Beyond Canonical Difference-of-Convex Programs: the Single Reverse Polar Problem

Giancarlo Bigi[1] Antonio Frangioni[2] Qinghua Zhang[3]

Abstract

We propose a novel generalization of the Canonical DC problem (CDC), and we study the convergence of outer approximation algorithms for its solution, which use an approximated oracle for checking the global optimality conditions. Although the approximated optimality conditions are similar to those of CDC, this new class of problems is shown to significantly differ from its special case. Indeed, outer approximation approaches for CDC need be substantially modified in order to cope with the more general problem, bringing to new algorithms. We develop a hierarchy of conditions that guarantee global convergence, and we build three different cutting plane algorithms relying on them.

Keywords: *Single reverse polar problems, approximate optimality conditions, cutting plane algorithms*

AMS Subject Classification: *90C26, 90C46, 90C33*

[1]Dipartimento di Informatica, Università di Pisa, Largo B.Pontecorvo, 3, 56127 Pisa, Italia.
email: giancarlo.bigi@di.unipi.it – corresponding author

[2]Dipartimento di Informatica, Università di Pisa, Largo B.Pontecorvo, 3, 56127 Pisa, Italia.
email: frangio@di.unipi.it

[3]Wuhan University, School of Mathematics and Statistics, Wuchang, Luojia Hill, 430072 Wuhan, China.
email: Qinghuazhang@whu.edu.cn

1 Introduction

DC optimization problems are those nonconvex programs in which the objective function is the difference of two convex functions and the constraint is given by the set difference of two convex sets. It is well-known that all of them can be transformed into the so-called Canonical DC problem [1,2]. Since a large number of nonconvex optimization problems can be reduced to DC optimization problems [3–12], the Canonical DC problem (CDC, shortly) has a wide set of applications.

In this paper, we introduce the Single Reverse Polar problem (SRP, shortly) as a generalization of CDC. In effect, SRP allows to directly address a host of different problems of practical interest other than all the applications that can be formulated as DC programs (see, for instance, [2,14,15]). Among them, Separable Linear Complementarity Programs (SLCP, shortly) have many practical applications. For instance, the l_0-norm could be easily reformulated as a linear complementarity form [16], thus convex programs with an additional l_0 constraint [17–19], such as compressed sensing problems, can be easily formulated as SLCPs. Many other structures, such as value-at-risk minimization problem in portfolio selection, also admit a linear complementarity reformulation (see, for instance, [16, 20, 21] and the references therein). Furthermore, *separable* bilevel problems, where leader and follower variables are only "tied" in the objective function, are also easily reduced to SLCPs; in turn, these can be used to reformulate several families of Mixed-Integer (Non)Linear Programs along the lines of [22].

We extend the outer approximation algorithms for CDC of [3] to the new class of problems, as this approach should have similar cost per iteration for SRP and CDC. In fact, these algorithms are based on an *approximated oracle* for checking the global optimality conditions, which is the most computationally demanding part of the approach. As shown in Section 2, the optimality conditions for SRP are a minimal modification of those for CDC. Both entail the solution of an optimization problem with a non-convex objective function and a convex feasible region. The "difficult" part (the objective function) of both is the same, while the "easy" part (the feasible set) is very similar; hence, it is likely that the difference does not substantially impact the practical cost of solving the problems. Furthermore, allowing not to solve them up to proven global optimality can substantially improve the efficiency of the overall cutting-plane method. Yet, this also requires to properly characterize the impact of approximations in the oracle on the quality of the obtained solution. After that is done, one can devise different ways to exploit the information produced by the oracle to construct globally convergent algorithms. This analysis gives rise to three different implementable algorithms for SRP. Our analysis shows that, despite the similarities, SRP have markedly different properties than CDC, which substantially impact on the way in which algorithms can be constructed, thereby shedding some new light on the algorithms for the original problem too.

The paper is organized as follows. In Section 2 we describe and analyze the main properties of SRP and contrast them with those of CDC. Then, in Section 3 we extend our approximate optimality conditions for CDC [3] to the new problem. In Section 4, we develop a hierarchy of conditions that guarantee the convergence of cutting plane algorithms; relying on these conditions, in Section 5 we build three cutting plane algorithms for solving SRP, and we discuss possible strategies to enhance their effectiveness in practice. Finally, Section 6 draws some conclusions.

2 The Single Reverse Polar Problem

In this paper, we study the following type of global optimization problem

$$(SRP) \qquad \min\{f(x) + g(w) \ \mid \ x \in \Omega,\ w \in \Gamma,\ wx \geq \alpha\},$$

where $f : \mathbb{R}^n \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ are convex extended-valued functions, $\Omega, \Gamma \subset \mathbb{R}^n$ are full-dimensional compact and convex sets, and wx denotes the scalar product of the vectors w and x. Without any loss of generality, we suppose $\Omega \subseteq \operatorname{dom} f$ and $\Gamma \subseteq \operatorname{dom} g$. Up to scaling, one can always assume $\alpha \in \{-1, 0, 1\}$.

Problem (SRP) arises as a generalization of the canonical DC problem

$$(CDC) \qquad \min\{dx \ \mid \ x \in \Omega \setminus \operatorname{int}\ \Gamma^*\},$$

where $d \in \mathbb{R}^n$ and $\Gamma^* := \{y \in \mathbb{R}^n \ \mid \ yw \leq 1, \ \ \forall w \in \Gamma\}$ is the polar set of Γ. In turn, (CDC) can be rewritten as

$$\min\{dx \ \mid \ x \in \Omega,\ w \in \Gamma,\ wx \geq 1\}, \tag{1}$$

i.e., a convex program with a *nonconvex constraint* [3, 13]. Thus, (CDC) can be considered a special case of (SRP) with $\alpha = 1$, $f(x) = dx$ and $g(w) = 0$. This is why we consider the single complicating constraint $wx \geq \alpha$ in (SRP) a *reverse polar* constraint, and we name the problem as the *Single Reverse Polar* problem.

In order to avoid that (SRP) could be reduced to a convex minimization problem, we also suppose that the set

$$A_\alpha = \{(x, w) \in \mathbb{R}^n \times \mathbb{R}^n \ \mid \ wx \geq \alpha\}$$

provides an essential constraint, i.e.,

$$\min\{f(x) + g(w) \ \mid \ (x, w) \in \Omega \times \Gamma\} < \gamma^* \tag{2}$$

where

$$\gamma^* = \min\{f(x) + g(w) \ \mid \ (x,w) \in (\Omega \times \Gamma) \cap A_\alpha\}.$$

Assumption (2) can be equivalently stated as the existence of a "low point" $(\bar{x}, \bar{w}) \in \Omega \times \Gamma$ satisfying

$$f(x) + g(w) > f(\bar{x}) + g(\bar{w}) = \bar{\gamma} \qquad \forall (x,w) \in (\Omega \times \Gamma) \cap A_\alpha. \tag{3}$$

Clearly, any pair minimizing $f+g$ over $\Omega \times \Gamma$ provides such a point if (2) holds. Since the interiors of Ω and Γ are not empty, then there exists also some $(\bar{x}, \bar{w}) \in \operatorname{int} \Omega \times \operatorname{int} \Gamma$ satisfying (3). Therefore, we suppose also this further condition throughout all the paper. These assumptions guarantee that any feasible solution $(x,w) \in (\Omega \times \Gamma) \cap \operatorname{int} A_\alpha$ provides a better feasible solution taking the unique intersection between the segment with $(\bar{x}, \bar{w})$ and (x,w) as end points and the boundary of A_α, i.e., $(x', w') \in \operatorname{conv}(\{(\bar{x}, \bar{w}), (x,w)\}) \cap \partial A_\alpha$ satisfies $f(x') + g(w') < f(x) + g(w)$. As a consequence, the nonconvex constraint $wx \geq \alpha$ is active at every optimal solution (x,w) of (SRP), i.e.,

$$wx = \alpha. \tag{4}$$

Therefore, one can equivalently assume that the reverse polar constraint is an *equality* constraint. This allows to formulate separable programs with linear complementarity constraints in the (SRP) format. In fact, Separable Linear Complementarity Programs of the form

$$(SLCP) \qquad \min\{f(x) + h(y) \ \mid \ (x,y) \in \Omega \times Y, \ (q + My)x = 0\},$$

where $\Omega, Y \subset \mathbb{R}^n$ are full-dimensional compact convex sets and $M \in \mathbb{R}^{n \times n}$ is full-rank, can be easily reduced to (SRP) by defining $\Gamma = \{q + My \ \mid \ y \in Y\}$ (which is full-dimensional since Y is and M is full-rank) and $g(w) = \min\{h(y) \ \mid \ y \in Y, \ w = q + My\}$ (which is convex and extended-valued). Thus, the complementary constraint can be brought in the reverse polar form (4).

Notice that the boundedness of Ω and Γ guarantees the existence of an optimal solution and therefore, due to (3), that $\gamma^* > \bar{\gamma}$ always hold: this property will be useful later on. Although optimal solutions must lie in the boundary of A_α, they don't necessarily belong to the boundary of $\Omega \times \Gamma$, as the following example shows.

