Outer Approximation Algorithms for Canonical DC Problems

Giancarlo Bigi [1] Antonio Frangioni Qinghua Zhang

Dipartimento di Informatica, Università di Pisa,
Largo B.Pontecorvo 3, 56127 Pisa, Italia

[1] Corresponding author: tel: +39 050 2213124; fax: +39 050 2212726. *e-mail address:* giancarlo.bigi@di.unipi.it.

Journal of Global Optimization manuscript No.
(will be inserted by the editor)

Outer Approximation Algorithms for Canonical DC Problems

Giancarlo Bigi · **Antonio Frangioni** ·
Qinghua Zhang

Received: 26 March 2008 / Accepted: date

Abstract The paper discusses a general framework for outer approximation type algorithms for the canonical DC optimization problem. The algorithms rely on a polar reformulation of the problem and exploit an approximated oracle in order to check global optimality. Consequently, approximate optimality conditions are introduced and bounds on the quality of the approximate global optimal solution are obtained. A thorough analysis of properties which guarantee convergence is carried out; two families of conditions are introduced which lead to design six implementable algorithms, whose convergence can be proved within a unified framework.

Keywords DC problems · polar set · approximate optimality conditions · cutting plane algorithms

1 Introduction

Nonconvex optimization problems often arise from applications in engineering, economics and other fields (see, for instance, [6,9]). Often, these problems either have a natural formulation or can be reformulated as *DC optimization* problems, that is nonconvex problems where the objective function is the difference of two convex functions and the constraint can be expressed as the set difference of two convex sets. In turn, every DC optimization problem can be reduced to the so-called *canonical DC* (shortly CDC) problem through standard transformations [16]. Several algorithms to solve it have been proposed [15,12,7,13,8,4]; many of them are modifications of the first cutting plane algorithm proposed by Tuy in [15].

Giancarlo Bigi · Antonio Frangioni · Qinghua Zhang
Dipartimento di Informatica, Università di Pisa, Largo B.Pontecorvo, 3, 56127 Pisa, Italia

Giancarlo Bigi
E-mail: giancarlo.bigi@di.unipi.it

Antonio Frangioni
E-mail: frangio@di.unipi.it

Qinghua Zhang
E-mail: zhang@di.unipi.it

In this paper, we consider the canonical DC problem relying on an alternative equivalent formulation based on a polar characterization of the constraint. We define a unified algorithmic framework for outer approximation type algorithms, which are based on an "oracle" for checking the global optimality conditions, and we study different sets of conditions which guarantee its convergence to an (approximated) optimal solution. As the oracle is the most computationally demanding part of the approach, we allow working with an *approximated oracle* which solves the related (nonconvex) optimization problem only approximately. Because of this, we provide an extensive analysis of approximate optimality conditions, which allow us to derive bounds on the quality of the obtained solution. Our analysis identifies two main classes of approaches, which give rise to six different implementable algorithms, four of which can't be reduced to the original cutting plane algorithm by Tuy and its modifications.

The paper is organized as follows. In Section 2 the *polar based* reformulation of the canonical DC problem is introduced, and the well-known optimality conditions are recalled. In Section 3 we propose a notion of approximate oracle and we define corresponding approximate optimality conditions, investigating the relationships between the exact optimal value and the approximate optimal values. In Section 4 a thorough convergence analysis is carried out for the "abstract" unified algorithmic framework, and then six *different implementable* algorithms are proposed which fit within the framework. Finally, in the last section the connections of these results with the existing algorithms in the literature are outlined.

2 The Canonical DC Problem

Throughout all the paper we focus on the canonical DC minimization problem

$$(CDC) \qquad \min\{\ dx \ \mid\ x \in \Omega \setminus \operatorname{int} C\ \}$$

where $\Omega \subseteq \mathbb{R}^n$ and $C \subseteq \mathbb{R}^n$ are full-dimensional closed convex sets, $d \in \mathbb{R}^n$ and dx denotes the scalar product between d and the vector of variables $x \in \mathbb{R}^n$.

The assumption on the dimension of the constraining sets is not restrictive. In fact, if Ω is not full-dimensional, the problem can be easily reformulated in the (affine) space generated by Ω. If C is not full-dimensional, then we have int $C = \emptyset$ and the problem is actually a convex minimization problem.

In order to avoid that (CDC) could be reduced to a convex minimization problem, we also suppose that the set C provides an essential constraint, i.e.

$$\min\{\ dx \ \mid\ x \in \Omega\ \} < \min\{\ dx \ \mid\ x \in \Omega \setminus \operatorname{int} C\ \}.$$

Relying on an appropriate translation, this assumption can be equivalently stated through the following two conditions

$$0 \in \operatorname{int} \Omega \cap \operatorname{int} C, \tag{1}$$

$$dx > 0 \qquad \forall x \in \Omega \setminus \operatorname{int} C. \tag{2}$$

Therefore, we assume that (1) and (2) hold. Notice that these assumptions guarantee that any feasible solution $x \in \Omega \setminus C$ provides a better feasible solution taking the unique intersection between the segment with 0 and x as end points and the boundary of C, i.e. $x' \in \mathrm{bd}\,(C) \cap (0, x)$ satisfies $dx' < dx$ where $\mathrm{bd}\,(C)$ denotes the boundary of C. As a consequence, all optimal solutions to (CDC) belong to the boundary of C.

In order to guarantee the existence of optimal solutions, we may assume the boundedness of the level sets

$$D(\gamma) := \{\ x \in \Omega \ \mid \ dx \leq \gamma\ \}$$

for the feasible values γ, i.e. those values $\gamma = dx \geq \gamma^*$ for some $x \in \Omega \setminus \mathrm{int}\, C$, where

$$\gamma^* := \min\{\ dx \ \mid \ x \in \Omega \setminus \mathrm{int}\, C\ \}.$$

Actually, such an assumption on the level sets is strictly related to the compactness of the reverse constraining set C as the following result shows.

Lemma 1 *Let γ be a feasible value.*

(i) If C is compact, then so is $D(\gamma)$.
(ii) If $D(\gamma)$ is compact, then

$$\gamma^* = \min\{\ dx \ \mid \ x \in \Omega \setminus \mathrm{int}\, \hat{C}\ \}$$

where $\hat{C} = C \cap B$ for any given compact set B such that $D(\gamma) \subseteq \mathrm{int}\, B$.

Proof (i) Assume by contradiction, suppose there exists a sequence $\{x^k\} \subseteq D(\gamma)$ such that $\|x^k\| \to +\infty$. Possibly taking a suitable subsequence, let $u = \lim_{k\to\infty} x^k \|x^k\|^{-1}$: clearly $du \leq 0$ and u belongs to the recession cone of Ω [10, Theorem 8.2]. Since $0 \in \Omega$ and C is bounded, there exists $\lambda > 0$ such that $x^0 = 0 + \lambda u \in \Omega \setminus C$. As $dx^0 \leq 0$, assumption (2) is contradicted.

(ii) Let $\bar{\gamma} := \min\{\ dx \ \mid \ x \in \Omega \setminus \mathrm{int}\, \hat{C}\ \}$. Since $\hat{C} \subseteq C$, then $\gamma^* \geq \bar{\gamma}$. Furthermore, $\gamma \geq \gamma^*$ and the compactness of $D(\gamma)$ guarantee the existence of $\bar{x} \in \Omega \setminus \mathrm{int}\, \hat{C}$ such that $\bar{\gamma} = d\bar{x}$. As $\mathrm{int}\, \hat{C} = \mathrm{int}\, C \cap \mathrm{int}\, B$ and $\bar{x} \in D(\gamma)$, then $\bar{x} \notin \mathrm{int}\, C$: $\bar{x}$ is feasible to (CDC) and therefore $\gamma^* \leq \bar{\gamma}$. □

Therefore, we assume that C is compact throughout all the paper. Moreover, this compactness assumption ensures existence of an optimal solution x^*, and therefore due to (2) we have $\gamma^* = dx^* > 0$, a property that will turn out to be very useful.

The level sets introduced above are also helpful to check whether a feasible value is optimal or not. In fact, it is straightforward that $\gamma = \gamma^*$ implies the following inclusion:

$$D(\gamma) \subseteq C. \tag{3}$$

Furthermore, it has been shown (see [23, Proposition 10]) that the necessary optimality condition (3) is also sufficient when problem (CDC) is *regular*, i.e.

$$\min\{\ dx \ \mid \ x \in \Omega \setminus \mathrm{int}\, C\ \} = \inf\{\ dx \ \mid \ x \in \Omega \setminus C\ \}. \tag{4}$$

The above regularity condition is strongly related to the existence of optimal solutions to (CDC) with additional properties (see the Lemma below). Furthermore, regularity will be exploited to prove that stopping criteria with finite tolerance yield approximate optimal solutions.

Lemma 2 *The regularity condition (4) holds if and only if (CDC) has an optimal solution $x^* \in \mathrm{bd}\,(\Omega \setminus C)$.*

Proof Given any optimal solution $x^* \in \mathrm{bd}\,(\Omega \setminus C)$, there exists a sequence $\{x^k\}$ such that $x^k \in \Omega \setminus C$ and $x^k \to x^*$; hence

$$\inf\{\ dx \ \mid\ x \in \Omega \setminus C\ \} \leq \lim_{k\to\infty} dx^k = dx^* = \min\{\ dx \ \mid\ x \in \Omega \setminus \mathrm{int}\, C\ \}.$$

As the reverse inequality always holds, the regularity condition (4) follows.

Vice versa, suppose the regularity condition (4) holds. Therefore, there exists a sequence $\{x^k\} \subseteq \Omega \setminus C$ such that $dx^k \downarrow \gamma^*$. By Lemma 1 the compactness of C guarantees that $D(\gamma)$ is compact for $\gamma = dx^1$. Therefore, the sequence $\{x^k\}$ admits at least one cluster point $x^* \in \mathrm{cl}\,(\Omega \setminus C)$. Since Ω is closed and $x^k \notin C$ for all k, we have $x^* \in \Omega$ and $x^* \notin \mathrm{int}\, C$. This implies that x^* is feasible and hence optimal as $dx^* = \gamma^*$. Since all optimal solutions belong to the boundary of C, then $x^* \notin \Omega \setminus C$ and therefore $x^* \in \mathrm{bd}\,(\Omega \setminus C)$. □

The constraint $x \notin$ int C is the source of nonconvexity in problem (CDC) and it is given just as a set relation. However, relying on the polarity between convex sets, we can express this nonconvex constraint in a different fashion. Let us recall that

$$C^* = \{\ w \in \mathbb{R}^n \ \mid\ wx \leq 1, \quad \forall x \in C\ \}$$

is the polar set of C and it is a closed convex set. Exploiting bipolarity relations (see, for instance, [10]), it is easy to check that the assumption $0 \in$ int C ensures that $x \notin$ int C if and only if $wx \geq 1$ for some $w \in C^*$. Therefore, problem (CDC) can be equivalently formulated as

$$\min\{\ dx \ \mid\ x \in \Omega,\ w \in C^*,\ wx \geq 1\ \} \tag{5}$$

where polar variables w have been introduced and the nonconvexity is given by the inequality constraint, which asks for some sort of reverse polar condition. Also, the assumption $0 \in$ int C ensures the compactness of C^*. The exploitation of polar variables will be an important tool to devise novel algorithms for (CDC) through its reformulation (5).

