A new approach for solving mixed integer DC programs using a continuous relaxation with no integrality gap and smoothing techniques

Takayuki Okuno Yoshiko T. Ikebe

Abstract

In this paper, we consider a class of mixed integer programming problems (MIPs) whose objective functions are DC functions, that is, functions representable in terms of the difference of two convex functions. These MIPs contain a very wide class of computationally difficult nonconvex MIPs since the DC functions have powerful expressability. Recently, Maehara, Marumo, and Murota provided a continuous reformulation without integrality gaps for discrete DC programs having only integral variables. They also presented a new algorithm to solve the reformulated problem. Our aim is to extend their results to MIPs and give two specific algorithms to solve them. First, we propose an algorithm based on DCA originally proposed by Pham Dinh and Le Thi, where convex MIPs are solved iteratively. Next, to handle nonsmooth functions efficiently, we incorporate a smoothing technique into the first method. We show that sequences generated by the two methods converge to stationary points under some mild assumptions.

Key words: mixed integer DC program, integrality gap, closed convex extension, smoothing method

1 Introduction

Let us consider the following optimization problem:

$$\min f(x) \quad \text{sub.to} \quad x = (x_M, x_N) \in S, \quad x_M \in \mathbb{R}^M, \quad x_N \in \mathbb{Z}^N.$$ 

Here $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a closed proper function, i.e., $f$ is lower semicontinuous and its effective domain $\text{dom } f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$ is not empty. Moreover, $S \subseteq \mathbb{R}^n$ is a nonempty closed convex set, and $M$ and $N$ are disjoint sets of indices such that $M \cup N = \{1, 2, \ldots, n\}$, and $x_M = (x_i)_{i \in M}$, $x_N = (x_i)_{i \in N}$.

*Department of Information and Computer Technology Tokyo University of Science 6-3-1 Niijuku, Katsushika-ku, Tokyo 125-8585, Japan (\{t.okuno,yoshiko\}@rs.tus.ac.jp)
In the case where $f$ is a linear or convex quadratic function and $S$ is represented with only linear or convex quadratic inequalities, the branch-and-bound method and cutting plane techniques work nicely in the practical sense. Indeed, there are many commercial and free solvers implementing them, e.g. CPLEX [18], gurobi [15] and SCIP [1]. On the other hand, for the general nonlinear case, the above problem is extremely difficult to solve. There are a number of ways to approach such mixed-integer nonlinear problems. One method is to extend the framework of branch-and-bound to the continuous spaces [12, 27, 28]. Another is to utilize sequential quadratic programming (SQP) [11, 19, 9, 10]. These algorithms incorporate such techniques as trust regions, outer approximations and branch-and-bound techniques to solve quadratic problems approximating the original one. In particular, in applying these SQP-type algorithms to mixed-integer convex problems with continuously differentiable convex functions, global convergence to an optimum can be proved. There are also algorithms which deal solely with mixed-integer nonlinear programs with convex $f$ [13, 14, 8, 30, 2]. See, for example, the surveys [3] and [5].

In this paper, we consider the case where $f$ is a so-called DC function, that is, a function representable as the difference of two convex functions:

$$\min_{x \in S, \ x_M \in \mathbb{R}^M, \ x_N \in \mathbb{Z}^N} f(x) = g(x) - h(x)$$

(1)

where $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions. Hereafter, we suppose $\infty - \infty = \infty$ for convention, whereby $\emptyset \neq \text{dom} \ g \subseteq \text{dom} \ h$ and $\text{dom} \ g = \text{dom} \ f$ naturally hold.

The class of DC functions covers a very wide range of functions. For example, any twice continuously differentiable function is DC, moreover, functions generated by applying operators such as $\sum$, $\Pi$, $| \cdot |$, and $\max(\cdot, \cdot)$ to DC functions also belong to the class DC [16, 17]. Hence, the problem of our focus, (1) covers a wide class of mixed integer programs. Note however, that given a DC function $f$, finding two explicit convex functions $g$ and $h$ representing $f$ is a hard open problem. Among the functions for which a DC representation is easily found, perhaps the most common are the quadratic functions. In this paper, we assume that one DC representation is explicitly given; how we obtain it will not enter our discussion.

The DC programming in continuous variables is an important field of research in continuous optimization, and theoretical and practical aspects have been extensively studied [25, 26]. For example, the global optimality condition is completely characterized by the Toland-Singer duality theorem. This duality theorem in turn forms the basis for the fundamental DC programming algorithm known as DCA [25], which is known to have nice convergence properties.

DC programming also has many useful applications. One example is in mixed-integer linear programs, where integer constraints on variables are incorporated into the objective functions via penalty functions [24]. Other notable results have been reported in sparse optimization [29, 21] and portfolio selection [20]. This is an active field, with remarkable recent progress in both theory and applications.
On the other hand, discrete DC programming, which concerns DC programs with integrally constrained variables, that is, \( \Pi \) with \( M = \emptyset \) and \( N = \{1, 2, \ldots, n\} \), is still a relatively unexplored area. Recently, a promising approach was proposed by Maehara and Murota [22], who showed how the framework of discrete convex analysis can be applied, to export results in continuous DC theory to a discrete setting. This was further pursued in Maehara, Marumo and Murota [23], who proved a powerful result in constructing continuous relaxations of discrete DC programs. The simplest continuous relaxation for \( \Pi \) may just replace \( \mathbb{Z}^N \) by \( \mathbb{R}^N \). As is well known, this does not work effectively in general, since an integrality gap usually occurs, that is, the optimal values of the original and relaxed problems do not coincide. On the other hand, the new continuous relaxation proposed in [23] replaces \( g \) with its closed convex closure (and \( h \) with an arbitrary relaxation). Its notable property is that no integrality gap is generated.

In this paper we extend the theorem of Maehara, Marumo and Murota to mixed integer DC programs of the form \( \Pi \), and propose two algorithms to solve them. Our first algorithm, a generic scheme based on the DCA originally proposed by Pham Dinh and Le Thi [25], iteratively solves a sequence of convex mixed integer programs. Our second algorithm, designed to deal with nonsmooth functions, is obtained by incorporating smoothing techniques [7] into the first. The sequences generated by both two methods converge to stationary points under some mild assumptions.

This paper is organized as follows. In Section 2 we briefly describe existing results in continuous and discrete DC programming, and in Section 3, we show how to extend the theorem of Maehara, Marumo and Murota to obtain a continuous relaxation of \( \Pi \) with no integrality gap. Next, in Section 4 we describe our basic first algorithm, and in Section 5 we give our second algorithm with smoothing techniques along with results in convergence properties.

Throughout the paper, we will use the following notations: For any \( x \in \mathbb{R}^n \), \( \|x\| \) represents the Euclidean 2-norm of \( x \). For any nonempty set \( X \subseteq \mathbb{R}^n \), we denote the convex hull and closure of \( X \) by \( co X \) and \( cl X \), respectively. Also, we denote the interior and relative interior of \( X \) by \( int X \) and \( ri X \), respectively. Let \( \varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) be a convex function. For \( x \in \text{dom} \varphi \), the subdifferential of \( \varphi \) at \( x \), that is, the set of all subgradients of \( \varphi \) at \( x \), is denoted by \( \partial \varphi(x) \). We write the conjugate of \( \varphi \) as \( \varphi^* \), that is, the function \( \varphi^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) defined by \( \varphi^*(y) = \sup_{x \in \mathbb{R}^n} \{ \langle y, x \rangle - \varphi(x) \} \) where \( \langle y, x \rangle \) stands for the canonical inner product of \( y \) and \( x \), i.e., \( \langle y, x \rangle = y^\top x \). For \( \psi : \mathbb{Z}^n \) (resp., \( \mathbb{R}^n \)) \rightarrow \mathbb{R} \cup \{+\infty\} \) the epigraph of \( \psi \) is the set \( \text{epi}_{\mathbb{Z}^n} \psi \) (resp., \( \text{epi}_{\mathbb{R}^n} \psi \)) := \{ (x, x_{n+1}) \mid x_{n+1} \geq \psi(x), x \in \mathbb{Z}^n \) (resp., \( \mathbb{R}^n \)) \} \subseteq \mathbb{R}^{n+1} \). Finally, \( \mathbb{R}^n_{++} \) is the non-negative (positive) orthant in \( \mathbb{R}^n \).
2 A brief review of continuous and discrete DC programings

We begin by considering (1) with \( S = \mathbb{R}^n \) and \( N = \emptyset \), more specifically,

\[
\min_{x \in \mathbb{R}^n} \{g(x) - h(x)\}.
\]

Then, the following proposition holds.