Example 2.1 Consider (SRP) with $n = 2$, $\alpha = 3$,

$$\Omega = \{x \in \mathbb{R}^2 \ \mid \ (x_1 - 1)^2 + (x_2 - 1)^2 \leq 2\}, \quad f(x) = 3(x_1 - 1)^2 + 2(x_2 - 1)^2,$$

$$\Gamma = \{w \in \mathbb{R}^2 \ \mid \ (w_1 - 1)^2 + (w_2 - 1)^2 \leq 2\}, \quad g(w) = 2(w_1 - 1)^2 + 5(w_2 - 1)^2.$$

The point $(\bar{x}, \bar{w}) \in \operatorname{int}\Omega \times \operatorname{int}\Gamma$ with $\bar{x} = \bar{w} = (1,1)$ satisfies (3) and therefore $\bar{\gamma} = 0$. All the other points $(x, w) \in \Omega \times \Gamma$ satisfy $f(x) + g(w) > 0$, thus $\gamma^* > \bar{\gamma}$ and condition (2) holds. If $x \in \partial\Omega$, then $f(x) = 2\big((x_1-1)^2 + (x_2-1)^2\big) + (x_1-1)^2 = 4 + (x_1-1)^2 \geq 4$. Similarly, if $w \in \partial\Gamma$, then $g(w) \geq 4$. Therefore, the objective value of all the feasible points $(x, w) \notin \operatorname{int}\Omega \times \operatorname{int}\Gamma$ is not less than 4. However, the point (x, w) with $x = w = (1.5, 1.5)$ is feasible and $f(x) + g(w) = 3$. Therefore, the set of all the optimal solutions is contained into $\operatorname{int}\Omega \times \operatorname{int}\Gamma$.

Example 2.1 relies on the strict convexity of f and g. Indeed, when the objective function of (SRP) is linear, then there has to be some optimal solution lying in the boundary of $\Omega \times \Gamma$.

Theorem 2.1 *Suppose that f and g are linear, i.e., $f(x) = dx$ and $g(w) = ew$ for some $d, e \in \mathbb{R}^n$. If $n \geq 2$, then at least one optimal solution of (SRP) belongs to $\partial\Omega \times \partial\Gamma$.*

Proof. Take any optimal solution (x^*, w^*). Suppose $x^* \in \operatorname{int}\Omega$: the compactness of Ω implies $\{x \in \mathbb{R}^n \mid w^*x = \alpha\} \cap \partial\Omega \neq \emptyset$. Take two points x^1 and x^2 in the above intersection such that $x^* \in [x^1, x^2]$: we have $\min\{dx^1, dx^2\} \leq dx^*$, so either (x^1, w^*) or (x^2, w^*) is optimal. The thesis can be proved analogously in case $w^* \in \operatorname{int}\Gamma$. □

The level set

$$R(\gamma) := \{(x, w) \in \Omega \times \Gamma \mid f(x) + g(w) \leq \gamma\},$$

which is bounded due to the compactness assumptions on Ω and Γ, is helpful to check whether or not a feasible value, i.e., a value $\gamma \geq \gamma^*$ is optimal. In fact, it is straightforward that $\gamma = \gamma^*$ implies the inclusion

$$R(\gamma) \cap \operatorname{int} A_\alpha = \emptyset. \tag{5}$$

The optimality condition (5) will be proved in Section 3.2 in a more general form, and it is equivalent to

$$v(OC_\gamma) = \max\{vz - \alpha \mid (z, v) \in R(\gamma)\} \leq 0. \tag{6}$$

This is analogous to the "optimization form" of the optimality condition

$$\{z \in \mathbb{R}^n \mid z \in \Omega,\ v \in \Gamma,\ dz \leq \gamma\} \subseteq \Gamma^*$$

of the canonical DC problem, which becomes sufficient under mild assumptions (see [2,3]). Similarly, the necessary optimality condition (5) is also sufficient when (SRP) is *regular*, i.e.,

$$\gamma^* = \inf\{f(x) + g(w) \mid (x, w) \in (\Omega \times \Gamma) \cap \operatorname{int} A_\alpha\}. \tag{7}$$

Furthermore, regularity will be exploited to relate approximations in the stopping criteria with the quality of the corresponding approximate optimal solutions, as discussed in the following section.

3 Approximate Optimality Conditions

In this section, we study the global optimality conditions (5) and (6) for (SRP), introducing their approximated forms and comparing the results with those available for the canonical DC problem.

3.1 Optimality Conditions and (Approximate) Oracles

As in the (CDC) case, we plan to approach the "geometrical" optimality condition (5) via its optimization counterpart (6). Since the latter's objective function is *not* concave, the problem is a difficult one. Yet, the advantage of employing it is that one can at least easily define a computationally relevant concept of *approximate* optimality conditions.

The first step towards this aim is to consider the relaxation of (6)

$$v(\overline{OC}_\gamma) \;=\; \max\{vz - \alpha \;\mid\; z \in S,\; v \in Q,\; f(z) + g(v) \leq \gamma\} \tag{8}$$

where Ω and Γ are replaced by two convex sets S and Q, respectively, satisfying

$$\Omega \subseteq S, \quad \Gamma \subseteq Q. \tag{9}$$

The idea is to start with some "rough estimate" of the original sets, e.g. where S and Q are polyhedra with few vertices or facets, and iteratively refine it as needed. Any choice of S and Q satisfying (9) ensures that $v(\overline{OC}_\gamma) \geq v(OC_\gamma)$ (cf. (6)); thus, the inequality $v(\overline{OC}_\gamma) \leq 0$ provides a convenient *sufficient* optimality condition for (SRP). Assuming that (8) is significantly easier to solve than (6), one can then devise iterative schemes that check that condition, and then either discover that S and Q are not appropriate approximations of (respectively) Ω an Γ, and improve them, or find that γ is not the optimal value and improve it. This is what has been done in [3] for (CDC), and is repeated here for (SRP). Our analysis will show that the different structural characteristics of the latter problem over the former imply that the strategies by which the information produced by an "oracle" for (8) can be used to devise a convergent algorithm for (SRP) are rather different from those that work for (CDC). However, what remains very similar is the actual form of the oracle, and therefore the appropriate notion of "approximate solution" that can be employed. Indeed, in the following we will make the following assumptions about the approximate oracle for (8), which are taken almost verbatim from [3]: a procedure Θ is available which, given S, Q, γ, and two positive tolerances ε and ε',

- *either* produces an upper bound

$$\varepsilon v(\overline{OC}_\gamma) \leq l \qquad \text{such that} \qquad l \leq \varepsilon' \tag{10}$$

- *or* produces a pair

$$(\bar{z}, \bar{v}) \text{ feasible for (8)} \quad \text{such that} \quad \bar{v}\bar{z} - \alpha \geq \varepsilon v(\overline{OC}_\gamma) > \varepsilon' . \tag{11}$$

This clearly is a pretty weak requirement about the way in which (8) need to be solved; that is, one has *either* to compute a feasible solution $(\bar{z}, \bar{v})$ that is "sufficiently close" to the optimal one, or prove that any such solution is uninteresting by computing an upper bound $l \geq v(\overline{OC}_\gamma)$ that demonstrates that the optimal value is "small". Algorithmically, the two parts of (10)—(11) usually even correspond to two entirely different classes of approaches: feasible solutions are produced by heuristics, while upper bounds are produced by solving suitable relaxations. We direct the interested reader to the discussion in [3, 13], which applies with relatively little changes to the current environment.

What does need a specific discussion, instead, is the fact that condition (10) requires the lower bound to be "small enough", but allows it to be positive. This means that when (10) holds, and the algorithm is (as we shall see) stopped, the obtained solution is not guaranteed to be optimal. Thus, as in [3] a study of the relationships between the tolerances ε and ε' and the quality of the obtained solution (actually, of the current upper bound γ on the true optimal value γ^*) is required. This turns out to be significantly different from the canonical DC case, as the next Section will show.

3.2 Approximate Optimality Conditions

We are interested in characterizing the values γ for which the stopping criterion (10) holds, which is better analyzed when rewritten as $v(OC_\gamma) \leq \delta$ in terms of the single parameter $\delta = \varepsilon'/\varepsilon$. These values are strictly related to the approximated problem

$$(SRP_\delta) \qquad \min\{ f(x) + g(w) \;\mid\; (x, w) \in (\Omega \times \Gamma) \cap A_{\alpha+\delta} \}$$

obtained by perturbing the right-hand side of the reverse-polar constraint in (SRP) by the critical parameter δ. In particular, it is convenient to consider the value function

$$\phi(\delta) := \inf\{ f(x) + g(w) \;\mid\; (x, w) \in (\Omega \times \Gamma) \cap \operatorname{int} A_{\alpha+\delta} \}, \tag{12}$$

which involves a further restriction of the feasible region to the interior of the nonconvex set $A_{\alpha+\delta}$. Clearly, $\phi(\delta)$ can be greater than the optimal value of (SRP_δ) as no regularity assumption is required in our analysis.

Proposition 3.1 *For any $\delta \geq 0$, the following statements are equivalent:*

(i) $v(OC_\gamma) \leq \delta$*;*

(ii) $R(\gamma) \cap \operatorname{int} A_{\alpha+\delta} = \emptyset$*;*

(iii) $\gamma \leq \phi(\delta)$*.*

Proof. The equivalence between (i) and (ii) readily follows from the definition of $v(OC_\gamma)$. Analogously, (ii) implies (iii) by the definition of $\phi(\gamma)$. We prove that (iii) implies (ii) by contradiction: if (ii) does not hold, there exists some $(x, w) \in R(\gamma) \cap \operatorname{int} A_{\alpha+\delta}$. Take any (x^1, w^1) in the intersection between $\operatorname{int} A_{\alpha+\delta}$ and the open line segment with (x, w) and $(\bar{x}, \bar{w})$ as its end points. Since $(x^1, w^1) \in \Omega \times \Gamma$, we get $\phi(\delta) \leq f(x^1) + g(w^1) < f(x) + g(w) \leq \gamma$ in contradiction with (iii). □

As an immediate consequence of the proposition, we also have

$$\phi(\delta) = \sup\{\gamma \mid R(\gamma) \cap \operatorname{int} A_{\alpha+\delta} = \emptyset\}.$$

Considering the optimal value of (SRP_δ) as γ in Proposition 3.1, we get that (ii) is a necessary optimality condition for (SRP_δ). Furthermore, the condition is also sufficient if (SRP_δ) is *regular*, i.e., $\phi(\delta)$ is actually the optimal value. When $\delta = 0$, (SRP_δ) coincides with (SRP) and therefore Proposition 3.1 provides optimality conditions for (SRP), too. In particular, (ii) provides the necessary condition (5) and (i) its equivalent "optimization form" $v(OC_\gamma) \leq 0$, while regularity, if it holds, guarantees that they are also sufficient. Therefore, inclusion (ii) can be considered as an approximate optimality condition for (SRP), and condition (iii) provides the adequate tool to evaluate the quality of the approximation. In fact, if (SRP) is regular, i.e., $\phi(0) = \gamma^*$,then

$$0 \leq \gamma - \gamma^* \leq \phi(\delta) - \gamma^* = \phi(\delta) - \phi(0)$$

holds for any feasible value γ which satisfies (i). The following result guarantees that the approximation approaches the optimal value as δ goes to 0.