Relying on bipolarity relations, the optimality condition (3) can be equivalently stated in a polar fashion as

$$D(\gamma) \times C^* \subseteq \{\ (x, w) \in \mathbb{R}^n \times \mathbb{R}^n \ \mid\ wx \leq 1\ \} \tag{6}$$

while the regularity condition (4) reads

$$\min\{\ dx \mid x \in \Omega,\ w \in C^*,\ wx \geq 1\ \} = \inf\{\ dx \mid x \in \Omega,\ w \in C^*,\ wx > 1\ \}. \tag{7}$$

As an immediate consequence of (6), any optimal solution (x^*, w^*) to (5) satisfies both $x^* \in \mathrm{bd}\,(C)$ and $w^* x^* = 1$.

3 Approximate Optimality Conditions

Given a feasible value γ, the optimality condition (3) or (6) should be checked in order to recognize whether or not γ is the optimal value. Unfortunately, there is no known efficient way to check the inclusion between two sets. Yet, any exact algorithm for (CDC) or (5) must eventually cope with this problem.

3.1 Optimality Conditions and (Approximate) Oracles

In order to make (3), or equivalently (6), more readily approachable, we consider the following "optimization version" of the optimality conditions:

$$\max\{\ vz - 1\ \mid\ z \in D(\gamma)\ ,\ v \in C^*\ \}. \tag{8}$$

Obviously, (6) holds if and only if the optimal value $v(OC_\gamma)$ of (8) is less or equal to 0. Thus the above problem provides a way for checking optimality of a given value γ. Since the objective function of (8) is not concave, there are no known efficient approaches for this problem as well. However, checking (6) through the optimization problem (8) has the advantage of making it easy to define a proper notion of approximate optimality conditions.

A first way of approximating problem (8) is to replace $D(\gamma)$ and C^* with two convex sets S and Q, respectively, satisfying

$$C^* \subseteq Q, \tag{9}$$

$$D(\gamma) \subseteq S. \tag{10}$$

This is a standard step in cutting plane (outer approximation) approaches, where S and Q are chosen to be "easier" than the original sets (e.g. polyhedra with possibly few vertices or facets) and iteratively refined to become better and better approximations of $D(\gamma)$ and C^* as needed. Hence, one considers the following *relaxation* of (8):

$$\max\{\ vz - 1\ \mid\ z \in S,\ v \in Q\ \}, \tag{11}$$

whose optimal value $v(\overline{OC}_\gamma)$ provides an upper bound on $v(OC_\gamma)$; thus, the inequality $v(\overline{OC}_\gamma) \leq 0$ provides a convenient *sufficient* optimality condition for (5). If it does not hold, then either γ is not the optimal value, or S and Q are not "good" approximations of $D(\gamma)$ and C^*, respectively. All the cutting plane algorithms presented in this work follow the same basic scheme: (11) is solved and its solution is used to improve S, or Q, or γ, in such a way to guarantee convergence of γ to the optimal value. The focus of the research is on devising a number of different ways to achieve a convergent algorithm for (5) out of an "oracle" for (11). However, it is likely that in any such approach the solution of (11) is going to be the computational bottleneck; therefore, it makes sense to consider solving (11) *only approximately.*

Solving (11) approximately may actually mean two different things:

1. computing a "large enough" lower bound on $v(\overline{OC}_\gamma)$, i.e. finding a feasible solution $(\bar{z}, \bar{v})$ of (11) "sufficiently close" to the optimal solution;
2. computing a "small enough" upper bound $l \geq v(\overline{OC}_\gamma)$.

Algorithmically, the two notions correspond to two entirely different classes of approaches: lower bounds are produced by *heuristics* computing feasible solutions, while upper bounds are produced by solving suitable *relaxations* of (11), e.g. replacing the non-concave objective function vz with a suitable concave upper approximation. Exact algorithms combining the two can then be used to push the lower bound and the upper bound arbitrarily close together. However, for the sake of our approaches only one of the two bounds is needed at any given time. In fact, $v(\overline{OC}_\gamma)$ is either positive or non-positive. To establish that the first case holds amounts to finding a feasible solution

$(\bar{z}, \bar{v})$ to (11) such that $\bar{z}\bar{v} - 1 > 0$, while for the second case one needs an upper bound $l \leq 0$.

This is the rationale behind our definition of an *approximate oracle* for (11). In our development we will assume availability of a procedure Θ which, given S, Q, γ, and two positive tolerances ε and ε'

– *either* produces an upper bound

$$\varepsilon v(\overline{OC}_\gamma) \leq l \qquad \text{such that} \qquad l \leq \varepsilon' \tag{12}$$

– *or* produces a pair

$$(\bar{z}, \bar{v}) \in S \times Q \qquad \text{such that} \qquad \bar{v}\bar{z} - 1 \geq \varepsilon v(\overline{OC}_\gamma) > \varepsilon'. \tag{13}$$

Clearly, (13) corresponds to a pretty weak requirement about the way in which (11) is solved: a solution, which is optimal only with *fixed but arbitrary relative* tolerance $\varepsilon > 0$ and *absolute* tolerance ε', is required. Condition (12) allows the upper bound to be "small enough" but positive, rather than non-negative; this is taken as the stopping condition of the approach, and we will show that the positive tolerance allows for finite termination of the algorithms even when γ is optimal. The drawback is that a feasible value γ needn't be optimal when (12) holds: the next subsection is devoted to the study of the relationships between the "quality" of γ and the tolerances ε and ε'.

3.2 Approximate Optimality Conditions

The stopping criterion (12) implies $v(OC_\gamma) \leq \varepsilon'/\varepsilon$: the tolerances provide the upper bound $\delta = \varepsilon'/\varepsilon$ for the optimal value of (8). The values γ for which this upper bound holds are strictly related to the following approximated problem

$$\min\{\ dx \ \mid\ x \in \Omega,\ w \in C^*,\ wx \geq 1 + \delta\ \}, \tag{14}$$

which is obtained by perturbing the right-hand side of the nonconvex constraint in (5). Our analysis does not require any regularity assumption on (14) and it is based on the following quantity

$$\phi(\delta) := \inf\{\ dx \ \mid\ x \in \Omega,\ w \in C^*,\ wx > 1 + \delta\ \}.$$

Obviously, $\phi(\delta)$ may be greater than the optimal value of (14). Anyway, the value function ϕ provides the right tool to disclose the connections between γ, (12) and (14).

Proposition 1 *Let $\delta \geq 0$. Then, the following statements are equivalent:*

(i) $v(OC_\gamma) \leq \delta$;
(ii) $D(\gamma) \times C^* \subseteq \{\ (x, w) \in \mathbb{R}^n \times \mathbb{R}^n \ \mid\ wx \leq 1 + \delta\ \}$;
(iii) $\gamma \leq \phi(\delta)$.

Proof The equivalence between (*i*) and (*ii*) follows immediately from the definition of $v(OC_\gamma)$. Analogously, (*ii*) implies (*iii*) by the definition of $\phi(\delta)$.

Suppose (*ii*) does not hold: there exist $x \in D(\gamma)$ and $w \in C^*$ such that $wx > 1 + \delta$. Take any $t \in (0, 1)$ large enough to have $w(tx) > 1 + \delta$. Since $0 \in \Omega$, the convexity of Ω implies $tx \in \Omega$; obviously $d(tx) < dx \leq \gamma$. Therefore, (tx, w) guarantees $\phi(\delta) < \gamma$ contradicting (*iii*). □

Considering the optimal value of (14) as γ in Proposition 1, we get that (*ii*) is a necessary optimality condition for (14). Furthermore, if the problem is regular (i.e. $\phi(\delta)$ is actually the optimal value), it is also sufficient. Choosing $\delta = 0$, the known optimality conditions for (5) follow too. Therefore, inclusion (*ii*) can be considered as an approximate optimality condition for (5). It is easy to check that (*ii*) is equivalent to the inclusion $D(\gamma) \subseteq (1+\delta)C$: perturbing the right-hand side of the nonconvex constraint in (5) corresponds to perturbing the reverse constraining set C in (CDC). As an immediate consequence of the proposition, we also have

$$\phi(\delta) = \sup\{ \ \gamma \ \mid \ D(\gamma) \times C^* \subseteq \{ \ (x,w) \in \mathbb{R}^n \times \mathbb{R}^n \ \mid \ wx \leq 1+\delta \ \} \ \}.$$

The stopping criterion (*i*) guarantees approximate optimality and condition (*iii*) provides the adequate tool to evaluate the quality of the approximation. In fact, supposing (5) to be regular, i.e. $\gamma^* = \phi(0)$, we have that

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

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 2 *The value function ϕ is right-continuous at 0, i.e.*

$$\lim_{\delta \downarrow 0} \phi(\delta) = \phi(0).$$

Proof Clearly ϕ is nondecreasing, 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 exist $\bar{\gamma} = \lim_{\delta \downarrow 0} \phi(\delta)$ and $\bar{\gamma} \geq \phi(0)$. Since $\bar{\gamma} \leq \phi(\delta)$ for any $\delta > 0$, Proposition 1 implies $v(OC_{\bar{\gamma}}) \leq \delta$ for any $\delta > 0$. Since $v(OC_{\bar{\gamma}})$ does not depend upon δ, we get $v(OC_{\bar{\gamma}}) \leq 0$. Therefore, Proposition 1 guarantees $\bar{\gamma} \leq \phi(0)$. □

Although the approximation always converges to the optimal value, the rate of convergence may be less than linear as the following example shows.

Example 1 Consider (14) with $n = 2$, $d = (-1,2)$, $\Omega = \{x \in \mathbb{R}^2 \mid -2 \leq x_1 \leq 0.1, x_1 + 2x_2 + 2 \geq 0\}$, and $C = \{ \ x \in \mathbb{R}^2 \ \mid \ x_1^2 + x_2^2 \leq 4 \ \}$. It is easy to check that (14) is regular for any $\delta \geq 0$ and that

$$(x^*(\delta), w^*(\delta)) = \big((-2, 2\sqrt{(1+\delta)^2 - 1}), (\frac{-1}{2(1+\delta)}, \frac{1}{2}\sqrt{1 - \frac{1}{(1+\delta)^2}})\big)$$

is an optimal solution to (14) for δ small enough. Therefore, we have $\phi(\delta) = 4\sqrt{(1+\delta)^2 - 1} + 2$ and

$$\lim_{\delta \downarrow 0} \ [\phi(\delta) - \phi(0)]/\delta = \lim_{\delta \downarrow 0} \ 4\sqrt{1 + 2/\delta} = +\infty.$$

Thus, regularity is not enough to achieve a linear rate of convergence. Additional assumptions on the problem are needed: the existence of an optimal solution with some particular properties guarantees the Lipschitz behavior of ϕ.