**Proposition 2.1** ([25]). Suppose that the DC program (2) has an optimal solution \( x^* \). Then, we have

1. \( \partial g(x^*) \supseteq \partial h(x^*) \),
2. \( \bar{y} \in \partial h(x^*) \iff x^* \in \partial h^*(\bar{y}) \), and
3. \( \bar{y} \in \partial h(x^*) \Rightarrow \bar{y} \) is an optimal solution of \( \inf_{y \in \mathbb{R}^n} \{h^*(y) - g^*(y)\} \).

The following theorem is known as Toland-Singer duality, and forms the basis for DC minimization algorithms.

**Theorem 2.2.** *(Toland-Singer duality)*

\[
\inf_{x \in \mathbb{R}^n} \{g(x) - h(x)\} = \inf_{y \in \mathbb{R}^n} \{h^*(y) - g^*(y)\}
\]

We next define stationary points for DC programs that contain global optima.

**Definition 2.3.** A stationary point for \( g - h \) is a point \( x^* \) such that

\[
\partial g(x^*) \cap \partial h(x^*) \neq \emptyset.
\]

Let us introduce an existing algorithm for solving the DC program which will become the base of our proposed algorithms, and cite its convergence results. For details we refer the reader to [25].

**Simplified DC Algorithm (DCA)**

**Step 0:** Choose \( x^0 \in \mathbb{R}^n \). Set \( k = 0 \)

**Step 1:** Choose \( y^k \in \partial h(x^k) \) and \( x^{k+1} \in \partial g^*(y^k) \)

**Step 2:** If stopping criterion is satisfied stop,

else set \( k = k + 1 \) and go to Step 1

**Theorem 2.4** ([25]). Let \( \{x^k\} \) and \( \{y^k\} \) be the sequences generated by the simplified DCA. Then, the following statements hold.

1. \( g(x^{k+1}) - h(x^{k+1}) \leq g(x^k) - h(x^k) \).
2. \( h^*(y^{k+1}) - g^*(y^{k+1}) \leq h^*(y^k) - g^*(y^k) \).

3. Every accumulation point \( x^* (y^*) \) of the sequence \( \{x^k\} (\{y^k\}) \) is a stationary point of \( g - h (h^* - g^*) \).

We now turn to DC programs with discrete variables. Before introducing the results of Maehara, Marumo and Murota, we define some concepts related to discrete functions. Consider a function on discrete variables, \( \varphi : \mathbb{Z}^n \rightarrow \mathbb{R} \cup \{+\infty\} \).

**Definition 2.5.** A convex function \( \hat{\varphi} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) is a convex extension of \( \varphi \) if

\[
\hat{\varphi}(x) = \varphi(x) \quad (x \in \mathbb{Z}^n).
\]

The convex closure of \( \varphi \) is the function \( \varphi^{cl} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) whose epigraph is equal to the closed convex hull of the epigraph of \( \varphi \), i.e.,

\[
\text{epi}_{\mathbb{R}^n} \varphi^{cl} = \text{cl co epi}_{\mathbb{Z}^n} \varphi.
\]

While the convex closure can be defined for any \( \varphi \), clearly, not all discrete functions have convex extensions. If the discrete function \( \varphi \) does have a convex extension \( \hat{\varphi} \), then we always have

\[
\varphi^{cl}(x) = \hat{\varphi}(x) \quad (x \in \mathbb{Z}^n).
\]

As, in this paper, we will be concerned only with discrete functions which are the restrictions of continuous convex functions on \( \mathbb{R}^n \) to \( \mathbb{Z}^n \), all discrete functions will trivially have convex extensions.

Let us consider the DC program (1) with \( S = \mathbb{R}^n \) in which all variables are restricted to integer values, i.e., \( M = \emptyset \) and \( N = \{1, 2, \ldots, n\} \):

\[
\min_{x \in \mathbb{Z}^n} \{g(x) - h(x)\}.
\]  

(3)

If we define the discrete functions \( g_Z, h_Z : \mathbb{Z}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) as the restrictions of \( g \) and \( h \) to \( \mathbb{Z}^n \):

\[
g_Z(x) = g(x), \quad h_Z(x) = h(x) \quad (x \in \mathbb{Z}^n)
\]

(4)

and let \( \hat{g}, \hat{h} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) be any convex extensions of \( g_Z \) and \( h_Z \), then the following continuous DC program is clearly a relaxation of (3)

\[
\min_{x \in \mathbb{R}^n} \{\hat{g}(x) - \hat{h}(x)\}.
\]  

(5)

The original functions \( g \) and \( h \) are obvious candidates for the convex extensions \( \hat{g} \) and \( \hat{h} \), but this is usually a poor choice as the two optimal values of (3) and (5) generally do not coincide. Maehara, Marumo and Murota [23] proved that the appropriate choice of \( \hat{g} \) ensures this will not happen.

**Theorem 2.6 ([23]).** If \( \hat{g} \) is the convex closure of \( g_Z \), then the optimal values of the two problems (3) and (5) coincide.

We now turn to our main concern, mixed integer DC programs.
3 Continuous relaxation with no integrality gap

We begin by rephrasing problem (1). By using the indicator function of set $S$, that is, the function $\delta_S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$
\delta_S(x) = \begin{cases} 
0 & (x \in S) \\
+\infty & (x \notin S)
\end{cases},
$$

(1) can be written as

$$
\min_{x : x = (x_M, x_N) \in \mathbb{R}^M \times \mathbb{Z}^N} \left( \delta_S(x) + g(x) \right) - h(x).
$$

(6)

Since $S$ is a closed convex set, $\delta_S$, and hence $\delta_S + g$ are closed proper convex functions. Now define $\tilde{g}$ and $\tilde{h}$ respectively as the restrictions of $\delta_S + g$ and $h$ to $\mathbb{R}^M \times \mathbb{Z}^N$, that is,

$$
\tilde{g} := (\delta_S + g)|_{\mathbb{R}^M \times \mathbb{Z}^N} \quad \text{and} \quad \tilde{h} := h|_{\mathbb{R}^M \times \mathbb{Z}^N}.
$$

We also denote the convex closure of $\tilde{g}$ by $\tilde{g}^c$. Convex extensions, epigraphs, and convex closures of $\tilde{g}$ and $\tilde{h}$ are defined in a manner analogous to the discrete functions in the last paragraph of Section 1 and Definition 2.5; for example, the epigraph of $\tilde{g}$ is defined as the set $\{(x_M, x_N, x_{n+1}) \in \mathbb{R}^M \times \mathbb{Z}^N \times \mathbb{R} \mid x_{n+1} \geq g(x_M, x_N)\} =: \text{epi}_{\mathbb{R}^M \times \mathbb{Z}^N} \tilde{g}$.

In the rest of this section, we extend the theorem of Maehara, Marumo and Murota, to mixed-integer DC programs (1), that is, DC programs involving both integer-valued and continuous variables. More precisely, we prove the following theorem.

**Theorem 3.1.** Let $\tilde{h} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be an arbitrary convex extension of $h$. Then, the following (continuous) DC program:

$$
\min_{x \in \mathbb{R}^n} \left\{ \tilde{g}^c(x) - \tilde{h}(x) \right\}
$$

(7)

has the same optimal value as the mixed-integer DC program (6), i.e., as (1). In particular, the optimal set of (6) is contained in that of (7).

