A Study of the Difference-of-Convex Approach for Solving Linear Programs with Complementarity Constraints

Francisco Jara-Moroni* Jong-Shi Pang† Andreas Wächter‡

February 11, 2016

Abstract

This paper studies the difference-of-convex (DC) penalty formulations and the associated difference-of-convex algorithm (DCA) for computing stationary solutions of linear programs with complementarity constraints (LPCCs). We focus on three such formulations and establish connections between their stationary solutions and those of the LPCC. Improvements of the DCA are proposed to remedy some drawbacks in a straightforward adaptation of the DCA to these formulations. Extensive numerical results, including comparisons with an existing nonlinear programming solver and the mixed-integer formulation, are presented to elucidate the effectiveness of the overall DC approach.

1 Introduction

Using the difference-of-convex programming approach, this paper investigates the computation of local solutions of linear programs with complementarity constraints (LPCCs) of the form

$$\begin{array}{ll} \underset{x,y,w}{\text{minimize}} & c^T x + d^T y \\ \text{subject to} & Ax + By \geq f \\ & Mx + Ny + q = w \\ \text{and} & 0 \leq y \perp w \geq 0, \end{array} \tag{1}$$

where $c \in \mathbb{R}^t$, $d, q \in \mathbb{R}^n$, $f \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times t}$, $B \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{n\times t}$ and $N \in \mathbb{R}^{n\times n}$. The expression "$y \perp w$" means that the vectors w and y are orthogonal, i.e., $w^T y = 0$. Beginning with the early work in the integer programming community on what was known then as the "complementary program" [15, 16, 17], and as a special case of a mathematical program with equilibrium constraints (MPCCs) [27] on which there is an extensive literature to date, the LPCC has grown in importance in applications [14] as complementarity constraints can be used to model many logical, piecewise, and nonconvex conditions [14], even discontinuous ones such as cardinality objective [6] and constraints [2]. In addition, the

*Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, USA. This author was supported in part by the U.S. National Science Foundation grant CMMI-1334639. *Email: francisco-jaramoroni2013@u.northwestern.edu*

†Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California 90089, USA. This author was partially supported by the U.S. National Science Foundation grant CMMI-1402052 and the Air Force Office of Sponsored Research grant FA9550-15-1-0126. *Email: jongship@usc.edu*

‡Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, USA. This author was supported in part by the U.S. National Science Foundation grant CMMI-1334639. *Email: andreas.waechter@northwestern.edu*

LPCC provides an interesting framework for the study of nonconvex quadratic programs [12]. In general, determining the global solution of LPCCs is NP-hard. While there has been recent research into the global resolution of the LPCC [13, 34] and its extension to a quadratic program with complementarity constraints [1], the task of designing a general-purpose procedure with a provable certificate to compute a globally optimal solution to the LPCC efficiently, if it exists, remains practically challenging.

The present paper extends the work of two references [22, 29] in which problems with complementarity constraints are solved via their reformulations as difference-of-convex (DC) programs; these reformulated DC programs are solved by the difference-of-convex algorithm (DCA) pioneered by Le Thi Hoai An and Pham Dinh Tao [19, 20, 21, 23, 31]. In [22], based on several DC decompositions, the DCA was applied to the standard linear complementarity problem [4]; in [29], the DC approach was embedded in a branch-and-bound scheme for solving mathematical programs with affine variational inequality constraints. Our work is in the spirit of [22]; specifically, we introduce several penalized formulations of the complementarity constraint and employ various DC decompositions of the resulting penalty functions, based on which the DCA is applied to the resulting penalized DC programs and enhancements to this basic algorithm are introduced. This approach is different from a nonlinear programming approach as implemented in the solvers FILTER [8, 10] and KNITRO [3], which are applicable to deal with the complementarity constraints [7, 25]. Recently, motivated by a sequential LPCC method for solving MPCCs [24], a pivoting method [5] is developed that includes an anticycling technique for verifying "B(ouligand)-stationarity". For a survey of algorithms for solving LPCCs, see [18].

The organization of the remainder of the paper is as follows. After formally defining the LPCC and introducing three main stationarity concepts for this problem in the next section, we briefly review the DCA for computing a critical point of a DC program in Section 3. Three penalty formulations of complementarity constraint are then discussed in detail in Section 4 that is divided into subsections, one for each penalty formulation. Section 5 describes enhancements to the DCA applied to the penalized formulations of the LPCC. Extensive computation results are reported in Section 6 to support the practical performance of the enhanced DCA. Overall, this paper has allowed us to gain a deeper understanding of the DC approach to the local solution of the LPCC.

2 Concepts of Stationarity

Stationarity concepts for mathematical programs with complementarity constraints have been well studied. For the LPCC, the references [9, Sections 3 and 4], [33] and [5, Section 2] are most pertinent to the following discussion. For concise notation, we define the set of linear constraints in the LPCC (1),

$$\Omega \triangleq \left\{ (x, y, w) \in \mathbb{R}^t \times \mathbb{R}^{2n}_+ \mid Ax + By \geq f,\, Mx + Ny + q = w \right\}.$$

Let $(\bar{x}, \bar{y}, \bar{w})$ be a feasible point of the LPCC (1). We define the following partition of the variable indices for y and w:

$$\mathcal{I}_y(\bar{y}, \bar{w}) \triangleq \{ i \mid \bar{y}_i = 0 < \bar{w}_i \}, \quad \mathcal{I}_w(\bar{y}, \bar{w}) \triangleq \{ i \mid \bar{y}_i > 0 = \bar{w}_i \}, \quad \mathcal{I}_0(\bar{y}, \bar{w}) \triangleq \{ i \mid \bar{y}_i = 0 = \bar{w}_i \}, \quad (2)$$

and refer to the indices in $\mathcal{I}_0(\bar{y}, \bar{w})$ as the degenerate (or bi-active) indices, for which strict complementarity fails to hold. Corresponding to these index sets, we define the *relaxed linear program*:

$$\begin{array}{llll} \text{RxLP:} & \underset{(x,y,w)\in\Omega}{\text{minimize}} & c^T x + d^T y & \\ & & y_i = 0 \leq w_i, & i \in \mathcal{I}_y(\bar{y}, \bar{w}) \\ & & y_i \geq 0 = w_i, & i \in \mathcal{I}_w(\bar{y}, \bar{w}) \\ & \text{and} & y_i, w_i \geq 0, & i \in \mathcal{I}_0(\bar{y}, \bar{w}) \end{array}$$

and the *restricted linear program*:

$$\begin{array}{llll} \text{RsLP:} & \underset{(x,y,w)\in\Omega}{\text{minimize}} & c^T x + d^T y & \\ & & y_i = 0 \leq w_i, & i \in \mathcal{I}_y(\bar{y}, \bar{w}) \\ & & y_i \geq 0 = w_i, & i \in \mathcal{I}_w(\bar{y}, \bar{w}) \\ & \text{and} & y_i, w_i = 0, & i \in \mathcal{I}_0(\bar{y}, \bar{w}) \end{array}$$

Moreover, for every subset $\mathcal{J} \subseteq \mathcal{I}_0(\bar{y}, \bar{w})$, we also define the *piecewise linear program*:

$$\begin{array}{llll} \text{LP}(\mathcal{J})\text{:} & \underset{(x,y,w)\in\Omega}{\text{minimize}} & c^T x + d^T y & \\ & & y_i = 0 \leq w_i, & i \in \mathcal{I}_y(\bar{y}, \bar{w}) \\ & & y_i \geq 0 = w_i, & i \in \mathcal{I}_w(\bar{y}, \bar{w}) \\ & & y_i = 0 \leq w_i, & i \in \mathcal{J} \\ & \text{and} & y_i \geq 0 = w_i, & i \in \mathcal{I}_0(\bar{y}, \bar{w}) \setminus \mathcal{J}, \end{array}$$

which is a *piece* of the LPCC at the given point $(\bar{x}, \bar{y}, \bar{w})$. Based on the LPs: (RxLP), (RsLP), and LP($\mathcal{J}$), we have the following definition.

Definition 1. A feasible triple $(\bar{x}, \bar{y}, \bar{w})$ of the LPCC (1) is said to be

- *strongly stationary* if it is a solution of the relaxed LP (RxLP);
- *weakly stationary* if it is a solution of the restricted LP (RsLP);
- *B(ouligand)-stationary* if it is a solution of the piecewise LP($\mathcal{J}$) for all $\mathcal{J} \subseteq \mathcal{I}_0(\bar{y}, \bar{w})$. □

These stationary concepts can be described in terms of multipliers of the constraints; in particular, using the notation $\circ$ for the Hadamard product of two vectors, we see that a triple $(\bar{x}, \bar{y}, \bar{w}) \in \Omega$ is weakly stationary if and only if it is feasible to (1) and there exist multipliers $\lambda \in \mathbb{R}^m$ and μ^y and $\mu^w \in \mathbb{R}^n$, so that

$$\begin{array}{lcl} A^T\lambda + M^T\mu^w & = & c \\ B^T\lambda + N^T\mu^w + \mu^y & = & d \\ 0 \leq \lambda \perp Ax + By - f & \geq & 0 \\ y \circ \mu^y = 0 & \text{and} & w \circ \mu^w = 0. \end{array} \tag{3}$$

Notice that there is no sign restriction on μ_i^y and μ_i^w for $i \in \mathcal{I}_0(\bar{y}, \bar{w}))$; if these bi-active multipliers are restricted to be nonnegative, then the resulting conditions are equivalent to the strong stationarity of the given triple $(\bar{x}, \bar{y}, \bar{w})$. In contrast, B-stationarity cannot be described in terms of a singe tuple of multipliers due to the multiple LPs that depend on the index subsets $\mathcal{J}$ of $\mathcal{I}_0(\bar{y}, \bar{w})$. The following proposition elucidates the connections among the above stationarity conditions and the local minimizing property of the triple. An example in [33] shows that a B-stationary point is not necessarily strongly stationary.

Proposition 2. Let $(\bar{x}, \bar{y}, \bar{w})$ be a given feasible solution of (1). The following implications hold:

$$\text{strong stationarity} \Rightarrow \text{local minimizing} \Leftrightarrow \text{B-stationarity} \Rightarrow \text{weak stationarity.}$$

Proof. It is clear that strong stationarity implies B-stationarity which implies weak stationarity. It remains to show that B-stationarity is equivalent to local minimizing. This is not difficult to see because locally near $(\bar{x}, \bar{y}, \bar{w})$, the feasible region Ω is the union of the feasible sets of the pieces of the LP($\mathcal{J}$) for all $\mathcal{J} \subseteq \mathcal{I}_0(\bar{y}, \bar{w})$. (This fact was noted in the early work of the MPCC as described in [27].) □

With the equivalence of B-stationarity and the locally minimizing property (for the LPCC), we see that this is the sharpest among the other stationarity concepts. Nevertheless, the computation of a B-stationary solution to date remains elusive, unless one aims to apply a global resolution scheme such as the logical Benders approach described in [13] or others global methods [29, 34] that invariably involve some kind of enumeration. This paper does not address these global schemes.

3 The Difference of Convex Functions Approach

The proposed method replaces the complementarity constraint $0 \le y \perp w \ge 0$ by a penalization term that is added to the objective function:

$$\underset{(x,y,w)\in\Omega}{\text{minimize}}\; c^T x + d^T y + \rho\, \phi(y, w). \tag{4}$$

Here, $\rho > 0$ is a penalty parameter, and the penalty function $\phi : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \longrightarrow \mathbb{R}_+$ is zero if and only if $y \perp w$. Starting from [27], penalty formulations have been studied extensively in the literature of the MPCCs. Our focus here is based on the assumption that $\phi(y, w)$ is a difference of two convex functions, i.e.

$$\phi(y, w) \;=\; \phi_+(y, w) - \phi_-(y, w), \tag{5}$$

where ϕ_+ and ϕ_- are convex. Writing the objective function in (4) as $f_+(x, y, w) - f_-(x, y, w)$ with $f_+(x, y, w) = c^T x + d^T y + \rho\phi_+(y, w)$ and $f_-(x, y, w) = \rho\phi_-(y, w)$, we see that (4) is a DC program which has the following general form:

$$\underset{z\in C}{\text{minimize}}\; f(z) \;\triangleq\; f_1(z) - f_2(z), \tag{6}$$

where $f_1,\ f_2 : \mathbb{R}^m \longrightarrow \mathbb{R}$ are convex and $C \subseteq \mathbb{R}^m$ is a closed convex set. The basic idea of the DCA [21, 31, 23] for solving (6) is as follows: At an iterate z^k, the algorithm overestimates the concave part of the objective function by a linear function, using a subgradient $g^k \in \partial f_2(z^k)$ of f_2 at z^k. An optimal solution of the resulting convex auxiliary problem

$$\min_{z\in C}\; f_1(z) - (g^k)^T (z - z^k) \tag{7}$$

provides the new iterate. Needless to say, the choice of the subgradient has an important effect on the practical efficiency of the algorithm; we will discuss more about such choices in specific contexts of the function f_2. The algorithm is formally stated in Algorithm 1.

Algorithm 1 Basic difference-of-convex functions algorithm

1: Choose termination tolerance $\varepsilon_{\text{tol}} > 0$.
2: Let $z^1 \in C$ and $k \leftarrow 0$.
3: **repeat**
4: Set $k \leftarrow k + 1$.
5: Compute subgradient $g^k \in \partial f_2(z^k)$ of f_2 at z^k.
6: Compute the new iterate z^{k+1} as solution of (7).
7: **until** $f(z^k) - f(z^{k+1}) \le \varepsilon_{\text{tol}}$
8: Return z^{k+1}.