Proposition 3 *If there exists an optimal solution (x^*, w^*) to (5) such that*

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

for some direction $u \in \mathbb{R}^n$, then 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 Let $\bar{\lambda} > 0$ be such that $x^* + \bar{\lambda} u \in \Omega$; the convexity of Ω implies $x(\lambda) := x^* + \lambda u \in \Omega$ for any $\lambda \in [0, \bar{\lambda}]$; furthermore, $w^*(x^* + \lambda u) = 1 + \lambda w^* u > 1$ if $\lambda > 0$ since optimality implies $w^* x^* = 1$. Thus, the sequence $(x(\lambda), w^*)$ shows that the regularity condition (7) holds. Therefore, we have $\phi(0) = dx^*$.

Chosen $\bar{\delta} := (w^* u/2)\bar{\lambda}$, let us consider $y(\delta) := x(2\delta/w^* u)$ for any $\delta \in (0, \bar{\delta}]$: we have $y(\delta) \in \Omega$ and

$$w^* y(\delta) = w^* x^* + (2\delta/w^* u) w^* u = 1 + 2\delta > 1 + \delta,$$

where the last equality holds. Therefore, $(y(\delta), w^*)$ provides an upper bound for $\phi(\delta)$, i.e. $\phi(\delta) \leq dy(\delta)$. Finally, we get

$$\phi(\delta) - \phi(0) \leq dy(\delta) - dx^* = (2du/w^* u)\delta. \qquad \square$$

Though regularity has not been explicitly required for (5), the assumption on the optimal solution implies it. A geometric view of this assumption can be achieved relying on the (Bouligand) tangent cone of C at x^*, namely the set

$$T(C, x) := \{\ u \in \mathbb{R}^n \ \mid\ \exists t_n \downarrow 0,\ u_n \to u \ \text{ s.t. } x + t_n u_n \in C\ \},$$

and its following characterization.

Lemma 3 *Let $x^* \in \mathrm{bd}\,(C)$. Then, the following statements are equivalent:*

(i) $u \in T(C, x^*)$;

(ii) $wu \leq 0$ *for all* $w \in C^*$ *such that* $wx^* = 1$.

Proof Take any $u \in T(C, x^*)$: there exist $t_n \downarrow 0$ and $u_n \to u$ such that $x^* + t_n u_n \in C$. Therefore, we have $w(x^* + t_n u_n) \leq 1$ for any $w \in C^*$. If $wx^* = 1$, we get $wu_n \leq 0$ and taking the limit $wu \leq 0$.

Vice versa, suppose u satisfies (ii) but $u \notin T(C, x^*)$. Since the tangent cone is a closed set, there exists $\varepsilon > 0$ such that $\hat{u} = u - \varepsilon x^* \notin T(C, x^*)$. Consider any $t_n \downarrow 0$ and $u_n \to \hat{u}$ such that $x^* + t_n u_n \notin C$. Therefore, there exist $w_n \in C^*$ such that $w_n(x^* + t_n u_n) > 1$. Assumption (1) implies that C^* is compact (see, for instance, [10, Corollary 14.5.1]). Thus, we can suppose $w_n \to \bar{w}$ for some $\bar{w} \in C^*$. Taking the limit in the above inequality, we get $\bar{w}x^* \geq 1$ and therefore $\bar{w}x^* = 1$. Since $t_n w_n u_n > 1 - w_n x^* \geq 0$, we also get $\bar{w}\hat{u} \geq 0$. The assumption on u guarantees also $\bar{w}u \leq 0$. Therefore, we get the contradiction $0 \leq \bar{w}\hat{u} = \bar{w}(u - \varepsilon x^*) \leq -\varepsilon$. $\square$

The following characterization allows to formulate the assumption of Proposition 3 in a geometric fashion.

Proposition 4 *Let $x^* \in \mathrm{bd}\,(C)$. Then, the following statements are equivalent:*

(i) *there exist* $w^* \in C^*$ *and* $u \in \mathbb{R}^n$ *such that* $w^* x^* = 1$ *and (15) holds;*

(ii) $T(\Omega, x^*) \nsubseteq T(C, x^*)$.

Proof Suppose (ii) does not hold and take any $w^* \in C^*$ and $u \in \mathbb{R}^n$ such that $w^* x^* = 1$ and $x^* + \bar{\lambda} u \in \Omega$ for some $\bar{\lambda} > 0$. The convexity of Ω implies $\Omega \subseteq x^* + T(\Omega, x^*)$ and therefore $\bar{\lambda} u \in T(\Omega, x^*) \subseteq T(C, x^*)$. By Lemma 3 we get $w^* u \leq 0$: hence (i) does not hold.

Vice versa, take any $u \in T(\Omega, x^*) \setminus T(C, x^*)$. Lemma 3 implies that there exists $w^* \in C^*$ such that $w^* x^* = 1$ and $w^* u > 0$. As $u \in T(\Omega, x^*)$, there exist $t_n \downarrow 0$ and $u_n \to u$ such that $x^* + t_n u_n \in \Omega$; if n is large enough, we also have $w^* u_n > 0$. Thus, w^* and u_n satisfy (15). $\square$

It is worth to note that (ii) depends upon x^* only. Indeed, the original formulation of the canonical DC problem does not have polar variables. Anyway, x^* is an optimal solution to (CDC) if and only if (x^*, w^*) is an optimal solution to (5) for any $w^* \in C^*$ such that $w^* x^* = 1$. As a consequence, Propositions 3 and 4 lead to the main result of the section.

Theorem 1 *If there exists an optimal solution (x^*, w^*) to (5) such that $T(\Omega, x^*) \nsubseteq T(C, x^*)$, then ϕ is locally Lipschitz at 0.*

The assumption on the tangent cones can be considered as a *strong regularity* condition. In fact, it implies regularity but they are not equivalent, as the problem of Example 1 shows for $\delta = 0$. Anyway, when C is a polyhedron, strong regularity collapses to regularity.

Theorem 2 *Suppose C is a polyhedron. Then, (5) is regular if and only if there exists an optimal solution (x^*, w^*) to (5) such that $T(\Omega, x^*) \nsubseteq T(C, x^*)$.*

Proof Suppose (5) is regular: Lemma 2 implies the existence of an optimal solution (x^*, w^*) to (5) such that $x^* \in \mathrm{bd}\,(\Omega \setminus C)$. Suppose $T(\Omega, x^*) \subseteq T(C, x^*)$. Since C is a polyhedron, there exists $\varepsilon > 0$ such that

$$[x^* + T(C, x^*)] \cap B(x^*, \varepsilon) = C \cap B(x^*, \varepsilon).$$

Since the convexity of Ω implies $\Omega \subseteq x^* + T(\Omega, x^*)$, we have

$$\Omega \cap B(x^*, \varepsilon) \subseteq C \cap B(x^*, \varepsilon)$$

in contradiction with $x^* \in \mathrm{bd}\,(\Omega \setminus C)$.

The if part follows from Proposition 4 and the proof of Proposition 3. □

Corollary 1 *Suppose C is a polyhedron. If (5) is regular, then ϕ is locally Lipschitz at 0.*

4 Conditions and Algorithms

In this section we present several algorithms which (approximately) solve (CDC) through its reformulation (5) if an approximated oracle Θ is available. We first establish a hierarchy of abstract conditions ensuring convergence; then, for each set of conditions we propose actual implementable procedures which realize it.

4.1 General Convergence Conditions

All the algorithms will follow the generic cutting plane scheme sketched in the previous section. More in details, a non increasing sequence of feasible values $\{\gamma^k\}$ is produced, and the oracle Θ is called for each γ^k, thereby producing either a value l^k such that condition (12) holds, or points z^k and v^k such that (13) are satisfied. By repeatedly calling the oracle, we can construct a procedure which either proves that γ^k satisfies condition (12) or produces a better feasible value $\gamma^{k+1} < \gamma^k$. In the latter case, γ^{k+1} is associated to (produced by) points x^k and w^k such that

$$x^k \in C, \quad w^k \in C^*, \quad w^k x^k = 1, \tag{16}$$

which implies also $(x^k, w^k) \in \mathrm{bd}\,(C) \times \mathrm{bd}\,(C^*)$. In fact, if $x^k \in \mathrm{int}\ C$(analogous to $w^k \in \mathrm{int}\ C^*$), then $w^k x^k < \max\{w^k x \mid x \in C\} \leq 1$ (see [10, Theorem 13.1]). The rationale for (16) is that any optimal solution must satisfy these conditions.

It must be stressed that the above conditions do not require $x \in \Omega$ and therefore (x^k, w^k) may be infeasible for the polar reformulation (5). Anyway, (5) can be equivalently stated as

$$\min\{\ \zeta(w)\ \mid\ w \in C^*\ \} \tag{17}$$

where

$$\zeta(w) = \min\{\ \theta(x)\ \mid\ wx \geq 1\ \}$$

and

$$\theta(x) = \begin{cases} dx & \text{if } x \in \Omega \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore, the polar variable w^k is always feasible for (17), though it may be $\theta(x^k) = +\infty$. Since $\zeta(w) \leq \theta(x)$ for all pairs (x, w) satisfying (16), we can choose $\gamma^{k+1} = \zeta(w^k)$ whenever $x^k \notin \Omega$. As $\zeta(w^k)$ is the optimal value of a convex problem, it can be assumed to be efficiently computable. Moreover, if γ^{k+1} turns out to be optimal, then w^k is the "polar part" of an optimal solution: in fact any

$$\bar{x} \in \operatorname{argmin}\{\ dx\ \mid\ x \in \Omega,\ w^k x \geq 1\ \}$$

provides the complementary x part of the optimal solution.

Thus, a given pair (x^k, w^k) can provide two (potentially) different feasible values: $\theta(x^k)$ which is essentially costless to compute but may be infinite, and $\zeta(w^k)$ whose computation requires the solution of a convex program. In general one may want to avoid the computation of $\zeta(w^k)$ unless it is strictly necessary; to allow a general treatment we will in the following indicate with $\gamma(x, w)$ a function taking a pair (x, w) satisfying (16) and producing a feasible value. Which of the two possible implementations (θ and ζ) is required will be discussed in the context of each implementable algorithm.

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 (3) holds, then γ^k is the optimal value: stop;
2. Select (x^k, w^k) satisfying (16) such that $\gamma^{k+1} = \gamma(x^k, w^k) < \gamma^k$;
   set $k = k + 1$; goto 1.

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

$$0 \leq \lim_{k \to \infty} \gamma^k = \gamma^\infty < \cdots < \gamma^{k+1} < \gamma^k < \cdots < \gamma^1.$$

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

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

Obviously, Algorithm 1 is too general to deduce any meaningful property; something more has to be said:

1. how exactly the optimality condition (3) is checked,
2. how (x^k, w^k) such that $\gamma(x^k, w^k) < \gamma^k$ is selected once one knows that (3) is not fulfilled.