**Proof.** If (1) is not bounded from below, then neither is (7) since (7) is a relaxation of (1). Hence, we only consider the case where (1), i.e., (6) has a finite optimal value. Let $\tilde{v}^*$ and $v^*$ respectively be the optimal values of problems (6) and (7). We will prove that $\tilde{v}^* = v^*$. Since (7) is a relaxation of (1), it suffices to show $\tilde{v}^* \leq v^*$. By optimality of $\tilde{v}^*$ for (6),

$$
\tilde{g}(x_M, x_N) \geq \tilde{h}(x_M, x_N) + \tilde{v}^* \quad (x_M, x_N) \in \mathbb{R}^M \times \mathbb{Z}^N
$$

implying that

$$
\emptyset \neq \text{epi}_{\mathbb{R}^M \times \mathbb{Z}^N} \tilde{g} \subseteq \text{epi}_{\mathbb{R}^M \times \mathbb{Z}^N} \tilde{h} + (0, \tilde{v}^*) \subseteq \text{epi}_{\mathbb{R}^n} \hat{h} + (0, v^*),
$$

(8)
where the leftmost inequality is from the assumption that \( \text{dom } g \neq \emptyset \). Notice that from (8) and \( \text{epi}_{\mathbb{R}^M \times \mathbb{Z}^N} \tilde{g} \subseteq \text{epi}_{\mathbb{R}^n} \tilde{g}^{cl} \), we obtain \( D := \text{epi}_{\mathbb{R}^n} \tilde{g}^{cl} \cap \left( \text{epi}_{\mathbb{R}^n} \tilde{h} + (0, \tilde{v}^*) \right) \neq \emptyset \). Now, we show that

\[
\text{epi}_{\mathbb{R}^n} \tilde{g}^{cl} \subseteq D \supseteq \text{epi}_{\mathbb{R}^M \times \mathbb{Z}^N} \tilde{g}.
\] (9)

Moreover, there must be some closed proper convex function \( \phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) satisfying \( \text{epi}_{\mathbb{R}^n} \phi = D \). Obviously \( \phi \) is a closed convex extension of \( \tilde{g} \) from \( \mathbb{R}^M \times \mathbb{Z}^N \) to \( \mathbb{R}^n \). However, this contradicts the fact that \( \tilde{g}^{cl} \) is the convex closure of \( \tilde{g} \), because \( \text{epi} \phi \) is a proper subset of \( \text{epi} \tilde{g}^{cl} \) by (10). Therefore, we have (9) which further yields \( \tilde{v}^* \leq v^* \). This completes the proof.

By the above result, it is justified to solve (7) instead of (1). In the remainder of the paper, we propose a specific algorithm for solving (7). In our algorithm, we choose \( h \) as \( \hat{h} \), a convex extension of \( \tilde{h} \). Therefore, our target is to solve the following problem

\[
\min_{x \in \mathbb{R}^n} \{ \tilde{g}^{cl}(x) - h(x) \}.
\] (11)

4 A basic algorithm for the mixed integer DC program

In this section, we formulate a basic algorithm based on the DCA of Section 2 for solving (11). For the DC program (11), recall that the DCA involves finding \( x_k \) and \( y_k \) with

\[
y^k \in \partial h(x^k), \text{ and } x^{k+1} \in \partial (\tilde{g}^{cl})^*(y^k).
\]

Finding \( x^{k+1} \in \partial (\tilde{g}^{cl})^*(y^k) \) can be accomplished by using the following relations:

\[
x^{k+1} \in \partial (\tilde{g}^{cl})^*(y^k) \iff \partial \tilde{g}^{cl}(x^{k+1}) \ni y^k
\]

\[
\iff x^{k+1} \text{ is a solution of } \inf_{w \in \mathbb{R}^n} (\tilde{g}^{cl}(w) - \langle y^k, w \rangle)
\]

The rightmost optimization problem involves minimizing a convex function. However, this cannot be solved by using standard convex optimization methodologies such as the interior point method, since we do not have an explicit expression of \( \tilde{g}^{cl} \) in general. This can be overcome by using Theorem 3.1 to note that it corresponds to solving the following convex mixed integer program:

\[
\min g(x) - \langle y^k, x \rangle
\]

\[
\text{sub.to } x \in S, \ x = (x_M, x_N) \in \mathbb{R}^M \times \mathbb{Z}^N.
\] (12)

Hence, by replacing \( x^{k+1} \in \partial (\tilde{g}^{cl})^*(y^k) \) with (12) in Step 2 of the simplified DCA, we gain a specific algorithm for solving (11) as below:
Sequential convex mixed-integer programming method (SCMIP)

Step 0: Choose \(x^0 \in \mathbb{R}^n\). Set \(k = 0\).

Step 1: Choose \(y^k \in \partial h(x^k)\) and solve (12) to obtain \(x^{k+1}\).

Step 2: If stopping criterion is satisfied, stop, else set \(k = k + 1\) and go to step 1.

Obviously, each iteration point \(x^k\) is feasible to (1). By applying existing results on convergence for the DCA \([25, 26]\), we can make some observations for the case that at least one of \(\tilde{g}^{cl}\) and \(h\) is a strongly convex function.

- Both \(\tilde{g}^{cl}(x^k) - h(x^k) = f(x^k)\) and \(h^*(y^k) - (\tilde{g}^{cl})^*(y^k)\) strictly decrease.
- If \(x^*(\text{resp., } y^*)\) is an accumulation point of \(\{x^k\}\) (resp., \(\{y^k\}\)), then \(x^*(\text{resp., } y^*)\) is a stationary point of \(\min \tilde{g}^{cl}(x) - h(x)\) (resp., \(\min h^*(y) - (\tilde{g}^{cl})^*(y)\)). That is to say, \(y^* \in \partial (\tilde{g}^{cl})(x^*) \cap \partial h(x^*)\) and \(x^* \in \partial (\tilde{g}^{cl})^*(y^*) \cap \partial h^*(y^*)\) hold.
- The \(x^N\)-part of \(x^k\) converges to some integer point within finitely many iterations.\(^1\)

In the above discussion, the assumption that at least one of \(\tilde{g}^{cl}\) and \(h\) is strongly convex is crucial. We place emphasis on the “at least one” phrase.

In problem (1), we did not assume strong convexity of either \(g\) or \(h\). Thus, at first glance, the above results may seem inapplicable, however, it can be easily overcome by considering the following equivalent problem for fixed \(\rho > 0\):

\[
\min \left( g(x) + \rho \frac{\|x\|^2}{2} \right) - \left( h(x) + \rho \frac{\|x\|^2}{2} \right) \text{ sub.to } x \in S, \quad (x_M, x_N) \in \mathbb{R}^M \times \mathbb{Z}^N. \quad (13)
\]

We note here that the convex closure of \(g(\cdot) + \rho \|\cdot\|^2/2\) is usually not strongly convex, whereas \(h(\cdot) + \rho \|\cdot\|^2/2\) always is. Thus it is important that we do not need the strong convexity of both \(g\) and \(h\).

Before ending this section, we make an important remark concerning the drawbacks of transforming (11) to (13). Consider two different DC-decompositions \((g_1, h_1)\) and \((g_2, h_2)\) for \(f\), i.e., \(f = g_1 - h_1 = g_2 - h_2\), and corresponding continuous DC programs of the form (11). Their two optimal sets are exactly the same. However, their sets of stationary points may possibly differ. This phenomenon does not occur in continuous DC programs without discrete variables, and thus it is characteristic of (11). To illustrate it, let us consider the following trivial mixed integer program:

**Example 1.**

\[
\min x \text{ sub.to } x \in \{-1, 0, 1\} \quad (14)
\]

\(^1\)This property is specific to the case where the DCA is applied to (11). For the proof, see the convergence analysis of the next algorithm.
Choose two DC-decompositions \((g_1, h_1) = (x, 0)\) and \((g_2, h_2) = (x^2 + x, x^2)\). Then, \(g_1^c(x) = x\), \(g_2^c(x) = x^2 + x\) is the polygonal line connecting the three points \((-1, 0), (0, 0)\) and \((1, 2)\), and \(\text{dom } g_1^c = \text{dom } g_2^c = [-1, 1]\). Thus the resulting optimization problems of the form (1) are:

\[
\min \begin{cases} x & (x < -1) \\ -x^2 & (-1 \leq x \leq 0) \\ 2x - x^2 & (0 \leq x \leq 1) \\ \infty & (1 < x) \end{cases} \quad \text{and} \quad \min \begin{cases} x & (x < -1) \\ 0 & (-1 \leq x \leq 0) \\ 1 - x & (0 \leq x \leq 1) \end{cases}
\]

The set of stationary points of the former problem is nothing but the optimal set \([-1]\) of (14), while that of the latter is \(\{0, -1\}\). This example indicates that the choice of DC decomposition may affect efficiency in finding the optima.