Properties of the DCA have been well known in the DC literature, see e.g. [19, 21]. In particular, it has been shown that when $z^* = z^k$ is returned by the algorithm for some finite k or if z^* is a limit point of an infinite sequence $\{z^k\}$ (with $\varepsilon_{\text{tol}} = 0$), then z^* is a critical point for (6) in the sense that $\partial f_2(z^*) \cap (\partial f_1(z^*) + \mathcal{N}(z^*; C)) \neq \emptyset$, where $\mathcal{N}(z^*; C)$ is the *normal cone* of C at z^*. Moreover, if C is polyhedral, as in the case of (4), and if either f_1 or f_2 is *polyhedral convex* [32], then the algorithm will terminate in a finite number of steps. Subsequently, we show in Proposition 6 that criticality in a penalized piecewise linear (thus non-differentiable) DC formulation of the LPCC is equivalent to weak stationarity of the LPCC. It has been noted in [30] that the condition of criticality is in general weaker than that of d(irectional)-stationarity for a convex constrained DC program. Specifically, $z^* \in C$ is d-stationary for (6) if $f'(z^*, z - z^*) \ge 0$ for all $z \in C$, where the prime $'$ notation denotes the directional derivative at z^* in the direction $z - z^*$; the stationarity condition is equivalent to the inclusion $\partial f_2(z^*) \subseteq (\partial f_1(z^*) + \mathcal{N}(z^*; C))$. In turn, in the terminology of DC programming [19, 21], the latter inclusion states that z^* satisfies the *generalized KKT condition* of (6). It is an elementary fact that d-stationarity is a necessary condition of the locally minimizing property for a convex constrained optimization problem with a directionally differentiable objective function.

4 DC Penalty Functions

We present three different choices for the penalty function ϕ in the following subsections. They correspond to the DC objective functions used in [22] for the solution of linear complementarity problems and are here extended to the LPCC. This reference presents also a fourth DC function (the "simple constrained quadratic program"); since the numerical results therein show rather poor performance with this formulation, we are not exploring it in this paper. Besides providing the foundation for the penalty-based algorithmic approach for the solution of the LPCC, the discussion in the subsections herein relate the stationarity conditions of the penalty formulation to the various LPCC stationarity conditions introduced in Definition 1. While connections like these have been discussed to some extent in the MPCC literature (see e.g. [5, 9]), there are two emphases in our presentation that are not fully transparent in this literature. First, we address the stationarity conditions the penalty formulations for fixed values of the penalty parameter ρ; second, being nonconvex optimization problems, practically computable solutions to the penalty formulations are typically of the stationary kind; it is therefore important to understand how such computed solutions translate into the LPCC stationarity solutions.

4.1 A piecewise linear penalty

The first penalty function that we consider is

$$\phi^{\rm PL}(y,w) \triangleq \sum_{i=1}^{n} \min\,(\,y_i, w_i\,) \tag{8}$$

which is clearly a concave, thus DC, function. This corresponds to the "concave separable minimization" formulation in [22]. The authors of this reference identified this as the most effective formulation among the ones they used for solving the standard LCP. Incidentally, this penalty function was used in a successive linearization algorithm [28] that reformulates the LCP as a concave minimization problem. With (8) as the penalty function, the problem (5) becomes

$$\underset{z \triangleq (x,y,w) \in \Omega}{\text{minimize}} \;\; \theta^{\rm PL}(z) \,\triangleq c^T x + d^T y + \rho \sum_{i=1}^{n} \min(\,y_i, w_i\,). \tag{9}$$

In what follows, we examine the stationarity and local minimizing properties of a triple $\bar{z} \triangleq (\bar{x}, \bar{y}, \bar{w}) \in \Omega$ that is feasible for (9) but may not be feasible to (1). We further show in Proposition 6 how stationarity in terms of the directional derivatives of the objective function, i.e., d-stationarity, is related to the criticalilty property in terms of the subdifferentials of the two convex components of the DC representation of the min function. We begin with a result showing that $\bar{z} \in \Omega$ is d-stationary for (9) if and only if is locally minimizing; we further relate this stationarity property in terms of the LP pieces of the LPCC (1), given by ${\rm LP}_\rho(\widehat{\mathcal{J}})$

$$\underset{x,y,w \in \Omega}{\text{minimize}} \; c^T x + d^T y + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y},\bar{w})} y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y},\bar{w})} w_i + \sum_{i \in \widehat{\mathcal{J}}} y_i + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y},\bar{w}) \setminus \widehat{\mathcal{J}}} w_i \right] \tag{10}$$

for all subsets $\widehat{\mathcal{J}}$ of $\widehat{\mathcal{I}}_=(\bar{y}, \bar{w})$, where

$$\widehat{\mathcal{I}}_y(\bar{y},\bar{w}) \triangleq \{\, i \mid \bar{y}_i < \bar{w}_i \,\}, \quad \widehat{\mathcal{I}}_w(\bar{y},\bar{w}) \triangleq \{\, i \mid \bar{w}_i < \bar{y}_i \,\}, \quad \widehat{\mathcal{I}}_=(\bar{y},\bar{w}) \triangleq \{\, i \mid \bar{y}_i = \bar{w}_i \,\}. \tag{11}$$

Proposition 3. Let $\rho > 0$ and a triple $\bar{z}^\rho \triangleq (\bar{x}^\rho, \bar{y}^\rho, \bar{w}^\rho) \in \Omega$ be given. The following three statements are equivalent.

(a) $\bar{z}^\rho$ is a local minimizer of (9);

(b) $\bar{z}^\rho$ is d-stationary for (9);

(c) $\bar{z}^\rho$ is a minimizer of the ${\rm LP}_\rho(\widehat{\mathcal{J}})$ for every subset $\widehat{\mathcal{J}}$ of $\widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)$.

Proof. (a) $\Rightarrow$ (b). This is trivially true in general for a directionally differentiable objective.

(b) ⇒ (c). Suppose that $\bar{z}^\rho$ is d-stationary for (9). Let $z = (x, y, w) \in \Omega$ be arbitrary. We then have

$$\begin{aligned}
0 &\leq \theta'(\bar{z}^\rho; z - \bar{z}^\rho) \\
&= c^T(\, x - \bar{x}^\rho\,) + d^T(\, y - \bar{y}^\rho\,) + \\
&\qquad \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} (\, y - \bar{y}_i^\rho\,) + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} (\, w - \bar{w}_i^\rho\,) + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \min(\, y_i - \bar{y}_i^\rho,\, w_i - \bar{w}_i^\rho\,) \right] \\
&\leq c^T(\, x - \bar{x}^\rho\,) + d^T(\, y - \bar{y}^\rho\,) + \\
&\qquad \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} (\, y - \bar{y}_i^\rho\,) + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} (\, w - \bar{w}_i^\rho\,) + \sum_{i \in \widehat{\mathcal{J}}} [\, y_i - \bar{y}_i^\rho\,] + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho) \setminus \widehat{\mathcal{J}}} [\, w_i - \bar{w}_i^\rho\,] \right]
\end{aligned}$$

for every $\widehat{\mathcal{J}} \subseteq \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)$. Thus (c) holds.

(c) ⇒ (a). Suppose that $\bar{z}^\rho$ is a local minimizer of the $\text{LP}_\rho(\widehat{\mathcal{J}})$ for every subset $\widehat{\mathcal{J}}$ of $\widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)$. Let $(x, y, w) \in \Omega$ be arbitrary. The vector $(x^\tau, y^\tau, w^\tau) \triangleq (\bar{x}^\rho, \bar{y}^\rho, \bar{w}^\rho) + \tau(x - \bar{x}^\rho, y - \bar{y}^\rho, w - \bar{w}^\rho)$ remains in the convex set Ω for all $\tau \in [0, 1]$. Moreover, for $\tau > 0$ sufficiently small,

$$\min(y_i^\tau, w_i^\tau) = \begin{cases} y_i^\tau & \text{if } i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho) \\ w_i^\tau & \text{if } i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho). \end{cases}$$

Let $\widehat{\mathcal{J}} \triangleq \left\{ i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho) \mid y_i^\tau \leq w_i^\tau \right\}$, we then have

$$\min(y_i^\tau, w_i^\tau) = \begin{cases} y_i^\tau & \text{if } i \in \widehat{\mathcal{J}} \\ w_i^\tau & \text{if } i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho) \setminus \widehat{\mathcal{J}}. \end{cases}$$

Hence,

$$\begin{aligned}
& c^T\bar{x}^\rho + d^T\bar{y}^\rho + \rho \sum_{i=1}^n \min(\, \bar{y}_i^\rho, \bar{w}_i^\rho\,) \\
&= c^T\bar{x}^\rho + d^T\bar{y}^\rho + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} \bar{y}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} \bar{w}_i^\rho + \sum_{i \in \widehat{\mathcal{J}}} \bar{y}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho) \setminus \widehat{\mathcal{J}}} \bar{w}_i^\rho \right] \\
&\leq c^T x^\tau + d^T y^\tau + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} y_i^\tau + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}, \bar{w})} w_i^\tau + \sum_{i \in \widehat{\mathcal{J}}} y_i^\tau + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho) \setminus \widehat{\mathcal{J}}} w_i^\tau \right] \\
&= c^T x^\tau + d^T y^\tau + \rho \sum_{i=1}^n \min(\, y_i^\tau, w_i^\tau\,),
\end{aligned}$$

where the inequality follows from the fact that $\bar{z}^\rho$ is a local minimizer of $\text{LP}_\rho(\widehat{\mathcal{J}})$. Passing to the limit $\tau \downarrow 0$ completes the proof of the proposition. □

In the above proposition, the triple $(\bar{x}^\rho, \bar{y}^\rho, \bar{w}^\rho)$ is not necessarily feasible to the LPCC (1). If it is, then it is a local minimizer of the LPCC; moreover, $(\bar{x}^\rho, \bar{y}^\rho, \bar{w}^\rho)$ remains locally minimizing for (9) for all ρ sufficiently large. These are the two assertions in the following result.

Proposition 4. Let $(\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) \in \Omega$ be a local minimizer of (9) for some $\bar{\rho} > 0$. Then $(\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}})$ is a local minimizer of the LPCC (1) if and only if $\min(\bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) = 0$. Moreover, if $\min(\bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) = 0$, then $(\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) \in \Omega$ is a local minimizer of (9) for all $\rho \geq \bar{\rho}$.

Proof. Clearly only the "if" statement requires a proof. Suppose $(\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) \in \Omega$ is a local minimizer of (9) satisfying $\min(\bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) = 0$ for some $\bar{\rho} > 0$. Let (x, y, w) be an arbitrary triple feasible to (1) that is sufficiently close to $(\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}})$. We then have

$$c^T x + d^T y \;=\; c^T x + d^T y + \bar{\rho} \sum_{i=1}^{n} \min(\, y_i, w_i \,) \;\geq\; c^T \bar{x}^{\bar{\rho}} + d^T y^{\bar{\rho}} + \bar{\rho} \sum_{i=1}^{n} \min(\, \bar{y}_i^{\bar{\rho}}, \bar{w}_i^{\bar{\rho}} \,) \;=\; c^T \bar{x}^{\bar{\rho}} + d^T \bar{y}^{\bar{\rho}},$$

where the inequality holds because of the local minimizing property of $(\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}})$. To prove the last assertion of the proposition, let $\rho \geq \bar{\rho}$ be given. We then have, for every $(x, y, w) \in \Omega$,

$$\begin{aligned}
& c^T x + d^T y + \rho \sum_{i=1}^{n} \min(\, y_i, w_i \,) \\
& \quad \geq\; c^T x + d^T y + \bar{\rho} \sum_{i=1}^{n} \min(\, y_i, w_i \,) \\
& \quad \geq\; c^T \bar{x}^{\bar{\rho}} + d^T y^{\bar{\rho}} + \bar{\rho} \sum_{i=1}^{n} \min(\, \bar{y}_i^{\bar{\rho}}, \bar{w}_i^{\bar{\rho}} \,) \quad \text{provided that } (x, y, w) \text{ is sufficiently close to } (\bar{x}^{\bar{\rho}}, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}}) \\
& \quad =\; c^T \bar{x}^{\bar{\rho}} + d^T y^{\bar{\rho}} + \rho \sum_{i=1}^{n} \min(\, \bar{y}_i^{\bar{\rho}}, \bar{w}_i^{\bar{\rho}} \,) \quad \text{because } \min(\, \bar{y}^{\bar{\rho}}, \bar{w}^{\bar{\rho}} \,) \;=\; 0,
\end{aligned}$$

completing the proof of the proposition. □

While Proposition 4 discusses that suitable local minimizers of (9) correspond to local minimizers of the LPCC (1), the following proposition shows that each local minimizer of (1) can be recovered as a local minimizer of (9) for sufficiently large values of the penalty parameter. We formulate this assertion as a necessary and sufficient condition which shows that (9) is an *exact* local penalty formulation of (1).

Proposition 5. Let $\bar{z} \triangleq (\bar{x}, \bar{y}, \bar{w}) \in \Omega$ be given. The following two statements are equivalent.

(a) $\min(\bar{y}, \bar{w}) = 0$ and there exists $\bar{\rho} > 0$ so that $\bar{z}$ is a local minimizer of (9) for all $\rho \geq \bar{\rho}$;

(b) $\bar{z}$ is a local minimizer of the LPCC (1).

Proof. (a) ⇒ (b). This follows from Proposition 4.