The two points are strictly interwoven: finding (x^k, w^k) such that $\gamma(x^k, w^k) < \gamma^k$ immediately proves that γ^k is not optimal; vice versa, assume that we have any constructive procedure that produces a point $z^k \in D(\gamma^k)\backslash C$ when γ^k is not optimal: there exists $w^k \in C^*$ such that $w^k z^k > 1$ and $x^k = (w^k z^k)^{-1} z^k$ satisfies both $x^k \in D(\gamma^k)$ and $\gamma(x^k, w^k) \leq dx^k < dz^k \leq \gamma^k$.
Then, a first question is if such a method provides a convergent algorithm; not surprisingly, without further qualification the answer is negative.

Example 2 Consider (5) with $n = 2$, $d = (0, 1)$ and the sets

$$\Omega = \{ \ x \in \mathbb{R}^2 \ \mid \ -1.8 \leq x_1 \leq 1.96, \ x_2 \geq -0.1 \ \}, \quad C = \{ \ x \in \mathbb{R}^2 \ \mid \ x_1^2 + x_2^2 \leq 4 \ \};$$

therefore, we have

$$C^* = \{ \ w \in \mathbb{R}^2 \ \mid \ 4(w_1^2 + w_2^2) \leq 1 \ \}.$$

Starting from any value $\gamma^1 > 0.87$ and applying the above procedure, we can find the sequences $z^k = (-1.8, \gamma^{k-1})$, $x^k = 2z^k/||z^k||$ and $w^k = z^k/2||z^k||$, which lead to a non-optimal solution $(x^\infty, w^\infty) \approx ((-1.8, 0.87), (-1.8, 0.87)/4)$, whereas the optimal solution is $(x^*, w^*) \approx ((1.96, 0.4), (1.96, 0.4)/4)$.

Thus, some care is needed in choosing the sequence w^k in Algorithm 1, as well as the accompanying sequences z^k and x^k if the mechanism illustrated above is to be used. Actually, our "more implementable" approximate optimality condition based on (8) indicates that a fourth sequence v^k, which "is to w^k what z^k is to x^k", should be taken into account as well. In fact, we propose the following general assumptions under which convergence can be proved:

$$v^k z^k - 1 \geq \varepsilon \ \max\{ \ vz - 1 \ \mid \ (z, v) \in D(\gamma^k) \times C^* \ \} \tag{18}$$

$$\liminf_{k \to \infty} \ v^k z^k \leq 1 \tag{19}$$

where $\varepsilon \in (0, 1)$. Condition (18) basically says that v^k and z^k must be produced by some process attempting to solve the nonconvex problem (8) for $\gamma = \gamma^k$, although the process may be "terminated early" due to the optimality tolerance ε. Condition (19) 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, (18) and (19) are sufficient to guarantee convergence to the optimal value.

Proposition 5 *If conditions* (18) *and* (19) *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

$$v^k z^k - 1 \geq \varepsilon \max\{ \; vz - 1 \;\; | \;\; (z,v) \in D(\gamma^\infty) \times C^* \; \}$$

for all k. Taking the limit, (19) implies

$$\max\{ \; vz - 1 \;\; | \;\; (z,v) \in D(\gamma^\infty) \times C^* \; \} \leq 0,$$

and therefore γ^∞ is the optimal value. □

When developing a "concrete" algorithm for (CDC), the abstract condition (19) shouldn't be directly imposed on the sequences $\{z^k\}$ and $\{v^k\}$. In fact, these are the results of a "complex" optimization process, i.e. approximately solving (8), 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 condition (19). One possibility is to rely on sequences of points x^k and w^k, which satisfy one of these pairs of conditions:

$$\begin{cases} \limsup\limits_{k\to\infty} \; v^k(z^k - x^k) \leq 0 & \text{(a)} \\ \\ \limsup\limits_{k\to\infty} \; v^k x^k \leq 1 & \text{(b)}, \end{cases} \tag{20}$$

$$\begin{cases} \limsup\limits_{k\to\infty} \; (v^k - w^k)z^k \leq 0 & \text{(a)} \\ \\ \limsup\limits_{k\to\infty} \; w^k z^k \leq 1 & \text{(b)}. \end{cases} \tag{21}$$

Both pairs of conditions clearly imply (19).

Lemma 4 *If either (20) or (21) hold, then (19) holds.*

Therefore, we can define the two sets of conditions which, separately, guarantee convergence of Algorithm 1:

$$B_1 \equiv (18) \wedge (20) \qquad\qquad B_2 \equiv (18) \wedge (21).$$

Though they look highly symmetric to each other, we will show that B_1 and B_2 are by no means equivalent. In fact, we will propose several different sets of conditions (in particular, four for B_1 and two for B_2) which imply one of them, and develop implementable subprocedures that attain these conditions, thereby defining six different *implementable algorithms.*

4.2 The Outer Approximation Machinery

As addressed in Section 3, one way to make (8) more tractable is to replace $D(\gamma)$ and C^* with two "simpler" convex sets Q and S such that $C^* \subseteq Q$ and $D(\gamma) \subseteq S$. Of course, this requires some appropriate machinery to update S and Q in order to make them "good enough" approximations of Ω and C^*. Convexity of both sets allows to rely on cutting procedures based on standard separation tools. In fact, the result below follows readily from the general Basic Outer Approximation Theorem [5, Theorem II.1].

Theorem 3 *Let r be a convex function such that $R = \{\ x \in \mathbb{R}^n \ \mid\ r(x) \leq 0\ \}$ satisfies $0 \in \operatorname{int} R$. Let $\{R^k\}$ be a sequence of convex sets and $\{x^k\}$ be a sequence of points which satisfy the following conditions:*

1. $x^k \in R^k \backslash R$
2. $R^{k+1} = R^k \cap \{\ x \in \mathbb{R}^n \ \mid\ p^k(x - y^k) + r(y^k) \leq 0\ \}$ *where* $p^k \in \partial r(y^k)$ *for some* $y^k \in [0, x^k) \setminus \operatorname{int} R$.

Then, any cluster point $\bar{x}$ of the sequence $\{x^k\}$ belongs to $\operatorname{bd}(R)$.

Theorem 3 suggests the standard cutting-plane procedure described in Subprocedure 1: it takes a "simple" representation S, typically a polyhedron, of the convex set R and a point x which proves the two are different; it "improves" S to a representation of R which does not contain x, and still is a polyhedron if S is, by intersecting S with a separating hyperplane which cuts off x but no point in R. Due to Theorem 3, iterating this process leads, in the limit, to a point in R; in other words, S becomes an "arbitrarily close" representation of R near a cluster point.

Subprocedure 1 Cutting-Plane subprocedure

Input: a closed convex set $R = \{\ x \in \mathbb{R}^n \ \mid\ r(x) \leq 0\ \}$ such that $0 \in \operatorname{int} R$,
a closed convex set S such that $R \subseteq S$ and a point $x \in S \backslash R$.

1. Select a point $y \in (0, x) \cap \operatorname{bd}(R)$ and a sub-gradient $p \in \partial r(y)$.
2. Set $S = S \cap \{\ x \ \mid\ p(x - y) + r(y) \leq 0\ \}$.

Output: S.

It is worth remarking that condition $0 \in \operatorname{int} R$ is required to ensure that $y \neq x$, and therefore that the hyperplane actually separates R and x strictly. In our setting, the condition is satisfied for $D(\gamma)$: this is due to (1) and to the fact that $\gamma \geq \gamma^* > 0$, itself a consequence of the boundedness of C as discussed in Section 2. Boundedness of C is also equivalent to $0 \in \operatorname{int} C^*$; therefore, the condition is *a fortiori* true for S and Q, the sets Subprocedure 1 will be called upon, due to (10) and (9), respectively.

4.3 A Generic Outer Approximation Subprocedure

We can now define a generic outer approximation procedure which, only provided with an approximate oracle Θ, allows implementations of Algorithm 1 which attain the convergence conditions introduced in Subsection 4.1. We call this a "generic" outer approximation procedure because it depends on two parameters: a selection rule Ψ for the x and w variables, and a stopping criterion Υ. In this subsection we will describe the properties of the subprocedure which are independent of the choices of Ψ and Υ; later on, we will show several different possible choices for these, leading to different implementable algorithms.

Conditions (10) and (9) guarantee that $D(\gamma)$ and C^* are included in S^i and Q^i, respectively, for $i = 1$. The cutting-plane Subprocedure 1 ensures this is still true for any i and therefore we get the following "non-increasing" sequences of sets:

$$D(\gamma) \subseteq \cdots \subseteq S^{i+1} \subseteq S^i \subseteq \cdots \subseteq S^1,$$

$$C^* \subseteq \cdots \subseteq Q^{i+1} \subseteq Q^i \subseteq \cdots \subseteq Q^1.$$

Subprocedure 2 Outer Approximation subprocedure

Input: Q and S, closed convex sets satisfying (9) and (10), a feasible value γ.

0. $S^1 = S$; $Q^1 = Q$; $i = 1$;
1. Call the oracle Θ for S^i, Q^i, γ. If the oracle produces an upper bound l^i satisfying condition (12), then stop.
2. Otherwise, Θ produces (z^i, v^i) satisfying (13); Select (x^i, w^i) satisfying (16) and condition Ψ;
3. If $z^i \notin D(\gamma)$ then use Subprocedure 1 with $D(\gamma)$, S^i and z^i to get S^{i+1}; else $S^{i+1} = S^i$;
4. If $v^i \notin C^*$ then use Subprocedure 1 with C^*, Q^i and v^i to get Q^{i+1}; else $Q^{i+1} = Q^i$;
5. If stopping criterion Υ holds then stop. else $i = i + 1$; goto 1.

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

We can now prove the basic properties of Subprocedure 2, which are independent of the choice of Ψ and Υ.

Lemma 5 *If Subprocedure 2 never ends, then all the cluster points of $\{z^i\}$ and $\{v^i\}$ belong to $D(\gamma)$ and C^*, respectively.*

Proof Subprocedure 2 generates two sequences of points $\{z^i\}$ and $\{v^i\}$ such that $z^i \in S^i$, $v^i \in Q^i$, and the hypotheses of Theorem 3 are satisfied; hence, all the cluster points of $\{z^i\}$ and $\{v^i\}$ belong to $D(\gamma)$ and C^*, respectively. □

It will be crucial to ensure that the sequences $\{z^i\}$ and $\{v^i\}$ do indeed have cluster points. As both $D(\gamma)$ and C^* are assumed to be compact, it is very natural to suppose also that

$$\{z^i\} \text{ and } \{v^i\} \text{ are bounded.} \tag{22}$$

In fact, this condition holds, for instance, if S^1 and Q^1 are compact, which is not a restrictive assumption as $D(\gamma)$ and C^* are compact too. Therefore, from now onwards we suppose that (22) always holds. Note that the sequences $\{x^i\}$ and $\{w^i\}$ are always bounded as due to (16) they belong to bounded sets C and C^*, respectively.