Proposition 3.2 *The value function ϕ is right-continuous at 0, i.e.* $\lim_{\delta \downarrow 0} \phi(\delta) = \phi(0)$.

Proof. Clearly ϕ is nonincreasing, that is $\phi(\delta^1) \geq \phi(\delta^2)$ whenever $\delta^1 \geq \delta^2 \geq 0$. As it is also bounded below by $\phi(0)$, there exists $\tilde{\gamma} = \lim_{\delta \downarrow 0} \phi(\delta)$ and $\tilde{\gamma} \geq \phi(0)$. Since $\tilde{\gamma} \leq \phi(\delta)$ for any $\delta > 0$, Proposition 3.1 implies $v(OC_{\tilde{\gamma}}) \leq \delta$ for any $\delta > 0$. Since $v(OC_{\tilde{\gamma}})$ does not depend upon δ, we get $v(OC_{\tilde{\gamma}}) \leq 0$. Therefore, Proposition 3.1 guarantees $\tilde{\gamma} \leq \phi(0)$. □

Although the approximation always converges to the optimal value, the rate of convergence may be less than linear. Indeed, the following example shows that not even regularity is enough to achieve a linear rate of convergence.

Example 3.1 Consider (SRP) with $n = 2$, $\alpha = 1$,

$$\Omega = \{(x_1, x_2) \in \mathbb{R}^2 \times \mathbb{R}^2 \ \mid \ x_1^2 + (x_2 - 1)^2 \leq 4\}, \quad f(x) = x_2,$$

$$\Gamma = \text{conv}\,\{(0,1), (0,-1/2), (1/2,0), (-1/2,0)\}, \quad g(w) = w_1.$$

A standard computation shows $\Gamma^* = \{(-2,-2), (2,-2), (-2,1), (2,1)\}$. Since any feasible solution (x, w) must satisfy $x \notin \text{int}\,\Gamma^*$, which implies $x_2 \geq 1$, then $x^* = (-2, 1)$ and $w^* = (-1/2, 0)$ provide the unique optimal solution and the optimal value is $\gamma^* = 1/2$. Therefore, the point $(0_2, 0_2) \in \text{int}\,\Omega \times \text{int}\,\Gamma$ satisfies (3) and thus the problem satisfies assumption (2). The problem is regular. In fact, choosing $x(\delta) = (-\sqrt{4-\delta^2}, 1+\delta)$ and $w(\delta) = (-0.5+\delta, 2\delta)$ with $0 \leq \delta \leq 1/2$, we have $(x(\delta), w(\delta)) \to (x^*, w^*)$ as $\delta \downarrow 0$ $(x(\delta), w(\delta)) \in \Omega \times \Gamma$, $w(\delta)x(\delta) > 1$ if $\delta > 0$ and moreover

$$\lim_{\delta \downarrow 0} \ f(x(\delta)) + g(w(\delta)) = \lim_{\delta \downarrow 0} \ 1/2 + 2\delta = 1/2 = \gamma^*.$$

Given any $\delta > 0$, any feasible solution (x, w) to problem (SRP_δ) satisfies $wx \geq 1 + \delta$ and therefore $x_2 \geq 1 + \delta$. Since $(\hat{x}, \hat{w})$ with $\hat{x} = (-|x_1|, x_2)$ and $\hat{w} = (-|w_1|, |w_2|)$ is also feasible and $f(x) + g(w) \geq f(\hat{x}) + g(\hat{x})$, then any optimal solution (x, w) to (SRP_δ) satisfies $x_1 \leq 0$, $w_2 \geq 0$ and $-1/2 \leq w_1 \leq 0$. By Theorem 2.1 there exists an optimal solution that satisfies also $x_1^2 + (x_2 - 1)^2 = 4$, and $w_2 = 2w_1 + 1$. Taking into account (4), we get $w_1 = (1 + \delta - x_2)/(x_1 + 2x_2)$. Therefore, we have

$$\begin{aligned}
\phi(\delta) &\geq \min\{x_2 + \frac{1+\delta-x_2}{x_1+2x_2} \ \mid \ 1+\delta \leq x_2 \leq 3, \ x_1^2 + (x_2-1)^2 \leq 4\} \\
&\geq \min\{x_2 + \frac{1+\delta-x_2}{2x_2-2} \ \mid \ 1+\delta \leq x_2 \leq 3\} \\
&\geq \min\{\ 1 + \tau + \frac{\delta - \tau}{2\tau} \mid \delta \leq \tau \leq 2\} \\
&\geq \ 1 + \sqrt{\delta/2} + \frac{\delta - \sqrt{\delta/2}}{2\sqrt{\delta/2}} \ = \ 1/2 + \sqrt{2\delta}.
\end{aligned}$$

Thus, the rate of convergence is not linear, i.e., the value function ϕ is not locally Lipschitz at 0, since

$$\lim_{\delta\to 0}\frac{\phi(\delta)-\phi(0)}{\delta}\geq \lim_{\delta\to 0}\frac{1/2+\sqrt{2\delta}-\gamma^*}{\delta}=\lim_{\delta\to 0}\frac{\sqrt{2\delta}}{\delta}=+\infty.$$

In the canonical DC problem both regularity and the locally Lipschitz continuity of the value function ϕ at 0 are guaranteed by the existence of an optimal solution x^* such that the following relationship

$$T(\Omega,x^*)\not\subseteq T(\Gamma^*,x^*) \tag{13}$$

holds between the (Bouligand) tangent cones of Ω and Γ at x^* [23, Theorem 3.7]. Besides, condition (13) is actually equivalent to regularity if Γ^* (or equivalently Γ) is a polyhedron [23, Theorem 3.8]. These relevant properties are lost in (SRP). In fact, Example 3.1 provides a case in which ϕ is not locally Lipschitz at 0 although (SRP) is regular and (13) holds. The following example provides a case in which the problem is not regular while (13) holds, though both Ω and Γ are polyhedra.

Example 3.2 Consider (SRP) with the same data of Example 3.1 except for

$$\Omega=\text{conv}\,\{(0,3),(0,-1),(2,1),(-2,1)\}.$$

A standard computation shows $\Omega^*=\{(-1,-1),(1,-1),(-1/3,1/3),(1/3,1/3)\}$. Since the feasible region is included in the one of the previous example and the point (x^*,w^*) with $x^*=(-2,1)$ and $w^*=(-1/2,0)$ keeps being feasible, then it is the unique optimal solution and the optimal value is $\gamma^*=1/2$. Again, the point $(0_2,0_2)\in\text{int}\,\Omega\times\text{int}\,\Gamma$ satisfies (3) and thus the problem satisfies assumption (2). The problem is not regular. In fact, $(x,w)\in(\Omega\times\Gamma)\cap\text{int}\,A_1$ implies both $x\notin\Gamma^*$ and $w\notin\Omega^*$ and therefore $x_2>1$ and $w_1>-1/3$. As a consequence, we have $\phi(0)\geq 1-1/3=2/3>\gamma^*$. On the contrary, condition (13) holds since $T(\Gamma^*,x^*)=\mathbb{R}_+\times\mathbb{R}_-$ while $T(\Omega,x^*)=\{\in\mathbb{R}^2 \ | \ x_1\geq 0,\ -x_1\leq x_2\leq x_1\}$.

When recast in the format (1), the canonical DC problem has a very peculiar structure: the variables w do not appear in the objective function. This feature is reflected also by condition (13), in which the role played by the sets Ω and Γ is not symmetric. Therefore, it is reasonable to expect (SRP) to call for additional conditions in order to guarantee regularity and the Lipschitz property.

Theorem 3.1 *If there exists an optimal solution (x^*,w^*) to (SRP) such that at least one of the following conditions*

$$\{x^*+\lambda u \ | \ \lambda>0\}\cap\Omega\neq\emptyset \quad \textit{and} \quad w^*u>0 \tag{14}$$

$$\{w^* + \lambda u \ \mid \ \lambda > 0\} \cap \Gamma \neq \emptyset \quad and \quad ux^* > 0 \tag{15}$$

holds for some direction $u \in \mathbb{R}^n$*, then the problem is regular and the value function* ϕ *is locally Lipschitz at 0, i.e., there exist* $L > 0$ *and* $\bar{\delta} > 0$ *such that*

$$\phi(\delta) - \phi(0) \leq L\delta \qquad \forall\ \delta \in [0, \bar{\delta}].$$

Proof. Suppose (14) holds. Therefore, there exists some $\bar{\lambda} > 0$ such that $x(\lambda) := x^* + \lambda u \in \Omega$ for all $\lambda \in [0, \bar{\lambda}]$ due to the convexity of Ω. Up to scaling the direction u, we can suppose $\bar{\lambda} = 1$. Since $w^*(x^* + \lambda u) = w^* x^* + \lambda w^* u > \alpha$, we have $(x(\lambda), w^*) \in (\Omega \times \Gamma) \cap \operatorname{int} A_\alpha$ and therefore (SRP) satisfies the regularity condition (7) as $x(\lambda) \to x^*$ for $\lambda \downarrow 0$. Thus, $\phi(0) = f(x^*) + g(w^*)$. Choosing any positive $\delta \leq w^* u/2$, we have $\lambda_\delta := 2\delta / w^* u \in (0, 1]$ and moreover

$$w^*(x^* + \lambda_\delta u) = w^* x^* + (2\delta / w^* u) w^* u = \alpha + 2\delta > \alpha + \delta.$$

Therefore, $(x^* + \lambda_\delta u, w^*) \in (\Omega \times \Gamma) \cap \operatorname{int} A_{\alpha+\delta}$. If δ is small enough, we have

$$\begin{aligned} \phi(\delta) - \phi(0) &\leq f(x^* + \lambda_\delta u) + g(w^*) - f(x^*) - g(w^*) \\ &\leq f(x^* + \lambda_\delta u) - f(x^*) \leq M \lambda_\delta \end{aligned}$$

for some suitable $M > 0$ due to the locally Lipschitz continuity of the convex function f. Setting $L = 2M / w^* u$, we get $\phi(\delta) - \phi(0) \leq L\delta$. The proof is analogous in case (15) holds, just exchanging the roles of the variables x and w. □

Remark 3.1 Relying on the reverse polar constraint, condition (14) can be equivalently formulated as $(\Omega \times \{w^*\}) \cap \operatorname{int} A_\alpha \neq \emptyset$ while condition (15) as $(\{x^*\} \times \Gamma) \cap \operatorname{int} A_\alpha \neq \emptyset$.