(b) ⇒ (a). For the purpose of deriving a contradiction, suppose that $\bar{z}$ is not a local minimizer of (9) for any ρ_k in a sequence of positive scalars $\{\rho_k\}$ tending to infinity. Proposition 3 then implies that for each k there exists $\widehat{\mathcal{J}}_k \subseteq \widehat{\mathcal{I}}_{=}(\bar{y}, \bar{w})$ so that $\bar{z}$ is not a minimizer of the $\mathrm{LP}_{\rho_k}(\widehat{\mathcal{J}}_k)$. Note that $\widehat{\mathcal{I}}_{=}(\bar{y}, \bar{w}) = \mathcal{I}_0(\bar{y}, \bar{w})$ since $\bar{z}$ is feasible to the LPCC. Because there are only finitely many subsets of $\mathcal{I}_0(\bar{y}, \bar{w})$, we may assume, without loss of generality, that $\widehat{\mathcal{J}}_k = \mathcal{J}$ for all k for some $\mathcal{J} \subseteq \mathcal{I}_0(\bar{y}, \bar{w})$. Since $\bar{z}$ is a local minimizer of the LPCC, Proposition 2 implies that $\bar{z}$ is a B-stationary point, and therefore a minimizer of the $\mathrm{LP}(\mathcal{J})$. Note that $\mathrm{LP}_{\rho_k}(\mathcal{J})$ is a standard penalty formulation of $\mathrm{LP}(\mathcal{J})$. Therefore $\bar{z}$ is also a minimizer to $\mathrm{LP}_{\rho_k}(\mathcal{J})$ for sufficiently large ρ_k, leading to a contradiction. □

4.1.1 Criticality and weak stationarity

Among many DC decompositions (6) of the piecewise linear DC objective $\theta^{\rm PL}$, we employ the one that is given by $f_1^{\rm PL}(z) \triangleq c^T x + d^T y$ and $f_2^{\rm PL}(z) \triangleq \sum_{i=1}^{n} \max(-y_i, -w_i)$. With this decomposition, we have the following result which re-affirms that criticality defined for this DC formulation is weaker than d-stationarity for the LPCC. Recall that criticality in this context means: $\partial f_2^{\rm PL}(z^*) \cap \left(\partial f_1^{\rm PL}(z^*) + \mathcal{N}(z^*;\Omega)\right) \neq \emptyset$.

Proposition 6. Let $\rho > 0$ and a triple $\bar{z}^\rho \triangleq (\bar{x}^\rho, \bar{y}^\rho, \bar{w}^\rho) \in \Omega$ be given. The following three statements are equivalent with respect to the DC decomposition $(f_1^{\rm PL}, f_2^{\rm PL})$ of $\theta^{\rm PL}$:

(a) $\bar{z}^\rho$ is a critical point of the DC program (9);

(b) for every $(x, y, w) \in \Omega$,

$$c^T (\, x - \bar{x}^\rho \,) + d^T (\, y - \bar{y}^\rho \,) + \\ \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} (\, y - \bar{y}_i^\rho \,) + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} (\, w - \bar{w}_i^\rho \,) + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \max(\, y_i - \bar{y}_i^\rho, \, w_i - \bar{w}_i^\rho \,) \right] \geq 0;$$

(c) $\bar{z}^\rho$ is an optimal solution of the following convex piecewise linear program:

$$\underset{(x,y,w)\in\Omega}{\text{minimize}} \; c^T x + d^T y + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} w_i + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \max(y_i, w_i) \right].$$

Moreover, if $\bar{z}^\rho$ is critical to (9) and feasible to the LPCC (1), then $\bar{z}^\rho$ is a weakly stationary solution of (1).

Proof. By definition, $\bar{z}^\rho$ is a critical point of the DC program (9) if and only if there exists $(0, \tilde{a}, \tilde{b})$ in $\partial f_2^{\rm PL}(\bar{z}^\rho)$ such that $(0, \tilde{a}, \tilde{b}) \in \partial f_1^{\rm PL}(\bar{z}^\rho) + \mathcal{N}(\bar{z}^\rho; \Omega)$. With $(a, b) \triangleq \rho^{-1}(\tilde{a}, \tilde{b})$, this implies that $\bar{z}^\rho$ is an optimal solution of the linear program:

$$\underset{(x,y,w)\in\Omega}{\text{minimize}} \; c^T x + d^T y - \rho \, [\, a^T y + b^T w \,] \, .$$

By the definition of the separable function $f_2^{\rm PL}$, it is then not difficult to see that $\bar{z}^\rho$ is a critical point of (9) if and only if there exist $\xi_i \in [0, 1]$ for $i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)$ such that $\bar{z}^\rho$ is an optimal solution of the linear program:

$$\underset{(x,y,w)\in\Omega}{\text{minimize}} \; c^T x + d^T y + \rho \left\{ \sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} w_i + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} [\, \xi_i \, y_i + (\, 1 - \xi_i \,) \, w_i \,] \right\}.$$

Based on this equivalence, we can now complete the proof of the proposition. Clearly statements (b) and (c) are equivalent. Suppose that (a) holds. Let $\xi_i \in [0, 1]$ for $i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)$ be as given above. We then have for any $(x, y, w) \in \Omega$

$$
\begin{aligned}
& c^T x + d^T y + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} w_i + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \max(y_i, w_i) \right] \\
& \quad \geq c^T x + d^T y + \rho \left\{ \sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} w_i + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} [\, \xi_i \, y_i + (\, 1 - \xi_i \,) \, w_i \,] \right\} \\
& \quad \geq c^T \bar{x}^\rho + d^T \bar{y}^\rho + \rho \left\{ \sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} \bar{y}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} \bar{w}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} [\, \xi_i \, \bar{y}_i^\rho + (\, 1 - \xi_i \,) \, \bar{w}_i^\rho \,] \right\} \\
& \quad = c^T \bar{x}^\rho + d^T \bar{y}^\rho + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} \bar{y}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} \bar{w}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \max(\, \bar{y}_i^\rho, \bar{w}_i^\rho \,) \right].
\end{aligned}
$$

Thus (c) holds. Conversely, suppose that (c) holds. If $\bar{\theta}(x, y, w)$ denotes the objective function in the LP in part (c), we then have $0 \in \partial \bar{\theta}(\bar{z}^\rho) + \mathcal{N}(\bar{z}^\rho; \Omega)$. Since $\partial \max(s, t) = \{(\lambda, 1 - \lambda) \mid \lambda \in [0, 1]\}$ at a pair (s, t) with $s = t$, statement (a) follows readily from the subdifferential characterization of optimality $\bar{z}$ to the LP in part (c). Therefore statements (a), (b), and (c) are equivalent.

To complete the proof of the proposition, suppose that $\bar{z}^\rho$ is critical to (9) and feasible to the LPCC (1). Let (x, y, w) be a feasible triple to the restricted LP (RsLP) at $\bar{z}^\rho$. We then have from (c) that

$$
\begin{aligned}
c^T x + d^T y &= c^T x + d^T y + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} w_i + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \max(y_i, w_i) \right] \\
&\geq c^T \bar{x}^\rho + d^T \bar{y}^\rho + \rho \left[\sum_{i \in \widehat{\mathcal{I}}_y(\bar{y}^\rho, \bar{w}^\rho)} \bar{y}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y}^\rho, \bar{w}^\rho)} \bar{w}_i^\rho + \sum_{i \in \widehat{\mathcal{I}}_=(\bar{y}^\rho, \bar{w}^\rho)} \max(\bar{y}_i^\rho, \bar{w}_i^\rho) \right] \\
&= c^T \bar{x}^\rho + d^T \bar{y}^\rho,
\end{aligned}
$$

establishing that $\bar{z}^\rho$ is an optimal solution of RsLP. □

4.1.2 The DCA applied to (9)

In the application of the DCA to (9), the subproblem (7) becomes the following LP, which we will denote as LP(y^k, w^k, ξ^k):

$$
\min_{(x,y,w) \in \Omega} \; c^T x + d^T y + \rho \left[\sum_{i \in \widehat{\mathcal{J}}_w^k} y_i + \sum_{i \in \widehat{\mathcal{J}}_y^k} w_i + \sum_{i \in \widehat{\mathcal{J}}_=^k} \left(\xi_i^k y_i + (1 - \xi_i^k) w_i \right) \right], \tag{12}
$$

where

$$
\widehat{\mathcal{J}}_y^k \triangleq \widehat{\mathcal{I}}_y(y^k, w^k), \qquad \widehat{\mathcal{J}}_w^k \triangleq \widehat{\mathcal{I}}_w(y^k, w^k), \qquad \widehat{\mathcal{J}}_=^k \triangleq \widehat{\mathcal{I}}_=(y^k, w^k), \tag{13}
$$

and $\xi_i \in [0, 1]$ indicates which particular subgradient of each max-function in f_2^{PL} is chosen at a non-differentiable pair (y_i, w_i). We note that the current iterate (x^k, y^k, w^k) enters only in the partition of the variable indices (13), and that the absolute values of the variables do not matter. From this point of view, we may reinterpret the DCA as an active-set method, in which the algorithm solves LPs corresponding to different guesses of the optimal active set, or piece, of the LPCC. If we assume that ξ_i is chosen only from a finite set Ξ, then this observation shows that the DC algorithm will terminate after a finite number of iterations, because there exist only a finite number of possible constraint partitions $\widehat{\mathcal{J}}_y^k$, $\widehat{\mathcal{J}}_w^k$, $\widehat{\mathcal{J}}_=^k$, and the objective function is strictly monotonically decreasing. In our implementation, we restrict the choice of subgradients to the finite set $\Xi = \left\{0, \frac{1}{2}, 1\right\}$. The follow proposition shows that the final iterates exhibit stationarity properties for the LPCC.

Proposition 7. Assume that the LPCC is bounded below and that ξ_i^k in (12) is chosen from a finite set $\Xi \subseteq [0, 1]$. Then the DC algorithm, with termination tolerance $\varepsilon_{\text{tol}} = 0$, will terminate after a finite number of iterations. If z^k at termination is feasible to the LPCC (1), then z^k is a weakly stationary point.

Proof. We already argued that the algorithm terminates after a finite number of iterations. If remains to show the second assertion of the proposition. Let z^k be as given. It follows that $z^{k+1} = (x^{k+1}, y^{k+1}, w^{k+1})$ is an optimal solution of $\text{LP}(y^k, w^k, \xi^k)$. The point $z^k \in \Omega$ is feasible for $\text{LP}(y^k, w^k, \xi^k)$, and by properties of the subgradient, the objective of this LP overestimates the DC objective $f(z)$ in (6). Therefore we have, using the definitions of $\widehat{\mathcal{J}}_w^k$, $\widehat{\mathcal{J}}_y^k$, and $\widehat{\mathcal{J}}_=^k$,

$$\begin{aligned} f(z^{k+1}) &\leq \text{optimal objective of } \text{LP}(y^k, w^k, \xi^k) \\ &\leq c^T x^k + d^T y^k + \rho \left[\sum_{i \in \widehat{\mathcal{J}}_w^k} y_i + \sum_{i \in \widehat{\mathcal{J}}_y^k} w_i + \sum_{i \in \widehat{\mathcal{J}}_=^k} \left(\xi_i^k y_i + (1 - \xi_i^k) w_i \right) \right] \\ &= c^T x^k + d^T y^k + \rho \sum_{i=1}^n \min\{y_i^k, w_i^k\} = f(z^k). \end{aligned}$$

Since the algorithm terminated, we have $f(z^{k+1}) = f(z^k)$, i.e., equalities hold in all relationships above, and therefore z^k is optimal for $\text{LP}(y^k, w^k, \xi^k)$. Consequently, there exist multipliers λ^f, λ^g, λ^q, μ^y, μ^w so that

$$\begin{aligned} A^T \lambda^f + M^T \lambda^q &= c \\ B^T \lambda^f + N^T \lambda^q + \mu^y &= d + \rho \left[\mathbb{I}_{y^k < w^k} + \xi \circ \mathbb{I}_{y^k = w^k} \right] \\ -\lambda^q + \mu^w &= \rho \left[\mathbb{I}_{y^k > w^k} + (e - \xi) \circ \mathbb{I}_{y^k = w^k} \right] \\ (Ax^k + By^k - f) \circ \lambda^f &= 0 \\ y^k \circ \mu^y &= 0 = w^k \circ \mu^w \\ \lambda^f,\ \mu^y,\ \mu^w &\geq 0 \\ \text{and} \quad (x^k, y^k, w^k) &\in \Omega, \end{aligned} \tag{14}$$

where e is the vector of all ones, and $\mathbb{I}_{\text{cond}}$ is a 0-1 vector in which those components are 1 for which the condition in the subscript holds. Now define

$$\bar{\mu}^y = \mu^y - \rho \left(\mathbb{I}_{y^k < w^k} + \xi \circ \mathbb{I}_{y^k = w^k} \right) \quad \text{and} \quad \bar{\mu}^w = \mu^w - \rho \left(\mathbb{I}_{y^k > w^k} + (e - \xi) \circ \mathbb{I}_{y^k = w^k} \right).$$

Since $y^k \circ w^k = 0$ by assumption, we have $w_k^i = 0$ for any component i for which $y_i^k > 0$. Therefore $y^k \circ \mathbb{I}_{y^k < w^k} = 0$. Similarly, if $y_k^i = w_k^i$, we must have $y_k^i = 0$ and hence $y^k \circ \mathbb{I}_{y^k = w^k} = 0$. Finally, since

$y^k \circ \mu^y = 0$, we obtain $\bar{\mu}_y \circ y^k = 0$. Similarly, we have $\bar{\mu}_w \circ w^k = 0$. Therefore, the system (14) can be written as:

$$\begin{array}{lll} A^T\lambda^f + M^T\lambda^q & = & c \\ B^T\lambda^f + N^T\lambda^q + \bar{\mu}^y & = & d \\ -\lambda^q + \bar{\mu}^w & = & 0 \\ (Ax^k + By^k - f) \circ \lambda^f & = & 0 \\ y^k \circ \bar{\mu}^y & = & 0 \, = \, w^k \circ \bar{\mu}^w \\ \lambda^f & \geq & 0 \\ \text{and} \quad (x^k, y^k, w^k) & \in & \Omega, \end{array}$$

i.e., the weak stationarity conditions hold, see (3). □

4.2 A bilinear penalty function in (y, w)

We next consider the bilinear penalty function:

$$\phi^{\mathrm{BL}}(y, w) \triangleq y^T w \tag{15}$$

that leads to the following penalized bilinear programming formulation of the LPCC (1): for $\rho > 0$,

$$\underset{z \triangleq (x,y,w) \in \Omega}{\text{minimize}} \; \theta^{\mathrm{BL}}(z) \triangleq c^T x + d^T y + \rho \sum_{i=1}^n y_i w_i. \tag{16}$$

Since the penalty function $\phi^{\mathrm{BL}}(y, w)$ is differentiable, the stationary solutions of the (nonconvex) program (16) follow the standard definition; namely, $\bar{z} \in \Omega$ is stationary if and only if $(z - \bar{z})^T \nabla \theta^{\mathrm{BL}}(\bar{z}) \geq 0$ for all $z \in \Omega$. The following result relates such a stationary solution, if feasible to the LPCC (1), to a strongly stationary solution of the LPCC.