Corollary 2 *If $\varepsilon' > 0$, and Subprocedure 2 never ends, then no cluster point of $\{z^i\}$ belongs to C.*

Proof By Lemma 5 all cluster points of $\{v^i\}$ belong to C^* and (22) guarantees that at least one exists. If there were a cluster point of $\{z^i\}$ in C, one would have that $\liminf_{i\to\infty} v^i z^i \leq 1$ in contradiction with $v^i z^i - 1 > \varepsilon'$, which is guaranteed by the oracle for any i (c.f. (13)). □

Proposition 6 *If $\varepsilon' > 0$ and $D(\gamma) \subseteq C$, then Subprocedure 2 stops after a finite number of iterations.*

Proof Suppose Subprocedure 2 never ends; due to (22), the sequence $\{(z^i, v^i)\}$ has at least a cluster point which belongs to $D(\gamma) \times C^*$ by Lemma 5. Since $D(\gamma) \subseteq C$, then all the cluster points actually belong to $C \times C^*$: therefore, we have $\liminf_{i\to\infty} v^i z^i \leq 1$ which yields a contradiction as in Corollary 2. □

The above proofs show the need of requiring $\varepsilon' > 0$, since for $\varepsilon' = 0$ the subprocedure may never stop. In other words, Subprocedure 2 can not *finitely* prove that the optimal value is optimal; this is why it is relevant to clarify the relationship between approximated optimal values and the optimal value.

Finally, it is useful to remark that while condition (19) is characteristic of optimizing sequences, it holds for every fixed γ by substituting x^i to z^i, even if $\varepsilon' = 0$.

Lemma 6 *If Subprocedure* 2 *never ends, then* $\limsup_{i\to\infty} v^i x^i \leq 1$.

Proof Lemma 5 guarantees that all the cluster points of $\{v^i\}$ belong to C^*. Since $x^i \in C$ for all i, the thesis follows immediately □

The subprocedure can then be used to define implementable versions of the Prototype Algorithm 1.

Algorithm 2 Implementable Outer Approximation Algorithm

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

Some remarks on Algorithm 2 are in order:

- Since $D(\gamma^k) \subseteq S^k$ and $C^* \subseteq Q^k$, (13) guarantees that condition (18) is always satisfied by all possible variants of the algorithm, i.e. irrespective of the concrete choices for Ψ and Υ;
- at Step 4, the obvious possibility for Q^{k+1} and S^{k+1} is to set them as the Q^i and S^i produced by Subprocedure 2; since this leads to accumulation in Q^k and S^k of *all* cutting planes generated along the iterates, and therefore possibly to "large" descriptions of Q^k and S^k;
- which implementation of $\gamma(x^k, w^k)$ has to be chosen depends on the properties of the points x^k and w^k (see Table 1 in Subsection 4.6) and therefore ultimately on Ψ.

The following subsections are devoted to the study of which conditions Ψ and Υ result in a convergent Algorithm 2.

4.4 Algorithms Exploiting the Set of Conditions B_1

While the oracle in Subprocedure 2 guarantees (18), condition (20) has to be achieved through additional properties. The algorithms of this subsection will require (20b) more or less directly and will obtain (20a) by imposing (21b) and one extra condition, which simply requires x^k and z^k to be collinear:

$$z^k = \mu_1^k x^k \qquad \text{for some } \mu_1^k > 0. \tag{23}$$

Lemma 7 *If (23) holds for all k, then (21b) implies (20a).*

Proof Due to (23) and $w^k x^k = 1$, (21b) reads $\limsup_{k\to\infty} \mu_1^k \leq 1$, thus we have

$$\limsup_{k\to\infty} v^k(z^k - x^k) = \limsup_{k\to\infty} (\mu_1^k - 1)v^k x^k \leq 0$$

where the inequality is due to boundedness of the sequences $\{v^k\}$ and $\{x^k\}$. □

All algorithms in this subsection will exploit condition (23). Together with (16), this forces to choose $x^k \in \{\alpha z^k \mid \alpha \geq 0\} \cap \mathrm{bd}\,(C)$, thereby basically making the choice of x^k automatic once z^k is known. Note that the intersection is nonempty due to boundedness of C, and therefore x^k is always well defined.

The easiest way to guarantee that the sequences generated by Algorithm 2 satisfy (23) is to impose that z^i and x^i are always collinear in Subprocedure 2. Furthermore, this allows to prove that Subprocedure 2 either attains a decrease of the objective function or detects approximate optimality in a finite number of steps, provided that $dz^i \leq \gamma$.

Lemma 8 *Suppose $S^1 \subseteq \{ z \in \mathbb{R}^n \mid dz \leq \gamma \}$ and set*

$$\Psi \;\equiv\; [z^i = \mu_1^i x^i \quad \textit{with} \quad \mu_1^i > 0].$$

If $\varepsilon' > 0$ and Subprocedure 2 never ends, then it produces iterates satisfying $x^i \in (0, z^i) \cap \Omega$, $z^i \notin C$ and $\gamma(x^i, w^i) < \gamma$ for sufficiently large i.

Proof Lemma 5 guarantees that all the cluster points of $\{z^i\}$ and $\{v^i\}$ belong to $D(\gamma)$ and C^*, and Corollary 2 guarantees that each cluster point $\bar{z}$ of $\{z^i\}$ does not belong to C, therefore $\bar{z} \in \Omega \backslash C$. Thus, there exists $\bar{x} \in (0, \bar{z})$ such that $\bar{x}$ is a cluster point of $\{x^i\}$. By eventually taking the appropriate subsequences, suppose $z^i \to \bar{z}$ and $x^i \to \bar{x}$. All the above implies that $x^i \in (0, z^i)$ and $z^i \notin C$ for all sufficiently large i. Since $0 \in \mathrm{int}\,\Omega$ and $\bar{z} \in \Omega$, we have also $\bar{x} \in \mathrm{int}\,\Omega$ and therefore, $x^i \in \Omega$ for all sufficiently large i. Hence, we have $\gamma(x^i, w^i) \leq dx^i < dz^i \leq \gamma$ as $z^i \in S^i \subseteq S^1$. □

The assumption on S^1 is actually a mild assumption on how S^k is updated in Algorithm 2: it is enough to keep the "objective cut" $dz \leq \gamma^k$ among the inequalities which define S^k and update it at each iteration to the current value of γ^k. Furthermore, this assumption implies that the membership test in Step 3 of Subprocedure 2 can be reduced to $z^i \notin \Omega$.

Some of the properties guaranteed by the above Lemma can be exploited in the stopping criterion Υ. Anyway, in order to guarantee that the decrease guaranteed by Subprocedure 2 under (23) is "sufficient", one has to prove also that the set of conditions B_1 is satisfied: this requires (20), which will be achieved through (20b) and (21b). In the next subsections we develop four different ways in which this can be done.

4.4.1 Algorithm C_1

The first possibility, directly inspired by the algorithms already proposed in the literature (see, for instance, [23]), is to resort to the following conditions:

$$dz^k \leq \gamma^k, \tag{24}$$

$$x^k \in (0, z^k) \cap \Omega \cap \mathrm{bd}\,(C). \tag{25}$$

Condition (25) implies (23) with $\mu_1^k > 1$. Actually, the two conditions are equivalent if $z^k \notin C$ and $x^k \in \Omega$ (since we always have $x^k \in \mathrm{bd}\,(C)$); anyway we don't ask for these two conditions. As (25) guarantees that the sequence of points $\{x^k\}$ is feasible, we can set $\gamma(x^k, w^k) = dx^k$.

Lemma 9 *If $\gamma^* > 0$ and (24), (25) hold for all k, then (21b) holds.*

Proof Since x^h is feasible, we have

$$dx^0 - \sum_{k=1}^{h}(dx^{k-1} - dx^k) = dx^h \geq \gamma^*$$

and therefore

$$dx^0 - \gamma^* \geq \sum_{k=1}^{h}(dx^{k-1} - dx^k) \geq \sum_{k=1}^{h}(dz^k - dx^k)$$

where the last inequality holds since (24) reads $dz^k \leq \gamma^k = dx^{k-1}$. Taking the limit, we get

$$\lim_{h\to+\infty}\sum_{k=1}^{h}(dz^k - dx^k) \leq dx^0 - \gamma^* < +\infty.$$

Since $\mu_1^k > 1$, (25) implies $dz^k - dx^k > 0$ and therefore we get $dz^k - dx^k = (\mu_1^k - 1)dx^k \to 0$, which implies that $\lim_{k\to\infty}\mu_1^k = 1$ since the feasibility of x^k gives $dx^k \geq \gamma^* > 0$. Therefore, we have

$$\limsup_{k\to\infty}\ w^k z^k = \limsup_{k\to\infty}\ \mu_1^k w^k x^k = \limsup_{k\to\infty}\ \mu_1^k \leq 1$$

since (16) guarantees $w^k x^k = 1$. □

Therefore, we can define the following set of conditions

$$C_1 \ \equiv\ (18) \wedge (20b) \wedge (24) \wedge (25)$$

which implies B_1 and thus guarantees convergence for Algorithm 2. The proper choice of Ψ and Υ ensures that these conditions are finitely attained within Subprocedure 2 except (20b), which requires the knowledge of the entire sequences generated by Algorithm 2. Therefore, we consider a positive sequence $\sigma^k \to 0$ and ask for the subprocedure to provide points v^i and x^i such that

$$v^i x^i \leq 1 + \sigma^k.$$

This condition can be considered an appropriate formulation of (20b) within Subprocedure 2 as in this way Algorithm 2 will surely satisfy (20b).

Proposition 7 *Suppose $S^1 \subseteq \{\ z \in \mathbb{R}^n \ \mid\ dz \leq \gamma\ \}$ and set*

$$\Psi \ \equiv\ [z^i = \mu_1^i x^i \quad \textit{with} \quad \mu_1^i > 0], \qquad \Upsilon \ \equiv\ [x^i \in \Omega] \ \wedge\ [v^i x^i \leq 1 + \sigma^k].$$

If $\varepsilon' > \sigma^k > 0$, then Subprocedure 2 ends in a finite number of steps; if it stops at Step 5, it reports points x^i, w^i, z^i and v^i satisfying the set of conditions C_1.

Proof Lemma 8 and Lemma 6 guarantee that the stopping criterion Υ will be satisfied for i large enough, independently from the choice of σ^k. Therefore, Subprocedure 2 ends in a finite number of steps. Suppose it ends at Step 5. The stopping criterion Υ directly guarantees (20b); (18) holds as all iterates satisfy (13); (24) follows immediately from the assumption on S^1 as $S^i \subseteq S^1$. Finally, the stopping criterion Υ and (13) allow to get

$$0 < v^i x^i \leq 1 + \sigma^k < 1 + \varepsilon' \leq v^i z^i = \mu_1^i v^i x^i$$

which implies $\mu_1^i > 1$ and thus we have $z^i \notin C$. Therefore, $x^i \in (0, z^i) \cap \mathrm{bd}\,(C)$ and hence (25) holds since the stopping criterion Υ provides $x^i \in \Omega$. □

For this algorithm to work, the sequence $\{\sigma^k\}$ has to be defined explicitly, either a-priori or dynamically as it is used to stop Subprocedure 2. Unlike most algorithms in the literature, it is not needed to require $\mu_1^i > 1$ at every iteration *within the subprocedure*, thus leaving a wider freedom of choice.