Notice that condition (14) cannot hold at any optimal solution (x^*, w^*) if $f \equiv 0$. Otherwise, $(x(\lambda), w^*)$ would be an optimal solution for any $\lambda > 0$ sufficiently small too and $w^* x(\lambda) > \alpha$ would contradict the property that all the optimal solutions satisfy (4). Similarly, condition (15) cannot hold at any optimal solution if $g \equiv 0$.

Since (14) and (13) are equivalent in the case of the canonical DC problem [23, Proposition 3.6], i.e., essentially (SRP) with $\alpha = 1$ and $g \equiv 0$, then Theorem 3.7 in [23] follows from Theorem 3.1. Notice that if $\alpha = 1$ but $g \not\equiv 0$ (14) implies (13) but not vice versa as shown in Example 3.2.

4 Convergence Conditions and the Basic Subprocedure

In this section, we first establish abstract conditions ensuring global (approximate) convergence of oracle-based algorithms for (SRP), and then present a generic subprocedure which is crucial for building actual implementable algorithms, discussing its properties.

4.1 General Convergence Conditions

The algorithms we will develop follow the generic cutting plane scheme sketched in Subsection 3.1. A non increasing sequence of feasible values $\{\gamma^k\}$ is produced, and the oracle Θ is called for each γ^k, providing either a value l^k such that condition (10) holds or points z^k and v^k satisfying conditions (11). By calling the oracle, repeatedly if needed, we can build a procedure which either proves that γ^k satisfies condition (10) or produces a better feasible value $\gamma^{k+1} < \gamma^k$. In the latter case, γ^{k+1} is produced (directly or indirectly) by points x^k and w^k such that

$$(x^k, w^k) \in (\Omega \times \Gamma) \cap \partial A_\alpha. \tag{16}$$

The rationale for (16) is that any optimal solution must satisfy this condition, due to feasibility and (4). A pair (x^k, w^k) may provide (potentially) different feasible values (see Subsection 5.5). Anyway, in the following we will always set $\gamma^{k+1} = f(x^k) + g(w^k)$ for the sake of simplicity.

With the above notation, we can introduce the prototype of our algorithms.

Algorithm 1 Prototype algorithm

---

0. $\gamma^1 = +\infty$; $k = 1$;
1. If the optimality condition (5) holds, then stop (γ^k is the optimal value);
2. Select (x^k, w^k) satisfying (16) such that $\gamma^{k+1} = f(x^k) + g(w^k) < \gamma^k$;
   set $k = k + 1$; goto 1.

Clearly, if at the initialization Step 0 a feasible pair (x^0, w^0) is known, one can alternatively set $\gamma^1 = f(x^0) + g(w^0)$. An important feature of Algorithm 1 is that $\{\gamma^k\}$ is a decreasing sequence bounded below:

$$\bar{\gamma} < \gamma^* \leq \gamma^\infty = \lim_{k \to \infty} \gamma^k < \cdots < \gamma^k < \gamma^{k-1} < \cdots < \gamma^1.$$

Therefore, $\{\ R(\gamma^k)\ \}$ is a "non-increasing sequence", i.e.,

$$R(\gamma^\infty) \subseteq \cdots \subseteq R(\gamma^{k+1}) \subseteq R(\gamma^k) \subseteq \cdots \subseteq R(\gamma^1).$$

Obviously, Algorithm 1 is too general to deduce any meaningful property. Indeed, something more has to be said about how exactly optimality condition (5) is checked, and how (x^k, w^k) such that $f(x^k) + g(w^k) < \gamma^k$ is selected once one knows that (5) is not fulfilled. Clearly, the two points are strictly interwoven, in that finding (x^k, w^k) such that $f(x^k)+g(w^k) < \gamma^k$ immediately proves that γ^k is not optimal. Vice versa, assume that we have any constructive procedure that produces a point $(z^k, v^k) \in R(\gamma^k)$ such that $v^k z^k > \alpha$ when γ^k is not optimal. Pick (x^k, w^k) in the intersection between ∂A_α and the segment with $(\bar{x}, \bar{w})$ and (z^k, v^k) as the end points: clearly, $(x^k, w^k) \in R(\gamma^k)$ and $f(x^k) + g(w^k) < f(z^k) + g(v^k) \leq \gamma^k$. Not surprisingly, without further qualification such a method does not provides a convergent algorithm ([3, Example 4.1]).

Thus some care is needed in choosing the sequences x^k and w^k. Indeed, the most general assumptions under which we can prove convergence are not stated in terms of x^k and w^k, but rather in terms of the two corresponding sequences z^k and v^k produced by (6), out of which x^k and w^k are constructed:

$$\liminf_{k\to\infty} v^k z^k \leq \alpha, \tag{17}$$

$$v^k z^k - \alpha \geq \varepsilon \max\{vz - \alpha \;\mid\; (z,v) \in R(\gamma^k)\}. \tag{18}$$

Condition (18), where $\varepsilon \in (0,1)$, basically says that v^k and z^k must be produced by some process attempting to solve the non-concave problem (6) with $\gamma = \gamma^k$, although the process may be "terminated early" due to the optimality tolerance ε. Condition (17) rather requires the two sequences to be asymptotically jointly feasible, and, as we will see, there are several different implementable ways for ensuring that this holds. Anyway, as far as abstract conditions go, (17) and (18) are already sufficient to guarantee convergence to the optimal value.

Proposition 4.1 *If* (17) *and* (18) *hold, then the sequence of feasible values* $\{\gamma^k\}$ *in Algorithm 1 converges to the optimal value* γ^*.

Proof. Since each γ^k is a feasible value, we have $\gamma^* \leq \gamma^\infty$, i.e. γ^∞ is a feasible value, too. Hence, (18) implies that $v^k z^k - \alpha \geq \varepsilon \max\{vz - \alpha \;\mid\; (z,v) \in R(\gamma^\infty)\}$ for all k. Taking the limit, (17) implies $\max\{vz - \alpha \mid (z,v) \in R(\gamma^\infty)\} \leq 0$, and therefore γ^∞ is the optimal value. □

When developing a "concrete" algorithm for (SRP), the abstract condition (17) can not be directly imposed on the sequences $\{z^k\}$ and $\{v^k\}$. In fact, these are the results of "complex" optimization process, i.e. approximately solving (6), upon which we want to impose as few conditions as possible, in order to leave as much freedom as possible to different implementations of this critical task. Therefore, we seek alternative ways for obtaining (17). However, given z^k and v^k as produced by the oracle we have full control on how x^k and w^k are constructed, provided that (16) is satisfied; we can use this to enforce (17) through either one of the following two pairs of conditions:

$$\begin{cases} \limsup_{k\to\infty} v^k(z^k - x^k) \leq 0 & \text{(a)} \\ \limsup_{k\to\infty} v^k x^k \leq \alpha & \text{(b)} \end{cases} \tag{19}$$

$$\begin{cases} \limsup_{k\to\infty} (v^k - w^k) z^k \leq 0 & \text{(a)} \\ \limsup_{k\to\infty} w^k z^k \leq \alpha & \text{(b)} \end{cases} \tag{20}$$

Lemma 4.1 *If either* (19) *or* (20) *hold, then* (17) *holds.*

Proof. Joining (19a) and (19b) we get $\limsup_{k\to\infty} v^k z^k - \alpha \leq 0$, whence (17); the proof for (20) is analogous. □

Therefore, we can define the two sets of conditions which, separately, guarantee convergence for Algorithm 1: $B_1 \equiv$ (18) and (19); $B_2 \equiv$ (18) and (20). Indeed, all the implementable algorithms we propose in the following imply at least one of these, and therefore provably solve (SRP) to optimality.

4.2 A Generic Outer Approximation Subprocedure

As already discussed in Subsection 3.1, one key idea to make (6) more tractable is to replace Γ and Ω by two "simpler" approximating convex sets Q and S. Clearly, this requires some appropriate machinery to update S and Q in order to make them "good enough" approximations of Ω and Γ as required. Convexity of both sets allows to rely on cutting plane procedures based on standard separation tools [24]. Given some point $x \in S \setminus \Omega$, we assume to be able to find an hyperplane *strictly separating* x from Ω. If the constraining function for Ω is known, for instance, this requires finding $s \in (\bar{x}, x) \cap \partial\Omega$ for some $\bar{x} \in \operatorname{int}\Omega$ and a subgradient of the constraining function at s. It is worth noting that the condition $\bar{x} \in \operatorname{int}\Omega$, which implies $\bar{x} \in \operatorname{int} S$ due to (9), is needed to ensure that $s \neq x$, and therefore that the hyperplane actually separates S and x strictly. Obviously, we make analogous assumptions for Γ.