Proposition 8. Let $\bar{z} \triangleq (\bar{x}, \bar{y}, \bar{w}) \in \Omega$ be given. The following two statements hold:

(a) If $\bar{z}$ is stationary for (16) for some $\rho > 0$ and is feasible to the LPCC (1), then $(\bar{x}, \bar{y}, \bar{w})$ is a strongly stationary point for (1).

(b) Conversely, if $\bar{z}$ is a strongly stationary point for (1), then $\bar{z}$ is a stationary point for (16) for all $\rho > 0$ sufficiently large.

Proof. Let $z = (x, y, w) \in \Omega$ be a feasible point for the relaxed LP (RxLP). The stationarity of $\bar{z}$ for (16) then yields

$$\begin{aligned} 0 \;&\leq\; (z - \bar{z})^T \nabla \theta^{\mathrm{BL}}(\bar{z}) \\ &=\; c^T(x - \bar{x}) + d^T(y - \bar{y}) + \rho \sum_{i=1}^n \left[\bar{w}_i (y_i - \bar{y}_i) + \bar{y}_i (w_i - \bar{w}_i) \right] \\ &=\; c^T(x - \bar{x}) + d^T(y - \bar{y}) + \rho \sum_{i \in \widehat{\mathcal{I}}_y(\bar{y},\bar{w})} \bar{w}_i \, y_i + \sum_{i \in \widehat{\mathcal{I}}_w(\bar{y},\bar{w})} \bar{y}_i \, w_i, \qquad \text{since } \bar{y} \text{ and } \bar{w} \text{ are complementary} \\ &=\; c^T(x - \bar{x}) + d^T(y - \bar{y}). \end{aligned}$$

The last equality follows from the fact that $\widehat{\mathcal{I}}_y(\bar{y}, \bar{w}) = \mathcal{I}_y(\bar{y}, \bar{w})$ and $\widehat{\mathcal{I}}_w(\bar{y}, \bar{w}) = \mathcal{I}_w(\bar{y}, \bar{w})$, given their definitions in (2) and (11), and because (x, y, w) is feasible to the LP (RxLP). Hence $\bar{z}$ is a minimizer of (RxLP) and (a) holds.

To prove (b), suppose $(\bar{x}, \bar{y}, \bar{w})$ is a strongly stationary point for (1). Let $(\bar{\lambda}, \mu^y, \mu^w)$ satisfy (3) along with the nonnegativity of μ_i^w and μ_i^y for all $i \in \mathcal{I}_0(\bar{y}, \bar{w})$. It remains to show that a multiplier $\bar{\mu}$ exists such that

$$\begin{array}{rcl} 0 & = & c - A^T\bar{\lambda} - M^T\bar{\mu} \\ 0 & \leq & \bar{y} \perp d + \rho\,\bar{w} - B^T\bar{\lambda} - N^T\bar{\mu} \geq 0 \\ 0 & \leq & \bar{w} \perp \rho\,\bar{y} + \bar{\mu} \geq 0 \\ 0 & \leq & \bar{\lambda} \perp A\bar{x} + B\bar{y} - f \geq 0. \end{array} \tag{17}$$

We claim that $\bar{\mu} \triangleq \mu^w$ does the job. Since (3) gives $d = B^T\bar{\lambda} + N^T\mu^w + \mu^y$, it suffices to show that $\bar{y} \perp \mu^y + \rho\bar{w} \geq 0$ and $\bar{w} \perp \mu^w + \rho\bar{y} \geq 0$ for all $\rho > 0$ sufficiently large. Defining

$$\bar{\rho} \triangleq \max\left\{0, \left\{-\frac{\mu_i^y}{\bar{w}_i} \mid \bar{w}_i > 0\right\}, \left\{-\frac{\mu_i^w}{\bar{y}_i} \mid \bar{y}_i > 0\right\}\right\}$$

we see that both $\mu^y + \rho\bar{w}$ and $\mu^w + \rho\bar{y}$ are nonnegative for $\rho > \bar{\rho}$. Lastly, the remaining two orthogonality conditions $\bar{y} \perp \mu^y + \rho\bar{w}$ and $\bar{w} \perp \mu^w + \rho\bar{y}$ follow from the strong stationarity conditions of $(\bar{x}, \bar{y}, \bar{w})$. □

While Proposition 4 has shown that the local minimizers of the penalized piecewise linear program (9) that are feasible to (1) are the same as the local minimizers of the LPCC (1), this equivalence of such minimizers between (16) and (1) is not addressed by Proposition 8. Borrowing an example from [5, 33] we show that a local (or even global) minimizer of (1) may not be obtainable from solutions of the bilinear penalty formulation for any fixed $\rho > 0$; thus, there are solutions to the original LPCC that are "elusive" from the penalized bilinear program.

Example 9. Consider

$$\begin{array}{ll} \underset{(x,y,w)\in\mathbb{R}^3}{\text{minimize}} & -x + y + w \\ \text{subject to} & -x + 4y \geq 0 \\ & -x + 4w \geq 0 \\ \text{and} & 0 \leq y \perp w \geq 0. \end{array}$$

It is not difficult to see that $(x, y, w) = (0, 0, 0)$ is a global and therefore local minimizer for this LPCC. Consider the penalized problem

$$\begin{array}{ll} \underset{(x,y,w)\in\mathbb{R}^3}{\text{minimize}} & -x + y + w + \rho\,yw \\ \text{subject to} & -x + 4y \geq 0 \\ & -x + 4w \geq 0 \\ \text{and} & y, w \geq 0 \end{array}$$

for $\rho > 0$. With $y = w = x/4$, the objective function is equal to $-x/2 + \rho x^2/16$ which is negative when $0 < x < 8/\rho$. Thus $(x, y, w) = (0, 0, 0)$ is not locally minimizing the penalized bilinear program for any $\rho > 0$. □

In [22], the bilinear function ϕ^{BL} is written as a DC function by adding and subtracting the norm of the vectors y and w:

$$\phi^{\mathrm{BL}}(y,w) \triangleq \underbrace{y^T w + \frac{\gamma}{2}\left(\|y\|_2^2 + \|w\|_2^2\right)}_{\triangleq\, \phi_+^{\mathrm{BL}_1}(y,w)} - \underbrace{\frac{\gamma}{2}\left(\|y\|_2^2 + \|w\|_2^2\right)}_{\triangleq\, \phi_-^{\mathrm{BL}_1}(y,w)}, \tag{18}$$

where $\phi_+^{\mathrm{BL}_1}$ is convex if $\gamma > 1$. This results in the DC program (6) having $f_1^{\mathrm{BL}_1}(z) \triangleq c^T x + d^T y + \rho\phi_+^{\mathrm{BL}_1}(y,w)$ and $f_2^{\mathrm{BL}_1}(z) \triangleq \rho\phi_-^{\mathrm{BL}_1}(y,w)$. The subproblem (7) in the DCA applied to the resulting penalized DC program is a convex quadratic program:

$$\underset{(x,y,w)\in\Omega}{\text{minimize}}\ c^T x + d^T y + \rho\left[y^T w + \frac{\gamma}{2}\left(\|\,y - y^k\,\|_2^2 + \|\,w - w^k\,\|_2^2\right)\right]. \tag{19}$$

In addition, we propose an alternative DC representation of the bilinear function ϕ^{BL}:

$$\phi^{\mathrm{BL}}(y,w) = \underbrace{\frac{1}{4}\|y+w\|_2^2}_{\triangleq\, \phi_+^{\mathrm{BL}_2}(y,w)} - \underbrace{\frac{1}{4}\|y-w\|_2^2}_{\triangleq\, \phi_-^{\mathrm{BL}_2}(y,w)}. \tag{20}$$

This DC decomposition has been used in [35] albeit not as a DC objective. The advantage of the DC pair $(\phi_+^{\mathrm{BL}_2}, \phi_-^{\mathrm{BL}_2})$ is that it does not depend on a parameter γ that might need to be adjusted for good performance. The subproblem (7) in the DCA applied to the resulting penalized DC program is the following convex quadratic program:

$$\underset{(x,y,w)\in\Omega}{\text{minimize}}\ c^T x + d^T y + \frac{\rho}{4}\|y+w\|_2^2 - \frac{\rho}{2}\left(y^k - w^k\right)^T (y-w), \tag{21}$$

which is different from (19). From Proposition 8, we may deduce that the DCA applied to the penalized bilinear formulation (9) for a given $\rho > 0$ will compute a strongly stationary solution of the LPCC (1) if the computed solution satisfies the complementarity condition.

4.3 Quadratic penalty function in (x, y)

An equivalent formulation of the LPCC (1) is obtained by eliminating the variable w. This leads to the following formulation:

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & c^T x + d^T y \\ \text{subject to} & Ax + By \geq f \\ \text{and} & 0 \leq y \perp Mx + Ny + q \geq 0. \end{array}$$

Using the penalty function

$$\phi^{-w}(x,y) \triangleq y^T(Mx + Ny + q) = q^T y + \tfrac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \underbrace{\begin{bmatrix} 0 & M^T \\ M & N + N^T \end{bmatrix}}_{\triangleq\, D} \begin{pmatrix} x \\ y \end{pmatrix},$$

where D is a symmetric indefinite matrix, we obtain the formulation

$$\underset{(x,y)\in\Omega_{xy}}{\text{minimize}}\ c^T x + d^T y + \rho\,\phi^{-w}(x,y) \tag{22}$$

with $\Omega_{xy} \triangleq \{(x,y) \in \mathbb{R}^{t\times n} \mid Ax + By \geq f,\ Mx + Ny + q \geq 0,\ y \geq 0\}$. Similar to Proposition 8, we have the relationship between stationary points.

Proposition 10. Let $(\bar{x}, \bar{y}) \in \Omega_{xy}$ be given. The following two statements hold:

(a) Suppose $(\bar{x}, \bar{y})$ is a stationary point for (22) and let $\bar{w} \triangleq q + M\bar{x} + N\bar{y}$. If $(\bar{x}, \bar{y}, \bar{w})$ is a feasible solution the LPCC (1) for some $\rho > 0$, then $(\bar{x}, \bar{y}, \bar{w})$ is a strongly stationary solution for (1).

(b) Conversely, if $(\bar{x}, \bar{y}, \bar{w})$ is a strongly stationary point for (1), then $(\bar{x}, \bar{y})$ is stationary for (22) for all $\rho > 0$ sufficiently large.

Proof. (a) If $(\bar{x}, \bar{y}) \in \Omega_{xy}$ is a stationary point for (22), then constraint multipliers $\bar{\lambda}$ and $\bar{\mu}$ exist such that

$$\begin{aligned} 0 &= c - A^T\bar{\lambda} - M^T(\,\bar{\mu} - \rho\,\bar{y}\,) \\ 0 &\leq \bar{y} \perp d + \rho\,[\,q + M\bar{x} + N\bar{y}\,] - B^T\bar{\lambda} - N^T(\,\bar{\mu} - \rho\,\bar{y}\,) \geq 0 \\ 0 &\leq \bar{\lambda} \perp A\bar{x} + B\bar{y} - f \geq 0 \\ 0 &\leq \bar{\mu} \perp M\bar{x} + N\bar{y} + q \geq 0. \end{aligned} \tag{23}$$

By letting $\mu^w \triangleq \bar{\mu} - \rho\bar{y}$ and $\mu^y \triangleq d - B^T\bar{\lambda} - N^T\mu^w$, we now show that (3) holds. The first two conditions in (23) can be written as

$$\begin{aligned} c &= A^T\bar{\lambda} + M^T\mu^w \\ d &= B^T\bar{\lambda} + N^T\mu^w + \mu^y \\ 0 &\leq \bar{y} \perp \rho\,\bar{w} + \mu^y \geq 0. \end{aligned}$$

Because $\bar{y}$ and $\bar{w}$ are feasible for the LPCC (1), we have $\bar{y} \circ \bar{w} = 0$, and hence $\bar{y} \circ \mu^y = \bar{y} \circ (\rho\,\bar{w} + \mu^y) = 0$. Furthermore, since $\bar{w}_i = 0$ for $i \in \mathcal{I}_0(\bar{y}, \bar{w})$, we also have $\mu^y_i \geq 0$ for $i \in \mathcal{I}_0(\bar{y}, \bar{w})$ from the last line above. Similarly, because $\bar{y} \circ \bar{w} = 0$ from (1), conditions (23) imply $\mu^w \circ \bar{w} = \bar{\mu} \circ (\,M\bar{x} + N\bar{y} + q\,) = 0$. Finally, since $\bar{y}_i = 0$ for $i \in \mathcal{I}_0(\bar{y}, \bar{w})$, we also have $\mu^w_i = \bar{\mu}_i \geq 0$ for $i \in \mathcal{I}_0(\bar{y}, \bar{w})$ by the definition of μ^w.