4.4.2 Algorithm C_2

An alternative way to obtain (20b) is to require

$$v^k x^h \leq 1 \qquad \text{for all } h < k. \tag{26}$$

Lemma 10 *If (26) holds for all* k*, then (20b) holds.*

Proof Assume by contradiction, suppose $v^k x^k > 1 + \delta$ for infinitely many k and a given $\delta > 0$. Since $\{v^k\}$ and $\{x^k\}$ are bounded, we can suppose $v^k \to \bar{v}$ and $x^k \to \bar{x}$ (eventually taking the appropriate subsequences). Condition (26) implies that $\bar{v} x^h \leq 1$ for all h and therefore $\bar{v}\bar{x} \leq 1$, a contradiction. □

Therefore, we can define the set of conditions

$$C_2 \;\equiv\; (18) \wedge (24) \wedge (25) \wedge (26)$$

which implies C_1 and therefore B_1, thus ensuring convergence for Algorithm 2.

Clearly, condition (26) is guaranteed if

$$Q^k \subseteq \bigcap_{h<k} \{\ v \in \mathbb{R}^n \ \mid\ v x^h \leq 1\ \}. \tag{27}$$

This can be easily achieved updating Q^{k+1} in Step 4 of Algorithm 2 as follows:

$$Q^{k+1} = Q^i \cap \{\ v \in \mathbb{R}^n \ \mid\ v x^i \leq 1\ \}, \tag{28}$$

where Q^i and x^i are those produced at the end Subprocedure 2.

Lemma 11 *If (28) holds, then* $C^* \subseteq Q^{k+1}$*.*

Proof Subprocedure 2 guarantees $C^* \subseteq Q^i$. If we consider the support function of C, namely

$$\sigma_C(v) := \max\{\ vx \ \mid\ x \in C\ \},$$

then we have

$$C^* = \{\ v \in \mathbb{R}^n \ \mid\ \sigma_C(v) - 1 \leq 0\ \}.$$

Since (16) guarantees $x^i \in C$, any $v \in C^*$ satisfies $v x^i \leq \sigma_C(v) \leq 1$. □

In this way all the inequalities produced by the Subprocedure 2 are kept: the "quality" of Q^{k+1} may improve, reducing the number of iterations required to stop the subprocedure, but it is likely to increase the cost of each iteration; the practical impact of this trade-off could be gauged only experimentally. In any case, in (28) it is always possible to replace Q^i with Q^k or any intermediate Q^j produced by the subprocedure since they both contain C^*.

Again, an implementable version of the Algorithm 2 can be obtained by choosing Ψ and Υ properly.

Proposition 8 *Set*

$$\Psi \;\equiv\; [z^i = \mu_1^i x^i \quad with \;\; \mu_1^i > 0], \qquad \Upsilon \;\equiv\; [x^i \in \Omega] \;\wedge\; [z^i \notin C].$$

If $\varepsilon' > 0$ and (27) holds, then Subprocedure 2 ends in a finite number of steps; if it stops at Step 5, it reports points x^i, w^i, z^i and v^i satisfying the set of conditions C_2.

Proof Analogous to that of Proposition 7, considering that (26) follows from (27) and that $x^i \in \Omega$ and $z^i \notin C$ imply (25). □

4.4.3 Algorithm C_3

Lemma 10 states that condition (20b) is implied by condition (26) under our boundedness assumptions. Symmetrically, we can prove the following result in the same way.

Lemma 12 *If*

$$z^k w^h \leq 1 \quad \textit{for all } h < k \tag{29}$$

hold for all k, then (21b) holds.

Therefore, we can define the set of conditions

$$C_3 \;\equiv\; (18) \wedge (20b) \wedge (23) \wedge (29)$$

which implies B_1 (and thus guarantees convergence for Algorithm 2) as (23) and (29) imply (20a) by combining Lemmas 12 and 7.

Clearly, (29) is guaranteed if

$$S^k \subseteq \bigcap_{h<k} \{\, z \in \mathbb{R}^n \;\mid\; w^h z \leq 1 \,\}. \tag{30}$$

This is easily obtained, for instance, by implementing Step 4 of Algorithm 2 as

$$S^{k+1} = S^i \cap \{\, z \in \mathbb{R}^n \;\mid\; w^i z \leq 1 \,\} \tag{31}$$

where S^i and w^i are those produced at the end Subprocedure 2, but it is always possible to replace S^i with S^k or any intermediate S^j produced by the subprocedure. Anyway, the current value has to be updated through ζ in order to guarantee that S^{k+1} outer approximates $D(\gamma^{k+1})$.

Lemma 13 *Suppose $\gamma(x, w) = \zeta(w)$. If (31) is used in Algorithm 2, then $D(\gamma^k) \subseteq S^k$ for all k.*

Proof The proof is by induction on the iterate index k. If $k = 1$, the thesis is guaranteed by the choice of the input data. Suppose the thesis holds for a given k and there exists $\bar{x} \in D(\gamma^{k+1})$ such that $\bar{x} \notin S^{k+1}$: we have

$$\bar{x} \in D(\gamma^{k+1}) \subseteq D(\gamma^k) \subseteq S^i$$

where the last inclusion is guaranteed by the way Subprocedure 2 updates S^k. Therefore, (31) implies $w^i\bar{x} > 1$. Since $\bar{x} \in \Omega$, then $\hat{x} := (w^i\bar{x})^{-1}\bar{x} \in \Omega$ (as $w^i\bar{x} > 1$ and $0 \in \Omega$). Moreover, $w^i\hat{x} = 1$ and therefore $\gamma^{k+1} \leq d\hat{x} < d\bar{x}$ providing the contradiction $\bar{x} \notin D(\gamma^{k+1})$. □

Again, an implementable version of Algorithm 2 can be obtained by choosing Ψ and Υ properly. Note that the correctness of this version requires $\gamma(x, w) = \zeta(w)$; besides, there is no guarantee that x^k is feasible.

Proposition 9 *Set*

$$\Psi \;\equiv\; [z^i = \mu_1^i x^i \quad \textit{with} \quad \mu_1^i > 0], \qquad \Upsilon \;\equiv\; [\zeta(w^i) < \gamma] \;\wedge\; [v^i x^i \leq 1 + \sigma^k],$$

If $\varepsilon', \sigma^k > 0$ and (30) holds, then Subprocedure 2 ends in a finite number of steps; if it stops at Step 5, it reports points x^i, w^i, z^i and v^i satisfying the set of conditions C_3.

Proof Analogous to that of Proposition 7, considering that (23) comes by Ψ and that (29) is implied by (30). □

Like Algorithm C_1, one has to use a sequence σ^k converging to zero explicitly; in this case, however, it is *not* required $\sigma^k < \varepsilon'$, at least initially.

4.4.4 Algorithm C_4

The sets of conditions C_2 and C_3 are two independent modifications of C_1; the specific update (28) for Q^{k+1} is exploited for the former, while the "symmetric" update (31) for S^{k+1} is exploited for the latter. The two modifications can be combined: the set of conditions

$$C_4 \;\equiv\; (18) \wedge (26) \wedge (23) \wedge (29)$$

implies B_1 thanks to Lemmas 10, 12 and 7, thus ensuring convergence for Algorithm 2. The following result provides an implementable version of the algorithm.

Proposition 10 *Set*

$$\Psi \;\equiv\; [z^i = \mu_1^i x^i \quad \textit{with} \quad \mu_1^i > 0], \qquad \Upsilon \;\equiv\; [\zeta(w^i) < \gamma].$$

If $\varepsilon' > 0$, (27) and (30) hold, then Subprocedure 2 ends in a finite number of steps; if it stops at Step 5, it reports points x^i, w^i, z^i and v^i satisfying the set of conditions C_4.

4.5 Algorithms Exploiting the Set of Conditions B_2

The algorithms of this subsection need (21) instead of (20). As (21b) has been exploited to achieve (20a), symmetrically (21a) can be obtained through (20b), relying on the "polar counterpart" of (23), namely

$$v^k = \mu_2^k w^k \qquad \text{for some } \mu_2^k > 0. \tag{32}$$

Together with (16), this forces to choose $w^k \in \{\ \alpha v^k \ \mid \ \alpha \geq 0\ \} \cap \operatorname{bd}(C^*)$, thereby basically fixing w^k once v^k is known. Note that this intersection is always nonempty since C^* is compact.

Lemma 14 *If (32) holds for all k, then (20b) implies (21a).*

Proof Due to (32) and $w^k x^k = 1$, (20b) reads $\limsup_{k\to\infty} \ \mu_2^k \leq 1$, thus we have

$$\limsup_{k\to\infty} \ (v^k - w^k)z^k = \limsup_{k\to\infty} \ (\mu_2^k - 1)w^k z^k \leq 0$$

where the inequality is due to the boundedness of $\{z^k\}$ and $\{w^k\}$. □

The algorithms of this subsection will exploit (32). The easiest way to guarantee that the sequences generated by Algorithm 2 satisfy it is to impose that w^i and v^i are always collinear in Subprocedure 2.

Lemma 15 *Suppose $S^1 \subseteq \{\ z \in \mathbb{R}^n \ \mid \ dz \leq \gamma\ \}$ and set*

$$\Psi \ \equiv \ [v^i = \mu_2^i w^i \quad \textit{with} \quad \mu_2^i > 0].$$

If $\varepsilon' > 0$ and Subprocedure 2 never ends, then it produces iterates satisfying $\zeta(w^i) < \gamma$ for sufficiently large i.

Proof Taking the appropriate subsequences, we can suppose $w^i \to \bar{w}$, $v^i \to \bar{v}$ and $z^i \to \bar{z}$. The collinearly assumption Ψ implies that $\bar{v} = \bar{\mu}\bar{w}$ for some $\bar{\mu} \geq 0$ and condition (13) guarantees $\bar{\mu} \neq 0$. Lemma 5 guarantees $\bar{v} \in C^*$; since $w^i \in \operatorname{bd}(C^*)$, we have $\bar{w} \in \operatorname{bd}(C^*)$ and thus $\bar{\mu} \in (0, 1]$. Therefore, we have

$$\lim_{i\to\infty} \ w^i z^i = \bar{w}\bar{z} = \bar{\mu}^{-1}\bar{v}\bar{z} \geq \lim_{i\to\infty} \ v^i z^i \geq 1 + \varepsilon'.$$

where the last inequality is due to (13). Therefore, $w^i z^i \geq 1 + \varepsilon'/2$ holds for all sufficiently large i. By Lemma 5 we have $\bar{z} \in \Omega$; since $0 \in \operatorname{int}\Omega$, we get $\bar{z}^i := (1 + \varepsilon'/2)^{-1} z^i \in \Omega$ for all sufficiently large i. Hence, we have $\zeta(w^i) \leq d\bar{z}^i < dz^i \leq \gamma$ as $w^i \bar{z}^i \geq 1$ and $z^i \in S^i \subseteq S^1$. □

Using the above results, we can develop versions of Algorithm 2, which are "symmetric" to those that rely on the set of conditions B_1. However, the polar reformulation (5) is asymmetric in the sense that only the "original"variables x appear in the objective function. Therefore, only two of those four algorithms can be mirrored in this case. Specifically, we will develop sets of conditions D_1 and D_2 corresponding to C_3 and C_4, respectively. No algorithms corresponding to C_1 and C_2 can be devised since they should exploit the condition

$$w^k \in (0, v^k) \cap C^* \cap \operatorname{bd}(\Omega^*),$$

which is "symmetric" to (25). However, it would imply the existence of an optimal solution (x^*, w^*) such that $w^* \in C^* \cap \operatorname{bd}(\Omega^*)$, which is not necessarily true: if you consider (5) with $n = 1$, $d = 1$ and $\Omega = C^* = [-1/2, 4]$, the unique optimal point is $(x^*, w^*) = (1/4, 4)$ while $C^* \cap \operatorname{bd}(\Omega^*) = 1/4$.