Exploiting the above separation tools and relying on an approximate oracle Θ, we can build a generic outer approximation procedure which allows implementations of Algorithm 1 satisfying the sufficient convergence conditions introduced in Subsection 4.1. We call this procedure "generic" because it does not provide any specific rule for selecting x^i and w^i from z^i and v^i. In the next section we will discuss different rules which lead to implementable algorithms.

Subprocedure 1 Outer approximation subprocedure

Input: S and Q, closed convex sets satisfying (9), a feasible value γ

0. $S^1 = S$; $Q^1 = Q$; $i = 1$;
1. call the oracle Θ on S^i, Q^i, and γ, with tolerances $\bar{\varepsilon}$ ($> \varepsilon$) and ε';
   if Θ produces an upper bound l^i satisfying (10)
   then stop.
   else Θ produces $(\bar{z}^i, \bar{v}^i)$ satisfying $\bar{v}^i \bar{z}^i - \alpha \geq \bar{\varepsilon} v(\overline{OC}_\gamma)$;
2. if $\bar{z}^i \notin \Omega$ then use S^i and $\bar{z}^i$ to produce $S^{i+1} \not\ni \bar{z}^i$;
   else $S^{i+1} = S^i$;
3. if $\bar{v}^i \notin \Gamma$ then use Q^i and $\bar{v}^i$ to produce $Q^{i+1} \not\ni \bar{v}^i$;
   else $Q^{i+1} = Q^i$;
4. let $(z^i, v^i) = (1 - \beta^i)(\bar{z}^i, \bar{v}^i) + \beta^i(\bar{x}, \bar{w})$ for the *smallest* $\beta^i \geq 0$
   such that $z^i \in \Omega$ and $v^i \in \Gamma$.
5. if $v^i z^i - \alpha < \varepsilon v(\overline{OC}_\gamma)$ then $i = i + 1$ and goto 1.
6. select $x^i \in \Omega$ and $w^i \in \Gamma$ such that $w^i x^i = \alpha$ relying on z^i and v^i; stop.

Output: Q^i and S^i; either l^i, or x^i, w^i, z^i, v^i.

The following properties are independent of the selection rule for x^i and w^i:

1. We assume (9) for S^1 and Q^1: adding cutting planes at steps 2 and/or 3 ensures (9) for any i, i.e., we get "non-increasing" sequences
$$\Omega \subseteq \cdots \subseteq S^{i+1} \subseteq S^i \subseteq \cdots \subseteq S^1, \tag{21}$$
$$\Gamma \subseteq \cdots \subseteq Q^{i+1} \subseteq Q^i \subseteq \cdots \subseteq Q^1. \tag{22}$$

2. The choice of (z^i, v^i) at step 4 guarantees
$$f(z^i) + g(v^i) \leq \max\{\, f(\bar{z}^i) + g(\bar{v}^i),\ f(\bar{x}) + g(\bar{w}) \,\} \leq \gamma,$$
i.e., z^i and v^i are also feasible for the maximization problem which is approximately solved by the oracle. Note that the step β^i can be 0: this happens when $(\bar{z}^i, \bar{v}^i) \in \Omega \times \Gamma$.

3. The choices of (z^i, v^i) and (x^i, w^i) guarantee they belong to $\Omega \times \Gamma$ and thus they have a finite value of the objective function since $\Omega \times \Gamma \subseteq \mathrm{dom}\, f \times \mathrm{dom}\, g$. This may not happen for $(\bar{z}^i, \bar{v}^i)$, but it has no influence on the algorithm.

4. The condition "$v^i z^i - \alpha < \varepsilon v(\overline{OC}_\gamma)$" at step 5 may be difficult to check directly, as the value of $v(\overline{OC}_\gamma)$ is not known (although a suitable upper bound must be computed by the oracle Θ). A stronger condition that can be surely checked is the following:

$$v^i z^i - \alpha \geq (\bar{v}^i \bar{z}^i - \alpha)(\varepsilon/\bar{\varepsilon}).$$

Indeed, if it holds, then we are guaranteed that (z^i, v^i) satisfies (11) and therefore (18), and the algorithm can advance to step 5, otherwise it loops.

5. If the algorithm loops at step 5, then at least one between $Q^i \neq Q^{i+1}$ and $S^i \neq S^{i+1}$ holds. In fact, if $\bar{z}^i \in \Omega$ and $\bar{v}^i \in \Gamma$ then $z^i = \bar{z}^i$ and $v^i = \bar{v}^i$, so that the condition at step 5 cannot be true.

In the algorithm, we are forced to require a "stricter" tolerance $\bar{\varepsilon}$ to the oracle Θ in order to be able to guarantee convergence to a solution that is optimal only to within $\delta = \varepsilon'/\varepsilon$ for the "looser" tolerance $\varepsilon < \bar{\varepsilon}$; the exact role of this assumption will be discussed in details later on. However, nothing is required to the ratio $\varepsilon/\bar{\varepsilon}$ except being smaller than one, so the two tolerances can be taken arbitrarily close to each other.

It is clear from the previous discussion that the subprocedure will never repeat the same iterates: if it does not stop, then at least one between the inclusions $S^i \setminus \{\bar{z}^i\} \supset S^{i+1}$ and $Q^i \setminus \{\bar{v}^i\} \supset Q^{i+1}$ holds, so at least one between $\bar{z}^{i+1} \neq \bar{z}^i$ and $\bar{v}^{i+1} \neq \bar{v}^i$ holds. We now prove the basic properties of the subprocedure, which lead to finite termination under $\varepsilon' > 0$.

Lemma 4.2 *If the subprocedure never ends, then all the cluster points of $\{\bar{z}^i\}$ and $\{\bar{v}^i\}$ belong to Ω and Γ, respectively.*

Proof. Consider the sequence $\{\bar{z}^i\}$: either $\bar{z}^i \in \Omega$ for all large enough i, and therefore the thesis follows from the closeness of Ω, or $\bar{z}^i \in S^i \setminus \Omega$ for infinitely many indices. In the latter case, the general Basic Outer Approximation Theorem [24, Theorem II.1] ensures that all the cluster points of $\{\bar{z}^i\}$ belong to Ω. The same reasoning works for $\{\bar{v}^i\}$. □

It is crucial to ensure that the sequences $\{\bar{z}^i\}$ and $\{\bar{v}^i\}$ do indeed have cluster points. Since both Ω and Γ are compact, then it is natural to assume that the sequences $\{\bar{z}^i\}$ and $\{\bar{v}^i\}$ are bounded; in view of (21) and (22), this holds e.g. if S^1 and Q^1 are compact. We therefore assume both sequences to be bounded in all the following development. Besides, all the other sequences are bounded; in fact they belong to the bounded sets Ω and Γ. We can now prove that the "original" sequence $\{(\bar{z}^i, \bar{v}^i)\}$ and the "modified" one $\{(z^i, v^i)\}$ share the same set of cluster points.

Lemma 4.3 *If infinitely many iterates i are produced, then there is a one-to-one correspondence between the cluster points of $\{(\bar{z}^i, \bar{v}^i)\}$ and those of $\{(z^i, v^i)\}$.*

Proof. It is enough to prove that 0 is the only cluster point of $\{\beta^i\}$. Assume by contradiction that $\beta^i \to \bar{\beta} \in (0, 1]$. By taking subsequences, if needed, let $(\bar{z}^i, \bar{v}^i) \to (\bar{z}, \bar{v})$. Then, we have $(z^i, v^i) \to (\hat{z}, \hat{v}) = (1-\bar{\beta})(\bar{z}, \bar{v}) + \bar{\beta}(\bar{x}, \bar{w})$. Since $\bar{\beta} > 0$, $(\bar{x}, \bar{w}) \in \operatorname{int}\Omega \times \operatorname{int}\Gamma$ and $(\bar{z}, \bar{v}) \in \Omega \times \Gamma$, then we get $(\hat{z}, \hat{v}) \in \operatorname{int} \Omega \times \operatorname{int} \Gamma$. Therefore, we have $(z^i, v^i) = \beta^i(\bar{x}, \bar{w}) + (1-\beta^i)(\bar{z}^i, \bar{v}^i) \in \operatorname{int}\Omega \times \operatorname{int}\Gamma$ for sufficiently large i, and thus the segment with (z^i, v^i) and $(\bar{z}^i, \bar{v}^i)$ as end points has a nonempty intersection with the set $(\operatorname{int}\Omega \times \operatorname{int}\Gamma)$. This contradicts the assumption that β^i is the smallest non-negative value such that $z^i \in \Omega$ and $v^i \in \Gamma$. □

Hence, the subprocedure cannot loop infinitely many times as step 5.

Proposition 4.2 *If $\varepsilon' > 0$, then the subprocedure finitely stops.*

Proof. If the subprocedure never ends, then

$$(v^i z^i - \alpha)/\varepsilon < v(\overline{OC}_\gamma) \leq (\bar{v}^i \bar{z}^i - \alpha)/\bar{\varepsilon}$$

holds for all the indices i. Take any common cluster point $(\bar{z}, \bar{v})$ of $\{(z^i, v^i)\}$ and $\{(\bar{z}^i, \bar{v}^i)\}$. Therefore, we get the contradiction $1 = (\bar{v}\bar{z} - \alpha)/(\bar{v}\bar{z} - \alpha) \leq \varepsilon/\bar{\varepsilon} < 1$, just taking the limit in the above chain of inequalities for the subsequence providing the cluster point. □

The above proof shows the need for requiring $\varepsilon' > 0$, since the subprocedure may never stop for $\varepsilon' = 0$: it can not *finitely* prove that the optimal value is optimal. That is why it is important to clarify the relationship between approximated optimal values and the optimal value. Furthermore, the proof also shows that requiring the "tighter" tolerance $\bar{\varepsilon}$ on $(\bar{z}^i, \bar{v}^i)$ is needed in order to ensure that the "looser" tolerance ε is attained on the modified iterates (z^i, v^i), and therefore to guarantee finite termination.