(b) Conversely, suppose $(\bar{x}, \bar{y})$ is a strongly stationary point for (1). Let $(\bar{\lambda}, \mu^y, \mu^w)$ satisfy (3) along with the nonnegativity of μ^w_i and μ^y_i for all $i \in \mathcal{I}_0(\bar{y}, \bar{w})$. We claim that by defining $\bar{\mu} \triangleq \mu^w + \rho\bar{y}$, (23) holds for all $\rho > 0$ sufficiently large. In turn, it suffices to show the second and fourth complementarity conditions in (23) for all $\rho > 0$ sufficiently large. Since $d = B^T\bar{\lambda} + N^T\mu^w + \mu^y$, we have $d + \rho\,(q + M\bar{x} + N\bar{y}) - B^T\bar{\lambda} - N^T(\bar{\mu} - \rho\,\bar{y}) = \rho\,(q + M\bar{x} + N\bar{y}) = \rho\,\bar{w}$; hence the second complementarity condition in (23) holds for all $\rho > 0$ because $\bar{y}$ and $\bar{w}$ are feasible for the LPCC (1). For the fourth complementarity condition to hold, choose $\rho > 0$ sufficiently large so that $\mu^w_i + \rho\bar{y}_i \geq 0$ if $\bar{y}_i > 0$. For the remaining indices with $\bar{y}_i = 0$ we have two cases: (i) If $\bar{w}_i > 0$, then $\bar{\mu}_i = \mu^w_i = 0$ by complementarity; (ii) if $\bar{w}_i = 0$, then $i \in= \mathcal{I}_0(\bar{y}, \bar{w})$ and $\bar{\mu}_i = \mu^w_i \geq 0$, because $(\bar{x}, \bar{y})$ is strongly stationary. $\square$

The penalty function ϕ^{-w} can be rewritten as DC function

$$\phi^{-w}(x, y) = \frac{1}{2}\left[\underbrace{\begin{pmatrix} x \\ y \end{pmatrix}^T D \begin{pmatrix} x \\ y \end{pmatrix} + \gamma\,(\,\|x\|_2^2 + \|y\|_2^2\,)}_{\phi_+^{-w}(x,y)}\right] - \underbrace{\frac{\gamma}{2}\,(\,\|x\|_2^2 + \|y\|_2^2\,)}_{\phi_-^{-w}(x,y)}. \tag{24}$$

The function ϕ_+^{-w} is convex if γ is larger than the absolute value of smallest eigenvalue of the symmetric indefinite matrix D. The resulting subproblems in DCA are of the form:

$$\underset{(x,y)\in\Omega_{xy}}{\text{minimize}}\; c^Tx + d^Ty + \frac{\rho}{2}\left[\begin{pmatrix} x \\ y \end{pmatrix}^T D \begin{pmatrix} x \\ y \end{pmatrix} + \gamma\,(\,\|x\|_2^2 + \|y\|_2^2\,)\right] - \rho\,\gamma\,\left[\,(\,y^k\,)^Ty + (\,x^k\,)^Tx\,\right]. \tag{25}$$

In summary, while the theoretical properties of the quadratic penalty ϕ^{-w} without the w-variable are the same as those of the bilinear penalty ϕ^{BL}, the respective subproblems in the DCA differ, and a different sequence of iterates is generated.

5 Enhancements of the DCA

The straightforward application of the DCA to the penalty formulations exhibits some inefficiencies. In the case of the piecewise linear function (Section 4.1), the method might terminate at a point that is weakly stationary but not strongly stationary and has an inferior objective value. When a quadratic penalty function (Sections 4.2 and 4.3) is used, we observed that the method could suffer from the slow (linear) convergence inherent to first-order methods, because the concave part of the DC function is approximated by a linear function. In what follows, we propose modifications to address these specific shortcomings.

5.1 Improved DCA for the piecewise linear penalty

In subproblem (12) with the piecewise linear penalty function, a parameter $\xi_i^k \in [0,1]$ has to be chosen for $i \in \mathcal{J}_=^k$. To avoid any bias for either one of the two complementary variables, we usually choose $\xi_i^k = \frac{1}{2}$. However, we noticed that in some instances improving directions existed at the points at which the DCA terminated and that were missed by the method. In following example, the LP subproblem (12) produces a zero step at an iterates that is not a strongly stationary point for the LPCC.

Example 11. Consider the problem

$$\begin{array}{ll} \underset{x,y,w}{\text{minimize}} & -x + 2y \\ \text{subject to} & -x \geq -10 \\ & -y \geq -10 \\ & x - w = 0 \\ \text{and} & 0 \leq y \perp w \geq 0. \end{array}$$

The unique strongly stationary point of this problem lies in $z^* = (x^*, y^*, w^*) = (10, 0, 10)$. Applying the DCA for the piecewise linear penalty formulation, with starting point $(x, y, w) = (5, 10, 5)$ and penalization parameter $\rho > 2$, the first linear subproblem solved has a negative objective gradient equal to d_0. Therefore, the DCA will reach the point $(0, 0, 0)$ at the first iteration. If, at the second subproblem, the chosen subgradient is $\xi^2 = (\frac{1}{2}, \frac{1}{2})$, the negative gradient becomes d_1 and, hence, the solution will not move and the method will stop at this weakly stationary point. Figure 1 illustrates how these two iterates are produced.

However, if a different element $\xi^2 = (0, 1)$ of the subgradient $\partial \min(y_i, w_i)$ is chosen, the DC subproblem produces z^* as the next iterate, and the algorithm terminates at a strongly stationary point. □

This example leads us to derive a heuristic that attempts to escape a weakly stationary point once the regular DCA has terminated. At the final iterate we choose different subgradients of $\min(y_i, w_i)$ that may not neutralize potential improving directions. The heuristic utilizes the multipliers $\bar{\mu}^{y,k}$ and $\bar{\mu}^{w,k}$ in the LP subproblem (12) for the non-negativity restrictions of y and w, respectively. In order to recover the quantities that correspond to the multipliers μ^y and μ^w in the stationarity conditions (3) for the LPCC,

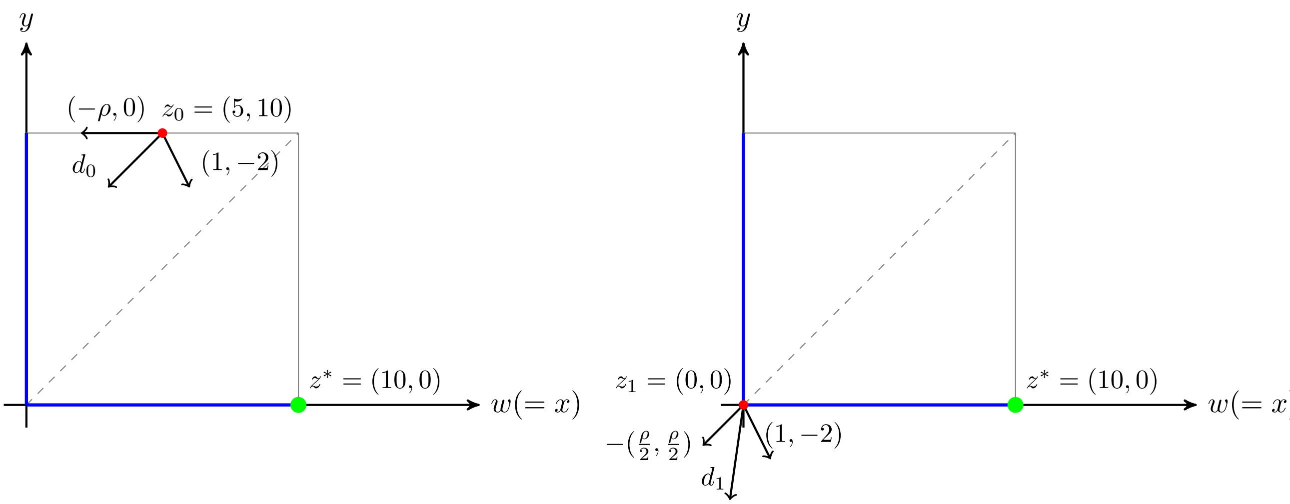


Figure 1: DCA for the piecewise linear penalty function converges to a weakly stationary point.

we need to subtract the influence of the penalty term and compute for each index $i \in \mathcal{I}_0(y^{k+1}, w^{k+1})$

$$\begin{aligned}
&\mu_i^{y,k} = \bar{\mu}_i^{y,k} - \rho && \mu_i^{w,k} = \bar{\mu}_i^{w,k} && \text{if } i \in \widehat{\mathcal{J}}_y^k, \\
&\mu_i^{y,k} = \bar{\mu}_i^{y,k} && \mu_i^{w,k} = \bar{\mu}_i^{w,k} - \rho && \text{if } i \in \widehat{\mathcal{J}}_w^k, \\
&\mu_i^{y,k} = \bar{\mu}_i^{y,k} - \rho\xi_i^k && \mu_i^{w,k} = \bar{\mu}_i^{w,k} - \rho(1 - \xi_i^k) && \text{if } i \in \widehat{\mathcal{J}}_=^k.
\end{aligned}$$

If $z^{k+1} = (x^{k+1}, y^{k+1}, w^{k+1})$ is feasible for the LPCC (1) and $\mu_i^{y,k} \geq 0$ and $\mu_i^{w,k} \geq 0$ for all $i \in \mathcal{I}_0(y^{k+1}, w^{k+1})$, then z^{k+1} is a strongly stationary point for the LPCC (1). Otherwise, we use the adjusted multipliers to choose the subgradient to restart the DCA: For each $i \in \mathcal{I}_0(y^{k+1}, w^{k+1})$ we set

$$\xi_i^{k+1} = \begin{cases} 1 & \text{if } \mu_i^{y,k} < \mu_i^{w,k} - \varepsilon_\xi \\ \frac{1}{2} & \text{if } \left| \mu_i^{y,k} - \mu_i^{w,k} \right| \geq \varepsilon_\xi \\ 0 & \text{if } \mu_i^{y,k} > \mu_i^{w,k} + \varepsilon_\xi \end{cases} \tag{26}$$

and $\xi_i^{k+1} = \frac{1}{2}$ if $i \in \widehat{\mathcal{J}}_=^{k+1} \setminus \mathcal{I}_0(y^{k+1}, w^{k+1})$. The small tolerance ε_ξ (chosen to be 10^{-8} in our experiments) accounts for potential numerical error of the subproblem solution. If the resulting step reduces the objective function, we resume the regular DCA; otherwise the improved method terminates.

5.2 Improved DCA for the quadratic penalty function in (y, w)

In the standard DCA for the quadratic penalty functions, the nonlinear concave part of the quadratic penalty problem is approximated by a linear function. We observed that this can lead to slow convergence. To accelerate the method, the improved DCA described next fixes variables to zero for which complementarity has already been achieved, and deletes the penalty terms for those components in the subproblem objective. In the extreme case, when one variable in all complementarities has been fixed, the subproblem becomes an LP, and the next iterate jumps immediately to the optimal solution in that piece. This saves many iterations compared to the original method whose convergence may be slowed by the penalty term. Even when not all complementarities have been fixed, the enhanced version has to resolve fewer complementarities with a linear approximation of the quadratic penalty function than the original version, and faster convergence can be expected.

We note that the subproblems (19) and (21) can be rewritten in a unified form with each complementarity having its own quadratic penalty function:

$$\min_{(x,y,w)\in\Omega} \; c^T x + d^T y + \rho \sum_{i=1}^{n} \psi_i^k(y_i, w_i).$$

By construction of the DC subproblem (7), we have that the gradient of the original DC function and that of its convex approximation coincide at the current iterate; i.e.

$$\nabla \psi_i^k(y_i^k, w_i^k) = \nabla \phi_i(y_i^k, w_i^k) = \begin{pmatrix} w_i^k \\ y_i^k \end{pmatrix} \tag{27}$$

where $\phi_i(y_i, w_i) = y_i \cdot w_i$. In our proposed variation of the DCA, we maintain the set $\mathcal{F}_y^k$ of indices i for which y_i^k is fixed at zero in iteration k. The index set $\mathcal{F}_w^k$ is defined analogously for w^k, and we let $\mathcal{F}^k = \mathcal{F}_y^k \cup \mathcal{F}_w^k$. These sets try to predict which sides of the complementarities are active (zero) at the solution. The modified subproblem that delivers the next iterate is then defined as

$$\begin{aligned} &\underset{(x,y,w)\in\Omega}{\text{minimize}} && c^T x + d^T y + \rho \sum_{i \notin \mathcal{F}^k} \psi_i^k(y_i, w_i) \\ &\text{subject to} && y_i = 0 \text{ for } i \in \mathcal{F}_y^k \\ &\text{and} && w_i = 0 \text{ for } i \in \mathcal{F}_w^k. \end{aligned} \tag{28}$$

Note that the more variables become fixed in $\mathcal{F}^k$, the fewer free variables there are in the subproblem, leading to a reduction in computation time. The sets $\mathcal{F}_y^k$ and $\mathcal{F}_w^k$ are updated throughout the course of the algorithm. An index i enters the set $\mathcal{F}_y^{k+1}$, if $i \notin \mathcal{F}_y^k$ and $y_i^{k+1} = 0$; i.e., the variable y_i has just become zero. The update rule for w is similar.

We must also account for the possibility that our prediction is incorrect and a variable has to be released because it might be nonzero at the optimal solution. Recall that, ultimately, we want to solve the penalty formulation (4) of the LPCC with $\phi(y, w) = \phi^{\text{BL}}(y, w) = y^T w$ defined in (15). The first-order optimality conditions for (16) are

$$\begin{array}{lcl} A^T\lambda^f + M^T\lambda^q & = & c \\ B^T\lambda^f + N^T\lambda^q + \mu^y & = & d + \rho\, w \\ -\lambda^q + \mu^w & = & \rho\, y \\ 0 \le \lambda^f \perp Ax + By - f & \ge & 0 \\ 0 \le y \perp \mu^y & \ge & 0 \\ 0 \le w \perp \mu^w & \ge & 0 \\ (x, y, w) & \in & \Omega. \end{array} \tag{29}$$

On the other hand, the first-order optimality conditions for (28) can be written as

$$
\begin{array}{lcl}
A^T\lambda^f + M^T\lambda^q & = & c \\
B^T\lambda^f + N^T\lambda^q + \bar{\mu}^y & = & d + \rho\sum_{i\notin\mathcal{F}^k} w_i^k e_i \\
-\lambda^q + \bar{\mu}^w & = & \rho\sum_{i\notin\mathcal{F}^k} y_i^k e_i \\
0 \le \lambda^f \perp Ax + By - f & \ge & 0 \\
0 \le y \perp \bar{\mu}^y & \ge & 0 \\
0 \le w \perp \bar{\mu}^w & \ge & 0 \\
y_i & = & 0 \text{ for } i \in \mathcal{F}_y^k \\
w_i & = & 0 \text{ for } i \in \mathcal{F}_w^k \\
(x, y, w) & \in & \Omega.
\end{array}
\tag{30}
$$

where we used (27) in the second and third equation, and $e_i \in \mathbb{R}^n$ denotes the i^{th} coordinate vector.