## 5 SCMIP with smoothing techniques

In this section, we focus on a particular case of (1) where either one of \(g\) and \(h\) is not differentiable on the effective domain. Such problems often occur in applications. For example, consider problem (1) equipped with a dc inequality constraint \(g_1(x) - g_2(x) \leq 0\) with \(g_1, g_2 : \mathbb{R}^n \to \mathbb{R}\) being convex functions. If \(g_1 - g_2\) is not convex, it is difficult to directly apply the SCMIP. One remedy for this is to lift the constraint \(g_1(x) - g_2(x) \leq 0\) into the objective function as a penalty term \(\tau \max(g_1 - g_2, 0)\) with \(\tau > 0\) being a penalty parameter, then further decomposing it as follows:

\[
\min (g(x) + \tau \max(g_1(x), g_2(x))) - (h(x) + \tau g_2(x))
\]

sub.to \(x \in S, \ x = (x_M, x_N) \in \mathbb{R}^M \times \mathbb{Z}^N\).

This problem has the form (1), since \(g + \tau \max(g_1, g_2)\) and \(h + \tau g_2\) are both convex. Obviously, \(g + \tau \max(g_1, g_2)\) is not differentiable in general, due to the existence of the max function. Theoretically, the SCMIP is applicable regardless of the differentiability of \(g\) and \(h\). However, in this case we must iteratively solve mixed-integer nonsmooth optimization problems. Practically, nonsmooth problems are not as tractable as smooth ones, even if they possess only continuous variables. Moreover, most available free or commercial solvers cannot deal with them. Thus, we employ the smoothing method, which is one of the most powerful techniques for solving optimization problems or nonlinear equations involving nondifferentiable functions. This method solves a sequence of approximated problems in which given nonsmooth functions are replaced by so-called smoothing functions. For a comprehensive survey on the smoothing method, refer to [7] and references therein. In this paper, the smoothing functions are defined as below.

**Definition 5.1.** Let \(\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) be a function such that \(\text{int dom } \phi \neq \emptyset\) and \(\phi\) is continuous on \(\text{dom } \phi\). We say that \(\Phi : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}\) is a smoothing function of \(\phi\) when...
Various kinds of smoothing functions for specialized problems have been studied extensively. For example, consider the plus function \((\cdot)_+ := \max(\cdot, 0)\). Many nonsmooth functions such as \(|x|, \max(x, y), \text{ and } \min(x, y)\) can be explicitly represented with the plus function, thus it is very versatile. One way of approximating the plus function is by using a piecewise continuous function \(p : \bbR \to \bbR_+\) such that 
\[
\int_{-\infty}^{+\infty} p(s) ds = 1, \quad p(s) = p(-s), \quad \text{and} \quad \int_{-\infty}^{+\infty} |s| p(s) ds < \infty.
\]
For such \(p\), it is well-known that \(\Phi(t, \mu) := \int_{-\infty}^{+\infty} \{(t - \mu s)_+ p(s)\} ds\) becomes a smoothing function for \((\cdot)_+\). This class of smoothing function is called the Chen-Mangasarian function \([6]\) and admits the following properties:

\((P1)\) For any \(\mu \geq 0\), \(\Phi(\cdot, \mu)\) is convex.

\((P2)\) For any \(x \in \text{dom } \phi\), \(\Phi(x, \cdot)\) is a nondecreasing function on \(\bbR_+\) and furthermore there exists some \(\kappa > 0\) such that 
\[
0 \leq \Phi(x, \mu_2) - \Phi(x, \mu_1) \leq \kappa(\mu_2 - \mu_1)
\]
holds for any \(x \in \bbR^n\) and \(0 < \mu_1 \leq \mu_2\).

Properties \((P1)\) and \((P2)\) often play a crucial role in establishing convergence analysis for the smoothing method. They are satisfied by many existing smoothing functions for \((x)_+, \max(x, y), \min(x, y)\) and \(|x|\). In the rest of this section, we will establish convergence properties under \((P1)\) and \((P2)\).

Now, let us turn back to (1) and let \(G, H : \bbR^n \times \bbR_+ \to \bbR\) respectively be smoothing functions for \(g\) and \(h\). Hereafter, to simplify notations, for any \(u \in \bbR^n_+\) and \(\{u^k\}_{k \geq 0} \subseteq \bbR^n_+\), we often write 
\[
u = (\mu_1, \mu_2), \quad u^k = (\mu_1^k, \mu_2^k).
\]
In addition, we denote
\[
G_\mu(\cdot) = G(\cdot, \mu), \quad H_\mu(\cdot) = H(\cdot, \mu).
\]
Following the terminology of Section \(\S\) let
\[
\tilde{G}_\mu = (G_\mu + \delta_S)|_{\bbR^n \times \bbZ^N},
\]
\(\tilde{G}_{\mu}^\text{cl}\) be the convex closure of \(\tilde{G}_\mu\), and then define \(f_u, \tilde{f}_u^\text{cl} : \bbR^n \to \bbR \cup \{+\infty\}\) by
\[
f_u := G_{\mu_1} - H_{\mu_2}, \quad \tilde{f}_u^\text{cl} := \tilde{G}_{\mu_1}^\text{cl} - H_{\mu_2},
\]
respectively.
We now propose a new algorithm incorporating the smoothing method into the SCMIP presented in the previous section. In our method, instead of solving (12), we execute the SCMIP solving the following mixed integer convex optimization problem with a smoothing parameter 
\[ u_k^2 \in \mathbb{R}_2^+ \] such that \( \lim_{k \to \infty} u_k = 0 \):
\[
\begin{align*}
\min & \quad G_{\mu_1^1}(x) - \langle y_k, x - x_k \rangle \\
\text{sub.to} & \quad x \in S, \ x = (x_M, x_N) \in \mathbb{R}^M \times \mathbb{Z}^N,
\end{align*}
\]
where \( y_k \in \partial H_{\mu_2^2}(x_k) \). This is equivalent to solving the continuous convex program:
\[
\min \tilde{G}_{\mu_1^1}(x) - \langle y_k, x - x_k \rangle.
\]
(16)

The overall framework of the smoothing SCMIP is described as follows:

**Smoothing SCMIP**

**Step 0:** Choose \( x^0 \in \mathbb{R}^n \), \( u^0 = (\mu_1^0, \mu_2^0) \in \mathbb{R}_2^+ \), and \( \gamma \in (0, 1) \). Set \( k = 0 \)

**Step 1:** Set \( y^k \in \partial H_{\mu_2^2}(x^k) \) (\( y^k = \nabla H_{\mu_2^2}(x^k) \) if \( \mu_2^k > 0 \)) and solve (15) to obtain \( x^{k+1} \).

**Step 2:** If stopping criterion is satisfied stop,
else set \( u^{k+1} = \gamma u^k \), and \( k = k + 1 \) and go to Step 1

**Remark 5.2.** Note that in Step 0 we do not restrict the scope of \( \{u^k\} \) to the positive orthant \( \mathbb{R}_2^+ \) for general versatility. If we set \( u^k = 0 \) for all \( k \geq 0 \), the above algorithm is nothing but the SCMIP without the smoothing technique. Also, if we set \( \mu_2^k = 0 \) for all \( k \geq 0 \), the smoothing is applied only to the function \( g \), leaving \( h \) as it is.

**Convergence analysis for the smoothing SCMIP**

In this section, we show that generated sequences \( \{x^k\} \) and \( \{y^k\} \) respectively converge to stationary points of the DC problem (11) and its Toland-Singer dual in the sense of Definition 2.3 replacing \( g \) with \( \tilde{g} \).