4.5.1 Algorithm D_1

We can define the set of conditions

$$D_1 \equiv (18) \wedge (20b) \wedge (29) \wedge (32)$$

in a "symmetric" way with respect to C_3. Due to Lemmas 12 and 14, D_1 implies B_2 and therefore it ensures convergence for Algorithm 2. An implementable version can be obtained by choosing Ψ and Υ as follows.

Proposition 11 *Set*

$$\Psi \equiv [v^i = \mu_2^i w^i \quad \textit{with} \quad \mu_2^i > 0], \qquad \Upsilon \equiv [\zeta(w^i) < \gamma] \;\wedge\; [v^i x^i \leq 1 + \sigma^k],$$

If $\varepsilon', \sigma^k > 0$ and (30) holds, then Subprocedure 2 ends in a finite number of steps; if it stops at Step 5, it reports points x^i, w^i, z^i and v^i satisfying the set of conditions D_1.

4.5.2 Algorithm D_2

We can define the set of conditions

$$D_2 \equiv (18) \wedge (26) \wedge (29) \wedge (32)$$

in a "symmetric" way with respect to C_4. Due to Lemmas 10, 12 and 14, D_2 implies B_2 and therefore it ensures convergence for Algorithm 2. An implementable version can be obtained by choosing Ψ and Υ as follows.

Proposition 12 *Set*

$$\Psi \equiv [v^i = \mu_2^i w^i \quad \textit{with} \quad \mu_2^i > 0], \qquad \Upsilon \equiv [\zeta(w^i) < \gamma].$$

If $\varepsilon' > 0$, (27) and (30) hold, then Subprocedure 2 ends in a finite number of steps; if it stops at Step 5, it reports points x^i, w^i, z^i and v^i satisfying the set of conditions D_2.

4.6 Summary

We have developed six different implementable versions of Algorithm 2: while they are all based on Subprocedure 2, they differ for the stopping criterion Ψ, the condition Υ on the iterations, how the evaluation function γ is implemented and how S^k and Q^k are updated. All the considered variants are summarized in Table 1.

	Ψ	Υ	γ	Q^k	S^k
C_1	$z^i = \mu_1^i x^i,\ \mu_1^i > 0$	$x^i \in \Omega \ \wedge\ v^i x^i \leq 1 + \sigma^k$	θ		
C_2	$z^i = \mu_1^i x^i,\ \mu_1^i > 0$	$x^i \in \Omega \ \wedge\ z^i \notin C$	θ	(28)	
C_3	$z^i = \mu_1^i x^i,\ \mu_1^i > 0$	$\zeta(w^i) < \gamma^k \ \wedge\ v^i x^i \leq 1 + \sigma^k$	ζ		(31)
C_4	$z^i = \mu_1^i x^i,\ \mu_1^i > 0$	$\zeta(w^i) < \gamma^k$	ζ	(28)	(31)
D_1	$v^i = \mu_2^i w^i,\ \mu_2^i > 0$	$\zeta(w^i) < \gamma^k \ \wedge\ v^i x^i \leq 1 + \sigma^k$	ζ		(31)
D_2	$v^i = \mu_2^i w^i,\ \mu_2^i > 0$	$\zeta(w^i) < \gamma^k$	ζ	(28)	(31)

Table 1 Summary of implementable versions of Algorithm 2

Now, we want to show that all these algorithms are indeed different, in the sense that they can produce different optimizing sequences even if the same instance and the same starting conditions are given. To this aim, we consider problem (CDC) with $n = 2$, $d = (0, 1)$ and

$$\Omega = \{ \ x \in \mathbb{R}^2 \ \mid \ -1 \le x_1 \le 2, \ -1 \le x_2 \le 5, \ 3x_1 - x_2 \le 4 \ \},$$

$$C = \{ \ x \in \mathbb{R}^2 \ \mid \ x_1^2 + x_2^2 \le 4 \ \}.$$

Notice that Ω is a bounded polyhedron, whose vertices provide the alternative description

$$\Omega = conv\Big(\{ \ (1,-1), \ (-1,-1), \ (-1,5), \ (2,5), \ (2,2) \ \}\Big).$$

It is easy to check that the unique optimal solution is the intersection between the segment $[(1,-1),(2,2)]$ (the boundary of the constraint $3x_1 - x_2 \le 4$) and the boundary of C, namely the point $x^* = (6+\sqrt{6}, 3\sqrt{6}-2)/5 \in \Omega \setminus \operatorname{int} C$. Therefore, the optimal value is $\gamma^* = (3\sqrt{6}-2)/5 \approx 1.0697$. Note that all standard assumptions are satisfied: (1) and (2) hold, C is compact while regularity follows from Lemma 2. Furthermore, the value function ϕ is locally Lipschitz at 0, as $(0,\delta) \in T(\Omega, x^*)$ and $(0,\delta) \notin T(C, x^*)$ for any $\delta > 0$ (see Theorem 1).

Considering the polar reformulation (5), we have

$$C^* = \{ \ w \in \mathbb{R}^2 \ \mid \ 4(w_1^2 + w_2^2) \le 1 \ \}.$$

Since any optimal solution of (5) must satisfy $w^* x^* = 1$ and $w^* \in \operatorname{bd}(C^*)$, we have that $w^* = (6+\sqrt{6}, 3\sqrt{6}-2)/20$ provides the only possibility for the polar part of the optimal solution.

In the following, we assume the oracle Θ to always choose the same (z, v) when S, Q and γ are the same; furthermore, we set $\varepsilon = 1$ so that the pairs (z, v) satisfying (13) must actually be optimal for (11). In this way, we eliminate the nondeterminism due to the fact that the oracle may return different $\varepsilon-$optimal solutions of (11), which may be "many" especially if $\varepsilon << 1$; nonetheless, the six algorithms all construct different optimizing sequences for this instance.

Consider the following starting situation:

$$\sigma^1 = 0.1, \quad \gamma^1 = +\infty, \quad Q^1 = [-1/2, 1/2] \times [-1/2, 1/2],$$

$$S^1 = \{ \ x \in \mathbb{R}^2 \ \mid \ -1 \le x_1 \le 2, \ -1 \le x_2 \le 10, \ 3x_1 - x_2 \le 4 \ \}$$

$$= \operatorname{conv}\Big(\{ \ (1,-1), \ (-1,-1), \ (-1,10), \ (2,10), \ (2,2) \ \}\Big).$$

All algorithms start call Subprocedure 2 with S^1, Q^1 and γ^1 as input data. The oracle provides an optimal solution of the certificate problem

$$\max \{ \ vz - 1 \ \mid \ (z, v) \in S^1 \times Q^1 \ \},$$

which can be easily obtained comparing the value $\bar{v}\bar{z}$ for all pairs where $\bar{z}$ is an extreme point of S^1 and $\bar{v}$ is an extreme point of Q^1. In this case, the unique optimal solution turns out to be $(z^1, v^1) = ((2, 10), (1/2, 1/2))$ with optimal value $v(\overline{OC}_{\gamma^1}) = 5$; thus, according to our assumptions, this is the pair the oracle Θ returns for all algorithms.

Algorithms implementing the set of conditions B_1. The four algorithms C_1, C_2, C_3, and C_4 ask for $x^i$and z^i to be collinear. Due to (16) the only possible choice is $x^1 = (2, 10)/\sqrt{26}$; since we have both $z^1 \notin C$ and $x^1 \in \Omega$, then the point satisfies also the more restrictive condition (25). Due to (16) the only choice for the corresponding polar point is $w^1 = (1, 5)/\sqrt{104}$.

The subprocedure stops at this first iteration for algorithms C_2 and C_4, since we have $x^1 \in \Omega$, $z^1 \notin C$ and $\zeta(w^1) \leq dx^1 < \gamma^1$. On the contrary, it does not stop for algorithms C_1 and C_3 since

$$v^1x^1 = 6/\sqrt{26} \approx 1.1767 > 1 + \sigma^1.$$

In algorithm C_2 the subprocedure provides the new current value $\gamma^2 = \theta(x^1) = dx^1 = 10/\sqrt{26} \approx 1.9612$ while in algorithm C_4 it provides γ^2 as

$$\zeta(w^1) = \min\{\ dx \ \mid \ x \in \Omega,\ x_1 + 5x_2 \geq \sqrt{104}\ \}.$$

The optimal solution of the above linear program is $\bar{x}^1 = (10 + \sqrt{26}, 3\sqrt{26} - 2)/8$ and therefore the current value will be updated to

$$\gamma^2 = \zeta(w^1) = d\bar{x}^1 = (3\sqrt{26} - 2)/8 \approx 1.6621 < 10/\sqrt{26}.$$

As for algorithms C_1 and C_3, the subprocedure performs one more iteration after the sets S^1 and Q^1 have been updated through subprocedure 1 (since $z^1 \notin \Omega$ and $v^1 \notin C^*$):

$$S^2 = S^1 \cap \{\ (x \in \mathbb{R}^2 \ \mid \ x_2 \leq 5\ \} = \Omega,$$

$$Q^2 = Q^1 \cap \{\ w \in \mathbb{R}^2 \ \mid \ \sqrt{2}(w_1 + w_2) \leq 1\ \}.$$

At the second iteration of the subprocedure the oracle returns the (unique) optimal solution of the certificate problem

$$\max\ \{\ vz - 1\ \mid (z, v)\ \in S^2 \times Q^2\ \},$$

which is $(z^2, v^2) = ((2, 5), (\sqrt{2} - 1, 1)/2)$. Therefore, the collinearity condition Ψ and (16) imply $x^2 = (4, 10)/\sqrt{29}$ and $w^2 = (2, 5)/2\sqrt{29}$. Since $x^2 \in \Omega$, $\zeta(w^2) \leq dx^2 < \gamma^1$ and

$$v^2x^2 = (3 + 2\sqrt{2})/\sqrt{29} \approx 1.0823 \leq 1 + \sigma^1,$$

the subprocedure stops: algorithm C_1 selects $\gamma^2 = \theta(x^2) = dx^2 = 10/\sqrt{29} \approx 1.6569$ while algorithm C_3 solves the linear program

$$\zeta(w^2) = \min\{\ dx \ \mid \ x \in \Omega,\ 2x_1 + 5x_2 \geq 2\sqrt{29}\ \}$$

in order to get the point $\bar{x}^2 = (20 + 2\sqrt{29}, 6\sqrt{29} - 8)/17)$ and set $\gamma^2 = \zeta(w^2) = d\bar{x}^2 = (6\sqrt{29} - 8)/17 \approx 1.4301$.