Consider an iterate (x^k, y^k, w^k), and let $(x^{k+1}, y^{k+1}, w^{k+1}, \lambda^f, \lambda^q, \bar{\mu}^y, \bar{\mu}^w)$ be a primal-dual solution of the subproblem satisfying the optimality conditions (30). With

$$
\begin{array}{ll}
\mu_i^y = \bar{\mu}_i^y \text{ if } i \notin \mathcal{F}_y^k & \qquad \mu_i^y = \bar{\mu}_i^y + \rho\, w_i^k \text{ if } i \in \mathcal{F}_y^k \\
\mu_i^w = \bar{\mu}_i^w \text{ if } i \notin \mathcal{F}_w^k & \qquad \mu_i^w = \bar{\mu}_i^w + \rho\, y_i^k \text{ if } i \in \mathcal{F}_w^k,
\end{array}
$$

the tuple $(x^{k+1}, y^{k+1}, w^{k+1}, \lambda^f, \lambda^q, \mu^y, \mu^w)$ satisfies the first-order optimality conditions (29) for the penalty problem (4), as long as $y^k = y^{k+1}$ and $w^k = w^{k+1}$ and

$$
\mu_i^y \ge 0 \text{ if } i \in \mathcal{F}_y^k \quad \text{ and } \quad \mu_i^w \ge 0 \text{ if } i \in \mathcal{F}_w^k. \tag{31}
$$

We use this observation to decide when to release a variable from the sets $\mathcal{F}_y^k$ and $\mathcal{F}_w^k$ of fixed components: If (31) does not hold, then there is a good chance that a multiplier for a fixed variable would have a negative sign at the solution, and the variable should therefore not forced to be zero. The overall method is summarized in Algorithm 2.

6 Computational Results

This section explores the practical performance of the DC decompositions described earlier for three different problem sets. The results confirm that the enhancements proposed in Section 5 contribute in one of two ways for the basic DCA applied to these decompositions: (a) for the piecewise linear penalty function, solutions with better objective values are found, or (b) for the bilinear penalty function, termination occurs faster, thus speeding up the algorithm. In short, the enhanced DC methods provide a competitive option for solving LPCCs.

The experiments were performed on a Linux workstation with 3.10GHz Xeon processors using 20 cores (40 cores hyperthreaded) and 256GB RAM. The DCA and its enhancements were implemented in MATLAB (R2014a). To compare the performance of our methods with efficient alternatives, we solved all instances also with the FILTER software, a sophisticated implementation of an active-set SQP method for nonlinear programs with special features to solve MPCCs [7, 8]. FILTER was called from AMPL [11], with AMPL's presolve disabled. We have considered the KNITRO nonlinear programming solver which has the capability to handle complementarity constraints with a penalty function [25]; but we do not describe its performance

Algorithm 2 Accelerated DCA for the quadratic penalty function in (y, w)

Choose termination tolerance $\varepsilon_{\text{tol}} > 0$.
Let $(x^1, y^1, w^1) \in \Omega$, $k \leftarrow 0$, $\mathcal{F}_y^1 = \emptyset$, and $\mathcal{F}_w^1 = \emptyset$.
repeat
    Set $k \leftarrow k + 1$.
    Solve subproblem (28).
    Let $(x^{k+1}, y^{k+1}, w^{k+1})$ be an optimal solution of (28), and let $\bar{\mu}_i^y$ and $\bar{\mu}_i^w$ be the multipliers in (30).
    Update the set of fixed variables:

$$\mathcal{F}_y^{k+1} = \left\{ i \notin \mathcal{F}_y^k : y_i^{k+1} = 0 \right\} \cup \left\{ i \in \mathcal{F}_y^k : \bar{\mu}_i^y + \rho\, w_i^{k+1} \geq 0 \right\}$$
$$\mathcal{F}_w^{k+1} = \left\{ i \notin \mathcal{F}_w^k : w_i^{k+1} = 0 \right\} \cup \left\{ i \in \mathcal{F}_w^k : \bar{\mu}_i^w + \rho\, y_i^{k+1} \geq 0 \right\}$$

until $c^T x^k + d^T y^k + \rho\, \phi(y^k, w^k) - (c^T x^{k+1} + d^T y^{k+1} + \rho\, \phi(y^{k+1}, w^{k+1})) \leq \varepsilon_{\text{tol}}$
Return $(x^{k+1}, y^{k+1}, w^{k+1})$.

here since it was unable to solve many of the test problems. All these approaches only attempt to find local optima. To assess the solution quality, we also consider the MILP reformulation of the LPCC,

$$\begin{array}{ll} \underset{x,y,w,z}{\text{minimize}} & c^T x + d^T y \\ \text{subject to} & Ax + By \geq f \\ & Mx + Ny + q = w \\ & 0 \leq y \leq u^y \circ z \\ & 0 \leq w \leq u^w \circ (1 - z) \\ \text{and} & z \in \{0, 1\}^n, \end{array} \tag{32}$$

that can compute the global optimum of the instances. Here, u^y and u^w are explicit upper bounds for y and w. In the case where such upper bounds were not explicitly available, a large value (1,000) was used. A time limit of 900 seconds (wall clock time) was imposed on the MILP solver, using 32 threads. This corresponds roughly to up to 8 hours of CPU time. If no optimal solution was found after this time, the solver returned the best incumbent. All LPs and MILPs in our experiments were solved with CPLEX 12.6.2 using default settings.

Three sets of problems were tested: (a) Linear Complementarity Problems (LCP) instances from [22] (Section 6.1), (b) linearizations of problems from the MacMPEC library [26] (Section 6.2), and (c) random instances of the inverse quadratic problem (Section 6.3). For easiness of notation, decompositions (9), (18), (20), and (24) are abbreviated as PL, BL1, BL2, and B-w, respectively. For each decomposition, two different starting points were tried. The first one, denoted by E, sets $y = w = \mathbf{1}$, and the other, called R, uses the solution to the LP relaxation of the LPCC, which is defined as (4) with $\rho = 0$. Note that the choice of the initial value of x is irrelevant, because it does not appear in $f_2(z)$ of the decomposition (5) and has no effect on the subproblem (7). Finally, enhanced methods are marked with a star (*). For example, BL2-E* refers to the enhanced DC approach based on decomposition (20), with starting point on $y = w = \mathbf{1}$. For decomposition (18), the value of γ is also specified. Decomposition (24) uses $\gamma = \max\{0, -\lambda_{\min}\} + 0.01$, where $\lambda_{\min}$ is the smallest eigenvalue of matrix D defined in Section 4.3.

The implemented strategy for the penalization parameter ρ is given in Algorithm 3. It starts with a small value (namely, 1 in our experiments) and solves the penalty problem (4) with a termination tolerance of $\varepsilon_{\text{tol}} = 10^{-8}$. If the solution is already considered complementary, i.e. the infeasibility measure $V_{\perp}(y, w) \triangleq \max_i \{\min\{y_i, w_i\}\}$ is below a tolerance of $\varepsilon_{\perp} = 10^{-8}$, increasing ρ would have no effect (see Propositions 4, 8, and 10), and therefore we stop in Step 4. Otherwise, we increase ρ by a constant factor (in our case, 10) and solve the DC problem with the new ρ. If ρ reaches a maximum allowed value of 10^9, the algorithm terminates in Step 9.

Algorithm 3 Overall algorithm with increasing penalty parameter

```
1: Choose starting penalization parameter ρ > 0, increase factor π > 1, upper bound ρ̄, termination
   tolerance ε_tol > 0, and complementarity tolerance ε_⊥ > 0.
2: repeat
3:     Solve DC problem (4) using the DC algorithm with current ρ and tolerance ε_tol. Obtain solution
   (x, y, w) ∈ Ω.
4:     if V_⊥(y, w) ≤ ε_⊥ then
5:         Exit loop
6:     else
7:         Update ρ := πρ
8:     end if
9: until ρ ≥ ρ̄
10: Return (x, y, w).
```

We distinguish the following outcomes of the algorithm. Counting one subproblem solution as one iteration, we terminate with "Maximum number of iterations reached" when the iteration limit of 100 subproblem solves is exceeded. The algorithm stops with "Successful termination" if Algorithm 3 terminates because the complementarity tolerance in Step 4 is satisfied. However, there are also cases in which Algorithm 3 terminates because ρ exceeded $\bar{\rho}$ in Step 9, which we label as "Maximum value of penalty parameter reached." The latter case deserves closer examination. It is the expected outcome when the problem is (locally) infeasible. However, we frequently also observed the following situation in our experiments: The inner DC algorithm in Step 3 of Algorithm 3 terminates because the penalized objective function is not changing by more than $\varepsilon_{\text{tol}} = 10^{-8}$ between iterations, but the feasibility test in Step 4 is not satisfied, even though $V_{\perp}(y, w)$ is quite small and it appears that the iterates are close to a stationary point. The method then repeatedly increases ρ, taking only one step per penalty parameter value, until ρ reaches its limit. We will examine this situation in more detail in Section 6.2. For now, we will count a run as converged, if the method does not run out of iterations, and if the infeasibility measure $V_{\perp}(y, w)$ is at most $\varepsilon_{\text{feas}}$, where $\varepsilon_{\text{feas}} = 10^{-5}$ unless otherwise specified.

6.1 Linear complementarity problem instances

The LCP consists in finding a vector $x \in \mathbb{R}^n$ which satisfies

$$0 \leq x \perp \tilde{A}x - \tilde{b} \geq 0,$$

where $\tilde{A} \in \mathbb{R}^{n \times n}$ and $\tilde{b} \in \mathbb{R}^n$ are given. We reformulate the LCP as an LPCC by setting c, d, M, A, B and f to zero and letting $N = \tilde{A}$ and $q = -\tilde{b}$. Notice that this is a feasibility problem, so reaching strongly stationary points here has no special meaning. Therefore, results for the enhanced piecewise linear penalty function are not reported. In [22], the authors proposed the solution of LCPs with the DC algorithm, decomposing different penalization functions for the complementarity term similar to those discussed here.

Method	Average time		Average iterations	
	Orig	Enhcd	Orig	Enhcd
PL-E	0.068	-	2.000	-
BL1-E-γ=1.01	4.906	1.099	8.497	3.533
BL1-E-γ=2	6.539	1.369	11.671	4.451
BL2-E	2.156	0.607	6.094	2.289
B-w-E	0.450	-	2.140	-
MILP	0.164	-	-	-
FILTER	0.281	-	-	-

Table 1: Geometric averages of computation times (in seconds) and iterations for the large LCP instances, including the original ("Orig") and enhanced ("Enhcd") variants of the bilinear decompositions.

Their methods DCA1, DCA2, and DCA4, correspond to B-w, PL, and BL1. The reference also contains a further penalty formulation, DCA3, which was shown therein not to perform well and so we did not include it in our approach for solving LPCCs.

In this first set of numerical experiments, we solved each of the twelve LCP instances given in [22]. The first six problems have dimensions $n \in \{2, 3, 4\}$. The remaining instances are scalable, and we chose $n = 1,000$, the largest size reported in this reference. Only the starting point E was used, since the LP relaxation is meaningless when no objective is given.

Similar to [22], most methods solve all the instances. The only exception is BL2-E in test case 5, where the final infeasibility is violated only slightly with $1.2 \cdot 10^{-5} > \varepsilon_{\text{feas}}$. Table 1 displays the geometric averages of CPU times and iterations for all methods aggregated over the six scalable problems. We used geometric averages because these are less biased towards the more time-consuming instances than the arithmetic averages. We excluded the small instances because their solution time was considered negligible (less than 0.15 seconds).

The piecewise linear method is clearly the fastest by an order of magnitude, because it requires only two LP solutions per problem. The B-w version also does well, with computation times comparable to the MILP formulation and the FILTER solver. Similar observations were made in [22]. Enhancements on the bilinear penalties reduce time and iterations by 70–78% and 58–63%, respectively. Note that relative decrease in computation time is larger than that in iteration count. This confirms the hypothesis in Section 5.2 that the fixing of the variables in the enhancement for the bilinear penalty function not only speeds up convergence, but also leads to smaller subproblems that can be solved more quickly. Nevertheless, the BL variants still require one to two orders of magnitude more computation time than the PL variants for these LCP instances.

6.2 MacMPEC library

The MacMPEC library [26] contains a collection of MPCCs, from which we constructed LPCC instances by linearizing the objective function and constraints around the origin. The original upper and lower variable bounds were kept, and set to 1,000 where no bounds were defined. A total of 128 instances were created. In order to select instances that are guaranteed to be feasible, we first solved the MILP formulations and identified 96 instances in which the MILP solver found a feasible point within a maximum time limit of 900 wall clock seconds. Thirteen of these 96 instances have complementarity dimension greater than 100.