Throughout our convergence analysis, we suppose that function \( g \) has an effective domain with nonempty interior and \( \text{dom} h \) is the full space for simplicity of expression. Moreover, we assume that the smoothing functions \( G \) and \( H \) are chosen so that they satisfy properties (P1) and (P2). That is to say, we suppose that the following Assumption (A0) holds in the subsequent analysis:

(A0) \( \text{int dom} \ g \neq \emptyset \), \( \text{dom} h = \mathbb{R}^n \), and \( G \) and \( H \) satisfy (P1) and (P2).

Furthermore, we make the following assumptions:

(A1) The set \( \Omega_0 := \{ x \in \mathbb{R}^n \mid f(x) \leq f(x^1) + \kappa(\mu_1^1 + \mu_2^1) \} \cap \mathcal{F} \) is nonempty and compact, where \( \mathcal{F} \) denotes the feasible region of (11), i.e., \( \mathcal{F} = S \cap (\mathbb{R}^M \times \mathbb{Z}^N) \) and \( \kappa \) is a positive constant prescribed in (P2).
(A2) For any $\mu \geq 0$, $H_\mu$ is strongly convex with a modulus $\tau > 0$, i.e.,
\[
\alpha H_\mu(x) + (1 - \alpha)H_\mu(z) - H_\mu(\alpha x + (1 - \alpha)z) \geq \alpha(1 - \alpha)\tau\|x - z\|^2/2
\]
holds for any $x, z \in \mathbb{R}^n$ and $\alpha \in [0, 1]$.

(A3) $\text{dom } f \cap \mathcal{F} \neq \emptyset$.

Notice the following: Assumption (A1) holds when $f$ is coercive on $\mathcal{F}$, that is, $f(x^\ell) \to \infty$ as $\ell \to \infty$ for any sequence $\{x^\ell\}$ such that $\|x^\ell\| \to \infty$ and $x^\ell \in \mathcal{F}$ for all $\ell \geq 1$. Assumption (A2) can be satisfied by considering an alternative DC-decomposition $f(\cdot) = (g(\cdot) + \tau\|\cdot\|^2/2) - (h(\cdot) + \tau\|\cdot\|^2/2)$ and smoothing the function $h$, although the set of stationary points may change (see Example 1 in Section 4). By (A3), we require that $\mathcal{F}$ has at least one feasible point taking a finite value.

We begin by proving finiteness of the optimal value of the problem of interest.

**Proposition 5.3.** Suppose that Assumptions (A1) and (A3) hold. Then, the optimal value of $\mathcal{I}$, i.e., $\mathcal{I}$, is finite.

**Proof.** We first show that $\inf_{x \in \mathcal{F}} f(x) > -\infty$. Suppose that the contrary is true. Then, there exists some sequence $\{x^\ell\} \subseteq \Omega_0$ such that $\lim_{\ell \to \infty} f(x^\ell) = -\infty$, where $\Omega_0$ is the nonempty and compact set defined in Assumption (A1). Therefore, from Assumption (A1), $\{x^\ell\}$ is bounded and has an accumulation point in $\Omega_0$, say $x^* \in \Omega_0$, and we can assume $\lim_{\ell \to \infty} x^\ell = x^*$ without loss of generality. As $f$ is lower semicontinuous, $f(x^*) \leq \lim_{k \to \infty} f(x^k) = -\infty$. However, this is a contradiction. Therefore, $\inf_{x \in \mathcal{F}} f(x) > -\infty$. We also obtain $\inf_{x \in \mathcal{F}} f(x) < \infty$ in view of Assumption (A3). Hence, we have the desired result. \hfill \Box

**Lemma 5.4.** For any $x \in \mathbb{R}^n$ and $u = (\mu_1, \mu_2) \in \mathbb{R}^2_+$, the following relations hold:
\[
g(x) \leq G_{\mu_1}(x) \leq g(x) + \kappa \mu_1, \quad h(x) \leq H_{\mu_2}(x) \leq h(x) + \kappa \mu_2,
\]
in addition,
\[
f(x) - \kappa \mu_2 \leq f_u(x) \leq f(x) + \kappa \mu_1.
\]

**Proof.** We first prove (17). Choose $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^2_+$ arbitrarily. Let us consider the case where $x \in \text{dom } f = \text{dom } g$. From (P2), we obtain $G_{\mu}(x) \leq G_{\mu_1}(x) \leq G_{\mu}(x) + \kappa(\mu_1 - \mu)$. Using these facts and $\lim_{\mu \to +0} G_{\mu}(x) = g(x)$, and $G_0(x) = g(x)$ by Definition 5.1, we have $g(x) \leq G_{\mu_1}(x) \leq g(x) + \kappa \mu_1$. When $x \in \text{dom } f = \text{dom } g$, by $\text{dom } G_{\mu} = \text{dom } g$ from the definition of $G_{\mu}$, it follows that $g(x) = G_{\mu_1}(x) = \infty$. Thus, $g(x) \leq G_{\mu_1}(x) \leq g(x) + \kappa \mu_1$ is obvious. Similarly, we have $h(x) \leq H_{\mu_2}(x) \leq h(x) + \kappa \mu_2$. Using (17) together with $f = g - h$ and $f_u = G_{\mu_1} - H_{\mu_2}$, we readily obtain (18). \hfill \Box

We next state some technical lemmas for establishing convergence properties of the proposed algorithm. The following lemma assures that the function $\tilde{G}_{\mu}^{\text{cl}}$ is well behaved.
Lemma 5.5. For any $\mu \geq 0$, the following statements hold.

(i) $\tilde{g}^{cl}(x) \leq \tilde{G}^{cl}_\mu(x) \leq \kappa \mu + \tilde{g}^{cl}(x)$ for $x \in \text{dom } g \cap S$ and

(ii) $\text{dom } g \cap S \supseteq \text{dom } \tilde{g}^{cl} = \text{dom } \tilde{G}^{cl}_\mu$.

(iii) Under Assumption (A3), $\text{ri } \text{dom } \tilde{G}^{cl}_\mu \neq \emptyset$.

Proof. (i) From Lemma 5.3, $g(x) \leq G_\mu(x) \leq g(x) + \kappa \mu$ for any $x \in \text{dom } g \cap S$ and $\mu \geq 0$. Thus we have $\text{epi}_{R^M \times Z^N} \tilde{g} \supseteq \text{epi}_{R^M \times Z^N} \tilde{G}_\mu \supseteq \text{epi}_{R^M \times Z^N} g + (0, \kappa \mu)$. Hence, $\text{cl } \text{co } \text{epi}_{R^M \times Z^N} \tilde{g} \supseteq \text{cl } \text{co } \text{epi}_{R^M \times Z^N} \tilde{G}_\mu \supseteq \text{cl } \text{co } \text{epi}_{R^M \times Z^N} g + (0, \kappa \mu)$ follows. We then obtain $\text{epi}_{R^M} \tilde{g}^{cl} \supseteq \text{epi}_{R^M} \tilde{G}^{cl}_\mu \supseteq \text{epi}_{R^M} \tilde{g}^{cl} + (0, \kappa \mu)$ implying $\tilde{g}^{cl}(x) \leq \tilde{G}^{cl}_\mu(x) \leq \kappa \mu + \tilde{g}^{cl}(x)$.

(ii) Choose $x \in \text{dom } \tilde{g}^{cl}$ arbitrarily. By the definition of $\tilde{g}^{cl}$, $x \in S$ follows. In addition, if $x \notin \text{dom } g$, $\tilde{g}^{cl}(x) = g(x) = +\infty$, which contradicts $x \in \text{dom } \tilde{g}^{cl}$ and thus $x \in \text{dom } g$ holds. Therefore, we obtain the first inclusion. The second equality is readily obtained from (i).