The four algorithms have all provided different values for γ^2 and therefore they are different from each other.

Algorithms implementing the set of conditions B_2. The algorithms D_1 and D_2 require w^i and v^i to be collinear. Due to (16) the only possible choice is $w^1 = (1,1)/2\sqrt{2}$ and the corresponding point in the original space can be only $x^1 = (\sqrt{2}, \sqrt{2})$. The subprocedure stops at this first iteration for algorithm D_2, since we have $x^1 \in \Omega$ and therefore $\zeta(w^1) \leq dx^1 < \gamma^1$. On the contrary, it does not stop for algorithm D_1 since

$$v^1 x^1 = \sqrt{2} \approx 1.4142 > 1 + \sigma^1.$$

In algorithm D_2 the subprocedure provides the new current value γ^2 as

$$\zeta(w^1) = \min\{\ dx \ \mid\ x \in \Omega,\ x_1 + x_2 \geq 2\sqrt{2}\ \} = (3 - \sqrt{2})/\sqrt{2} \approx 1.1213.$$

and the corresponding optimal solution $\bar{x}^1 = (1 + \sqrt{2}, 3 - \sqrt{2})/\sqrt{2}$ is the best achieved point. Since this value for γ^2 is different from all those seen so far, D_2 is yet another different algorithm.

In algorithm D_1 the subprocedure performs a second iteration after the sets S^1 and Q^1 have been updated exactly in the same way as in algorithms C_1 and C_3 (since z^1 and v^1 are indeed the same). Therefore, the oracle provides the same $z^2 = (2, 5)$ and $v^2 = (\sqrt{2} - 1, 1)/2$. Due to the collinearity condition Ψ and (16), we get $w^2 = (\sqrt{2} - 1, 1)/2\sqrt{4 - 2\sqrt{2}}$ and $x^2 = \sqrt{2 - \sqrt{2}}(1, 1 + \sqrt{2})$. Since

$$v^2 x^2 = \sqrt{4 - 2\sqrt{2}} \approx 1.0824 \leq 1 + \sigma^1$$

the subprocedure ends. The value it returns as γ^2 is

$$\zeta(w^2) = \min\{\ dx \ \mid\ x \in \Omega,\ (\sqrt{2} - 1)x_1 + x_2 \geq 2\sqrt{4 - 2\sqrt{2}}\ \} \approx 1.4169.$$

and the corresponding optimal solution

$$\bar{x}^2 = \left(\frac{4 + 2\sqrt{4 - 2\sqrt{2}}}{2 + \sqrt{2}}, \frac{4 + 6\sqrt{4 - 2\sqrt{2}} - 4\sqrt{2}}{2 + \sqrt{2}} \right)$$

is the best achieved point. Once again, this value for γ^2 is different from all previous ones: all the six algorithms are different.

5 Comparisons and Conclusions

The algorithms proposed in this paper are inspired by the seminal works of Tuy [15, 16], in which the canonical DC problem has been introduced, it has been shown how any *DC* problem can be reduced to it, and the first cutting plane algorithm has been proposed. The initial algorithm had less refined convergence properties; by cutting off points such that $dx > \gamma^k - \alpha$, for a feasible tolerance $\alpha \geq 0$, the algorithm may terminate with only an α-optimal solution. More refined versions of the algorithms, more akin to those presented in this paper, were presented later. The *polyhedral annexation method*, proposed in [21, 25] for the special case of (CDC) where Ω is a polyhedron, is the first where the exact form

$$v^k z^k \geq \max\{\ vz \ \mid\ (z, v) \in D(\gamma^k) \times C^*\ \}$$

of the approximate optimality conditions (12) (see also Proposition 1) has been introduced; afterwards, [27,23] showed that this algorithm can be extended to any (CDC) problem. In [22], the non-slackened "objective cut" (24) was introduced, and $\gamma^1 = +\infty$ was first allowed. A further variant was developed in [17] for the "more general" case where dx is replaced by a convex finite-valued function $f(x)$ although this can also be recast as a canonical DC program.

Several attempts at generalizing the results in the above papers were not entirely successful. A variant of [17] has been proposed in [7], where a binary search on the value of γ is proposed; this, however, is unnecessary. The algorithm proposed in [13], a modified form of the ones in [15,16], as well as its modified form in [5], were later shown not to guarantee convergence [22]. Similarly, a counter example disproving convergence was developed in [3] for the cutting plane algorithms of [2,1]. Finally, the analogous algorithm of [11], based on a slightly modified form of the classical optimality condition (3), was also shown not to be always convergent [14]; besides, the modified optimality condition is not easier to check than (3).

All the converging algorithms in the above papers satisfy the set of conditions C_1 or C_2, and are special cases of those presented in this paper. Furthermore, it is basically given for granted that the"oracle" for checking the optimality conditions is realized through enumeration of vertices. The contributions of the present paper are the following:

- The introduction of "approximate oracle" conditions (12)–(13), which are designed to allow for more sophisticated and efficient solution procedures, with respect to pure vertex enumeration, to tackle the problem of checking the optimality condition, arguably the computational bottleneck in this type of approaches.
- A thorough study of the impact of approximations in the optimality conditions onto the quality of the approximately optimal solutions satisfying them.
- Full exploitation of the "primal-polar" formulation of the optimality conditions based on (8) in order to derive a very general hierarchy of conditions ensuring convergence.
- A general algorithmic scheme based on the developed hierarchy which gives rise to six different implementable algorithms, four of which (C_3, C_4, D_1 and D_2) do not seem to have previously been considered in the literature; each of these algorithms can generate an approximate optimal value in a finite number of steps, where the error can be managed and controlled.

It may be worth remarking that the "new" algorithms C_3, C_4, D_1 and D_2 all use $\gamma(x, w) = \zeta(w)$. This has been inspired by the reformulation of (CDC) as the *quasi-concave minimization* problem (17) already proposed in [26]. However, in that paper a "cut and split" method was used, that is entirely different from the outer approximation algorithms proposed in this paper. Indeed, that method belongs to the main other family of algorithms for canonical DC problems, that of *branch and bound* methods (see, for instance, [18,19,20]). So, this research has shown how concepts developed for one family of approaches can be useful even for an entirely unrelated one.

While this paper seems to offer a quite comprehensive convergence theory for "oracle-based" outer approximation algorithms for canonical DC programs, much still needs to be done before these algorithms become widely used and accepted as those based on the branch and bound paradigm. In particular, more work is needed to identify practically efficient ways to implement the oracle, at least on special types of canonical

DC programs in which the sets Ω and C have some form of exploitable structure; this will be the focus of further research.

References

1. Ben Saad, S.: A new cutting plane algorithm for a class of reverse convex 0-1 integer programs. In: C.A. Floudas, Pardalos, P.M. (eds.): Recent Advances in Global Optimization, pp. 152–164. Princeton University Press, Princeton (1992)
2. Ben Saad, S., Jacobsen, S.E.: A level set algorithm for a class of reverse convex programs. Ann. Oper. Res. 25, 19–42 (1990)
3. Ben Saad, S., Jacobsen, S.E.: Comments on a reverse convex programming algorithm. J. Global Optim. 5, 95–96 (1994)
4. Fulop, J.: A finite cutting plane method for solving linear programs with an additional reverse constraint. European J. Oper. Res. 44, 395–409 (1990)
5. Horst, R., Tuy, H.: Global Optimization. Springer, Berlin (1990)
6. Horst, R., Pardalos, P.M. (eds.): Handbook of Global Optimization. Kluwer Academic Publishers, Dordrecht (1995)
7. Nghia, M.D., Hieu, N.D.: A method for solving reverse convex programming problems. Acta Math. Vietnam. 11, 241–252 (1986)
8. Pham, D.T., El Bernoussi, S.: Numerical methods for solving a class of global nonconvex optimization problems. International Series of Numerical Mathematics 87, 97–132 (1989)
9. Pintér, J.D. (ed.): Global Optimization: Scientific and Engineering Case Studies. Springer, Berlin (2006)
10. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
11. Strekalovsky, A.S., Tsevendorj, I.: Testing the $\mathbb{R}$-strategy for a reverse convex problem. J. Global Optim. 13, 61–74 (1998)
12. Thach, P.T.: Convex programs with several additional reverse convex constraints. Acta Math. Vietnam. 10, 35–57 (1985)
13. Thoai, N.V.: A modified version of Tuy's method for solving d.c. programming problems. Optimization 19, 665–674 (1988)
14. Tuan, H.D.: Remarks on an algorithm for reverse convex programs. J. Global Optim. 16, 295–297 (2000)
15. Tuy, H.: Global minimization of a difference of two convex functions. Math. Programming Studies 30, 150–182 (1987)
16. 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, pp. 273–303. North-Holland, Amsterdam (1986)
17. Tuy, H.: Convex programs with an additional reverse convex constraint. J. Optim. Theory Appl. 52, 463–486 (1987)
18. Tuy, H., Horst, R.: Convergence and restart in branch-and-bound algorithms for global optimization. Application to concave minimization and D.C. optimization problems. Math. Programming 41, 161-183 (1988)
19. Tuy, H.: Normal conical algorithm for concave minimization over polytopes. Math. Programming 51, 229-245 (1991)
20. Tuy, H.: Effect of the subdivision strategy on convergence and efficiency of some global optimization algorithms. J. Global Optim. 1. 23–36 (1991)
21. Tuy, H.: On nonconvex optimization problems with separated nonconvex variables. J. Global Optim. 2, 133–144 (1992)
22. Tuy, H.: Canonical DC programming problem: outer approximation methods revisited. Oper. Res. Lett. 18, 99–106 (1995)
23. Tuy, H.: D.C. optimization: theory, methods and algorithms. In: Horst, R., Pardalos, P.M. (eds.): Handbook of Global Optimization, pp. 149–216. Kluwer Academic Publishers, Dordrecht (1995)
24. Tuy, H.: Convex Analysis and Global Optimization. Kluwer Academic Publishers, Dordrecht (1998)
25. Tuy, H., Al-Khayyal, F.A.: Global optimization of a nonconvex single facility location problem by sequential unconstrained convex minimization. J. Global Optim. 2, 61–71 (1992)
26. Tuy, H., Migdalas, A., Varbrand, P.: A quasiconcave minimization method for solving linear two-level programs. J. Global Optim. 4, 243–263 (1994)
27. 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)