If the subprocedure stops at step 1, then γ is approximately optimal with positive tolerances $\bar{\varepsilon}$ and ε'; if it stops at step 6, the existence and convergence properties of (x^i, w^i) would depend on the exact choice of the selection rule. A detailed discussion on this issue and the corresponding convergence proofs will be given in the following sections.

We end this section with a further result which greatly simplifies the analysis of convergence of the algorithms. For several of them, it is necessary to impose a further condition (other than

$\gamma^1 \geq \gamma^*$) on the initial value, i.e.,

$$\gamma^1 \leq \min\{\gamma_{\bar{x}}, \gamma_{\bar{w}}\}, \tag{23}$$

where

$$\gamma_{\bar{x}} = f(\bar{x}) + \min\{g(w) \ \mid w \ \in \Gamma, \ w\bar{x} \geq \alpha\},$$

$$\gamma_{\bar{w}} = g(\bar{w}) + \min\{f(x) \ \mid \ x \in \Omega, \ \bar{w}x \geq \alpha\}.$$

Note that $\gamma_{\bar{x}}$ and $\gamma_{\bar{w}}$ are the optimal values of two convex problems, hence "easily" available. Furthermore, if (23) holds at the first iteration, then it automatically holds at all subsequent ones since $\{\gamma^k\}$ is a decreasing sequence. However, it has to be remarked that either one (and even both) can be $+\infty$, as there is no guarantee that the corresponding feasible regions are nonempty. This surely happens if $\alpha > 0$ and $(\bar{x}, \bar{w}) = (0, 0)$, which can always be assumed without loss of generality in the (CDC) case [3].

Lemma 4.4 *If* (23) *holds, then* $g(v^i) \geq g(\bar{w})$ *implies* $\bar{w}z^i \leq \alpha$ *and* $f(z^i) \geq f(\bar{x})$ *implies* $v^i\bar{x} \leq \alpha$.

Proof. We only prove the first implication, as the proof of the other is symmetric. By contradiction, suppose that both $g(v^i) \geq g(\bar{w})$ and $\bar{w}z^i > \alpha$ hold. Since $\bar{w}\bar{x} < \alpha$, the mean value theorem implies that there exists some $\tilde{x} \in (\bar{x}, z^i)$ such that $\bar{w}\tilde{x} = \alpha$. Hence, we have $f(\bar{x}) + g(\bar{w}) < \gamma \leq \gamma_{\bar{w}} \leq f(\tilde{x}) + g(\bar{w})$ and therefore $f(\bar{x}) < f(\tilde{x}) < f(z^i)$. Thus, we have the contradictory chain of inequalities $\gamma_{\bar{w}} \leq f(\tilde{x}) + g(\bar{w}) < f(z^i) + g(v^i) \leq \gamma \leq \gamma_{\bar{w}}$. □

5 Implementable Algorithms

The results of Section 4 can be exploited to define implementable versions of the prototype Algorithm 1, as described in Algorithm 2.

Algorithm 2 Implementable outer approximation algorithm

0. $\gamma^1 = +\infty$; Select $S^1 \supseteq \Omega$, $Q^1 \supseteq \Gamma$; $k = 1$;
1. Call subprocedure 1 with S^k, Q^k, and γ^k;
2. If subprocedure 1 stops at Step 1, then stop.
3. Set x^k, w^k, z^k and v^k as the output of subprocedure 1;
4. Set Q^{k+1} and S^{k+1}, possibly using the output of subprocedure 1;
5. Set $\gamma^{k+1} = f(x^k) + g(w^k) < \gamma^k$; set $k = k + 1$; goto 1.

At Step 4 the obvious possibility for Q^{k+1} and S^{k+1} is to choose the sets Q^i and S^i produced by subprocedure 1. However, this leads to accumulation of *all* the cutting planes generated along the iterates in Q^k and S^k so that one may want to "purge" some of the accumulated cutting planes. This can always be done since only (9) needs to be satisfied.

In order for Algorithm 2 to work, at least one between the set of conditions B_1 and B_2 must hold. This requires appropriate ways of constructing x^k and w^k out of z^k and v^k, which have not been specifed in the subprocedure. In all the concrete algorithms below x^k and w^k are obtained from z^k and v^k moving along the *directions which "lead towards the low point"* $(\bar{x}, \bar{w})$ with appropriate non-negative stepsizes, i.e.,

$$x^k = z^k - \lambda_1^k y^k, \quad w^k = v^k - \lambda_2^k u^k$$

with $\lambda_1^k, \lambda_2^k \geq 0$ and

$$y^k = N(z^k - \bar{x}), \quad u^k = N(v^k - \bar{w}), \tag{24}$$

where $N(q)$ denotes the normalized direction associated to any $q \in \mathbb{R}^n$, i.e., $N(q) = q/\|q\|$ if $q \neq 0$ while $N(q) = 0$ if $q = 0$. The algorithms differ just by the selection of the stepsizes. Anyway, all the selection rules impose $\lambda_1^k \in [0, \|z^k - \bar{x}\|]$ and $\lambda_2^k \in [0, \|v^k - \bar{w}\|]$ so that

$$x^k \in [\bar{x}, z^k], \quad w^k \in [\bar{w}, v^k] \tag{25}$$

hold. In this way the required conditions $x^k \in \Omega$ and $w^k \in \Gamma$ are guaranteed by the convexity of the sets Ω and Γ.

Thanks to (11), (21) and (22), condition (18) is always satisfied by all possible variants of Algorithm 2. Therefore, B_1 and B_2 actually reduce to (19) and (20), respectively. Indeed, a sufficient condition for them to hold is that the *steps vanish*, i.e.,

$$\lambda_1^k \to 0, \quad \lambda_2^k \to 0, \tag{26}$$

provided that $w^k x^k = \alpha$. In fact, (26) implies $v^k(z^k - x^k) = v^k \lambda_1^k y^k \to 0$, and therefore, $v^k x^k = (w^k + \lambda_2^k u^k) x^k = \alpha + \lambda_2^k u^k x^k \to \alpha$, i.e., condition (19). Analogously, it ensures also (20). Notice that convergence is achieved if *either* condition B_1 or B_2 hold, while (26) implies *both*. Of course, if steps vanish so does the difference in objective function value, i.e., (26) implies

$$(f(z^k) + g(v^k)) - (f(x^k) + g(w^k)) \to 0 \tag{27}$$

while the converse need not be true. In the next subsections we discuss different *stepsize rules* which guarantee the *existence* of (x^k, w^k) satisfying $w^k x^k = \alpha$ and the vanishing step condition (26), therefore ensuring the convergence of Algorithm 2.

5.1 Algorithm R_1

The first implementable algorithm we propose employs the stepsizes

$$\lambda_1^k = \lambda^k \|z^k - \bar{x}\|, \quad \lambda_2^k = \lambda^k \|v^k - \bar{w}\|, \tag{28}$$

where $\lambda^k \in]0,1[$. The rationale of (28) is that we want to *reduce the relative distance between z^k and v^k and their respective low points ($\bar{x}$ and $\bar{w}$) at the same rate*; in fact, λ^k is the fraction of the distance that is travelled, and it must be the same in the x-space and in the w-space.

The existence of (x^k, w^k) satisfying $w^k x^k = \alpha$ is obvious: in fact, we get $w^k x^k = v^k z^k > \alpha$ if $\lambda^k = 0$ while $w^k x^k = \bar{w}\bar{x} < \alpha$ if $\lambda^k = 1$, and the result follows by continuity. An explicit formula can be easily derived for the correct value of λ^k, but it has no relevance in the analysis of convergence.

Lemma 5.1 *If* (24) *and* (28) *hold, then* (26) *holds.*

Proof. The assumptions guarantee $(x^k, w^k) = \lambda^k(\bar{x}, \bar{w}) + (1 - \lambda^k)(z^k, v^k)$, and therefore

$$\bar{\gamma} = f(\bar{x}) + g(\bar{w}) < f(x^k) + g(w^k) \leq f(z^k) + g(v^k) \leq \gamma^k \tag{29}$$

follows from the convexity of f and g. Since $\gamma^k \leq f(x^{k-1}) + g(w^{k-1})$, then $\{f(x^k) + g(w^k)\}$ is a non-increasing sequence bounded below and hence convergent. As a consequence, we have $(f(x^{k-1}) + g(w^{k-1})) - (f(x^k) + g(w^k)) \to 0$ and thus (27) follows from (29). Furthermore, the convexity of f and g implies

$$f(z^k) - f(x^k) + g(v^k) - g(w^k) \geq \lambda^k[f(z^k) - f(\bar{x}) + g(z^k) - g(\bar{w})] \geq \lambda^k(\gamma^* - \bar{\gamma}).$$

Thus, (27) implies $\lambda^k \to 0$ and (26) follows immediately from (28) since the sequences $\{\|z^k - \bar{x}\|\}$ and $\{\|v^k - \bar{w}\|\}$ are bounded. $\square$

5.2 Algorithm R_2

"Abstract" conditions ensuring (26) are the following

$$\lambda_1^k > 0 \implies f(z^k) - f(x^k) \geq \tau \|z^k - x^k\|, \tag{30}$$

$$\lambda_2^k > 0 \implies g(v^k) - g(w^k) \geq \tau \|v^k - w^k\|, \tag{31}$$

where τ is a small enough positive value.

Lemma 5.2 *If* (30) *and* (31) *hold, then* (26) *holds.*

Proof. Conditions (30) and (31) imply that $f(z^k) \geq f(x^k)$ and $g(v^k) \geq g(w^k)$, which in turn guarantee $f(z^k) + g(v^k) \geq f(x^k) + g(w^k)$ and therefore (27) follows just arguing as in the proof of Lemma 5.1. Since both terms are non-negative, we have $f(z^k) - f(x^k) \to 0$ and $g(v^k) - g(w^k) \to 0$. Therefore, (30) and (31) imply $\lambda_1^k \|y^k\| \to 0$ and $\lambda_2^k \|u^k\| \to 0$, which guarantee (26). □

At first glance, it is not obvious how the rather abstract conditions (30) and (31) can be guaranteed. However, any stepsize rule providing (25) ensures that at least a suitable τ exists.