(iii) We can easily derive $\text{dom } \tilde{g}^{cl} \neq \emptyset$ from $\text{dom } g \cap F \neq \emptyset$ by (A3). From the nonemptiness and convexity of $\text{dom } \tilde{g}^{cl}$, we obtain $\text{ri } \text{dom } \tilde{g}^{cl} \neq \emptyset$. Then, (ii) yields $\text{ri } \text{dom } \tilde{G}^{cl}_\mu \neq \emptyset$.

\[\square\]

We now give one proposition and one lemma concerning accumulation points of values and subgradients of $\tilde{G}^{cl}_\mu$ and $H_\mu$.

Lemma 5.6. Let $x^* \in \text{dom } \tilde{g}^{cl}$ and consider sequences $\{x^\ell\} \subseteq \text{dom } \tilde{g}^{cl}$ and $\{u^\ell\} \subseteq R^2_+$ such that $x^\ell \to x^*$ and $u^\ell \to 0$ as $\ell$ tends to $\infty$. Then, $\lim_{\ell \to \infty} \tilde{G}^{cl}_\mu(x^\ell) = \tilde{g}^{cl}(x^*)$ holds.

Proof. From Lemma 5.5(ii) and $\{x^\ell\} \subseteq \text{dom } \tilde{g}^{cl}$, $\{x^\ell\} \subseteq \text{dom } g \cap F$ follows. Then, by using Lemma 5.5(i), we have

$$\tilde{g}^{cl}(x^\ell) \leq \tilde{G}^{cl}_\mu(x^\ell) \leq \kappa \mu_1 + \tilde{g}^{cl}(x^\ell).$$

Letting $\ell \to \infty$ here leads to

$$\tilde{g}^{cl}(x^*) \leq \lim_{\ell \to \infty} \tilde{G}^{cl}_\mu(x^\ell) \leq \tilde{g}^{cl}(x^*),$$

where we use continuity of $\tilde{g}^{cl}$ on $\text{dom } \tilde{g}^{cl}$ and $\lim_{\ell \to \infty} x^\ell = x^* \in \text{dom } \tilde{g}^{cl}$. Hence $\lim_{\ell \to \infty} \tilde{G}^{cl}_\mu(x^\ell) = \tilde{g}^{cl}(x^*)$ holds. \[\square\]

Lemma 5.7. Let $x^* \in \text{dom } \tilde{g}^{cl}$, $v^* \in R^n$, and $w^* \in R^n$. Consider sequences $\{x^\ell\} \subseteq \text{dom } \tilde{g}^{cl}$, $\{u^\ell\} \subseteq R^2_+$, $\{v^\ell\} \subseteq R^n$ and $\{w^\ell\} \subseteq R^n$ satisfying the next conditions:

- $v^\ell \in \partial \tilde{G}^{cl}_\mu(x^\ell)$ and $w^\ell \in \partial H_\mu(x^\ell)$ for any $\ell \geq 1$, and

...
• \( \lim_{\ell \to \infty} (x^\ell, u^\ell, v^\ell, w^\ell) = (x^*, 0, v^*, w^*) \).

Then, \( v^* \in \partial \tilde{g}^\cl(x^*) \) and \( w^* \in \partial h(x^*) \).

**Proof.** It suffices to show

\[
\tilde{g}^\cl(x^* + z) - \tilde{g}^\cl(x^*) \geq \langle v^*, z \rangle \quad (19)
\]

\[
h(x^* + z) - h(x^*) \geq \langle w^*, z \rangle \quad (20)
\]

for any \( z \). Since (20) can be shown similarly to (19), we show only (19). Choose \( z \in \mathbb{R}^n \) arbitrarily. If \( x^* + z \notin \text{dom} \tilde{g}^\cl \), (19) is true since \( \tilde{g}^\cl(x^* + z) = \infty \) while \( \tilde{g}^\cl(x^*) < \infty \). Hence, we consider the case where \( x^* + z \in \text{dom} \tilde{g}^\cl \) and let \( z^\ell := x^* + z - x^\ell \) for \( \ell \). Then, \( x^\ell + z^\ell = x^* + z \in \text{dom} \tilde{g}^\cl \) for any \( \ell \). Since \( \tilde{G}_{\mu_1}^\cl \) is convex and \( v^\ell \in \partial \tilde{G}_{\mu_1}^\cl(x^\ell) \), we have

\[
\tilde{G}_{\mu_1}^\cl(x^\ell + z^\ell) - \tilde{G}_{\mu_1}^\cl(x^\ell) \geq \langle v^\ell, z^\ell \rangle \quad (21)
\]

for any \( \ell \). Letting \( \ell \) tend to \( \infty \) in (21) and using Lemma 5.6 imply (19). The proof is complete. \( \square \)

Now, with help of the above lemmas we next prove a proposition concerning well-definedness of the proposed algorithm.

**Proposition 5.8.** Under Assumptions (A1) and (A2), the smoothing SCMIP is well-defined in the sense that \( \partial H_{\mu_2}^h(x^k) \) is nonempty and

\[
\min G_{\mu_1}^h(x) - \langle y^k, x - x^k \rangle \text{ sub. to } x \in \mathcal{F} \quad (22)
\]

has at least one optimum for any \( k \geq 0 \). Furthermore, under Assumption (A3), the optimal value is finite.

**Proof.** The nonemptiness of \( \partial H_{\mu_2}^h(x^k) \) readily follows from \( \text{dom} H_{\mu_2}^h = \text{dom} h = \mathbb{R}^n \) following from (A0). We will show the latter statement. Let \( \phi_k(x) := G_{\mu_1}^h(x) - \langle y^k, x - x^k \rangle \). For any \( \{z^l\} \subseteq \mathcal{F} \) such that \( ||z^l|| \to \infty \), it holds that

\[
\phi_k(z^l) = f_{u^k}(z^l) + H_{\mu_2}^h(z^l) - \langle y^k, z^l - x^k \rangle \to \infty (l \to \infty),
\]

since \( H_{\mu_2}^h \) is strongly convex by Assumption (A2) and \( \{f_{u^k}(z^l)\} \) can be shown to be bounded from below by using Proposition 5.3 and Lemma 5.4. Hence, \( \phi_k \) is coercive and convex in \( \mathcal{F} \), from which we can conclude that (22) has an optimum. It remains to show that the optimal value is finite. Notice that Lemma 5.4 yields dom \( g = \text{dom} G_{\mu_1} \) and thus dom \( f = \text{dom} g = \text{dom} G_{\mu_1} \) holds. Then, from Assumption (A3), there exists some feasible point \( x \in \mathcal{F} \) such that \( G_{\mu_1}(x) < \infty \). Therefore, we have

\[
\min_{x \in \mathcal{F}} G_{\mu_1}(x) - \langle y^k, x - x^k \rangle < \infty. \quad \square
\]

The next proposition concerns the decrement of the objective value per iteration.
Proposition 5.9. Suppose that Assumption (A1) holds. Then, for any \( k \geq 1 \), it holds that

\[
f_{u_{k+1}}(x^{k+1}) \leq f_u(x^k) - \frac{\tau}{2}\|x^{k+1} - x^k\|^2 + \kappa(\mu - \mu_k),
\]

Moreover, the same can be said for the function \( \tilde{f}_{u_{k+1}} \):

\[
\tilde{f}_{u_{k+1}}(x^{k+1}) \leq \tilde{f}_{u_k}(x^k) - \frac{\tau}{2}\|x^{k+1} - x^k\|^2 + \kappa(\mu - \mu_k).
\]

Proof. For simplicity of expression, we consider only the case where \( \{u^k\} \subseteq \mathbb{R}^2_+ \).
Hence, \( \partial H_{\mu_2}(x^k) = \nabla H_2(x^k) \) for all \( k \geq 0 \). The argument below can be directly extended to the case where \( \mu_k = 0 \) or \( \mu_k = 0 \) for any \( k \geq 0 \).

