut stronger than that with the 1-norm.

Acknowledgements

We are grateful to Giancarlo Bigi for useful discussions. The third author gratefully acknowledges financial support from the following grants: ANR 07-JCJC-0151 "ARS", Digiteo Chair 2009-14D "RMNCCO", Digiteo Emergence 2009-55D "ARM". The first and the last author are partially supported by Project DecisOpElet 2006 of Università di Bologna which is kindly acknowledged. Finally, we like to thank the Associate Editor for very useful comments that led to an improved version of the paper.

References

[1] F.A. Al-Khayyal and H.D. Sherali. On finitely terminating branch-and-bound algorithms for some global optimization problems. *SIAM Journal of Optimization*, 10:1049–1057, 2000.

[2] A. Atamturk and V. Narayanan. Conic mixed-integer rounding cuts. *Mathematical Programming*, 122:1–20, 2010.

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

[4] E. Balas and R. Jeroslow. Canonical cuts on the unit hypercube. *SIAM Journal on Applied Mathematics*, 23:61–69, 1972.

[5] P. Belotti, S. Cafieri, J. Lee, and L. Liberti. On the convergence of feasibility based bounds tightening. In U. Faigle, editor, *Proceedings of the* 8^{th} *Cologne-Twente Workshop on Graphs and Combinatorial Optimization*, Köln, 2010. Köln University.

[6] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter. Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods and Software*, 24(4):597–634, 2009.

[7] M.S. Boddy and D.P. Johnson. A new method for the global solution of large systems of continuous constraints. In Ch. Bliek, Ch. Jermann, and A. Neumaier, editors, *Global Optimization and Constraint Satisfaction*, volume 2861 of *Lecture Notes in Computer Science*, pages 142–156. Springer, Berlin, 2003.

[8] M. Bruglieri and L. Liberti. Optimal running and planning of a biomass-based energy production process. *Energy Policy*, 36:2430–2438, 2008.

[9] A. Caprara and M. Locatelli. Global optimization problems and domain reduction strategies. *Mathematical Programming*, to appear.

[10] M.T. Çezik and G. Iyengar. Cuts for mixed 0-1 conic programming. *Mathematical Programming*, 104:179–202, 2005.

[11] S. Ceria and J. Soares. Convex programming for disjunctive convex optimization. *Mathematical Programming*, 86:595–614, 1999.

[12] F.H. Clarke. *Optimization and Nonsmooth Analysis*. Classics in Applied Mathematics 5. SIAM, Philadelphia, 1990.

[13] G. Codato and M. Fischetti. Combinatorial Benders' cuts. *Operations Research*, 54:756–766, 2006.

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

[15] A. Frangioni and C. Gentile. A Computational Comparison of Reformulations of the Perspective Relaxation: SOCP vs. Cutting Planes. *Operations Research Letters*, 37:206–210, 2009.

[16] F. Kılınç Karzan, G.L. Nemhauser, and M.W.P. Savelsbergh. Information-based branching schemes for binary linear mixed-integer programs. *Mathematical Programming Computation*, 1:249–293, 2009.

[17] H.-J. Kim and J.N. Hooker. Solving fixed-charge network flow problems with a hybrid optimization and constraint programming approach. *Annals of Operations Research*, 115:95–124, 2002.

[18] L. Liberti. Writing global optimization software. In L. Liberti and N. Maculan, editors, *Global Optimization: from Theory to Implementation*, pages 211–262. Springer, Berlin, 2006.

[19] L. Liberti and C.C. Pantelides. An exact reformulation algorithm for large nonconvex NLPs involving bilinear terms. *Journal of Global Optimization*, 36:161–189, 2006.

[20] G. Nannicini and P. Belotti. Rounding-based heuristics for nonconvex minlps with binary variables. Technical report, Working paper, 2009.

[21] I. Nowak. *Relaxation and Decomposition Methods for Mixed Integer Nonlinear Programming.* International Series of Numerical Mathematics, Birkhäuser Verlag, 2005.

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

[23] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[24] H. Schichl and A. Neumaier. Interval analysis on directed acyclic graphs for global optimization. *Journal of Global Optimization*, 33(4):541–562, 2005.

[25] H. Sherali and W. Adams. A reformulation-linearization technique (rlt) for semi-infinite and convex programs under mixed 0-1 and general discrete restrictions. *Discrete Applied Mathematics*, 157:1319–1333, 2009.

[26] H. Sherali and A. Alameddine. A new reformulation-linearization technique for bilinear programming problems. *Journal of Global Optimization*, 2:379–410, 1992.

[27] M. Tawarmalani and N.V. Sahinidis. Semidefinite relaxations of fractional programs via novel convexifications techniques. *Journal of Global Optimization*, 20:137–158, 2001.

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