Table 2 shows the number of instances that reach the different outcomes for each method. We note that the piecewise linear decomposition never exceeds the iteration limit, and that the enhancement for the bilinear decompositions helps to reduce the number of iterations and allows more instances to be solved within the limit. First we consider the quality of the returned final iterates in terms of feasibility. As mentioned at the end of the introduction of Section 6, the bilinear methods terminate when the progress in the penalty function becomes very small (below $\varepsilon_{\text{tol}} = 10^{-8}$), but this does not imply that complementarity variables that are zero at a solution are very small as well. For example, consider a complementary pair with $y_i = w_i = 10^{-4}$. The corresponding penalty term is $y_i \cdot w_i = 10^{-8}$, but the feasibility violation is $V_\perp(y_i, w_i) = 10^{-4}$. Therefore, the choice of the feasibility tolerance $\varepsilon_{\text{feas}}$ is very crucial when deciding which final iterates constitute a successful run of a method in our statistics. In Table 2 we see that the number of runs that are counted as successful increases significantly with $\varepsilon_{\text{feas}}$ for the bilinear methods, whereas it has very little effect for the piecewise linear methods. This demonstrates that the bilinear versions are less capable of resolving complementarities to high accuracy. One may argue that a much tighter tolerance ε_{tol} for the DCA should be chosen, but most likely this will lead to numerical difficulties since the subproblems are solved only up to a certain tolerance as well (CPLEX has a default tolerance of 10^{-6}). We also observe, however, that the enhancement for the bilinear versions, which relies on fixing variables to 0 exactly, clearly improves this aspect of the solutions. Given the observations in Table 2, we chose the feasibility tolerance $\varepsilon_{\text{feas}} = 10^{-5}$ for all our statistics.

Next we examine the relative performance of the methods in terms of objective function values, first among only the DC methods. Again, we see that the enhancements for the bilinear methods lead to a significant improvement in solution quality. The enhancement for the piecewise linear method has less of an effect. We also observe that, for this problem set, the unbiased E starting point typically leads to better solutions than the one based on the LP relaxation, and that the perturbation parameter $\gamma = 1.01$ appears to be the better choice for the BL1 variants. The B-w variant, missing an enhanced version, does not perform well. Finally, we assess the solution quality compared to the global optima (obtained by the MILP solver) and relative to the FILTER code. The "E*" variants find the global solution in 70-77% of the instances. Overall, among the DC methods, PL-E* and BL2-E* are the more effective variants for this problem set in terms of solution quality. A comparison with the FILTER code shows that this MPCC solver finds a better solution than the best DC variants more often than not.

Table 3 compares the geometric averages of computation time and iterations between original and enhanced methods. We only included the 52 instances where at least one method took over 1 second of wall clock time to solve. For each method, averages are computed over all runs with successful outcomes. Comparing original and improved versions, we again observe that the enhancement for the BL methods leads to a significant reduction in computation time and iteration count. As expected, the PL methods increase the times slightly when the improvement is enabled, because it is based on continuing the DCA with additional iterations that use different choices of the subgradients. We see that the average computation time for PL methods is slightly higher than that of MILP and FILTER, and the enhanced BL variants take 5-8 times as long as the PL versions. We also see that for these problems, the average number of iteration is around 3 for the PL methods, and 5-7 for the well-performing BL variants.

6.3 Medium-scale inverse QP instances

Consider the convex quadratic program (QP):

$$\begin{array}{ll} \underset{y}{\text{minimize}} & \frac{1}{2} y^T Q y + c^T y \\ \text{subject to} & Ay \geq b, \end{array} \tag{33}$$

Method	Max iter	$\varepsilon_{\text{feas}} = 10^{-8}$	$\varepsilon_{\text{feas}} = 10^{-5}$	$\varepsilon_{\text{feas}} = 10^{-2}$	Best DC	Global found	FILTER worse	FILTER better
PL-E	0	83	83	85	73	71	12	24
PL-R	0	72	72	78	68	66	7	28
PL-E*	0	83	83	85	76	74	12	21
PL-R*	0	72	72	78	68	66	7	28
BL1-E-γ=1.01	11	39	58	83	33	32	3	59
BL1-R-γ=1.01	6	32	47	85	22	21	6	67
BL1-E*-γ=1.01	2	82	83	91	76	73	8	17
BL1-R*-γ=1.01	2	75	75	89	66	63	9	28
BL1-E-γ=2	13	24	42	74	30	30	3	64
BL1-R-γ=2	7	21	40	80	19	19	6	71
BL1-E*-γ=2	2	66	76	85	70	67	5	23
BL1-R*-γ=2	3	72	73	85	64	61	9	32
BL2-E	10	45	66	83	43	41	8	55
BL2-R	9	42	62	84	35	33	7	64
BL2-E*	1	81	84	92	77	74	11	17
BL2-R*	2	77	78	91	71	68	12	26
B-w-E	58	28	28	30	27	27	3	65
B-w-R	31	46	46	50	45	45	5	46
FILTER	-	79	92	95	-	76	-	-

Table 2: Number of problems with specific outcomes. "Max iter": Maximum number of iterations exceeded (excluded from remaining statistics); "$\varepsilon_{\text{feas}} =$": Returned iterate satisfies $V_\perp \leq \varepsilon_{\text{feas}}$; "Best DC": as good as best DC method; "Global found": Global solution found; "FILTER better": Number of instances in which FILTER objective is better; "FILTER worse": Number of instances in which FILTER objective is worse.

Method	Average time		Average iterations		Successful	
	Orig	Enhcd	Orig	Enhcd	Orig	Enhcd
PL-E	0.031	0.032	2.768	3.268	46	46
PL-R	0.025	0.028	2.631	2.857	41	41
BL1-E-γ=1.01	0.552	0.140	16.784	5.026	29	45
BL1-R-γ=1.01	0.436	0.147	14.443	5.653	28	42
BL1-E-γ=2	0.739	0.153	23.449	5.912	21	42
BL1-R-γ=2	0.539	0.202	21.769	7.378	24	40
BL2-E	0.449	0.121	13.721	4.448	33	45
BL2-R	0.356	0.114	11.898	4.610	31	41
MILP	0.022	-	-	-	52	-
FILTER	0.019	-	-	-	52	-

Table 3: Geometric averages of computation times (in seconds) and iterations for MacMPEC instances, including the original ("Orig") and enhanced ("Enhcd") variants of the decompositions. The number of instances with successful outcomes is given in the last columns.

where $Q \in \mathbb{R}^{n\times n}$ is symmetric positive definite, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. For given matrices Q and A, the inverse convex quadratic problem, as described in [14], consists in finding vectors x, b, and c

that solve the forward problem (33) and are the least deviated from some given vectors $\bar{x}, \bar{b}$, and $\bar{c}$; that is (x, b, c) solves

$$\begin{array}{ll} \underset{x,b,c}{\text{minimize}} & \left\|(x,b,c) - (\bar{x}, \bar{b}, \bar{c})\right\|_1 \\ \text{subject to} & \left\{ \begin{array}{lll} x \in & \underset{y}{\arg\min} & \frac{1}{2} y^T Q y + c^T y \\ & \text{subject to} & Ay \geq b. \end{array} \right. \end{array}$$

By stating the KKT conditions in the second-level problem and reformulating the ℓ_1-norm with slack variables, we obtain the LPCC

$$\begin{array}{llllllll} \underset{x,b,c,z^x,z^b,z^c,\lambda}{\text{minimize}} & \sum_{i=1}^{n} z_i^x + \sum_{j=1}^{m} z_j^b + \sum_{i=1}^{n} z_i^c & & & & & \\ \text{subject to} & Qx + c - A^T\lambda = 0 & & & & & \\ & -z^x \leq & x - \bar{x} & \leq z^x, & -u^x \leq & x & \leq u^x \\ & -z^b \leq & b - \bar{b} & \leq z^b, & -u^b \leq & b & \leq u^b \\ & -z^c \leq & c - \bar{c} & \leq z^c, & -u^c \leq & c & \leq u^c \\ \text{and} & 0 \leq \lambda \perp & Ax - b & \geq 0, & & \lambda & \leq u^\lambda \end{array} \tag{34}$$

For our experiments, feasible random inverse QP instances were generated in MATLAB with the following procedure, given dimensions m and n, and a sparsity level s. The parameter s was chosen so that, on average, every matrix had at most 10 non-zero elements. We include explicit bounds u^x, u^b, u^c, and u^λ to compare with the MILP formulation.

1: Generate a sparse random symmetric positive definite matrix $Q \in \mathbb{R}^{n\times n}$ and a sparse matrix A in $\mathbb{R}^{m\times n}$, using the MATLAB commands SPRANDSYM$(n, s, 0.5, 1)$ and SPRAND(m, n, s).
2: Generate a random vector $x \in \mathbb{R}^n$ with components following a normal distribution $N(0, 1)$.
3: Generate vectors $\widetilde{\lambda}$ and $\widetilde{w} \in \mathbb{R}^m$ uniformly at random between 0 and 10.
4: Generate a random binary vector $v \in \{0, 1\}^m$ and set $\lambda = \widetilde{\lambda} \circ v$ and $w = \widetilde{w} \circ (\mathbf{1} - v)$, so that $\lambda \perp w$.
5: Define $b \triangleq Ax - w$ and $c \triangleq A^T\lambda - Qx$.
6: Perturb x, b and c with normally distributed $N(0, 1)$ noise to obtain vectors $\bar{x}, \bar{b}$ and $\bar{c}$.
7: Set upper bounds $u^x \triangleq 10 \max\{|x_i|\}$, $u^b \triangleq 10 \max\{|b_i|\}$, $u^c \triangleq 10 \max\{|c_i|\}$, $u^\lambda \triangleq 10 \max\{|\lambda_i|\}$.

By construction, vectors x, b, c, and λ are feasible to (34). A total of 120 instances were generated with complementarity dimension m equal to 10, 25, 50, 100, 250, and 500, and 20 instances per size. Dimension n was chosen as $0.75 \times m$. The experiments are summarized in Table 4 and Figure 2. In terms of the robustness of the methods, the final columns in Table 4 show once again that the enhancements for the BL methods are essential for good performance. We also see that the PL variants are successful in all instances. Since decomposition B-w showed poor results on these instances (solving less than 25% of them), it is not included in these statistics.

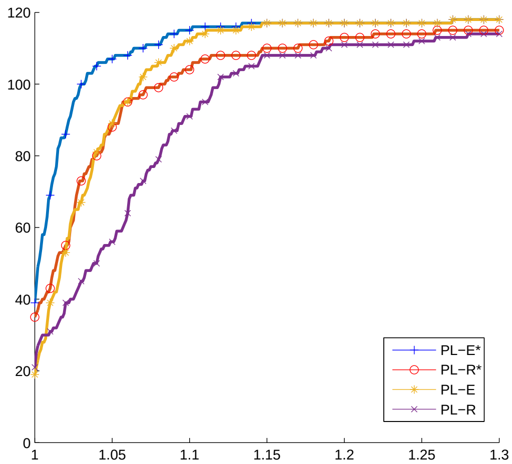
The plots in Figure 2 compare solution quality obtained by the various methods. The performance measure per instance is calculated as a ratio between the objective value of each method and the best solution among all methods in Table 4. If a method failed on an instance, its ratio is considered infinite. The profiles in the graphs count how many instances have a performance measure below a threshold, given on the x-axis. Note that the x-axis scale differs from one plot to another. Plot 2a illustrates that the enhancement for the PL methods shifts the curves upwards, implying that solution quality improves. In fact, the number of cases in which the method terminates at a point that is proven to be strongly

Method	$m=10$	$m=25$	$m=50$	$m=100$	$m=250$	$m=500$	It	Inf
PL-E	0.03 (20)	0.04 (20)	0.05 (20)	0.1 (20)	0.64 (20)	5.17 (20)	0	0
PL-R	0.03 (20)	0.03 (20)	0.04 (20)	0.08 (20)	0.45 (20)	3.81 (20)	0	0
PL-E*	0.03 (20)	0.05 (20)	0.07 (20)	0.12 (20)	0.72 (20)	6.60 (20)	0	0
PL-R*	0.03 (20)	0.04 (20)	0.05 (20)	0.11 (20)	0.53 (20)	5.87 (20)	0	0
BL1-E-γ=1.01	1.61 (17)	2.26 (3)	-	-	-	-	95	5
BL1-R-γ=1.01	0.80 (16)	1.60 (9)	-	-	-	-	87	8
BL1-E*-γ=1.01	0.18 (20)	0.32 (19)	0.65 (20)	2.28 (18)	7.62 (18)	22.99 (20)	0	5
BL1-R*-γ=1.01	0.14 (20)	0.24 (19)	0.62 (20)	2.09 (19)	6.33 (18)	20.88 (20)	0	4
BL1-E-γ=2	1.69 (6)	-	-	-	-	-	113	1
BL1-R-γ=2	1.27 (14)	1.70 (4)	-	-	-	-	99	3
BL1-E*-γ=2	0.27 (20)	0.51 (19)	1.06 (20)	3.46 (18)	12.3 (19)	32.63 (17)	3	4
BL1-R*-γ=2	0.16 (20)	0.36 (19)	0.91 (20)	3.09 (19)	9.57 (20)	29.88 (19)	1	2
BL2-E	0.99 (20)	1.60 (11)	2.78 (8)	6.62 (4)	-	-	71	6
BL2-R	0.69 (18)	1.06 (10)	2.50 (8)	-	-	-	75	9
BL2-E*	0.16 (20)	0.28 (19)	0.51 (20)	1.60 (19)	6.58 (20)	18.64 (20)	0	2
BL2-R*	0.11 (20)	0.21 (20)	0.36 (20)	1.43 (18)	4.79 (20)	15.11 (19)	0	3
MILP	0.17 (20)	0.49 (20)	0.87 (20)	147.16 (20)	904.38 (20)	900.58 (20)	0	0
FILTER	0.01 (20)	0.03 (20)	0.23 (20)	1.35 (20)	44.59 (20)	547.96 (20)	0	0

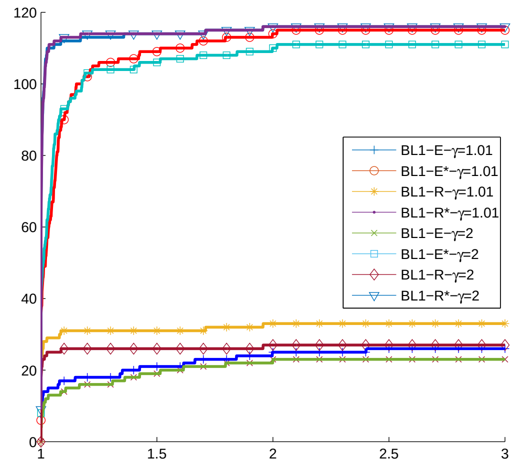
Table 4: Computation times on Inverse QP instances. For each method and problems size, the arithmetic averages of the computation times (in seconds) are given. The numbers within the parentheses list the number of successfully solved problems over which the averages are taken. The "It" column displays the number of instances in which a method ran out of iterations, and the "Inf" column the number of problems in which the complementarity violation is not below $\varepsilon_{\perp} = 10^{-5}$.

stationary increases from 28 (31) to 116 (115) out of 120 for the E (R) starting point. This demonstrates the effectiveness of the procedure in Section 5.1 that aims at escaping non-strongly stationary points by choosing different subgradients. The best performing method is PL-E*, which solves 100 problems with an objective value within 3% of the best solution, and 114 within 10%. Plots 2b and 2c confirm that the enhancement for the bilinear methods increases robustness. Using the optimal solution of the LP relaxation as starting point provides better objective values. The choice of the perturbation parameter γ appears to be less relevant. Plot 2d compares the best methods of the three previous plots with the MILP and FILTER solutions. The former finds the best solution in 98 cases and the latter in 74. FILTER, MILP, BL1-R*-$\gamma = 1.01$, BL2-R* and PL-E* solve 110, 103, 96, 95, 69 of the problems within 1% of the best solution, and 117, 119, 110, 110, 107 of the problems within 5%, respectively. This demonstrates the competitiveness of the DC approach in terms of solution quality.