Since \( x^{k+1} \) is an optimum of (13) for \( k \geq 0 \) and \( x^k \) is feasible to (15) for \( k \geq 1 \), we have

\[
G_{\mu_1}(x^{k+1}) - \langle y^k, x^{k+1} - x^k \rangle \leq G_{\mu_1}(x^k) - \langle y^k, x^k - x^k \rangle = G_{\mu_1}(x^k).
\]

On the other hand, we derive \( H_{\mu_2}(x^{k+1}) - H_{\mu_2}(x^k) \geq \langle y^k, x^{k+1} - x^k \rangle + \tau\|x^{k+1} - x^k\|^2/2 \) from \( y^k \in \nabla H_{\mu_2}(x^k) \) and the strong convexity of \( H_{\mu_2} \) with a modulus \( \tau > 0 \) (Assumption (A2)), and thus we get

\[
-H_{\mu_2}(x^{k+1}) + \langle y^k, x^{k+1} - x^k \rangle \leq -H_{\mu_2}(x^k) - \frac{\tau}{2}\|x^{k+1} - x^k\|^2.
\]

Summing up the both sides of (25) and (26) yields

\[
G_{\mu_1}(x^{k+1}) - H_{\mu_2}(x^{k+1}) \leq G_{\mu_1}(x^k) - H_{\mu_2}(x^k) - \frac{\tau}{2}\|x^{k+1} - x^k\|^2.
\]

Combining this in turn with \( G_{\mu_1}(x^{k+1}) \leq G_{\mu_1}(x^{k+1}) \) and \( H_{\mu_2}(x^{k+1}) \leq H_{\mu_2}(x^{k+1}) + \kappa(\mu_k - \mu_k) \) following from (P2) yields (23). By noting that \( \tilde{G}_{\mu_1}(x^k) = G_{\mu_1}(x^k) \) for any \( k \geq 1 \) since \( x^k \) is feasible to (1), we further derive from (23) that

\[
\tilde{G}_{\mu_1}(x^{k+1}) - H_{\mu_2}(x^{k+1}) \leq \tilde{G}_{\mu_1}(x^k) - H_{\mu_2}(x^k) - \frac{\tau}{2}\|x^{k+1} - x^k\|^2 + \kappa(\mu_k - \mu_k).
\]

Therefore, from \( \tilde{f}_{u} = \tilde{G}_{\mu_1} - H_{\mu_2} \), we conclude (24).

We are now ready to give the final convergence property of the smoothing SCMIP.

Theorem 5.10. Suppose that Assumptions (A1)–(A3) hold. Then,

\( i \) the generated sequences \( \{x^k\} \) and \( \{y^k\} \) are bounded, and

\( ii \) \( \lim_{k \to \infty} \|x^{k+1} - x^k\| = 0 \).

Let arbitrary accumulation points of \( \{x^k\} \) and \( \{y^k\} \) be \( x^* := (x_M^*, x_N^*) \in \mathbb{R}^M \times \mathbb{Z}^N \) and \( y^* \in \mathbb{R}^n \), respectively. Then, we have

\( iii \) \( x^* \) is feasible to (1),
(iv) \( \{x_N^k\} \) converges to \( x_N^* \) in finitely many iterations, and

(v) \( x^* \) is a stationary point of DC program (11). In addition, \( y^* \) is a stationary point of the Toland-Singer dual of (11).

Proof. (i) Let \( X_1 := \{(x,u) \in \mathcal{F} \times \mathbb{R}^2 \mid f_u(x) + \kappa \mu_2 \leq f_u(x^1) + \kappa \mu_2, \ 0 \leq u \leq u^1 \} \) and \( X_2 := \{(x,u) \in \mathcal{F} \times \mathbb{R}^2 \mid f(x) \leq f(x^1) + \kappa (\mu_1^+ + \mu_2^+), \ 0 \leq u \leq u^1 \} \). Notice that \( X_2 \) is bounded from Assumption (A1). Choose \((x,u) \in X_1 \) arbitrarily. Then, from Lemma 5.4 we have

\[
f(x) \leq f_u(x) + \kappa \mu_2 \leq f_u(x^1) + \kappa \mu_2 \leq f(x^1) + \kappa (\mu_1^+ + \mu_2^+)
\]

where the first inequality follows from Lemma 5.4 [18], the second one from \((x,u) \in X_1 \), and the third by using [18] again with \( x = x^1 \) and \( u = u^1 \). This together with \( 0 \leq u \leq u^1 \) implies \((x,u) \in X_2 \). Therefore, we have \( X_1 \subseteq X_2 \), and thus \( X_1 \) is bounded. Since \( \{(x^k,u^k)\}_{k \geq 1} \subseteq X_1 \) from Proposition 5.9 we then conclude that \( \{x^k\} \) is bounded. To prove the boundedness of \( \{y^k\} \), suppose that \( \{y^k\} \) is unbounded for contradiction. Let \((\bar{x},\bar{y})\) be an arbitrary accumulation point of \( \{(x^k,y^k/\|y^k\|)\} \). We may assume that \( (x^k, y^k/\|y^k\|) \to (\bar{x}, \bar{y}) \) as \( k \to \infty \) without loss of generality. From \( y^k \in \partial H_{\mu_2^k}(x^k) \), we have

\[
H_{\mu_2^k}(x^k + \bar{y}) - H_{\mu_2^k}(x^k) \geq \langle \bar{y}, y^k \rangle.
\]  

Then, by Definition 5.13 with \( \Phi = H \), \( \lim_{k \to \infty} H_{\mu_2^k}(x^k + \bar{y}) = h(\bar{x} + \bar{y}) < \infty \) and \( \lim_{k \to \infty} H_{\mu_2^k}(x^k) = h(\bar{x}) < \infty \), which together with dividing the both sides of (27) by \( \|y^k\| \) and driving \( k \to \infty \) entail \( 0 \geq \|\bar{y}\|^2 \), which contradicts \( \|\bar{y}\| = 1 \). We thus see that \( \{y^k\} \) is bounded.

(ii) By Proposition 5.9 and Lemma 5.4 with \( x = x^{k+1} \) and \( u = u^{k+1} \), it holds that

\[
f(x^{k+1}) \leq f(x^{k+1} + \kappa \mu_2) \leq f(x^k) + \kappa \mu_2 - \frac{\tau}{2} \|x^{k+1} - x^k\|^2 < \infty
\]  

for any \( k \geq 1 \), where the last inequality holds since \( f(x^k) \leq f(x^k + \kappa \mu_1) \) follows from Lemma 5.4 and \( x^k \in \text{dom} G_{\mu_1} = \text{dom} f \ (k \geq 1) \) from Proposition 5.8 under Assumption (A3). Then, by noting that \( \{f(x^k)\} \) is bounded from below according to Proposition 5.3 and the fact that \( x^k \in \mathcal{F} \) for \( k \geq 1 \), (28) implies that \( \{f(x^k) + \kappa \mu_2\} \) is a bounded monotone nonincreasing sequence, and hence a convergent sequence. Therefore, \( \lim_{k \to \infty} \|x^{k+1} - x^k\| = 0 \) follows from the second and third inequalities of (28).

(iii) By noting \( \{x^k\}_{k \geq 1} \subseteq \mathcal{F} = S \cap (\mathbb{R}^M \times \mathbb{Z}^N) \) and the closedness of \( \mathcal{F} \), \( x^* \in \mathcal{F} \) is easily derived.

(iv) For contradiction, suppose that \( \{x_N^k\} \) does not converge in finitely many iterations. Then, there exist infinitely many \( k \) such that \( \|x_N^{k+1} - x_N^k\| \geq 1 \), since \( x_N^k \in \mathbb{Z}^N \) for any \( k \geq 1 \). However, it contradicts the fact of \( \lim_{k \to \infty} \|x^k - x_N^k\| = 0 \).
We first prove that \( x^* \in \text{dom } \tilde{g}^{\text{cl}} \). Supposing to the contrary, we have \( \tilde{g}^{\text{cl}}(x^*) = \infty \). Then, from the definition of \( \tilde{g}^{\text{cl}} \) and feasibility of \( x^* \), we obtain \( g(x^*) = \tilde{g}(x^*) = \infty \), and hence \( f(x^*) = \infty \). On the other hand, (28) and (13) yield that \( f(x^*) = \lim_{k \to \infty} f_{u,k}(x^k) + \kappa \mu_2^k < \infty \). This is a contradiction.