Lemma 5.3 *If* (25) *holds, then there exists* $\tau > 0$ *such that either* $f(z^k) - f(x^k) \geq \tau \|z^k - x^k\|$ *or* $g(v^k) - g(w^k) \geq \tau \|v^k - w^k\|$ *holds at each iteration.*

Proof. If $x^k = z^k$ or $w^k = v^k$, then the thesis is obvious. Otherwise, consider the value $M = \sup_k \{\max\{\|x^k - \bar{x}\|, \|w^k - \bar{w}\|\}\}$, which is finite since the sequence $\{(x^k, w^k)\}$ is bounded. Taking any positive $\tau \leq (\gamma^* - \bar{\gamma})/2M$ we have

$$(f(x^k) + g(w^k)) - (f(\bar{x}) + g(\bar{w})) \geq \gamma^* - \bar{\gamma} \geq 2\tau M \geq \tau \|x^k - \bar{x}\| + \tau \|w^k - \bar{w}\|,$$

and therefore either $f(x^k) - f(\bar{x}) \geq \tau \|x^k - \bar{x}\|$ or $g(w^k) - g(\bar{w}) \geq \tau \|w^k - \bar{w}\|$ holds. Since (25) holds, then $x^k = \bar{x} + \lambda^k (z^k - \bar{x})$ with $\lambda^k = \|x^k - \bar{x}\| / \|z^k - \bar{x}\| \in]0, 1[$. Therefore, the convexity of f implies

$$[f(z^k) - f(x^k)] / \|z^k - x^k\| \geq [f(x^k) - f(\bar{x})] / \|x^k - \bar{x}\|.$$

Similarly, we can prove the corresponding inequality for v^k, w^k and g and therefore the thesis follows immediately exploiting these two inequalities. □

Although a suitable τ exists, it remains to show how the stepsizes can be chosen to guarantee (30) and (31). In particular, one has to detect which of the two directions must be given a nonzero stepsize. Just arguing as in Lemma 5.3, we can show also that $f(z^k) - f(\bar{x}) \geq \tau \|z^k - \bar{x}\|$ or $g(v^k) - g(\bar{w}) \geq \tau \|v^k - \bar{w}\|$ holds at each step for any small enough positive τ. Therefore, we propose the following rule:

$$(\lambda_1^k, \lambda_2^k) = \begin{cases} \left(0, \ \frac{v^k z^k - \alpha}{(v^k - \bar{w}) z^k} \|v^k - \bar{w}\|\right), & \text{if } \frac{f(z^k) - f(\bar{x})}{\|z^k - \bar{x}\|} \leq \frac{g(v^k) - g(\bar{w})}{\|v^k - \bar{w}\|}, \\ \left(\frac{v^k z^k - \alpha}{v^k (z^k - \bar{x})} \|z^k - \bar{x}\|, 0\right) & \text{otherwise.} \end{cases} \tag{32}$$

The positive stepsize is chosen in such a way that the corresponding (x^k, w^k) satisfies $w^k x^k = \alpha$. Under an additional condition on the initial value, the selection rules (24) and (32) provide a convergent algorithm.

Lemma 5.4 *Suppose* (23) *holds. If* (24) *and* (32) *hold, then also* (25)*,* (30) *and* (31) *hold.*

Proof. Suppose $[f(z^k)-f(\bar{x})]/\|z^k-\bar{x}\| \leq [g(v^k)-g(\bar{w})]/\|v^k-\bar{w}\|$. Then, we have $g(v^k) \geq g(\bar{w})$ and Lemma 4.4 implies $\bar{w}z^k \leq \alpha$. Since $v^k z^k > \alpha$, the stepsize rule (32) implies $0 < \lambda_2^k \leq \|v^k-\bar{w}\|$, and therefore (25) holds. The thesis follows immediately as Lemma 5.3 guarantees that the inequality $g(v^k) - g(\bar{w}) \geq \tau\|v^k - \bar{w}\|$ holds for some $\tau > 0$, while $\lambda_2^k > 0$, i.e., condition (31), and $\lambda_1^k = 0$ guarantees (30). The proof of the other case is analogous. □

5.3 Algorithm R_3

Yet another different choice for the stepsizes would be to take them equal, i.e.,

$$\lambda_1^k = \lambda_2^k = \lambda^k, \tag{33}$$

which would imply $\lambda^k = \|z^k - x^k\| = \|v^k - w^k\|$. Anyway, this choice does not fall within the framework of Algorithm 2 as the following example shows.

Example 5.1 Consider (SRP) with $n = 1$, $\alpha = 1$,

$$\Omega = [-1/2, 20], \quad f(x) = x,$$

$$\Gamma = [-1/2, 1/2], \quad g(w) = -2w,$$

together with the low point $(\bar{x}, \bar{w}) = (0, 0) \in \operatorname{int}\Omega \times \Gamma$. The optimal solution is $(x^*, w^*) = (2, 1/2)$, with optimal value $\gamma^* = 1$. Take $S^1 = \Omega$ and $Q^1 = \Gamma$ and set $\varepsilon = 0.9$, $\bar{\varepsilon} = 1$ and $\gamma^1 = 19$. The subprocedure gives $(z^1, v^1) = (\bar{z}^1, \bar{v}^1) = (20, 1/2)$. The directions (24) are $y^1 = u^1 = 1$ and the stepsize rule (33) chooses $\lambda_1^1 = \lambda_2^1 \approx 0.44885$. Then, $(x^1, w^1) \approx (19.55115, 0.05115)$ and $f(x^1) + g(w^1) \approx 19.44885 > \gamma^1$, which does not satisfy the monotonicity of the objective value required by Algorithm 2.

It is worth noting that in the above example we have $g(\bar{w}) > g(v^1)$ and hence moving from v^1 towards $\bar{w}$ cannot lead to any improvement for "g-part" of the objective function. This simple remark suggests to modify the rule (33) in the following way:

$$\lambda_1^k = \begin{cases} \lambda^k, & \text{if } f(z^k) > f(\bar{x}), \\ 0 & \text{otherwise} \end{cases} \qquad \lambda_2^k = \begin{cases} \lambda^k, & \text{if } g(v^k) > g(\bar{w}), \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

Indeed, this modification guarantees convergence.

Lemma 5.5 *If* (24) *and* (34) *hold, then* (26) *holds.*

Proof. The stepsize rule (34) guarantees $f(z^k) \geq f(x^k)$ and $g(v^k) \geq g(w^k)$. In fact, it implies $x^k = z^k$ whenever $f(z^k) \leq f(\bar{x})$ and $v^k = w^k$ whenever $g(v^k) \leq g(\bar{w})$ while in the other cases the inequalities follow from the convexity of f and g in view of (25). Moreover, we clearly have $f(z^k)+g(v^k) > f(x^k)+g(w^k)$, otherwise it would be $f(z^k)+g(v^k) = \bar{\gamma}$ contradicting (3). Therefore, the thesis follows just arguing as in the proof of Lemma 5.1. □

Anyway, there is no guarantee that choosing the stepsizes according to (34) allows to get a point $(x^k, w^k) \in \Gamma \times \Omega$ satisfying $w^k x^k = \alpha$. This is true taking an initial value satisfying (23).

Lemma 5.6 *If (23) holds, then there exists* $\lambda^k > 0$ *such that (24) and (34) guarantee (25) and* $w^k x^k = \alpha$.

Proof. If $f(z^k) \leq f(\bar{x})$, then $\lambda_1^k = 0$, i.e., $x^k = z^k$, and also $g(v^k) > g(\bar{w})$, which implies $\bar{w} x^k = \bar{w} z^k \leq \alpha$ by Lemma 4.4. Since $v^k z^k > \alpha$, then there exists $w^k \in [\bar{w}, v^k[$ such that $w^k x^k = \alpha$ and therefore $\lambda^k = \|v^k - w^k\|$ provides the required stepsize. Analogously, if $g(v^k) \leq g(\bar{w})$, then we have $w^k = v^k$ and $\lambda^k = \|z^k - x^k\|$ for some suitable $x^k \in [\bar{x}, z^k[$.

Finally, if both $f(z^k) > f(\bar{x})$ and $g(v^k) > g(\bar{w})$ hold, consider the real-valued function $\zeta(\lambda) = (v^k - \lambda u^k)(z^k - \lambda y^k)$. Clearly, we have both $\zeta(0) = v^k z^k > \alpha$ and $\zeta(\lambda) < \alpha$ for $\lambda = \min\{\|z^k - \bar{x}\|, \|v^k - \bar{w}\|\}$. In fact, $\zeta(\lambda) = w\bar{x}$ for some $w \in [\bar{w}, v^k[$ or $\zeta(\lambda) = \bar{w}x$ for some $x \in [\bar{x}, z^k[$ and $\bar{w}\bar{x} < \alpha$. Lemma 4.4 guarantees $\bar{w} z^k \leq \alpha$ and $v^k \bar{x} \leq \alpha$. Thus, the continuity of ζ implies the existence of λ^k such that $w^k x^k = \zeta(\lambda^k) = \alpha$. Since $\lambda^k \leq \lambda$, then (25) follows too. □

Remark 5.1 The idea behind the stepsize rule (34) can be applied to modify rule (28) correspondingly, i.e.,

$$\lambda_1^k = \begin{cases} \lambda^k \|z^k - \bar{x}\|, & \text{if } f(z^k) > f(\bar{x}), \\ 0 & \text{otherwise} \end{cases}$$

$$\lambda_2^k = \begin{cases} \lambda^k \|v^k - \bar{w}\|, & \text{if } g(v^k) > g(\bar{w}), \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, this new rule provides a convergent algorithm. Actually, it is a sort of combination of the algorithms R_1 and R_3: if $f(z^k) > f(\bar{x})$ and $g(v^k) > g(\bar{w})$ hold, then it performs the same iteration that R_1 would, otherwise the one that R_3 would.