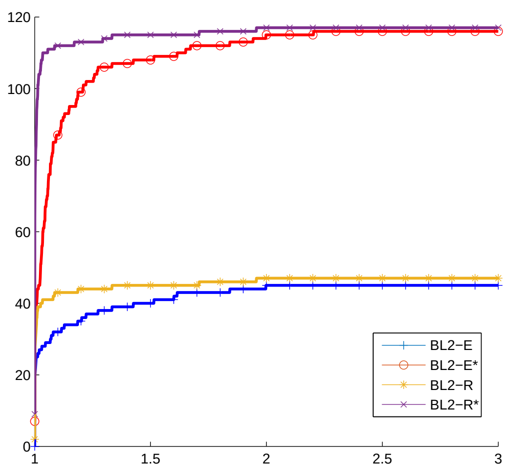
Although the MILP formulation and the FILTER solver provide slightly better solution quality, Table 4 considers a different perspective. Even for the largest case $m = 500$, the PL variants solve the instances in an average of 6.6 seconds, and the enhanced BL methods within 32 seconds. In contrast, FILTER requires on average almost 10 minutes. Even more dramatically, the CPLEX solver runs out of the wall clock time limit of 15 min in all instances using 32 cores, while the other approaches run on only a single core. This shows a very clear advantage of the DC approach, particularly since there is only a relatively small loss in solution quality.



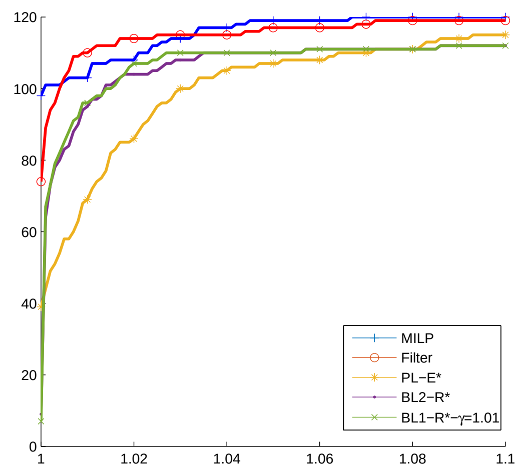
(a) PL variants



(b) BL1 variants



(c) BL2 variants



(d) Best DC variants vs MILP vs FILTER

Figure 2: Inverse QP Performance Profiles.

6.4 Large-scale inverse QP instances

Our final set of experiments considers large-scale instances of the inverse QP problem with up to 5,000 complementarity constraints. We compare the relative performance of the best methods identified in the previous section. The FILTER solver was not able to solve any instance of that size, and the MILP formulation was not attempted because already the $m = 500$ instances ran out of time on 32 cores. Table 5 shows the average computation time (in seconds) and solution quality, defined as the ratio of the objective value of a method over the best objective value, taken over the successfully solved instances (given in the "Feas" columns). Whenever BL2-R* solves an instance, it obtains the best objective value. The largest deviation from the best solution among all instances is 7%, 11%, 6%, and 7% for PL-E*, PL-R*, BL1-R*-$\gamma = 1.01$, BL1-R*-$\gamma = 2$, respectively. Again, only the PL variants solve all of the instances to the desired level of feasbility. Considering solution time, even the largest instances are solved within at most 1 hour on average. The slowest successful run was BL1-R*-$\gamma = 2$ for one $m = 5,000$ instance with 66 minutes. In contrast to the smaller instances in the previous section, the computation times for the PL variants, which require only LP solves, is no longer shorter than those for the BL methods, which require the solution of QPs. This is likely due to the different scalability of the simplex LP solver and the interior point QP solver, which are the default methods in CPLEX. Overall, variant BL2 based on the newly proposed decomposition of the bilinear penalty function is the most successful option in this test set.

Method	$m = 1,000$			$m = 2,500$			$m = 5,000$		
	Avrg time	Avrg ratio	Feas	Avrg time	Avrg ratio	Feas	Avrg time	Avrg ratio	Feas
PL-E*	23.72	1.02	5	316.88	1.02	5	3101.21	1.01	5
PL-R*	27.85	1.03	5	279.48	1.04	5	2875.57	1.02	5
BL1-R*-γ=1.01	48.37	1.01	5	384.30	1.02	5	2577.43	1.00	2
BL1-R*-γ=2	77.14	1.02	5	472.40	1.01	2	3452.60	1.00	3
BL2-R*	41.03	1.00	5	315.85	1.00	5	2156.08	1.00	4

Table 5: Computation time and solution quality for large-scale inverse QP instances.

7 Conclusions and Further Research

We examined four different DC decompositions of penalty function reformulations for LPCCs. Three of these correspond to decompositions previously proposed in [22] for solving LCPs. We established relationships between the stationary points and local minima of the LPCCs and those of the penalty problems. Numerical experiments demonstrate that the proposed improvement of the piecewise linear penalty variation can significantly improve the solution quality by escaping non-strongly stationary points. Similarly, the enhancement of the bilinear penalty variant is shown to reduce the number of iterations significantly. Overall, the DC algorithm and its enhancements are competitive in terms of objective function values compared to state-of-the-art solvers, and are the only option for the large-scale test problems examined here. With these encouraging results of the DC approach for solving LPCCs, further research is warranted to study the application of this approach to MPCCs with DC objectives and linear complementarity constraints.

References

[1] L. Bai, J.E. Mitchell, and J.S. Pang. On convex quadratic programs with linear complementarity constraints. *Computational Optimization and Applications* 54(3) 517–544 (2013). 2

[2] O. Burdakov, Ch. Kanzow, and A. Schwartz. Mathematical programs with cardinality constraints: Reformulation by complementarity-type constraints and a regularization method. Preprint 324, Institute of Mathematics, University of Würzburg (Germany, July 2014). http://www.optimization-online.org/DB_FILE/2014/07/4450.pdf (last revised February 2015). 1

[3] R.H. Byrd, J. Nocedal, and R.A. Waltz. KNITRO: An integrated package for nonlinear optimization. In G. di Pillo and M. Roma, editors. *Large-Scale Nonlinear Optimization* pp. 35-59, Springer Verlag (New York 2006). 2

[4] R.W. Cottle, J.S. Pang, and R.E. Stone. *The Linear Complementarity Problem*. SIAM Classics in Applied Mathematics 60 (Philadelphia 2009). [Originally published by Academic Press, Boston (1992)] 2

[5] H.R. Fang, S. Leyffer, and T.S. Munson. A pivoting algorithm for linear programs with complementarity constraints. *Optimization Methods and Software* 27(1) 89–114 (2012). 2, 5, 13

[6] M. Feng, J.E. Mitchell, J.S. Pang, A. Wächter, and X. Shen. Complementarity formulations of ℓ_0-norm optimization problems. http://www.optimization-online.org/DB_FILE/2013/09/4053.pdf (last revised January 2015). 1

[7] R. Fletcher and S. Leyffer. Solving mathematical program with complementarity constraints as nonlinear programs. *Optimization Methods and Software* 19(1) 15-40 (2004). 2, 19

[8] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming* 91(2) 239–270 (2002). 2, 19

[9] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes. Local convergence of SQP methods for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization* 17(1) 259-286 (2006). 2, 5

[10] R. Fletcher, S. Leyffer, and Ph.L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM Journal on Optimization* 13(1) 44-59 (2002). 2

[11] R. Fourer, D.M. Gay, and B.W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. Boyd & Fraser (2003). 19

[12] J. Hu, J.E. Mitchell, and J.S. Pang. An LPCC approach to nonconvex quadratic programs. *Mathematical Programming* 133(1) 243–277 (2012). 2

[13] J. Hu, J.E. Mitchell, J.S. Pang, and B. Yu. On the global solution of linear programs with linear complementarity constraints. *SIAM Journal on Optimization* 19(1) 445–471 (2008). 2, 4

[14] J. Hu, J.E. Mitchell, J.S. Pang, and B. Yu. On linear programs with linear complementarity constraints. *Journal of Global Optimization* 53(1) 29–51 (2012). 1, 24

[15] T. Ibaraki. Complementary programming. *Operations Research* 19(6) 1523–1529 (1971). 1

[16] T. Ibaraki. The use of cuts in complementary programming. *Operations Research* 21(1) 353–359 (1973). 1

[17] R.G. Jeroslow. Cutting planes for complementarity constraints. *SIAM Journal on Control and Optimization* 16(1) 56–62 (1978). 1

[18] J.J. Judice. Algorithms for linear programming with linear complementarity constraints. *TOP* 20(1) 4–25 (2012). 2

[19] H.A. Le Thi, V.N. Huynh, and T. Pham Dinh. DC programming and DCA for general DC programs. *Advances in Intelligent Systems and Computing*, pp. 15–35, (Springer 2014). 2, 5

[20] H.A. Le Thi and T. Pham Dinh. Recent advances in DC programming and DCA. *Transactions on Computational Collective Intelligence* 8342 1–37 (2014). 2

[21] H.A. Le Thi and T. Pham Dinh. The state of the art in DC programming and DCA. Research Report (60 pages), Lorraine University (2013). 2, 4, 5

[22] H.A. Le Thi and T. Pham Dinh. On solving linear complemetarity problems by DC programming and DCA. *Computational Optimization and Applications* 50(3) 507–524 (2011). 2, 5, 6, 14, 20, 21, 22, 28

[23] H.A. Le Thi and T. Pham Dinh. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research* 133 23-46 (2005). 2, 4

[24] S. Leyffer and T.S. Munson. A globally convergent Filter method for MPECs. Preprint ANL/MCSP1457-0907, Argonne National Laboratory, Mathematics and Computer Science Division (revised April 2009). 2

[25] S. Leyffer, G. Lopez-Calva, and J. Nocedal. Interior point methods for mathematical programs with complementarity constraints. *SIAM Journal on Optimization* 17(1) 52–77 (2006). 2, 19

[26] S. Leyffer. MacMPEC: AMPL collection of MPECs (2000). Available at http://www.mcs.anl.gov/~leyffer/MacMPEC/. 20, 22

[27] Z.Q. Luo, J.S. Pang, and D. Ralph. *Mathematical Programs With Equilibrium Constraints*. Cambridge University Press (Cambridge, England 1996). 1, 4

[28] O.L. Mangasarian. Solution of general linear complementarity problems via nondifferentiable concave minimization. *Acta Mathematica Vietnamica* 22(1) 199–205 (1997). 6

[29] L.D. Muu, Q.T. Dinh, H.A. Le Thi, and T. Pham Dinh. A new decomposition algorithm for globally solving mathematical programs with affine equilibrium constraints. *Acta Mathematica Vietnamica* 37(2) 201–218 (2012). 2, 4

[30] J.S. Pang, M. Razaviyayn, and A. Alvarado. Computing B-stationary points of nonsmooth DC programs. *Mathematics of Operations Research*, acceptance subject to minor revision (January 2015; original submitted October 2014). 5

[31] T. Pham Dinh and H.A. Le Thi. Convex analysis approach to DC programming: Theory, algorithm and applications. *Acta Mathematica Vietnamica* 22(1) 289–355 (1997). 2, 4

[32] R.T. Rockafeelar. *Convex Analysis*. Princeton University Press (Princeton 1970). 5

[33] H. Scheel and S. Scholtes. Mathematical program with complementarity constraints: Stationarity, optimality and sensitivity. *Mathematics of Operations Research* 25(1) 1–22 (2000). 2, 4, 13

[34] B. Yu. *A Branch and Cut Approach to Linear Programs with Linear Complementarity Constraints*. Ph.D. thesis. Department of Decision Sciences and Engineering Systems, Rensselaer Polytechic Institute (August 2011). 2, 4

[35] B. Yu, J.E. Mitchell, and J.S. Pang. Obtaining tighter relaxations of mathematical programs with complementarity constraints. In T. Terlaky and F. Curtis, editors. *Modeling and Optimization: Theory and Applications*. Springer Proceedings in Mathematics and Statistics (New York 2012) Volume 21, pp. 1–23. 14