We next show the main claim. As \( \text{ri dom } \tilde{G}^{\text{cl}}_{\mu_1^k} \neq \emptyset \) from Lemma 5.5, we obtain
\[
\partial \left( \tilde{G}^{\text{cl}}_{\mu_1^k}(x) - \langle y^k, x - x^k \rangle \right) = \partial \tilde{G}^{\text{cl}}_{\mu_1^k}(x) - y^k.
\]
By using this equality and the fact that \( x^{k+1} \) is an optimum of the convex program (16)
\[
\partial H_{\mu_2^k}(x^k) \ni y^k, \quad \partial \tilde{G}^{\text{cl}}_{\mu_1^k}(x^{k+1}) - y^k \ni 0.
\]
Without loss of generality (if necessary by taking a subsequence), we can assume that \( (x^k, y^k) \) converges to \( (x^*, y^*) \) as \( k \) tends to \( \infty \). Since \( \lim_{k \to \infty} x^{k+1} = x^* \) is derived from (ii) and \( x^* \in \text{dom } \tilde{g}^{\text{cl}} \) holds from the first-half argument, Lemma 5.7 and (29) yield that \( \partial \tilde{g}^{\text{cl}}(x^*) \ni y^* \) and \( y^* \in \partial h(x^*) \), which means \( y^* \in \partial \tilde{g}^{\text{cl}}(x^*) \cap \partial h(x^*) \). Therefore, \( x^* \in \partial (\tilde{g}^{\text{cl}})(y^*) \cap \partial h^*(y^*) \). This completes the proof.

\[ \square \]

6 Concluding remarks

In this paper, we have considered mixed integer programs having DC objective functions and closed convex constraints. For these problems, we have extended the result of Maehara, Marumo, and Murota concerning continuous relaxations of discrete DC programs and obtained a continuous DC program whose optimal value is exactly equal to the original one. We have also proposed two algorithms to solve the obtained relaxed problem. The first is a fundamental algorithm based on the DCA, which is a well-known algorithm for continuous DC programs. In the second, we incorporate a smoothing method into the first, so that we can handle nonsmooth functions efficiently. For both methods, we proved that the generated sequence converges to a stationary point under some mild assumptions.

Our contribution can be summarized as follows:

- We have proposed a new framework for solving mixed integer DC programs. These are a wide class of problems containing many mixed integer nonlinear programs which are notorious as being extremely difficult. Although our method involves the computationally costly routine of repeatedly solving convex MIPs, it still has significant merit, since it provides a practical way of dealing with these tough problems.

- We have theoretically proved convergence of generated sequences, thus the solutions provided by our algorithms are stationary points which have good chances of being the global optimum.
We conclude this paper by mentioning that while our method does not obtain polynomial complexity, there may be some specific problems for which it is tractable. For example, Maehara, Marumo and Murota [23] showed that their DCA-based algorithm can efficiently solve the degree-concentrated spanning tree problem. The search for such problems is a possible direction for future work.

References

[1] T. Achterberg, SCIP: solving constraint integer programs, Math. Prog. Comp. 1 (2009) 1–41.

[2] P. Bonami, L. T. Biegler, A. R. Conn, G. Cornéjols, I. E. Grossmann, C. D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, A. Wächter, An algorithmic framework for convex mixed integer nonlinear programs, Discrete Optim. 53 (2009) 78–87.

[3] P. Bonami, M. Kilinç, J. Linderoth, Algorithms and software for mixed integer nonlinear programs, in Mixed integer nonlinear programming, The IMA volumes in mathematics and its applications (J. Lee and S. Leyffer eds.) 154 (2012).

[4] P. Briest, P. Krysta, Single-minded unlimited supplou pricing on sparse instances, Proc. 17th Annu ACM-SIAM Symp Discrete Algorithms, (2006) 1093–1102.

[5] S. Burer, A. N. Letchford, Non-convex mixed-integer nonlinear programming: A survey, Surv. Oper. Res. Manag. Sci. 17 (2012) 97–106.

[6] C. Chen, O. L. Mangasarian, A class of smoothing functions for nonlinear and mixed complementarity problems, Comput. Optim. Appl. 5 (1996) 97–138.

[7] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, Math. Prog. Series B 134 (2012) 71–99.

[8] M. A. Duran, I. E. Grossmann, An outer-approximation algorithm for a class of mixed-integer nonlinear programs, Math. Prog. Series A 36 (1986) 307–339.

[9] O. Exler, T. Lehmann, K. Schittkowski, A trust region SQP algorithm for mixed integer nonlinear programming, Optim. Lett. 1 (2007) 269–280.

[10] O. Exler, T. Lehmann, K. Schittkowski, A comparative study of SQP-type algorithms for nonlinear and nonconvex mixed-integer optimization, Math. Prog. Comp. 4 (2012) 383–412.

[11] R. Fletcher, S. Leyffer, Solving mixed integer nonlinear programs by outer approximation, Math. Prog. Series A 66 (1994) 327–349.
[12] C. A. Floudas, Deterministic global optimization: Theory, algorithms and applications, Kluwer Academic Publishers, Dordrecht (1999).

[13] A. M. Geoffrion, Generalized Benders decomposition, J. Optim. Theory Appl. 4 (1972) 237–260.

[14] O. K. Gupta, A. Ravindran, Branch and bound experiments in convex non-linear integer programming, Manage. Sci. 31 (1985) 1533–1546.

[15] Gurobi optimizer reference manual, Optimization, Gurobi and others, URL:http://www.gurobi.com 2 (2012) 1–3.

[16] R. Horst, N. V. Thoai, DC programming: overview, J. Optim. Theory Appl. 103 (1999) 1–43.

[17] R. Horst, P. M. Pardalos, N. V. Thoai, Introduction to Global Optimization 2nd Edition, Springer Science+Business Media, B.V., 2000.

[18] ILOG CPLEX V12.1, User’s Manual for CPLEX, International Business Machines Corporation, 46 (2009) 157.

[19] S. Leyffer, Integrating SQP and branch-and-bound for mixed integer nonlinear programming, Comput. Optim. Appl. 18 (2001), 295–309.

[20] H. A. Le Thi, H. M. Le, T. Pham Dinh, Portfolio selection under downside risk measures and cardinality constraints based on DC programming and DCA, Comput. Manag. Sci. 6 (2009) 459–475.

[21] H. A. Le Thi, H. M. Le, T. Pham Dinh, Feature selection in machine learning: an exact penalty approach using a Difference of Convex function algorithm, Mach. Learn. 101 (2015) 163–186.

[22] T. Maehara, K. Murota, A framework of discrete DC programming by discrete convex analysis, Math. Prog. Series A 152 (2015) 435–466.

[23] T. Maehara, N. Marumo, K. Murota, Continuous relaxation for discrete DC programming, in Modelling, computation and optimization in information systems and management sciences, AISC 359 (2015) 181–190.

[24] Y. S. Niu, T. Pham Dinh, A DC programming approach for mixed-integer linear programs, in Modelling, computation and optimization in information systems and management sciences, CCIS 14 (2008) 244–253.

[25] T. Pham Dinh, H. A. Le Thi, Convex analysis approach to D.C. programming: theory, algorithms and applications, Acta Mathematica Vietnamica 22 (1997) 289–355.

[26] T. Pham Dinh, H. A. Le Thi, Recent advances in DC programming and DCA, in Lecture Notes in Computer Science 8342 (2014) 1–37.
[27] H. D. Sherali, H. Wang, Global optimization of nonconvex factorable programming problems, Math. Prog. Series A 89 (2001) 459–478.

[28] M. Tawarmalani, N. V. Sahinidis, Global optimization of mixed–integer nonlinear programs: a theoretical and computational study, Math. Prog. Series A 99 (2004) 563–591.

[29] M. Thiao, T. Pham Dingh, H. A. Le Thi, A DC programming approach for space eigenvalue problem, Proc. ICML2010 (2010), 1063–1070.

[30] T. Westerlund, F. Pettersson, An extended cutting plane method for solving convex MINLP problems, Comput. Chemical Engrg. 19 (1995) 131–136.