5.4 Numerical Illustrations

The algorithms R_1, R_2 and R_3 are indeed different, in the sense that they may produce different optimizing sequences even if the same problem and the same starting conditions are given. We suppose the oracle Θ to always choose the same (z,v) when S, Q and γ are the same; nonetheless, the three algorithms construct different optimization sequences for (SRP) with $\alpha = 1$, $n = 2$,

$$\Gamma = \{(w_1, w_2) \ \mid \ w_1^2 + w_2^2 \le 1/4\}, \quad f(x) = x_2,$$

$$\Omega = \{(x_1, x_2) \ \mid \ -1 \le x_1 \le 1, \ x_2 \ge -1\}, \quad g(w) = w_1/2.$$

The point $(\bar{x}, \bar{w}) \in \operatorname{int}\Omega \times \operatorname{int}\Gamma$ with $\bar{x} = \bar{w} = 0$ satisfies (3), and therefore $\bar{\gamma} = 0$. Take $S^1 = \Omega$ and $Q^1 = \Gamma$, and set $\bar{\varepsilon} = 0.8$, $\varepsilon = 0.5$ and $\gamma^1 = 4$. All the algorithms start calling the subprocedure with S^1, Q^1 and γ^1 as input data. Suppose the oracle outputs the solution $\bar{z}^1 = (1,2)$, $\bar{v}^1 = (2/9, 4/9)$, which indeed satisfies

$$\bar{\varepsilon} v(\overline{OC}_{\gamma^1}) = \bar{\varepsilon}(\sqrt{5}/2 - 1) = 0.8(-1 + \sqrt{5}/2) \simeq 0.094 \le 0.111 \simeq (\bar{v}^1 \bar{z}^1 - 1).$$

Since $\bar{z}^1 \in \Omega$ and $\bar{v}^1 \in \Gamma$, then $\beta^1 = 0$, i.e., $(z^1, v^1) = (\bar{z}^1, \bar{v}^1)$. Since $\varepsilon v(\overline{OC}_{\gamma^1}) \simeq 0.059 \le (v^1 z^1 - 1)$, the subprocedure immediately stops. However, the new iterates x^1 and w^1 and the corresponding new feasible value γ^2 are different in the three cases: algorithm R_1 provides $x^1 = (3,6)/\sqrt{10}$ and $w^1 = \sqrt{10}(1/15, 2/15)$, therefore $\gamma^2 = 19\sqrt{10}/30 \simeq 2.003$; algorithm R_2 provides and $x^1 = (9/10, 9/5)$ and $w^1 = (2/9, 4/9)$, therefore $\gamma^2 = 86/45 \simeq 1.911$; algorithm R_3 provides $x^1 \simeq 0.98(1,2)$ and $w^1 \simeq 0.202(1,2)$, therefore $\gamma^2 \simeq 2.061$. Thus, the three algorithms produce different values γ^2 starting from the same situation. Furthermore, it is possible to prove [13, Examples 3.4.4-5] that these algorithms are also different from those for (CDC) described in [3, 13].

5.5 Possible Practical Improvements

While all the algorithms proposed in this section are (approximately) globally convergence, little is known about their actual rate of convergence. Interestingly, a simple technique can be used to try to improve the convergence speed of γ^k to the optimal value γ^*; this can be done, at each iteration, by finding a "better" feasible value than $\gamma^{k+1} = f(x^k) + g(w^k)$. In fact, once x^k and w^k are selected, we can fix one of the two and optimize over the other; in other words, one may consider solving the two convex minimization problems

$$\bar{w}^k \in \operatorname{argmin}\{g(w) \ \mid \ w \in \Gamma, \ wx^k \ge \alpha\}$$

$$\bar{x}^k \in \operatorname{argmin}\{ f(x) \;\mid\; x \in \Omega,\; w^k x \geq \alpha \}$$

and set $\gamma^{k+1} = \min\{ f(x^k) + g(\bar{w}^k), f(\bar{x}^k) + g(w^k) \}$. Since $w^k x^k = \alpha$, both problems are feasible and this definition of γ^{k+1} cannot provide a larger (hence worse) value than the standard one. Therefore, it is easy to see that this modification retains the global convergence of the original algorithms, although it would not in general be easy to prove this convergence with the modified sequences. Remarkably, the process can be iterated: whenever, say, $f(x^k) + g(\bar{w}^k) < f(x^k) + g(w^k)$, one may re-solve the second optimization problem above with $w^k = \bar{w}^k$, and keep doing so (alternating between the two blocks of variables) until no further improvement is possible.

6 Conclusions

We have shown how to extend the oracle-based outer approximation solution methods developed for the canonical DC problem to the larger class of the single reverse polar problems, which comprises interesting problems such as separable linear complementarity and separable bilevel ones. As this class seems to be new, a thorough analysis (approximate) of optimality condition and properties of optimal solutions in (SRP) has been performed, as well as the comparison with the corresponding features of (CDC). The concept of approximated oracle devised for (CDC) directly extends to (SRP); this has the potential to make oracle-based algorithms practical even for large-scale instances, in contrast to the vertex enumeration techniques usually touted for the (CDC) case. To this purpose we have developed a general hierarchy of conditions ensuring convergence of oracle-based outer approximation algorithms for (SRP), a general algorithmic scheme based on the hierarchy, and three different implementable algorithms which can generate an approximate optimal value in a finite number of steps, where the error can be managed and controlled. To the best of our knowledge, there are no existing algorithms devoted to (SRP). Despite the fact that (CDC) is just a special case of (SRP) with $\alpha = 1$, $f(x) = dx$, $g(w) = 0$ and $\Gamma = C^*$, oracle-based outer approximation algorithms for (CDC) can not be applied to (SRP) directly: some crucial properties of (CDC) are lost in (SRP), which requires a significant update both for the theory and for the algorithms.

References

[1] Tuy, H.: Canonical DC programming problem: outer approximation methods revisited. Oper. Res. Lett. **18**, 99–106 (1995)

[2] Tuy, H.: D.C. optimization: theory, methods and algorithms. In: Horst, R., Pardalos, P.M. (eds.): Handbook of global optimization, Nonconvex optimization and its applications, vol. 2, pp. 149–216. Kluwer Academic Publishers, Dordrecht (1995)

[3] Bigi, G., Frangioni, A., Zhang Q.H.: Outer approximation algorithms for canonical DC problems. J. Global Optim. **46**, 163–189 (2010)

[4] Fulop, J.: A finite cutting plane method for solving linear programs with an additional reverse constraint. Eur. J. Oper. Res. **44**, 395–409 (1990)

[5] Nghia, M.D., Hieu, N.D.: A method for solving reverse convex programming problems. Acta Math. Vietnam. **11**, 241–252 (1986)

[6] Tao, P.D., El Bernoussi, S.: Numerical methods for solving a class of global nonconvex optimization problems. International Series of Numerical Mathematics **87**, 97–132 (1989)

[7] Thach, P.T.: Convex programs with several additional reverse convex constraints. Acta Math. Vietnam. **10**, 35–57 (1985)

[8] Tuy, H.: A general deterministic approach to global optimization via d.c. programming. In: Hiriart-Urruty, J.B. (ed.) FERMAT days 85: mathematics for optimization, North-Holland Math. Stud., 129, pp. 273–303. North-Holland Publishing Co., Toulouse (1986)

[9] Tuy, H.: Convex programs with an additional reverse convex constraint. J. Optim. Theory Appl **52**, 463–486 (1987)

[10] Tuy, H.: On nonconvex optimization problems with separated nonconvex variables. J. Global Optim. **2**, 133–144 (1992)

[11] Tuy, H.: Convex analysis and global optimization. Kluwer Academic Publishers, Dordrecht (1998)

[12] Tuy, H., Tam, B.T.: Polyhedral annexation vs outer approximation for the decomposition of monotonic quasiconcave minimization problems. Acta Math. Vietnam. **20**, 99–114 (1995)

[13] Zhang, Q.H.: Outer approximation algorithms for DC programs and beyond. Ph.D. thesis, Università di Pisa (2008)

[14] Blanquero, R., Carrizosa, E., Hansen, P.: Locating objects in the plane using global optimization techniques. Math. Oper. Res. **34**, 837–858 (2009)

[15] Wozabal, D., Hochreiter, R., Pflug, G.C.: A difference of convex formulation of value-at-risk constrained optimization. OR. Spectrum. **3**, 377–400 (2010)

[16] Hu, J.: On linear programs with linear complementarity constraints. Ph.D. thesis, Rensselaer Polytechnic Institute (2008)

[17] Lu, Z.S., Zhang, Y.: Penalty Decomposition Methods for l_0-Norm Minimization. Computing Research Repository arxiv.org/abs/1105.2782, Cornell University Library (2010)

[18] Miller, A.: Subset Selection in Regression. Chapman and Hall, London (2002)

[19] Tropp, J.: Convex programming methods for identifying sparse signals in noise. IEEE T. Inform. Theory. **52**, 1030–1051 (2006)

[20] Han, L.S., Tiwari, A., Camlibel, M.K., Pang, J.S.: Convergence of time-stepping schemes for passive and extended linear complementarity systems. SIAM J. Numer. Anal. **47**, 3768–3796 (2009)

[21] Hu, J., Mitchell, J.E., Pang, J.S., Bennett, K.P., Kunapuli, G.: On the global solution of linear programs with linear complementarity constraints. SIAM J. on Optim. **19**, 445–471 (2008)

[22] Frangioni, A.: On a new class of bilevel programming problems and its use for reformulating mixed integer problems. Eur. J. Oper. Res. **82**, 615–646 (1995)

[23] Bigi, G., Frangioni, A., Zhang, Q.H.: Approximate optimality conditions and stopping criteria in canonical DC programming. Optim. Method. Softw. **25**, 19–27 (2010)

[24] Horst, R., Tuy, H.: Global optimization. Springer, Berlin (1990)